# A linear regression model for imprecise response

Maria Brigida Ferraro*†, Renato Coppi, Gil González-Rodríguez and Ana Colubi

SUMMARY

A linear regression model with imprecise response and $p$ real explanatory variables is analyzed. The imprecision of the response variable is functionally described by means of certain kinds of fuzzy sets, the $LR$ fuzzy sets. The $LR$ fuzzy random variables are introduced to model usual random experiments when the characteristic observed on each result can be described with fuzzy numbers of a particular class, determined by 3 random values: the center, the left spread and the right spread. In fact, these constitute a natural generalization of the interval data. To deal with the estimation problem the space of the $LR$ fuzzy numbers is proved to be isometric to a closed and convex cone of $\mathbb{R}^3$ with respect to a generalization of the most used metric for $LR$ fuzzy numbers. The expression of the estimators in terms of moments is established, their limit distribution and asymptotic properties are analyzed and applied to the determination of confidence regions and hypothesis testing procedures. The results are illustrated by means of some case-studies.

*KEY WORDS* Least-squares approach; Asymptotic distribution; $LR$ fuzzy data; Interval data; Regression Models

## 1 Introduction

Different elements of a statistical problem may be imprecisely observed or defined. This has led to the development of various theories able to cope with an uncertainty which is not necessarily due to randomness: e.g. the methods based on imprecise probabilities (see, for instance, Walley [31]), or on the use of subjective probabilities (see, for instance, Singpurwalla & Booker [30]) or diverse approaches for fuzzy statistical analysis (see, for instance, Coppi [4], Colubi [5] or Denoeux *et al.* [7]). In this paper we will consider a regression problem for a random experiment in which a fuzzy response and real-valued explanatory variables are observed.

Actually, in many practical applications in public health, medical science, ecology, social

*Correspondence to: M. B. Ferraro, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Sapienza Università di Roma, P.le Aldo Moro, 5 - 00185, Rome, Italy

†E-mail: mariabrigida.ferraro@uniroma1.it

or economic problems, many useful variables are vague, and the researchers find it easier to reflect the vagueness through fuzzy data than to discard the vagueness and obtain precise data. In addition it is often less expensive to obtain an imprecise observation than to look for precise measurements of the variable of interest (see, for instance, Heagerty & Lele [15]).

In order to handle a typical kind of imprecision the so-called $LR$ fuzzy sets are often used. They are determined by three values: the center, the left spread and the right spread. For example, in agriculture quantitative soil data are unavailable over vast areas and imprecise measures, that can be modelled through $LR$ fuzzy sets, are used (see Lagacherie *et al.* [21]). Also in medical science symptoms, diagnosis and phenomena of disease may often lead to $LR$ data (see, for instance, Di Lascio *et al.* [6]). $LR$-type fuzzy data may also arise in other contexts, like image processing or artificial intelligence (see, for instance, Sezgin & Sankur [29], Ranilla & Rodríguez-Muñiz [28]).

The $LR$ fuzzy sets are a generalization of the intervals. Epidemiological research often entails the analysis of failure times subject to grouping, and the analysis with interval-grouped data is numerically simple and statistically meaningful (see Pipper & Ritz [26], Gil *et al.* [12], Billard & Diday [2]).

Several regression studies involving fuzzy random variables to model imprecise data have been developed (see, for instance, Näther [25], Krätschmer [19], González-Rodríguez *et al.* [14], etc).

Coppi *et al.* [3] have proposed a linear regression model with $LR$ fuzzy response. The basic idea consists in modelling the centers of the response variable by means of a classical regression model, and simultaneously modelling the left and the right spread of the response through simple linear regressions on its estimated centers. The study in Coppi *et al.* [3] is mainly descriptive, and the authors impose a non-negativity condition to the numerical minimization problem to avoid negative estimated spreads. In this work we propose an alternative model to overcome the non-negativity condition, because the inferences for models with non-negativity restrictions are more complex and less efficient (see, for instance, Liew [22] and Gallant & Gerig [10]).

In Section 2 the way of modelling the imprecise response through $LR$ fuzzy random variables is formalized. In Section 3 the variance of an $LR$ fuzzy random variable is defined and some properties are proved. In Section 4 the new linear regression model is introduced, and the least squares estimators of the parameters are found and analyzed. Section 5 deals with asymptotic confidence regions and asymptotic hypothesis tests for the regression parameters. In section 6 a real-life example with $LR$ fuzzy data and another with interval data are illustrated. Finally, Section 7 contains some remarks and future directions.

# 2 Modelling the imprecise data

## 2.1 Fuzzy sets

In this work a fuzzy set $A$ of $\mathbb{R}$ will be simply defined as a mapping $A : \mathbb{R} \to [0,1]$ verifying some conditions. Let $\mathcal{K}_c(\mathbb{R})$ be the class of nonempty compact convex subsets of $\mathbb{R}$, we will consider the *class of fuzzy sets* $\mathcal{F}_c(\mathbb{R}) = \{A : \mathbb{R} \to [0,1] | A_\alpha \in \mathcal{K}_c(\mathbb{R})\}$, where $A_\alpha$ is the $\alpha$-level of fuzzy set $A$, that is, $A_\alpha = \{x \in \mathbb{R} | A(x) \geq \alpha\}$, for $\alpha \in (0,1]$, and $A_0 = cl(\{x \in \mathbb{R} | A(x) > 0\})$ (Zadeh [33]).
In practice there are some experiments whose results can be described by means of fuzzy sets of a particular class, determined by 3 values: the center, the left spread and the right spread. This type of fuzzy datum is called *LR fuzzy number* and it is defined such that (see Fig. 1)

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

where $A^m \in \mathbb{R}$ is the center, $A^l \in \mathbb{R}^+$ and $A^r \in \mathbb{R}^+$ are, respectively, the left and the right spread, $L$ and $R$ are functions verifying the properties of the class of fuzzy sets $\mathcal{F}_c(\mathbb{R})$, such that $L(0) = R(0) = 1$ and $L(x) = R(x) = 0$, $\forall x \in \mathbb{R} \setminus [0,1]$.

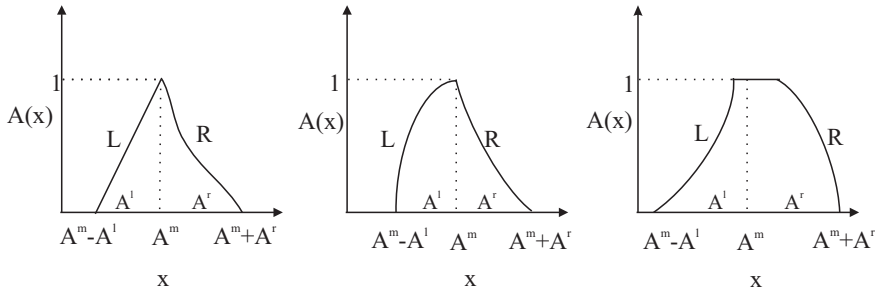


Figure 1: Examples of LR membership functions

**Remark 1.** An interval $I$ is a particular kind of $LR$ fuzzy set where the membership function is the characteristic function $1_I$, that is equal to 1, for all $x \in I$, and 0 otherwise ($L = R = I_{[0,1]}$, $A^m = (\inf I + \sup I)/2$ and $A^l = A^r = (\sup I - \inf I)/2$).

Let $\mathcal{F}_{LR}$ be the class of $LR$ fuzzy numbers. Since any $A \in \mathcal{F}_{LR}$ can be represented by means of a 3-tuple $(A^m, A^l, A^r)$, we define the mapping $s : \mathcal{F}_{LR} \to \mathbb{R}^3$ such that $s(A) = s_A = (A^m, A^l, A^r)$.

