

A multiple linear regression model for LR fuzzy random variables

Maria Brigida Ferraro*, Paolo Giordani*

*Department of Statistics, Probability and Applied Statistics
SAPIENZA University of Rome, Rome, Italy*

Abstract

In standard regression analysis the relationship between one (response) variable and a set of (explanatory) variables is investigated. In a classical framework the response is affected by probabilistic uncertainty (randomness) and, thus, treated as a random variable. However, the data can also be subjected to other kinds of uncertainty, such as imprecision. A possible way to manage all of these uncertainties is represented by the concept of fuzzy random variable (FRV). The most common class of FRVs is the LR family (LR FRV), which allows us to express every FRV in terms of three random variables, namely, the center, the left spread and the right spread. In this work, limiting our attention to the LR FRV class, we address the linear regression problem in presence of one or more imprecise random elements. The procedure for estimating the model parameters and the determination coefficient are discussed, and the hypothesis testing problem is addressed following a bootstrap approach. Furthermore, in order to illustrate how the proposed model works in practice, the results of a real-life example are given.

Keywords: LR fuzzy data, regression models, least squares approach, bootstrap procedure

*Corresponding author

Email addresses: mariabrigida.ferraro@uniroma1.it (Maria Brigida Ferraro),
paolo.giordani@uniroma1.it (Paolo Giordani)

1. Introduction

In the literature a great deal of attention has been paid to the management of uncertain information. We can roughly distinguish two sources of uncertainty, namely, randomness and imprecision. In the case of randomness the available information is uncertain because we do not know the (precise) outcome of a (random) mechanism. In general, randomness is limited to the data generation process and it can be dealt with probability theory (probabilistic uncertainty). In contrast with randomness, imprecision is connected to the uncertainty concerning the placement of an outcome in a given class and, thus, it can be seen as non-probabilistic uncertainty.

The different sources of uncertainty are not exclusive but can occur together. A possible way to cope with imprecision is represented by fuzzy set theory (Zadeh, 1965). This allows us to express the imprecise information in terms of fuzzy sets. When such information is also affected by randomness, the concept of fuzzy random variable (FRV) can be adopted (Puri & Ralescu, 1986). In this work we aim at investigating the linear regression problem when the data are random and imprecise. This problem has been deeply analyzed by coping with the different sources of uncertainty in a separate way. With particular reference to imprecise data, at least three approaches can be distinguished. The first one is the possibilistic approach. Originally introduced by Tanaka *et al.* (1982), its basic idea is that the regression model is intrinsically fuzzy because there does not exist a “true” relationship between the response variable and the explanatory ones. This is done by detecting fuzzy regression coefficients such that the fuzziness of the estimated response variable is minimized. Other works about possibilistic regression can be found in, e.g., Tanaka & Watada (1988), Tanaka *et al.* (1995) and Guo & Tanaka (2006). Another approach is the least squares one, in which a suitable dissimilarity measure between the observed and the estimated response variable must be introduced and the model parameters are estimated by minimizing such a dissimilarity measure. See, for instance, Celminš (1987), Diamond (1988), Chang & Lee (1996), D’Urso (2003), Coppi *et al.* (2006), Bargiela *et al.* (2007), Lu & Wang (2009). Generally speaking, the possibilistic and least squares approaches could also be used when the fuzzy data are affected by randomness. Unfortunately, this is simply done by overlooking it. The third line of research, which we may call fuzzy-probabilistic approach, consists of explicitly taking into account randomness for estimating the regression parameters and assessing their properties. Works belonging to this approach

can be found in, e.g., Körner & Näther (1998), Krätschmer (2006a, 2006b), Näther (2006), Ferraro *et al.* (2009a), González-Rodríguez *et al.* (2009). Note, however, that a few assignments of the previously mentioned papers to a given approach could be debatable.

The available proposals for linear regression in the presence of imprecise data can also be distinguished with respect to the nature (precise or imprecise) of the response and explanatory variables. In this paper, we assume to deal with imprecise response and explanatory variables, affected by randomness, and we approach the linear regression problem from the fuzzy-probabilistic viewpoint. Limiting our attention to the so-called *LR* fuzzy family, this is achieved by proposing a new linear regression model exploiting the potentialities of FRVs. As we will see, the parameters can be expressed in terms of the moments of real random variables. To estimate the parameters a closed form solution will be provided and their statistical properties will be investigated. The paper is organized as follows. In the next section, the concepts of (*LR*) fuzzy sets and FRVs are recalled. Then, Section 3 focuses on the proposed linear regression model. The estimation of the model parameters is discussed in Section 4. Section 5 addresses the hypothesis testing problem. To do it, the bootstrap approach is adopted and a simulation experiment is carried out in order to evaluate the quality of the bootstrap tests. Finally, the results of a real-life application are reported in Section 6 and some concluding remarks are given in Section 7.

2. Preliminaries

Given a universe U of elements, a fuzzy set \tilde{A} is a subset of U defined through the so-called *membership function* $\mu_{\tilde{A}}(x)$, $\forall x \in U$. For a generic $x \in U$, the membership function expresses the extent to which x belongs to \tilde{A} . Such a degree ranges from 0 (complete non-membership) to 1 (complete membership). The fuzzy set \tilde{A} is interpreted as a property (for instance the concept of “good”) and the membership function gives how well a generic x in the universe U (i.e., the scale $[0, 10]$) is able to characterize the property (i.e., what “good” is). In other words, $\mu_{\tilde{A}}(8) = 0.9$ is the degree of truth (0.9) of “good” concerning number 8 (8 characterizes “good” with a degree equal to 0.9).

A particular class of fuzzy sets is the LR family, whose members are the so-called *LR fuzzy numbers*. The space of the LR fuzzy numbers is denoted by \mathcal{F}_{LR} . A nice property of the LR family is that its elements can be deter-

mined uniquely in terms of the mapping $s : \mathcal{F}_{LR} \rightarrow \mathbb{R}^3$, i.e., $s(\tilde{A}) = s_{\tilde{A}} = (A^m, A^l, A^r)$. This implies that \tilde{A} can be expressed by means of three real-valued parameters, namely, the center (A^m) and the (non-negative) left and right spreads (A^l and A^r , respectively). In what follows it is indistinctly used $\tilde{A} \in \mathcal{F}_{LR}$ or (A^m, A^l, A^r) .

