UNIVERSITA' CA' FOSCARI VENEZIA
DIPARTIMENTO DI MATEMATICA APPLICATA

RENDICONTI PER GLI STUDI ECONOMICI QUANTITATIVI

2002 Volume unico

Titolo della rivista e ISSN

RENDICONTI PER GLI STUDI ECONOMICI QUANTITATIVI 1591-9773

Option pricing via regime switching models and multilayer perceptrons: a comparative approach

Monica Billio, Marco Corazza, and Michele Gobbo

[1] Department of Economic Sciences
University Ca' Foscari of Venice
Cannaregio, 873
30121 Venice - Italy
[2] Department of Applied Mathematics
University Ca' Foscari of Venice
Dordoduro, 3825/E
30123 Venice - Italy
[3] GRETA - Gruppi di Ricerca Economica Teorica e Applicata
San Marco, 3870
30124 Venice - Italy

Abstract. Several empirical investigations have emphasized that the Black and Scholes option pricing formula shows systematic biases when compared to the market prices. In order to face the determination of option pricing models alternative to the Black and Scholes one, in this paper we propose two approaches. In the first one we take into account a particular parametric stochastic model for volatility: a regime switching model for the dynamics of the logarithmic returns of the underlying. In the second approach we assume that the biases observed for the Black and Scholes prices depend on the violation of more than one of the hypotheses of the related model; in particular, as it is often difficult to a priori determine the violated assumptions, we consider a non-parametric tool: the multilayer perceptron artificial neural networks. Both the approaches we developed has been carefully tested on *call* options written on the UK stock index FTSE-100. The results we obtained show that the considered models allow improvements with respect to the Black and Scholes pricing (that we consider as benchmark). In short details: the regime switching models are able to explain the *smile effects* empirically observed, and the multilayer perceptron models show good capabilities in miming the market option pricing mechanism.

KEYWORDS: option pricing, Black and Scholes formula, regime switching model, multilayer perceptron artificial neural network, UK stock index FTSE-100 .

JEL CLASSIFICATION: G13, C51, G45.
AMS CLASSIFICATION: 91B28, 62P05, 82C32.

1 Introduction and motivations

Financial options are primary example of derivative securities which entitles the owner the ability - but not the obligation - to trade in the future the underlying securities at a given price, called strike or *exercise* price. In the last 30 years, options have known an exponential growth since they are used by operators for hedging versus various forms of risks, and for speculation purposes (see, for example, [Sartore, 1999]).

Much of the initial success, and of the following propagation of these financial instruments may be also explained by the existence of a valuation model which is easy to implement: the closed form option pricing formula proposed in [Black *et al.*, 1973] (on following: BS) for European *call* options, formula extended in various directions in [Merton, 1973]). However, the assumptions underlying the BS formula are rarely met. In fact, many empirical investigations have shown that the logarithmic returns of many financial time series are characterized by probability distributions which differ substantially from the normal one supposed by BS. In particular, among the other features, the empirical distributions are leptokurtic, asymmetric, and show fat tails; moreover, various forms of frictions (like, for instance, the transaction costs) exist in the markets, and a continuous trading is - of course - not possible.

The violation of some of the assumptions underlying the BS model necessarily involves that the model valuations show systematic biases when compared to the market prices. Several empirical analyzes realized to verify the accuracy of the BS formula (see, for example, [Rubinstein, 1994]) agree in evidencing two kinds of errors: the one with respect to the time to *maturity*, and the one with respect to the moneyness[1].

In spite of the mentioned biases, the BS valuation model is widely used for pricing standard European options because of its simplicity; in fact, it needs the estimation of only a parameter, the one concerning the market volatility. However, the operators adjust the BS prices in order to correct both the time to *maturity* and the moneyness biases; in this way they replicate the so called *smile effect* empirically observable. By utilizing such an adjusted valuation formula, the operators "approximate" an (unknown) option pricing model which differs form the BS one; in fact, by so doing, they operate as if the dynamics of the price of the underlying follows a stochastic process different from the geometric Brownian motion assumed in the BS valuation models. In other words, the operators continuously deal with the determination

[1] The moneyness is given by the ratio between the current price of the underlying asset and the *exercise* price.

of models, alternative to the BS one, able to properly incorporates the features characterizing the option markets.

In order to face the determination of option pricing models alternative to the BS one, in this paper we propose the following two approaches:

- a first one in which we take into account a particular parametric stochastic model for volatility: a regime switching model (on following: RSM) for the dynamics of the logarithmic return of the underlying;
- a second one in which we assume that the biases observed for the BS prices depend on the violation of more than one of the hypotheses of the related model; in particular, as it is often difficult to a priori determine the violated assumptions, we consider a non-parametric tool: the multilayer perceptron (on following: MLP) artificial neural networks.

Both the approaches we developed has been carefully tested on the *call* options written on the United Kingdom (on following: UK) stock index FTSE-100. The results we obtained show that the considered models allow improvements with respect to the BS pricing (that we consider as benchmark). In short details:

- the RSMs are particularly able to explain the *smile effects* empirically observed[2];
- the MLP models show good capabilities in miming the market option pricing mechanism.

The remainder of the paper is organized as follows: in section **2** we shortly recall some basics on the option pricing; in sections **3** and **4** we, respectively, introduce the RSMs and MLP models, we describe the related pricing procedures, and we present the results; finally, in section **5** we give some final remarks.

2 Stock index written option pricing: a short recall

As well known, financial options are derivative securities that give the right to buy (*call* options), or to sell (*put* options), at (*European*-like options), or within (*American*-like options), a certain *expiry* date T a

[2] Notice that, in some cases, even the use of "standard" mixture of Gaussian probability distribution results restrictive.

given underlying asset for an agreed amount X (*exercise* price) (see, for details, [Hull, 2003]).

