Bootstrap confidence intervals for the parameters of a linear regression model with fuzzy random variables

Maria Brigida Ferraro[1], Renato Coppi[1], and Gil González-Rodríguez[2]

Abstract Confidence intervals for the parameters of a linear regression model with a fuzzy response variable and a set of real and/or fuzzy explanatory variables are investigated. The family of *LR* fuzzy random variables is considered and an appropriate metric is suggested for coping with this type of variables. A class of linear regression models is then proposed for the center and for suitable transforms of the spreads in order to satisfy the non-negativity conditions for the latter ones. In order to estimate the regression parameters confidence intervals are introduced and discussed. Since there are no parametric models for the imprecise variables, a bootstrap approach has been used. The empirical behavior of the procedure is analyzed by means of simulated data and a real-case study.

Key words: Fuzzy random variables, Linear regression analysis, Confidence regions, Bootstrap approach

1 Introduction

The study of relationships between variables is a crucial point in the investigation of natural and social phenomena. A particular relevance, in this connection, must be given to the analysis of the link between a "response" variable, say Y, and a set of "explanatory" variables, say X_1, X_2,...,X_p.
When approaching this problem from a statistical viewpoint, we realize that several sources of uncertainty may affect the analysis. These concern: a)sampling variability; b)partial or total ignorance about the kind of relationship between Y and $(X_1,...,X_p)$; c)imprecision/vagueness in the way statistical data concerning these variables are measured (see, for instance, Coppi, 2008, for a detailed examination

Dipartimento di Scienze Statistiche - SAPIENZA Università di Roma, Italy mariabrigida.ferraro@uniroma1.it, renato.coppi@uniroma1.it · Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Spain gil@uniovi.es

of the various sources of uncertainty in Regression Analysis).
In the present paper we focus our attention on sources a) and c). To this purpose the notion of Fuzzy Random Variable (FRV) is utilized (Puri & Ralescu, 1986). By means of it, we take simultaneously into account the uncertainty due to the randomness and that pertaining to the imprecision/vagueness of the observed data. A regression model, in this context, aims at establishing a link between a random fuzzy response variable, say $\widetilde{Y}$, and a set of random fuzzy explanatory variables, say $\widetilde{X}_1,...,\widetilde{X}_p$. When this link is expressed in functional terms, the model may be written in the following way:

$$\widetilde{E}\left(\widetilde{Y}\right) = \widetilde{f}\left(\widetilde{X}_1,...,\widetilde{X}_p\right),$$

where both $\widetilde{E}(\cdot)$ and $\widetilde{f}(\cdot)$ are fuzzy sets.
A complete treatment of the above model in the framework of parametric inference would require the explicitation of the family of joint densities

$$p\left(\widetilde{Y},\widetilde{X}_1,...,\widetilde{X}_p/\underline{\theta}\right),$$

where $\underline{\theta}$ is a vector of parameters, or at least of the conditional densities

$$p\left(\widetilde{Y}/\widetilde{X}_1,...,\widetilde{X}_p,\underline{\theta}\right).$$

An attempt in this direction has been made by Näther and co-authors (Näther, 2000, 2006; Wünsche & Näther, 2002). Unfortunately several limitations have been found, when trying to extend the classical theory of linear models to the case of fuzzy variables. One of the causes of these limitations consists in the lack of an appropriate family of sampling models for FRVs supporting the development of a complete inferential theory (consider, for example, the difficulty in defining a suitable notion of Normal FRVs).
In the present paper a different approach is adopted. First, the membership functions of the involved variables are formalized in terms of *LR* fuzzy numbers (in particular triangular fuzzy numbers characterized by three quantities: center, left spread, right spread). Then an appropriate distance between triangular fuzzy numbers is introduced and an isometry is found between the space of triangular fuzzy numbers and $\mathbb{R}^3$. This allows the construction of a parametric regression model linking respectively the center, left spread and right spread of the response variable to the centers and spreads of the explanatory variables. Suitable transforms of the spreads of the response variable are utilized in order to ensure the non-negativity of the estimated spreads.
While the point estimation problem concerning this model has been dealt with in previous works (see Ferraro *et al.*, 2010, 2011; Ferraro & Giordani, 2011), the main objective of the present paper consists in the evaluation of the sampling variation of the estimated regression parameters. This is achieved by means of confidence intervals, which are constructed by applying an appropriate bootstrap procedure.

