



WP 15-13

Nikola Gradojevic

IÉSEG School of Management (LEM-CNRS), Lille Catholic University, France
The Rimini Centre for Economic Analysis, Italy

MULTI-CRITERIA CLASSIFICATION FOR PRICING EUROPEAN OPTIONS

Copyright belongs to the author. Small sections of the text, not exceeding three paragraphs, can be used provided proper acknowledgement is given.

The *Rimini Centre for Economic Analysis* (RCEA) was established in March 2007. RCEA is a private, nonprofit organization dedicated to independent research in Applied and Theoretical Economics and related fields. RCEA organizes seminars and workshops, sponsors a general interest journal *The Review of Economic Analysis*, and organizes a biennial conference: *The Rimini Conference in Economics and Finance* (RCEF). The RCEA has a Canadian branch: *The Rimini Centre for Economic Analysis in Canada* (RCEA-Canada). Scientific work contributed by the RCEA Scholars is published in the RCEA Working Papers and Professional Report series.

The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Rimini Centre for Economic Analysis.

Multi-criteria Classification for Pricing European Options

Nikola Gradojevic*

2015

Abstract

This paper builds a novel multi-criteria, non-parametric classification framework in order to improve the accuracy of pricing European options. The proposed approach is based on classifying financial options according to their implied volatility, time to maturity and moneyness. Using a recent data set for the daily S&P 500 index call options traded in 2012, the multi-criteria modular neural network model demonstrates its superior out-of-sample pricing performance relative to competing parametric and non-parametric models. By observing the model's pricing errors across various option types, the analysis provides additional insights into pricing biases and stresses the importance of selecting appropriate classification criteria.

Keywords: Finance; Option Pricing; Neural Networks; Implied Volatility; Classification; Non-parametric Methods.

JEL No: C45; G12

*IÉSEG School of Management (LEM-CNRS), Lille Catholic University, 3, rue de la Digue, 59000 Lille, France. Email: n.gradojevic@ieseg.fr; The Rimini Centre for Economic Analysis, Italy.

1. Introduction

A time series of financial option prices is in general subject to discrete regime shifts that reflect changes in business cycles, government policy and random shocks, originating from systemic and firm-specific events. To handle the problem of option pricing under regime switching, various approaches have been developed. They include the regime switching model by Hamilton (1989), GARCH option pricing models by Duan (1995), Markov switching-type models (Elliott and Siu, 2013), non-parametric models (Kukolj *et al.*, 2012) and Bayesian approaches (Martin *et al.*, 2005).

In addition to potential structural breaks resulting from temporal regime shifts, the estimation (or training) of option pricing models on heterogeneous data can be problematic for several reasons. First, the cross-sectional heterogeneity is associated with option traders who have different time horizons and risk preferences. At the more static level of the trading mechanism are the fundamental traders who trade on longer time horizons. At dynamic layers, there are high-frequency (day traders) and short-term traders with a time horizon of a few days. Each of these types of traders may be interested in buying and selling options of specific maturity and “moneyness”.¹ Also, some traders may trade by observing the implied volatility movements. For instance, trading strategies that exploit the so-called “volatility skew” involve the situation when a higher-strike option is overpriced relative to lower-strike options, and vice-versa. Overall, it is the combination of traders’ preferences and their trading activities across trading horizons that generates market prices.

Consequently, during its estimation, an option pricing model could become trapped in local minima or may generalize poorly (Ronco and Gawthrop, 1995). Moreover, it could be susceptible to the “recency effect”: the model parameters adapted unduly in favor of the most recent training data. Feldkamp and Puskorius (1998) approach the recency effect through multistream learning where the data set is split into multiple files (streams) and, in each training cycle, the estimated parameters in streams are updated independently. This procedure improves the generalization properties of a non-linear model and is an increasing function of the number of streams.

This paper builds on the recent literature that documents the effectiveness of the notion of modularity in option pricing, i.e., decomposing the data into modules organized with regard

¹Options are financial derivatives whose value (i.e., price) depends on the value of the underlying security. For example, a call (put) option based on a stock provides its buyer with a right to purchase (sell) a predetermined amount of stocks at a contracted price (“strike price”) on or by a specific date (“maturity”). For this right the buyer of an option pays a price called the premium. Moneyness is defined as the ratio of the stock price to the strike price.

to certain classification criteria (Gradojevic *et al.*, 2009; Gradojevic and Kukolj, 2011; Kukolj *et al.*, 2012). Each module represents a sub-model that is estimated independently. Finally, option price predictions from the sub-models are combined into a global option pricing function. The novelty of the current paper is that it improves upon the previous research contributions by taking them to a multi-criteria setting, thus providing valuable insights into option pricing and its biases. Specifically, the proposed multi-criteria method classifies options while taking into account the following three dimensions: moneyness, maturity and implied volatility.

The findings confirm the usefulness of the non-parametric, multi-criteria methodology that results in the most accurate out-of-sample pricing performance, relative to the competing parametric and non-parametric models. Compared to the Black-Scholes model, the average out-of-sample accuracy gains are statistically significant and range from roughly 90% to 98% (in the last quarter of 2012) and from 81% to 85% (in the first quarter of 2012). The improvements in pricing accuracy over the alternative models are also remarkable: about 80% for the feedforward neural network (NN) model with the “hint” (Garcia and Gençay, 2000) and about 95% for the semi-parametric model by Ait-Sahalia and Lo (1998). Given that Gençay and Gibson (2009) showed that the out-of-sample pricing performance of the NN model dominates standard parametric models for the S&P 500 index, it can be concluded that the proposed multi-criteria approach exhibits superior forecasting ability.

Further analysis presents various biases in option pricing that are of significant practical importance to investors. First, while the increased complexity of the three-criteria option pricing model is in general beneficial, certain option types such as the long-term and in-the-money options with low implied volatility have been found difficult to price accurately. In this context, an empirical investigation that compares several multi-criteria settings shows that pricing biases in a two-criteria model that omits the volatility input are not as large as the ones in a three-criteria model. In all, the largest pricing biases arise from the implied volatility, which is followed by the biases associated with moneyness. Both of these cases produce more serious biases when compared the pricing errors related to the maturity of options.

The remainder of the paper is laid out as follows. Section 2 briefly reviews the relevant option pricing literature. Section 3 explains the multi-criteria modular NN model as well as the competing models. Section 4 describes the data and the results of the out-of-sample pricing exercises. Section 5 provides concluding remarks.

