SOME RESULTS ON A MULTIVARIATE GENERALIZATION OF THE FUZZY LEAST SQUARE REGRESSION

Francesco Campobasso, Annarita Fanizzi and Marina Tarantini
Department of Statistical Sciences "Carlo Cecchi", University of Bari, Italy

Keywords: Fuzzy least square regression, Multivariate generalization, Total deviance, Decomposition, Goodness of fitting.

Abstract: Fuzzy regression techniques can be used to fit fuzzy data into a regression model, where the deviations between the dependent variable and the model are connected with the uncertain nature either of the variables or of their coefficients. P.M. Diamond (1988) treated the case of a simple fuzzy regression of an uncertain dependent variable on a single uncertain independent variable, introducing a metrics into the space of triangular fuzzy numbers. In this work we managed more than a single independent variable, determining the corresponding estimates and providing some theoretical results about the decomposition of the sum of squares of the dependent variable according to Diamond's metric, in order to identify its components.

1 INTRODUCTION

Modalities of quantitative variables are commonly given as exact single values, although sometimes they cannot be precise. The imprecision of measuring instruments and the continuous nature of some observations, for example, prevent researcher from obtaining the corresponding true values.

On the other hand qualitative variables are commonly expressed using common linguistic terms, which also represent verbal labels of sets with uncertain borders. This is the case of the answers provided in the customer satisfaction surveys, which are collected through ordered categories from "not at all" to "completely".

The appropriate way to manage such an uncertainty of observations is provided by fuzzy theory.

In 1988 P. M Diamond introduced a metric onto the space of triangular fuzzy numbers and derived the expression of the estimated coefficients in a simple fuzzy regression of an uncertain dependent variable on a single uncertain independent variable.

Starting from a multivariate generalization of this regression, we give important results about the decomposition of the deviance of the dependent variable according to Diamond's metric.

2 THE FUZZY LEAST SQUARE REGRESSION

A triangular fuzzy number $\tilde{X}=(x,x^L,x^R)_T$ for the variable X is characterized by a membership function $\mu_{\tilde{X}}(x_i)$ like the one represented in Fig.1.

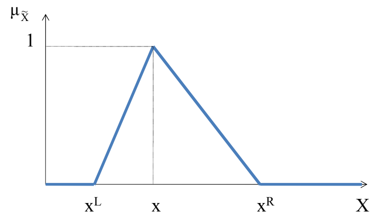

Figure 1: Representation of a triangular fuzzy number.

The accumulation value x is considered the centre of the fuzzy number, while $x-x^L$ and x^R-x are considered the left spread and the right spread respectively. Note that x belongs to $\tilde{X}$ with the highest degree, while the other values included between the extremes x^L and x^R belong to $\tilde{X}$ with a gradually lower degree.

The set of triangular fuzzy numbers is closed with respect to sum: given two triangular fuzzy numbers $\tilde{X}=(x,x^L,x^R)_T$ and $\tilde{Y}=(y,y^L,y^R)_T$, their sum is still a triangular fuzzy number $\tilde{Z}=\tilde{X}+\tilde{Y}=(x+y,x^L+y^L,x^R+y^R)_T$. Moreover the

Campobasso F., Fanizzi A. and Tarantini M. (2009).
SOME RESULTS ON A MULTIVARIATE GENERALIZATION OF THE FUZZY LEAST SQUARE REGRESSION.
In *Proceedings of the International Joint Conference on Computational Intelligence*, pages 75-78
DOI: 10.5220/0002321200750078

product of a triangular fuzzy number $\tilde{X}=(x,x^L,x^R)_T$ and a real number k depends on the sign of the latter, resulting equal to $k\tilde{X}=(kx,kx^L,kx^R)_T$ if k is positive or $k\tilde{X}=(kx,kx^R,kx^L)_T$ if k is negative.

P.M. Diamond (1988) introduced a metric onto the space of triangular fuzzy numbers; according to this metric, the distance between $\tilde{X}$ and $\tilde{Y}$ is $d(\tilde{X},\tilde{Y})^2=d((x,x^L,x^R)_T,(y,y^L,y^R)_T)^2=(x-y)^2+(x^L-y^L)^2+(x^R-y^R)^2$.