The work is organized as follows. In Section 2 some preliminary notions concerning fuzzy sets, *LR* fuzzy numbers and their distance, the definition of FRVs and their properties are briefly recalled. In Section 3 the linear regression model for *LR* fuzzy variables is introduced along with the procedure for estimating its parameters. The construction of bootstrap confidence intervals is illustrated in Section 4. A simulation study and a real-case analysis are described in Section 5, while concluding remarks are made in Section 6.

2 Preliminaries

Given a universe U of elements, a fuzzy set $\widetilde{A}$ is defined through the so-called *membership function* $\mu_{\widetilde{A}} : U \to [0,1]$. For a generic $x \in U$, the membership function expresses the extent to which x belongs to $\widetilde{A}$. Such a degree ranges from 0 (complete non-membership) to 1 (complete membership).

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 $\mathscr{F}_{LR}$. A nice property of the LR family is that its elements can be determined uniquely in terms of the mapping $s : \mathscr{F}_{LR} \to \mathbb{R}^3$, i.e., $s(\widetilde{A}) = s_{\widetilde{A}} = (A^m, A^l, A^r)$. This implies that $\widetilde{A}$ 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 $\widetilde{A} \in \mathscr{F}_{LR}$ or (A^m, A^l, A^r). The membership function of $\widetilde{A} \in \mathscr{F}_{LR}$ can be written as

$$\mu_{\widetilde{A}}(x) = \begin{cases} L\left(\frac{A^m - x}{A^l}\right) & x \le A^m,\ A^l > 0, \\ 1_{\{A^m\}}(x) & x \le A^m,\ A^l = 0, \\ R\left(\frac{x - A^m}{A^r}\right) & x > A^m,\ A^r > 0, \\ 0 & x > A^m,\ A^r = 0, \end{cases} \tag{1}$$

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 $L(x) = R(x) = 0, \forall x \in \mathbb{R} \setminus [0,1]$, and 1_I is the indicator function of a set I. $\widetilde{A}$ is a *triangular* fuzzy number if $L(z) = R(z) = 1 - z$, for $0 \le z \le 1$.

The operations considered in $\mathscr{F}_{LR}$ are the natural extensions of the Minkowski sum and the product by a positive scalar for intervals. Going into detail, the sum of $\widetilde{A}$ and $\widetilde{B}$ in $\mathscr{F}_{LR}$ is the *LR* fuzzy number $\widetilde{A} + \widetilde{B}$ 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 $\widetilde{A} \in \mathscr{F}_{LR}$ by a positive scalar γ is

$$\gamma(A^m, A^l, A^r) = (\gamma A^m, \gamma A^l, \gamma A^r).$$

Yang & Ko [13] defined a distance between two *LR* fuzzy numbers $\widetilde{X}$ and $\widetilde{Y}$ as follows

$$D^2_{LR}(\widetilde{X},\widetilde{Y})=(X^m-Y^m)^2+[(X^m-\lambda X^l)-(Y^m-\lambda Y^l)]^2+[(X^m+\rho X^r)-(Y^m+\rho Y^r)]^2,$$

where the parameters $\lambda = \int_0^1 L^{-1}(\omega)\mathrm{d}\omega$ and $\rho = \int_0^1 R^{-1}(\omega)\mathrm{d}\omega$ play the role of taking into account the shape of the membership function. For instance, in the triangular case, it is $\lambda = \rho = \frac{1}{2}$. For what follows it is necessary to embed the space $\mathscr{F}_{LR}$ into $\mathbb{R}^3$ by preserving the metric. For this reason a generalization of the Yang and Ko metric can be derived [6]. Given $a = (a_1,a_2,a_3)$ and $b = (b_1,b_2,b_3) \in \mathbb{R}^3$, it is

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

where $\lambda,\, \rho \in \mathbb{R}^+$. $D^2_{\lambda\rho}$ will be used in the sequel as a tool for quantifying errors in the regression models we are going to introduce.