2. Literature review

From the finance perspective, accurate option pricing is important for several reasons: 1) an accurate pricing model enables traders to avoid selling underpriced or buying overpriced options, 2) it provides an arbitrage-free method of pricing other less liquid options given the observed pricing function on e.g., the liquid S&P 500 index options, 3) to hedge options investors need a model that is able to explain the behavior of option prices across different strike prices and maturities, and 4) the information from the option market helps the understanding of the underlying asset dynamics.

For the above reasons, since the seminal work by Black and Scholes (1973), option pricing has become the focus of scholarly attention. However, past literature suggests that the standard assumptions of the Black-Scholes model are rarely satisfied. For instance, the well-documented “volatility smile” and “volatility smirk” pricing biases violate the Black-Scholes model assumption of constant volatility. Additionally, stock returns have been shown to exhibit non-normality and jumps. Finally, biases also occur across option maturities.²

In order to address the biases of the Black-Scholes model, research efforts have focused on developing parametric and non-parametric models. With regard to parametric models, the research has mainly focused on three models: the stochastic volatility (SV), stochastic volatility random jump (SVJ) and stochastic interest rate (SI) parametric models. All three models have been shown to be superior to the Black-Scholes model in out-of-sample pricing and hedging exercises (Bakshi *et al.*, 1997). Specifically, the SV model has been shown to have first-order importance over the Black-Scholes model (Gençay and Gibson, 2009). The SVJ model further enhances the SV model for pricing short-term options, while the SI model extends the SVJ model in regards to the pricing of long-term options.

Although parametric models appear to be successful in relaxing the assumptions that underlie the Black-Scholes model, while simultaneously improving pricing accuracy, these models exhibit some moneyness-related biases for short-term options. In addition, the pricing improvements produced by the parametric models are generally not robust and are inferior to non-parametric approaches (Gençay and Gibson, 2009; Gradojevic *et al.*, 2009). For example, the out-of-sample performance of an artificial NN model was compared to the SVJ, SI and SV parametric approaches for the S&P 500 stock market index over the 1989-1991 data span. The comparison concludes that the NN pricing model with the GARCH (1, 1) volatility dominates all parametric models. The non-parametric approaches to option pricing have also been used by Hutchinson *et al.* (1994), Garcia and Gençay (2000), Qi and Maddala (1996), Gençay and Qi (2001), Andreou *et al.* (2008) and Gençay and Altay-Salih (2003).

²For an excellent recent review of the option pricing literature please see Renault (2010).

Some other important contributions in this respect are the mixture of distributions model by Melick and Thomas (1997) and Bhat and Kumar (2012), and the semi-parametric estimator by Ait-Sahalia and Lo (1998). These models have shown sizable improvements in option pricing accuracy compared to the Black-Scholes model; however, their out-of-sample pricing is inferior to non-parametric modular NN models (Gradojevic *et al.*, 2009). Some recent non-parametric approaches also include the affine jump-diffusion models (Carr and Wu, 2004) and the normal inverse Gaussian models (Eriksson *et al.*, 2009; Barndorff-Nielsen and Shephard, 2001). More recently, fuzzy logic has been shown to be very useful for option pricing (Agliardi and Agliardi, 2009; Yoshida, 2003).

3. Option pricing methodologies

3.1 Multi-criteria modular neural network

To explain the concept of a multi-criteria modular NN (MCMNN) modeling of option prices, it will be assumed that a function of four variables c_t ($c_t = \phi(x_{1t}, x_{2t}, x_{3t}, x_{4t}) + \varepsilon_t$) is driven by different functions defined over known domains of x_{1t} , x_{2t} , x_{3t} and x_{4t} , where t denotes the time index. Namely, the function $\phi(x_{1t}, x_{2t}, x_{3t}, x_{4t})$ can be written as a linear combination of functions $c_{1t}, c_{2t}, \dots, c_{Mt}$ as follows (Jang *et al.*, 1997):

$$\phi(x_{1t}, \dots, x_{4t}) = \omega_1 c_{1t}(x_{1t}, \dots, x_{4t}) + \dots + \omega_M c_{Mt}(x_{1t}, \dots, x_{4t}) = \sum_{k=1}^M \omega_k c_{kt}(x_{1t}, \dots, x_{4t}). \quad (1)$$

This function can be approximated by a MCMNN (Figure 1). A meaningful decomposition of the function $\phi(x_{1t}, \dots, x_{4t})$ would be to approximate it locally by M modules $c_{1t}, c_{2t}, \dots, c_{Mt}$ and then to sum the output of each module, while applying the appropriate weights $\omega_1, \omega_2, \dots, \omega_M$. In this work, a partitioning approach that sets only one $\omega_i = 1$, while all other $\omega_k = 0$ for $k \neq i$, $k = 1, \dots, M$, is applied. This can be viewed as a “crude partitioning” of the (x_{1t}, \dots, x_{4t}) space, i.e., the weights are redistributed to only one module at the time.³

Modules and weights are specified by using the moneyness (S_t/K), implied volatility (σ_{IV}) and time to maturity (τ) criteria. The modules are feedforward (backpropagation) NNs and are in general estimated as:

³The combination weights ω_k ($k = 1, \dots, M$) can also be estimated using a “softmax” method (McCullagh and Nelder, 1989; Bridle, 1989) or Gaussian mixture models (Haykin, 1999).

$$c_{kt} = g_k \left(\beta_{k0} + \sum_{j=1}^{q_k} \beta_{kj} \psi_k \left(\alpha_{kj0} + \sum_{i=1}^s \alpha_{kij} x_{it} \right) \right), \quad k = 1, \dots, M. \quad (2)$$

In this example, $s=4$ ($i=1,2,3,4$) and q_k is the number of hidden nodes for the k^{th} module. The single hidden and the output layers of the modules are characterized by two flexible classes of non-linearities: ψ_k and g_k , respectively. The backpropagation learning algorithm requires continuous differentiable non-linearities. The types used in this paper are the sigmoid logistic or hyperbolic tangent functions in the hidden layer, and the linear function in the output layer. α_{kij} and β_{kj} denote appropriate connection weights between the adjacent layers for the modules. Subscripts 0 for α and β stand for NN biases.