The arithmetics considered in \mathcal{F}_{LR} are the natural extensions of the Minkowski sum and the product by a positive scalar for interval. Going into detail, the sum of \tilde{A} and \tilde{B} in \mathcal{F}_{LR} is the LR fuzzy number $\tilde{A} + \tilde{B}$ so that

$$(A^m, A^l, A^r) + (B^m, B^l, B^r) = (A^m + B^m, A^l + B^l, A^r + B^r),$$

and the product of $\tilde{A} \in \mathcal{F}_{LR}$ by a scalar γ is the LR fuzzy number $\gamma\tilde{A}$ so that

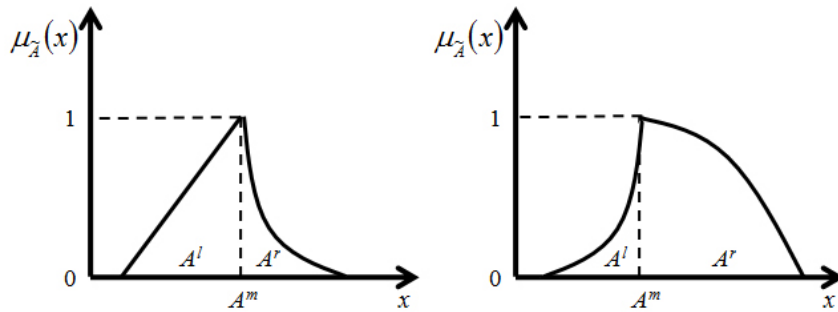
$$\gamma(A^m, A^l, A^r) = \begin{cases} (\gamma A^m, \gamma A^l, \gamma A^r) & \gamma > 0, \\ (\gamma A^m, -\gamma A^r, -\gamma A^l) & \gamma < 0, \\ 1_{\{0\}} & \gamma = 0. \end{cases}$$

The membership function of $\tilde{A} \in \mathcal{F}_{LR}$ can be written as

$$\mu_{\tilde{A}}(x) = \begin{cases} L\left(\frac{A^m - x}{A^l}\right) & x \leq A^m, \\ R\left(\frac{x - A^m}{A^r}\right) & x \geq A^m, \end{cases} \quad (1)$$

where the functions L and R are particular decreasing shape functions from \mathbb{R}^+ to $[0, 1]$ such that $L(0) = R(0) = 1$ and $L(x) = R(x) = 0, \forall x \in \mathbb{R} \setminus [0, 1]$ (see Fig. 1).

Figure 1: Examples of LR fuzzy numbers



\tilde{A} is a *triangular* fuzzy number if (1) takes the form

$$\mu_{\tilde{A}}(x) = \begin{cases} 0 & x \leq A^m - A^l, \\ 1 - \frac{A^m - x}{A^l} & A^m - A^l \leq x \leq A^m, \\ 1 - \frac{x - A^m}{A^r} & A^m \leq x \leq A^m + A^r, \\ 0 & x \geq A^m + A^r. \end{cases} \quad (2)$$

The α -level set ($0 < \alpha \leq 1$) of \tilde{A} can be defined as the non-empty compact convex subset of \mathbb{R} , A_α , such that $A_\alpha = \{x \in U : \mu_{\tilde{A}}(x) \geq \alpha\}$. If $\alpha = 0$, $A_0 = cl(x \in \mathbb{R} : \mu_{\tilde{A}}(x) > 0)$

The definition of α -level set is connected with that of fuzzy random variable (FRV) in the Puri and Ralescu sense (Puri & Ralescu, 1986). Note that in the sequel we limit our attention to FRVs of LR type (in brief *LR FRV*). Let (Ω, A, P) be a probability space, an *LR FRV* is a mapping $\tilde{X} : \Omega \rightarrow \mathcal{F}_{LR}$ such that the α -level set X_α is a random compact convex set for any $\alpha \in [0, 1]$. As for non-fuzzy random variables, it is possible to determine the moments for a FRV. To this purpose, it is necessary to introduce a suitable metric $D_{LR}^2(\tilde{X}, \tilde{Y}) = \langle s_{\tilde{X}} - s_{\tilde{Y}}, s_{\tilde{X}} - s_{\tilde{Y}} \rangle_{LR}$ for LR fuzzy numbers, where $\langle \cdot \rangle_{LR}$ denotes the inner product, so that $(\mathcal{F}_{LR}, D_{LR})$ is a metric space. Note that we can also express the moments according to the mapping s . The expectation of a FRV \tilde{X} is the unique fuzzy set $E(\tilde{X}) \in \mathcal{F}_{LR}$ such that $(E(\tilde{X}))_\alpha = E(X_\alpha)$ provided that $E\|\tilde{X}\|_{D_{LR}}^2 < \infty$. Also, on the basis of the mapping s , we can observe that $s_{E(\tilde{X})} = (E(X^m), E(X^l), E(X^r))$. The variance of \tilde{X} can be defined as $\sigma_{\tilde{X}}^2 = var(\tilde{X}) = E[(D_{LR}^2(\tilde{X}, E(\tilde{X})))]$. In terms of s , it is $\sigma_{\tilde{X}}^2 = var(\tilde{X}) = E \langle s_{\tilde{X}} - s_{E(\tilde{X})}, s_{\tilde{X}} - s_{E(\tilde{X})} \rangle_{LR}$. In a similar way, the covariance between two FRVs \tilde{X} and \tilde{Y} is $\sigma_{\tilde{X}, \tilde{Y}} = cov(\tilde{X}, \tilde{Y}) = E \langle s_{\tilde{X}} - s_{E(\tilde{X})}, s_{\tilde{Y}} - s_{E(\tilde{Y})} \rangle_{LR}$. In order to better characterize the moments of FRVs, we now introduce a particular distance measure. Following Yang & Ko (1996), we have

$$D_{LR}^2(\tilde{X}, \tilde{Y}) = (X^m - Y^m)^2 + [(X^m - \lambda X^l) - (Y^m - \lambda Y^l)]^2 + [(X^m + \rho X^r) - (Y^m + \rho Y^r)]^2. \quad (3)$$

In (3), the parameters $\lambda = \int_0^1 L^{-1}(\omega) d\omega$ and $\rho = \int_0^1 R^{-1}(\omega) d\omega$ play the role of taking into account the shape of the membership function. For instance, if the membership function takes the form reported in (2), it is $\lambda = \rho = \frac{1}{2}$. For what follows it is necessary to embed the space \mathcal{F}_{LR} into \mathbb{R}^3 by preserving

the metric. For this reason a generalization of the Yang and Ko metric has been derived (see Ferraro *et al.* 2009a). Given $a = (a_1, a_2, a_3)$ and $b = (b_1, b_2, b_3) \in \mathbb{R}^3$, it is

$$D_{\lambda\rho}^2(a, b) = (a_1 - b_1)^2 + ((a_1 - \lambda a_2) - (b_1 - \lambda b_2))^2 + ((a_1 + \rho a_3) - (b_1 + \rho b_3))^2, \quad (4)$$

where $\lambda, \rho \in \mathbb{R}^+$. The distance in (4) will be used in the sequel as a tool for quantifying errors in the regression model we are going to introduce.

3. The linear regression model for *LR FRVs*

The available information refers to an *LR* fuzzy response variable \tilde{Y} and p *LR* fuzzy explanatory variables $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p$ observed on a sample of n statistical units, $\{\tilde{Y}_i, \tilde{X}_{1i}, \tilde{X}_{2i}, \dots, \tilde{X}_{pi}\}_{i=1, \dots, n}$. We are interested in analyzing the relationship between \tilde{Y} and $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p$. The idea is to model the center and the spreads of \tilde{Y} by means of the centers and the spreads of $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p$. However, in doing so, attention should be paid to the non-negativity of the spreads of \tilde{Y} . To overcome this problem one can either solving a non-negative regression problem (see, e.g., Lawson & Hanson, 1995) or modelling a transformation of the spreads of \tilde{Y} (the new “response variable”) by means of the centers and the spreads of $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p$. The former choice is a numerical procedure yielding a dependence between the errors and the explanatory variables (Liew, 1976) and not allowing to formalize a realistic theoretical model and to obtain a complete analytical solution. We thus propose to consider the latter choice introducing two invertible functions $g : (0, +\infty) \rightarrow \mathbb{R}$ and $h : (0, +\infty) \rightarrow \mathbb{R}$. The linear regression model can be formalized as