The standard pricing approach for *European*-like *call* options is founded on the seminal valuation model proposed in [Black *et al.*, 1973]; in particular, in this model one assumes that the price of the underlying asset $S(t)$ is driven by the following continuous-time stochastic dynamics:

$$dS(t) = \mu S(t)dt + \sigma S(t)dz,$$

where
μ is the instantaneous expected return rate;
σ is the standard deviation of the instantaneous expected return rate;
z is a random variable following a Wiener process.

A stock index is a suitable weighted average of the prices of single stocks; because of it, the related option pricing formula is not particularly different from the one originally proposed in [Black *et al.*, 1973]. The main difference consists in the fact that the stocks composing an index usually pay dividends, and that the value of such an index is not always adjusted for considering the related dividend effect. In details, the pricing formula for *European*-like *call* options characterized by continuous distribution of dividends at a known and constant rate q is the following one, given in [Merton, 1973]:

$$C_{BS}(t) = S(t)e^{-q\tau}\Phi(d_1) - Xe^{-r\tau}\Phi(d_2) \tag{1}$$

where
$C_{BS}(t)$ is the BS-like price at time t;
$\tau = T - t$ is the time to *maturity*, in which T is the *expiration* date;
$\Phi(d) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{d} e^{-0.5t^2}du;$
$d_1 = \dfrac{\log\left(S(t)/X\right) + \left(r - q + \sigma^2/2\right)\tau}{\sigma\sqrt{\tau}}$, in which r is the free risk interest rate;
$d_2 = d_1 - \sigma\sqrt{\tau}$.

The "inputs" of this BS-like pricing formula are: the underlying asset price $S(t)$ (in our case the UK stock index FTSE-100); the *exercise* price X; the time to *maturity* τ; the free risk interest rate r; the dividend yield rate q; the volatility σ.

The first three "inputs" can be easily obtained by consulting financial journals[3].

With regard to the assessment of the free risk interest rate, in the BS approach r is assumed known and constant during the lifetime of the option. But, in practice, the free risk interest rate varies over time. So, in order to estimate a proper r, at first we took into account the interbank interest rates[4] at one week, at one month, at three months, at six months, and at twelve months; then we determined a free risk interest rate for each option time to *maturity* by linearly interpolating the neighbouring interbank interest rates.

As far as it is concerned the dividend yield rate, we obtained its time series by consulting the web site of the London Stock Exchange[5]. In particular, a preliminary investigation of the data pointed out that the dividend yield rate is nearly constant for the lifetime of all the options we took into account; because of that, we can consider q constant.

With regard to the volatility, its estimation was obtained by averaging suitable implicit standard deviations achieved from the market option prices. In particular, for a given time to *maturity*, we calculated the volatility by a weighted average of the implicit standard deviations determined from the market price time series of the two options whose *exercise* prices are closest to the price of the underlying of the investigated option; by so doing, we conjecture to be able to mitigate the "impact" of the noise present in the data on the option pricing. Moreover, if options with different times to *maturity* were traded in the same day, then the volatility estimate was obtained by a further weighted average of the volatilities calculated in correspondence of each of the time to *maturity*. By the use of the latest weighted average, we conjecture to be able to provide meaningful estimations of the volatility whatever is the time to *maturity*. In any case, notice that

- at-the-money *call* options[6] are usually the most traded; therefore, their implicit standard deviation incorporates highly reliable information;

[3] Notice that the stock market and the option one close at the same time; so, it is assured the simultaneous determination of both the daily closing prices.

[4] In details, the London InterBank Offer Rate (LIBOR) on sterling.

[5] The address is "http://www.londonstockexchange.com".

[6] A *call* option is said to be out-of-the-money if $S(t) < X$; it is said to be at-the-money if $S(t) \approx X$; it is said to be in-the-money if $S(t) > X$.

- at-the-money *call* option prices are highly sensitive to change in volatility; therefore, poor estimation of the volatility itself involves significative error in pricing this category of option.

Finally, as far as it is concerned the "output" of the BS-like pricing formula (1), we consider the daily closing prices of the European-like *call* options written on the UK stock index FTSE-100, traded at the London International Financial Futures and Options Exchange (also known as LIFFE) from January 1, 1999 to December 27, 1999 (source: Bloomberg).

3 Option pricing via RSMs

Simply speaking, when RSMs are utilized in order to model the stochastic behaviour of a given financial asset (in our case the UK stock index FTSE-100), generally one assumes that the returns of this asset are characterized by a suitable mixture of Gaussian probability distributions with different means and different standard deviations, whose transition probabilities from each to other depend on a non observable process describable by a Markov chain (see, for details, [Hamilton, 1994] and [Krolzig, 1997]). Therefore, an option valuation model based on a switching regime approach is necessarily a stochastic volatility pricing model[7].

3.1 RSMs: some basics

Suppose that the returns of a given financial asset are distributed as

$$y_t = \mathcal{N}\left(\mu_{s(t)}, \sigma_{s(t)}\right), \text{ with } s(t) = 1, \ldots, i, \ldots, N,$$

where
$s(t)$ is the outcome of an unobservable $N-$ state Markov chain characterized by the transition probability matrix which follows

$$P = \begin{pmatrix} p_{11} & \cdots & p_{i1} & \cdots & p_{N1} \\ \vdots & & \vdots & & \vdots \\ p_{1j} & \cdots & p_{ij} & \cdots & p_{Nj} \\ \vdots & & \vdots & & \vdots \\ p_{1N} & \cdots & p_{iN} & \cdots & p_{NN} \end{pmatrix},$$

[7] See, for a review of the stochastic volatility pricing models, [Billio *et al.*, 2003].

in which p_{ij} is the probability that the $i-$ th state is followed by the $j-$ th one.