In particular Diamond analysed the regression model of a fuzzy dependent variable $\tilde{Y}$ on a single fuzzy independent variable $\tilde{X}$:

$$\tilde{Y} = a + b\,\tilde{X} + \varepsilon.$$

The expression of the corresponding parameters is derived from minimizing the sum $\sum_{i=1}^{n} d(a+b\tilde{X}_i,\tilde{Y}_i)^2$ of the squared distances between theoretical and empirical values in n observed units of the fuzzy dependent variable $\tilde{Y}$ with respect to a and b.

Such a sum takes different forms according to the signs of the coefficient b, as the product of a fuzzy number $\tilde{X}=(x,x^L,x^R)_T$ and a real number k depends on whether the latter is positive or negative. Therefore, multiplying by a negative real number, the right extreme of the fuzzy number is obtained by adding the left spread to the centre, while its left extreme is obtained by subtracting the right spread from the centre.

Diamond demonstrated that the optimization problem has a unique solution under certain conditions.

3 A MULTIVARIATE GENERALIZATION OF THE REGRESSION MODEL

Recently we generalized this estimation procedure to the case of k independent variables (k≥1). Let's assume to observe a fuzzy dependent variable $\tilde{Y}_i=(y_i,\underline{\mu}_i,\overline{\mu}_i)_T$ and two fuzzy independent variables, $\tilde{X}_i=(x_i,\underline{\xi}_i,\overline{\xi}_i)_T$ and $\tilde{Z}_i=(z_i,\underline{\delta}_i,\overline{\delta}_i)_T$, on a set of n units. The linear regression model is given by

$$\tilde{Y}_i{}^* = a + b\,\tilde{X}_i + c\,\tilde{Z}_i, \quad i=1,2,\ldots,n;\ a,b,c \in \mathrm{IR}.$$

The corresponding parameters are determined by minimizing the sum of Diamond's distances between theoretical and empirical values of the dependent variable

$$\sum_{i=1}^{n} d(a+b\tilde{X}_i+c\tilde{Z}_i,\tilde{Y}_i)^2 \qquad (1)$$

respect to a, b and c. As we stated above, such a sum assumes different expressions according to the signs of the regression coefficients b and c. This generates the following four cases

<u>*Case 1*</u>: b≥0, c≥0

$$\sum_{i=1}^{n} d(a+b\tilde{X}_i+c\tilde{Z}_i,\tilde{Y}_i)^2 =$$

$$=\sum_{i=1}^{n}\left[(y_i^L-a-bx_i^L-cz_i^L)^2+(y_i-a-bx_i-cz_i)^2+(y_i^R-a-bx_i^R-cz_i^R)^2\right]$$

where $y_i^L = y_i - \underline{\mu}_i$, $y_i^R = y_i + \overline{\mu}_i$ and x_i^L, x_i^R, z_i^L, z_i^R have similar meanings.

<u>*Case 2*</u>: b<0, c≥0

$$\sum_{i=1}^{n} d(a+b\tilde{X}_i+c\tilde{Z}_i,\tilde{Y}_i)^2 =$$

$$=\sum_{i=1}^{n}\left[(y_i^L-a-bx_i^R-cz_i^L)^2+(y_i-a-bx_i-cz_i)^2+(y_i^R-a-bx_i^L-cz_i^R)^2\right]$$

<u>*Case 3*</u>: b≥0, c<0

$$\sum_{i=1}^{n} d(a+b\tilde{X}_i+c\tilde{Z}_i,\tilde{Y}_i)^2 =$$

$$\sum_{i=1}^{n}\left[(y_i^L-a-bx_i^L-cz_i^R)^2+(y_i-a-bx_i-cz_i)^2+(y_i^R-a-bx_i^R-cz_i^L)^2\right]$$

