DEMB Working Paper Series

N. 28

**A comparison of fuzzy regression methods for the estimation
of the implied volatility smile function**

Silvia Muzzioli[*]
A. Ruggeri[**]
Bernard De Baets[***]

November 2013

[*]   Department of Economics and CEFIN, University of Modena and Reggio Emilia, Italy
       silvia.muzzioli@unimore.it
[**]   Department of Economics and Social Sciences, Sapienza University of Rome
[***] KERMIT, Department of Mathematical Modelling, Statistics and Bioinformatics,
       Ghent University, Belgium

A comparison of fuzzy regression methods for the estimation of the implied volatility smile function

S. Muzzioli[a*], A. Ruggieri[b], B. De Baets[c]

[a]*Department of Economics and CEFIN,University of Modena and Reggio Emilia, Italy*

[b]*Department of Economics and Social Sciences, Sapienza University of Rome*

[c]*KERMIT, Department of Mathematical Modelling, Statistics and Bioinformatics,*
Ghent University, Belgium

Abstract

The information content of option prices on the underlying asset has a special importance in finance. In particular, with the use of option implied trees, market participants may price other derivatives, estimate and forecast volatility (see e.g. the volatility index VIX), or higher moments of the underlying asset distribution. A crucial input of option implied trees is the estimation of the smile (implied volatility as a function of the strike price), which boils down to fitting a function to a limited number of existing knots. However, standard techniques require a one-to-one mapping between volatility and strike price, which is not met in the reality of financial markets, where, to a given strike price, two different implied volatilities are usually associated (coming from different types of options: call and put).

In this paper we compare the widely used methodology of discarding some implied volatilities and interpolating the remaining knots with cubic splines, to a fuzzy regression approach which does not require an a-priori choice of implied volatilities. To this end, we first extend some linear fuzzy regression methods to a polynomial form and we apply them to the financial problem. The fuzzy regression methods used range from the possibilistic regression method of Tanaka, Uejima and Asai [14], the least squares fuzzy regression method of Savic and Pedrycz [13] and the hybrid method of Ishibuchi and Nii [4].

Keywords: Fuzzy statistics and data analysis, Finance, Fuzzy regression methods, Non-linear programming, Implied volatility.

---

* Corresponding author: Department of Economics, University of Modena and Reggio Emilia, V.le Berengario 51, 41121 Modena, Italy, tel. +39-059-2056771, fax. +39-059-2056947 e-mail: silvia.muzzioli@unimore.it.

1. Introduction

The information content of option prices on the underlying asset has a special importance in finance. In particular, with the use of option implied trees, market participants may price other derivatives, estimate and forecast volatility (see e.g. the volatility index VIX), or higher moments of the underlying asset distribution. An option gives the holder the right to buy (call option) or to sell (put option) a financial instrument (the underlying asset) for a pre-specified price (strike price), on a given date (expiry date). Option prices are quoted in the market for a discrete number (e.g. 15 in the Italian market) of different strike prices K, usually equally spaced, ranging from K_{min} (the minimum quoted strike price) to K_{max} (the maximum quoted strike price). An option is said to be at-the-money, out-of-the-money or in-the-money, if it generates a zero, negative, positive payoff respectively, if exercised immediately.

A crucial input of option implied trees is the estimation of the smile (implied volatility as a function of the strike price), which boils down to fitting a function to a limited number of existing knots (pairs of strike price and implied volatility). The main issue with the use of option prices is the generation of the missing prices for strikes that are not quoted, but are necessary in order to derive option implied trees or volatility forecasts. The way in which implied volatility varies with strike price is referred to as the "smile" (or smirk) effect, since depending on the market under scrutiny, it can be depicted with a smile (if implied volatility is higher at the edges of the strike price interval than it is in the middle) or a smirk (if implied volatility is higher for low strike prices than it is for high strike prices).

The no arbitrage argument would imply that it is indifferent to obtain an implied volatility from a call or a put with the same strike price. Empirically, due to market frictions and the impossibility to perfectly replicate every claim, the two implied volatilities are different. Therefore, it is market practice to keep the implied volatility of put options for strikes below the current value of the underlying asset and the one of call options for strikes above (those options are called out-of-the-money, since if exercised they would deliver no positive payoff). The latter market practice is based on the observation that the options retained are the most exchanged and thus the most informative. For the at-the-money strikes (the one right below and right above the underlying asset value), an average of call and put implied volatilities is used.

The aim of this paper is to investigate the potential of fuzzy regression for the estimation of the smile. With fuzzy regression we should be able to use all the information of both call and put options, without having to make an a priori choice. Given that the majority of the papers in the literature concentrates on fuzzy linear regression and there is no ready-to-use model which can be

adapted to our application, as a first step, we extend three of the most used linear regression methods to a polynomial form. The methods range from the possibilistic regression method of Tanaka, Uejima and Asai [14], the least squares fuzzy regression method of Savic and Pedrycz [13] and the hybrid method of Ishibuchi and Nii [4]. Second, the usefulness of fuzzy regression is evaluated by constructing an option implied tree with the estimated smile function and assessing the accuracy of the tree in pricing the same options used for its construction. Third, to leave nothing in doubt, we also assess the usefulness of the fuzzy regression methods in forecasting the real moments of the distribution (variance, skewness and kurtosis).

The results are particularly useful in at least two aspects. First, the extension to a polynomial form of the fuzzy regression methods may challenge the application to other interesting problems in finance and other disciplines, characterized by a non-linear relationship between dependent and independent variable. Second, at the practitioner level, we show that indeed, fuzzy regression could provide important improvements over standard techniques.

