ON THE GENERALIZATION PERFORMANCE OF A
REGRESSION MODEL WITH IMPRECISE ELEMENTS

MARIA BRIGIDA FERRARO

Department of Statistical Sciences, Sapienza University of Rome, P.le A. Moro 5
00185, Rome, Italy
mariabrigida.ferraro@uniroma1.it

A linear regression model for imprecise random variables is considered. The imprecision of a random element has been formalized by means of the LR fuzzy random variable, characterized by a center, a left and a right spread. In order to avoid the non-negativity conditions the spreads are transformed by means of two invertible functions. To analyze the generalization performance of that model an appropriate prediction error is introduced, and it is estimated by means of a bootstrap procedure. Furthermore, since the choice of response transformations could affect the inferential procedures a computational proposal is introduced for choosing from a family of parametric link functions, the Box-Cox family, the transformation parameters that minimize the prediction error of the model.

Keywords: LR fuzzy random variable; linear regression model; prediction error; bootstrap approach; Box-Cox transforms.

1. Introduction and motivation

In many contexts the statistical information is imprecise. In order to manage it the fuzzy sets are used (see, for more details, Zadeh [33]). In the literature, different statistical procedures for imprecise information are proposed (see, for example, Hung [25], Sun and Wu [31], Sinova et al. [30]). In the regression context in the last years the number of publications is growth (see, An *et al.* [1], Blanco-Fernández et al. [5], Cattaneo and Wiencierz [8], D'Urso *et al.* [15], Giordani [22], Körner and Näther [26]). In this paper we restrict our attention to a family of regression models with imprecise information previously introduced: Ferraro *et al.* [18,19] and Ferraro and Giordani [20]. 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. LR fuzzy sets are a generalization of intervals. 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. The idea to consider a modelling structure based on three connected (or not connected) sub-models has been already proposed by D'Urso and Gastaldi [13] and D'Urso [12]. Those proposals consist 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 center. Those studies are mainly descriptive, and the parameters are estimated by means of a least squares procedure. In D'Urso [12] also an unconstrained approach is addressed by transforming the spreads, but it is important to note that in that case, unlike our proposal, the transforms are related to the estimated center. In our previous work, by introducing the transforms, we have avoided a restricted procedure and we have obtained analytical solutions.

The aim of this paper is twofold. On one hand we aim to analyze the generalization performance of that model. This represents a crucial point in many practical applications and allows us to assess the quality of the model. So, we introduce a specific prediction error. It is defined taking into account the information related to three joint models by using an appropriate distance measure. The squared distance is used as loss function in this context. On the other hand, we study the relationship between the response transformations and the results of the inferential procedures. Response transformation is a usual approach used in the linear regression context (see, for example, Atkinson and Riani [3]). In practice, the parametric power transform proposed by Box and Cox [6] is the most used in the linear regression model context. In the literature there are many works dealing with this kind of problem (see, for example, Scallan *et al.* [29], Edwards and Hamilton [17], Foster *et al.* [21], Marazzi and Yohai [27], Hamasaki and Kim [23]). 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 we propose a computational procedure based on a grid search method in order to look for the parameters of a parametric family that minimize 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 and the basis of this work: a linear regression model for imprecise random variables. In Section 3 an appropriate prediction error is introduced and estimated by means of a bootstrap procedure. Section 4 contains some simulation studies putting in evidence the relationship between the choice of the transformation functions and the inferential procedures. Section 5 focuses on the new computational procedure for fitting the best parameters of the link functions family in terms of prediction error. In addition, in order to illustrate the empirical behaviour of this method, simulation studies are reported in Subsection 5.1. Section 6 contains some real-case applications. 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*. Denoting by $\mathcal{F}_{LR}$ the space of the LR fuzzy numbers, the membership function of $\widetilde{A} \in \mathcal{F}_{LR}$ can be written as