Let $(\Omega,\mathscr{A},P)$ be a probability space. In this context, a mapping $\widetilde{X} : \Omega \to \mathscr{F}_{LR}$ is an *LR* FRV if the *s*-representation of $\widetilde{X}$, $(X^m,X^l,X^r) : \Omega \to \mathbb{R}\times\mathbb{R}^+\times\mathbb{R}^+$ is a random vector [11]. As for non-fuzzy random variables, it is possible to determine the moments for an *LR* FRV. The expectation of an *LR* FRV $\widetilde{X}$ is the unique fuzzy set $E(\widetilde{X})$ ($\in \mathscr{F}_{LR}$) such that $(E(\widetilde{X}))_\alpha = E(X_\alpha)$ provided that $E\|\widetilde{X}\|^2_{D_{LR}} = E(X^m)^2 + E(X^m-\lambda X^l)^2 + E(X^m+\rho X^r)^2 < \infty$, where X_α is the α-level of fuzzy set $\widetilde{X}$, that is, $X_\alpha = \{x \in \mathbb{R} | \mu_{\widetilde{X}}(x) \geq \alpha\}$, for $\alpha \in (0,1]$, and $X_0 = cl(\{x \in \mathbb{R} | \mu_{\widetilde{X}} \geq 0\})$. Moreover, on the basis of the mapping *s*, we can observe that $s_{E(\widetilde{X})} = (E(X^m),E(X^l),E(X^r))$.

3 A linear regression model with LR fuzzy variables

In our previous works, Ferraro *et al.* (2010a, 2011) and Ferraro & Giordani (2011), we introduced a linear regression model for imprecise information. In the general case an *LR* fuzzy response variable $\widetilde{Y}$ and *p* *LR* fuzzy explanatory variables $\widetilde{X}_1,\widetilde{X}_2,\ldots,\widetilde{X}_p$ observed on a random sample of *n* statistical units, $\{\widetilde{Y}_i,\widetilde{X}_{1i},\widetilde{X}_{2i},\ldots,\widetilde{X}_{pi}\}_{i=1,\ldots,n}$, have been taken into account. We consider the shape of the membership functions as fixed, so the fuzzy response and the fuzzy explanatory variables are determined only by means of three parameters, namely the center and the left and right spreads. We faced the non-negativity constraints of the spreads of the response variable by introducing two invertible functions $g : (0,+\infty) \longrightarrow \mathbb{R}$ and $h : (0,+\infty) \longrightarrow \mathbb{R}$, in order to make the spreads assuming all the real values. In that way we didn't solve a numerical procedure, we formalized a theoretical model and we got a complete solution for the model parameters. The model is formalized as