The paper proceeds as follows. In Section 2 we provide a brief introduction to fuzzy regression methods. In Section 3 we recall the linear fuzzy regression methods of Tanaka, Uejima and Asai [14], Savic and Pedrycz [13] and Ishibuchi and Nii [4] and in Section 4 we extend them to the polynomial case. In Section 5 we present a simple example of estimation of the smile function with the different methods. Section 6 presents the data set, the methodology and the results. The last section concludes and provides some hints for future research.

2. Fuzzy regression

Fuzzy regression methods can be used to fit both crisp and fuzzy data, handling both imprecision of measurements and fuzziness of the relationship among variables. For crisp data, they are particularly useful when ordinary regression is not appropriate because of the impossibility of verifying distributional assumptions or deriving a valid statistical relationship.

The aim of fuzzy regression is to incorporate all the vagueness embedded in the data, without losing the information which is inevitably overridden when the original data is arbitrarily modified or the imprecision cancelled. Therefore, it could be particularly useful in this financial application, characterized by imprecise data (two implied volatilities are associated to the same strike price), which is normally made crisp, by either averaging the two implied volatilities in a single estimate (for at-the-money strikes) or retaining only one of the two and discarding the other (for out-of-the-money options).

Among fuzzy regression models we distinguish between models where the relationship between the variables is fuzzy and models where the variables themselves are fuzzy. In the first case, we have crisp inputs, crisp outputs (CICO) and a fuzzy system structure, while in the second case, the system structure is fuzzy, the output is fuzzy and the input can be fuzzy or crisp (crisp input and fuzzy output (CIFO) or fuzzy input and fuzzy output (FIFO)). A second classification employs the two basic approaches used in fuzzy regression: the so-called possibilistic regression which aims at minimizing the fuzziness in the model (Tanaka, Uejima and Asai [14], a linear programming approach), and fuzzy least squares regression, which uses least squares of errors as a fitting criterion (Diamond [1]). Also hybrid methods have been proposed, which use both the possibilistic and the least squares approach (Ishibuchi and Nii [4]). For a comprehensive literature review of fuzzy regression methods and their applications see Kahraman, Beşkese and Bozbura [6] and Muzzioli and De Baets [11].

To the best of our knowledge, the majority of the papers in the literature concentrates on fuzzy linear regression. However, if the relationship among variables is not linear, it could be useful to extend the fuzzy linear regression model to the non-linear case, permitting a more accurate fit of the data. Only a few papers (see e.g. Hong and Do [3], D'Urso and Gastaldi [2], Mosleh et al. [10]) address the non-linear case, mainly with the use of neural networks, and only for the CIFO and FIFO case.

Given that the financial application addressed is characterized by crisp input and outputs (CICO) and given that the shape of the smile function is usually modelled with a second degree polynomial, neither one of the existing models can be readily applied. Therefore an extension of the fuzzy linear regression models of Tanaka, Uejima and Asai [14], Savic and Pedricz [13] and Ishibuchi and Nii [4] to a fuzzy polynomial regression model of order two is needed. In the following, we briefly review in Section 3 the linear case, and we introduce, in Section 4 the quadratic case.

3. The linear case

Let us start with the fuzzy linear regression model, and let us focus on the CICO case. The goal of fuzzy regression is to determine a linear fuzzy model which includes all the given (x_p, y_p) pairs, where $x_p = (x_{p1}, x_{p2}, ..., x_{pn})$, $p=1,...,m$ (see e.g. Tanaka, Uejima and Asai [14]) at a given confidence level h, where $Y_h = [F(x)]_h$ is the alpha-cut of the fuzzy output $F(x)$, i.e. $y_p \in [F(x_p)]_h$. The deviations between observed and estimated points are viewed as the fuzziness

of the model structure, therefore, no assumptions concerning the errors are made since they are not part of the fuzzy regression model:

$$Y = A_0 + A_1 x_1 + \ldots + A_n x_n \tag{1}$$

where $Y=F(x)$ is a fuzzy output, $x = (x_1, x_2, \ldots, x_n)$ is a non-fuzzy input vector and A_i, $i=0,\ldots, n$, are the fuzzy coefficients.

The fuzzy coefficients are determined in such a way that the estimated $F(x)$ has minimum fuzzy width at a target degree of belief h, i.e. the membership degree of each observation should be greater than the threshold value h. The parameter h can be chosen by the decision maker (see also Moskowitz and Kim [9] for the assessment of h) and represents the degree of belief desired: if the degree of belief is set to zero, the fuzzy output will exactly embed all the observations at the 0-level set; if a higher degree of belief ($h>0$) is set, upper and lower fuzzy bands are widened in order to embed all the observations at the h-level set. A level $h>0$ is usually set when we are unsure if additional information could lie outside the existing input points. For this financial application a value $h=0$ is used since we believe that the input data are sufficient to describe the fuzzy regression model.

The Tanaka, Uejima and Asai method for equation (1) assumes the fuzzy coefficients to have a symmetric triangular membership function: $A_i = (a_i^C, a_i^w)$, where a_i^C and a_i^w are the center and the spread respectively of the symmetric triangular fuzzy number A_i. In order to determine the coefficients A_i, the following linear programming problem has to be solved. Minimize the total spread of the fuzzy output:

$$\min z = \sum_{p=1}^{m} \left\{ a_0^w + a_1^w \left| x_{p1} \right| + \ldots + a_n^w \left| x_{pn} \right| \right\} \tag{2}$$

subject to:

$$a_0^C + \sum_{i=1}^{n} a_i^C x_{pi} - (1-h) \left[a_0^w + \sum_{i=1}^{n} a_i^w \left| x_{pi} \right| \right] \leq y_p, \quad p=1,\ldots, m$$

$$a_0^C + \sum_{i=1}^{n} a_i^C x_{pi} + (1-h) \left[a_0^w + \sum_{i=1}^{n} a_i^w \left| x_{pi} \right| \right] \geq y_p, \quad p=1,\ldots, m$$

$$a_i^w \geq 0$$

Note that in problem (2) the function to be minimized is the total spread of the fuzzy output, as proposed by Tanaka [15], instead of the total spread of the fuzzy coefficients as in the original Tanaka, Uejima and Asai method.