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

where the functions L and R are particular non-increasing shape functions from $\mathbb{R}^+$ to $[0,1]$ such that $L(0) = R(0) = 1$ and $L(x) = R(x) = 0, \forall x \in \mathbb{R} \setminus [0,1]$, 1_I is the indicator function of a set I and A^m, A^l (≥ 0) and A^r (≥ 0) are three real-valued parameters, namely, the center, the left spread and the right spread, respectively. $\widetilde{A}$ is a *triangular* fuzzy number if $L(z) = R(z) = 1 - z$, for $0 \leq z \leq 1$. Given the shape of the membership function, $\widetilde{A} \in \mathcal{F}_{LR}$ can be determined uniquely in terms of the mapping $s : \mathcal{F}_{LR} \to \mathbb{R}^3$, i.e., $s(\widetilde{A}) = s_{\widetilde{A}} = (A^m, A^l, A^r)$. In what follows it is indistinctly used $\widetilde{A} \in \mathcal{F}_{LR}$ or (A^m, A^l, A^r). Note that an interval I is a particular kind of LR fuzzy number where the membership function is 1_I, that is equal to 1, for all $x \in I$, and 0 otherwise. Each fuzzy number can be also defined in terms of α-level sets. The α-level set $(0 < \alpha \leq 1)$ of $\widetilde{A}$ can be also defined as the non-empty compact convex subset of $\mathbb{R}$, A_α, such that $A_\alpha = \{x \in U : \mu_{\widetilde{A}}(x) \geq \alpha\}$. If $\alpha = 0$, $A_0 = cl(\{x \in \mathbb{R} : \mu_{\widetilde{A}}(x) > 0\})$.

A distance for LR fuzzy numbers has been introduced by Yang and Ko [32]:

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

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 case of triangular fuzzy numbers, 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.* [18]). Given $\underline{a} = (a_1, a_2, a_3)$ and $\underline{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$, it is

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

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

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

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 and Ralescu [28]).

2.2. *Linear regression model for fuzzy random variables*

To study the relationship between an LR fuzzy response variable $\widetilde{Y}$ and p LR fuzzy explanatory variables $\widetilde{X}_1, \widetilde{X}_2, ..., \widetilde{X}_p$, 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, a linear regression model with imprecise elements has been proposed by Ferraro and Giordani [20]. The involved variables are observed on a random sample of n statistical units, $\{\widetilde{Y}_i, \widetilde{X}_{1i}, \widetilde{X}_{2i}, ..., \widetilde{X}_{pi}\}_{i=1,...,n}$. We considered 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 did not solve a numerical procedure, we formalized a theoretical model and we got a complete solution for the model parameters. The model is formalized as

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

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 $E(\varepsilon_m|\underline{X}) = E(\varepsilon_l|\underline{X}) = E(\varepsilon_r|\underline{X}) = 0$, $\underline{a}_m = (a_{mm}^1, a_{ml}^1, a_{mr}^1, ..., a_{mm}^p, a_{ml}^p, a_{mr}^p)$, $\underline{a}_l = (a_{lm}^1, a_{ll}^1, a_{lr}^1, ..., a_{lm}^p, a_{ll}^p, a_{lr}^p)$ 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 $\widetilde{Y}^T = (Y^m, g(Y^l), h(Y^r))$ (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). Finally, b_m, b_l, b_r denote the inter-

cepts. The model is well defined also for real (crisp) explanatory variables (see, for more details, Ferraro *et al.* [18,19]). In that case all the vectors have length equal to p. 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 Ferraro *et al.* [18,19] and Ferraro and Giordani [20] we have fixed the transformation functions g and h 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 [6]) and, by means of an algorithm, choosing the optimal parameters of the family. In general, the transformed spreads, $g(Y^l)$ and $h(Y^r)$ in model (2), could be expressed as

$$g(Y^l) = \begin{cases} \frac{(Y^l)^{k_1}-1}{k_1}, & k_1 \neq 0 \\ \\ log(Y^l), & k_1 = 0 \end{cases} \tag{3}$$

and

$$h(Y^r) = \begin{cases} \frac{(Y^r)^{k_2}-1}{k_2}, & k_2 \neq 0 \\ \\ log(Y^r), & k_2 = 0 \end{cases} \tag{4}$$

(see, for more details, Box and Cox [6]).