[Insert Figure 1 about here]

The option pricing formula is defined in the spirit of Hutchinson *et al.* (1994) and Garcia and Gençay (2000):

$$C_t = \phi(S_t, K, \tau, \sigma_{IV}, r) \quad (3)$$

where C_t is the call option price, S_t is the price of the underlying asset, K is the strike price, τ is the time to maturity, r is the risk-free interest rate and σ_{IV} is the implied volatility. Assuming the homogeneity of degree one of the pricing function ϕ with respect to S_t and K , one can write the option pricing function as follows:

$$\underbrace{\frac{C_t}{K}}_{c_t} = \phi \left(\underbrace{\frac{S_t}{K}}_{x_1}, 1, \underbrace{\tau}_{x_2}, \underbrace{\sigma_{IV}}_{x_3}, \underbrace{r}_{x_4} \right) = \phi(x_1, x_2, x_3, x_4). \quad (4)$$

In general, options are often referred to as plain vanilla derivatives because their payoff (or price) is determined by the so-called underlying, which is in our case the S&P-500 stock market index. Call options are more profitable for the buyer when, *ceteris paribus*, the price of the underlying (S_t) increases or the strike price (K) decreases. Therefore, intuitively, these two variables must be integral parts of the option pricing formula. Further, when time to maturity (τ) increases, call options become more valuable. This is explained by the fact that it is more likely that the option will be in the money ($S_t - K > 0$) and, thus, worthwhile exercising at maturity. The preceding explanatory variables are extended with the implied volatility and the risk-free interest rate that are considered standard inputs to an option pricing model.

The pricing function ϕ is approximated by a MCMNN with 18 modules determined by the three criteria (S_t/K , τ and σ_{IV}). The time to maturity cutoff points are selected as follows: $\tau < 0.1$ (short-term options), $0.1 \leq \tau \leq 0.2$ (medium-term options) and $\tau > 0.2$ (long-term options).⁴ The selection of modules according to the moneyness criterion is performed as follows: $(S_t/K) < 0.98$ (out-of-the-money options), $0.98 \leq (S_t/K) \leq 1.02$ (near-the-money options) and $(S_t/K) > 1.02$ (in-the-money options). The implied volatility criteria classifies options across two ranges: $\sigma_{IV} \leq 0.15$ (low volatility options) and $\sigma_{IV} > 0.15$ (high volatility options)⁵. Thus, the pricing function is decomposed along the three dimensions into 18 separate non-linearities ($3 \times 3 \times 2 = 18$). The modules are trained independently on the data for the 18 option types and, during prediction, based on the values of x_1 , x_2 and x_3 , only one is active: when $\omega_i = 1$, ($i = 1, \dots, 18$), $\omega_k = 0$ ($k \neq i$, $k = 1, \dots, 18$). Table 1 describes the modules specified over different ranges of S_t/K , τ and σ_{IV} that are used to estimate functions c_{kt} ($k = 1, \dots, 18$).⁶

[Insert Table 1 about here]

The choice of the number of modules, first, follows the logic of categorizing options based on maturity (short-term, medium-term, long-term), then, moneyness (out-of-the-money, near-the-money, in-the-money), and, finally, by their implied volatility (low, high). The breakpoints of the categories are adapted from Table 4 of Garcia and Gençay (2000). Therefore, a “hard-partitioning” approach is used to choose both the number of modules and the cut-off points. This approach is by no means considered an optimal multi-criteria option pricing model and the goal of the paper is not to search for an optimal number of modules (and their boundaries). Rather, the intention is to demonstrate that even a crude multi-criteria classification scheme can result in a superior forecasting performance. Furthermore, this setting allows a thorough examination of the role of implied volatility in option pricing.

3.2 Alternative estimators

The out-of-sample pricing performance of an MCMNN model is first compared to the Black-Scholes model. The Black-Scholes call prices (C_t) are computed using the standard formula:

⁴The units for τ are the number of days to maturity divided by the number of days in a year (366 for 2012).

⁵0.15 is the median value of implied volatility across all traded call options in the first two quarters of the year (training data).

⁶Increasing the number of modules leads to insufficient data for training of NNs at a daily frequency, but this is certainly a viable research avenue in a high-frequency setting, when a large number of financial products are traded.

$$C_t = S_t N(d) - K e^{-r\tau} N(d - \sigma\sqrt{\tau}) \quad \text{where } d = \frac{\ln(S_t/K) + (r + 0.5\sigma^2)\tau}{\sigma\sqrt{\tau}} \quad (5)$$

where N is the cumulative normal distribution, S_t is the price of the underlying asset, K is the strike price, τ is the time to maturity, r is the risk-free interest rate, and σ is the volatility of the underlying asset.

A popular non-parametric alternative to NNs to estimate the pricing function is kernel regression (Aït-Sahalia and Lo, 1998). In contrast to NNs, where observational errors are reduced by averaging the data based on a recursive error minimization procedure, for kernel regression this is achieved by local averaging (smoothing). The kernel method estimates the price based on the weighted sum of the information from the in-sample data. More specifically, given some vector $(x_{1t}; x_{2t}; x_{3t}; x_{4t})$, to estimate the price, more weight is assigned to the observations at locations that are closer to the vector. Suppose that an option pricing model has four regressors. In that case, the appropriate kernel pricing function would be constructed as a product of four univariate kernels. The usual trade-off between smoothness and goodness-of-fit is achieved by the choice of the bandwidth of the kernel function. Aït-Sahalia and Lo (1998) show that the kernel estimation errors for the SPD and option delta lie within one per cent of the theoretical values. The out-of-sample pricing performance of their model is also impressive and superior over the NN model. It is noteworthy that the kernel pricing function from Aït-Sahalia and Lo (1998) that is used for forecasting is semi-parametric. It is informed by dividend yield, τ , K , S_t , risk-free interest rate, and non-parametrically estimated volatility. To forecast option prices out-of-sample, a semi-parametric kernel approach from Aït-Sahalia and Lo (1998) is followed. First, implied volatilities were estimated based on futures price, the exercise price and time to maturity using the Nadaraya-Watson kernel estimator. Then, call option prices were computed from the estimated volatility function via the Black-Scholes formula.

Another competing model that is estimated is the Garcia and Gençay (2000)'s model with the hint. The "hint" involves utilizing additional prior information about the properties of an unknown (pricing) function that is used to guide the learning process. In the context of Garcia and Gençay (2000), this means breaking up the pricing function into four parts, controlled by x_1 (S_t/K), x_2 (τ), x_3 (σ_{IV}) and x_4 (r). Each part contains a cumulative distribution function which is estimated non-parametrically through NN models:

$$\begin{aligned}
f(x_1, x_2, x_3, x_4; \theta) = & \beta_0 + x_1 \left(\sum_{j=1}^d \beta_j^1 \frac{1}{1 + \exp(-\gamma_{j0}^1 - \gamma_{j1}^1 x_1 - \gamma_{j2}^1 x_2 - \gamma_{j3}^1 x_3 - \gamma_{j4}^1 x_4)} \right) \\
& + x_2 \left(\sum_{j=1}^d \beta_j^2 \frac{1}{1 + \exp(-\gamma_{j0}^2 - \gamma_{j1}^2 x_1 - \gamma_{j2}^2 x_2 - \gamma_{j3}^2 x_3 - \gamma_{j4}^2 x_4)} \right) \\
& + x_3 \left(\sum_{j=1}^d \beta_j^3 \frac{1}{1 + \exp(-\gamma_{j0}^3 - \gamma_{j1}^3 x_1 - \gamma_{j2}^3 x_2 - \gamma_{j3}^3 x_3 - \gamma_{j4}^3 x_4)} \right) \\
& + x_4 \left(\sum_{j=1}^d \beta_j^4 \frac{1}{1 + \exp(-\gamma_{j0}^4 - \gamma_{j1}^4 x_1 - \gamma_{j2}^4 x_2 - \gamma_{j3}^4 x_3 - \gamma_{j4}^4 x_4)} \right) \quad (6)
\end{aligned}$$

where θ denotes the parameters of the NN model that are to be estimated (β and γ) and d is the number of hidden units in the NN model, which is set according to the best performing NN model in terms of the magnitude of the mean-squared prediction error (MSPE) on the validation data. To control for possible sensitivity of the NNs to the initial parameter values, the estimation is performed from ten different random seeds and the average MSPE values are reported. The averaging approach is also applied to the MCMNN model and the average measures of forecasting performance are reported.

The statistical significance of the difference in the out-of-sample (testing set) performance of alternative models is tested using the Diebold-Mariano test (Diebold and Mariano, 1995). The null hypothesis that there is no difference in the MSPE of the two alternative models is tested.

4. Results

The data are provided by DeltaNeutral and represent the daily S&P 500 index European call option prices, taken from the Chicago Board Options Exchange. Call options across different strike prices and maturities are considered for 2012. Since it is one of the deepest and the most liquid option markets in the United States, the S&P 500 index option market is sufficiently close to the theoretical setting of the Black-Scholes model. The implied volatility used in the estimations is a proprietary mean estimate provided by DeltaNeutral. The risk-free rate is approximated by the monthly yield of the U.S. Treasury bills.

To reduce the size of the data set, options with zero volume on a given day were eliminated, and, due to theoretical considerations, the focus was only on the near-the-money

options (with strike prices between 95% and 105% of the underlying S&P 500 index). The data were divided into three parts: the first two quarters of 2012 (estimation data), third quarter of 2012 (validation data) and fourth quarter of 2012 (testing data). This produced the following non-overlapping sub-samples:

- Training sample: 15001 observations;
- Validation sample: 10542 observations;
- Testing sample: 10457 observations.

The optimal NN architecture for each module is determined from the out-of-sample performance on the validation set with respect to the MSPE. To keep the complexity of the NNs close to Garcia and Gençay (2000) and Gradojevic *et al.* (2009), the modules are single hidden layer NNs with either sigmoid logistic or hyperbolic tangent activation functions. Hence, the search for an optimal NN architecture involves specifying the number of hidden nodes (from 1-15 hidden nodes) and their activation functions. The parameters are estimated using the standard Levenberg-Marquardt algorithm. Overfitting is prevented by early stopping, i.e., stopping the training process when the validation set error starts to increase. To control for possible sensitivity of the NNs to the initial parameter values, the training is performed from ten different random seeds and the average MSPE and the mean absolute percentage error (MAPE) values are reported.

Table 2 presents the pricing, out-of-sample performance of the MCMNN model with 18 modules. The average pricing errors with respect to the average MSPE of each module is also reported. The total average MSPE is 0.102×10^{-4} , which is comparable to the most accurate performances for the S&P 500 call options in Garcia and Gençay (2000) and Gradojevic *et al.* (2009). The MSPEs across the modules, however, exhibit some variability, with the largest errors observed in modules (6), (8) and (9), which are underlined in the table. It appears that pricing biases arise mainly due to the difficulties in pricing specific moneyness and volatility (in-the-money, low volatility options). The most elusive option type to price is an in-the-money, low volatility, long-term option (module 9). The median values of the errors confirm the findings, while the large standard deviation values for modules (6) and (9) can be interpreted as a substantial instability in the performance of NN models in those modules.

[Insert Table 2 about here]

To investigate in more detail the pricing biases of the MCMNN model, Figure 2 plots all out-of-sample $(c_t - \hat{c}_t)^2$ ($t = 1, \dots, 10457$) pricing errors on the Z-axis in a three-dimensional

space against the following predictors on the X- and Y-axes: 1) implied volatility (σ_{IV}) and maturity (τ) in Panel A, 2) implied volatility (σ_{IV}) and moneyness (S_t/K) in Panel B, and 3) maturity (τ) and moneyness (S_t/K) in Panel C. Panel A of Figure 2 shows relatively large pricing biases for the low volatility options. Panel B confirms that the low implied volatility options are difficult to price, but it also points to another source of large pricing errors, namely the in-the-money options. The area of large pricing errors due to moneyness and volatility is represented in Panel B by a sizeable vertical spike. As the area of mispricing in the last panel is located identically, it can be concluded that the biases related to moneyness are systematic. In all, the model frequently misprices some in-the-money and low volatility options, while the degree of mispricing seems unrelated to the maturity. Hence, the shape of the error surface in Figure 2 complements the above analysis based on Table 2.

[Insert Figure 2 about here]

Table 3 compares the out-of-sample pricing performance of the MCMNN model with 18 modules, the NN model with the hint, the kernel pricing model from Ait-Sahalia and Lo (1998) and the Black-Scholes model in terms of the MSPE and MAPE performance measures. The MCMNN and the NN model with the hint models are estimated ten times from ten different sets of starting values and the average MSPE and MAPE are reported. The pricing errors uncover the superiority of the MCMNN approach, followed by the NN model with the hint as the best alternative. The Black-Scholes model performed especially poorly and its pricing performance was statistically significantly worse (at the 1% significance level) according to the Diebold-Mariano statistic than the non-parametric models. The pricing performance of the MCMNN model in the fourth quarter is remarkable considering that the third quarter (validation) data are not included in the training set.⁷

[Insert Table 3 about here]

Next, to gain additional insight into the generalization properties of the MCMNN model, its pricing ability is compared to all variants of the model that employ two classification criteria. This exercise will also provide a rationale for extending an MNN model to a MCMNN model. Three classification schemes with nine modules each are considered: 1) MNN with moneyness (S_t/K) and volatility (σ_{IV}) classification criteria, 2) MNN with maturity (τ) and

⁷Similar pricing performance was obtained by using the third and the fourth quarter of 2012 as training data, the second quarter as validation data and the first quarter as testing data. This “reverse” pricing exercise demonstrated the robustness of the MCMNN methodology. For brevity reasons, the results are not included in the current paper, but they can be obtained from the author by request.

volatility (σ_{IV}) classification criteria, and 3) MNN with moneyness (S_t/K) and maturity (τ) classification criteria. The first specification uses the two variables that produced the largest pricing biases in the MCMNN model and its pricing errors for the modules are listed in Table 4. The overall average MSPE can be found in the bottom-left corner of the table and it is substantially greater than the one for the MCMNN model. Clearly, the average MSPE is inflated by the poor pricing of the in-the-money, low volatility options (module 3). Hence, for this MNN specification, it is beneficial to extend the set of classification criteria with the maturity variable to a MCMNN model.