<u>*Case 4*</u>: b<0, c<0

$$\sum_{i=1}^{n} d(a+b\tilde{X}_i+c\tilde{Z}_i,\tilde{Y}_i)^2 =$$

$$\sum_{i=1}^{n}\left[(y_i^L-a-bx_i^R-cz_i^R)^2+(y_i-a-bx_i-cz_i)^2+(y_i^R-a-bx_i^L-cz_i^L)^2\right]$$

Let's consider, as an example, case 3. The expression to be minimized is given by

$$(\mathbf{y}^L-\mathbf{X}^{LR}\boldsymbol{\beta})'(\mathbf{y}^L-\mathbf{X}^{LR}\boldsymbol{\beta})+(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+(\mathbf{y}^R-\mathbf{X}^{RL}\boldsymbol{\beta})'(\mathbf{y}^R-\mathbf{X}^{RL}\boldsymbol{\beta}) \quad (2)$$

in matricial terms, where

$\mathbf{y}^L$ e $\mathbf{y}^R$ are n-dimensional vectors, whose elements are the lower extremes $y_i^L = y_i - \underline{\mu}_i$ and the upper extremes $y_i^R = y_i + \overline{\mu}_i$ respectively;

$\mathbf{X}^{LR}$ is the n×3 matrix, formed by vectors **1**, $\mathbf{x}^L=[x_i^L = x_i - \underline{\xi}_i]$ and $\mathbf{z}^R=[z_i^R = z_i + \overline{\delta}_i]$;

$\mathbf{X}^{RL}$ is the n×3 matrix (analogous to $\mathbf{X}^{LR}$), formed by vectors **1**, $\mathbf{x}^R$, $\mathbf{z}^L$;

y is the n-dimensional vector of centres y_i;

X is the n×3 matrix formed by vectors **1**, $\mathbf{x}^C=[x_i]$, $\mathbf{z}^C=[z_i]$;

β is the vector (a, b, c) '.

Similarly to OLS estimation procedure, the optimization problem admits a single and finite solution if $[(\mathbf{X}^{LR})'\,\mathbf{X}^{LR} + (\mathbf{X})'\,\mathbf{X} + (\mathbf{X}^{RL})'\,\mathbf{X}^{RL}]$ is invertible and the hessian matrix $[2(\mathbf{X}^{LR})'\,\mathbf{X}^{LR} + 2(\mathbf{X})'\,\mathbf{X} + 2(\mathbf{X}^{LR})'\,\mathbf{X}^{RL}]$ is definite positive. The matricial expression of the fuzzy least square (FLS) estimator is given by

$$\boldsymbol{\beta}=[(\mathbf{X}^{LR})'\mathbf{X}^{LR}+\mathbf{X}'\mathbf{X}+(\mathbf{X}^{RL})'\mathbf{X}^{RL}]^{-1}[(\mathbf{X}^{LR})'\mathbf{y}^L+\mathbf{X}'\mathbf{y}+(\mathbf{X}^{RL})'\mathbf{y}^R].$$

It's worth noticing that the FLS estimator would equal the OLS one if the observed variables were crisp. The found solution $\beta^*=(a^*, b^*, c^*)$ is admissible if the signs of the regression coefficients are coherent with basic assumptions ($b^*\geq 0$ and $c^*<0$).

In the remaining three cases the expression (2) to be minimized is obtained after replacing $\mathbf{X}^{LR}$ and $\mathbf{X}^{RL}$ by $\mathbf{X}^{LL}$ and $\mathbf{X}^{RR}$ (case 1), by $\mathbf{X}^{RL}$ and $\mathbf{X}^{LR}$ (case 2), $\mathbf{X}^{RR}$ and $\mathbf{X}^{LL}$ (case 4) respectively.

The optimum solution corresponds to that (admissible) one which makes minimum (1) among all.

Note that the generalization of such a procedure to the case of several independent variables is immediate and that the number of solutions to analyse, in order to identify the optimum one, growths exponentially with the considered number of variables. For example, if the model includes k independent variables, 2^k possible cases must be taken into account, which derive from combining the signs of the regression coefficients.