In order to price the *European*-like *call* options written on the UK stock index FTSE-100, we proceed as follows: at first, we model via RSMs the stochastic behaviours of the daily returns of the considered stock index; then, we utilize this modeling for developing a suitable numerical approach for pricing the considered options (see the next subsection).

With regard to the modeling of the behaviour of the daily returns of the investigated stock index, we take into account RSMs characterized by two regimes[8] ($N = 2$); in both regimes the stock index returns are distributed in Gaussian way with constant mean $\mu_{s(t)}$ and constant standard deviation $\sigma_{s(t)}$, with $s(t) = 1, 2$. In particular, the first model we consider (that we call *RSM2-M-SD*) is

$$y_t \sim \begin{cases} \mathcal{N}(\mu_1, \sigma_1) \text{ if } s(t) = 1 \\ \mathcal{N}(\mu_2, \sigma_2) \text{ if } s(t) = 2 \end{cases} \tag{2}$$

with transition probability matrix:

$$P = \begin{pmatrix} p_{11} & p_{21} = 1 - p_{11} \\ p_{12} = 1 - p_{22} & p_{22} \end{pmatrix};$$

the relative parameters to estimate are: μ_1, μ_2, σ_1, σ_2, p_{11}, and p_{22}. The second - and last - model we take into account (that we call *RSM2-SD*) is a simplified version of the RSM (2),

$$y_t \sim \begin{cases} \mathcal{N}(\mu, \sigma_1) \text{ if } s(t) = 1 \\ \mathcal{N}(\mu, \sigma_2) \text{ if } s(t) = 2 \end{cases},$$

with transition probability matrix analogous to the previous one; now, the relative parameters to estimate are: μ, σ_1, σ_2, p_{11}, and p_{22}.

The results relative to the models *RSM2-M-SD* and *RSM2-SD* are reported in Tables 1 and 2, respectively. Notice that in both the models, mainly the standard deviation allows to discriminate the two regimes, that is allows to distinguishes between normal market phases (characterized by "low" volatility) and turbulent market phases (characterized by "high" volatility)[9].

[8] During the empirical analyses we also tested RSMs characterized by three regimes, but without any meaningful improvement in the experimental outcomes. Moreover, notice that the number of parameters to estimate increases as the number of the regimes grows up.

[9] Our results are similar to the ones presented in [Shaller *et al.*, 1997].

Table 1. Estimates of the parameters of model *RSM2-M-SD*

Parameter	μ_1	μ_2	σ_1	σ_2	p_{11}	p_{22}
Estimate	0.0000822	0.0005973	0.0173909	0.01050513	0.9860428	0.9960097
Standard error	0.0016793	0.0005997	0.0013850	0.00046825	0.0143420	0.0046330
t-value	0.0489891	0.9959378	12.556619	22.43481700	68.7521150	214.979110

Table 2. Estimates of the parameters of model *RSM2-SD*

Parameter	μ	σ_1	σ_2	p_{11}	p_{22}
Estimate	0.00052840	0.01737708	0.01050302	0.98650976	0.99614086
Standard error	0.00054660	0.00133744	0.00045973	0.01322692	0.00437072
t-value	0.96670556	12.99277100	22.84623400	74.58347200	227.91225000

3.2 Option evaluation phase

Starting from the parameters whose estimates are reported in the previous subsection, here we implement a numerical approach - methodologically based on the multinomial one proposed in [Cox *et al.*, 1979] (on following: CRR) - in order to price the investigated options.

As known, in the CRR model one assumes that the price of the underlying asset is driven by a discrete-time multiplicative binomial process, *i.e.*

$$S_{t+1} = \begin{cases} S(t) \cdot u \cdot p \\ S(t) \cdot d \cdot (1-p) \end{cases}$$

where
$u \in (1, +\infty)$ and $d \in (0, 1)$ are the usual multiplicative factors;
p and $1-p$ are, respectively, the probabilities the price $S(t)$ increases and decreases in the next time instant $t+1$.

Given this model for the price dynamics of the underlying asset, the CRR numerical approach for option valuation can be itemized as follows:

<u>step 1</u>: by using the current underlying asset price S_0, the multiplicative factors u and d, and the probabilities p and $1-p$, one generates the tree of the asset prices from t to *maturity*;
<u>step 2</u>: for each underlying asset price at *maturity*, one calculates the relative option payoff at the *expiration* date;

step 3: by suitably folding back all the option payoffs calculated at the *expiration* date, one determines the current option value.

In order to implement our discrete-time numerical approach for option pricing when the dynamics of the returns of the underlying asset are describable by a RSM, notice that

- at the *expiration* date, the underlying asset returns are not distributed in Gaussian way but as a mixture of Gaussian probability distributions;
- in folding back the option payoffs from the *expiration* date to the current time, the transition probabilities of switching between the two considered regimes have to be take into account;
- the transition probabilities are relative to a source of uncertainty which is new with respect to the ones considered in the usual models for option pricing: the determination of the probability distribution characterizing the current and the future dynamics of the underlying asset returns. Because of it, in the RSM environment it is not possible to determine an hedging portfolio; moreover, all the "simplifications" coming from the use of standard risk neutrals arguments are not more reasonable.

In our option evaluating approach we operate in a risk neutral framework[10], assuming that the risks additionally coming from the source of uncertainty reported in the latest point are not priced in the market. Thus, the discrete-time processes related to the returns are determined by setting the means of each regime distribution equal to $r - q$, and the discount phase occurs at the interest rate r. In order to assure the convergence of the implied discrete-time processes, and in order to obtain a good approximation of the probability distribution of the underlying asset returns, we take into account a pentanomial tree[11].

In correspondence of each intermediate node of the considered tree, the value of the option is determined by following the usual CRR approach. In particular, in our option pricing model we have also to take into account the probabilities to transit from a regime to another; because of that, at each intermediate node we determine two different prices of the option, both conditional to the regime generating the data (*i.e.* the "low" variance regime, or the "high" variance regime).