In what follows we use without distinction $A \in \mathcal{F}_{LR}$ or its $s$-representation $(A^m, A^l, A^r)$. The natural sum and the product by a scalar in $\mathcal{F}_{LR}$ extend the Minkowski sum and the product by a positive scalar for intervals, that is, for all $\alpha \in [0,1]$ we have:

$$(A+B)_\alpha = \{a+b \,|\, a \in A_\alpha, b \in B_\alpha\}, \qquad (\gamma A)_\alpha = \{\gamma a | a \in A_\alpha\},$$

These operations can be alternatively expressed considering the $s$-representation, that is, $A+B$ is the fuzzy set in $\mathcal{F}_{LR}$ such 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 $\gamma A$ is the fuzzy set in $\mathcal{F}_{LR}$ such 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 function $s$ is obviously *semi-linear*, because $s(A) + s(B) = s(A+B)$ and $\gamma s(A) = s(\gamma A)$, if $\gamma > 0$.
Yang and Ko [32] have defined a distance $D^2_{LR}$ between two $LR$ fuzzy numbers $A$, $B \in \mathcal{F}_{LR}$ as follows

$$\begin{aligned} D^2_{LR}(A,B) = (A^m - B^m)^2 &+ ((A^m - \lambda A^l) - (B^m - \lambda B^l))^2 \\ &+ ((A^m + \rho A^r) - (B^m + \rho B^r))^2, \end{aligned} \tag{1}$$

where $\lambda = \int_0^1 L^{-1}(\omega)d\omega$ and $\rho = \int_0^1 R^{-1}(\omega)d\omega$ represent the influence of the shape of the membership function on the distance. As a result $(\mathcal{F}_{LR}, D^2_{LR})$ is a metric space.

## 2.2 The isometry

In order to embed the space $\mathcal{F}_{LR}$ into $\mathbb{R}^3$ by preserving the metric, we will define a metric in $\mathbb{R}^3$ and we will show that this metric endows $\mathbb{R}^3$ with a Hilbertian structure.

**Proposition 1.** *Given $a = (a_1, a_2, a_3)$, $b = (b_1, b_2, b_3) \in \mathbb{R}^3$ and $\lambda$, $\rho \in \mathbb{R}^+$, $(\mathbb{R}^3, D_{\lambda\rho})$ is a metric space, where*

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

*takes inspiration from the Yang-Ko distance. Moreover*

$$\langle a, b\rangle_{\lambda\rho} = \langle a_1, b_1\rangle_{\mathbb{R}} + \langle (a_1 - \lambda a_2), (b_1 - \lambda b_2)\rangle_{\mathbb{R}} + \langle (a_1 - \lambda a_3), (b_1 - \lambda b_3)\rangle_{\mathbb{R}}$$

*is an inner product.*

The next proposition states that $\mathcal{F}_{LR}$ is isometric to a closed convex cone of the Hilbert space $(\mathbb{R}^3, \langle\cdot,\cdot\rangle_{\lambda\rho})$.

**Proposition 2.** *We consider the space* $\mathcal{F}_{LR}$, $\lambda = \int_0^1 L^{-1}(\omega)d\omega$ *and* $\rho = \int_0^1 R^{-1}(\omega)d\omega$. *Then* $\mathcal{F}_{LR}$ *is isometric to a closed convex cone of* $\mathbb{R}^3$ *endowed with the inner product* $\langle\cdot,\cdot\rangle_{\lambda\rho}$.

From now on, we will consider the operation $\langle A, B\rangle_{LR} = \langle s_A, s_B\rangle_{LR}$, which is not exactly an inner product due to the lack of linearity, but it has interesting properties.

### 2.3 Fuzzy random variables

Kwakernaak [20], Puri & Ralescu [27] and Klement *et al.* [16] have introduced the concept of Fuzzy Random Variable (FRV) as an extension of both, random variables and random sets.
Let $(\Omega, \mathcal{A}, P)$ be a probability space. According to Puri and Ralescu, the mapping $X : \Omega \to \mathcal{F}_{LR}$ is an FRV if for any $\alpha \in [0,1]$ the $\alpha-$cut $X_\alpha$ is a convex compact random set. This is equivalent to requiring that the $s$-representation of $X$, $(X^m, X^l, X^r) : \Omega \to \mathbb{R}\times\mathbb{R}^+\times\mathbb{R}^+$ be a random vector. It should be noted that in our approach $X$ is not an ill-measured real random variable but a random element assuming "purely" fuzzy values (see, also, González-Rodríguez *et al.* [13]),

**Example 1.** An example of FRVs is introduced in Colubi [5]. In a recent study about the reforestation in a given area of Asturias (Spain), carried out in the INDUROT institute (University of Oviedo), the quality of the trees has been analyzed. This characteristic has not been assigned on the basis of an underlying real-valued magnitude, but rather on the basis of subjective judgements/perceptions, through the observation of the leaf structure, the root system, the relationship height/diameter, and so on. The experts used a fuzzy-valued scale to represent their perceptions, besides linguistic labels, because the usual categorical scale (very low, low, medium, high, very high) was not able to capture the perceptions. The considered support goes from 0 (absence of quality) to 100 (perfect quality). It is possible to have different values for the same linguistic label. Some possible fuzzy values are represented in Fig. 2. This variable has been observed on 238 trees. Thus $\Omega = \{$*sets of trees in a given area of Asturias*$\}$ endowed with the Borel $\sigma$-field. Since the observations were arbitrarily chosen, $P$ is the uniform distribution over $\Omega$. For any $i \in \Omega$, several characteristics are to be observed. In particular, the quality, $Y_i$, has been considered as an $LR$ triangular fuzzy variable ($\lambda = \rho = 1/2$) (see Table 1).

The expected value of an FRV is defined by means of the generalized Aumann integral (Aumann [1]), that is, the expected value of the FRV $X$ is the unique fuzzy set $E(X)$ in

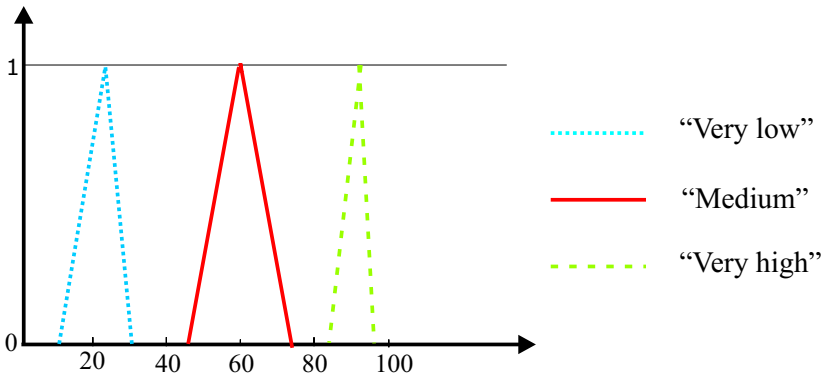


Figure 2: Values of the "Quality" of three different trees

$\mathcal{F}_{LR}$, s.t., for all $\alpha \in [0,1]$,

$$(EX)_\alpha = EX_\alpha,$$

if $E\|X\|^2_{LR} < \infty$ (see Puri & Ralescu [27]). Equivalently, $E(X)$ is the fuzzy set in $\mathcal{F}_{LR}$ whose $s$-representation is equal to $(EX^m, EX^l, EX^r)$.

## 3 The variance and the covariance

The notion of variance for FRVs has been previously established in terms of several metrics (see Körner [17] and Lubiano *et al.* [23]). By following the same ideas, we can also consider it in the sense of the $D_{LR}$ metric.

**Definition 1.** The variance of an $LR$ fuzzy random variable $X = (X^m, X^l, X^r)$ with $E\|X\|^2_{LR} < \infty$ is defined by $Var(X) = ED^2_{LR}(X, EX) = E\langle s_X - s_{EX}, s_X - s_{EX}\rangle_{LR}$.

It can be easily checked that

$$\begin{aligned} Var(X) &= E\left[3(X^m - EX^m)^2 + \lambda^2(X^l - EX^l)^2 + \rho^2(X^r - EX^r)^2\right] \\ &\quad + E\left[-2\lambda(X^m - EX^m)(X^l - EX^l) + 2\rho(X^m - EX^m)(X^r - EX^r)\right] \\ &= 3Var(X^m) + \lambda^2 Var(X^l) + \rho^2 Var(X^r) - 2\lambda Cov(X^m, X^l) + 2\rho Cov(X^m, X^r). \end{aligned}$$

This notion of variance satisfies the same suitable properties of the usual variance in $\mathbb{R}$, that is,

**Proposition 3.** *Let $X$ and $Y$ be LR fuzzy random variables, $A \in \mathcal{F}_{LR}$ and $\gamma \in \mathbb{R}^+$. Then*

1. $Var(X) = E\|X\|^2_{LR} - \|EX\|^2_{LR}$,
2. $Var(\gamma X) = \gamma^2 Var(X)$,
3. $Var(A + X) = Var(X)$,
4. $Var(X + Y) = Var(X) + Var(Y)$ *if* $X$ *and* $Y$ *are independent,*
5. *if* $A \in \mathcal{F}_{LR}$, *then* $\Delta_X(A) = E\left[D^2_{LR}(X, A)\right] = Var(X) + D^2_{LR}(A, EX)$.