4 DECOMPOSITION OF TOTAL DEVIANCE OF THE DEPENDENT VARIABLE

In this section two important theoretical results will be demonstrated: the first one regards the inequality between theoretical and empirical values of the average fuzzy dependent variable (unlike in the OLS estimation procedure for crisp variables); the second one regards the decomposition of the total deviance of the dependent variable, which involves other two additive components besides the regression and the residual deviances.

It is necessary to obtain preliminary results for this purpose. After considering, only for example, the case 3 and in particular rewriting (2) as

$[(\mathbf{X}^{LR})'\mathbf{X}^{LR}+(\mathbf{X})'\mathbf{X}+(\mathbf{X}^{RL})'\mathbf{X}^{RL}]\beta=[(\mathbf{X}^{LR})'\mathbf{y}^{L}+(\mathbf{X})'\mathbf{y}+(\mathbf{X}^{RL})'\mathbf{y}^{R}]$,

we can obtain the following system of equations:

$$\begin{cases} 3na+b\sum(x_i^L+x_i+x_i^R)+c\sum(z_i^R+z_i+z_i^L)=\sum(y_i^L+y_i+y_i^R) \\ a\sum(x_i^L+x_i+x_i^R)+b\sum((x_i^L)^2+(x_i)^2+(x_i^R)^2)+c\sum(x_i^Lz_i^R+x_iz_i+x_i^Rz_i^L)= \\ \quad =\sum(y_i^Lx_i^L+y_ix_i+y_i^Rx_i^R) \\ a\sum(z_i^R+z_i+z_i^L)+b\sum(x_i^Lz_i^R+x_iz_i+x_i^Rz_i^L)+c\sum((z_i^R)^2+(z_i)^2+(z_i^L)^2)= \\ \quad =\sum(y_i^Lz_i^R+y_iz_i+y_i^Rz_i^L) \end{cases}$$

$$\Leftrightarrow$$

$$\begin{cases} \sum(a+bx_i^L+cz_i^R)+\sum(a+bx_i+cz_i)+\sum(a+bx_i^R+cz_i^L)= \\ \quad =\sum y_i^L+\sum y_i+\sum y_i^R \\ \sum(a+bx_i^L+cz_i^R)x_i^L+\sum(a+bx_i+cz_i)x_i+\sum(a+bx_i^R+cz_i^L)x_i^R= \\ \quad =\sum y_i^Lx_i^L+\sum y_ix_i+\sum y_i^Rx_i^R \\ \sum(a+bx_i^L+cz_i^R)z_i^R+\sum(a+bx_i+cz_i)z_i+\sum(a+bx_i^R+cz_i^L)z_i^L= \\ \quad =\sum y_i^Lz_i^R+\sum y_iz_i+\sum y_i^Rz_i^L \end{cases}$$

$$\Leftrightarrow$$

$$\begin{cases} \sum y_i^{*L}+\sum y_i^*+\sum y_i^{*R}=\sum y_i^L+\sum y_i+\sum y_i^R & (3) \\ \sum y_i^{*L}x_i^L+\sum y_i^*x_i+\sum y_i^{*R}x_i^R=\sum y_i^Lx_i^L+\sum y_ix_i+\sum y_i^Rx_i^R & (4) \\ \sum y_i^{*L}z_i^R+\sum y_i^*z_i+\sum y_i^{*R}z_i^L=\sum y_i^Lz_i^R+\sum y_iz_i+\sum y_i^Rz_i^L & (5) \end{cases}$$

Equation (3) shows that the total sum of lower extremes, centres and upper extremes of the theoretical values of the dependent variable coincides with the same amount referred to the empirical values. Such an equation does not allow us to say that theoretical and empirical values of the average fuzzy dependent variable coincide.