[10] Notice that this assumption is usual in such approaches.
[11] See, for related models, [Bollen, 1999] and [Billio *et al.*, 1997].

Formally, in correspondence of the generic intermediate node, one has:

$$
\begin{aligned}
&C\left(t|S(t)=1\right)= \\
&\quad\left[p_{11}\cdot C\left(t+1|s_{t+1}=1\right)+\left(1-p_{11}\right)\cdot C\left(t+1|s_{t+1}=2\right)\right]e^{-r}, \\
&C\left(t|S(t)=2\right)= \\
&\quad\left[\left(1-p_{22}\right)\cdot C\left(t+1|s_{t+1}=1\right)+p_{22}\cdot C\left(t+1|s_{t+1}=2\right)\right]e^{-r}.
\end{aligned}
$$

Of course, also at the starting node we determine two different conditional option values, $C\left(0|s_0=1\right)$ and $C\left(0|s_0=2\right)$; in order to obtain the (unique) price of the option in $t=0$, $C_{RSM}\left(0\right)$, we utilized the pricing approach proposed in ([Bollen, 1999]), following which

$$C_{RSM}\left(0\right)=p_1\cdot C\left(0|s_0=1\right)+p_2\cdot C\left(0|s_0=2\right),$$

where
$C_{RSM}(t)$ is the RSM price at time t;
p_1 and p_2 are the unconditional probabilities.

Notice that the choice to use p_1 and p_2 as weights in the previous pricing formula is criticizable because the unconditional probabilities can significantly vary over time; an alternative possibility consists in using as weights the filtered probabilities (see, for details, [Billio *et al.*, 2003]).

3.3 Results

In order to compare the RSM performances with the BS ones and with the market happenings, we consider the relative implied volatilities. In particular, here we present some applications representative of the whole data set relative to the out-of-sample options priced.

Generally speaking, the results we obtained are not particularly satisfying. In fact, from Figures 1 and 2 one notices that the RSM implied volatilities are not always close enough to the market ones, and that their deviations are - on the average - similar to the deviations shown by the BS implied volatilities; moreover, the RSM pricing models only partially are able to replicate the market *smile effect*. We conjecture that such results strongly depend on the assumptions relative to the modeling of the underlying asset price dynamics; in particular, we deem necessary to relax at least the assumption relative the constant transition probabilities.

Another remarkable feature characterizing the RSM implied volatilities consists in the fact that they underestimate - on the average

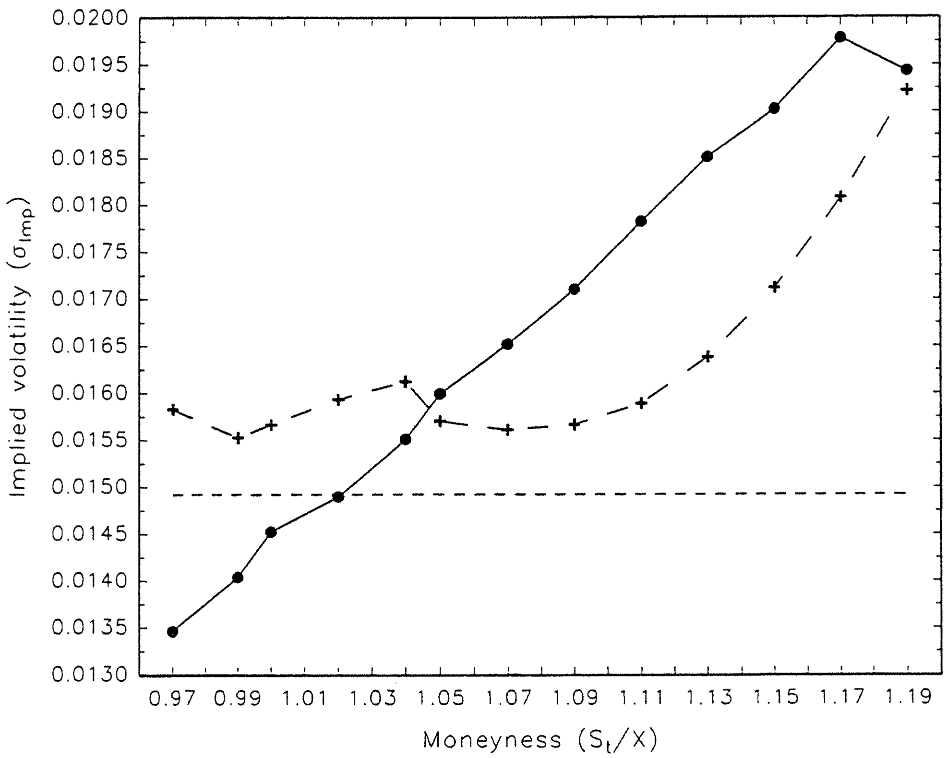

Fig. 1. Market implied volatility: solid line; BS implied volatility: horizontal line; RSM implied volatility: dashed line.

- the market ones; therefore, the market option prices are - on the average - greater than the ones coming from our RSMs. Some explanations about this topic are given in the literature (see, for example, ([Bates, 1996]). In particular, we recall that we developed our RSM in a risk neutral framework, assuming that the risks additionally coming from the determination of the probability distribution characterizing the dynamics of the underlying asset returns are not priced in the market; because of that, the investors could require an additional risk premium in order to compensate the risks associated to the RSM approach. Therefore, such differences in volatility risk premiums could explicate the smaller option prices determined by our RSMs with respect to the market ones.