The method of Savic and Pedrycz combines ordinary least squares regression and the minimum fuzziness principle, by pursuing a two-stage methodology. In the first stage only the center of the fuzzy model is fixed by using ordinary least squares regression. In the second stage the minimum fuzziness criterion is used in order to find the spread of the fuzzy regression coefficients, by solving problem (2), where the center of each fuzzy coefficient is imposed to be equal to the ordinary least squares coefficient computed in the first stage. In particular, for the first stage, the least squares equations for the general linear regression model are used:

$$(x'x)\boldsymbol{a}^C = x'y$$

and the least square estimator for the central values is derived as follows:

$$\boldsymbol{a}^C = (x'x)^{-1}x'y$$

In the second stage, the same problem in equation (2) is solved with the difference that the vector $\boldsymbol{a}^C$ is not unknown, rather it is pre-determined in the first stage. Therefore we solve:

$$\min z = \sum_{p=1}^{m} \left\{ a_0^w + a_1^w |x_{p1}| + \ldots + a_n^w |x_{pn}| \right\} \tag{3}$$

subject to:

$$\boldsymbol{a_0^C} + \sum_{i=1}^{n} \boldsymbol{a_i^C} x_{pi} - (1-h)\left[a_0^w + \sum_{i=1}^{n} a_i^w |x_{pi}| \right] \le y_p , \quad p = 1, \ldots, m$$

$$\boldsymbol{a_0^C} + \sum_{i=1}^{n} \boldsymbol{a_i^C} x_{pi} + (1-h)\left[a_0^w + \sum_{i=1}^{n} a_i^w |x_{pi}| \right] \ge y_p , \quad p = 1, \ldots, m$$

$$a_i^w \ge 0$$

where $\boldsymbol{a}^C$ is predetermined in phase 1.

Ishibuchi and Nii [4] proposed a hybrid method which computes the central values of the linear fuzzy model by means of ordinary least squares regression and the upper and lower bounds of the fuzzy model by minimizing the total spread of the fuzzy output. They extend the Tanaka, Uejima and Asai [14] method to non-symmetric triangular fuzzy coefficients. In order to introduce the Ishibuchi and Nii [4] method, we cannot rely anymore on center and spread, therefore we denote each asymmetric triangular fuzzy number A_i as a triplet $A_i = (a_i^L, a_i^C, a_i^U)$ $i=0,\ldots, n$, where a_i^L is the lower bound, a_i^C is the central value and a_i^U is the upper bound. Let us write the fuzzy regression model as a triplet as follows:

$$Y = F(x) = \left(f^L(x), f^C(x), f^U(x) \right)$$

6

where $f^L(x)$ is the lower bound, $f^U(x)$ the upper bound, and $f^C(x)$ the central value. From fuzzy arithmetic it follows that:

$$f^L(x) = a_0^L + \sum_{\substack{i=1 \\ x_i \geq 0}}^{n} a_i^L x_i + \sum_{\substack{i=1 \\ x_i < 0}}^{n} a_i^U x_i$$

$$f^C(x) = a_0^C + \sum_{i=1}^{n} a_i^C x_i$$

$$f^U(x) = a_0^U + \sum_{\substack{i=1 \\ x_i \geq 0}}^{n} a_i^U x_i + \sum_{\substack{i=1 \\ x_i < 0}}^{n} a_i^L x_i$$

For a given confidence level h, the fuzzy linear regression model $F(x)$ can be expressed in terms of h-cuts as follows:

$$[F(x)]_h = \left[hf^C(x) + (1-h)f^L(x), \ hf^C(x) + (1-h)f^U(x) \right]$$

In the first stage, in order to determine the central value $f^C(x)$, the least squares equations for the general linear regression model are used:

$$(x'x)a^C = x'y$$

and the least square estimator for the central values is obtained as follows:

$$a^C = (x'x)^{-1}x'y$$

In the second stage, $f^L(x)$ and $f^U(x)$ are determined by solving the following linear programming problem:

$$\min z = \sum_{p=1}^{m} f^U(x_p) - f^L(x_p) \tag{4}$$

subject to:

$$h f^C(x_p) + (1-h)f^L(x_p) \leq y_p, \ p = 1, \ldots, m$$

$$h f^C(x_p) + (1-h)f^U(x_p) \geq y_p, \ p = 1, \ldots, m$$

$$a_i^L \leq a_i^C \leq a_i^U, \quad i = 0, 1, \ldots, n$$

where a_i^C is predetermined in phase 1.

4. The quadratic case

Let us now introduce the quadratic case in the three models. For simplicity, let us focus on a fuzzy regression model with only one explanatory variable and, given that our input data is strictly positive (both strike prices and implied volatilities), we can simplify formulas by overriding the case of negative input and\or output. The polynomial regression model takes the following form:

$$Y = A_0 + A_1 x + A_2 x^2 \tag{5}$$

where Y is the fuzzy output, x is a non-fuzzy input vector and A_i, $i=0,...,2,$ are the fuzzy coefficients of the second order polynomial. Given that x is crisp, the triangular form is preserved in the right hand side.