Property 5 of Proposition 3 shows that $Var(X)$ verifies the *Fréchet principle* (see Fréchet [9]) because $E\left[D^2_{LR}(X, A)\right]$ is minimized, for $A = EX$, which makes coherent the application of least-squares techniques in regression problems.
Taking inspiration from the expression of the variance in terms of the inner product we can also define the covariance as follows.

**Definition 2.** The covariance between two LR fuzzy random variables $X = (X^m, X^l, X^r)$ and $Y = (Y^m, Y^l, Y^r)$ is defined by $Cov(X, Y) = E\langle s_X - s_{EX}, s_Y - s_{EY}\rangle_{LR}$.

In this case it is easy to prove that

$$\begin{aligned} Cov(X,Y) &= 3Cov(X^m, Y^m) + \lambda^2 Cov(X^l, Y^l) + \rho^2 Cov(X^r, Y^r) \\ &\quad -\lambda Cov(X^m, Y^l) - \lambda Cov(X^l, Y^m) + \rho Cov(X^m, Y^r) + \rho Cov(X^r, Y^m). \end{aligned}$$

Due to the lack of linearity of $\mathcal{F}_{LR}$, the covariance does not have the same meaning and all the properties of the covariance in $\mathbb{R}$.

## 4 Least squares estimators

Consider a random experiment in which an $LR$ fuzzy response variable $Y$ and $p$ real explanatory variables $X_1$, $X_2$,...,$X_p$ are observed on $n$ statistical units, $\{Y_i, \underline{X}_i\}_{i=1,...,n}$, where $\underline{X}_i = (X_{1i}, X_{2i}, ..., X_{pi})$, or in a compact form $(\underline{Y}, \mathbf{X})$. Since $Y$ is determined by $(Y^m, Y^l, Y^r)$, the proposed regression model concerns the real-valued random variables in this tuple. The center $Y^m$ can be related to the explanatory variables $X_1, X_2, ..., X_p$ through a classical regression model. However, the restriction of non-negativity satisfied by $Y^l$ and $Y^r$ entails some difficulties (see Coppi *et al.* [3]). One solution is to consider a model with the restriction of non-negativity but, when a variable has this kind of restrictions, the errors of the model may be dependent on the explanatory variable, and the classical methods are not efficient (see, for instance, Liew [22], Gallant & Gerig [10]). In addition, in presence of non-negativity restrictions most of works in literature are numerical procedures

while in this paper the idea is to formalize a realistic theoretical model and to obtain a complete analytical solution.
We propose modelling a transformation of the left spread and a transformation of the right spread of the response through simple linear regressions (on the explanatory variables $X_1$, $X_2$,...,$X_p$). This can be represented in the following way, letting $g : (0, +\infty) \longrightarrow \mathbb{R}$ and $h : (0, +\infty) \longrightarrow \mathbb{R}$ be invertible:

$$\begin{cases} Y^m = \underline{X}\,\underline{a}_m^{'} + b_m + \varepsilon_m \\ g(Y^l) = \underline{X}\,\underline{a}_l^{'} + b_l + \varepsilon_l \\ h(Y^r) = \underline{X}\,\underline{a}_r^{'} + b_r + \varepsilon_r \end{cases} \tag{2}$$

where $\varepsilon_m$, $\varepsilon_l$ and $\varepsilon_r$ are real-valued random variables with $E(\varepsilon_m|\underline{X}) = E(\varepsilon_l|\underline{X}) = E(\varepsilon_r|\underline{X}) = 0$, $\underline{a}_m = (a_{m1}, ..., a_{mp})$, $\underline{a}_l = (a_{l1}, ..., a_{lp})$ and $\underline{a}_r = (a_{r1}, ..., a_{rp})$ are the $(1 \times p)$-vectors of the parameters related to the vector $\underline{X}$. The covariance matrix of the vector of explanatory variables $\underline{X}$ will be denoted by $\Sigma_{\underline{X}}$ and $\Sigma$ will stand for the covariance matrix of $(\varepsilon_m, \varepsilon_l, \varepsilon_r)$, whose variances are strictly positive and finite.

**Remark 2.** Since the expected values of $\varepsilon_m$, $\varepsilon_l$ and $\varepsilon_r$ given $\underline{X}$ are equal to 0 it results that $\varepsilon_m$, $\varepsilon_l$ and $\varepsilon_r$ are uncorrelated with the explanatory variables.

**Remark 3.** From an econometric point of view the model (2) can be seen as a simultaneous equation system. It should be underlined that in this case the parameter identification problem does not affect the model, due to the way it has been defined (see, for instance, Mardia *et al.* [24]).

**Remark 4.** In practice, particularly in the socio-economical domain, it is possible to have restrictions on the center $Y^m$ or on the explanatory variables $X_1, X_2, ..., X_p$. In this case it is possible to transform these variables too. It results a non linear model.

**Example 2.** We consider a simplification of the data introduced in Colubi [5] (see Table 1). We use the new linear regression model to analyze the part of the *quality*, $Y$, of 238 trees explained by the *height*, $X$.
In presence of constrained variables, a common approach consists in transforming the constrained variable into an unconstrained one by means of the logarithmic transformation (that is $g$=$h$=ln). We will use this approach in this example to transform the spreads into real variables without the restriction of non-negativity.

Table 1: Quality ($Y^m$, $Y^l$, $Y^r$) and Height ($X$) of 238 trees in Asturias.

| $Y^m$(center) | $Y^l$(left spread) | $Y^r$(right spread) | $X(cm)$ |
|---|---|---|---|
| 45 | 12.5 | 15 | 170 |
| 25 | 15 | 12.5 | 245 |
| 17.5 | 7.5 | 12.5 | 190 |
| 20 | 11.25 | 15 | 130 |
| 55 | 15 | 12.5 | 230 |
| 23.75 | 11.25 | 18.75 | 90 |
| 56.25 | 18.75 | 13.75 | 195 |
| 13.75 | 8.75 | 8.75 | 75 |
| 26.25 | 13.75 | 8.75 | 184 |
| 62.5 | 10 | 7.5 | 215 |
| 75 | 12.5 | 10 | 245 |
| 67.5 | 12.5 | 12.5 | 220 |
| 32.5 | 22.5 | 10 | 195 |
| 40 | 15 | 10 | 160 |
| 52.5 | 12.5 | 17.5 | 213 |
| 55 | 15 | 17.5 | 215 |
| 77.5 | 12.5 | 12.5 | 370 |
| 85 | 5 | 5 | 230 |
| 50 | 20 | 20 | 234 |
| ... | ... | ... | ... |

In Proposition 4 we show that the population parameters can be expressed, as usual, in terms of some moments involving the considered random variables.

**Proposition 4.** *Let $Y$ be an LR fuzzy random variable and $\underline{X}$ the vector of $p$ real random variables satisfying the linear model (2), then we have that*

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

*where* $\Sigma_{\underline{X}} = E\left[(\underline{X}-E\underline{X})^{'}(\underline{X}-E\underline{X})\right]$

The estimators of the population parameters will be based on the Least Squares (LS) criterion. As it was above-mentioned, the use of this criterion is justified by the properties of the variance proved in Proposition 3, among which we find the *Fréchet principle*. In addition, it should be remarked that the lack of realistic parametric models for the distribution of FRVs prevents us from using other approaches, as maximum likelihood. In this case, using the generalized Yang-Ko metric $D^2_{\lambda\rho}$ written in vector terms, the LS problem consists in looking for $\underline{\hat{a}}_m$, $\underline{\hat{a}}_l$, $\underline{\hat{a}}_r$, $\hat{b}_m$, $\hat{b}_l$ and $\hat{b}_r$ in order to

$$\min \Delta^2_{\lambda\rho} \quad = \quad \min D^2_{\lambda\rho}((\underline{Y}^m, g(\underline{Y}^l), h(\underline{Y}^r)), ((\underline{Y}^m)^*, g^*(\underline{Y}^l), h^*(\underline{Y}^r))) \tag{3}$$

where $(\underline{Y}^m)^* = \mathbf{X}\underline{a}_m^{'} + \underline{1}b_m$, $g^*(\underline{Y}^l) = \mathbf{X}\underline{a}_l^{'} + \underline{1}b_l$ and $h^*(\underline{Y}^r) = \mathbf{X}\underline{a}_r^{'} + \underline{1}b_r$ are the $(n \times 1)$-vectors of the predicted values.
The function to minimize