$$\begin{cases} Y^m = \underline{X} \underline{a}'_m + b_m + \varepsilon_m, \\ g(Y^l) = \underline{X} \underline{a}'_l + b_l + \varepsilon_l, \\ h(Y^r) = \underline{X} \underline{a}'_r + b_r + \varepsilon_r, \end{cases} \quad (5)$$

where $\underline{X} = (X_1^m, X_1^l, X_1^r, \dots, X_p^m, X_p^l, X_p^r)$ is the row-vector of length $3p$ of all the components of the explanatory variables, $\varepsilon_m, \varepsilon_l$ and ε_r are real-valued random variables with $E(\varepsilon_m|\underline{X}) = E(\varepsilon_l|\underline{X}) = E(\varepsilon_r|\underline{X}) = 0$, $\underline{a}_m = (a_{mm}^1, a_{ml}^1, a_{mr}^1, \dots, a_{mm}^p, a_{ml}^p, a_{mr}^p)$, $\underline{a}_l = (a_{lm}^1, a_{ll}^1, a_{lr}^1, \dots, a_{lm}^p, a_{ll}^p, a_{lr}^p)$ and $\underline{a}_r = (a_{rm}^1, a_{rl}^1, a_{rr}^1, \dots, a_{rm}^p, a_{rl}^p, a_{rr}^p)$ are row-vectors of length $3p$ of the parameters related to \underline{X} . The generic a_{ij}^t is the regression coefficient between the component $i \in \{m, l, r\}$ of \tilde{Y} (where m, l and r refer to the center Y^m and

the transformations of the spreads $g(Y^l)$ and $h(Y^r)$, respectively) and the component $j \in \{m, l, r\}$ of the explanatory variables \tilde{X}^t , $t = 1, \dots, p$, (where m , l and r refer to the corresponding center, left spread and right spread). For example, a_{mr}^3 represents the relationship between the right spread of the explanatory variable \tilde{X}^3 (X_3^r) and the center of the response, Y^m . The covariance matrix of \underline{X} is denoted by $\Sigma_{\underline{X}} = E [(\underline{X} - E\underline{X})'(\underline{X} - E\underline{X})]$ and Σ stands for the covariance matrix of $(\varepsilon_m, \varepsilon_l, \varepsilon_r)$, with variances, $\sigma_{\varepsilon_m}^2$, $\sigma_{\varepsilon_l}^2$ and $\sigma_{\varepsilon_r}^2$, strictly positive and finite. The population parameters can then be expressed, as usual, in terms of some moments related to real random variables. We get

$$\begin{aligned} \underline{a}'_m &= \{\Sigma_{\underline{X}}\}^{-1} E \left[(\underline{X} - E\underline{X})' (Y^m - EY^m) \right], \\ \underline{a}'_l &= \{\Sigma_{\underline{X}}\}^{-1} E \left[(\underline{X} - E\underline{X})' (g(Y^l) - Eg(Y^l)) \right], \\ \underline{a}'_r &= \{\Sigma_{\underline{X}}\}^{-1} E \left[(\underline{X} - E\underline{X})' (h(Y^r) - Eh(Y^r)) \right], \\ b_m &= E(Y^m | \underline{X}) - E\underline{X} \{\Sigma_{\underline{X}}\}^{-1} E \left[(\underline{X} - E\underline{X})' (Y^m - EY^m) \right], \\ b_l &= E(g(Y^l) | \underline{X}) - E\underline{X} \{\Sigma_{\underline{X}}\}^{-1} E \left[(\underline{X} - E\underline{X})' (g(Y^l) - Eg(Y^l)) \right], \\ b_r &= E(h(Y^r) | \underline{X}) - E\underline{X} \{\Sigma_{\underline{X}}\}^{-1} E \left[(\underline{X} - E\underline{X})' (h(Y^r) - Eh(Y^r)) \right]. \end{aligned}$$

The above expressions are useful to prove some statistical properties of the estimators introduced in the next section.

Remark 1. *In the simple case, that is, $p = 1$, the model (5) takes the form*

$$\begin{cases} Y^m = a_{mm}X^m + a_{ml}X^l + a_{mr}X^r + b_m + \varepsilon_m, \\ g(Y^l) = a_{lm}X^m + a_{ll}X^l + a_{lr}X^r + b_l + \varepsilon_l, \\ h(Y^r) = a_{rm}X^m + a_{rl}X^l + a_{rr}X^r + b_r + \varepsilon_r. \end{cases}$$

Remark 2. *When the explanatory variables are real-valued, the model (5) reduces to the regression model proposed by Ferraro et al. (2009a).*

3.1. The determination coefficient

Since the total variation of the response can be written in terms of variances and covariances of real random variables, by taking advantage of their properties it can be decomposed in the variation not depending on the model

and that explained by the model. In particular, let \tilde{Y} and $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p$ be LR FRVs satisfying the linear model (5) so that the errors are uncorrelated with \underline{X} , by indicating $Y^T = (Y^m, g(Y^l), h(Y^r))$, we obtain

$$\begin{aligned} E [D_{\lambda\rho}^2(Y^T, E(Y^T))] &= E [D_{\lambda\rho}^2(Y^T, E(Y^T|\underline{X}))] \\ &+ E [D_{\lambda\rho}^2(E(Y^T|\underline{X}), E(Y^T))]. \end{aligned} \quad (6)$$

Based on the decomposition of the total variation (6), it is possible to define the following determination coefficient,

Definition 1. Let \tilde{Y} be the LR FRV of the linear model (5), the determination coefficient can be defined as

$$R^2 = \frac{E [D_{\lambda\rho}^2(E(Y^T|\underline{X}), E(Y^T))]}{E [D_{\lambda\rho}^2(Y^T, E(Y^T))]} = 1 - \frac{E [D_{\lambda\rho}^2(Y^T, E(Y^T|\underline{X}))]}{E [D_{\lambda\rho}^2(Y^T, E(Y^T))]} \quad (7)$$

This coefficient measures the degree of linear relationship. As in the classical case, it takes values in $[0, 1]$. In particular, $R^2 = 0$ indicates linear independence and when R^2 reaches the value 1, it shows that the variability of the response is completely explained by the model.

4. The estimation problem

4.1. Estimation of the regression parameters

The estimation problem of the regression parameters is faced by means of the Least Squares (LS) criterion. By using the generalized Yang-Ko metric $D_{\lambda\rho}^2$ written in matrix notation, the LS problem consists in looking for \hat{a}_m , \hat{a}_l , \hat{a}_r , \hat{b}_m , \hat{b}_l and \hat{b}_r such that

$$\Delta_{\lambda\rho}^2 = D_{\lambda\rho}^2((\underline{Y}^m, g(\underline{Y}^l), h(\underline{Y}^r)), ((\underline{Y}^m)^*, g^*(\underline{Y}^l), h^*(\underline{Y}^r))) \quad (8)$$

is minimized, where \underline{Y}^m , $g(\underline{Y}^l)$ and $h(\underline{Y}^r)$ are the $n \times 1$ vectors of the observed values and $(\underline{Y}^m)^* = \mathbf{X}\underline{a}'_m + \underline{1}b_m$, $g^*(\underline{Y}^l) = \mathbf{X}\underline{a}'_l + \underline{1}b_l$ and $h^*(\underline{Y}^r) = \mathbf{X}\underline{a}'_r + \underline{1}b_r$ are the theoretical ones being $\mathbf{X} = (\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n)'$ the $n \times 3p$ matrix of the explanatory variables.