4 Option pricing via MLP models

Generally speaking, the MLP models are quantitative tools originally inspired to the biological neural networks and to their learning capabilities. In particular, starting from a data set of input-output pat-

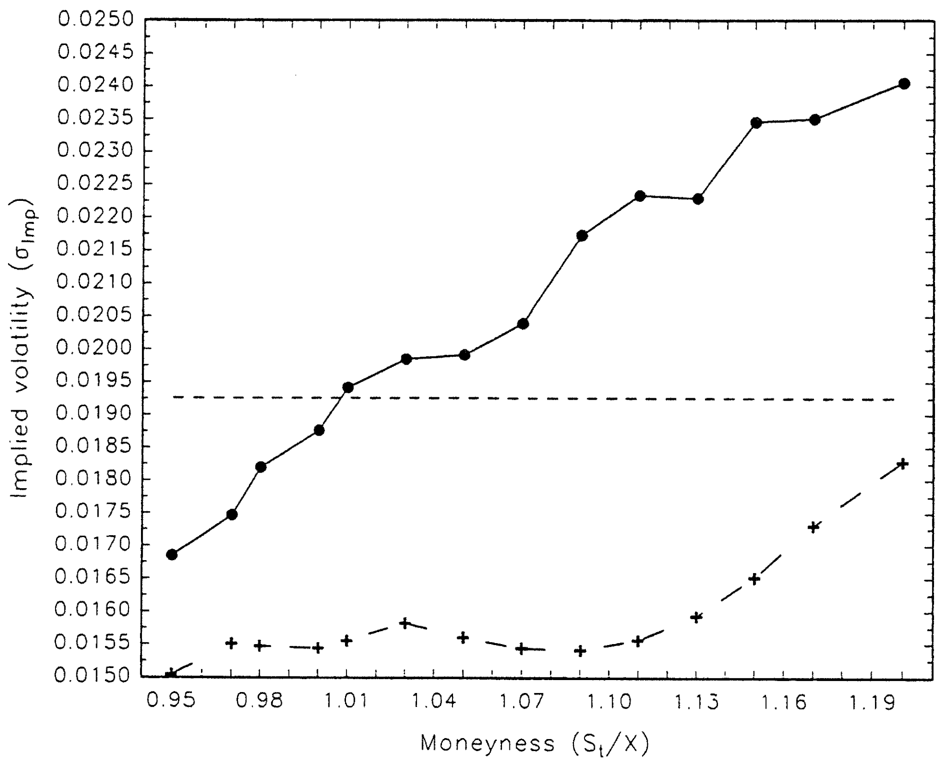

Fig. 2. Market implied volatility: solid line; BS implied volatility: horizontal line; RSM implied volatility: dashed line.

terns generated by an unknown process, the MLP models are able to represent, in a non-parametric way, such a process (see, for details, [Hornik *et al.*, 1989]).

More in details, these artificial neural networks have the same structure of direct arch weighted graphs whose nodes are arranged in one input layer, one or more hidden layers[12], and one output layer. Each layer is then fully connected with the subsequent one (the input layer with the first hidden layer, the first hidden layer with the second hidden layer, ..., and the last hidden layer with the output layer) without any intra-layer connection. Because of that, the input signal is feed-forwardly propagated from the input layer through the hidden ones to the output layer.

Each single node, except the input ones, is able to perform data calculation and transformation. In particular, each "computing" node is characterized by two functions: the first one determines the node

[12] In general, the number of nodes of each hidden layer can not be a priori determined.

input by calculating a suitable aggregation of the signals incoming in the node itself[13]; the second function determines the node output by suitably transforming the node input[14] (see, for more details, [Hertz *et al.*, 1991]).

Thanks to such a parallel and distributed structure, the MLP models show generalization capabilities after the "learning", and robustness to the noise in the data; because of that, these models seem to be particularly able in performing financial input-output patterns.

With regard to our option pricing problem, the MPL approach presents some advantages with respect to several parametric models (see, for example, [Hutchinson *et al.*, 1994]):

- the MLP approach does not need a priori detailed assumptions relative to the variables of the investigated input-output relationship;
- the MLP models are robust to specification errors;
- the MLP approach provides an approximate closed form option pricing formula (see subsection **4.2**).

On the other hand, the MLP models require large data set of historical input-output patterns in order to achieve successful results; moreover, as the MLP models are black-box tools, the interpretation of the approximate closed form formulas provided by these models are generally not easy to give.

As far as it is concerned the set-up of a MLP model, several aspects have to be taken into account. In particular, among the other topics (see, for more details, [Belcaro *et al.*, 1996]):

- the preprocessing of the data set of the input-output patterns generated by the unknown process;
- the determination of the optimal MLP model architecture;
- the selection of the training and testing procedures (like, for instance, the choice of the learning algorithm, the minimization of the generalization error[15], ...).

---

13 The most common used aggregation function is the summation one.

14 One of the most commonly used transformation function is the logistic one.

15 The generalization error is the error an MLP model should ideally commit in representing the input-output relationship if it were "learning" from the data set of all the possible input-output patterns, *i.e.* from the universe data set. Of course, as the available data set is always a subset of the universe one, the generalization error is always unknown. In order to avoid any biasing possibility coming from such an "ignorance", we utilize a learning criterion based on the so called *concurrent descent methodology*: first, the input-output data set is suitably

4.1 Preprocessing and modeling phases

MLP models are data-driven tools; because of it, in order to determine their optimal architecture is essential to perform a "good" preprocessing phase. In particular, in developing the MLP pricing model, we devote our attention to the selection of a suitable subset D from the available data set of the historical input-output patterns. By so doing, we can exclude from the processing phase the uninformative input-output patterns; moreover, as positive consequence, the reduction of the size of the data set to process implies the reduction of the related computational costs.

In details, the criteria we utilized to exclude the uninformative patterns are similar to the ones proposed in [Anders *et al.*, 1998]:

- exclusion of the deep-out-of-the-money and the deep-in-the-money options because they are usually poorly traded;
- exclusion of the options characterized by less than 15 days to *maturity* because they are usually traded at their intrinsic value;
- exclusion of the options traded at prices smaller than 100 basis points because there are several, and quite different, options traded at a same low price.