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

becomes

$$\begin{aligned}\Delta^2_{\lambda\rho} \quad &= \quad 3\left(\underline{Y}^m - \mathbf{X}\underline{a}_m^{'} - \underline{1}b_m\right)^{'}\left(\underline{Y}^m - \mathbf{X}\underline{a}_m^{'} - \underline{1}b_m\right) \\ &+ \quad \lambda^2\left(g(\underline{Y}^l) - \mathbf{X}\underline{a}_l^{'} - \underline{1}b_l\right)^{'}\left(g(\underline{Y}^l) - \mathbf{X}\underline{a}_l^{'} - \underline{1}b_l\right) \\ &+ \quad \rho^2\left(h(\underline{Y}^r) - \mathbf{X}\underline{a}_r^{'} - \underline{1}b_r\right)^{'}\left(h(\underline{Y}^r) - \mathbf{X}\underline{a}_r^{'} - \underline{1}b_r\right) \\ &- \quad 2\lambda\left(\underline{Y}^m - \mathbf{X}\underline{a}_m^{'} - \underline{1}b_m\right)^{'}\left(g(\underline{Y}^l) - \mathbf{X}\underline{a}_l^{'} - \underline{1}b_l\right) \\ &+ \quad 2\rho\left(\underline{Y}^m - \mathbf{X}\underline{a}_m^{'} - \underline{1}b_m\right)^{'}\left(h(\underline{Y}^r) - \mathbf{X}\underline{a}_r^{'} - \underline{1}b_r\right).\end{aligned} \tag{4}$$

**Proposition 5.** *The solutions of the LS problem are*

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

*where, as usual,* $\overline{Y^m}$, $\overline{g(Y^l)}$, $\overline{h(Y^r)}$ *and* $\overline{\underline{X}}$ *are, respectively, the sample means of* $Y^m$,

$g(Y^l)$, $h(Y^r)$ *and* $\underline{X}$,

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

*are the centered values of the response and*

$$\widetilde{\boldsymbol{X}} = \boldsymbol{X} - \underline{1}\,\overline{\underline{X}}$$

*the centered matrix of the explanatory variables.*

**Proposition 6.** *The estimators* $\widehat{\underline{a}}_m$, $\widehat{\underline{a}}_l$, $\widehat{\underline{a}}_r$, $\widehat{b}_m$, $\widehat{b}_l$ *and* $\widehat{b}_r$ *are unbiased and strongly consistent.*

For inferential purposes it is useful to provide an approximation to the distribution of the estimators. The above-mentioned lack of realistic parametric models for the distribution of the FRVs makes worth to look for the asymptotic distribution of the estimators.

**Proposition 7.** *Under the assumptions of model (2), as* $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} \widehat{\underline{a}}_m' - \underline{a}_m' \\ \widehat{\underline{a}}_l' - \underline{a}_l' \\ \widehat{\underline{a}}_r' - \underline{a}_r' \end{pmatrix} \xrightarrow{D} N\left(\underline{0}', \left(\Sigma_{\underline{X}}\right)^{-1}\Sigma\right). \tag{5}$$

Since the probability distribution function that has generated the data set is unknown, in practice we propose to use a bootstrap procedure to evaluate the accuracy of the estimators, by means of the estimates of the standard errors (see Efron & Tibshirani [8]).

## 5 Confidence regions and hypothesis testing on the regression parameters

In addition to the estimation of the regression parameters, the confidence regions and the hypothesis testing procedures are introduced. Starting from the asymptotic distribution (5) it is easily obtained the following $100(1-\alpha)$ confidence region for the parameters $(\underline{a}_m', \underline{a}_l', \underline{a}_r')'$

$$\left[\begin{pmatrix} \widehat{\underline{a}}_m' \\ \widehat{\underline{a}}_l' \\ \widehat{\underline{a}}_r' \end{pmatrix} - \frac{c_{\alpha/2}}{\sqrt{n}}, \begin{pmatrix} \widehat{\underline{a}}_m' \\ \widehat{\underline{a}}_l' \\ \widehat{\underline{a}}_r' \end{pmatrix} + \frac{c_{\alpha/2}}{\sqrt{n}}\right]$$

where $c_{\alpha/2}$ is a $\alpha/2$-quantile of a $N\left(\underline{0}', \left(\Sigma_{\underline{X}}\right)^{-1}\Sigma\right)$.
In order to test the null hypothesis $H_0 : (\underline{a}_m', \underline{a}_l', \underline{a}_r')' = (\underline{k}_m', \underline{k}_l', \underline{k}_r')'$ against the alternative $H_1 : (\underline{a}_m', \underline{a}_l', \underline{a}_r')' \neq (\underline{k}_m', \underline{k}_l', \underline{k}_r')'$, where $\underline{k}_m$, $\underline{k}_l$, and $\underline{k}_r$ are vectors of constant values in $\mathbb{R}$, the test statistic $T_n = V_n' V_n$, where

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

can be used. It is possible to define a rejection region for the null hypothesis, that is

**Proposition 8.** *In testing the above-defined null hypothesis at the nominal significance level $\alpha$, $H_0$ should be rejected if $T_n > c_\alpha$, where $c_\alpha$ is a $\alpha$-quantile of the asymptotic distribution of $T_n$, that is $f_1(V)$ ($V \sim N\left(\underline{0}', \left(\Sigma_{\underline{X}}\right)^{-1}\Sigma\right)$ and $f_1(A) = A'A$).*

The unknown population variance $\left(\Sigma_{\underline{X}}\right)^{-1}\Sigma$ can be approximated by means of the sample one and the Slutzky's theorem guarantees the asymptotic convergence of the standardized statistic to a normal distribution.

## 6 Empirical results

To illustrate the application of the regression model introduced in this work we consider the following examples.

**Example 3.** We consider the data of Example 1. For analyzing the part of the quality explained by the height of the trees we use the new regression model and we obtain the following estimated models

$$\begin{cases} \widehat{Y^m} &= 0.1558X + 18.7497 \\ \widehat{Y^l} &= \exp(-0.00017X + 2.5780) \\ \widehat{Y^r} &= \exp(-0.00067X + 2.6489) \end{cases} \tag{6}$$

The value of the estimated parameter $\hat{a}_m$ equal to 0.1558 represents a positive linear relationship between the response and the explanatory variable. In particular, the quality is expected to increase of about 0.16 for any additional cm of the height.
The estimated spreads of the response variable, $\widehat{Y^l}$ and $\widehat{Y^r}$, represent the imprecision of the quality estimated by the new model. In Fig. 3 the extreme values of the 0-level and the single-value of the 1-level of the quality by the height are indicated, respectively, by means of the vertical segments and the dots, while the estimated centers and the estimated spreads are represented by the solid line and the dash line.

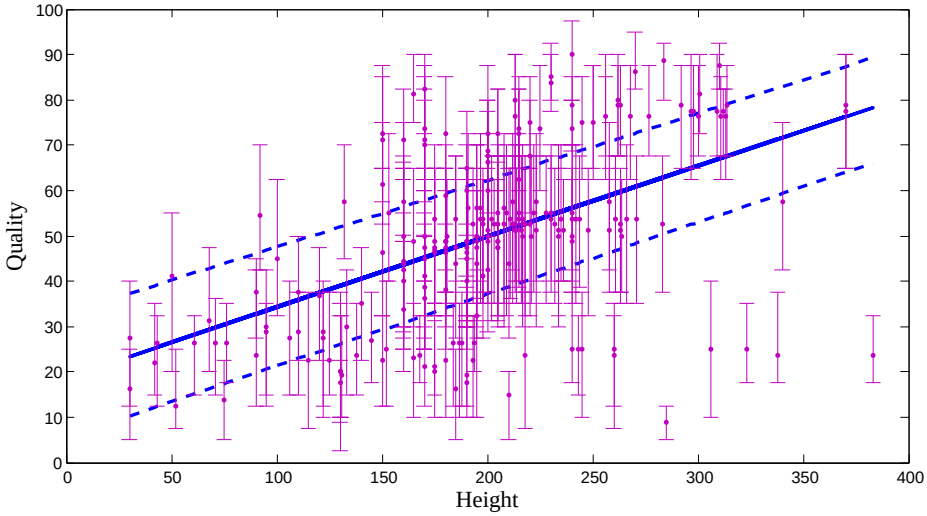


Figure 3: The observed extreme values of the 0-level and the single-value of the Quality by the Height of the trees, and the estimated linear regression models

To evaluate the accuracy of these estimates we draw 800 bootstrap samples of size $n = 238$ with replacement from our data set. For each bootstrap replication we calculate the estimate of the parameters of the linear regression model. By means of the 800 replications of the estimation procedure we compute the estimate of the standard errors $\widehat{se}$ of the parameters and we check that

$$\widehat{es}(\hat{a}_m) = 0.0210, \qquad \widehat{es}(\hat{a}_l) = 0.0004, \qquad \widehat{es}(\hat{a}_r) = 0.0004,$$
$$\widehat{es}(\hat{b}_m) = 3.9745, \qquad \widehat{es}(\hat{b}_l) = 0.0821, \qquad \widehat{es}(\hat{b}_r) = 0.0839.$$

Hence two kinds of uncertainty have been taken into account: the imprecision of the estimated quality and the stochastic uncertainty of the regression model represented by the above values.