Proposition 1. *The solution of the LS problem is*

$$\begin{aligned}
\hat{\underline{a}}'_m &= (\mathbf{X}^{c'} \mathbf{X}^c)^{-1} \mathbf{X}^{c'} \underline{Y}^{mc}, \\
\hat{\underline{a}}'_l &= (\mathbf{X}^{c'} \mathbf{X}^c)^{-1} \mathbf{X}^{c'} g(\underline{Y}^l)^c, \\
\hat{\underline{a}}'_r &= (\mathbf{X}^{c'} \mathbf{X}^c)^{-1} \mathbf{X}^{c'} h(\underline{Y}^r)^c, \\
\hat{b}_m &= \overline{Y^m} - \underline{\overline{X}} \hat{\underline{a}}'_m, \\
\hat{b}_l &= \overline{g(Y^l)} - \underline{\overline{X}} \hat{\underline{a}}'_l, \\
\hat{b}_r &= \overline{h(Y^r)} - \underline{\overline{X}} \hat{\underline{a}}'_r,
\end{aligned}$$

where

$$\begin{aligned}
\underline{Y}^{mc} &= \underline{Y}^m - \underline{\overline{1}} \overline{Y^m}, \\
g(\underline{Y}^l)^c &= g(\underline{Y}^l) - \underline{\overline{1}} \overline{g(Y^l)}, \\
h(\underline{Y}^r)^c &= h(\underline{Y}^r) - \underline{\overline{1}} \overline{h(Y^r)}
\end{aligned}$$

are the centered values of the response variables,

$$\mathbf{X}^c = \mathbf{X} - \underline{\overline{1}} \underline{\overline{X}}$$

is the centered matrix of the explanatory variables and, $\overline{Y^m}$, $\overline{g(Y^l)}$, $\overline{h(Y^r)}$ and $\underline{\overline{X}}$ denote, respectively, the sample means of Y^m , $g(Y^l)$, $h(Y^r)$ and \underline{X} .

Proof. In order to solve the minimization problem and to find the parameters estimators, we follow the usual procedure of equating to zero the partial derivatives of the objective function with respect to (w.r.t.) the parameters to be estimated, although we have to take into account that the regression parameters are related to some others. The objective function (8) can be exploited as

$$\begin{aligned}
\Delta_{\lambda\rho}^2 &= \|\underline{Y}^m - (\underline{Y}^m)^*\|^2 + \|(\underline{Y}^m - \lambda g(\underline{Y}^l)) - ((\underline{Y}^m)^* - \lambda g^*(\underline{Y}^l))\|^2 \\
&\quad + \|(\underline{Y}^m + \rho h(\underline{Y}^r)) - ((\underline{Y}^m)^* + \rho h^*(\underline{Y}^r))\|^2
\end{aligned}$$

and, after a little algebra, it can be written as

$$\begin{aligned}
\Delta_{\lambda\rho}^2 &= 3 \left(\underline{Y}^m - \mathbf{X}\underline{a}'_m - \underline{1}b_m \right)' \left(\underline{Y}^m - \mathbf{X}\underline{a}'_m - \underline{1}b_m \right) \\
&+ \lambda^2 \left(g(\underline{Y}^l) - \mathbf{X}\underline{a}'_l - \underline{1}b_l \right)' \left(g(\underline{Y}^l) - \mathbf{X}\underline{a}'_l - \underline{1}b_l \right) \\
&+ \rho^2 \left(h(\underline{Y}^r) - \mathbf{X}\underline{a}'_r - \underline{1}b_r \right)' \left(h(\underline{Y}^r) - \mathbf{X}\underline{a}'_r - \underline{1}b_r \right) \\
&- 2\lambda \left(\underline{Y}^m - \mathbf{X}\underline{a}'_m - \underline{1}b_m \right)' \left(g(\underline{Y}^l) - \mathbf{X}\underline{a}'_l - \underline{1}b_l \right) \\
&+ 2\rho \left(\underline{Y}^m - \mathbf{X}\underline{a}'_m - \underline{1}b_m \right)' \left(h(\underline{Y}^r) - \mathbf{X}\underline{a}'_r - \underline{1}b_r \right). \tag{9}
\end{aligned}$$

Starting from the estimation of b_l and b_r , we equate to zero the partial derivatives w.r.t b_l and b_r , respectively. It is easy to find that the minimum is attained at

$$b_l = \overline{g(Y^l)} - \overline{\mathbf{X}}\underline{a}'_l - \frac{1}{\lambda}\overline{Y^m} + \frac{1}{\lambda}\overline{\mathbf{X}}\underline{a}'_m + \frac{1}{\lambda}b_m, \tag{10}$$

$$b_r = \overline{h(Y^r)} - \overline{\mathbf{X}}\underline{a}'_r + \frac{1}{\rho}\overline{Y^m} - \frac{1}{\rho}\overline{\mathbf{X}}\underline{a}'_m - \frac{1}{\rho}b_m. \tag{11}$$

Since b_l and b_r depend on b_m , we have to substitute (10) and (11) in (9) before equating to zero the partial derivative of the objective function w.r.t. b_m . As a result, we obtain

$$b_m = \overline{Y^m} - \overline{\mathbf{X}}\underline{a}'_m.$$

Since the parameters b_m , b_l and b_r are expressed in terms of \underline{a}_m , \underline{a}_l and \underline{a}_r , to go on with the estimation procedure it is important to take this into account by substituting b_m , b_l and b_r in the objective function.

We consider the centered vectors \underline{Y}^{mc} , $g(\underline{Y}^l)^c$, $h(\underline{Y}^r)^c$ and the centered matrix \mathbf{X}^c to make it simpler to analyze the objective function that can be expressed

as follows

$$\begin{aligned}
\Delta_{\lambda\rho}^2 &= 3(\underline{Y}^{mc} - \mathbf{X}^c \underline{a}'_m)' (\underline{Y}^{mc} - \mathbf{X}^c \underline{a}'_m) \\
&+ \lambda^2 \left(g(\underline{Y}^l)^c - \mathbf{X}^c \underline{a}'_l \right)' \left(g(\underline{Y}^l)^c - \mathbf{X}^c \underline{a}'_l \right) \\
&+ \rho^2 \left(h(\underline{Y}^r)^c - \mathbf{X}^c \underline{a}'_r \right)' \left(h(\underline{Y}^r)^c - \mathbf{X}^c \underline{a}'_r \right) \\
&- 2\lambda (\underline{Y}^{mc} - \mathbf{X}^c \underline{a}'_m)' \left(g(\underline{Y}^l)^c - \mathbf{X}^c \underline{a}'_l \right) \\
&+ 2\rho (\underline{Y}^{mc} - \mathbf{X}^c \underline{a}'_m)' \left(h(\underline{Y}^r)^c - \mathbf{X}^c \underline{a}'_r \right). \tag{12}
\end{aligned}$$