[Insert Table 4 about here]

With the second MNN variant, the goal is to improve upon the first model variant. This is to a certain extent expected because one of the problematic classification criteria (moneyness) is replaced by the “well-behaved” maturity classification criterion. This resulted in a reduction of the average MSPE from 0.357×10^{-4} to 0.153×10^{-4} (Table 5). Nevertheless, the pricing error is still greater than the MCMNN model’s error. Module (3) involved a relatively large average MSPE that contributed the most to the deterioration of the model’s performance. It can be concluded that the long-term, low volatility options are somewhat difficult to price. Noteworthy, it is advantageous to extend the MNN model with an additional classification criterion, which increases the pricing accuracy.

[Insert Table 5 about here]

The last MNN variant utilizes the most informative classification criteria (moneyness and maturity) that do not seem to exhibit the strongest pricing biases. Table 6 suggests that this improves the average overall MSPE to 0.075×10^{-4} , which becomes smaller than that of the MCMNN model. Although the pricing improvement relative to the MCMNN model is marginal (roughly 2.5%), the results stress that increasing the classification complexity does not necessarily imply superior performance. Therefore, supplying an uninformative classification criterion such as the volatility to an MNN model may worsen its generalization ability.

[Insert Table 6 about here]

Based on the above results, it is apparent that, for a given set of model inputs, the choice of the classification scheme is vital for the MCMNN model’s superior performance. This evidence is in line with some other studies that focused on pricing of certain types of options

(i.e., deep out-of-the-money options) and found that the classification of the data increases the pricing accuracy of the NN models (Gençay and Altay-Salih, 2003; Bennell and Sutcliffe, 2004). However, as shown in Table 6, one has to be cautious when using available domain knowledge to identify modules. Over-classification may result in pricing biases due to an inappropriate partitioning of the input space and potentially insufficient amount of training data in certain modules.

5. Conclusions

The purpose of this paper is to propose an extension of the MNN model to a multi-criteria setting called the MCMNN model. This model is estimated for the S&P 500 European call options in 2012. The use of recent data is advantageous since it allows for a comparative study across an array of competing models such as the feedforward NN model with the “hint” (Garcia and Gençay, 2000), the semi-parametric model of Aït-Sahalia and Lo (1998), the two-criteria MNN model (Gradojevic *et al.*, 2009) and the standard Black-Scholes model.⁸ Also, it provides a valuable insight into the models’ out-of-sample performance relative to the past studies in the field of option pricing.

The findings show that the MCMNN approach yields substantial pricing improvements over all the competing models. The accuracy of the model is, however, somewhat plagued by certain biases that arise from the implied volatility classification criterion. These biases are most pronounced when estimating the in-the-money, low volatility options that prove to be the most difficult to price. In all, the evidence favors the multi-criteria classification, but warns that the classification criteria have to be selected with care. Another point that requires further consideration is the amount data available for training of individual modules. Setting too many classification criteria reduces the size of the training set and may impede the learning ability of the NN models in some modules. Thus, it is of utmost importance to balance the complexity of the model with its forecasting ability. Yet, the success of the MCMNN model rests upon its multi-dimensional modularity feature that allows NNs to “specialize” in pricing certain types of options.

With all the preceding issues in perspective, it can be concluded that the idea of multi-dimensional modularity represents a promising research route and its potential applicability

⁸Gençay and Gibson (2009) demonstrated that NN models outperform the standard parametric extensions of the Black-Scholes model in terms of their out-of-sample pricing performance. For that reason, this paper concentrates on non-parametric option pricing and does not include the SV and SVJ models into the set of alternative models. In general, there are two possible explanations for the superiority of NN models: flexible functional forms that capture non-normality of return distributions and the adaptive learning ability.

should not be limited only to option pricing. Furthermore, the current paper's choice for the number of classification criteria, the number of modules and the breakpoints between the modules is arbitrary, but sensible and intuitive in the context of pricing financial options. Consequently, in future research, the attention should be placed on classification algorithms and selection of modules based on 'intelligent' methodologies from the statistical learning theory.

		$(S_t/K) < 0.98$	$0.98 \leq (S_t/K) \leq 1.02$	$(S_t/K) > 1.02$
$\sigma_{IV} \leq 0.15$	$\tau < 0.1$	c_{1t} (module 1)	c_{2t} (module 2)	c_{3t} (module 3)
	$0.1 \leq \tau \leq 0.2$	c_{4t} (module 4)	c_{5t} (module 5)	c_{6t} (module 6)
	$\tau > 0.2$	c_{7t} (module 7)	c_{8t} (module 8)	c_{9t} (module 9)
$\sigma_{IV} > 0.15$	$\tau < 0.1$	c_{10t} (module 10)	c_{11t} (module 11)	c_{12t} (module 12)
	$0.1 \leq \tau \leq 0.2$	c_{13t} (module 13)	c_{14t} (module 14)	c_{15t} (module 15)
	$\tau > 0.2$	c_{16t} (module 16)	c_{17t} (module 17)	c_{18t} (module 18)

Table 1: MODULE SELECTION CRITERIA FOR THE MCMNN OPTION PRICING MODEL WITH 18 MODULES.

Notes: The options are categorized based on maturity (short-term, medium-term, long-term), moneyness (out-of-the-money, near-the-money, in-the-money) and implied volatility (low volatility, high volatility). The three criteria generate the out-of-sample predictions that are estimated from a MCMNN with 18 modules. Specifically, ω_i 's for all 18 modules are specified in the following fashion: when $\omega_i = 1$ ($i = 1, \dots, 18$), $\omega_k = 0$ ($k \neq i, k = 1, \dots, 18$).