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

where $\underline{X} = (X_1^m, X_1^l, X_1^r, ..., X_p^m, X_p^l, X_p^r)$ is the row-vector of length $3p$ of all the components of the explanatory variables, ε_m, ε_l and ε_r are real-valued random variables with $E(\varepsilon_m|\underline{X}) = E(\varepsilon_l|\underline{X}) = E(\varepsilon_r|\underline{X}) = 0$, $\underline{a}_m = (a^1_{mm}, a^1_{ml}, a^1_{mr}, ..., a^p_{mm}, a^p_{ml}, a^p_{mr})$, $\underline{a}_l = (a^1_{lm}, a^1_{ll}, a^1_{lr}, ..., a^p_{lm}, a^p_{ll}, a^p_{lr})$ and $\underline{a}_r = (a^1_{rm}, a^1_{rl}, a^1_{rr}, ..., a^p_{rm}, a^p_{rl}, a^p_{rr})$ are row-vectors of length $3p$ of the parameters related to $\underline{X}$. The generic $a^t_{jj'}$ is the regression coefficient between the component $j \in \{m, l, r\}$ of $\widetilde{Y}$ (where m, l and r refer to the center Y^m and the transforms of the spreads $g(Y^l)$ and $h(Y^r)$, respectively) and the component $j' \in \{m, l, r\}$ of the explanatory variables $\widetilde{X}^t$, $t = 1, ..., p$, (where m, l and r refer to the corresponding center, left spread and right spread). For example, a^2_{ml} represents the relationship between the center of the response, Y^m, and the left spread of the explanatory variable $\widetilde{X}^2$ (X^l_2). Finally, b_m, b_l, b_r denote the intercepts. Therefore, by means of (2), we aim at studying the relationship between the response and the explanatory variables taking into account not only the randomness due to the data generation process, but also the information provided by the spreads of the explanatory variables (the imprecision of the data), which are usually arbitrarily ignored.
The covariance matrix of $\underline{X}$ is denoted by $\Sigma_{\underline{X}} = E\left[(\underline{X} - E\underline{X})'(\underline{X} - E\underline{X})\right]$ and Σ stands for the covariance matrix of $(\varepsilon_m, \varepsilon_l, \varepsilon_r)$, with variances, $\sigma^2_{\varepsilon_m}$, $\sigma^2_{\varepsilon_l}$ and $\sigma^2_{\varepsilon_r}$, strictly positive and finite.

3.1 Estimation problem

The estimation problem of the regression parameters is faced by means of the Least Squares (LS) criterion. Accordingly, the parameters of model (2) are estimated by minimizing the sum of the squared distances between the observed and theoretical values of the response variable. However, as already noted, suitable transforms of the spreads are considered in (2). This allows us to use of the generalized metric $D^2_{\lambda\rho}$ in the objective function of the problem. Therefore, the LS problem consists in looking for $\widehat{\underline{a}}_m$, $\widehat{\underline{a}}_l$, $\widehat{\underline{a}}_r$, $\widehat{b}_m$, $\widehat{b}_l$ and $\widehat{b}_r$ minimizing

$$\begin{aligned} \Delta^2_{\lambda\rho} &= D^2_{\lambda\rho}((\underline{Y}^m, g(\underline{Y}^l), h(\underline{Y}^r)), ((\underline{Y}^m)^*, g(\underline{Y}^l)^*, h(\underline{Y}^r)^*)) \\ &= \sum_{i=1}^{n} D^2_{\lambda\rho}((Y^m_i, g(Y^l_i), h(Y^r_i)), ((Y^m_i)^*, g(Y^l_i)^*, h(Y^r_i)^*)) \end{aligned} \tag{3}$$

In order to estimate the regression parameters we consider a least squares criterion and we obtain the following solution

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

where

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

are the centered values of the response variables,

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

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

4 Bootstrap confidence intervals

As in classical Statistics, in this case it is useful to estimate the regression parameters not only by a single value but by a confidence interval too. These intervals represent the reliability of the estimates. How likely the interval is to contain the parameter is determined by the confidence level $1-\alpha$.
Since in the context of FRVs there are not realistic parametric models, we introduce a bootstrap approach. Different approach could be used to construct bootstrap confidence intervals. In this work, we consider confidence intervals based on bootstrap percentiles (see, for more details, Efron & Tibshirani, 1993, and Blanco *et al.*, 2010).
We consider B bootstrap samples drawn with replacement from the observed sample $\{\widetilde{Y}_i, \widetilde{X}_{1i}, \widetilde{X}_{2i}, \ldots, \widetilde{X}_{pi}\}_{i=1,\ldots,n}$. For each sample we compute the estimators of the regression parameters. In this way we obtain sequences of B bootstrap estimators, that represent the empirical distributions of the estimators. Let $\widehat{F}$ be the cumulative distribution function of the bootstrap replications of each estimator. The $1-\alpha$ percentile interval is defined by means of the percentiles of $\widehat{F}$. For example, for the estimator a^1_{ml}, $\widehat{F}^{-1}(\alpha/2)$ is equal to $\widehat{a}^{1\,*(\alpha/2)}_{ml}$, that is, the $100\cdot(\alpha/2)$th percentile of the bootstrap distribution. In details, $\widehat{a}^{1\,*(\alpha/2)}_{ml}$ is the $B\cdot(\alpha/2)$th value in the ordered