Following the usual reasoning it is easy to check that

$$\underline{a}'_l = (\mathbf{X}^{c'} \mathbf{X}^c)^{-1} \mathbf{X}^{c'} g(\underline{Y}^l)^c - \frac{1}{\lambda} (\mathbf{X}^{c'} \mathbf{X}^c)^{-1} \mathbf{X}^{c'} \underline{Y}^{mc} + \frac{\underline{a}'_m}{\lambda}, \tag{13}$$

$$\underline{a}'_r = (\mathbf{X}^{c'} \mathbf{X}^c)^{-1} \mathbf{X}^{c'} h(\underline{Y}^r)^c + \frac{1}{\rho} (\mathbf{X}^{c'} \mathbf{X}^c)^{-1} \mathbf{X}^{c'} \underline{Y}^{mc} - \frac{1}{\rho} \underline{a}'_m. \tag{14}$$

The last step is the estimation of \underline{a}_m . Since this vector appears in (13) and (14) we need to substitute (13) and (14) in (12). By equating to 0 the partial derivative of (12) w.r.t. \underline{a}_m we get

$$\widehat{\underline{a}}'_m = (\mathbf{X}^{c'} \mathbf{X}^c)^{-1} \mathbf{X}^{c'} \underline{Y}^{mc}.$$

By making all the appropriate substitutions we also find

$$\begin{aligned}
\widehat{\underline{a}}'_l &= (\mathbf{X}^{c'} \mathbf{X}^c)^{-1} \mathbf{X}^{c'} g(\underline{Y}^l)^c, \\
\widehat{\underline{a}}'_r &= (\mathbf{X}^{c'} \mathbf{X}^c)^{-1} \mathbf{X}^{c'} h(\underline{Y}^r)^c, \\
\widehat{b}_m &= \overline{Y^m} - \overline{X} \widehat{\underline{a}}'_m, \\
\widehat{b}_l &= \overline{g(Y^l)} - \overline{X} \widehat{\underline{a}}'_l, \\
\widehat{b}_r &= \overline{h(Y^r)} - \overline{X} \widehat{\underline{a}}'_r.
\end{aligned}$$

□

Remark 3. *Since the LS estimators are written in terms of sample moments and taking into account the expression of the theoretical values, it can be shown that they are unbiased and strongly consistent.*

4.2. Estimation of the determination coefficient

In order to estimate the determination coefficient, it is worth introducing the next proposition about the decomposition of the total sum of squares.

Proposition 2. *Let \tilde{Y} and $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p$ be LR FRVs satisfying the linear model (5) observed on n statistical units, $\{\tilde{Y}_i, \tilde{X}_{1i}, \tilde{X}_{2i}, \dots, \tilde{X}_{pi}\}_{i=1, \dots, n}$. The total sum of squares, SST , is equal to the sum of the residual sum of squares, SSE , and the regression sum of squares, SSR , that is,*

$$SST = SSE + SSR. \quad (15)$$

In details,

(i) *the total sum of squares (SST) is*

$$\begin{aligned} SST = \|\underline{Y}^m - \underline{1}\overline{Y}^m\|^2 &+ \left\| (\underline{Y}^m - \lambda g(\underline{Y}^l)) - (\underline{1}\overline{Y}^m - \lambda \underline{1} \overline{g(\underline{Y}^l)}) \right\|^2 \\ &+ \left\| (\underline{Y}^m + \rho h(\underline{Y}^r)) - (\underline{1}\overline{Y}^m + \rho \underline{1} \overline{h(\underline{Y}^r)}) \right\|^2, \end{aligned}$$

(ii) *the residual sum of squares (SSE) is*

$$\begin{aligned} SSE = \|\underline{Y}^m - \widehat{\underline{Y}}^m\|^2 &+ \left\| (\underline{Y}^m - \lambda g(\underline{Y}^l)) - (\widehat{\underline{Y}}^m - \lambda \widehat{g(\underline{Y}^l)}) \right\|^2 \\ &+ \left\| (\underline{Y}^m + \rho h(\underline{Y}^r)) - (\widehat{\underline{Y}}^m + \rho \widehat{h(\underline{Y}^r)}) \right\|^2, \end{aligned}$$

(iii) *the regression sum of squares (SSR) is*

$$\begin{aligned} SSR = \|\widehat{\underline{Y}}^m - \underline{1}\overline{Y}^m\|^2 &+ \left\| (\widehat{\underline{Y}}^m - \lambda \widehat{g(\underline{Y}^l)}) - (\underline{1}\overline{Y}^m - \lambda \underline{1} \overline{g(\underline{Y}^l)}) \right\|^2 \\ &+ \left\| (\widehat{\underline{Y}}^m + \rho \widehat{h(\underline{Y}^r)}) - (\underline{1}\overline{Y}^m + \rho \underline{1} \overline{h(\underline{Y}^r)}) \right\|^2, \end{aligned}$$

where $\widehat{\underline{Y}}^m$, $\widehat{g(\underline{Y}^l)}$, $\widehat{h(\underline{Y}^r)}$ are the vectors of the estimated values, that is,

$$\widehat{\underline{Y}}^m = \mathbf{X}\widehat{\underline{a}}'_m + \underline{1}\widehat{b}_m, \quad \widehat{g(\underline{Y}^l)} = \mathbf{X}\widehat{\underline{a}}'_l + \underline{1}\widehat{b}_l, \quad \widehat{h(\underline{Y}^r)} = \mathbf{X}\widehat{\underline{a}}'_r + \underline{1}\widehat{b}_r.$$

Proof. The total sum of squares can be written as

$$\begin{aligned}
SST &= 3(\underline{Y}^m - \underline{1}\overline{Y}^m)'(\underline{Y}^m - \underline{1}\overline{Y}^m) \\
&+ \lambda^2 \left(g(\underline{Y}^l) - \underline{1}g(\overline{Y}^l) \right)' \left(g(\underline{Y}^l) - \underline{1}g(\overline{Y}^l) \right) \\
&+ \rho^2 \left(h(\underline{Y}^r) - \underline{1}h(\overline{Y}^r) \right)' \left(h(\underline{Y}^r) - \underline{1}h(\overline{Y}^r) \right) \\
&- 2\lambda (\underline{Y}^m - \underline{1}\overline{Y}^m)' \left(g(\underline{Y}^l) - \underline{1}g(\overline{Y}^l) \right) \\
&+ 2\rho (\underline{Y}^m - \underline{1}\overline{Y}^m)' \left(h(\underline{Y}^r) - \underline{1}h(\overline{Y}^r) \right). \tag{16}
\end{aligned}$$

By subtracting and adding \widehat{Y}^m in $(\underline{Y}^m - \underline{1}\overline{Y}^m)$, we get that $(\underline{Y}^m - \underline{1}\overline{Y}^m)'(\underline{Y}^m - \underline{1}\overline{Y}^m)$ is equal to

$$\begin{aligned}
& \left(\underline{Y}^m - \widehat{Y}^m + \widehat{Y}^m - \underline{1}\overline{Y}^m \right)' \left(\underline{Y}^m - \widehat{Y}^m + \widehat{Y}^m - \underline{1}\overline{Y}^m \right) \\
&= \left(\underline{Y}^m - \widehat{Y}^m \right)' \left(\underline{Y}^m - \widehat{Y}^m \right) + \left(\widehat{Y}^m - \underline{1}\overline{Y}^m \right)' \left(\widehat{Y}^m - \underline{1}\overline{Y}^m \right) \tag{17}
\end{aligned}$$

$$+ 2 \left(\underline{Y}^m - \widehat{Y}^m \right)' \left(\widehat{Y}^m - \underline{1}\overline{Y}^m \right) \tag{18}$$