Finally, starting from the "reduced" input-output data set, we split it - as usual in the MLP applications - in the following subsets:

- a training one, containing 2,000 input-output patterns selected in a random way from the "reduced" data set D;
- a validation one, containing 500 input-output patterns selected in a random way from the "reduced" data set D;
- an out-of-sample testing one, containing 530 input-output patterns relative to the options traded in the last time period considered in our investigation (November 1, 1999 to December 24, 1999)[16].

In general terms, with respect to the non-parametric representation of the unknown process generating the input-output patterns, it can be simply formalized as follows:

---

split in two not-overlapping subsets (the training and the validation ones); then, the training algorithm is iterated on the training subset as long as the (absolute) minimum of a pre-establish cost function on the testing subset is reached (see, for more details, subsection **4.2**).

[16] Usually, the performances of a MLP model are evaluated by using the patterns belonging to this subset.

$$C_{ANN}(t) = f\left(S(t), X, \tau = T - t, \sigma_{Imp}, r, q\right) \quad (3)$$

where
$C_{ANN}(t)$ is the ANN price at time t;
σ_{Imp} is the implied volatility.

Now, in order to make as easy as possible the computational task of the learning algorithm, we proceed by dropping the number of inputs (and, consequently, by reducing the number of the MLP model parameters to estimate) on the basis of the following remarks:

- in our MLP option pricing approach we operate in a risk neutral framework; thus, we can incorporate the dividend effect in the free risk interest rate by setting this interest rate equal to $r - q$;
- in our modeling it is reasonable to assume the independence between the prices and the return probability distribution of the underlying stock index; because of that, the hypotheses of the **Theorem 9** presented in [Merton, 1973] hold, and the related closed form option pricing formulas results homogeneous of degree 1 both in the asset price, and in the *exercise* one (with respect to the *exercise* price itself).

So, the non-parametric representation (3) can be sinthetically reformulated as follows:

$$\frac{C_{ANN}(t)}{X} = f\left(\frac{S(t)}{X}, 1, T - t, \sigma_{Imp}, r - q, 0\right) =$$
$$= g\left(\frac{S(t)}{X}, T - \tau, \sigma_{Imp}, r - q\right).$$

4.2 Learning phase in short

In general, the optimal values of the weights of a MLP model are determined by using an iterative estimation procedure. The most common learning algorithms are based on the *error back-propagation* method; in such algorithms, after a starting random initialization of the weights, the input-output patterns belonging to the training set are iteratively presented to the MLP model as long as, by a proper weight updating, the (absolute) minimum of a pre-established cost function is reached[17]. The weight updating is obtained by suitably

[17] Usually, the cost function defines a distance measure between the MLP model output and the desired one.

back-propagating the cost function value, *i.e.* the output error level, from the output layer through the hidden ones to the input layer (see, for more details, [Hertz *et al.*, 1991] and [Belcaro *et al.*, 1996]).

In these learning algorithms a crucial role is played by the stop-learning criterion. The classical version of this criterion is satisfied at the training algorithm iteration in which the calculated error level is not higher than a pre-established threshold. An unfavorable phenomenon associated to this version of the stop-learning criterion is the so called over-training (or over-fitting) problem, *i.e.* the possibility that the learning algorithm detect unexisting relationships between the inputs and the outputs; in such a case, the generalization error function does not result minimized[18]. In order to avoid this possibility, several Authors suggest to use a stop-learning criterion based on the cross validation technique known as *concurrent descent methodology*.

In implementing our MLP pricing model, we considered an only hidden layer; the numerical optimization algorithm we used at each step of the *error back-propagation* iterative procedure is the Levenberg-Marquardt one; as cost function we utilized the standard *root mean square error* (on following: RMSE) one; in order to avoid the over-training problem, we took into account the *concurrent descent methodology*.

4.3 Results

In order to determine the optimal MLP model architecture, we developed several one-hidden-layer MLP models, each of them characterized by a different number of hidden artificial neurons. From Table 3 it results that the smallest value reached by the RMSE function on the validation set is in correspondence of the MLP model with 8 hidden nodes (on following: MLP8). Then, in order to compare the performances of the best MLP model, MLP8, with the BS ones, we calculated for both these pricing approaches the value reached by the RMSE function on the out-of-sample testing set: $1.155 \cdot 10^{-3}$ for MLP8, and $4.242 \cdot 10^{-3}$ for the BS formula.

Such satisfying results we obtained from the out-of-sample analysis can be partially explicated, beyond the effectiveness of the learning capabilities of the MLP models, also by the feature that the out-of-sample options priced are "similar" among them; in fact, they are all

[18] Recall that such a minimization is the goal of the learning phase.

traded in the last time period considered in our investigation (November 1, 1999 to December 24, 1999). So, in order to deepen the comparison between the MLP8 performances and the BS one, we extend our analysis to all the options traded in the first period taken into account in our investigation (January 1, 1999 to October 31, 1999), included the input-output patterns which had been excluded from the processing phase as uninformative (see subsection **4.1**). In particular, we detail this new analysis both with respect to the moneyness (see Table 4) and with respect to the time to *maturity* (see Table 5).

As far as it is concerned the results coming from this new analysis:

- the MLP8 performances are always better than the BS ones, both with respect to the moneyness and with respect to the time to *maturity*;
- the behaviour of the RSME function is similar for both the considered pricing approaches: its values raises as time to *maturity* increases, and as moneyness moves away from 1.