list of the B bootstrap estimators $\left(\widehat{a}^{1*(1)}_{ml}, \widehat{a}^{1*(2)}_{ml}, \ldots, \widehat{a}^{1*(B)}_{ml}\right)$. The bootstrap percentile interval for a^1_{ml} is defined as:

$$\begin{aligned} CI_P(a^1_{ml}) &= \left[\widehat{F}^{-1}(\alpha/2), \widehat{F}^{-1}(1-\alpha/2)\right] \\ &= \left[\widehat{a}^{1*(\alpha/2)}_{ml}, \widehat{a}^{1*(1-\alpha/2)}_{ml}\right] \end{aligned}$$

The bootstrap percentile confidence interval for a^1_{ml} is obtained by means of the following algorithm

Algorithm

Step 1: Draw a sample of size n with replacement

$$\left\{(\underline{X}^*_i, Y^{m*}_i, Y^{l*}_i, Y^{r*}_i)\right\}_{i=1,\ldots,n},$$

from the original sample $\left\{(\underline{X}_i, Y^m_i, Y^l_i, Y^r_i)\right\}_{i=1,\ldots,n}$.

Step 2: Compute the bootstrap estimate $\widehat{a}^{1}_{ml}{}^*$.

Step 3: Repeat Steps 3 and 4 a large number B of times to get sets of B estimators for the regression parameter.

Step 4: Approximate the lower and upper limits of the interval by means the quantiles of the empirical distribution obtained at Step3. That is, the values in position $[(\alpha/2)B]+1$ and $[(1-\alpha/2)B]$ of the ordered empirical distribution. We indicate those values as $\widehat{a}^{1}_{mlL}{}^*$ and $\widehat{a}^{1}_{mlU}{}^*$. Thus the percentile confidence interval for a^1_{ml} at the confidence level $1-\alpha$ is

$$CI_P(a^1_{ml}) = \left[\widehat{a}^{1}_{mlL}{}^*, \widehat{a}^{1}_{mlU}{}^*\right]$$

An analogous algorithm could be used to construct the bootstrap percentile confidence intervals for all the regression parameters.

5 Empirical results

In order to check the empirical behaviour of the bootstrap approach to construct confidence intervals for the regression parameters some simulation studies and a real-case example have been developed.

5.1 Simulation studies

We consider a theoretical situation in which an *LR* fuzzy response $\widetilde{Y}$, an *LR* fuzzy explanatory variable $\widetilde{X}_1$ and a real explanatory variable X_2 are taken into account. We deal with the following real random variables: X_1^m behaving as $Norm(0,1)$ random variable, X_1^l and X_1^r as χ_1^2 and χ_2^2, respectively, X_2 as $U(-2,2)$, ε_m as $Norm(0,1)$, ε_l and ε_r as $Norm(0,0.5)$. The response variables are constructed in the following way:

$$\begin{cases} Y^m = 2X_1^m + 0.5X_1^l + 0.4X_1^r + X_2 + \varepsilon_m, \\ Y^2 = g(Y^l) = -1X_1^m + 0.3X_1^l - 0.4X_1^r + 2X_2 + +\varepsilon_l, \\ Y^3 = h(Y^r) = 1.2X_1^m + X_1^l - 0.7X_1^r - X_2 + +\varepsilon_r, \end{cases}$$

During the experiment we employ $B = 1000$ replications of the bootstrap estimator and we carry out $N = 10.000$ iterations of the bootstrap algorithm with the confidence level $\alpha = 0.95$ for different sample sizes ($n = 30, 50, 100, 200, 300$). We compute the empirical confidence levels as the proportion of bootstrap confidence intervals that include the theoretical parameter (on N). The empirical values are reported in Table 1. Since the values gathered in Table 1 tend to the nominal confidence level, as n increases, we can conclude that the bootstrap algorithm perform well in this context.