The first two terms of (17) are the first terms of SSE and SSR, respectively. Now we prove that the term in (18) is equal to 0. Since $\widehat{Y}^m = \mathbf{X}\widehat{a}'_m + \underline{1}\widehat{b}_m$ where $\widehat{a}'_m = (\mathbf{X}^c\mathbf{X}^c)^{-1}\mathbf{X}^c\underline{Y}^{mc}$ and $\widehat{b}_m = \overline{Y}^m - \overline{X}\widehat{a}'_m$, it results

$$\begin{aligned}
& \left(\underline{Y}^m - \widehat{Y}^m \right)' \left(\widehat{Y}^m - \underline{1}\overline{Y}^m \right) \\
&= \left(\underline{Y}^m - \mathbf{X}\widehat{a}'_m - \underline{1}\overline{Y}^m + \underline{1}\overline{X}\widehat{a}'_m \right)' \left(\mathbf{X}\widehat{a}'_m + \underline{1}\overline{Y}^m - \underline{1}\overline{X}\widehat{a}'_m - \underline{1}\overline{Y}^m \right) \\
&= \left(\underline{Y}^{mc} - \mathbf{X}^c\widehat{a}'_m \right)' \left(\mathbf{X}^c\widehat{a}'_m \right) = \widehat{a}'_m\mathbf{X}^c\underline{Y}^{mc} - \widehat{a}'_m\mathbf{X}^c\underline{Y}^{mc} = 0
\end{aligned}$$

By using the same procedure for the other terms in (16), namely by subtracting and adding the corresponding estimate in each term, the thesis follows. \square

Definition 2. Let \widetilde{Y} and $\widetilde{X}_1, \widetilde{X}_2, \dots, \widetilde{X}_p$ be LR FRVs satisfying the linear model (5) observed on n statistical units, $\{\widetilde{Y}_i, \widetilde{X}_{1i}, \widetilde{X}_{2i}, \dots, \widetilde{X}_{pi}\}_{i=1, \dots, n}$. The estimator of the determination coefficient R^2 is

$$\widehat{R}^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}.$$

It represents the part of total sum of squares explained by the regression model, so it can be considered as a goodness-of-fit measure and it takes values in $[0, 1]$. Furthermore, it can be shown that \widehat{R}^2 is a strong consistent estimator.

5. Hypothesis testing

5.1. Hypothesis testing on the regression parameters

The parameters \underline{a}_m , \underline{a}_l and \underline{a}_r express the strength of the relationship between the response variable and the explanatory ones. Testing the explicative power of \underline{X} consists in testing that the vectors of coefficients \underline{a}_m , \underline{a}_l and \underline{a}_r are equal to $\underline{0}$. In general it is possible to test the null hypothesis

$$H_0 : \begin{pmatrix} \underline{a}'_m \\ \underline{a}'_l \\ \underline{a}'_r \end{pmatrix} = \begin{pmatrix} \underline{k}'_m \\ \underline{k}'_l \\ \underline{k}'_r \end{pmatrix}$$

against the alternative

$$H_1 : \begin{pmatrix} \underline{a}'_m \\ \underline{a}'_l \\ \underline{a}'_r \end{pmatrix} \neq \begin{pmatrix} \underline{k}'_m \\ \underline{k}'_l \\ \underline{k}'_r \end{pmatrix},$$

where \underline{k}_m , \underline{k}_l , and \underline{k}_r are real-valued vectors. Starting from Ferraro *et al.* (2009a), the test statistic to be used is $T_n = V'_n V_n$, where

$$V_n = \sqrt{n} \begin{pmatrix} \widehat{\underline{a}}'_m - \underline{k}'_m \\ \widehat{\underline{a}}'_l - \underline{k}'_l \\ \widehat{\underline{a}}'_r - \underline{k}'_r \end{pmatrix}.$$

It is important to stress that, since there are not generalized models for FRVs that can be used in practice and an asymptotic test works suitably for large size samples, the hypothesis testing problem has been approached by bootstrapping. The non-parametric bootstrap test is based on the following algorithm:

Bootstrap algorithm

Step 1: Compute the estimates $\hat{a}_m, \hat{a}_l, \hat{a}_r$ and the value of the statistic

$$T_n = V_n' V_n.$$

Step 2: Compute the bootstrap population fulfilling the null hypothesis,

$$\{(\underline{X}_i, Z_i^m, Z_i^l, Z_i^r)\}_{i=1, \dots, n}, \quad (19)$$

where

$$\begin{aligned} Z_i^m &= Y_i^m - \underline{X}_i \hat{a}_m' + \underline{X}_i k_m', \\ Z_i^l &= g(Y_i^l) - \underline{X}_i \hat{a}_l' + \underline{X}_i k_l', \\ Z_i^r &= h(Y_i^r) - \underline{X}_i \hat{a}_r' + \underline{X}_i k_r'. \end{aligned}$$

Step 3: Draw a sample of size n with replacement

$$\{(X_i^*, Z_{i^*}^m, Z_{i^*}^l, Z_{i^*}^r)\}_{i=1, \dots, n},$$

from the bootstrap population (19).

Step 4: Compute the bootstrap estimates $\hat{a}_m^*, \hat{a}_l^*, \hat{a}_r^*$ and the value of the bootstrap statistic

$$T_n^* = V_n^{*'} V_n^*.$$

Step 5: Repeat Steps 3 and 4 a large number B of times to get a set of B estimators, denoted by $\{T_{n1}^*, \dots, T_{nB}^*\}$.

Step 6: Compute the bootstrap p -value as the proportion of values in $\{T_{n1}^*, \dots, T_{nB}^*\}$ being greater than T_n .

5.2. Hypothesis testing on a single parameter

A particular case of the above hypothesis test on the regression parameters is referred to testing the significance of a single regression parameter. In this way it is possible to check if a given component of the explanatory variables is significantly related to the LR fuzzy response variable. For example, let \tilde{Y} and $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p$ be LR FRVs satisfying the linear model (5), to test

the significance of the left spread of the explanatory variable \tilde{X}_1 w.r.t. the center of the response variable \tilde{Y} , it is tested the following hypothesis

$$H_0 : a_{ml}^1 = 0$$

against the alternative

$$H_1 : a_{ml}^1 \neq 0.$$

As for the previous hypothesis test, according to the bootstrap approach, the above described algorithm can be adopted. The most relevant difference consists in considering a bootstrap population $\{(\underline{X}_i, Z_i^m, Z_i^l, Z_i^r)\}_{i=1, \dots, n}$, where

$$\begin{aligned} Z_i^m &= Y_i^m - \hat{a}_{ml}^1 X_1^l, \\ Z_i^l &= g(Y_i^l), \\ Z_i^r &= h(Y_i^r). \end{aligned}$$

5.3. Linear independence test

In this section a bootstrap linear independence test is introduced on the basis of Ferraro *et al.* (2009b). To test the null hypothesis $H_0 : R^2 = 0$ against the alternative $H_1 : R^2 > 0$, the test statistic $T_n = n\hat{R}^2$ is used. Once again, a bootstrap algorithm can be adopted. To obtain a bootstrap population fulfilling the null hypothesis, the residual variables $Z^m = Y^m - \underline{X} \hat{a}_m$, $Z^l = g(Y^l) - \underline{X} \hat{a}_l$ and $Z^r = h(Y^r) - \underline{X} \hat{a}_r$ must be considered. A sample of size n with replacement $\{(\underline{X}_i^*, Z_i^{m*}, Z_i^{l*}, Z_i^{r*})\}_{i=1, \dots, n}$ from the bootstrap population is drawn and the bootstrap statistic to be used is

$$T_n^* = n \frac{\sum_{i=1}^n D_{\lambda\rho}^2(\widehat{Z}_i^{*T}, \overline{Z^{*T}})}{\sigma_{YT}^2},$$

where $Z_i^{*T} = (Z_i^{m*}, Z_i^{l*}, Z_i^{r*})$.

