Pricing and hedging derivative securities with neural networks and a homogeneity hint[☆]

René Garcia[a],*, Ramazan Gençay[b]

[a]*Département de Sciences Économiques et C.R.D.E., Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, Qué., Canada H3C 3J7*
[b]*Department of Economics, University of Windsor, Windsor Ont., Canada N9B 3P4*

Abstract

We estimate a generalized option pricing formula that has a *functional shape* similar to the usual Black–Scholes formula by a feedforward neural network model. This functional shape is obtained when the option pricing function is homogeneous of degree one with respect to the underlying asset price (S_t) and the strike price (K). We show that pricing accuracy gains can be made by exploiting this generalized Black–Scholes shape. Instead of setting up a learning network mapping the ratio S_t/K and the time to maturity (τ) directly into the derivative price, we break down the pricing function into two parts, one controlled by the ratio S_t/K, the other one by a function of time to maturity. The results indicate that the homogeneity hint always reduces the out-of-sample mean squared prediction error compared with a feedforward neural network with no hint. Both feedforward network models, with and without the hint, provide similar delta-hedging errors that are small relative to the hedging performance of the Black–Scholes model.

However, the model with hint produces a more stable hedging performance. © 2000 Elsevier Science S.A. All rights reserved.

1. Introduction

In a recent paper, Hutchinson et al. (1994) demonstrated that learning networks can be used successfully to estimate a pricing formula for options, with good out-of-sample pricing and delta-hedging performance. This nonparametric pricing method has the distinct advantage of not relying on specific assumptions about the underlying asset price dynamics and is therefore robust to specification errors that might affect adversely parametric models. Hutchinson et al. (1994) assume that their option pricing network formula is homogeneous of degree one in the underlying stock price and the strike price which enables them to use a smaller number of inputs in learning the nonparametric pricing function. This parsimony is an advantage since the rate of convergence of nonparametric estimators slows down considerably as the number of inputs increases. Broadie et al. (1996a), who also use nonparametric methods to estimate an option pricing function, invoke the nonstationarity of option and stock prices to justify such a homogeneity property.

This homogeneity assumption is not consistent with any asset price dynamics. Merton (1973) shows that serial independence of asset returns for the data generating process is a sufficient condition for homogeneity. In a non-arbitrage context, Garcia and Renault (1995) establish that conditional independence under the pricing probability measure between future returns and the current price is a necessary and sufficient condition for homogeneity of the option pricing function. Several processes obey these objective or risk-neutral distributional assumptions and lead therefore to homogeneous option pricing formulas.[1] These formulas can be characterized as generalizations of the Black–Scholes formula in the sense that the normal distribution function is replaced by another distribution function in an otherwise similarly shaped formula which stems from the convexity of the terminal payoff. Garcia and Renault (1995) also provide a dynamic asset pricing equilibrium model in a general stochastic framework that leads to a homogeneous option pricing

---

[1] For example, jump processes (see Hull, 1993, p. 454) or stable distributions (McCulloch, 1996).

formula which keeps the *main functional shape* of the usual Black–Scholes formula and nests most of the usual parametric option pricing formulas.[2]

In this paper, we show that pricing accuracy gains can be made by exploiting the implications of this homogeneity property in terms of functional shape. Instead of setting up a learning network mapping the stock price to strike price ratio (S_t/K) and the time to maturity (τ) directly into the derivative price, we break down the pricing function into two parts, one controlled by the ratio S_t/K, the other one by a function of time to maturity. In each part, a learning network is fit with S_t/K and τ as inputs. We just mentioned that this separation into two blocks is consistent not only with the Black–Scholes model, but with an array of other models which keep the homogeneity property of the option pricing function. It is this homogeneity restriction that we call a hint. In general, hints based on additional prior information about the properties of the unknown function to be learned guide the learning process. Because hints impose additional constraints on the set of allowable solutions to which the learning process may converge, they may tend to worsen the in-sample performance by excluding some solutions that might otherwise fit the data better. This constraint clearly helps to avoid overfitting in the learning algorithms. The main purpose of using hints is to improve the out-of-sample performance of the learning algorithms.

To assess the potential gains that can be made by using the homogeneity hint in setting up the nonparametric model, we simulate option prices that obey the Black–Scholes formula. In this experimental setting, we obtain out-of-sample pricing accuracy gains of about 25% in average. To assess the empirical relevance of this additional structure consistent with homogeneity, we estimate pricing functions for European call options on the S&P 500 index for various sampling periods between 1987 and 1994. The homogeneity hint always reduces the out-of-sample mean squared prediction error compared with a feedforward neural network with no hint. The feedforward network models provide smaller delta-hedging errors relative to the Black–Scholes model. Between the feedforward network models, the models with hint provide more stable average delta hedging errors relative to the networks without the homogeneity hint.

Recently, a number of papers have used nonparametric methods to price options. Ghysels et al. (1997) provide a survey of this literature. Two papers appeal to financial theory to complement a strictly nonparametric approach. Gouriéroux et al. (1994) apply a Kernel M-estimator methodology to the option

---

[2] They obtain as special cases the formula derived by Amin and Ng (1993) and a fortiori all the other pricing formulas that were nested in the latter: of course the Black–Scholes formula, but also the Hull and White (1987) and Bailey and Stulz (1989) stochastic volatility option pricing formulas and the Merton (1973), Turnbull and Milne (1991), and Amin and Jarrow (1992) stochastic interest rate option pricing formulas for equity options.

pricing problem by extending the Black–Scholes formulation.[3] In doing so, they recognize that the Black–Scholes formula is not strictly valid, but that its shape can still be useful to recover a pricing formula more in line with observed data. More precisely, they keep the Black–Scholes functional shape but they make the volatility parameter which is assumed fixed in the Black–Scholes model a function of some observable state variables such as the S_t/K ratio. Apart from the fact that we use a feedforward neural network estimation technique instead of a kernel estimator, our approach can be seen as a generalization of their approach since we dispense with the log-normality assumption of the asset underlying the Black–Scholes model. Aït-Sahalia and Lo (1996) use kernel estimation techniques for the option pricing function. They also point out that several of the partial derivatives of the option pricing function are of special interest. Apart from the well-known delta of the option, i.e. the first derivative of the option pricing formula with respect to the stock price, it is possible to recover the state price density (SPD) through the second derivative of the option pricing function with respect to the strike price.[4] In the context of complete markets, this state price density is very useful since it provides an arbitrage-free method of pricing complex or less liquid options given observed prices on liquid basic options.