Let's examine how the total deviance of Y can be decomposed according to Diamond's metric:

$$\mathrm{Dev(Tot)}=\sum[(y_i^L-\bar{y}^L)^2+(y_i-\bar{y})^2+(y_i^R-\bar{y}^R)^2].$$

Adding and subtracting the corresponding theoretical value within each square and developing all the squares, the total deviance can be expressed as

$$\begin{aligned} \mathrm{Dev(Tot)}=&\sum[(y_i^L-y_i^{*L})^2+(y_i^{*L}-\bar{y}^L)^2+2(y_i^L-y_i^{*L})(y_i^{*L}-\bar{y}^L)+ \\ &+(y_i-y_i^*)^2+(y_i^*-\bar{y})^2+2(y_i-y_i^*)(y_i^*-\bar{y})+(y_i^R-y_i^{*R})^2+ \\ &+(y_i^{*R}-\bar{y}^R)^2+2(y_i^R-y_i^{*R})(y_i^{*R}-\bar{y}^R)]= \end{aligned}$$

$$\begin{aligned} &=\left[\sum(y_i^L-y_i^{*L})^2+\sum(y_i-y_i^*)^2+\sum(y_i^R-y_i^{*R})^2\right]+ \\ &+\sum\left[(y_i^{*L}-\bar{y}^L)^2+(y_i^*-\bar{y})^2+(y_i^{*R}-\bar{y}^R)^2\right]+ \\ &+\sum\left[2(y_i^L-y_i^{*L})(y_i^{*L}-\bar{y}^L)+2(y_i-y_i^*)(y_i^*-\bar{y})+2(y_i^R-y_i^{*R})(y_i^{*R}-\bar{y}^R)\right] \end{aligned}$$

Adding and subtracting the theoretical average values of the lower extremes, of the centres and of the upper extremes of the dependent variable within each square and solving all the squares, the previous expression becomes

$$\begin{aligned} \mathrm{Dev(Tot)}=&\left[\sum(y_i^L-y_i^{*L})^2+\sum(y_i-y_i^*)^2+\sum(y_i^R-y_i^{*R})^2\right]+ \\ &+\sum\left[(y_i^{*L}-\bar{y}^{*L}+\bar{y}^{*L}-\bar{y}^L)^2+(y_i^*-\bar{y}^*+\bar{y}^*-\bar{y})^2+(y_i^{*R}-\bar{y}^{*R}+\bar{y}^{*R}-\bar{y}^R)^2\right]+ \\ &+\sum\left[2(y_i^L-y_i^{*L})(y_i^{*L}-\bar{y}^L)+2(y_i-y_i^*)(y_i^*-\bar{y})+2(y_i^R-y_i^{*R})(y_i^{*R}-\bar{y}^R)\right]= \\ =&\sum\left[(y_i^L-y_i^{*L})^2+(y_i-y_i^*)^2+(y_i^R-y_i^{*R})^2\right]+ \\ &+\sum\left[(y_i^{*L}-\bar{y}^{*L})^2+(y_i^*-\bar{y}^*)^2+(y_i^{*R}-\bar{y}^{*R})^2\right]+ \\ &+n[(\bar{y}^{*L}-\bar{y}^L)^2+(\bar{y}^*-\bar{y})^2+(\bar{y}^{*R}-\bar{y}^R)^2]+ \\ &+\sum\ [2(y_i^{*L}-\bar{y}^{*L})(\bar{y}^{*L}-\bar{y}^L)+2(y_i^*-\bar{y}^*)(\bar{y}^*-\bar{y})+ \\ &+2(y_i^{*R}-\bar{y}^{*R})(\bar{y}^{*R}-\bar{y}^R)+2(y_i^L-y_i^{*L})(y_i^{*L}-\bar{y}^L)+ \\ &+2(y_i-y_i^*)(y_i^*-\bar{y})+2(y_i^R-y_i^{*R})(y_i^{*R}-\bar{y}^R)] \end{aligned} \qquad (6)$$

where

$$\Sigma[(y_i-y_i^*)^2+(y_i^L-y_i^{*L})^2+(y_i^R-y_i^{*R})^2]=\mathrm{Dev(Res)}$$

represents the residual deviance,

$$\Sigma[(y_i^{*L}-\bar{y}^{*L})^2+\Sigma(y_i^*-\bar{y}^*)^2+\Sigma(y_i^{*R}-\bar{y}^{*R})^2]=\mathrm{Dev(Regr)}$$

represents the regression deviance and

$$[(\bar{y}^{*L}-\bar{y}^L)^2+(\bar{y}^*-\bar{y}\ )^2+(\bar{y}^{*R}-\bar{y}^R)^2]=d(\overline{Y},\overline{Y}^*)^2$$

represents the distance between theoretical and empirical average values of Y.