5.4. Simulation study

Several bootstrap algorithms have been proposed to obtain bootstrap p -values for testing hypotheses about the determination coefficient and the regression parameters of (5). By means of a simulation experiment we aimed at investigating whether the obtained p -values work as such, that is, if we find a bootstrap p -value equal to 0.05 we would like to conclude from this that the

true p -value (i.e., that obtained if we knew the distribution function) is 0.05. The simulation study concerned the test on a single regression parameter and the linear independence test. During the experiment we employed $B = 1000$ replications of the bootstrap estimator and we carried out 10.000 iterations of the test at three different nominal significance levels ($\alpha = 0.01, 0.05, 0.1$) for different sample sizes ($n = 30, 50, 100, 200, 300$). We considered the case of two LR fuzzy explanatory variables \tilde{X}_1 and \tilde{X}_2 . We dealt with the following real random variables: X_1^m and X_2^m , behaving as $Norm(0, 1)$ random variables, X_1^l and X_2^l as χ_1^2 , X_1^r and X_2^r as χ_2^2 , Y_m as $Norm(0, 1)$, $Y_2 = g(Y_l)$ and $Y_3 = h(Y_r)$ as $Norm(0, 0.5)$.

With respect to the hypothesis testing on a single parameter, we considered the test $H_0 : a_{mm}^1 = 0$ against the alternative $H_1 : a_{mm}^1 \neq 0$. The empirical percentages of rejection under H_0 are given in Table 1.

Table 1: Empirical percentages of rejection under the hypothesis $H_0 : a_{mm}^1 = 0$.

$n \setminus \alpha \times 100$	1	5	10
30	0.75	4.13	8.81
50	1.06	5.67	10.14
100	1.28	5.55	10.85
200	1.28	5.42	10.57
300	1.16	5.62	10.12

With respect to the linear independence test ($H_0 : R^2 = 0$ against $H_1 : R^2 > 0$), the empirical percentages of rejection under H_0 are reported in Table 2.

Table 2: Empirical percentages of rejection under the hypothesis of linear independence.

$n \setminus \alpha \times 100$	1	5	10
30	0.31	2.6	6.87
50	0.79	4.74	9.59
100	1.13	5.65	10.63
200	1.31	5.35	10.77
300	1.09	4.92	10.01

All in all, from Tables 1 and 2 we can conclude that the bootstrap p -values are fairly good approximations of the true p -values in most cases. As one may expect, this especially holds for increasing values of n ($n > 30$).

6. A real-case study

In a recent study about the student satisfaction of a course the subjective judgements/perceptions were observed on a sample of $n = 64$ students. To formalize the problem we defined $\Omega = \{\text{sets of students that attend the course}\}$ endowed with the Borel σ -field. Since the observations were arbitrarily chosen, P is the uniform distribution over Ω . For any $i \in \Omega$, three characteristics were observed. These were the overall assessment of the course, the assessment of the teaching staff and the assessment of the course content. Such an information was managed in terms of triangular fuzzy variables (hence $\lambda = \rho = 1/2$). In fact, to represent the subjective judgements/perceptions, the students were invited to draw a triangular fuzzy number for every characteristic. The considered support went from 0 (dissatisfaction) to 100 (full satisfaction). The students were informed to place the center where they wished to represent their average judgement/perception and the lower and upper bounds of the triangular fuzzy number where they wished to represent their worst and best judgement/perception, respectively. Note that the students were informed to compute the average, minimum and maximum values w.r.t. the variability of their subjective judgements/perceptions depending on the different course contents and/or members of the teaching staff. For analyzing the linear relationship of the overall assessment of the course (\tilde{Y}) on the assessment of the teaching staff (\tilde{X}_1) and the assessment of the course contents (\tilde{X}_2) (see Table 3), the proposed linear regression model was employed. To overcome the problem about the non-negativity of spreads estimates, we used the logarithmic transformation (that is, $g = h = \ln$). Through the LS procedure we obtained the following estimated model

$$\begin{cases} \widehat{Y}^m = 1.07X_1^m + 0.15X_1^l - 0.05X_1^r \\ \quad - 0.18X_2^m - 0.88X_2^l + 0.77X_2^r + 2.98 \\ \widehat{Y}^l = \exp(0.01X_1^m + 0.02X_1^l + 0.02X_1^r \\ \quad + 0.00X_2^m + 0.03X_2^l + 0.01X_2^r + 0.62) \\ \widehat{Y}^r = \exp(0.00X_1^m + 0.03X_1^l - 0.02X_1^r \\ \quad - 0.01X_2^m + 0.03X_2^l + 0.01X_2^r + 2.10) \end{cases}$$

To test the significance of every single regression parameter we computed the bootstrap p -values given in Table 4. Note that we set $B = 1000$. With respect to the model for the center of \tilde{Y} , we can see that, considering a significance level $\alpha = 0.05$, both the centers of \tilde{X}_1 and \tilde{X}_2 are significant. Furthermore, also the spreads of \tilde{X}_2 significantly affect the response Y^m . Thus,

Table 3: Overall assessment of the course (Y^m, Y^l, Y^r), Assessment of the teaching staff (X_1^m, X_1^l, X_1^r), Assessment of the course content (X_2^m, X_2^l, X_2^r) of the course.

Y^m	Y^l	Y^r	X_1^m	X_1^l	X_1^r	X_2^m	X_2^l	X_2^r
93	7	7	87	9	7	75	10	8
90	10	10	80	10	10	60	10	30
80	20	10	80	10	20	40	20	13
76	18	14	77	17	15	50	15	15
52	11	12	75	10	5	88	18	2
90	10	10	86	12	11	80	13	17
90	10	10	94	7	6	67	10	14
80	10	20	90	10	10	81	16	19
80	10	10	80	10	10	80	10	10
70	10	15	80	10	20	50	10	10
80	3	3	93	4	7	72	6	8
..

we can conclude that taking into account the spreads information of the explanatory variables is a value added. By inspecting Table 4 we can observe that, considering $\alpha = 0.05$, only some of the components of \tilde{X}_2 are significantly related to the transformed spreads of \tilde{Y} . When considering $\alpha = 0.10$, we find that also some components of \tilde{X}_1 play a significant role in explaining the transformed spread of \tilde{Y} ($g(Y^l)$). For the estimated model it resulted $\hat{R}^2 = 0.7526$, hence approximately almost 75.26% of the total variation of the overall assessment of the course is explained by the model. Furthermore, by applying the bootstrap procedure to test the linear independence (with $B = 1000$) a p -value equal to 0 was obtained, so the null hypothesis should be rejected.