By contrast, the two latter papers underline an important issue. In Gouriéroux et al. (1994), the parameters of the volatility function correcting the Black and Scholes model are estimated according to a specifically chosen objective function based on an empirical criterion, what they call objective-driven inference. In Aït-Sahalia and Lo (1996), although the ultimate objective might be to extract the risk neutral density, it is the pricing function that is estimated through kernel methods. Although kernel methods or feedforward networks can estimate consistently the derivatives of a function, it might be important to target directly the statistical or financial criterion of interest (see Bengio, 1996). If the goal is to hedge a portfolio, the training or the validation of the learning network should be done according to the hedging criterion which involves the first derivative of the function. As we will see, the architectures of the networks selected for pricing and hedging will be very different. Choosing the best pricing model for hedging purposes could lead to important financial losses, especially in networks without the homogeneity hint.

To show the usefulness of a learning network in pricing and hedging options, Hutchinson et al. (1994) looked at several techniques for modelling nonlinear

---

[3] Aït-Sahalia and Lo (1996) also use the same semiparametric approach, along with their purely nonparametric approach.

[4] Clément et al. (1993) and Patilea and Renault (1995) estimate an equivalent martingale measure by a nonparametric Bayesian method. Jondeau and Rockinger (1997) survey and compare several methods to extract risk neutral densities with an applicaton to exchange rate options.

statistical relationships nonparametrically: radial basis functions, projection pursuit regression, and multilayer perceptrons. We use only the last of these techniques to illustrate the accuracy gains that could be made by exploiting the homogeneity property of the formula. Whether similar gains can be made with the other techniques or any other nonparametric approach remains to be investigated.

Section 2 discusses the nonparametric approach to option pricing and the restrictions implied by the often assumed homogeneity of degree one of the pricing function in the underlying stock price and the strike price. Section 3 presents the feedforward neural networks used for estimating the option pricing function. In Section 4, we report the results of a Monte-Carlo experiment aimed at assessing the pricing and hedging accuracy gains provided by the homogeneity restriction. Section 5 mirrors Section 4 with actual price data on options written on the S&P 500 index. Section 6 concludes.

2. Nonparametric option pricing with homogeneity

A natural nonparametric function for pricing a European call option on a non-dividend paying asset will relate the price of the option to the set of variables which characterize the option, i.e. the price of the underlying asset S_t, the strike price K, and the time to maturity τ. Therefore, the option pricing function can be written as:

$$C_t = f(S_t, K, \tau). \tag{1}$$

This approach is followed by Hutchinson et al. (1994). The function will also be valid to learn prices generated by a Black–Scholes model as the interest rate and volatility parameters present in the formula are constant and cannot be identified by a nonparametric estimator of the function f. It is generally more difficult to estimate nonparametrically such a function when the number of input variables is large. To reduce the number of inputs, Hutchinson et al. (1994) divide the function and its arguments by K and write the pricing function as follows:

$$\frac{C_t}{K} = f\left(\frac{S_t}{K}, 1, \tau\right). \tag{2}$$

This form assumes the homogeneity of degree one in the asset price and the strike price of the pricing function f. Another technical reason for dividing by the strike price is that the process S_t is nonstationary while the variable $\frac{S_t}{K}$ is stationary as strike prices bracket the underlying asset price process. This point is emphasized in Ghysels et al. (1997). The crucial question is to determine to

what extent this homogeneity property is restrictive for the nonparametric learning of the option pricing function. From Merton (1973), we know that the call pricing function is homogeneous of degree one in the asset price and the strike price when the unconditional distribution of returns is independent of the level of the asset price. In Garcia and Renault (1995), Proposition 2 establishes that a necessary and sufficient condition for homogeneity is the conditional independence (under the pricing probability measure) between future returns and the current price, given the currently available information other than the history of the underlying asset price. This property must be understood as a noncausality relationship in the Granger sense from the current price to future returns (for a given informational setting) and not as an independence property. This characterization of homogeneity is more general than the sufficient condition proposed by Merton (1973), not only since the independence requirement is replaced by a more specific noncausality assumption, but also since it is stated in terms of the pricing probability measure rather than the data generating process. In such a setting, risk premiums may depend on the level of the asset price S_t. Very general processes are also admissible for the underlying asset, such as for example a stochastic volatility model, except that the volatility function cannot be a function of the asset price level as in implied tree models (see Rubinstein (1994)).

Garcia and Renault (1995) further propose an equilibrium model that ensures the homogeneity property.[5] Given a conditional (on state variables) log-normality assumption about the fundamentals of the economy, they derive an extended Black–Scholes option pricing formula. They stress that such an additional assumption is not really restrictive once the assumptions required for homogeneous option pricing are maintained since log-normality follows from a standard central limit argument. Therefore, the Black–Scholes shape of the option pricing formula will be robust when one remains true to homogeneity. Their general pricing formula for European call options is given by

$$\frac{C_t}{K} = E_t\left\{\frac{S_t}{K}Q_{XY}(t,\ T)\Phi(d_1) - \tilde{B}(t,\ T)\Phi(d_2)\right\}, \tag{3}$$

where $Q_{XY}(t,\ T)$ is a function of preference parameters and of conditional moments of future growth rates of consumption (X) and dividends (Y), $\tilde{B}(t,\ T)$ a stochastic discount factor also a function of preference parameters and future growth rates of consumption (X), $\Phi(.)$ the cumulative normal distribution

---

[5] Corollary 2.1 in Broadie et al. (1996b) also states that such a homogeneity property holds in a fairly general stochastic volatility model. One restriction is that the drift and diffusion functions of the stochastic volatility process may not depend on the asset price itself.

function and:

$$d_1 = \frac{\ln\left[\dfrac{S_t Q_{XY}(t,\,T)}{K\tilde{B}(t,\,T)}\right]}{(\sum_{\tau=t+1}^{T}\sigma_{Y\tau}^2)^{1/2}} + \frac{1}{2}\left(\sum_{\tau=t+1}^{T}\sigma_{Y\tau}^2\right)^{1/2},$$

$$d_2 = d_1 - \left(\sum_{\tau=t+1}^{T}\sigma_{Y\tau}^2\right)^{1/2}.$$

The expression inside the expectation in (3) keeps the Black–Scholes functional shape. For a nonparametric characterization of the option pricing function, we therefore use a generalized Black–Scholes formula:

$$\frac{C_t}{K} = \frac{S_t}{K}f_1\!\left(\frac{S_t}{K},\,\tau\right) - b(\tau)f_2\!\left(\frac{S_t}{K},\,\tau\right). \tag{4}$$

Theoretical restrictions stemming from the absence of arbitrage or from equilibrium constrain the functions f_1 and f_2 to approximate the same function (for example the normal distribution function in the Black–Scholes formula), with possibly different arguments or different signs for the same arguments. We account for these theoretical restrictions by constraining the neural network structure to be the same for f_1 and f_2. Since we limit the arguments of the functions to $\frac{S_t}{K}$ and τ, we let the signs of the inputs unconstrained within each function. The function $b(\tau)$ is a general function of the maturity of the option.