Let us write the fuzzy regression model as a triplet as follows:

$$Y = F(x) = \left(f^L(x), f^C(x), f^U(x) \right) \tag{6}$$

where $f^L(x)$ is the lower bound, $f^U(x)$ the upper bound, and $f^C(x)$ the central value. From fuzzy arithmetic (recall that x_p are strictly positive) it follows that:

$$f^L(x) = a_0^L + a_1^L x + a_2^L x^2$$
$$f^C(x) = a_0^C + a_1^C x + a_2^C x^2$$
$$f^U(x) = a_0^U + a_1^U x + a_2^U x^2$$

For a given confidence level h, the fuzzy linear regression model $F(x)$ can be expressed in terms of h-cuts as follows:

$$[F(x)]_h = \left[hf^C(x) + (1-h)f^L(x), \ hf^C(x) + (1-h)f^U(x) \right] \tag{7}$$

For fuzzy regression methods that use symmetric fuzzy numbers, it is useful to think about the fuzzy regression model in terms of central value and spread, as follows:

$$Y = F(x) = \left(f^C(x), f^w(x) \right) \tag{8}$$

where $f^C(x)$ is the central value and $f^w(x)$ is the spread.

It follows that:

$$f^U(x) = f^C(x) + f^w(x) \tag{9}$$

$$f^D(x) = f^C(x) - f^w(x) \tag{10}$$

$$[F(x)]_h = \left[f^C(x) + (1-h)f^w(x), \ f^C(x) - (1-h)f^w(x) \right] \tag{11}$$

The Tanaka, Uejima and Asai method for equation (5) assumes the fuzzy coefficients to have a symmetric triangular membership function: $A_i = (a_i^C, a_i^w)$, where a_i^C and a_i^w are the center and the spread respectively of the symmetric triangular fuzzy number A_i.

Given that some of the volatility observations (y_p) share the same strike price (x_p) (call and put implied volatilities), we have two different y_p associated to the same strike price x_p. Therefore, in order to include all the observations in the fuzzy model, we first compute the minimum and the maximum volatilities for each strike price x_p, $p=1,...,n$ ($\sigma_{min}(x_p) = min(\sigma_C(x_p), \sigma_P(x_p))$; $\sigma_{max}(x_p) = max(\sigma_C(x_p), \sigma_P(x_p))$) where $\sigma_C(x_p)$ is the call implied volatility and $\sigma_P(x_p)$ is the put implied volatility.

In order to determine the coefficients A_i, the following non-linear programming problem has to be solved. Minimize the total spread of the fuzzy output:

$$\min z = \sum_{p=1}^{m} a_0^w + a_1^w x_p + a_2^w x_p^2 \tag{12}$$

subject to:

$$a_0^C + a_1^C x_p + a_2^C x_p^2 - (1-h)\left[a_0^w + a_1^w x_p + a_2^w x_p^2\right] \le y_p = \sigma_{min}(x_p),$$
$$p = 1, ..., m$$
$$a_0^C + a_1^C x_p + a_2^C x_p^2 + (1-h)\left[a_0^w + a_1^w x_p + a_2^w x_p^2\right] \ge y_p = \sigma_{max}(x_p),$$
$$p = 1, ..., m$$
$$a_i^w \ge 0$$

The Savic and Pedrycz method for Eq. (5) is divided into two steps. In the first step, we determine the coefficients a_0^C, a_1^C, a_2^C of the central regression $f^C(x) = a_0^C + a_1^C x + a_2^C x^2$, by using the least squares estimation method:

$$\min z = \sum_{p=1}^{m} \left[y_p - (a_0^C + a_1^C x_p + a_2^C x_p^2) \right]^2 \tag{13}$$

where: $y_p = (\sigma_C(x_p) + \sigma_P(x_p))/2$.

Note that, in order to have a one-to-one mapping between strikes and implied volatilities and to compute the least squares central equation, for each strike price x_p, an average of the two implied volatilities has been used.

In the second step, the same problem in Eq. (12) is solved with the difference that the vector $\boldsymbol{a}^C$ is not an unknown, rather it is pre-determined in the first stage. Therefore we solve:

$$\min z = \sum_{p=1}^{m} a_0^w + a_1^w x_p + a_2^w x_p^2 \tag{14}$$

subject to:

9

$$a_0^C + a_1^C x_p + a_2^C x_p^2 - (1-h)[a_0^w + a_1^w x_p + a_2^w x_p^2] \leq y_p = \sigma_{\min}(x_p),$$

$$p = 1, \dots, m$$

$$a_0^C + a_1^C x_p + a_2^C x_p^2 + (1-h)[a_0^w + a_1^w x_p + a_2^w x_p^2] \geq y_p = \sigma_{\max}(x_p),$$

$$p = 1, \dots, m$$

$$a_i^w \geq 0$$

where a^C is predetermined in phase 1.

For the Ishibuchi and Nii method, in the first stage, we determine the coefficients a_0^C, a_1^C, a_2^C of the central regression $f^C(x) = a_0^C + a_1^C x + a_2^C x^2$, by using least squares estimation, Eq. (13). In the second stage, we determine $f^L(x)$ and $f^U(x)$, by solving the following problem:

$$\min z = \sum_{p=1}^{m} f^U(x_p) - f^L(x_p) \qquad (15)$$

where $f^U(x) = a_0^U + a_1^U x + a_2^U x^2$ and $f^L(x) = a_0^L + a_1^L x + a_2^L x^2$, subject to:

$$h f^C(x_p) + (1-h) f^L(x_p) \leq y_p = \sigma_{\min}(x_p), \quad p = 1, \dots, m$$

$$h f^C(x_p) + (1-h) f^U(x_p) \geq y_p = \sigma_{\max}(x_p), \quad p = 1, \dots, m$$

$$a_i^L \leq a_i^C \leq a_i^U, \quad i = 0,1,2,$$

where a_i^C is predetermined in phase 1.