Table 3. Value of the RMSE function on the training and the validation sets

# hidden nodes	On the training set	On the validation set
2	$3.796 \cdot 10^{-3}$	$3.680 \cdot 10^{-3}$
3	$1.818 \cdot 10^{-3}$	$1.909 \cdot 10^{-3}$
4	$1.597 \cdot 10^{-3}$	$1.736 \cdot 10^{-3}$
5	$1.459 \cdot 10^{-3}$	$1.604 \cdot 10^{-3}$
6	$1.440 \cdot 10^{-3}$	$1.604 \cdot 10^{-3}$
7	$1.376 \cdot 10^{-3}$	$1.514 \cdot 10^{-3}$
8	$1.341 \cdot 10^{-3}$	$1.451 \cdot 10^{-3}$
9	$1.331 \cdot 10^{-3}$	$1.466 \cdot 10^{-3}$

Table 4. Value of the RMSE function with respect to the moneyness $S(t)/X$

Moneyness	$S(t)/X$	# observations	MLP8	BS model
Deep-out-of-the-money	0.80-0.90	1,221	$1.531 \cdot 10^{-3}$	$9.169 \cdot 10^{-3}$
Out-of-the-money	0.90-0.97	1,815	$1.445 \cdot 10^{-3}$	$4.595 \cdot 10^{-3}$
Near-the-money	0.97-1.03	1,417	$1.241 \cdot 10^{-3}$	$3.263 \cdot 10^{-3}$
In-the-money	1.03-1.10	1,650	$1.404 \cdot 10^{-3}$	$6.460 \cdot 10^{-3}$
Deep-in-the-money	1.10-1.20	1,128	$1.543 \cdot 10^{-3}$	$7.603 \cdot 10^{-3}$
All the sample	0.80-1.20	7,231	$1.429 \cdot 10^{-3}$	$6.402 \cdot 10^{-3}$

Table 5. Value of the RMSE function with respect to the time to *maturity* τ

Time to *maturity*	Months	# observations	MLP8	BS model
Short-term	0-4	2,071	$1.311 \cdot 10^{-3}$	$3.764 \cdot 10^{-3}$
Medium-term	4-8	2,535	$1.415 \cdot 10^{-3}$	$6.560 \cdot 10^{-3}$
Long-term	8-12	2,625	$1.529 \cdot 10^{-3}$	$7.759 \cdot 10^{-3}$
All the sample	0-12	7,231	$1.429 \cdot 10^{-3}$	$6.402 \cdot 10^{-3}$

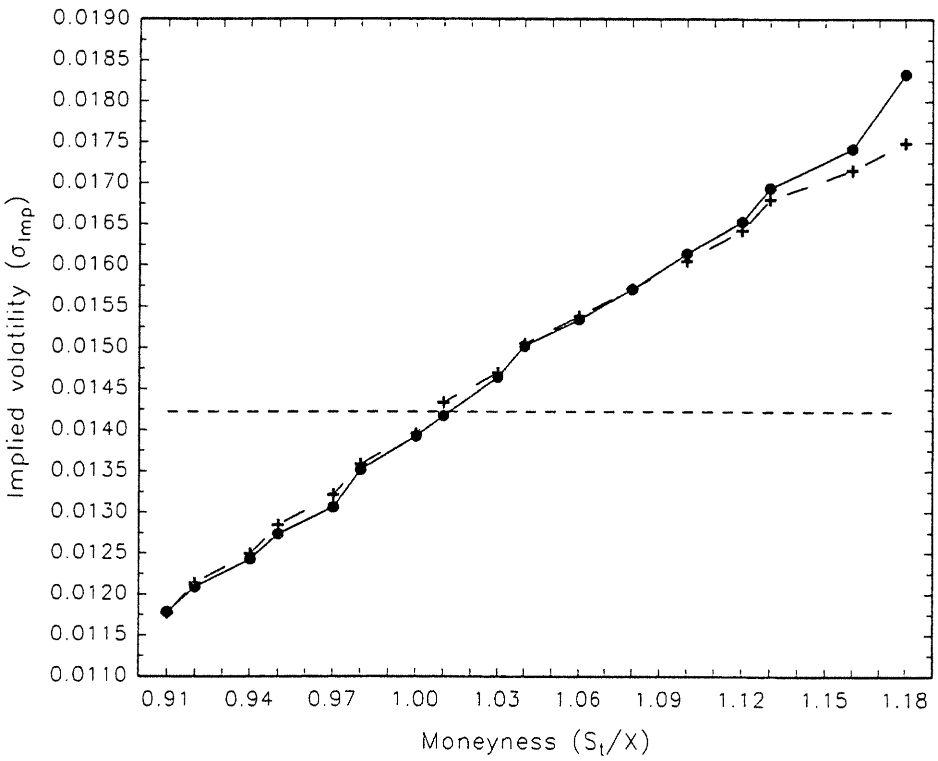

Fig. 3. Market implied volatility: solid line; BS implied volatility: horizontal line; RSM implied volatility: dashed line.

Finally, with regard to the comparison of the MLP8 implied volatility with the market one, notice that:

- in general, the MLP model is able to quite well replicate the dynamics of the market implied volatility (see Figures 3 and 4); by so doing, it shows its ability in correctly "learning" the features characterizing the market input-output relationships;
- the main errors the MLP model commits in replicating the market implied volatility are all in correspondence of poorly traded