Synthetically the expression (6) can be written as:

$$\mathrm{Dev(Tot)}=\mathrm{Dev(Res)}+\mathrm{Dev(Regr)}+nd(\overline{Y},\overline{Y}^*)^2+\eta$$

where:

$$\eta=2\Sigma\ [(y_i^{*L}-\bar{y}^{*L})(\bar{y}^{*L}-\bar{y}^L)+(y_i^*-\bar{y}^*)(\bar{y}^*-\bar{y})+(y_i^{*R}-\bar{y}^{*R})(\bar{y}^{*R}-\bar{y}^R)]+$$
$$+2\Sigma[(y_i^L-y_i^{*L})(y_i^{*L}-\bar{y}^L)+(y_i-y_i^*)(y_i^*-\bar{y}\ )+(y_i^R-y_i^{*R})(y_i^{*R}-\bar{y}^R)].$$

As the sums of deviations of each component from its average equal zero, then it is

$$2\Sigma\left[(y_i^{*L}-\bar{y}^{*L})(\bar{y}^{*L}-\bar{y}^L)+(y_i^*-\bar{y}^*)(\bar{y}^*-\bar{y})+(y_i^{*R}-\bar{y}^{*R})(\bar{y}^{*R}-\bar{y}^R)\right]=$$
$$=2[(\bar{y}^{*L}-\bar{y}^L)\Sigma(y_i^{*L}-\bar{y}^{*L})+(\bar{y}^*-\bar{y})\Sigma(y_i^*-\bar{y}^*)+(\bar{y}^{*R}-\bar{y}^R)\Sigma(y_i^{*R}-\bar{y}^{*R})]=0$$

and the amount η is reduced to

$$\eta=2\Sigma(y_i^L-y_i^{*L})(y_i^{*L}-\bar{y}^L)+2\Sigma(y_i-y_i^*)(y_i^*-\bar{y}\ )+$$
$$+2\Sigma(y_i^R-y_i^{*R})(y_i^{*R}-\bar{y}^R)=$$
$$=2\Sigma(y_i^L-y_i^{*L})y_i^{*L}-2\Sigma(y_i^L-y_i^{*L})\bar{y}^L+2\Sigma(y_i-y_i^*)y_i^*+$$
$$-2\Sigma(y_i-y_i^*)\bar{y}\ +2\Sigma(y_i^R-y_i^{*R})y_i^{*R}-2\Sigma(y_i^R-y_i^{*R})\bar{y}^R. \tag{7}$$

Moreover, as it is

$$y_i^{*L}=a+bx_i^L+cz_i^R,\quad y_i^*=a+bx_i+cz_i,$$
$$y_i^{*R}=a+bx_i^R+cz_i^L$$

it is also

$$2\Sigma(y_i^L-y_i^{*L})y_i^{*L}+2\Sigma(y_i-y_i^*)y_i^*+2\Sigma(y_i^R-y_i^{*R})y_i^{*R}=0\ .$$

By replacing expressions of the theoretical values in the latter equation, we obtain

$$=2[\Sigma(y_i^L-y_i^{*L})(a+bx_i^L+cz_i^R)+\Sigma(y_i-y_i^*)(a+bx_i+cz_i)+$$
$$+\Sigma(y_i^R-y_i^{*R})(a+bx_i^R+cz_i^L)]=$$
$$=2\{a[(\Sigma y_i^L+\Sigma y_i+\Sigma y_i^R)-(\Sigma y_i^{*L}+\Sigma y_i^*+\Sigma y_i^{*R})]+$$
$$+b[(\Sigma y_i^L x_i^L+\Sigma y_i x_i+\Sigma y_i^R x_i^R)-(\Sigma y_i^{*L}x_i^L+\Sigma y_i^* x_i+\Sigma y_i^{*R}x_i^R)]+$$
$$+c[(\Sigma y_i^L z_i^R+\Sigma y_i z_i+\Sigma y_i^R z_i^L)-(\Sigma y_i^{*L}z_i^R+\Sigma y_i^* z_i+\Sigma y_i^{*R}z_i^L)]\}$$