3. Learning networks

Hutchinson et al. (1994) used learning networks based on three techniques: radial basis functions, multilayer perceptrons, and projection pursuit regression. Our ultimate goal is to assess the usefulness of the homogeneity hint for learning the option pricing function. We select the multilayer perceptron or feedforward neural network technique for this purpose, hoping that similar results could be obtained with the other techniques.

3.1. Feedforward neural networks and hints

Let a typical regression function be written as $f(x,\,\theta)$, where x stands for the *explanatory variables*, θ is a vector of parameters and the function f determines how x and θ interact. This representation is identical to the output function of a feedforward network such that the network inputs are interpreted as the explanatory variables and the weights in the network are interpreted as the parameters, θ. In a typical feedforward network, the input units send signals x_j across weighted connections to intermediate or *hidden* units. Any given

hidden unit j receives the sum of all the p weighted inputs, $\gamma_{j0} + \sum_{i=1}^{p} \gamma_{ji} x_i$. The first term γ_{j0} is an intercept or a *bias* term. The weights γ_{ji} are the weights to the jth hidden unit from the ith input. The hidden unit j outputs a signal $h_j = G(\gamma_{j0} + \sum_{i=1}^{p} \gamma_{ji} x_i)$ where the *activation function G* is

$$G(x) = \frac{1}{1 + e^{-\alpha x}},$$

a logistic function which has the property of being a sigmoidal[6] function. The signals from the hidden units $j = 1, \ldots, d$ are sent to the output unit across weighted connections in a manner similar to what happens between the input and hidden layers. The output unit receives the sum of the weighted hidden units, $\beta_0 + \sum_{j=1}^{d} \beta_j h_j$. If the expression for h_j is substituted into the latter expression, it yields the output of a single layer feedforward network

$$f(x, \theta) = \Phi\left(\beta_0 + \sum_{j=1}^{d} \beta_j G\left(\gamma_{j0} + \sum_{i=1}^{p} \gamma_{ji} x_i\right)\right) \tag{5}$$

as a function of inputs and weights.

Many authors have investigated the universal approximation properties of neural networks (Gallant and White, 1988, 1992; Cybenko, 1989; Funahashi, 1989; Hornik et al., 1989, 1990). Using a wide variety of proof strategies, all have demonstrated that under general regularity conditions, a sufficiently complex single hidden-layer feedforward network can approximate a large class of functions and their derivatives to any desired degree of accuracy where the complexity of a single hidden layer feedforward network is measured by the number of hidden units in the hidden layer. One of the requirements for this universal approximation property is that the activation function has to be a sigmoidal such as the logistic function presented above.[7] We estimate θ by nonlinear least squares. Gallant and White (1992) show that the least squares estimates are consistent in the Sobolev norm, provided that the number of hidden units increases with the size of the data set.

All learning network methods share the same fundamental premise of learning from input and output pairs. For a method to learn an unknown function from data, it must be able to make generalizations to the out-of-sample setting from the limited input–output pairs upon which it is trained. In general, a learning network technique knows nothing about the unknown function it is trying to learn, except what is provided in the sample of input–output pairs. If

---

[6] *G* is a sigmodial function if $G : R \to [0,1]$, $G(a) \to 0$ as $a \to -\infty$, $G(a) \to 1$ as $a \to \infty$ and G is monotonic.

[7] For an excellent survey of the feedforward and recurrent network models, the reader may refer to Kuan and White (1994) and White (1992).

the provided pairs contain irrelevant information or a substantial amount of noise, the fit will be poor. Accordingly, the method will provide poor out-of-sample generalizations.

A method which provides additional information to the learning algorithm is the method of *hints*. A method with hints describes a situation where, in addition to the set of input–output pairs of an unknown function, there is additional prior information about the properties of the unknown function which is provided to the learning algorithm. In general, hints provide auxiliary information about the unknown function which can be used to guide the learning process. The idea of using auxiliary information about the target function to help the learning process is clearly a basic one, and has been used in the literature under different names such as hints, prior knowledge and explicit rules. Furthermore, a model with hint provides additional guidance to the learning algorithm in the presence of noisy data and a limited number of observations.

There are different types of hints common to different applications. Invariance hints of Duda and Hart (1973), Hinton (1987), Hu (1962) and Minsky and Papert (1988) are the most common types of hints in pattern recognition applications. An invariance hint asserts that the target function is invariant under certain transformations of the input. Monotonicity hints, as in Abu-Mostafa (1993), are common in applications such as medical diagnosis and credit rating where the target function is assumed to be monotonic in certain variables.[8] Symmetry hints are commonly used in foreign exchange predictions by technical analysts. Abu-Mostafa (1994, 1995) indicate that appropriately placed restrictions may lead to improved out-of-sample generalizations.

3.2. The estimated networks

For the architecture of our networks, with and without the homogeneity hint, we follow the general choice of the identity function for Φ and the logistic function for G. Therefore, we will estimate the following two models, respectively for the model without hint (6) and with hint (7):

$$f^{NN}(S_t/K, \tau; \theta) = \beta_0 + \sum_{j=1}^{d} \beta_j \frac{1}{1 + \exp(-\gamma_{j0} - \gamma_{j1}(S_t/K) - \gamma_{j2}\tau))}, \qquad (6)$$

$$f^{\mathbf{WH}}(S_t/K, \tau; \theta) = \beta_0 + \frac{S_t}{K}\left(\sum_{j=1}^{d} \beta_j^1 \frac{1}{1 + \exp(-\gamma_{j0}^1 - \gamma_{j1}^1(S_t/K) - \gamma_{j2}^1\tau))} \right)$$

$$- e^{-\alpha\tau}\left(\sum_{j=1}^{d} \beta_j^2 \frac{1}{1 + \exp(-\gamma_{j0}^2 - \gamma_{j1}^2(S_t/K) - \gamma_{j2}^2\tau))} \right). \quad (7)$$

---

[8] The methodology proposed by Abu-Mostafa (1993) is to create virtual examples from the observed data and add them to the training set as hints to improve the out-of-sample predictability. A similar methodology could be used to impose homogeneity.

