Fitting parametric link functions in a regression model with imprecise random variables

Maria Brigida Ferraro

Department of Statistical Sciences
Sapienza University of Rome
P.le Aldo Moro, 5 - I-00185 Rome
mariabrigida.ferraro@uniroma1.it

Abstract

In our previous works a new regression model for imprecise random variables has been introduced. The imprecision of a random element has been formalized by means of the fuzzy random variable (FRV). In details, a particular case of FRVs characterized by a center, a left and a right spread, the LR family (LR FRV), has been considered. The idea is to jointly consider three regression models in which the response variables are the center, and two transforms of the left and the right spreads in order to overcome the non-negativity conditions of the spreads. Response transformations could be fixed, as we have done so far, but all inferential procedures, such as estimation, hypothesis tests on the regression parameters, linearity test etc., could be affected by this choice. For this reason in this work we consider a family of parametric link functions, based on the Box-Cox transforms, and by means of a computational procedure we will look for the transformation parameters that minimize the prediction error of the model.

Keywords: LR fuzzy random variable, linear regression model, prediction error, Box-Cox transforms

1. Introduction and motivation

In many contexts the statistical information might be imprecise. In order to manage it the fuzzy sets could be used (see, for more details, Zadeh, 1965). In literature there are different statistical procedures for imprecise

information. In this paper we restrict our attention to a family of regression models with imprecise information previously introduced: Ferraro *et al.* (2010a, 2011) and Ferraro & Giordani (2011). In those works the imprecise elements have been represented by means of a particular kind of fuzzy sets, the LR family, determined by means of three parameters, the center, the left and the right spread, and a particular kind of membership function. The main difficulty when we treat with these data is the non-negativity condition of the spreads. The new family of regression models considers jointly three classical regression models whose responses are, respectively, the center and the transforms of the left and the right spread of the fuzzy response variable, and the explanatory variables are the center, the left and the right spread of each fuzzy explanatory variable. By introducing the transforms we have avoided a restricted procedure and we have obtained analytical solutions.
Response transformation is a usual approach used in the linear regression context (see, for example, Atkinson & Riani, 2000). In practice, the parametric power transform proposed by Box & Cox (1964) is the most used in the linear regression model context. In literature there are many works dealing with this kind of problem (see, for example, Scallan *et al.*, 1984, Edwards & Hamilton, 1995, Foster *et al.*, 2001, Marazzi & Yohai, 2006, Hamasaki & Kim, 2007). This approach is used in order to adjust data to a linear regression model. Response transformations could be fixed, as we have done in our previous works, but all inferential procedures, such as estimation, hypothesis tests on the regression parameters, linearity test etc., could be affected by this choice. For this reason this paper arises in order to overcome this problem. A computational procedure will be introduced by means of a grid search method in order to look for the transformation of the parameters that minimizes the prediction error of the model.
The paper is organized in the following way. In the next section some preliminaries are recalled, in details, the space of fuzzy sets, the concept of fuzzy random variable, an appropriate distance and the basis of this work: a linear regression model for imprecise random variables. In Section 3 the estimation problem and hypothesis testing procedure without fixing transformation functions are described. Section 4 contains some simulation studies that motivate the introduction of a procedure to obtain the transformation functions parameters. A prediction error for a model with imprecise elements and a cross validation procedure to estimate it are introduced and discussed in Section 5. Section 6 focuses on the new computational procedure, described also by means of an algorithm. In order to illustrate the empirical behaviour of

this method, simulation and real case studies are reported, respectively, in Section 6.1 and Section 6.2. Finally, in Section 7 there are some concluding remarks.

2. Preliminaries

2.1. Fuzzy sets and fuzzy random variables

A fuzzy set $\widetilde{A}$ is a subset of the universe U defined through the so-called *membership function* $\mu_{\widetilde{A}}(x)$, $\forall x \in U$, expressing 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 $\mathcal{F}_{LR}$. A nice property of the LR family is that its elements can be determined uniquely in terms of the mapping $s : \mathcal{F}_{LR} \to \mathbb{R}^3$, 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 $\widetilde{A} \in \mathcal{F}_{LR}$ 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 $\widetilde{A}$ and $\widetilde{B}$ in $\mathcal{F}_{LR}$ 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 $\widetilde{A} \in \mathcal{F}_{LR}$ by a scalar $\gamma > 0$ is

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

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

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

where the functions L, $R : \mathbb{R} \to [0, 1]$ are convex upper semi-continuous functions so that $L(0) = R(0) = 1$ and $L(z) = R(z) = 0$, for all $z \in \mathbb{R} \setminus [0, 1]$,

and 1_I is the indicator function of a set I.
$\widetilde{A}$ is a *triangular* fuzzy number if (1) takes the form

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

The α-level set ($0 < \alpha \leq 1$) of $\widetilde{A}$ can be defined as the non-empty compact convex subset of $\mathbb{R}$, A_α, such that $A_\alpha = \left\{x \in U : \mu_{\widetilde{A}}(x) \geq \alpha\right\}$. If $\alpha = 0$, $A_0 = cl(\{x \in \mathbb{R} : \mu_{\widetilde{A}}(x) > 0\})$. For more details one can refer to Zimmermann (2001).
A distance for LR fuzzy numbers has been introduced by Yang & Ko (1996). It is

$$\begin{aligned} D^2_{LR}(\widetilde{A}, \widetilde{B}) = (A^m - B^m)^2 &+ [(A^m - \lambda A^l) - (B^m - \lambda B^l)]^2 \\ &+ [(A^m + \rho A^r) - (B^m + \rho B^r)]^2. \end{aligned} \tag{3}$$