Note that the function to be minimized in the three models is the same, since $f^U(x_p) - f^L(x_p) = 2f^w(x_p)$ when the model is symmetrical.

5. A simple example

In this section we provide a simple example of the smile estimation on one single date, January, 2, 2008, by using the three different fuzzy regression approaches, and the standard classical approach (see e.g. Jiang and Tian [5]) of interpolating the knots by cubic splines. The data is shown in Table 1. The FTSE MIB index is worth 38035 and the interval of traded strike prices ranges from

33500 to 42000. We note that for five strike prices we have both call (σ_C) and put (σ_P) implied volatilities.

Standard market practice averages the call and the put implied volatilities with strike closest to the underlying index value in a single estimate (as the underlying index is 38035, the 38000 and the 38500 strikes) and retains only out-of-the-money options (put options with strike X<38000, the strike immediately lower than the underlying index value and call options with strike X>38500, the strike immediately above the underlying index value). The resulting knots are then interpolated by means of cubic splines, as shown in Figure 1.

By using fuzzy regression we take into account all the information provided by option prices. The three fuzzy regression methods are depicted in Figure 1. The red dotted lines represent the lower and the upper bounds of the three fuzzy regression models, which have been derived by using h=0, since in our opinion the existing knots are sufficient to describe the regression model. The purple line represents the central value of the three fuzzy regression models. The central line is the same for the Savic and Pedricz and the Ishibuchi and Nii models and it has been computed by averaging in a single estimate the implied volatilities of call and put from strike 37000 to strike 39000 in order to have a one to one mapping between strikes and implied volatilities and compute the least squares estimator. The green and the blue lines represent the upper and the lower bounds, respectively, of the *h*-cut (*h=0.7*) of the fuzzy regression model. By varying *h* between zero and one, the *h*-cuts span from the upper and lower bounds of the model (*h=0*), to the central line (*h=1*).

We can note that the Tanaka, Uejima and Asai (from now on, TU&A) method puts the central line exactly in the middle of the lower and upper bound, while the Savic and Pedricz (from now on, S&P) method starts from the central regression and finds symmetric upper and lower bounds which embed all observations and are therefore wider than the ones in the Tanaka, Uejima and Asai method. Ishibuchi and Nii (from now on, I&N) allow for asymmetric spreads, therefore it is able to shrink towards the existing knots at the edges of the strike price interval.

6. The data set, the methodology and the results

The data set consists of closing prices on FTSE MIB-index options (MIBO), recorded from 1 January 2008 to 31 December 2008, and is available upon request. MIBO are European options on the FTSE MIB index, which is a capital weighted index composed of 40 major stocks quoted on the Italian market. The data set for the FTSE MIB index and the MIBO is kindly provided by Borsa Italiana S.p.A.

In order to evaluate the accuracy of the different methodologies (standard market practice and the three different fuzzy regression methods) used to estimate the smile function, we construct an option implied tree (the Enhanced Derman and Kani's (EDK) implied tree, we refer the interested reader to Moriggia et al. [8]) and on the latter, each day, as an in-sample test, we price options traded on that day and as an out-of-sample test, we compute the implied moments of the distribution and compare the latter to subsequently realised physical moments in order to assess the forecasting power of the former.

In order to assess the accuracy of the different methods of estimation of the smile function, we resort to the following metrics widely used in the literature (see e.g. Lim and Zhi [7]). In particular, we use the Mean Absolute Percentage Error (MAPE) and the Mispricing Index (MISP) defined as follows:

$$\text{MAPE} = \frac{1}{m} \sum_{k=1}^{m} \frac{|P_k^T - P_k^M|}{P_k^M} \quad , \tag{16}$$

$$\text{MISP} = \frac{\sum_{k=1}^{m} \left(\frac{P_k^T - P_k^M}{P_k^M} \right)}{\sum_{k=1}^{m} \left| \frac{P_k^T - P_k^M}{P_k^M} \right|} , \tag{17}$$

where P^T and P^M indicate respectively the theoretical (computed with the estimated smile) and market price of the options and m is the number of options in the class. MAPE measures the accuracy of the model by means of absolute percentage errors, whereas MISP indicates the average underpricing or overpricing of the model. All the indexes have been computed both for the entire sample and for each option class: call versus put.

In order to obtain implied moments we use the risk-neutral densities estimated with the implied tree (see e.g. Tian [16]). Implied moments are computed as integrals of the risk-neutral density as follows:

$$m_\alpha = \int_{-\infty}^{\infty} x^\alpha f(x)\mathrm{d}x \tag{18}$$

with $\alpha = 1,2,3,4$, $x = \ln\frac{S_n}{S_t}$ and $f(x)$ is the risk-neutral density. As the implied tree yields a discrete cumulative distribution, a discrete summation over all nodes approximates the continuous integral in Eq. (18).

With these moments, variance, skewness and kurtosis are easily obtained as follows:

$$\text{VAR}(t,n) = m_2 - m_1^2 \tag{19}$$

$$\text{SKEW}(t,n) = \frac{m_3 - 3m_1 m_2 + 2m_1^3}{\left(m_2 - m_1^2\right)^{3/2}} \tag{20}$$

$$\text{KURT}(t,n) = \frac{m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4}{\left(m_2 - m_1^2\right)^2} \tag{21}$$

As the EDK implied tree requires a single smile function for input, we compute the implied tree by using three different smile functions: the upper and lower bounds and the central line of the fuzzy regression methods.