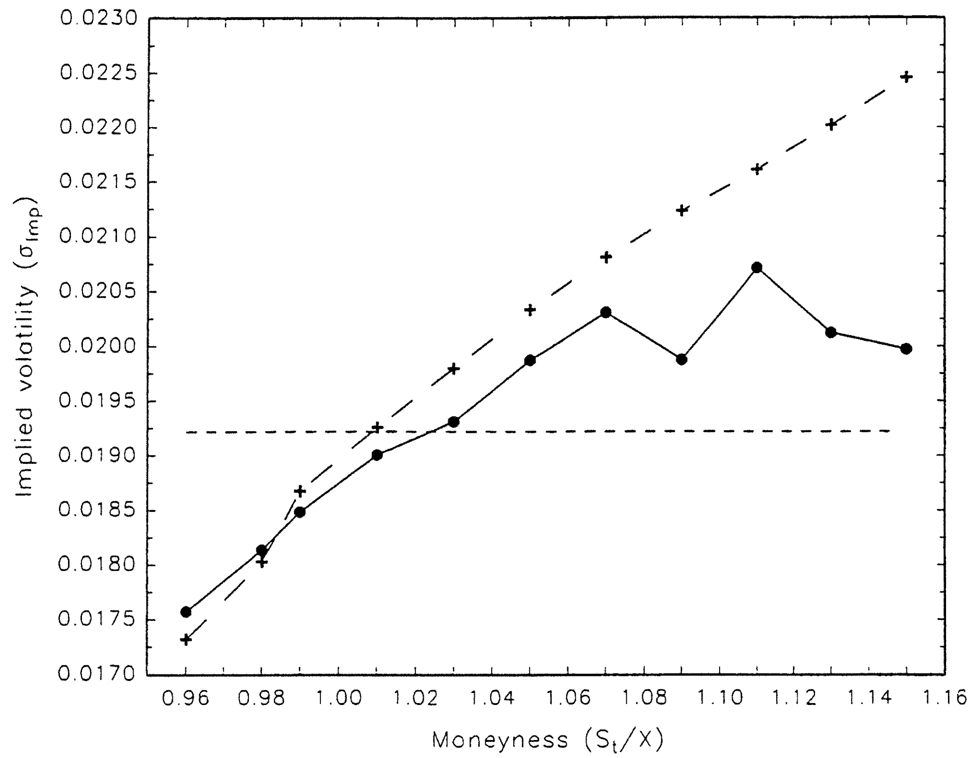

Fig. 4. Market implied volatility: solid line; BS implied volatility: horizontal line; RSM implied volatility: dashed line.

options, like the out-of-the-money and the in-the-money ones (see Figure 4); in such a case, the noise present in the data plays a crucial role.

5 Concluding remarks

In this work we face the determination of option pricing models alternative to the BS one. In particular, we propose two approaches: a parametric one based on the RSMs, and a non-parametric one based on the MLP models.

In general, a finished comparison between these two pricing approaches is not easy; moreover, it could be significantly affected by the available data set. In any case, notice that

- as far as it is concerned the RSM-based approach, results better than the one we obtained could be reached by detecting - and assuming - an appropriate model for the stochastic dynamics of the volatility, and by introducing a risk premium for the additional risks associated to the RSM approach itself;

– with regard to the MLP models, although the relative results are satisfying, they are black-box tools; therefore, the interpretation of the approximate close form formula provided by this approach is not easy to give; moreover, performances better than the one we obtained could be reached by using some of the so called *hybrid approach*, in which a MLP model and an efficient parametric pricing formula are jointly utilized.

References

[Anders *et al.*, 1998] Anders, U., Korn, O. and Schimitt, C. (1998). Improving the Pricing of Options: a Neural Network Approach. *Journal of Forecasting*, 17:369–3880.

[Bates, 1996] Bates, D.S. (1996). Testing Options Pricing Models. In Maddala, G.S. and Rao, C.R. (eds.) *Statistical Methods in Finance*. Elsevier.

[Belcaro *et al.*, 1996] Belcaro, P.L., Canestrelli E. and Corazza, M. (1996). Artificial Neural Network Forecasting Models: an Application to the Italian Stock Market. *Badania Operacyjne i Decyzje*, 3-4:29-48.

[Billio *et al.*, 1997] Billio, M. and Pelizzon, L. (1997), Pricing Options with Switching Volatility. In Hipp, C. (ed.) *Money, Finance, Banking and Insurance*. Verlag.

[Billio *et al.*, 2003] Billio, M. and Sartore, D. (2003), Stochastic Volatility Models: a Survey with Applications to Option Pricing and Value at Risk. In Dunis, C. and Laws, J. (eds.) *Quantitative Methods for Trading and Investment*. John Wiley & Sons.

[Black *et al.*, 1973] Black, F. and M. Scholes (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81:637–659.

[Bollen, 1999] Bollen, N. (1999). A Lattice for Valuing Options in Regime-Switching Models. *Journal of Derivatives*, 6:38–49.

[Cox *et al.*, 1979] Cox, J.C., Ross, S.A. and Rubinsten, M. (1979). Option Pricing: a Simplified Approach. *Journal of Financial Economics*, 7:229–263.

[Hamilton, 1994] Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.

[Hertz *et al.*, 1991] Hertz, J., Krogh, A. and Palmer, R. (1991). *Introduction to the Theory of Neural Computing*. Addison-Wesley.

[Hornik *et al.*, 1989] Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer Feedforward Networks are Universal Approsimators. *Neural Networks*, 2:359-366.

[Hull, 2003] Hull, J.C. (2003). *Options, Futures and Other Derivatives*. Prentice Hall.

[Hutchinson *et al.*, 1994] Hutchinson, J.M., Lo, A. and Poggio, T. (1994). A Non-parametric Approach to Pricing and Hedging Derivatives Securities via Learning Networks. *Journal of Finance*, 94:851-889.

[Krolzig, 1997] Krolzig, H.M. (1997). *Markov-Switching Vector Autoregression*. Springer.

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

[Rubinstein, 1994] Rubinstein, M. (1994). Implied Binomial Tree. *Journal of Finance*, 49:771-818.

[Sartore, 1999] Sartore, D. (1999). *Gli Strumenti Derivati. Analizzare, Prevedere e Coprire i Rischi Finanziari nelle Imprese.* IPSOA (in Italian).
[Shaller *et al.*, 1997] Shaller, H. and van Norden, S. (1997). Regime Switching in Stock Market Returns. *Applied Financial Economics*, 7:177-191.