From a statistical point of view, one drawback of the feedforward neural network technique is the virtual absence of inferential procedures to determine the best model specification. There are a number of information theoretic criteria such as the Schwarz Information Criteria (SIC) or the Akaike Information Criteria (AIC) which could be used for this purpose, but they support the choice of feedforward network models which do not generalize well. Swanson and White (1995) report that the SIC fails to select sufficiently parsimonious models in terms of being a reliable guide to the out-of-sample performance. Cross-validation based methods are also available and require heavy computational time to determine the network complexity.

We select the complexity of the networks based on their performance in an out-of-sample validation period. From equations (6) and (7), it can be seen that the networks for the model with hint will always have about twice as many parameters as the networks without the hint for a given number of hidden units. To compare fairly the performance of the two networks, we adopt a three-step strategy. First, we estimate networks with 1–9 hidden units[9] for the regular neural networks and 1–5 hidden units for the networks with hint over half of the data points for a particular sample, the training period. Next, we choose the network in each family that gives the best mean square prediction error (MSPE) over half of the remaining data points in the sample, called the validation period. Finally, we assess the prediction performance (MSPE) of the best model chosen in the previous step for the models with and without the homogeneity hint over the last quarter of data, the prediction period.

To gauge the improvement achieved with the homogeneity hint in a model kept as simple as possible, we limit our investigation to two inputs, namely $\frac{S_t}{K}$ and τ. It also makes our results comparable to the study of Hutchinson et al. (1994) who used the same inputs. Of course, the introduction of various estimates of the volatility of the underlying assets could further improve results but this avenue will not be pursued here. The interest rate is estimated by the parameter α in the model with hint (7) but is absent from (6).[10] Again, introducing the observed interest rate could add useful information.

Hutchinson et al. (1994) also evaluated their pricing model in terms of hedging performance. We compare the two families of neural networks, with and without the hint, according to this criterion. We proceed in the same way as for the MSPE criterion. We choose the best model on an intermediate validation period and evaluate the forecasted average hedging error over a final prediction period. We also present the percentage of options, in this final period, for which the hedging error is less than the Black–Scholes hedging error.

---

[9] We also experimented with feedforward network models with 10 hidden units, but they were never selected in the validation period.

[10] We impose the positivity of the corresponding parameter in the networks with hint.

4. Learning with a homogeneity hint: A Monte Carlo experiment

To run our Monte Carlo experiment, we adopt a Black–Scholes framework. The price of the underlying asset on which the option is written follows a geometric Brownian motion:

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t.$$

We adopt the same setting as Hutchinson et al. (1994) with S_0, the initial price of the stock set equal to \$50, μ equal to 10% and a yearly volatility σ to 20%. To make our experiment comparable to our performance assessment strategy with actual daily data, we simulate a year of daily-return data by drawing 253 random normal variates Z_t with mean $\mu/253$ and standard deviation $\sigma/\sqrt{253}$. The price series $\{S_t, t = 1, \ldots, 253\}$ is obtained as follows:

$$S_t = S_0 \exp\left(\sum_{i=1}^{t} Z_i \right).$$

Given this path of daily stock prices, we create stock options according to the rules of the Chicago Board Options Exchange (CBOE), summarized in the Appendix, and generate 3333 data points. As described in the previous section, we train the two families of networks over the data points that correspond to the first six months of the sample, which amounts to 1612 points in our experiment. Given the estimated networks, we predict the option prices over the next three months, a period which represents 916 data points. Based on these predicted prices, we compute the MSPE over this validation period in order to select in each family the network that delivers the lowest MSPE. We have now selected two competing models, one for each family of neural networks, with and without the homogeneity hint. We compute for each model the MSPE over the last part of the sample which counts 805 data points.[11] Since the estimation result of the networks depends on the random seed from which initial values are drawn for the parameters, we run this procedure five times.[12] Table 1 presents the average MSPE over the prediction period for each family of networks, along with the

---

[11] The out-of-sample forecast is therefore done over these 805 data points. Notice that the same estimated parameters are kept for the whole period. Of course, in a real-time forecast exercise, the networks should be reestimated as new information becomes available.

[12] For each experiment and each network (1–5 units for the networks with the homogeneity hint and 1–9 units for the networks without the hint), we draw 200 sets of parameters starting from different seeds and select the set with the smallest mean squared error as the starting values for our nonlinear least square estimation over the first half of the sample for each year. The number of units reported in the table corresponds to the networks that obtained the lowest MSPE over the validation period, averaged over the five runs.

Table 1
Out-of-sample mean square prediction errors on simulated call option Black–Scholes prices
(Total sample: 1612, Validation sample: 916, Prediction sample: 805)

Statistics	MSPE with hint	MSPE no hint	Ratio	MSPE linear
$\bar{x}$	0.1708 (3)	0.2321 (6)	0.74	37.35
σ	0.0623	0.0360	1.73	
DM	20.57			

Notes: This table presents the out-of-sample mean square prediction error (MSPE) performance of a neural network with a homogeneity hint, a regular feedforward network with no hint, and of a linear model for call option prices generated with the Black–Scholes formula. The table reports the average ($\bar{x}$) of the five MSPEs corresponding to five networks estimated from different seeds. The average number of hidden units of the five runs are reported between parentheses next to the average MSPEs. σ is the standard deviation of the five MSPEs of the estimated networks. The *Ratio* is the ratio between the corresponding statistics between the feedforward network model with hint and without hint. *DM* refers to the Diebold and Mariano (1995) test for a mean loss differential. This test statistic is distributed standard normal in large samples. All *DM* test statistics are calculated from the loss differential of the mean square prediction errors between the feedforward network models with and without the homogeneity hint. MSPE reported figures have been multiplied by 10^5.

average complexity in parentheses, as well as the standard deviation over the five experiments.