In (3), 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, if the membership function takes the form reported in (2), it is $\lambda = \rho = \frac{1}{2}$. As it will be clear, 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.* 2010a). Given $\underline{a} = (a_1, a_2, a_3)$ and $\underline{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$, it is

$$\begin{aligned} D^2_{\lambda\rho}(\underline{a}, \underline{b}) = (a_1 - b_1)^2 &+ ((a_1 - \lambda a_2) - (b_1 - \lambda b_2))^2 \\ &+ ((a_1 + \rho a_3) - (b_1 + \rho b_3))^2, \end{aligned} \tag{4}$$

where λ, $\rho \in \mathbb{R}^+$. The distance in (4) will be used in the following as a tool for quantifying errors in the regression model we are going to introduce.
In order to jointly consider two kinds of uncertainty, randomness and imprecision, the concept of of fuzzy random variable (FRV) arises. In what follow 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 $\widetilde{X} : \Omega \to \mathcal{F}_{LR}$ such that the α-level set X_α is a random compact convex set for any $\alpha \in [0, 1]$ (see, for further details, Puri & Ralescu, 1985, 1986). As for non-fuzzy random variables, it is possible to determine the moments of a FRV. The expectation of an LR FRV $\widetilde{X}$, $E(\widetilde{X})$, is the fuzzy set in $\mathcal{F}_{LR}$ (EX^m, EX^l, EX^r). With respect to (3) the variance of $\widetilde{X}$ is $\sigma^2_{\widetilde{X}} = var(\widetilde{X}) = E[(D^2_{LR}(\widetilde{X}, E(\widetilde{X}))]$ (see Ferraro *et al.*, 2010a).

2.2. Linear regression model for fuzzy random 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 $\widetilde{X}_1, \widetilde{X}_2, ..., \widetilde{X}_p$ observed on a random sample of n statistical units, $\{\widetilde{Y}_i, \widetilde{X}_{1i}, \widetilde{X}_{2i}, ..., \widetilde{X}_{pi}\}_{i=1,...,n}$, 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

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

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 $E(\varepsilon_m|\underline{X}) = E(\varepsilon_l|\underline{X}) = E(\varepsilon_r|\underline{X}) = 0$, $\underline{a}_m = (a_{mm}^1, a_{ml}^1, a_{mr}^1, ..., a_{mm}^p, a_{ml}^p, a_{mr}^p)$, $\underline{a}_l = (a_{lm}^1, a_{ll}^1, a_{lr}^1, ..., a_{lm}^p, a_{ll}^p, a_{lr}^p)$ and $\underline{a}_r = (a_{rm}^1, a_{rl}^1, a_{rr}^1, ..., a_{rm}^p, a_{rl}^p, a_{rr}^p)$ are row-vectors of length $3p$ of the parameters related to $\underline{X}$. The generic $a_{jj'}^t$ 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_{ml}^2 represents the relationship between the center of the response, Y^m, and the left spread of the explanatory variable $\widetilde{X}^2$ (X_2^l). Finally, b_m, b_l, b_r denote the intercepts. Therefore, by means of (5), 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.
In this context the dependence relationship is strictly related to the shape of the functions g and h, so we aim at studying the gh-linear dependence between the fuzzy response and the fuzzy explanatory variables. In this connection we defined a determination coefficient taking into account the decomposition of the total variation. By indicating $Y^T = (Y^m, g(Y^l), h(Y^r))$, we obtain

$$R^2 = \frac{E\left[D^2_{\lambda\rho}(E(Y^T|\underline{X}), E(Y^T))\right]}{E\left[D^2_{\lambda\rho}(Y^T, E(Y^T))\right]} = 1 - \frac{E\left[D^2_{\lambda\rho}(Y^T, E(Y^T|\underline{X}))\right]}{E\left[D^2_{\lambda\rho}(Y^T, E(Y^T))\right]}. \quad (6)$$

It represents the part of total variation of the gh-scale transformation of $\widetilde{Y}$ explained by the model. This coefficient measures the degree of gh-linear relationship. As in the classical case, it takes values in $[0, 1]$ (see, for more details, Ferraro *et al.*, 2011, and Ferraro & Giordani, 2011). In the sequel, when referring to gh-linear independence we will drop the prefix gh for the sake of brevity.