To construct a confidence band for the vector of parameters $(a_m, a_l, a_r)$, the covariance matrix of the vector $(\varepsilon_m, \varepsilon_l, \varepsilon_r)$ has been replaced by the covariance matrix of the residuals $\widehat{\varepsilon_{mi}} = \widehat{Y_i^m} - Y_i^m$, $\widehat{\varepsilon_{li}} = \widehat{g(Y_i^l)} - g(Y_i^l)$, $\widehat{\varepsilon_{ri}} = \widehat{h(Y_i^r)} - h(Y_i^r)$, and the variance of the explanatory variable, $\sigma_X^2$, has been estimated by means of the sample variance $\widehat{\sigma_X}^2 = 3715.9$. A confidence band of approximate level $\alpha = 0.05$ has been found, that is,

$$\left[ \begin{pmatrix} -28.9355 \\ -0.0133 \\ -0.0122 \end{pmatrix}, \begin{pmatrix} 29.2470 \\ 0.0130 \\ 0.0109 \end{pmatrix} \right].$$

When testing if the vector of regression parameters $(a_m, a_l, a_r)'$ is equal to $(0, 0, 0)'$, a $p$-value equal to 0 is obtained. Hence this hypothesis (related to the linear independence)

should be rejected.

**Example 4.** In this example we are interested in analyzing the dependence relationship of the Retail Trade Sales (in millions of dollars) of the U.S. in 2002 by kind of business on the number of employees (see http://www.census.gov/econ/www/). The Retail Trade Sales are intervals in the period: January 2002 through December 2002 (see Table 2). For each interval we consider the center and the spreads and we apply the new regression model in order to evaluate the dependence relationship. As in Example 3 we have transformed the spreads by means of the logarithmic transformation.

Table 2: The Retail Trade Sales and the Number of Employees of 22 kinds of Business in the U.S. in 2002.

| Kind of Business | Retail Trade Sales | Number of Employees |
|---|---|---|
| Automotive parts, acc., and tire stores | 4638-5795 | 453468 |
| Furniture stores | 4054-4685 | 249807 |
| Home furnishings stores | 2983-5032 | 285222 |
| Household appliance stores | 1035-1387 | 69168 |
| Computer and software stores | 1301-1860 | 73935 |
| Building mat. and supplies dealers | 14508-20727 | 988707 |
| Hardware stores | 1097-1691 | 142881 |
| Beer, wine, and liquor stores | 2121-3507 | 133035 |
| Pharmacies and drug stores | 11964-14741 | 783392 |
| Gasoline stations | 16763-23122 | 926792 |
| Men's clothing stores | 532-1120 | 62223 |
| Family clothing stores | 3596-9391 | 522164 |
| Shoe stores | 1464-2485 | 205067 |
| Jewelry stores | 1304-5810 | 148752 |
| Sporting goods stores | 1748-3404 | 188091 |
| Book stores | 968-1973 | 133484 |
| Discount dept. stores | 9226-17001 | 762309 |
| Department stores | 5310-14057 | 668459 |
| Warehouse clubs and superstores | 13162-22089 | 830845 |
| All other gen. merchandize stores | 2376-4435 | 263116 |
| Miscellaneous store retailers | 7862-10975 | 792361 |
| Fuel dealers | 1306-3145 | 98574 |

By means of the least squares estimation we obtain the following predicted values

$$\begin{aligned}\widehat{Y^m} &= 0.0181X - 672.731\\ \widehat{Y^l} &= \exp(0.000002482X + 5.9244)\\ \widehat{Y^r} &= \exp(0.000002482X + 5.9244)\end{aligned}$$

The value 0.0181 indicates the strength of the relationship between the response and the explanatory variable, in particular, the retail trade sales are expected to increase of about 18.100 dollars for any additional employee.
Also in this case we evaluate the accuracy of the estimators by means of a bootstrap procedure with 800 replications. It is easy to check that

$$\begin{aligned}&\widehat{es}(\hat{a}_m) = 0.0015, \qquad \widehat{es}(\hat{a}_l) = 0.0000, \qquad \widehat{es}(\hat{a}_r) = 0.0000,\\ &\widehat{es}(\hat{b}_m) = 412.0407, \qquad \widehat{es}(\hat{b}_l) = 0.2151, \qquad \widehat{es}(\hat{b}_r) = 0.2151.\end{aligned}$$

The intercept term $\hat{b}_m$ is affected by a high degree of uncertainty, while the uncertainty of $\hat{a}_l$ and $\hat{a}_r$, which represent the relationship between the explanatory variable and the logarithmic transformation of the spreads of the response, is practically equal to 0. As Fig.

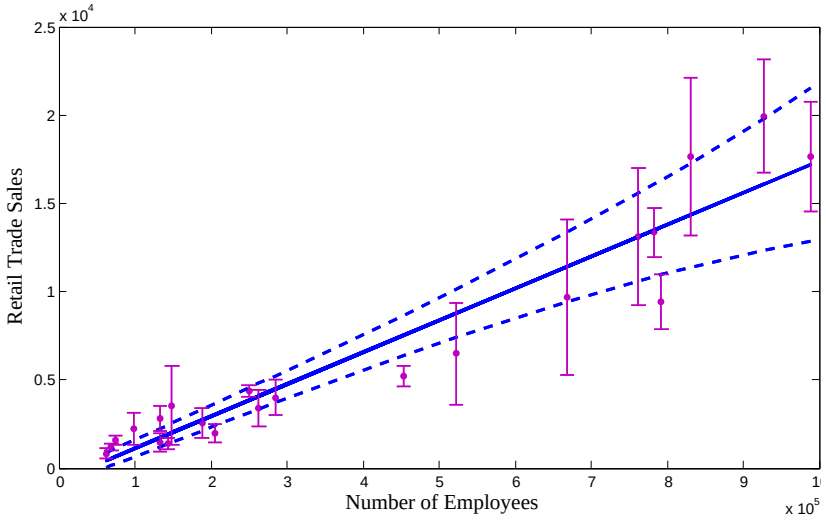


Figure 4: The observed interval Retail Trade Sales by Number of Employees and the estimated linear regression models

4 shows, the predicted values of the spreads grow as the number of employees increases. Also in this case the null hypothesis that all the regression parameters are equal to 0 should be rejected.

## 7 Concluding remarks

When modelling statistical relationships between imprecise and real elements by means of classical techniques, one of the main difficulties is related to the condition of non-negativity of the spreads. In this paper by means of the introduction of the functions $g$ and $h$ which transform the spreads into real numbers and through an appropriate metric, we have obtained a simple solution, expressed as a function of the sample moments, which furthermore is unbiased, consistent and useful in practice.

Based on an asymptotic distribution of the parameters, confidence regions have been constructed and hypothesis testing procedures have been analyzed. Since the asymptotic procedures work suitably for samples with very large size, it could be interesting to develop bootstrap procedures.

This new linear regression model can be used for all kinds of $LR$ functional data and in particular for interval-grouped data.

The linear regression model proposed in this paper can be generalized to other useful types of random sets, e.g. trapezoidal fuzzy sets, or considering nonlinear regression.

A further field of research consists in the study of appropriate functions $g$ and $h$ that can be used for a wide class of practical problems, by considering the model, for instance, in a semiparametric setting.