First, it should be noticed that the procedure selects in average a lower complexity than the maximum hidden units allowed for both models. The complexity is also roughly equivalent in terms of parameters for each family of models. The average MSPE is about 20 percent lower for the model with hint. To see if this difference is statistically significant, we compute the Diebold and Mariano (1995) test statistic (DM hereafter).[13] This statistic tests the null hypothesis of no difference in the forecasting accuracy of the two models. Given that the statistic is distributed as a $N(0, 1)$ variable, the value of 20.57 means that the equality of the two forecasts is overwhelmingly rejected in favor of the model with hint.

Hutchinson et al. (1994) argue that a more meaningful performance measure for a given option pricing formula is the tracking error associated with a replicating portfolio that delta-hedges an option position. The hedging ratios (the derivative of the option price with respect to the stock price) are computed based on the pricing formulas. To compute the tracking errors, we follow the procedure described in detail in Hutchinson et al. (1994) and summarized in the Appendix.

---

[13] With this statistic, forecast errors can be serially correlated and contemporaneously correlated. The test statistic is computed by averaging the forecast error differerences of the 25 pairs of models that result from the five estimations of each type of network.

Table 2
Average hedging errors on simulated call option Black–Scholes prices
(Number of options hedged: 18)

Statistics	AHE with hint	$\% < BS$	AHE no hint	$\% < BS$	Ratio	AHE linear	AHE BS
$\bar{x}$	0.2153 (3)	0.21	0.2022 (7)	0.17	1.06	12.28	0.0874
σ	0.0299		0.0086		3.50		

Notes: This table presents the average hedging error (AHE) of a neural network with a homogeneity hint, of a regular feedforward network with no hint, of the Black–Scholes model (BS) and of the linear model for call option prices generated with the Black–Scholes formula. $\bar{x}$ corresponds to the average of the five different AHEs estimated networks from five different seeds. The average number of hidden units of the five runs are reported between parentheses next to the average AHEs. σ is the standard deviation of the five AHEs of the estimated networks. The column ($\% < BS$) is the average of the percentage of options where the delta hedging error was less than the BS delta hedging error.

An average hedging error is reported, where the average is computed across all options hedged in the prediction sample. We also report the number of times that ε computed from the networks is less than the ε obtained with the Black–Scholes formula, as a percentage of the number of options hedged over the prediction period. For computing the hedging performance, the sample is split into three parts as we did for pricing. In the validation period though, for reasons that we will elaborate on in the next section, we select as the best model in each family of networks the one that minimizes the average hedging error and not the MSPE. Table 2 reports the average tracking error over the five runs for both models as well as the average complexity of the networks. Contrary to the pricing performance, the networks without hint produce a slightly lower average hedging error. As it was the case for pricing, the complexity is similar in both networks. The percentage of cases where the hedging error is less than the Black–Scholes hedging error is around 20% for both networks, which is expected since the Black–Scholes model is the true model and the only source of error comes from discrete hedging. In the next section, we assess to what extent these simulation results are confirmed with actual data.

5. Assessment of the relevance of the homogeneity hint for pricing and hedging S&P 500 call options

The data are daily S&P 500 Index European options obtained from the Chicago Board Options Exchange for the period January 1987 to October 1994. The S&P 500 index option market is extremely liquid and it is one of the most active options markets in the United States. This market is the closest to the theoretical setting of the Black–Scholes model. In constructing the data used in

the estimation, options with zero volume are not used. For each year, the sample is split into three parts: first half of the year (training period), third quarter (validation period) and fourth quarter (prediction period). One possible drawback of such a setup is that we will always evaluate the predictive ability of our networks on the last quarter of the year. The advantage is that it will facilitate comparison between years, especially with reference to the last quarter of 1987, when a market crash occurred. Since our main purpose is to compare the feedforward networks with and without the homogeneity hint, the last quarter of each year is as good for prediction purposes as any other period. Moreover, since our methodology for choosing the best architecture involves estimating numerous networks for each family of models and repeating the estimation for five different seeds, it requires a lot of computation and forces us to limit somewhat the scope of our investigation. In the next two subsections, we will analyze the results in terms of predictive performance for pricing and hedging respectively.

5.1. Pricing errors

The network pricing performance measure is the Mean Squared Prediction Error (MSPE) in the prediction sample. Results are presented in Table 3. For each year, we report the average MSPE obtained over the five experiments for each family of networks, along with the average number of hidden units selected. First, it should be observed that the average MSPE for the models with the homogeneity hint (WH) is always smaller than the MSPE of the models with no hint (NN). The average MSPE ratios of the models with and without hint for 1987 to 1993 are 46%, 89%, 98%, 92%, 93%, 92% and 72%, respectively. Furthermore, the ratio of the MSPE standard deviations across the five experiments substantially favours the model with the homogeneity hint. In most years, this ratio is lower than 50%. The values of the DM statistic are all large and positive, which means that we strongly reject the equality of the forecast errors in favor of the feedforward neural networks with hint.

The performance of the linear and the Black-and-Scholes models are also presented in Table 3. Not surprisingly, the linear model provides the poorest performance in terms of MSPE. The MSPE performance of the Black and Scholes (BS) model is significantly better than the linear model but worse than the feedforward network models. Over all years except 1987, the BS MSPE is 3–10 times larger than the feedforward networks. The result in 1987 might then appear surprising. It is less so when one realizes that the BS model incorporates information from the third quarter of the year that is not part of the training sample of the networks. Indeed, the volatility in the BS model is based on the stock returns over the last sixty days preceding the first day of the last quarter. If we give a comparable information to the networks, in the sense that we train them over the first nine months of 1987 and forecast the option prices over the

Table 3
Out-of-sample mean square prediction errors of the SP500 call options
(1987, Total sample: 3610, Validation sample: 2010, Prediction sample: 2239)

Statistics	MSPE with hint	MSPE no hint	Ratio	MSPE linear/BS
$\bar{x}$	16.70 (3)	36.23 (6)	0.46	98.96/4.38
σ	9.51	43.20	0.22	
DM	7.86			

(1988, Total sample: 3434, Validation sample: 1642, Prediction sample: 1479)

Statistics	MSPE with hint	MSPE no hint	Ratio	MSPE linear/BS
$\bar{x}$	0.7114 (4)	0.7959 (7)	0.89	8.41/2.07
σ	0.0429	0.0931	0.46	
DM	9.91			

(1989, Total sample: 3052, Validation sample: 1565, Prediction sample: 1515)

Statistics	MSPE wth hint	MSPE no hint	Ratio	MSPE linear/BS
$\bar{x}$	0.4138 (4)	0.4206 (8)	0.98	3.75/1.42
σ	0.0068	0.0160	0.43	
DM	6.99			