Module (i)	\overline{MSPE}_i	$median(MSPE_i)$	$std(MSPE_i)$	Type
(1)	0.002	0.002	0.001	OTM, ST, LIV
(2)	0.071	0.003	0.208	NTM, ST, LIV
(3)	0.055	0.018	0.110	ITM, ST, LIV
(4)	0.005	0.001	0.013	OTM, MT, LIV
(5)	0.051	0.002	0.152	NTM, MT, LIV
(6)	<u>0.772</u>	<u>0.557</u>	<u>0.728</u>	ITM, MT, LIV
(7)	0.153	0.111	0.109	OTM, LT, LIV
(8)	<u>0.557</u>	<u>0.527</u>	<u>0.169</u>	NTM, LT, LIV
(9)	<u>4.91</u>	<u>1.59</u>	<u>4.42</u>	ITM, LT, LIV
(10)	0.013	0.004	0.027	OTM, ST, HIV
(11)	0.027	0.002	0.041	NTM, ST, HIV
(12)	0.026	0.012	0.032	ITM, ST, HIV
(13)	0.014	0.004	0.023	OTM, MT, HIV
(14)	0.020	0.002	0.055	NTM, MT, HIV
(15)	0.004	0.004	0.002	ITM, MT, HIV
(16)	0.273	0.004	0.661	OTM, LT, HIV
(17)	0.033	0.021	0.025	NTM, LT, HIV
(18)	0.099	0.038	0.106	ITM, LT, HIV
Total MCMNN	0.102	0.090	0.121	

Table 2: PREDICTION PERFORMANCE OF THE MCMNN OPTION PRICING MODEL WITH 18 MODULES.

Notes: The average MSPE (\overline{MSPE}_i), its median value ($median(MSPE_i)$) and the standard deviation ($std(MSPE_i)$) have been obtained for each module i ($i = 1, \dots, 18$) and presented as averages across ten different random training seeds. The last row (“Total MCMNN”) represents the average measures of forecasting performance for the MCMNN option pricing model that combines all modules together and uses all data. The abbreviations for the option types are as follows: out-of-the-money (OTM), near-the-money (NTM), in-the-money (ITM), short-term maturity (ST), medium-term maturity (MT), long-term maturity (LT), low implied volatility (LIV) and high implied volatility (HIV). The figures in the table have been multiplied by 10^4 . The largest $MSPE$ -related values are underlined.

Model	\overline{MSPE}	$median(MSPE)$	$std(MSPE)$	\overline{MAPE}	$median(MAPE)$	$std(MAPE)$
MCMNN	0.102	0.090	0.121	5.27	5.39	1.48
NN with hint	0.521	0.101	0.838	12.40	12.95	6.87
A-LO	1.90	-	-	9.51	-	-
BS model	3.24	-	-	15.67	-	-
p-value(DM)	0.000	-	-	0.000	-	-

Table 3: COMPARATIVE PREDICTION PERFORMANCE.

Notes: Forecasting performance of the following option pricing models is considered: Garcia and Gençay (2000)'s feedforward NN model with the hint (NN with hint), multi-criteria modular neural network (MCMNN), kernel regression from Ait-Sahalia and Lo (1998) (A-LO) and the Black-Scholes model (BS model). The average statistics for the MCMNN and NN with the hint models have been obtained as averages across ten different random training seeds. p-value(DM) denotes the p-value of the Diebold and Mariano (1995) test statistic. This test is used to assess the statistical significance of the MCMNN model's forecast gains relative to the BS model. All $MSPE$ -related figures have been multiplied by 10^4 . All $MAPE$ -related figures have been multiplied by 10^2 .

Module (i)	\overline{MSPE}_i	$median(MSPE_i)$	$std(MSPE_i)$	Type
(1)	0.153	0.104	0.115	OTM, LIV
(2)	<u>1.07</u>	<u>0.533</u>	<u>1.06</u>	NTM, LIV
(3)	<u>26.4</u>	<u>6.23</u>	<u>39.61</u>	ITM, LIV
(4)	0.228	0.003	0.709	OTM, MIV
(5)	0.018	0.017	0.007	NTM, MIV
(6)	<u>1.26</u>	<u>0.670</u>	<u>1.27</u>	ITM, MIV
(7)	0.048	0.038	0.044	OTM, HIV
(8)	0.098	0.069	0.110	NTM, HIV
(9)	0.104	0.084	0.094	ITM, HIV
Total MNN	0.357	0.290	0.264	

Table 4: PREDICTION PERFORMANCE OF THE MNN OPTION PRICING MODEL WITH 9 MODULES - MONEYNES AND VOLATILITY.

Notes: The call options are classified according to the moneyness and implied volatility criteria. The average MSPE (\overline{MSPE}_i), its median value ($median(MSPE_i)$) and the standard deviation ($std(MSPE_i)$) have been obtained for each module i ($i = 1, \dots, 9$) and presented as averages across ten different random training seeds. The last row (“Total MNN”) represents the average measures of forecasting performance for the MNN option pricing model that combines all modules together and uses all data. The abbreviations for the option types are as follows: out-of-the-money (OTM, when $S_t/K < 0.98$), near-the-money (NTM, when $0.98 \leq S_t/K \leq 1.02$), in-the-money (ITM, when $S_t/K > 1.02$), low implied volatility (LIV, when $\sigma_{IV} < 0.12$), medium implied volatility (MIV, when $0.12 \leq \sigma_{IV} \leq 0.17$) and high implied volatility (HIV, when $\sigma_{IV} > 0.17$). All figures in the table have been multiplied by 10^4 . The largest MSPE values are underlined.

Module (i)	\overline{MSPE}_i	$median(MSPE_i)$	$std(MSPE_i)$	Type
(1)	0.080	0.047	0.113	ST, LIV
(2)	0.311	0.057	0.420	MT, LIV
(3)	<u>1.70</u>	<u>1.47</u>	<u>0.612</u>	LT, LIV
(4)	0.037	0.004	0.102	ST, MIV
(5)	0.341	0.001	1.04	MT, MIV
(6)	0.165	0.054	0.218	LT, MIV
(7)	0.028	0.016	0.042	ST, HIV
(8)	0.021	0.019	0.020	MT, HIV
(9)	0.037	0.029	0.026	LT, HIV
Total MNN	0.153	0.102	0.150	

Table 5: PREDICTION PERFORMANCE OF THE MNN OPTION PRICING MODEL WITH 9 MODULES - MATURITY AND VOLATILITY.