3. Estimation problem and hypothesis testing procedure without fixing transformation functions

In Ferraro *et al.* (2010a, 2011) and Ferraro & Giordani (2011) we have fixed the transformation functions f and g and then we have estimated the regression parameters and the determination coefficient. In this paper the aim is considering a family of transforms, the Box-Cox transformation model (see, for more details, Box and Cox, 1964). The idea is to not fix a priori the transforms fitting instead, by means of an algorithm, the optimal parameters of the family. In general, the transformed spreads, $g(Y^l)$ and $h(Y^r)$ in model (5), could be expressed as

$$g(Y^l) = \begin{cases} \frac{(Y^l)^{k_1}-1}{k_1}, & k_1 \neq 0 \\ log(Y^l), & k_1 = 0 \end{cases} \quad (7)$$

and

$$h(Y^r) = \begin{cases} \frac{(Y^r)^{k_2}-1}{k_2}, & k_2 \neq 0 \\ log(Y^r), & k_2 = 0 \end{cases} \quad (8)$$

(see, for more details, Box & Cox, 1964).
In order to estimate the regression parameters we consider a least squares criterion and we obtain the following solution

$$\begin{aligned}
\widehat{\underline{a}}_m^{'} &= (\mathbf{X}^{c'}\mathbf{X}^c)^{-1}\mathbf{X}^{c'}\underline{Y}^{mc}, \\
\widehat{\underline{a}}_l^{'}(k_1) &= (\mathbf{X}^{c'}\mathbf{X}^c)^{-1}\mathbf{X}^{c'}g(\underline{Y}^l)^c, \\
\widehat{\underline{a}}_r^{'}(k_2) &= (\mathbf{X}^{c'}\mathbf{X}^c)^{-1}\mathbf{X}^{c'}h(\underline{Y}^r)^c, \\
\widehat{b}_m &= \overline{Y^m} - \overline{\underline{X}}\,\widehat{\underline{a}}_m^{'}, \\
\widehat{b}_l(k_1) &= \overline{g(Y^l)} - \overline{\underline{X}}\,\widehat{\underline{a}}_l^{'}, \\
\widehat{b}_r(k_2) &= \overline{h(Y^r)} - \overline{\underline{X}}\,\widehat{\underline{a}}_r^{'},
\end{aligned}$$

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 $\underline{X}$. In details, $\widehat{\underline{a}}_l^{'}(k_1)$ and $\widehat{b}_l(k_1)$ are functions of the transformation parameter k_1, and $\widehat{\underline{a}}_r^{'}(k_2)$ and $\widehat{b}_r(k_2)$ of the parameter k_2.
By taking into account the decomposition of the total sum of squares (SST), an estimator of the determination coefficient, $\widehat{R}^2(k_1, k_2)$, which is function of k_1 and k_2, can be defined as

$$\widehat{R}^2(k_1, k_2) = 1 - \frac{SSE(k_1, k_2)}{SST(k_1, k_2)} = \frac{SSR(k_1, k_2)}{SST(k_1, k_2)},$$

where

$$\begin{aligned}
SST(k_1, k_2) = \left\|\underline{Y}^m - \underline{1}\,\overline{\underline{Y}^m}\right\|^2 &+ \left\|\left(\underline{Y}^m - \lambda g(\underline{Y}^l)\right) - \left(\underline{1}\,\overline{\underline{Y}^m} - \lambda\underline{1}\,\overline{g(\underline{Y}^l)}\right)\right\|^2 \\
&+ \left\|(\underline{Y}^m + \rho h(\underline{Y}^r)) - \left(\underline{1}\,\overline{\underline{Y}^m} + \rho\underline{1}\,\overline{h(\underline{Y}^r)}\right)\right\|^2,
\end{aligned}$$

is the total sum of squares,

$$SSE(k_1,k_2) = \left\|\underline{Y}^m - \widehat{\underline{Y}^m}\right\|^2 + \left\|\left(\underline{Y}^m - \lambda g(\underline{Y}^l)\right) - \left(\widehat{\underline{Y}^m} - \lambda\widehat{g(\underline{Y}^l)}\right)\right\|^2 + \left\|(\underline{Y}^m + \rho h(\underline{Y}^r)) - \left(\widehat{\underline{Y}^m} + \rho\widehat{h(\underline{Y}^r)}\right)\right\|^2,$$

is the residual sum of squares,

$$SSR(k_1,k_2) = \left\|\widehat{\underline{Y}^m} - \underline{1}\,\overline{\underline{Y}^m}\right\|^2 + \left\|\left(\widehat{\underline{Y}^m} - \lambda\widehat{g(\underline{Y}^l)}\right) - \left(\underline{1}\,\overline{\underline{Y}^m} - \lambda\underline{1}\,\overline{g(\underline{Y}^l)}\right)\right\|^2 + \left\|\left(\widehat{\underline{Y}^m} + \rho\widehat{h(\underline{Y}^r)}\right) - \left(\underline{1}\,\overline{\underline{Y}^m} + \rho\underline{1}\,\overline{h(\underline{Y}^r)}\right)\right\|^2,$$

is the regression sum of squares, with $\widehat{\underline{Y}^m}$, $\widehat{g(\underline{Y}^l)}$, $\widehat{h(\underline{Y}^r)}$ being the vectors of the estimated values, that is,

$$\widehat{\underline{Y}^m} = \mathbf{X}\widehat{\underline{a}}_m^{'} + \underline{1}\,\widehat{b}_m, \quad \widehat{g(\underline{Y}^l)} = \mathbf{X}\widehat{\underline{a}}_l^{'}(k_1) + \underline{1}\,\widehat{b}_l(k_1), \quad \widehat{h(\underline{Y}^r)} = \mathbf{X}\widehat{\underline{a}}_r^{'}(k_2) + \underline{1}\,\widehat{b}_r(k_2).$$

$\widehat{R}^2$ represents the part of the total sum of squares explained by the regression model, so it can be considered as a goodness-of-fit measure, taking values in $[0,1]$.