(1990, Total sample: 3605, Validation sample: 2075, Prediction sample: 2166)

Statistics	MSPE with hint	MSPE no hint	Ratio	MSPE linear/BS
$\bar{x}$	0.6761 (3)	0.7253 (6)	0.92	8.15/2.62
σ	0.0763	0.1222	0.62	
DM	5.04			

(1991, Total sample: 4481, Validation sample: 1922, Prediction sample: 2061)

Statistics	MSPE with hint	MSPE no hint	Ratio	MSPE linear/BS
$\bar{x}$	0.3498 (4)	0.3775 (8)	0.93	3.45/1.73
σ	0.0148	0.0336	0.44	
DM	11.57			

(1992, Total sample: 4374, Validation sample: 1922, Prediction sample: 1848)

Statistics	MSPE with hint	MSPE no hint	Ratio	MSPE linear/BS
$\bar{x}$	0.1511 (4)	0.1649 (7)	0.92	2.39/1.36
σ	0.0115	0.0126	0.91	
DM	14.97			

Table 3 (*Continued*)

Statistics	MSPE with hint	MSPE no hint	Ratio	MSPE linear/BS
(1993, Total sample: 4214, Validation sample: 1973, Prediction sample: 2030)				
Statistics	MSPE with hint	MSPE no hint	Ratio	MSPE linear/BS
$\bar{x}$	0.1054 (4)	0.1453 (6)	0.72	2.28/0.74
σ	0.0222	0.0498	0.44	
DM	11.24			

Notes: This table presents the out-of-sample mean square prediction error (MSPE) performance of a neural network with a homogeneity hint, a regular feedforward network with no hint, and of a linear model for call option prices from the SP500 call options. The table reports the average ($\bar{x}$) of the five MSPEs corresponding to five networks estimated from different seeds. The average number of hidden units of the five runs are reported between parentheses next to the average MSPEs. σ is the standard deviation of the five MSPEs of the estimated networks. The *Ratio* is the ratio between the corresponding statistics between the feedforward network model with hint and without hint. *DM* refers to the Diebold and Mariano (1995) test for a mean loss differential. This test statistic is distributed standard normal in large samples. All *DM* test statistics are calculated from the loss differential of the mean square prediction errors between the feedforward network models with and without the homogeneity hint. MSPE reported figures have been multiplied by 10^4.

last quarter, the MSPEs are 6.50 and 4.11 for the NN and WH models respectively. Therefore, with comparable information, the model with hint does slightly better than the BS model in 1987. This emphasizes the fact that in Table 3 the BS model is given an informational advantage compared with the other three models. The relative pricing performance of the networks is all the more remarkable.

To investigate for which options the two types of networks differ in their out-of-sample pricing performance, we report in Table 4 the out-of-sample MSPE for various categories of options based on maturity and moneyness for the year 1993. The ratios of the means and standard deviations of the feedforward networks with and without the homogeneity hint are lowest for the two ends of the spectrum, the short-term out-of-the money options and the longterm in-the-money options. The networks predict with the least difference for the medium-term near-the-money options. This breakdown of predictions by maturity and moneyness emphasizes that the homogeneity hint appears most useful to generalize out of sample when there are less observations in the learning sample. The number of options in sample for each category is very similar to the number reported for the prediction period. The results reported for the year 1993 are representative of what is obtained for the other years.

Table 4
Out-of-sample mean square prediction errors per maturity and moneyness S&P 500 call options
– Year 1993

MSPE $\times 10^4$	No. options	Mean no hint	Mean with hint	Ratio	SD no hint	SD with hint	Ratio
Short term							
Out	128	0.0720	0.0492	0.68	0.0271	0.0149	0.55
Near	473	0.1168	0.0910	0.78	0.0155	0.0149	0.96
In	143	0.2160	0.2323	1.08	0.0834	0.0385	0.46
Medium term							
Out	198	0.0292	0.0216	0.74	0.0068	0.0064	0.94
Near	409	0.0511	0.0503	0.91	0.0067	0.0060	0.92
In	72	0.1962	0.1907	0.97	0.0140	0.0176	1.25
Long term							
Out	186	0.0672	0.0512	0.76	0.0131	0.0193	1.47
At	337	0.0526	0.0439	0.84	0.0118	0.0016	0.14
In	84	1.5109	0.8149	0.54	1.0780	0.3484	0.32

Notes: This table presents a comparison between the out-of-sample mean square prediction error (MSPE) performance of a neural network with a homogeneity hint (WH) and of a regular feedforward network (NN) for S&P 500 European call options of different maturity and moneyness. The means and standard deviations are computed over five MSPEs obtained from five estimated networks starting from five different seeds. The maturity cutoff points are: less than 0.1 (short term), between 0.1 and 0.2 (medium term), above 0.2 (long term). The moneyness cutoff points are: less than 0.97 (out of the money), between 0.97 and 1.05 (near the money), above 1.05 (in the money).

To conclude this assessment of the pricing performance of the competing feedforward network models, we can safely say that the networks with hint predict better than the networks without hint.

5.2. Average hedging errors

Does a good pricing model produce a small average hedging error? One certainly gets this impression when looking at the results in Hutchinson et al. (1994). However, it should be remembered that they keep the architecture of the learning networks fixed. There was no procedure to select the optimal number of units. To answer the question in our setting, let us evaluate the average hedging error that would be obtained when the best pricing models selected in the previous section are used to compute the delta hedging ratios. In the last quarter of the year 1988, the model without hint would have produced an average

hedging error over the five runs greater than 30, compared with 4.50 for the linear model. The model with hint does much better comparatively with an average hedging error of 3.80. Still, both feedforward network models fail to do better than the BS model. The same pattern emerges for the other years.

Then, the best model in terms of out-of-sample pricing performance is not necessarily a good tool to delta hedge a portfolio. This result might suggest that one should train the networks based on the performance criterion ultimately used. The idea is basically the same as the objective-driven inference concept introduced by Gouriéroux et al. (1994) and has also been put forward in the learning network literature (see Bengio, 1996). However, it might be numerically difficult to estimate the parameters of the networks by minimizing an average hedging error criterion. We therefore suggest below a computationally easier route.