7. Concluding remarks

In this work a new linear regression model for LR fuzzy response and explanatory variables has been introduced and analyzed, by taking into account different kinds of uncertainty. In particular, through a formalization in terms of FRVs, we have coped with the randomness and the imprecision of the data. Furthermore, it has been dealt the problem of the non-negativity

Table 4: Hypothesis testing on each regression parameter. The underlined values are significant at $\alpha = 0.05$, whereas the values in bold also at $\alpha = 0.10$.

	estimate	<i>p</i> -value		estimate	<i>p</i> -value
\hat{a}_{mm}^1	<u>1.07</u>	<u>0.00</u>	\hat{a}_{mm}^2	<u>-0.18</u>	<u>0.02</u>
\hat{a}_{ml}^1	0.15	0.70	\hat{a}_{ml}^2	<u>-0.88</u>	<u>0.00</u>
\hat{a}_{mr}^1	-0.05	0.83	\hat{a}_{mr}^2	<u>0.77</u>	<u>0.00</u>
\hat{a}_{lm}^1	0.01	0.10	\hat{a}_{lm}^2	0.00	0.86
\hat{a}_{ll}^1	0.02	0.39	\hat{a}_{ll}^2	<u>0.03</u>	<u>0.02</u>
\hat{a}_{lr}^1	0.02	0.08	\hat{a}_{lr}^2	0.01	0.01
\hat{a}_{rm}^1	0.00	0.59	\hat{a}_{rm}^2	<u>-0.01</u>	<u>0.03</u>
\hat{a}_{rl}^1	0.03	0.21	\hat{a}_{rl}^2	0.03	0.09
\hat{a}_{rr}^1	-0.02	0.25	\hat{a}_{rr}^2	0.01	0.21

of the spreads of the response by using suitable transformation functions. In this way, by a least squares approach, analytic estimators of the regression parameters, fulfilling some statistical properties, have been obtained. Some inferential procedures have been developed. In particular, tests on the significance of the regression parameters have been stated by bootstrapping. Furthermore, a determination coefficient and an appropriate estimator have been introduced and the corresponding bootstrap linear independence test has been carried out. The suitability of the obtained results has been analyzed by means of a simulation study and a real-life application.

Future research can be done introducing a selection procedure to obtain the appropriate number of explanatory variables and addressing the multicollinearity problem. Moreover, it could be interesting to develop a fuzzy regression model able to suitably taking into account, in a different way, fuzzy sets representing an intrinsically *imprecise* or intrinsically *precise* yet ill-known property (see, e.g., Dubois & Prade, 2009). Specifically, it would be advisable to suggest a fuzzy regression model based on tools (distance functions, dissimilarity measures, variances) defined in different ways according to the nature of the property. With particular reference to the concept of variance for FRV, one can refer to Couso & Dubois (2009) in which three different definitions of variance are discussed according to the nature of the

modelled quantity.

References

- [1] Bargiela, A., Pedrycz, W., Nakashima, T., 2007. Multiple regression with fuzzy data. *Fuzzy Sets and Systems* 158, 2169–2188.
- [2] Celminš, A., 1987. Multidimensional least-squares fitting of fuzzy models. *Mathematical Modelling* 9, 669–690.
- [3] Chang, P.T., Lee, E.S., 1996. A generalized fuzzy weighted least-squares regression. *Fuzzy Sets and Systems* 82, 289–298.
- [4] Coppi, R., D’Urso, P., Giordani, P., Santoro, A., 2006. Least squares estimation of a linear regression model with LR fuzzy response. *Computational Statistics and Data Analysis* 51, 267–286.
- [5] Couso, I., Dubois, D., 2009. On the variability of the concept of variance for fuzzy random variables. *IEEE Transactions on Fuzzy Systems* 17, 1070–1080.
- [6] D’Urso, P., 2003. Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. *Computational Statistics and Data Analysis* 42, 47–72.
- [7] Diamond, P., 1988. Fuzzy least squares. *Information Sciences* 46, 141–157.
- [8] Dubois, D., Prade, H., 2009. Gradualness, uncertainty and bipolarity: making sense of fuzzy sets. 30th Linz Seminar on Fuzzy Set Theory.
- [9] Ferraro, M.B., Coppi, R., Gonzalez-Rodriguez, G., Colubi, A., 2009. A linear regression model for imprecise response. Tech. Rep. n. 15, Dept. Statistics, Probability and Applied Statistics, Sapienza University of Rome. Submitted.
- [10] Ferraro, M.B., Colubi, A., Gonzalez-Rodriguez, G., Coppi, R., 2009. A determination coefficient for a linear regression model with imprecise response. Tech. Rep. n. 16, Dept. Statistics, Probability and Applied Statistics, Sapienza University of Rome. Submitted.

- [11] González-Rodríguez, G., Blanco, A., Colubi, A., Lubiano, M.A., 2009. Estimation of a simple linear regression model for fuzzy random variables. *Fuzzy Sets and Systems* 160, 357–370.
- [12] Guo, P., Tanaka, H., 2006. Dual models for possibilistic regression analysis. *Computational Statistics and Data Analysis* 51, 253–266.
- [13] Körner, R., Näther, W., 1998. Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates. *Information Sciences* 109, 95–118.
- [14] Krätschmer, V., 2006a. Strong consistency of least-squares estimation in linear regression models with vague concepts. *Journal of Multivariate Analysis* 97, 633–654.
- [15] Krätschmer, V., 2006b. Limit distributions of least squares estimators in linear regression models with vague concepts. *Journal of Multivariate Analysis* 97, 1044–1069.
- [16] Lawson, C.L., Hanson, R.J., 1995. *Solving Least Squares Problems*. Classics in Applied Mathematics 15. SIAM. Philadelphia, PA.
- [17] Liew, C.K., 1976. Inequality constrained least-squares estimation. *Journal of the American Statistical Association* 71, 746–751.
- [18] Lu, J., Wang, R., 2009. An enhanced fuzzy linear regression model with more flexible spreads. *Fuzzy Sets and Systems* 160, 2505–2523.
- [19] Näther, W., 2006. Regression with fuzzy random data. *Computational Statistics and Data Analysis* 51, 235–252.
- [20] Puri, M.L., Ralescu, D.A., 1986. Fuzzy random variables. *Journal of Mathematical Analysis and Applications* 114, 409–422.
- [21] Tanaka, H., Ishibuchi, H., Yoshikawa, S., 1995. Exponential possibility regression analysis. *Fuzzy Sets and Systems* 69, 305–318.
- [22] Tanaka, H., Uejima, S., Asai, K., 1982. Linear regression analysis with fuzzy model. *IEEE Transactions on Systems, Man and Cybernetics* 12, 903–907.

- [23] Tanaka, H., Watada, J., 1988. Possibilistic linear systems and their application to the linear regression model. *Fuzzy Sets and Systems* 27, 275–289.
- [24] Yang, M.S., Ko, C.H., 1996. On a class of fuzzy c -numbers clustering procedures for fuzzy data. *Fuzzy Sets and Systems* 84, 49–60.
- [25] Zadeh, L.A., 1965. Fuzzy sets. *Information and Control* 8, 338–353.