For model (5) we have introduced different inferential procedures, in particular, a linear independence test has been analyzed in Ferraro *et al.* (2011) and in Ferraro & Giordani (2011), and a linearity test in Ferraro *et al.* (2010b). In order to test the null hypothesis of linear independence, $H_0 : R^2 = 0$, against the alternative $H_1 : R^2 > 0$, the test statistic $T_n = n\widehat{R}^2(k_1,k_2)$ is used. A bootstrap algorithm can be adopted. In order to obtain a bootstrap population fulfilling the null hypothesis, the residual variables $Z^m = Y^m - \underline{X}\,\widehat{\underline{a}}_m^{'}$, $Z^l = g(Y^l) - \underline{X}\,\widehat{\underline{a}}_l^{'}(k_1)$ and $Z^r = h(Y^r) - \underline{X}\,\widehat{\underline{a}}_r^{'}(k_2)$ must be considered. A sample of size n with replacement $\left\{(\underline{X}_i^*, Z_i^{m*}, Z_i^{l*}, Z_i^{r*})\right\}_{i=1,\ldots,n}$ from the bootstrap population is drawn and the bootstrap statistic to be used is

$$T_n^*(k_1,k_2) = n\frac{\sum_{i=1}^{n} D^2_{\lambda\rho}(\widehat{Z_i^{*T}}, \overline{Z^{*T}})}{\sigma^2_{Y^T}},$$

where $Z_i^{*T} = (Z_i^{m*}, Z_i^{l*}, Z_i^{r*})$. The non-parametric bootstrap test is based on the following algorithm:

Algorithm

Step 1: Compute the estimates $\widehat{\underline{a}}_m$, $\widehat{\underline{a}}_l(k_1)$, $\widehat{\underline{a}}_r(k_2)$ and the value of the statistic $T_n(k_1, k_2)$

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

$$\left\{(\underline{X}_i, Z_i^m, Z_i^l, Z_i^r)\right\}_{i=1,\ldots,n}, \tag{9}$$

Step 3: Draw a sample of size n with replacement

$$\left\{(X_i^*, Z_i^{m*}, Z_i^{l^*}, Z_i^{r*})\right\}_{i=1,\ldots,n},$$

from the bootstrap population (9).

Step 4: Compute the bootstrap estimates $\widehat{\underline{a}}_m^*$, $\widehat{\underline{a}}_l^*(k_1)$, $\widehat{\underline{a}}_r^*(k_2)$ and the value of the bootstrap statistic $T_n^*(k_1, k_2)$

Step 5: Repeat Steps 3 and 4 a large number B of times to get a set of B estimators, denoted by $\{T_{n1}^*(k_1, k_2), ..., T_{nB}^*(k_1, k_2)\}$.

Step 6: Compute the bootstrap p-value as the proportion of values in the sequence $\{T_{n1}^*(k_1, k_2), ..., T_{nB}^*(k_1, k_2)\}$ being greater than $T_n(k_1, k_2)$.

4. Synthetic example

In this section we consider synthetic data in order to show the influence of the shape of the transformation functions on some inferential procedures. Both in the context of hypothesis test procedures and in the analysis of the power function, we refer to a specific class of dependence model (borrowed from the shape of the Box-Cox transform).
The choice of the transformation parameters could affect the results of an hypothesis test. Consider the following variables: an LR fuzzy response variable $\widetilde{Y}$, a real explanatory variable X_1 and an LR fuzzy explanatory variable $\widetilde{X}_2$. In details, we deal with the following real random variables: X_1, behaving as $Unif(-2, 2)$ random variable, X_2^m behaving as $Unif(-1, 1)$ random variable, X_2^l and X_2^r as χ_1^2, and ε behaving as a $N(0, 0.2)$, and we construct the center, the left and the right spreads as

$$\begin{aligned}
Y^m &= X_1 + X_2^m + X_2^l + X_2^r + \varepsilon \\
Y^l &= \left[2\left(X_1 + X_2^m + X_2^l + X_2^r + \varepsilon\right) + 1\right]^{\frac{1}{2}} \\
Y^r &= \left[-2\left(X_1 + X_2^m + X_2^l + X_2^r + \varepsilon\right) + 1\right]^{\frac{1}{-2}}
\end{aligned}$$