The networks are still trained by minimizing a mean square criterion over the first half of the sample, as we did in the previous section. However, in the validation phase over the third quarter of each year, we choose as best models among the two families of feedforward network models the ones that minimize the average hedging error. The average number of units for the selected models appear in parentheses on the average performance $\bar{x}$ line in Table 5. It should be noticed that the models selected are more parsimonious than the models selected on a pricing criterion. For the feedforward networks without hint, the average number of units is either 4 or 5, while it was close to 7 for the pricing criterion (see Table 3). Similarly, for the networks with the homogeneity hint, the average is close to 3 units instead of 4.

Based on this selection procedure on a hedging criterion over the validation sample, the hedging performance of the selected models is evaluated for the options traded in the last quarter of the year. Contrary to what was obtained for the MSPE performance criterion, the ratios of the average hedging errors between the two families of feedforward neural networks are very close to one in most years. Therefore, the result found in the simulation exercise based on Black–Scholes prices is confirmed in the data. However, the standard deviation of the average hedging error over the five runs is often much smaller for the networks with the homogeneity hint, except in 1991. This result shows that the models with hint produce a more consistent hedging performance. The average hedging errors of the networks with the homogeneity hint are more stable across different starting values of the nonlinear optimizers and models with different numbers of hidden units.

It should also be emphasized that both feedforward neural networks are better for hedging than the BS model as the ratio is as low as 67% in 1991. This is also reflected in the percentages of options for which the AHE of the networks is less than the BS AHE. This statistic is always greater than 50% and exceeds70% in 1991.

Table 5
Average hedging errors on SP500 call options
(1988, Number of options hedged: 110)

Statistics	AHE with hint	% < BS	AHE no hint	% < BS	Ratio	AHE BS	AHE linear
$\bar{x}$	1.5911 (3)	0.5764	1.9901 (5)	0.5036	0.80	1.8433	4.5110
σ	0.0507		0.6183		0.08		

(1989, Number of options hedged: 108)

Statistics	AHE with hint	% < BS	AHE no hint	% < BS	Ratio	AHE BS	AHE linear
$\bar{x}$	3.014 (3)	0.5463	2.8060 (5)	0.5685	1.07	3.2343	4.7572
σ	0.1335		0.3255		0.41		

(1990, Number of options hedged: 132)

Statistics	AHE with hint	% < BS	AHE no hint	% < BS	Ratio	AHE BS	AHE linear
$\bar{x}$	2.6260 (3)	0.5818	2.6175 (4)	0.6091	1.003	3.2422	6.6553
σ	0.1266		0.5865		0.22		

(1991, Number of options hedged: 131)

Statistics	AHE with hint	% < BS	AHE no hint	% < BS	Ratio	AHE BS	AHE linear
$\bar{x}$	1.6844 (4)	0.7267	1.7349 (5)	0.7023	0.97	2.5043	4.6615
σ	0.1974		0.07		2.69		

(1992, Number of options hedged: 100)

Statistics	AHE with hint	% < BS	AHE no hint	% < BS	Ratio	AHE BS	AHE linear
$\bar{x}$	2.2100 (2)	0.5780	2.2406 (4)	0.6060	0.99	2.4522	5.8540
σ	0.0848		0.1055		0.80		

(1993, Number of options hedged: 145)

Statistics	AHE with hint	% < BS	AHE no hint	% < BS	Ratio	AHE BS	AHE linear
$\bar{x}$	1.4259 (3)	0.5324	1.4470 (4)	0.5255	0.99	1.5046	2.8114
σ	0.0212		0.0934		0.23		

Notes: This table presents the average hedging error (AHE) of a neural network with a homogeneity hint, of a regular feedforward network with no hint and the linear model from the SP500 call options. $\bar{x}$ corresponds to the average of the five different AHEs estimated networks from five different seeds. The average number of hidden units of the five runs are reported between parentheses next to the average AHEs. σ is the standard deviation of the five AHEs of the estimated networks. The column (% < BS) is the average of the percentage of options where the delta hedging error was less than the BS delta hedging error.

6. Conclusions

In this paper, we show that pricing accuracy gains can be made by exploiting the implications of the homogeneity property of the options prices in a neural network framework. Instead of setting up a learning network mapping the ratio S_t/K and the time to maturity (τ) directly into the derivative price, we break down the pricing function into two parts, one controlled by the ratio S_t/K, the other one by a function of time to maturity. The results indicate that the homogeneity hint always reduces the out-of-sample mean squared prediction error compared with a feedforward neural network with no hint. Our study confirms that the feedforward network models provide smaller delta-hedging errors relative to the Black-and-Scholes model, but does not reveal any significant difference between the feedforward network models themselves. However, the hedging performance of the networks with hint appears to be more stable across the estimated models.

The study of the average hedging error showed that the performance criterion should be taken into account to select the best models. In Hutchinson et al. (1994) the model selection issue was not dealt with as the complexity of the learning networks was fixed. In practice, the complexity of the model is always chosen based on various criteria such as information criteria or out-of-sample prediction errors. Naturally, the same idea will apply to another important use of the estimated pricing function which is the extraction of the risk neutral density. Estimating this density involves computing the second derivative of the option pricing function with respect to the strike price. The selection of the best pricing model might cause a poor estimation of the derivatives of the option pricing function and affect negatively the estimation of the hedging ratio or of the risk neutral density. These issues will be investigated in future research.

Appendix A

A.1. Rules for creation of the options

Stock options are on a January, February, or March cycle. The January cycle consists of the months of January, April, July, and October. The February cycle consists of the months of February, May, August, and November. The March cycle consists of the months of March, June, September, and December. If the expiration date for the current month has not yet been reached, options trade with expiration dates in the current month, the following month, and the next two months in its cycle. If the expiration date of the current month has passed, options trade with expiration dates in the next month, the next-but-one month,

and the next two months of the expiration cycle. When one option reaches expiration, trading in another is started.

To set the strike prices at which options can be written, the rule followed by the exchange is to use a \$5 spacing when the stock price is between \$25 and \$200. When options expire, a new expiration date is introduced. The two strike prices closest to the current stock price are usually selected by the exchange. If one of these is very close to the existing stock price, the third strike price closest to the current stock price may also be selected. If the stock price moves outside the range defined by the highest and the lowest strike price, trading is usually introduced in an option with a new strike price. We set the price of each option according to the Black–Scholes formula.