3. Prediction error

To assess the generalization performance of a learning method is very important in practice [24]. This is related to the accuracy and prediction capability and represents a measure of the quality used for choosing the best model. Hence, to check the adequacy of our model it is important to introduce a gh-prediction error. The aim is to predict the values of the response $\widetilde{Y}^T = (Y^m, g(Y^l), h(Y^r)) \in \mathbb{R}^3$ (gh-scale transformation of $\widetilde{Y}$) by means of the predictors $\widehat{\underline{f}}(\underline{X}) = (\widehat{f}_m(\underline{X}), \widehat{f}_l(\underline{X}), \widehat{f}_r(\underline{X})) \in \mathbb{R}^3$ ($\underline{X} \in \mathbb{R}^{3p}$), where $\widehat{\underline{f}}$ is estimated on the basis of the training set $\mathcal{T}$. Given the training set $\mathcal{T}$, the (gh-)prediction error over an independent test sample (also known as generalization error) is defined as

$$PE_{\mathcal{T}} = E\left(L(\widetilde{Y}^T, \widehat{\underline{f}}(\underline{X}))/\mathcal{T}\right), \tag{5}$$

where L is the loss function. The expected (gh-)prediction error is defined as

$$PE = E(PE_{\mathcal{T}}) = E\left(L(\widetilde{Y}^T, \widehat{\underline{f}}(\underline{X}))\right) \tag{6}$$

It is important to note that the expectation in (6) averages over all the random entities, including the randomness in the training set $\mathcal{T}$. In our case, we consider $D^2_{\lambda\rho}$ as loss function L and a linear model as predictor. Hence, the expected (gh-)prediction

6   *Maria Brigida Ferraro*

error is the expected value of the squared distance between the (unobservable) response $\widetilde{Y}^T = (Y^m, g(Y^l), h(Y^r))$ and the linear regression model, defined in (2), fitted on the basis of the training set, $\widehat{\underline{f}} = \widehat{\widetilde{Y}^T} = (\widehat{Y}^m, \widehat{g(Y^l)}, \widehat{h(Y^r)})$:

$$
\begin{aligned}
PE(k_1, k_2) &= E(D^2_{\lambda\rho}(\widetilde{Y}^T, \widehat{\underline{f}}(\underline{X}))) = E(D^2_{\lambda\rho}(\widetilde{Y}^T, \widehat{\widetilde{Y}^T}(k_1, k_2))) \qquad (7) \\
&= E\left(\left(Y^m - \widehat{Y}^m(k_1, k_2)\right)^2 \right. \\
&\quad + \left((Y^m - \lambda g(Y^l)) - \left(\widehat{Y}^m(k_1, k_2) - \lambda\widehat{g(Y^l)}(k_1, k_2)\right) \right)^2 \\
&\quad \left. + \left((Y^m + \rho h(Y^r)) - \left(\widehat{Y}^m(k_1, k_2) + \rho\widehat{h(Y^r)}(k_1, k_2)\right) \right)^2 \right).
\end{aligned}
$$

The expected $(gh\text{-})$prediction error, $PE(k_1, k_2)$, depends on the parameters k_1 and k_2 of the transformations g and h.

In general, if we consider the expected $(gh\text{-})$prediction error at an input point $\underline{x}_0$, it can be decomposed in the sum of an irreducible error, the squared bias and the variance:

Proposition 1. *Given an input point $\underline{x}_0$, the expected $(gh\text{-})$prediction error, $PE(\underline{x}_0)$, can be decomposed as*

$$
\begin{aligned}
PE(\underline{x}_0) &= E(D^2_{\lambda\rho}(\widetilde{Y}^T, \widehat{\underline{f}}(\underline{x}_0))/\underline{X} = \underline{x}_0) \qquad (8) \\
&= 3\sigma^2_{\varepsilon_m} + \lambda^2\sigma^2_{\varepsilon_l} + \rho^2\sigma^2_{\varepsilon_r} - 2\lambda\sigma_{\varepsilon_m, \varepsilon_l} + 2\rho\sigma_{\varepsilon_m, \varepsilon_r} \\
&\quad + D^2_{\lambda\rho}(E\widehat{\underline{f}}(\underline{x}_0), \underline{f}(\underline{x}_0)) \\
&\quad + E(D^2_{\lambda\rho}(\widehat{\underline{f}}(\underline{x}_0), E\widehat{\underline{f}}(\underline{x}_0))),
\end{aligned}
$$