Table 1 Empirical confidence level of the bootstrap CIs for the regression parameters.

n	30	50	100	200	300
$CI(a^1_{mm})$	.9440	.9352	.9410	.9390	.9475
$CI(a^1_{ml})$	.9381	.9378	.9351	.9382	.9443
$CI(a^1_{mr})$	.9384	.9392	.9410	.9408	.9463
$CI(a^2_{m})$	.9408	.9411	.9431	.9469	.9464
$CI(a^1_{lm})$	.9427	.9348	.9407	.9429	.9484
$CI(a^1_{ll})$	.9363	.9377	.9330	.9394	.9444
$CI(a^1_{lr})$	.9361	.9341	.9400	.9410	.9413
$CI(a^2_{l})$	.9357	.9397	.9551	.9485	.9489
$CI(a^1_{rm})$	.9401	.9364	.9383	.9466	.9466
$CI(a^1_{rl})$	.9371	.9324	.9344	.9405	.9404
$CI(a^1_{rr})$	.9375	.9383	.9403	.9425	.9430
$CI(a^2_{r})$	.9365	.9479	.9450	.9456	.9457
$CI(b_m)$	.9444	.9405	.9467	.9450	.9517
$CI(b_l)$	.9424	.9421	.9475	.9479	.9486
$CI(b_r)$	.9409	.9435	.9471	.9469	.9453

5.2 *A real-case study*

We consider the students' satisfaction of a course. In order to evaluate it, their subjective judgments/ perceptions are observed on a sample of $n = 64$ students (see, for more details, Ferraro & Giordani, 2011). For any student, four characteristics are observed: the overall assessment of the course, the assessment of the teaching staff, the assessment of the course content and the average mark (single-valued variable).

We managed them in terms of fuzzy variables, in particular of triangular type (hence $\lambda = \rho = 1/2$). For analyzing the linear relationship of the overall assessment of the course ($\widetilde{Y}$) on the assessment of the teaching staff ($\widetilde{X}_1$), the assessment of the course contents ($\widetilde{X}_2$) and the average mark (X_3), the proposed linear regression model is employed based on a sample of 64 students. In order to overcome the problem about the non-negativity of spreads estimates, we fix the logarithmic transformation (that is, $g = h = \ln$). Through the *LS* procedure we obtain the following estimated model

$$\begin{cases} \widehat{Y^m} = 1.08X_1^m + 0.13X_1^l - 0.07X_1^r \\ \quad -0.17X_2^m - 0.89X_2^l + 0.66X_2^r - 1.12X_3 + 34.06 \\ \widehat{Y^l} = exp(0.01X_1^m + 0.02X_1^l + 0.02X_1^r \\ \quad +0.00X_2^m + 0.03X_2^l + 0.01X_2^r - 0.00X_3 + 0.67) \\ \widehat{Y^r} = exp(0.00X_1^m + 0.03X_1^l - 0.02X_1^r \\ \quad -0.01X_2^m + 0.03X_2^l + 0.01X_2^r + 0.04X_3 + 1.01) \end{cases}$$

For each regression parameters we obtain the bootstrap percentile confidence intervals reported in Table 2.

Table 2 Bootstrap percentile CIs for the regression parameters at a confidence level equal to 0.95.

$CI(a^1_{mm})$	$[.7888, 1.3403]$	$CI(a^1_{lm})$	$[-.0018, .0199]$	$CI(a^1_{rm})$	$[-.0086, .0142]$
$CI(a^1_{ml})$	$[-.6060, .8087]$	$CI(a^1_{ll})$	$[-.0314, .0556]$	$CI(a^1_{rl})$	$[-.0161, .0633]$
$CI(a^1_{mr})$	$[-.4848, .5013]$	$CI(a^1_{lr})$	$[-.0052, .0358]$	$CI(a^1_{rr})$	$[-.0487, .0101]$
$CI(a^2_{mm})$	$[-.2878, .0324]$	$CI(a^2_{lm})$	$[-.0069, .0071]$	$CI(a^2_{rm})$	$[-.0154, -.0004]$
$CI(a^2_{ml})$	$[-1.4890, -.4884]$	$CI(a^2_{ll})$	$[.0092, .0579]$	$CI(a^2_{rl})$	$[.0021, .0688]$
$CI(a^2_{mr})$	$[.3474, .9626]$	$CI(a^2_{lr})$	$[-.0021, .0249]$	$CI(a^2_{rr})$	$[-.0002, .0330]$
$CI(a^3_m)$	$[-4.3814, .4688]$	$CI(a^3_l)$	$[-.0962, .0953]$	$CI(a^3_r)$	$[-.0473, .1964]$
$CI(b_m)$	$[5.1405, 121.9076]$	$CI(b_l)$	$[-2.0829, 3.16179]$	$CI(b_r)$	$[-3.7414, 3.5874]$