A.2. Procedure for computing the tracking errors

At date 0, we make up a portfolio by selling one call option and simultaneously purchasing Δ (the derivative of the option pricing function with respect to the stock price) shares of stock. Since the stock purchase is financed through the sale of the option and riskless borrowing, the value of the portfolio at time 0 is zero:

$$V(0) = V_S(0) + V_B(0) + V_C(0) = 0,$$

where $V_S(0)$, $V_B(0)$, and $V_C(0)$ are the values of stocks, bonds and options at time 0. At any time t, the value of the portfolio is given by

$$V(t) = V_S(t) + V_B(t) + V_C(t).$$

We want to evaluate the value of the portfolio at time T, that is the difference between the terminal value of the option and the value of the positions in bonds and stocks. To make this measure comparable over all options hedged in a given sampling period, we take the present value of its absolute value. For a particular option, the tracking error is therefore given by

$$\varepsilon = e^{-rT}|V(T)|.$$

The value of the stock position $V_S(T)$ is equal to $S(T)\Delta$, where Δ is the delta hedge ratio corresponding to the option pricing formula. In our comparison, we assess the average tracking error of four models: the neural networks with and without the homogeneity hint, the linear model and the Black–Scholes model. Even if we generate our option prices according to the Black–Scholes model, there is a tracking error for this latter model caused by the fact that our hedging strategy is discrete and not continuous. For the linear model, Δ is the estimated parameter of the variable S_t/K in the regression of C_t/K on a constant, the ratio

S_t/K and the time to maturity of the option. For the Black–Scholes model, it is given by $\Phi(d_1)$, where Φ is the standard normal cumulative distribution function and

$$d_1 = \frac{\ln(S_t/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}.$$

The value of the bond position $V_B(T)$ is constructed by recursion. We start with $V_B(0)$ from the equation above and compute recursively for each period until T

$$V_B(t) = \exp(r)V_B(t-1) - S(t)(\Delta(t) - \Delta(t-1)).$$

References

Abu-Mostafa, Y., 1993. A method for learning from hints. In: Hanson, S. et al. (Eds.), Advances in Neural Information Processing Systems, vol. 5. Morgan Kaufmann, San Mateo, CA, pp. 73–80.

Abu-Mostafa, Y., 1994. Learning from hints. Journal of Complexity 10, 165–178.

Abu-Mostafa, Y., 1995. Financial market applications of learning from hints. In: Refenes, A. (Ed.), Neural Networks in the Capital Markets. Wiley, London, pp. 221–232.

Aït-Sahalia, Y., Lo, A., 1996. Nonparametric estimation of state-price densities implicit in financial asset prices. Working Paper LFE-1015-96, MIT Laboratory for Financial Engineering.

Amin, K.I., Ng, V.K., 1993. Option valuation with systematic stochastic volatility. Journal of Finance XLVIII (3), 881–909.

Amin, K.I., Jarrow, R., 1992. Pricing options in a stochastic interest rate economy. Mathematical Finance 3 (3), 1–21.

Bailey, W., Stulz, R., 1989. The pricing of stock index options in a general equilibrium model. Journal of Financial and Quantitative Analysis 24, 1–12.

Bengio, Y., 1996. Using a financial training criterion rather than a prediction criterion. International Journal od Neural Systems, in press.

Broadie, M., Detemple, J., Ghysels, E., Torrès, O., 1996a. Nonparametric estimation of American options exercise boundaries and call prices. CIRANO Working Paper 96s-24, Montréal.

Broadie, M., Detemple, J., Ghysels, E., Torrès, O., 1996b. American options with stochastic dividends and volatility: a nonparametric investigation. CIRANO Working Paper 96s-26, Montréal.

Cybenko, G., 1989. Approximation by superposition of a sigmoidal function. Mathematics of Control, Signals and Systems 2, 303–314.

Duda, R., Hart, P., 1973. Pattern classification and scene analysis. Wiley, New York.

Funahashi, K.-I., 1989. On the approximate realization of continuous mappings by neural networks. Neural Networks 2, 183–192.

Gallant, A.R., White, H., 1988. There exists a neural network that does not make avoidable mistakes. Proceedings of the 2nd Annual IEEE Conference on Neural Networks, San Diego, CA. IEEE Press, New York; pp. I.657–I.664.

Gallant, A.R., White, H., 1992. On learning the derivatives of an unknown mapping with multilayer feedforward networks. Neural Networks 5, 129–138.

Garcia, R., Renault, E., 1995. Risk aversion, intertemporal substitution and option pricing. Working Paper, Université de Montréal.

Ghysels, E., Patilea, V., Renault, E., Torrès, O., 1997. Nonparametric methods and option pricing. CIRANO Working Paper 97s-19, Montréal.

Gouriéroux, C., Monfort, A., Tenreiro, C., 1994. Kernel M-estimators: nonparametric diagnostics for structural models. Working Paper 9405, CEPREMAP, Paris.

Hinton, G., 1987. Learning translation invariant recognition in a massively parallel network. Proceedings of Conference on Parallel Architectures and Languages Europe, pp. 1–13.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 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 3, 551–560.

Hu, M., 1962. Visual pattern recognition by moment invariants. IRE Transactions of Information Theory 8, 179–187.

Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. Journal of Finance XLII, 281–300.

Hutchinson, J.M., Lo, A.W., Poggio, T., 1994. A nonparametric approach to pricing and hedging derivative securities via learning network. Journal of Finance 3, 851–889.

Jondeau, E., Rockinger, M., 1997. Reading the smile: the message conveyed by methods which infer risk neutral densities. Mimeo, Banque de France and HEC-School of Management, Paris.

Kuan, C.-M, White, H., 1994. Artificial neural networks: an econometric perspective. Econometric Reviews 13, 1–91.

McCulloch, J.H., 1996. Financial applications of stable distributions, in: Maddala, G.F., Rao, C.R. (Eds.), Handbook of Statistics, vol. 14, Statistical Methods in Finance, Elsevier, Amsterdam.

Merton, R.C., 1973. Rational theory of option pricing. Bell Journal of Economics and Management Science 4, 141–183.

Minsky, M., Papert, S., 1988. Perceptrons. MIT Press, Cambridge, MA.

Patilea, V., Renault, E., 1995. Random probabilities for option pricing. Discussion paper, GREMAQ, Toulouse.

Swanson, N., White, H., 1995. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. Journal of Business and Economics Statistics 13, 265–275.

Turnbull, S., Milne, F., 1991. A simple approach to interest-rate option pricing. Review of Financial Studies 4, 87–121.

White, H., 1992. Artificial Neural Networks: Approximation and Learning. Blackwell, Cambridge.