The results of the pricing accuracy of each methodology used to estimate the smile are reported in Table 2. The cubic spline methodology obtains a better performance than using the extremes (the upper and lower bounds $f^U(0)$ and $f^D(0)$, respectively) of each fuzzy regression model; however, it obtains a worse performance than using the corresponding central values (f^C). If we look at different options' classes, for call options the cubic splines methodology is still better than any other model, except the TU&A.

As confirmed by the results, we expect the lower (upper) bound of each fuzzy regression method to underprice (overprice) on average each option class, since options are increasing functions in volatility. If we look to the mispricing index, the cubic splines methodology underprices severely each option class: the aggregate MISP is better only than the lower bounds ($f^D(0)$) of each fuzzy regression method, and the same result holds for both call and put options.

Among the fuzzy regression models, Ishibuchi and Nii share with Savic and Pedricz the same central regression, which yields a far better result than the one of Tanaka, Uejima and Asai. The latter is, however, the best in terms of mispricing index (which is close to zero) for put options. Moreover, if we look at the MAPE, the Tanaka, Uejima and Asai model has also the lowest variability (it ranges from 0.189 to 0.229) between the upper and lower regression bounds if compared with the other models (I&N: 0.198-0.257, S&P: 0.112-0.253). The same pattern holds for MISP. Therefore we could say that the Tanaka, Uejima and Asai model is the one which yields the narrowest interval of errors. Last, the MISP index show that while the central line of the Ishibuchi and Nii and Savic and Pedricz models underprice on average both call and put options, the central line of the Tanaka, Uejima and Asai model underprices more calls, but correctly prices puts. In terms of MAPE, the upper bound of the I&N model performs better than S&P, and better than TU&A only for puts. The lower bound of the I&N performs better than both lower bounds of S&P and TU&A. The analysis of the MISP confirms the pattern. Based on these results, the Ishibuchi and Nii model is clearly superior to the Savic and Pedricz one (same central line, but lower variability between upper and lower bound), both for calls and puts. It is also superior to the Tanaka, Uejima and Asai method, in the central line and with respect the boundaries, only for puts.

As Ishibuchi and Nii is the preferred methodology, to leave nothing in doubt we pursue in Table 3 a sensitivity analysis of the choice of *h*, by using *h=0.7, 0.8, 0.9*, which shows that *0.8* is the preferred cut.

Let us turn to the moments' estimation: the results in Table 4 highlight that the cubic spline methodology yields quite lower values for variance than the fuzzy regressions do. We can say that the fuzzy regression methods overestimate the subsequently realized variance. As in the pricing performance exercise, the narrowest band for variance estimation is provided by the Tanaka, Uejima and Asai method. For all the fuzzy regression methods, the lower bound yields the better estimation of variance. While physical skewness is slightly positive, all the estimation methods (including cubic splines) yield a negative skewness, which is the highest in absolute terms for the Tanaka, Uejima and Asai method and the lowest in absolute terms for the cubic splines methodology. All the methods overestimate physical kurtosis, the best estimation technique being the cubic splines method.

7. Conclusions

In this paper we have extended to a quadratic case (with positive input variables) the linear fuzzy regression methods of Tanaka, Uejima and Asai, Savic and Pedricz and Ishibuchi and Nii and applied them to an important problem in finance: the estimation of the smile function.

We have evaluated both the in-sample pricing performance of the different estimation methods and the out-of-sample forecasting performance of the moments, by using as a benchmark the standard cubic spline interpolation approach. The results highlight that, in sample, by using fuzzy regression, we can have a far better estimation than the classical approach based on cubic splines. In particular, the best estimation method is the Ishibuchi and Nii one with the preferred h-cut at $h=0.8$. However, out-of-sample, the cubic splines methodology is superior in the forecasting of the subsequent realised moments.

The results of the paper are very important both for spreading the use of fuzzy regression to other problems that can be characterized by positive variables and a quadratic fitting function, and for showing the utility of fuzzy regression with respect to standard practice in the estimation of the implied volatility smile function.

Future research is needed in two directions. First, it is necessary to investigate methods to optimally choose the crisp smile function (such as an optimal defuzzification method) among all the possible ones. Second, we could investigate how to transfer the ambiguity in the smile function into a fuzzy implied tree (following the preliminary approach in Muzzioli and Torricelli [12]), and get a weighted interval of prices for the options, to be defuzzified ex-post.

References

[1] Diamond P. (1988). Fuzzy least squares. *Information Sciences*, 46, 141-157.

[2] D'Urso P., Gastaldi T. (2000). An orderwise polynomial regression procedure for fuzzy data. *Fuzzy Sets and Systems*, 130 (1) 1-19.

[3] Hong D. H., Do H. Y. (2001). Fuzzy polynomial regression analysis using shape preserving operation. *Korean Journal of Computational and Applied Mathematics*, 8 (3), 645-656.

[4] Ishibuchi H., Nii M. (2001). Fuzzy regression using asymmetric fuzzy coefficients and fuzzified neural networks. *Fuzzy Sets and Systems,* 119, 273-290.

[5] Jiang G.J., Tian Y.S. (2007). Extracting model-free volatility from option prices: an examination of the VIX index. *Journal of Derivatives*, 14 (3), 35-60.

[6] Kahraman C., Beşkese A., Bozbura F.T. (2006). Fuzzy Regression Approaches and Applications. *Studies in Fuzziness and Soft Computing*, 201, 589-615.

[7] Lim K., Zhi D. (2002). Pricing options using implied trees: Evidence from FTSE-100 options. *Journal of Future Markets*, 22, 601-626.