It could be noted from Table 2 that the parameters that are significant are the same obtained by means of a bootstrap test on the regression parameters in Ferraro & Giordani (2011). In details, these are: a^1_{mm}, a^2_{ml}, a^2_{mr}, a^2_{ll}, a^2_{rm} and a^2_{rl}.

6 Concluding remarks

In this paper a linear regression model for *LR* fuzzy variables has been addressed. Along with the least squares estimators, confidence intervals have been introduced and discussed. The results obtained by means of a bootstrap approach are those expected in this context. In details, a bootstrap algorithm to approximate the bootstrap percentile confidence intervals of the parameters has been described and employed to simulated and real data.

Acknowledgements The research in this paper has been partially supported by the Spanish Ministry of Education and Science Grant MTM2009-09440-C02-02 and the COST Action IC0702. Their financial support is gratefully acknowledged.

References

1. Blanco, A., Corral, N., González-Rodríguez, G., Palacio, A.: On some confidence regions to estimate a linear regression model for interval data. In: Borgelt, C., González-Rodríguez, G., Trutschnig, W., Lubiano, M.A., Gil, M.A., Grzegorzewski, P., Hryniewicz, O. (eds.) Combining soft computing and statistical methods in data analysis. Advances in Intelligent and soft computing, **77**, 33–40 (2010).
2. Coppi, R.: Management of uncertainty in statistical reasoning: the case of regression analysis. Int. J. Approx. Reason. **47**, 284–305 (2008).
3. Coppi, R., D'Urso, P., Giordani, P., Santoro, A.: Least squares estimation of a linear regression model with LR fuzzy response. Computational Statistics and Data Analysis **51**, 267–286 (2006).
4. Coppi, R., Gil, M.A, Kiers, H.A.L.: The fuzzy approach to statistical analysis. Comput. Statist. Data Anal. **51**, 1–14 (2006).
5. Efron, B., Tibshirani, R.J.: An introduction to the bootstrap. Chapman & Hall, New York (1993)
6. Ferraro, M.B., Coppi, R., Gonzalez-Rodriguez, G., Colubi, A.: A linear regression model for imprecise response. Int. J. Approx. Reason. **51**, 759–770 (2010).
7. Ferraro, M.B., Colubi, A., Gonzalez-Rodriguez, G., Coppi, R.: A determination coefficient for a linear regression model with imprecise response. Environmetrics **22**, 487–596 (2011).
8. Ferraro, M.B., Giordani, P.: A multiple linear regression model for LR fuzzy random variables. Metrika DOI: 10.1007/s00184-011-0367-3 (in press.) (2011).
9. Näther, W.: On random fuzzy variables of second order and their application to linear statistical inference with fuzzy data. Metrika **51**, 201–221 (2000).
10. Näther, W.: Regression with fuzzy random data. Comp. Stat. Data Anal. **51**, 235–252 (2006).
11. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. J. Math. Anal. Appl. **114**, 409–422 (1986).
12. Wünsche, A., Näther, W.: Least-squares fuzzy regression with fuzzy random variables. Fuzzy Sets Syst. **130**, 43–50 (2002).
13. Yang, M.S., Ko, C.H.: On a class of fuzzy c-numbers clustering procedures for fuzzy data. Fuzzy Sets Syst. **84**, 49–60 (1996).
14. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965).