where, $3\sigma^2_{\varepsilon_m} + \lambda^2\sigma^2_{\varepsilon_l} + \rho^2\sigma^2_{\varepsilon_r} - 2\lambda\sigma_{\varepsilon_m, \varepsilon_l} + 2\rho\sigma_{\varepsilon_m, \varepsilon_r}$ is the irreducible error, $D^2_{\lambda\rho}(E\widehat{\underline{f}}(\underline{x}_0), \underline{f}(\underline{x}_0))$ is the squared bias and $E(D^2_{\lambda\rho}(\widehat{\underline{f}}(\underline{x}_0), E\widehat{\underline{f}}(\underline{x}_0)))$ is the variance of $\widehat{\underline{f}}(\underline{x}_0)$.

Proof. The expected $(gh\text{-})$prediction error at an input point $\underline{x}_0$, $PE(\underline{x}_0)$, is defined as

$$
\begin{aligned}
PE(\underline{x}_0) &= E(D^2_{\lambda\rho}(\widetilde{Y}^T, \widehat{\underline{f}}(\underline{x}_0))/\underline{X} = \underline{x}_0) \\
&= E\left(\left(Y^m - \widehat{f}_m(\underline{x}_0)\right)^2 \right. \\
&\quad + \left((Y^m - \lambda g(Y^l)) - \left(\widehat{f}_m(\underline{x}_0) - \lambda\widehat{f}_l(\underline{x}_0)\right) \right)^2 \\
&\quad \left. + \left((Y^m + \rho h(Y^r)) - \left(\widehat{f}_r(\underline{x}_0) + \rho\widehat{f}_r(\underline{x}_0)\right) \right)^2 /\underline{X} = \underline{x}_0 \right).
\end{aligned}
$$

After a little algebra, it can be written as

$$PE(\underline{x}_0) = 3E\left(\left(Y^m - \widehat{f}_m(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right)$$

$$+\lambda^2 E\left(\left(g(Y^l) - \widehat{f}_l(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right)$$

$$+\rho^2 E\left(\left(h(Y^r) - \widehat{f}_r(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right)$$

$$-2\lambda E\left(\left(Y^m - \widehat{f}_m(\underline{x}_0)\right)\left(g(Y^l) - \widehat{f}_l(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right)$$

$$+2\rho E\left(\left(Y^m - \widehat{f}_m(\underline{x}_0)\right)\left(h(Y^r) - \widehat{f}_r(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right)$$

Since $Y^m = f_m(\underline{X}) + \varepsilon_m$, $g(Y^l) = f_l(\underline{X}) + \varepsilon_l$, $h(Y^r) = f_r(\underline{X}) + \varepsilon_r$ and $E(\varepsilon_m/\underline{X} = \underline{x}_0) = E(\varepsilon_l/\underline{X} = \underline{x}_0) = E(\varepsilon_r/\underline{X} = \underline{x}_0) = 0$, we obtain

$$PE(\underline{x}_0) = 3E\left(\left(f_m(\underline{x}_0) - \widehat{f}_m(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + 3E\left(\varepsilon_m^2 / \underline{X} = \underline{x}_0\right)$$

$$+\lambda^2 E\left(\left(f_l(\underline{x}_0) - \widehat{f}_l(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + \lambda^2 E\left(\varepsilon_l^2 / \underline{X} = \underline{x}_0\right)$$

$$+\rho^2 E\left(\left(f_r(\underline{x}_0) - \widehat{f}_r(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + \rho^2 E\left(\varepsilon_r^2 / \underline{X} = \underline{x}_0\right)$$

$$-2\lambda E\left(\left(f_m(\underline{x}_0) - \widehat{f}_m(\underline{x}_0)\right)\left(f_l(\underline{x}_0) - \widehat{f}_l(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right) - 2\lambda E\left(\varepsilon_m\varepsilon_l / \underline{X} = \underline{x}_0\right)$$

$$+2\rho E\left(\left(f_m(\underline{x}_0) - \widehat{f}_m(\underline{x}_0)\right)\left(f_r(\underline{x}_0) - \widehat{f}_r(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right) + 2\rho E\left(\varepsilon_m\varepsilon_r / \underline{X} = \underline{x}_0\right)$$