Notes: The call options are classified according to the maturity and implied volatility criteria. The average MSPE (\overline{MSPE}_i), its median value ($median(MSPE_i)$) and the standard deviation ($std(MSPE_i)$) have been obtained for each module i ($i = 1, \dots, 9$) and presented as averages across ten different random training seeds. The last row (“Total MNN”) represents the average measures of forecasting performance for the MNN option pricing model that combines all modules together and uses all data. The abbreviations for the option types are as follows: short-term (LT, when $\tau < 0.10$), medium-term (MT, when $0.10 \leq \tau \leq 0.20$), long-term (LT, when $\tau > 0.20$), low implied volatility (LIV, when $\sigma_{IV} < 0.12$), medium implied volatility (MIV, when $0.12 \leq \sigma_{IV} \leq 0.17$) and high implied volatility (HIV, when $\sigma_{IV} > 0.17$). All figures in the table have been multiplied by 10^4 . The largest MSPE values are underlined.

Module (i)	\overline{MSPE}_i	$median(MSPE_i)$	$std(MSPE_i)$	Type
(1)	0.003	0.002	0.004	ST, OTM
(2)	0.003	0.002	0.003	ST, NTM
(3)	0.123	0.084	0.137	ST, ITM
(4)	0.002	0.000	0.005	MT, OTM
(5)	0.026	0.004	0.047	MT, NTM
(6)	0.007	0.003	0.008	MT, ITM
(7)	0.230	0.197	0.170	LT, OTM
(8)	0.307	0.328	0.098	LT, NTM
(9)	<u>0.491</u>	<u>0.295</u>	<u>0.450</u>	LT, ITM
Total MNN	0.075	0.066	0.020	

Table 6: PREDICTION PERFORMANCE OF THE MNN OPTION PRICING MODEL WITH 9 MODULES - MONEYNES AND MATURITY.

Notes: The call options are classified according to the maturity and moneyness criteria. The average MSPE (\overline{MSPE}_i), its median value ($median(MSPE_i)$) and the standard deviation ($std(MSPE_i)$) have been obtained for each module i ($i = 1, \dots, 9$) and presented as averages across ten different random training seeds. The last row (“Total MNN”) represents the average measures of forecasting performance for the MNN option pricing model that combines all modules together and uses all data. The abbreviations for the option types are as follows: short-term (LT, when $\tau < 0.10$), medium-term (MT, when $0.10 \leq \tau \leq 0.20$), long-term (LT, when $\tau > 0.20$), out-of-the-money (OTM, when $S_t/K < 0.98$), near-the-money (NTM, when $0.98 \leq S_t/K \leq 1.02$) and in-the-money (ITM, when $S_t/K > 1.02$). All figures in the table have been multiplied by 10^4 . The largest MSPE values are underlined.

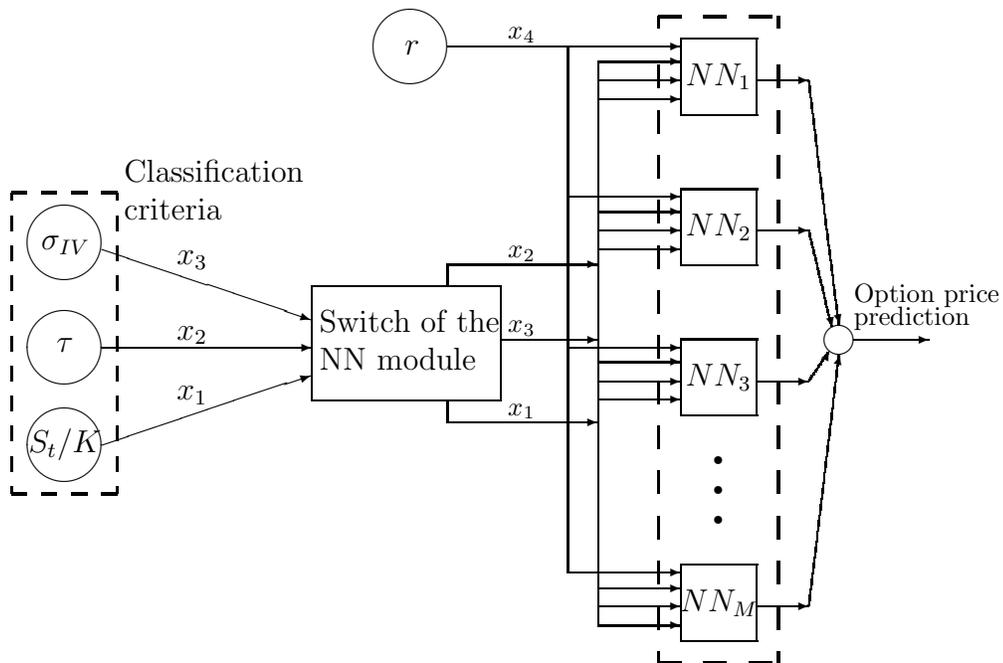


Figure 1: Multi-criteria modular neural network model.

Notes: The pricing function is decomposed into M separate non-linearities (NN_1, \dots, NN_M) called modules. The modules are trained independently on the data for the M option types and during prediction, based on the values of σ_{IV} , τ and S_t/K , only one NN is active (i.e., when $\omega_i = 1$, all $\omega_k = 0$; $k \neq i$, $k = 1, \dots, M$). More specifically, the out-of-sample predictions (C_t/K) are generated from 18 different NN models that span the whole testing set.