A sample of $n = 50$ units is drawn from the above variables. If we fix the following transforms

$$g(Y^l) = \frac{(Y^l)^2 - 1}{2}$$

and

$$h(Y^r) = \frac{(Y^r)^{-2} - 1}{-2}$$

by means of a bootstrap linear independence testing procedure (see, for more details, Ferraro *et al.*, 2011, and Ferraro & Giordani, 2011), we obtain a p-value equal to 0, hence we should reject the null hypothesis of linear independence. For different parameters of the Box-Cox transform we could reach the same conclusions but, if for example we use the following parameters

$$g(Y^l) = \frac{(Y^l)^{-2} - 1}{-2}$$

and

$$h(Y^r) = \frac{(Y^r)^2 - 1}{2}$$

we obtain a bootstrap p-value equal to .4520, hence in this case the null hypothesis could not be rejected.
We analyze now the power of the linear independence test. We have drawn a sample of size 50 from the following real random variables: X, behaving as $N(0,1)$ random variable, ε_m behaving as $N(0,1)$ random variable, ε_l and ε_r as $N(0,0.5)$. We construct the center, the left and the right spreads in the following way:

$$\begin{aligned} Y^m &= a_m X + \varepsilon_m \\ Y^l &= \left[2\left(a_l X + \varepsilon_l\right) + 1\right]^{\frac{1}{2}} \\ Y^r &= \left[-2\left(a_r X + \varepsilon_r\right) + 1\right]^{-\frac{1}{2}} \end{aligned} \tag{10}$$

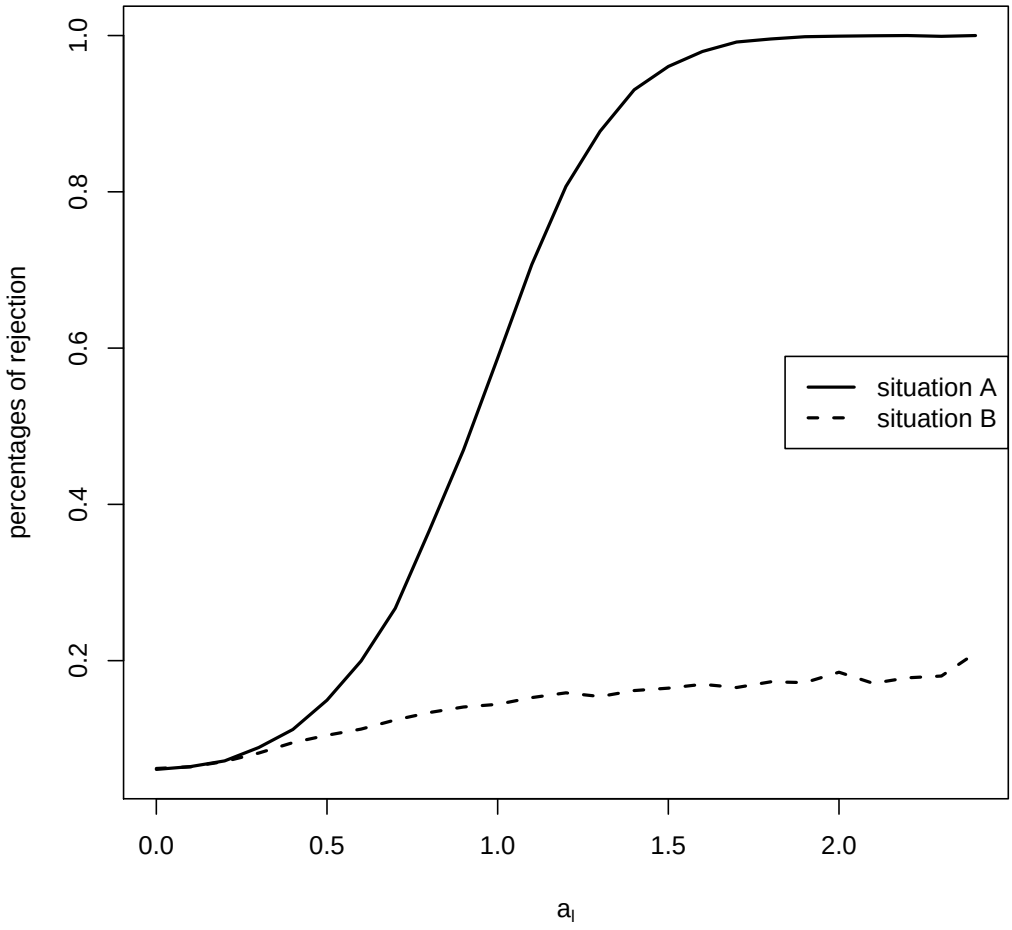
As the values of the parameters a_m, a_l and a_r get large the models tend to the alternative hypothesis so the percentages of rejection approximate the power of the test. According to the way we have constructed the data the logical choices of the parameters of the Box-Cox transforms are, respectively,

Table 1: Empirical percentages of rejection ($n = 50$).

a_m	a_l	a_r	situation A	situation B
0	0	0	6.08	6.20
0	.1	0	6.44	6.41
0	.2	0	7.17	7.09
0	.3	0	8.88	8.17
0	.4	0	11.19	9.51
0	.5	0	14.94	10.45
0	.6	0	19.96	11.24
0	.7	0	26.70	12.41
0	.8	0	36.67	13.36
0	.9	0	46.97	14.07
0	1.0	0	58.73	14.41
0	1.1	0	70.69	15.27
0	1.2	0	80.69	15.88
0	1.3	0	87.74	15.41
0	1.4	0	93.06	16.17
0	1.5	0	96.04	16.48
0	1.6	0	97.96	16.97
0	1.7	0	99.17	16.56
0	1.8	0	99.56	17.29
0	1.9	0	99.85	17.19
0	2.0	0	99.92	18.52
0	2.1	0	99.97	17.10
0	2.2	0	100	17.80
0	2.3	0	100	18.03

$k_1 = 2$ and $k_2 = -2$. We consider two situations: situation A with $k_1 = 2$ and $k_2 = -2$ and situation B with $k_1 = 0$ and $k_2 = 0$ (usual choice in our previous works). The values of the percentages of rejection when $a_m = a_r = 0$ and for increasing values of a_l are reported in table 1. Furthermore, from Figure 1 it is evident that the choice of k_1 and k_2 affects the power of the test. In particular, the power function tends quickly to 1 when the appropriate transforms are used. It seems to slowly increase to 1 with logarithmic transforms.