where

$$\left(\Sigma y_i^L+\Sigma y_i+\Sigma y_i^R\right)-\left(\Sigma y_i^{*L}+\Sigma y_i^*+\Sigma y_i^{*R}\right)=0$$

for (3),

$$\left(\Sigma y_i^L x_i^L+\Sigma y_i x_i+\Sigma y_i^R x_i^R\right)-\left(\Sigma y_i^{*L}x_i^L+\Sigma y_i^* x_i+\Sigma y_i^{*R}x_i^R\right)=0$$

for (4),

$$\left(\Sigma y_i^L z_i^R+\Sigma y_i z_i+\Sigma y_i^R z_i^L\right)-\left(\Sigma y_i^{*L}z_i^R+\Sigma y_i^* z_i+\Sigma y_i^{*R}z_i^L\right)=0$$

for (5).

Finally the expression (7) can be reduced to

$$\eta=-2\Sigma(y_i^L-y_i^{*L})\bar{y}^L-2\Sigma(y_i-y_i^*)\bar{y}\ -2\Sigma(y_i^R-y_i^{*R})\bar{y}^R=$$
$$-2(\bar{y}^L\Sigma e_i^L+\bar{y}\ \Sigma e_i+\bar{y}^R\Sigma e_i^R)\ .$$

Note that, if the residual deviance equals zero, also η and $d(\overline{Y},\overline{Y}^*)^2$ equal zero, because theoretical and empirical average values of Y coincide for each observation.

Therefore:

- if the regression deviance equals zero, then the model has no forecasting ability, because the sum of the components of the i-th estimated fuzzy value equal the sum of the sample average components (i = 1 ,..., n). Actually, if Dev (regr) = 0, for each i we have

$$\Sigma y_i^{*L}+\Sigma y_i^*+\Sigma y_i^{*R}=\Sigma y_i^L+\Sigma y_i+\Sigma y_i^R \quad =>$$
$$=> \quad ny_i^{*L}+ny_i^*+ny_i^{*R}=n\bar{y}^L+n\bar{y}\ +n\bar{y}^R \quad =>$$
$$=> \quad y_i^{*L}+y_i^*+y_i^{*R}=\bar{y}^L+\bar{y}\ +\bar{y}^R\ ;$$

- if the residual deviance equals zero, the relationship between dependent variable and independent ones is well represented by the estimated model. In this case, the total deviance is entirely explained by the regression deviance.

As usual, the largest the regression deviance is (the smallest the residual deviance is), the better the model fits data.

5 CONCLUSIONS

In this work, starting from a multivariate generalization of the Fuzzy Least Square Regression, we have decomposed the total deviance of the dependent variable according to the metric proposed by Diamond (1988). In particular we have obtained the expression of two additional components of variability, besides the regression deviance and the residual one, which arise from the inequality between theoretical and empirical values of the average fuzzy dependent variable (unlike in the OLS estimation procedure for crisp variables).

REFERENCES

Campobasso, F., Fanizzi, A., Tarantini, M., 2008, Fuzzy Least Square Regression, *Annals of Department of Statistical Sciences, University of Bari, Italy, 229-243.*

Diamond, P. M., 1988. Fuzzy Least Square, *Information Sciences*, 46:141-157.

Kao, C.,Chyu, C.L., 2003, Least-squares estimates in fuzzy regression analysis, *European Journal of Operational Research*, 148:426-435.

Takemura, K., 2005. Fuzzy least squares regression analysis for social judgment study, *Journal of Advanced Intelligent Computing and Intelligent Informatics*, 9(5), 461:466.