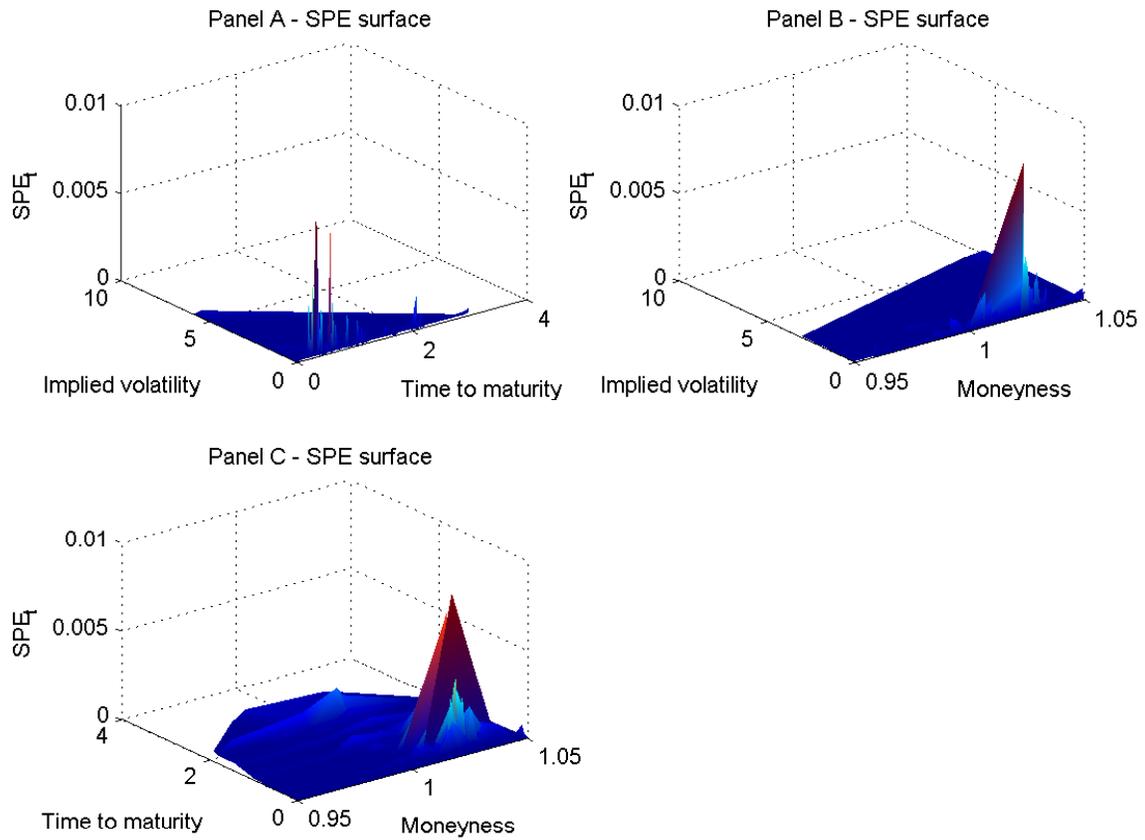


Figure 2: ERROR SURFACE OF THE MCMNN OPTION PRICING MODEL.

Notes: Out-of-sample squared errors ($SPE_t = (c_t - \hat{c}_t)^2$, $t = 1, \dots, 10457$) for the last quarter of 2012 are plotted against: 1) implied volatility (σ_{IV}) and maturity (τ) in Panel A, 2) implied volatility (σ_{IV}) and moneyness (S_t/K) in Panel B, and 3) maturity (τ) and moneyness (S_t/K) in Panel C.

References

- Agliardi, E. and Agliardi, R. (2009). Fuzzy defaultable bonds. *Fuzzy Sets and Systems*, **160**(18), 2597–2607.
- Aït-Sahalia, Y. and Lo, A. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, **53**, 499–547.
- Andreou, P., Charalambous, C., and Martzoukos, S. (2008). Pricing and trading European options by combining artificial neural networks and parametric models with implied parameters. *European Journal of Operational Research*, **185**(3), 1415–1433.
- Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, **52**(5), 2003–2049.
- Barndorff-Nielsen, O. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society B*, **63**, 167–241.
- Bennell, J. and Sutcliffe, C. (2004). Black-Scholes versus artificial neural networks in pricing FTSE 100 options. *International Journal of Intelligent Systems in Accounting, Finance and Management*, **12**, 243–260.
- Bhat, H. and Kumar, N. (2012). Option pricing under a normal mixture distribution derived from the Markov tree model. *European Journal of Operational Research*, **223**(3), 762–774.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637–659.
- Bridle, J. S. (1989). *Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition*, pages 227–236. Springer-Verlag, New York.
- Carr, P. and Wu, L. (2004). Time-changed Lévy processes and option pricing. *Journal of Financial Economics*, **71**, 113–141.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263.
- Duan, J. (1995). The GARCH option pricing model. *Mathematical Finance*, **5**, 13–32.
- Elliott, R. and Siu, T. (2013). Option pricing and filtering with hidden Markov-modulated pure-jump processes. *Applied Mathematical Finance*, **20**(1), 1–25.

- Erikkson, A., Ghysels, E., and Wang, F. (2009). The normal inverse Gaussian distribution and the pricing of derivatives. *Journal of Derivatives*, **16**, 23–38.
- Feldkamp, L. and Puskorius, G. (1998). A signal processing framework based on dynamic neural networks with application to problems in adaptation, filtering, and classification. In *Proceedings of IEEE*, volume 86, pages 2259–2277.
- Garcia, R. and Gençay, R. (2000). Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics*, **94**(1/2), 93–115.
- Gençay, R. and Altay-Salih, A. (2003). Degree of mispricing with the Black-Scholes model and nonparametric cures. *Annals of Economics and Finance*, **4**, 73–101.
- Gençay, R. and Gibson, R. (2009). Model risk for European-style stock index options. *IEEE Transactions on Neural Networks*, **18**, 193–202.
- Gençay, R. and Qi, M. (2001). Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping and bagging. *IEEE Transactions on Neural Networks*, **12**(4), 726–734.
- Gradojevic, N. and Kukulj, D. (2011). Parametric option pricing: A divide-and-conquer approach. *Physica D: Nonlinear Phenomena*, **240**(19), 1528–1535.
- Gradojevic, N., Gençay, R., and Kukulj, D. (2009). Option pricing with modular neural networks. *IEEE Transactions on Neural Networks*, **20**(4), 626–637.
- Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357–384.
- Haykin, S. (1999). *Neural Networks*. Prentice Hall, Inc., NJ.
- Hutchinson, J. M., Lo, A. W., and Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, **49**(3), 851–889.
- Jang, J. R., Sun, C., and Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing*. Prentice Hall, Inc., NJ.
- Kukulj, D., Gradojevic, N., and Lento, C. (2012). Improving non-parametric option pricing during the financial crisis. In *2012 IEEE Conference on Computational Intelligence for Financial Engineering and Economics, CIFE 2012 - Proceedings*, pages 93–99.
- Martin, G., Forbes, C., and Martin, V. (2005). Implicit Bayesian inference using option prices. *Journal of Time Series Analysis*, **26**(3), 437–462.

- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, London.
- Melick, W. R. and Thomas, C. P. (1997). Recovering an asset's implied PDF from option prices: An application to crude oil during the gulf crisis. *Journal of Financial and Quantitative Analysis*, **32**(1), 91–115.
- Qi, M. and Maddala, G. S. (1996). Option pricing using artificial neural networks: The case of S&P 500 index call options. In A. P. N. Refenes, Y. Abu-Mostafa, J. Moody, and A. Weigend, editors, *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets*, pages 78–91, New York. World Scientific.
- Renault, E. (2010). *Econometrics of option pricing*, volume 2, pages 518–528. Wiley, New York.
- Ronco, E. and Gawthrop, P. J. (1995). Modular neural networks: A state of the art. Technical Report CSC-95026, Centre for System and Control, University of Glasgow.
- Yoshida, Y. (2003). The valuation of European options in uncertain environment. *European Journal of Operational Research*, **145**(1), 221–229.