Figure 1: Empirical percentages of rejection for increasing values of a_l in situation A and situation B



5. Prediction error

In a regression context there are two viewpoints: the structural and the predictive one. From a predictive point of view to check the adequacy of our model it is important to introduce a prediction error. We should have a training set to estimate the regression parameters and a test set to evaluate the regression model by means of the prediction error. We indicate with $\left\{\widetilde{Y}_i^{TR}, \widetilde{X}_{1i}^{TR}, \widetilde{X}_{2i}^{TR}, ..., \widetilde{X}_{pi}^{TR}\right\}_{i=1,...,n_{TR}}$ the training set and with $\{\widetilde{Y}_i^{TS}, \widetilde{X}_{1i}^{TS}, \widetilde{X}_{2i}^{TS}, ..., \widetilde{X}_{pi}^{TS}\}_{i=1,...,n_{TS}}$ the test set. By means of the distance $D^2_{\lambda\rho}$, the prediction error is defined as the expected value of the distance between the observed values of the fuzzy response in the test set and the fitted values of the response constructed by means of the estimators obtained in the training set and the explanatory variables observed in the test set. In details,

$$
\begin{aligned}
PE(k_1,k_2) = E\Bigg(&\left\|\underline{Y}^{mTS} - \left(\mathbf{X}^{TS}\underline{\widehat{a}}_m^{TR\prime} + \underline{1}\,\widehat{b}_m^{TR}\right)\right\|^2 \\
&+ \left\|\left(\underline{Y}^{mTS} - \lambda g(\underline{Y}^{lTS})\right) - \left(\mathbf{X}^{TS}\underline{\widehat{a}}_m^{TR\prime} + \underline{1}\,\widehat{b}_m^{TR} - \lambda\left(\mathbf{X}^{TS}\underline{\widehat{a}}_l^{TR\prime}(k_1) + \underline{1}\,\widehat{b}_l^{TR}(k_1)\right)\right)\right\|^2 \\
&+ \left\|\left(\underline{Y}^{mTS} + \rho h(\underline{Y}^{rTS})\right) - \left(\mathbf{X}^{TS}\underline{\widehat{a}}_m^{TR\prime} + \underline{1}\,\widehat{b}_m^{TR} + \rho\left(\mathbf{X}^{TS}\underline{\widehat{a}}_r^{TR\prime}(k_2) + \underline{1}\,\widehat{b}_r^{TR}(k_2)\right)\right)\right\|^2\Bigg),
\end{aligned}
\tag{11}
$$

where $\underline{a}_m^{TR}$, $\underline{a}_l^{TR}(k_1)$, $\underline{a}_r^{TR}(k_2)$, b_m^{TR}, $b_l^{TR}(k_1)$ and $b_r^{TR}(k_2)$ are the estimators of the regression parameters obtained in the training set.
In practice, there are different approaches to estimate the prediction error. In this work the K-fold cross-validation procedure is performed (see, for more details, Hastie *et al.*, 2009). It consists in splitting the data into K roughly equal-sized parts. For the k-th part we calculate the predicted/fitted values of the response considering the regression parameters estimated by using the remaining $K-1$ parts. That is, the k-th part is considered as test set and the remaining $K-1$ parts are the training set. We repeat this procedure K times. In details, the estimated prediction error by means of a cross validation procedure is

$$
\widehat{PE}_{CV}(k_1,k_2) = \frac{1}{K}\sum_{k=1}^{K} Err_k, \tag{12}
$$

where

$$Err_k = \left(\left\| \underline{Y}_k^{mTS} - \left(\mathbf{X}_k^{TS} \widehat{\underline{a}}_m^{TR-k'} + \underline{1}\, \widehat{b}_m^{TR-k} \right) \right\|^2 \right. \tag{13}$$
$$+ \left\| \left(\underline{Y}_k^{mTS} - \lambda g(\underline{Y}_k^{l\,TS}) \right) - \left(\mathbf{X}_k^{TS} \widehat{\underline{a}}_m^{TR-k'} + \underline{1}\, \widehat{b}_m^{TR-k} - \lambda \left(\mathbf{X}_k^{TS} \widehat{\underline{a}}_l^{TR-k'} + \underline{1}\, \widehat{b}_l^{TR-k} \right) \right) \right\|^2$$
$$\left. + \left\| \left(\underline{Y}_k^{mTS} + \rho h(\underline{Y}_k^{r\,TS}) \right) - \left(\mathbf{X}_k^{TS} \widehat{\underline{a}}_m^{TR-k'} + \underline{1}\, \widehat{b}_m^{TR-k} + \rho \left(\mathbf{X}_k^{TS} \widehat{\underline{a}}_r^{TR-k'} + \underline{1}\, \widehat{b}_r^{TR-k} \right) \right) \right\|^2 \right),$$