[8] Moriggia V., Muzzioli S., Torricelli C. (2009). On the no-arbitrage condition in option implied trees. *European Journal of Operational Research*, 193(1), 212-221.

[9] Moskowitz H., K.J. Kim (1993). On assessing the H value in fuzzy linear regression. *Fuzzy Sets and Systems,* 58, 303-327.

[10] Mosleh M., Otadi M., Abbasbandy S. (2011). Fuzzy polynomial regression with fuzzy neural networks. *Applied Mathematical Modelling*, 35, 5400-5412.

[11] Muzzioli S., De Baets B. (2013). A comparative assessment of different fuzzy regression methods for volatility forecasting. *Fuzzy Optimization and Decision Making*, 12 (4), 433-450.

[12] Muzzioli S., Torricelli C. (2005). The pricing of options on an interval binomial tree: an application to the DAX-index option market. *European Journal of Operational Research*, 163, 192-200.

[13] Savic D.A., Pedrycz W. (1991). Evaluation of fuzzy linear regression models. *Fuzzy Sets and Systems,* 39, 51-63.

[14] Tanaka H., Uejima S., Asai K. (1982). Linear regression analysis with fuzzy model. *IEEE Transactions on Systems, Man and Cybernetics,* 12, 903-907.

[15] Tanaka H. (1987). Fuzzy data analysis by possibilistic linear models. *Fuzzy Sets and Systems,* 24, 363-375.

[16] Tian Y. S. (2012). Implied Binomial Trees with Cubic Spline Smoothing. *Available at SSRN: http://ssrn.com/abstract=2004605 or http://dx.doi.org/10.2139/ssrn.2004605.*
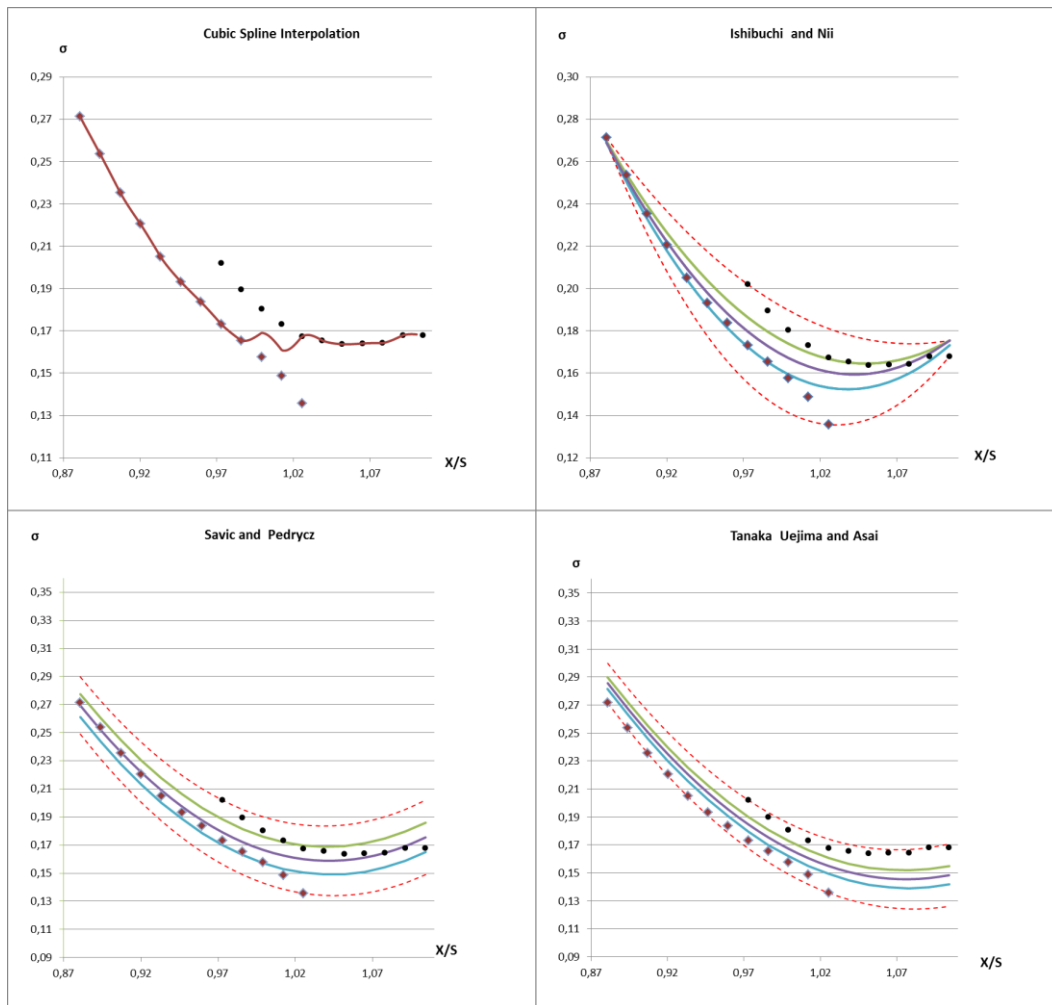
Figure 1. Cubic spline interpolation and the three fuzzy regression methods on January 02, 2008.

This figure shows the cubic spline interpolation, the hybrid model of Ishibuchi and Nii, the least squares fuzzy regression model of Savic and Pedrycz and the possibilistic fuzzy regression model of Tanaka, Uejima and Asai. For the three fuzzy regression methods, each figure reports the central value of the fuzzy output (in blue) and the upper and lower bounds of the fuzzy output for two different degrees of belief: $h=0$ and $h=0.7$. The crisp data pairs (x,y) are for call options (purple squares) and put options (black dots).