### Appendix

**Proof of Proposition 1.** It is clear that, $D_{\lambda\rho}(a,b) = D_{\lambda\rho}(b,a) \geq 0$ and it is null if and only if $a = b$. Concerning the triangle inequality we have that

$$\begin{aligned} D^2_{\lambda\rho}(a,b) &= (a_1-b_1)^2 + ((a_1-\lambda a_2)-(b_1-\lambda b_2))^2 + ((a_1+\rho a_3)-(b_1+\rho b_3))^2 \\ &= (a_1-c_1+c_1-b_1)^2 + ((a_1-\lambda a_2)-(c_1-\lambda c_2)+(c_1-\lambda c_2)-(b_1-\lambda b_2))^2 \\ &\quad +((a_1+\rho a_3)-(c_1+\rho c_3)+(c_1+\rho c_3)-(b_1+\rho b_3))^2 \\ &= D^2_{\lambda\rho}(a,c) + D^2_{\lambda\rho}(c,b) + 2(a_1-c_1)(c_1-b_1) \\ &\quad +2[(a_1-\lambda a_2)-(c_1-\lambda c_2)][(c_1-\lambda c_2)-(b_1-\lambda b_2)] \\ &\quad +2[(a_1+\rho a_3)-(c_1+\rho c_3)][(c_1+\rho c_3)-(b_1+\rho b_3)]. \end{aligned}$$

By Cauchy-Schwarz inequality, we obtain

$$D^2_{\lambda\rho}(a,b) \leq D^2_{\lambda\rho}(a,c) + D^2_{\lambda\rho}(c,b) + 2D^2_{\lambda\rho}(a,c)D^2_{\lambda\rho}(c,b) = (D_{\lambda\rho}(a,c) + D_{\lambda\rho}(c,b))^2.$$

Thus $D_{\lambda\rho}(a,b) \leq D_{\lambda\rho}(a,c) + D_{\lambda\rho}(c,b)$.

As a result we obtain that $D_{\lambda\rho}(a,b)$ is a metric in $\mathbb{R}^3$.

Since the terms defining $\langle\cdot,\cdot\rangle_{\lambda\rho}$ are based on $\langle\cdot,\cdot\rangle_{\mathbb{R}}$, it is easy to check that $\langle\cdot,\cdot\rangle_{\lambda\rho}$ verifies all the properties of an inner product.

**Proof of Proposition 2.** As $\mathcal{S} = \{s(A) : A \in \mathcal{F}_{LR}\}$ is $\mathbb{R} \times [0, \infty) \times [0, \infty)$, $\mathcal{S}$ is clearly a closed convex cone.

**Proof of Proposition 3.** Since the variance and the covariance in $\mathcal{F}_{LR}$ can be expressed in terms of the usual variance and covariance of real-valued variables, we can derive the thesis of this proposition through the basic properties of the variance and the covariance for real-valued random variables. In particular, we have that

$$\begin{aligned} Var(X) &= 3Var(X^m)+\lambda^2 Var(X^l) + \rho^2 Var(X^r) - 2\lambda Cov(X^m, X^l) + 2\rho Cov(X^m, X^r) \\ &= 3(E(X^m)^2 - (EX^m)^2) + \lambda^2(E(X^l)^2 - (EX^l)^2) + \rho^2(E(X^r)^2 - (EX^r)^2) \\ &\quad -2\lambda(E(X^m X^l) - EX^m EX^l) + 2\rho(E(X^m X^r) - EX^m EX^r) \\ &= E\|X\|^2_{LR} - \|EX\|^2_{LR}. \end{aligned}$$

Furthermore, the variance and the covariance in $\mathbb{R}$ satisfy simple properties in connection with the product by scalars and the sum that lead to

$$\begin{aligned} Var(\gamma X) &= 3Var(\gamma X^m) + \lambda^2 Var(\gamma X^l) + \rho^2 Var(\gamma X^r) \\ &\quad -2\lambda Cov(\gamma X^m, \gamma X^l) + 2\rho Cov(\gamma X^m, \gamma X^r) \\ &= 3\gamma^2 Var(X^m) + \lambda^2\gamma^2 Var(X^l) + \rho^2\gamma^2 Var(X^r) \\ &\quad -2\lambda\gamma^2 Cov(X^m, X^l) + 2\rho\gamma^2 Cov(X^m, X^r) = \gamma^2 Var(X) \end{aligned}$$

and

$$\begin{aligned} Var(A + X) &= 3Var(X^m + A^m) + \lambda^2 Var(X^l + A^l) + \rho^2 Var(X^r + A^r) \\ &\quad -2\lambda Cov(X^m + A^m, X^l + A^l) + 2\rho Cov(X^m + A^m, X^r + A^r) \\ &= 3Var(X^m) + \lambda^2 Var(X^l) + \rho^2 Var(X^r) \\ &\quad -2\lambda Cov(X^m, X^l) + 2\rho Cov(X^m, X^r) = Var(X) \end{aligned}$$

By means of similar reasoning, it is easy to check that

$$Var(X + Y) = Var(X) + Var(Y)$$

when $X$ and $Y \in \mathcal{F}_{LR}$ are independent. The quantity $\Delta_X(A)$ can be expressed as sum of terms that depend on real-valued random variables in the following way

$$\begin{aligned} \Delta_X(A) &= 3E\left[(X^m - A^m)^2\right] + \lambda^2 E\left[(X^l - A^l)^2\right] + \rho^2 E\left[(X^r - A^r)^2\right] \\ &\quad -2\lambda E\left[(X^m - A^m)(X^l - A^l)\right] + 2\rho E\left[(X^m - A^m)(X^r - A^r)\right]. \end{aligned}$$

Since the second moment of a real-valued random variable attains the minimum value when taken around the mean, we can easily obtain that

$$\begin{aligned} E\left[(X^m - A^m)^2\right] &= Var(X^m) + E\left[(EX^m - A^m)^2\right], \\ E\left[(X^l - A^l)^2\right] &= Var(X^l) + E\left[(EX^l - A^l)^2\right], \\ E\left[(X^r - A^r)^2\right] &= Var(X^r) + E\left[(EX^r - A^r)^2\right]. \end{aligned}$$

In an analogous way we can check that

$$\begin{aligned} E\left[(X^m - A^m)(X^l - A^l)\right] &= Cov(X^m, X^l) + E\left[(EX^m - A^m)(EX^l - A^l)\right], \\ E\left[(X^m - A^m)(X^r - A^r)\right] &= Cov(X^m, X^r) + E\left[(EX^m - A^m)(EX^r - A^r)\right]. \end{aligned}$$

By the composition of all these terms and taking into account that $VarX = 3Var(X^m) + \lambda^2 Var(X^l) + \rho^2 Var(X^r) - 2\lambda Cov(X^m, X^l) + 2\rho Cov(X^m, X^r)$, we obtain the thesis

$$E\left[D^2_{LR}(X, A)\right] = Var(X) + D^2_{LR}(A, EX).$$

**Proof of Proposition 4.** Under the assumptions in this theorem, it can be simply checked that

$$\begin{aligned} E\left[(\underline{X} - E\underline{X})'(Y^m - EY^m)\right] &= E\left[(\underline{X} - E\underline{X})'(\underline{X}\,\underline{a}'_m + b_m + \varepsilon_m - E\underline{X}\,\underline{a}'_m + b_m + E\varepsilon_m)\right] \\ &= E\left[(\underline{X} - E\underline{X})'(\underline{X} - E\underline{X})\underline{a}'_m + (\underline{X} - E\underline{X})'(\varepsilon_m - E\varepsilon_m)\right] \\ &= E\left[(\underline{X} - E\underline{X})'(\underline{X} - E\underline{X})\underline{a}'_m\right] + E\left[(\underline{X} - E\underline{X})'(\varepsilon_m - E\varepsilon_m)\right]. \end{aligned}$$

Since $\varepsilon_m$ is uncorrelated with the vector of explanatory variables $\underline{X}$, it results that

$$\underline{a}'_m = \left\{\Sigma_{\underline{X}}\right\}^{-1} E\left[(\underline{X} - E\underline{X})'(Y^m - EY^m)\right]$$

and

$$b_m = E(Y^m|\underline{X}) - E\underline{X}\left\{\Sigma_{\underline{X}}\right\}^{-1} E\left[(\underline{X} - E\underline{X})'(Y^m - EY^m)\right].$$

Analogously, following the same reasoning we obtain the remaining expressions.