with $\widehat{\underline{a}}_m^{TR-k}$, $\widehat{\underline{a}}_l^{TR-k}(k_1)$, $\widehat{\underline{a}}_r^{TR-k}(k_2)$, $\widehat{b}_m^{TR-k}$, $\widehat{b}_l^{TR-k}(k_1)$, $\widehat{b}_r^{TR-k}(k_2)$ that are the regression parameters estimated on the training set obtained by removing the k-th part and $\{\underline{Y}_k^{mTS}, \underline{Y}_k^{l\,TS}, \underline{Y}_k^{rTS}, \mathbf{X}_k^{TS}\}$ is the test set obtained by considering the k-th part.

6. Fitting the parameters of the Box-Cox transformations

All the inferential procedures related to model (5) could be influenced by the choice of the transforms. For this reason it is important to take into account the choice of these functions. That is, it should be introduced a procedure for looking for the transformation parameters. In this work the idea is to get the transformations in the Box-Cox family that minimize the prediction error of the model.

We introduce a standard grid search method in this context (see, for example, Foster *et al.*, 2001). The grid is usually defined by a multidimensional array (in our case we use two dimensions). Each dimension has a range of values. Each range is divided into a set of equal-valued intervals. In our case, the two dimensions are represented by the transformation parameters, k_1 and k_2. For different values of k_1 and k_2 we obtain the estimated regression parameters, $\widehat{a}_m$, $\widehat{b}_m$, $\widehat{a}_l(k_1)$, $\widehat{b}_l(k_1)$, $\widehat{a}_r(k_2)$ and $\widehat{b}_r(k_2)$, and the prediction error estimated by means of cross validation $\widehat{PE}_{CV}(k_1, k_2)$ reported in a matrix/grid whose rows and columns represent the values of the parameters k_1 and k_2. In practice, we consider a specific range of the values of the parameters. Suitable values for k_1 and k_2 are in the compact interval $[-2, 2]$ (see, for more details, Carroll, 1982). The aim is checking the minimum values in the grid/matrix that represent the minimum prediction error.

In order to obtain the expected results we consider the following algorithm

Algorithm

Step 1: For $k_1 = -2$ and $k_2 = -2$ compute the transformed spreads $g(Y^l)$ and $h(Y^r)$, the estimates $\widehat{\underline{a}}_m$, $\widehat{\underline{a}}_l(k_1)$, $\widehat{\underline{a}}_r(k_2)$, $\widehat{b}_m$, $\widehat{b}_l(k_1)$, $\widehat{b}_r(k_2)$ and the value of $\widehat{PE}_{CV}(k_1, k_2)$

Step 2: Repeat Step 1 for k_1 and k_2 from -2 to 2 with increments of 0.1 and obtain a grid/matrix of size 41×41, where the rows represent different values of k_1 and the columns different values of k_2.

Step 3: Choose the minimum in the matrix obtained in Step 2

Step 4: Select the row and the column of the minimum obtained in Step 3. These represent the optimal values of the parameters k_1 and k_2 of the Box-Cox family

6.1. Empirical results

In order to illustrate the empirical behaviour of the algorithm we have analyzed a Monte Carlo simulation. We have created a data set in which the spreads of the fuzzy response are linearly related with the explanatory ones by means of specific transforms. We have generated the following variables: an LR fuzzy response variable $\widetilde{Y}$, a real explanatory variable X_1 and an LR fuzzy explanatory variable $\widetilde{X}_2$. We deal with the following real random variables: X_1, behaving as $Unif(-2, 2)$ random variable, X_2^m behaving as $Unif(-1, 1)$ random variable, X_2^l and X_2^r as χ_1^2, and ε_m, ε_l, ε_r behaving as a $N(0, 0.2)$. We construct the center, the left and the right spreads in the following way:

$$\begin{aligned} Y^m &= X_1 + 1.2X_2^m + 0.3X_2^l + 0.5X_2^r + \varepsilon_m \\ Y^l &= \left[1.2\left(0.7X_1 + X_2^m + 0.4X_2^l + 0.3X_2^r + \varepsilon_l\right) + 1\right]^{\frac{1}{1.2}} \\ Y^r &= \left[-1\left(-0.8X_1 + 1.3X_2^m + X_2^l + 0.4X_2^r + \varepsilon_r\right) + 1\right]^{-1} \end{aligned} \tag{14}$$

We draw N random samples of size n and for each one we estimate the parameters k_1 and k_2 of the transforms by means of the introduced computational procedure. By considering the sequence of N values of the estimated parameters, that is an empirical distribution, we compute the estimated mean and mean squared error for different sample sizes n. In details, for each estimated parameter, we compute $\widehat{E}(\widehat{k}) = \sum_{j=1}^{N} \widehat{k}_j/N$ and $\widehat{MSE}(\widehat{k}) = \sum_{j=1}^{N} \left(\widehat{k}_j - k\right)^2/N$.