X	X/S	σ_C	σ_P
33500	0.88		0.2715
34000	0.89		0.2538
34500	0.91		0.2354
35000	0.92		0.2206
35500	0.93		0.205
36000	0.95		0.1933
36500	0.96		0.1837
37000	0.97	0.202	0.1732
37500	0.99	0.1897	0.1654
38000	1.00	0.1805	0.1577
38500	1.01	0.1731	0.1487
39000	1.03	0.1674	0.1357
39500	1.04	0.1655	
40000	1.05	0.1637	
40500	1.06	0.1641	
41000	1.08	0.1643	
41500	1.09	0.168	
42000	1.10	0.168	

Table 1. Option data on January 02, 2008.

	MAPE	MAPE Call	MAPE Put	MISP	MISP Call	MISP Put
CSPLINE	0.180	0.158	0.201	-0.930	-0.883	-0.975
I&N f^C	0.107	0.085	0.130	-0.368	-0.337	-0.3998
I&N $f^U(0)$	0.257	0.328	0.185	0.591	0.737	0.445
I&N $f^L(0)$	0.198	0.221	0.175	-0.991	-0.984	-0.998
S&P f^C	0.107	0.085	0.130	-0.368	-0.337	-0.399
S&P $f^U(0)$	0.385	0.578	0.192	0.819	0.947	0.690
S&P $f^L(0)$	0.253	0.270	0.236	-0.999	-0.999	-0.999
TU&A f^C	0.164	0.180	0.149	-0.394	-0.815	0.026
TU&A $f^U(0)$	0.189	0.168	0.210	0.600	0.629	0.572
TU&A $f^L(0)$	0.230	0.281	0.179	-0.995	-0.999	-0.990

Table 2. The pricing errors of the different smile estimation methodologies.

This table reports the pricing errors for the cubic spline (CSPLINE), Ishibuchi and Nii (I&N), Savic and Pedricz (S&P) and Tanaka, Uejima and Asai (TU&A) methods. Pricing errors are expressed by $MAPE = \frac{1}{m}\sum_{k=1}^{m}\frac{|P_k^T-P_k^M|}{P_k^M}$ and $MISP = \frac{\sum_{k=1}^{m}\left(\frac{P_k^T-P_k^M}{P_k^M}\right)}{\sum_{k=1}^{m}\left|\frac{P_k^T-P_k^M}{P_k^M}\right|}$, where P^T and P^M indicate respectively the theoretical (computed with the estimated smile) and market prices of the options and m is the number of options in the class. For each fuzzy regression model, $f^C(0)$ denotes the central line, $f^U(0)$ and $f^L(0)$ denote the upper and lower bounds of the *0*-cut, respectively.

	MAPE	MAPE Call	MAPE Put	MISP	MISP Call	MISP Put
I&N f^U (0.5)	0.157	0.168	0.145	0.304	0.407	0.201
I&N f^L (0.5)	0.151	0.155	0.149	-0.894	-0.875	-0.914
I&N f^U (0.7)	0.125	0.120	0.131	0.1017	0.147	0.057
I&N f^L (0.7)	0.127	0.123	0.132	-0.779	-0.762	-0.796
I&N f^U (0.8)	0.107	0.086	0.127	-0.031	-0.022	-0.040
I&N f^L (0.8)	0.125	0.118	0.131	-0.669	-0.648	-0.690
I&N f^U (0.9)	0.112	0.104	0.121	-0.163	-0.136	-0.190
I&N f^L (0.9)	0.115	0.106	0.125	-0.543	-0.528	-0.557

Table 3. A sensitivity analysis of the accuracy of the Ishibuchi and Nii method with respect to h.

This table reports the pricing errors of the Ishibuchi and Nii (I&N) method expressed by $\text{MAPE} = \frac{1}{m}\sum_{k=1}^{m}\frac{|P_k^T-P_k^M|}{P_k^M}$ and

$\text{MISP} = \frac{\sum_{k=1}^{m}\left(\frac{P_k^T-P_k^M}{P_k^M}\right)}{\sum_{k=1}^{m}\left|\frac{P_k^T-P_k^M}{P_k^M}\right|}$, where P^T and P^M indicate respectively the theoretical (computed with the estimated smile) and

market prices of the options and m is the number of options in the class. $f^U(h)$ and $f^L(h)$ denote the upper and lower bounds of the h-cut, h=0.5, 0.7, 0.8, 0.9.

		f^C	$f^U(0)$	$f^L(0)$	
I&N	**VAR**	0.149	0.210	0.132	
	SKEW	-0.303	-0.345	-0.386	
	KURT	3.614	3.693	3.990	
S&P	**VAR**	0.148	0.182	0.124	
	SKEW	-0.305	-0.346	-0.272	
	KURT	3.623	3.758	3.514	
TU&A	**VAR**	0.164	0.188	0.144	
	SKEW	-0.412	-0.412	-0.378	
	KURT	3.941	3.820	3.817	

	CSPLINE	**physical**	
VAR	0.097	0.012	
SKEW	-0.131	0.110	
KURT	3.075	3.028	

Table 4. Estimation of the moments.

This table reports the estimation of the moments (VAR=variance, SKEW=skewness, KURT=kurtosis) for the cubic splines (CSPLINE), Ishibuchi and Nii (I&N), Savic and Pedricz (S&P), Tanaka, Uejima and Asai (TU&A) methods, along with subsequently realized physical moments. For each fuzzy regression model, f^C denotes the central line, $f^U(0)$ and $f^L(0)$ denote the upper and lower bounds of the 0-cut, respectively.