**Proof of Proposition 5.** In order to solve the minimization problem and to check 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 some of the regression parameters are related to others. Starting from the estimation of $b_l$ and $b_r$, we equate to zero the partial derivatives, respectively, w.r.t $b_l$ and $b_r$. It is easy to find that the minimum is attained at

$$\begin{aligned} b_l &= \overline{g(Y^l)} - \overline{\underline{X}}\,\underline{a}'_l - \frac{1}{\lambda}\overline{Y^m} + \frac{1}{\lambda}\overline{\underline{X}}\,\underline{a}'_m + \frac{1}{\lambda}b_m \\ b_r &= \overline{h(Y^r)} - \overline{\underline{X}}\,\underline{a}'_r + \frac{1}{\rho}\overline{Y^m} - \frac{1}{\rho}\overline{\underline{X}}\,\underline{a}'_m - \frac{1}{\rho}b_m \end{aligned}$$

Since $b_l$ and $b_r$ depend on the term $b_m$, we have to substitute $b_l$ and $b_r$, as obtained above, in the objective function before equating to zero the partial derivative of the objective function w.r.t. $b_m$. As a result it will be obtained

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

Since the parameters $b_m$, $b_l$ and $b_r$ are expressed, respectively, 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 $\widetilde{\underline{Y}^m}$, $\widetilde{g(\underline{Y}^l)}$, $\widetilde{h(\underline{Y}^r)}$ and the centered matrix $\widetilde{\mathbf{X}}$ to make it simpler to analyze the objective function that can be expressed as follows

$$\begin{aligned}
\Delta^2_{\lambda\rho} &= 3(\widetilde{\underline{Y}^m} - \widetilde{\mathbf{X}}\underline{a}_m^{'})^{'}(\widetilde{\underline{Y}^m} - \widetilde{\mathbf{X}}\underline{a}_m^{'}) \\
&+ \lambda^2 \left(\widetilde{g(\underline{Y}^l)} - \widetilde{\mathbf{X}}\underline{a}_l^{'}\right)^{'} \left(\widetilde{g(\underline{Y}^l)} - \widetilde{\mathbf{X}}\underline{a}_l^{'}\right) \\
&+ \rho^2 \left(\widetilde{h(\underline{Y}^r)} - \widetilde{\mathbf{X}}\underline{a}_r^{'}\right)^{'} \left(\widetilde{h(\underline{Y}^r)} - \widetilde{\mathbf{X}}\underline{a}_r^{'}\right) \\
&- 2\lambda(\widetilde{\underline{Y}^m} - \widetilde{\mathbf{X}}\underline{a}_m^{'})^{'} \left(\widetilde{g(\underline{Y}^l)} - \widetilde{\mathbf{X}}\underline{a}_l^{'}\right) \\
&+ 2\rho(\widetilde{\underline{Y}^m} - \widetilde{\mathbf{X}}\underline{a}_m^{'})^{'} \left(\widetilde{h(\underline{Y}^r)} - \widetilde{\mathbf{X}}\underline{a}_r^{'}\right)
\end{aligned}$$

Following the usual reasoning it is easy to check that

$$\begin{aligned}
\underline{a}_l^{'} &= (\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{g(\underline{Y}^l)} - \frac{1}{\lambda}(\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\underline{Y}^m} + \frac{\underline{a}_m^{'}}{\lambda} \\
\underline{a}_r^{'} &= (\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{h(\underline{Y}^r)} + \frac{1}{\rho}(\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\underline{Y}^m} - \frac{\underline{a}_m^{'}}{\rho}
\end{aligned}$$

The last part is the estimation of $\underline{a}_m$. Since this vector appears in all the expressions of the other parameters we have to take this into account, and then by following the usual steps we can easily find that the minimum is attained at

$$\widehat{\underline{a}}_m^{'} = (\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\underline{Y}^m}$$

By making all the appropriate substitutions we obtain the other solutions, namely

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

**Proof of Proposition 6.** To prove the unbiasedness of the estimators we have to analyze their expected values. Starting from $\hat{a}_m^{'}$ we have

$$E\left(\widehat{\underline{a}}_m^{'}|\underline{X}\right) = E\left[(\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\underline{Y}^m}|\underline{X}\right]$$

Since $\widetilde{\underline{Y}^m} = \widetilde{\mathbf{X}}\underline{a}_m^{'} + \widetilde{\underline{\varepsilon}}_m$, where $\widetilde{\underline{\varepsilon}}_m$ is the $(n \times 1)$-vector of the centered errors, we obtain

$$\begin{aligned}
E\left(\widehat{\underline{a}}_m^{'}|\underline{X}\right) &= E\left[(\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}(\widetilde{\mathbf{X}}\underline{a}_m^{'} + \widetilde{\underline{\varepsilon}}_m)|\underline{X}\right] \\
&= E\left[(\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}}\underline{a}_m^{'}|\underline{X}\right] + E\left((\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\underline{\varepsilon}}_m|\underline{X}\right)
\end{aligned}$$

and, taking into account that the errors are uncorrelated with the explanatory variables, the thesis is proved, that is,

$$E\left(\underline{\widehat{a}}_m^{'}\right) = \underline{a}_m^{'}$$

Analogously, it is possible to check that $E\left(\underline{\widehat{a}}_l^{'}|\underline{X}\right) = \underline{a}_l^{'}$ and $E\left(\underline{\widehat{a}}_r^{'}|\underline{X}\right) = \underline{a}_r^{'}$.
Furthermore

$$E\left(\widehat{b}_m|\underline{X}\right) = E\left[\overline{Y^m}|\underline{X}\right] - E\left[\overline{\underline{X}}\,\underline{\hat{a}}_m^{'}|\underline{X}\right]$$

and since the sample means are unbiased estimators of the expectations, it is checked that $E\left(\widehat{b}_m|\underline{X}\right) = b_m$, and, by means of similar reasoning, the unbiasedness of $\widehat{b}_l$ and $\widehat{b}_r$.
The consistency is easily deduced from the expressions of the estimators and from the properties of population moments.

**Proof of Proposition 7.** Starting from the expression of $\underline{\widehat{a}}_m^{'}$, $\underline{\widehat{a}}_l^{'}$ and $\underline{\widehat{a}}_r^{'}$ in terms of sample moments

$$\begin{pmatrix} \underline{\widehat{a}}_m^{'} \\ \underline{\widehat{a}}_l^{'} \\ \underline{\widehat{a}}_r^{'} \end{pmatrix} = \begin{pmatrix} (\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\underline{Y}^m} \\ (\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{g(\underline{Y}^l)} \\ (\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{h(\underline{Y}^r)} \end{pmatrix},$$

and taking into account that $\widetilde{\underline{Y}^m} = \widetilde{\mathbf{X}}\underline{a}_m^{'} + \widetilde{\underline{\varepsilon}}_m$, $\widetilde{g(\underline{Y}^l)} = \widetilde{\mathbf{X}}\underline{a}_l^{'} + \widetilde{\underline{\varepsilon}}_l$ and $\widetilde{h(\underline{Y}^r)} = \widetilde{\mathbf{X}}\underline{a}_r^{'} + \widetilde{\underline{\varepsilon}}_r$, it is easy to check that

$$\begin{pmatrix} \underline{\widehat{a}}_m^{'} \\ \underline{\widehat{a}}_l^{'} \\ \underline{\widehat{a}}_r^{'} \end{pmatrix} = \begin{pmatrix} \underline{a}_m^{'} + (\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\underline{\varepsilon}}_m \\ \underline{a}_l^{'} + (\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\underline{\varepsilon}}_l \\ \underline{a}_r^{'} + (\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\underline{\varepsilon}}_r \end{pmatrix}.$$

In this way, we have that

$$\sqrt{n}\begin{pmatrix} \underline{\widehat{a}}_m^{'} - \underline{a}_m^{'} \\ \underline{\widehat{a}}_l^{'} - \underline{a}_l^{'} \\ \underline{\widehat{a}}_r^{'} - \underline{a}_r^{'} \end{pmatrix} = \left((\underline{1}^{'}\underline{1})^{-1}\widetilde{\mathbf{X}}^{'}\widetilde{\mathbf{X}}\right)^{-1}(\underline{1}^{'}\underline{1})^{-1/2}\begin{pmatrix} \widetilde{\mathbf{X}}^{'}\widetilde{\underline{\varepsilon}}_m \\ \widetilde{\mathbf{X}}^{'}\widetilde{\underline{\varepsilon}}_l \\ \widetilde{\mathbf{X}}^{'}\widetilde{\underline{\varepsilon}}_r \end{pmatrix},$$

and then,

$$
\begin{aligned}
(\underline{1}'\underline{1})^{-1/2}\begin{pmatrix} \widetilde{\mathbf{X}}'\widetilde{\underline{\varepsilon}}_m \\ \widetilde{\mathbf{X}}'\widetilde{\underline{\varepsilon}}_l \\ \widetilde{\mathbf{X}}'\widetilde{\underline{\varepsilon}}_r \end{pmatrix} &= (\underline{1}'\underline{1})^{-1/2}\begin{pmatrix} (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{\varepsilon}_m \\ (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{\varepsilon}_l \\ (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{\varepsilon}_m \end{pmatrix} \\
&+ (\underline{1}'\underline{1})^{-1/2}\begin{pmatrix} ((\underline{1}E\underline{X})'-(\underline{1}\,\overline{X})')\underline{\varepsilon}_m \\ ((\underline{1}E\underline{X})'-(\underline{1}\,\overline{X})')\underline{\varepsilon}_l \\ ((\underline{1}E\underline{X})'-(\underline{1}\,\overline{X})')\underline{\varepsilon}_m \end{pmatrix} \\
&- (\underline{1}'\underline{1})^{-1/2}\begin{pmatrix} (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{1}\overline{\varepsilon_m} \\ (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{1}\overline{\varepsilon_l} \\ (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{1}\overline{\varepsilon_r} \end{pmatrix} \\
&- (\underline{1}'\underline{1})^{-1/2}\begin{pmatrix} ((\underline{1}E\underline{X})'-(\underline{1}\,\overline{X})')\underline{1}\overline{\varepsilon_m} \\ ((\underline{1}E\underline{X})'-(\underline{1}\,\overline{X})')\underline{1}\overline{\varepsilon_l} \\ ((\underline{1}E\underline{X})'-(\underline{1}\,\overline{X})')\underline{1}\overline{\varepsilon_r} \end{pmatrix}
\end{aligned}
$$