Table 2: Estimated mean and mean square error for the estimated transformation parameters $\widehat{k}_1$ and $\widehat{k}_3$ for different sample sizes.

n	$\widehat{E}(\widehat{k}_1)$	$\widehat{MSE}(\widehat{k}_1)$	$\widehat{E}(\widehat{k}_2)$	$\widehat{MSE}(\widehat{k}_2)$
30	1.1681	.0651	-.9959	.0025
50	1.1778	.0221	-.9962	.0012
100	1.1804	.0180	-.9966	.0009
200				
300				

As reported in Table 2, the estimated means tend to the real values of the parameters and the estimated mean square errors tend to 0 as n increases.

6.2. Real case studies

We consider two examples analyzed in two previous work in which the transforms have been considered as fixed. The first one concerns a study about a reforestation in a given area of Asturias (Spain), carried out in the INDUROT institute (University of Oviedo), in which the quality of the trees has been analyzed (see, for more details, Colubi, 2009, and Ferraro *et al.*, 2011). This characteristic has been determined on the basis of subjective judgments/perceptions, through the observation of some characteristics of the trees related to the quality (the leaf structure, the root system, the relationship height/diameter, and so on). The experts perceptions are represented by means of a fuzzy-valued scale, in particular, by means of LR triangular fuzzy numbers. In order to analyze the linear relationship between this characteristic and the height (X_1) and the diameter (X_2)of the trees, the values related to 238 trees have been observed. By means of the new procedure we obtain $k_1 = 1.9$ e $k_2 = -2$ as optimal parameters of Box-Cox family and the corresponding estimated prediction error is equal to 154.3763.

The second example is about 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. By means of the introduced fitting parameters procedure it results that the optimal parameters k_1 and k_2 are, respectively, 1 and -2 and the corresponding estimated prediction error is 41.7363.

7. Concluding remarks

In this paper we have introduced a computational procedure in order to optimize the behaviour of a linear regression model with LR fuzzy elements from a structural and predictive point of view. We have referred to a family of transforms of the left and the right spread of the fuzzy response, the Box-Cox family, we have reported the estimation problem and the hypothesis test procedure without fixing a specific transformation function and we have analyzed and discussed the influence of this choice on some inferential procedures. From this analysis, the necessity of introduced a procedure in order to choose the transformation parameters arises. In order to construct it, a prediction error has been defined. The results obtained seem to be appropriate in this context.

References

[1] Atkinson, A., Riani, M., 2000. Robust diagnostic regression analysis. Springer, New York.

[2] Billard, L., Diday, E., 2000. Regression analysis for interval-valued data. In: Kiers, H.A.L., Rasson, J.-P., Groenen, P.J.F., Schader, M., (Eds.), Data Analysis, Classification and Related Methods. Springer-Verlag, Heidelberg, pp. 369–374.

[3] Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Association 26, 211–252.

[4] Colubi, A., 2009. Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data. Fuzzy Sets and Systems 160, 344-356.

[5] Carroll, R.J., 1982. Prediction and power transformations when the choice of power is restricted to a finite set. Journal of the American Statistical Association 77, 908–915.

[6] Edwards, L.J., Hamilton, S.A., 1995. Errors-in-variables and the Box-Cox transformation. Computational statistics and data analysis 20, 131–140.

[7] Ferraro, M.B., Coppi, R., González-Rodríguez, G., Colubi, A., 2010. A linear regression model for imprecise response. International Journal of Approximate Reasoning 51, 759–770.

[8] Ferraro, M.B., Colubi, A., Giordani, P., 2010. A Linearity Test for a Simple Regression Model with LR Fuzzy Response. In: Borgelt, C.; Gonzlez-Rodrguez, 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, 263–271.

[9] Ferraro, M.B., Colubi, A., González-Rodríguez, G., Coppi, R., 2011. A determination coefficient for a linear regression model with imprecise response. Environmetrics 22, 487-596.

[10] Ferraro, M.B., Giordani, P., 2011. A multiple linear regression model for imprecise information. Metrika DOI: 10.1007/s00184-011-0367-3 (in press.)

[11] Foster, A.M., Tian, L., Wei, L.J., 2001. Estimation for the Box-Cox transformation model without assuming parametric error distribution. Journal of the American Statistical Association 96, 1097–1101.

[12] Hamasaki, T., Kim, S.Y., 2007. Box and Cox power-transformation to confined and censored nonnormal responses in regression. Computational statistics and data analysis 51, 3788–3799.

[13] Hastie T, Tibshirani RJ, Friedman J (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York

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

[15] Marazzi, A., Yohau, V.J., 2006. Robust Box-Cox transformations based on minimum residual autocorrelation. Computational statistics and data analysis 50, 2572–2768.

[16] Puri, M.L., Ralescu, D.A., 1986. Fuzzy random variables. Journal of Mathematical Analysis and Applications 114, 409–422.

[17] Scallan, A., Gilchrist, R., Green, M., 1984. Fitting parametric link functions in generalised linear models. Computational Statistics and Data Analysis 2, 37–49.

[18] 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

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

[20] Zimmermann, H.J., 2001. Fuzzy set theory and its applications. Kluwer Academic, Dordrecht