Furthermore, as $n \to \infty$, the last three terms of the sum tend almost sure to $\underline{0}'$ ($(3p\times 1)$-null vector) and

$$
\left\{ \begin{pmatrix} (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{\varepsilon}_m \\ (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{\varepsilon}_l \\ (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{\varepsilon}_m \end{pmatrix} \right\}_{i=1,\ldots,n}
$$

is a sequence of random vectors i.i.d., centered at $\underline{0}'$, whose covariance matrix is $\Sigma_{\underline{X}}\Sigma$, so applying the Central Limit Theorem it results that

$$
(\underline{1}'\underline{1})^{-1/2}\begin{pmatrix} (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{\varepsilon}_m \\ (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{\varepsilon}_l \\ (\mathbf{X}'-(\underline{1}E\underline{X})')\underline{\varepsilon}_m \end{pmatrix} \xrightarrow{D} N(\underline{0}', \Sigma_{\underline{X}}\Sigma).
$$

Hence

$$
\sqrt{n}\begin{pmatrix} \widehat{\underline{a}}_m' - \underline{a}_m' \\ \widehat{\underline{a}}_l' - \underline{a}_l' \\ \widehat{\underline{a}}_r' - \underline{a}_r' \end{pmatrix} \xrightarrow{D} N\left(\underline{0}', \left(\Sigma_{\underline{X}}\right)^{-1}\Sigma\right).
$$

**Acknowledgements** Part of this research was carried out while Maria Brigida Ferraro worked at Departamento de Estadística e I. O. y D. M., University of Oviedo, Spain. We would like to acknowledge our colleagues in the SMIRE group (http://bellman.ciencias.uniovi.es/smire/) for making possible this stay and for their suggestions to improve this paper. The research in this paper has been partially supported by the Spanish Ministry of Education and Science Grants MTM2005-00045 and MTM2006-07501 and the Italian Ministry of Education, University and Research Grant PRIN 2005. Their financial support is gratefully acknowledged.

## References

[1] R.J. Aumann, Integrals of set-valued functions, J. Math. Anal. Appl. 12 (1965) 1–12.

[2] L. Billard, E. Diday, From statistics of data to the statistics of knowledge: symbolic data analysis, Journal of the American Statistical Association 98 (2003) 470–487.

[3] R. Coppi, P. D'Urso, P. Giordani, A. Santoro, Least squares estimation of a linear regression model with LR fuzzy response, Comp. Stat. Data Anal. 51 (2006) 267–286.

[4] R. Coppi, Management of uncertainty in statistical reasoning: the case of regression analysis, Int. J. Approx. Reason 47 (2008) 284–305.

[5] Colubi, A., Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data, Fuzzy sets and systems 160 (2009) 344–356.

[6] L. Di Lascio, L. Ginolfi, A. Albunia, G. Galardi, F. Meschi, A fuzzy-based methodology for the analysis of diabetic neuropathy, Fuzzy Sets and Systems 129 (2002) 203–228.

[7] T. Denoeux, M.H. Masson, P.A. Hérbert, Nonparametric rank-based statistics and significance tests for fuzzy data, Fuzzy Sets and Systems 153 (2005) 1–28.

[8] B. Efron, R.J. Tibshirani, An introduction to the bootstrap, Chapman & Hall, New York (1993).

[9] M. Fréchet, Les éléments aléatoires de natures quelconque dans un áspace distancié, Ann. Inst. H. Poincaré 10 (1948) 215–310.

[10] A.R.Gallant, T.M. Gerig, Computations for constrained linear models, Journal of Econometrics 12 (1980) 59–89.

[11] M.A. Gil, M. Montenegro, G. González-Rodríguez, A. Colubi, M.R. Casals, Bootstrap approach to the multi-sample test of means with imprecise data, Comp. Stat. Data Anal. 51 (2006) 148–162.

[12] M.A. Gil, G. González-Rodríguez, A. Colubi, M. Montenegro, Testing linear independence in linear models with interval-valued data, Comp. Stat. Data Anal. 51 (2007) 3002–3015.

[13] G. González-Rodríguez, A. Colubi, P. D'Urso, M. Montenegro, Multi-sample test-based clustering for fuzzy random variables, Int. J. Approx. Reason (2009), doi:10.1016/j.ijar.2009.01.003.

[14] G. González-Rodríguez, A. Blanco, M.A. Lubiano, Estimation of a simple linear regression model for fuzzy random variables, Fuzzy Sets and Systems 160 (2009) 357–370.

[15] P.J. Heagerty, S.R. Lele, A composite likelihood approach to binary spatial data, Journal of the American Statistical Association 93 (1998) 1099–1111.

[16] E. Klement, M.L. Puri, D.A. Ralescu, Limit theorems for fuzzy random variables, Proc. Roy. Soc. London Ser. A 1832 (1986) 171–182.

[17] R. Körner, Linear models with random fuzzy variables, PhD Thesis, Faculty of Mathematics and Computer Science, Freiberg University of Mining and Technology (1997a)

[18] R. Körner, On the variance of fuzzy random variables, Fuzzy Sets and Systems 92 (1997b) 83–93.

[19] V. Krätschmer, Limit distribution of least squares estimators in linear regression models with vague concepts, Journal of Multivariate Analysis 97 (2006) 1044–1069.

[20] H. Kwakernaak, Fuzzy random variables-I. Definitions and theorems, Inform. Sci. 15 (1978) 1–29.

[21] P. Lagacherie, D.R. Cazemier, R. Martin-Clouaire, T. Wassenaar, A spatial approach using imprecise soil data for modelling crop yields over vast areas, Agriculture, Ecosystems and Environment 81 (2000) 5–16.

[22] C.K. Liew, Inequality constrained least-squares estimation, Journal of the American Statistical Association 71 (1976) 746–751.

[23] M.A. Lubiano, M.A. Gil, M. López-Díaz, M.T. López, The $\overrightarrow{\lambda}$-mean squared dispersion associated with a fuzzy random variable, Fuzzy Sets and Systems111 (2000) 307–317.

[24] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, In: Z.W. Birnbaum, E. Lukacs (Eds), Probability and Mathematical Statistics. Academic Press, Harcourt Brace & Company, Publishers (1995).

[25] W. Näther, Regression with fuzzy random data, Comp. Stat. Data Anal. 51 (2006) 235–252.

[26] C.B. Pipper, C. Ritz, Checking the grouped data version of the Cox model for interval-grouped survival data, Scandinavian Journal of Statistics 34 (2007) 405–418.

[27] M.L. Puri, D.A. Ralescu, Fuzzy random variables, J. Math. Anal. Appl. 114 (1986) 409–422.

[28] J. Ranilla, L.J. Rodríguez-Muñiz, A heuristic approach to learning rules from fuzzy database, IEEE Intelligent Systems 22 (2007) 62–68.

[29] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation,Journal of Electronic Imaging 13 (2004 )146–168.

[30] N.D. Singpurwalla, J.M. Booker, Membership functions and probability measures of fuzzy sets, Journal of the American Statistical Association 99 (2004) 867–877.

[31] P. Walley, A bounded derivative model for prior ignorance about a real-valued parameter, Scandinavian Journal of Statistics 24 (1996) 463–483.

[32] M.S. Yang, C.H. Ko, On a class of fuzzy $c$-numbers clustering procedures for fuzzy data, Fuzzy Sets and Systems 84 (1996) 49–60.

[33] L.A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353.