The above expression is equal to

$$3E\left(\left(\widehat{f}_m(\underline{x}_0) - E\widehat{f}_m(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + 3E\left(\left(E\widehat{f}_m(\underline{x}_0) - f_m(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + 3\sigma_{\varepsilon_m}^2$$

$$+\lambda^2 E\left(\left(\widehat{f}_l(\underline{x}_0) - E\widehat{f}_l(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + \lambda^2 E\left(\left(E\widehat{f}_l(\underline{x}_0) - f_l(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + \lambda^2 \sigma_{\varepsilon_l}^2$$

$$+\rho^2 E\left(\left(\widehat{f}_r(\underline{x}_0) - E\widehat{f}_r(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + \rho^2 E\left(\left(E\widehat{f}_r(\underline{x}_0) - f_r(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + \rho^2 \sigma_{\varepsilon_l}^2$$

$$-2\lambda E\left(\left(\widehat{f}_m(\underline{x}_0) - E\widehat{f}_m(\underline{x}_0)\right)\left(\widehat{f}_l(\underline{x}_0) - E\widehat{f}_l(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right)$$

$$-2\lambda E\left(\left(E\widehat{f}_m(\underline{x}_0) - f_m(\underline{x}_0)\right)\left(E\widehat{f}_l(\underline{x}_0) - f_l(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right) - 2\lambda\sigma_{\varepsilon_m,\varepsilon_l}$$

$$+2\rho E\left(\left(\widehat{f}_m(\underline{x}_0) - E\widehat{f}_m(\underline{x}_0)\right)\left(\widehat{f}_r(\underline{x}_0) - E\widehat{f}_r(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right)$$

$$+2\rho E\left(\left(E\widehat{f}_m(\underline{x}_0) - f_m(\underline{x}_0)\right)\left(E\widehat{f}_r(\underline{x}_0) - f_r(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right) + 2\rho\sigma_{\varepsilon_m,\varepsilon_r}$$

8   *Maria Brigida Ferraro*

We note that

$$3E\left(\left(\widehat{f}_m(\underline{x}_0) - E\widehat{f}_m(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + \lambda^2 E\left(\left(\widehat{f}_l(\underline{x}_0) - E\widehat{f}_l(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right)$$

$$+\rho^2 E\left(\left(\widehat{f}_r(\underline{x}_0) - E\widehat{f}_r(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right)$$

$$-2\lambda E\left(\left(\widehat{f}_m(\underline{x}_0) - E\widehat{f}_m(\underline{x}_0)\right)\left(\widehat{f}_l(\underline{x}_0) - E\widehat{f}_l(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right)$$

$$+2\rho E\left(\left(\widehat{f}_m(\underline{x}_0) - E\widehat{f}_m(\underline{x}_0)\right)\left(\widehat{f}_r(\underline{x}_0) - E\widehat{f}_r(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right)$$

$$= E(D^2_{\lambda\rho}(\widehat{f}(\underline{x}_0), E\widehat{f}(\underline{x}_0)))$$

and

$$3E\left(\left(E\widehat{f}_m(\underline{x}_0) - f_m(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right) + \lambda^2 E\left(\left(E\widehat{f}_l(\underline{x}_0) - f_l(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right)$$

$$+\rho^2 E\left(\left(E\widehat{f}_r(\underline{x}_0) - f_r(\underline{x}_0)\right)^2 / \underline{X} = \underline{x}_0\right)$$

$$-2\lambda E\left(\left(E\widehat{f}_m(\underline{x}_0) - f_m(\underline{x}_0)\right)\left(E\widehat{f}_l(\underline{x}_0) - f_l(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right)$$

$$+2\rho E\left(\left(E\widehat{f}_m(\underline{x}_0) - f_m(\underline{x}_0)\right)\left(E\widehat{f}_r(\underline{x}_0) - f_r(\underline{x}_0)\right) / \underline{X} = \underline{x}_0\right)$$

$$= D^2_{\lambda\rho}(E\widehat{f}(\underline{x}_0), f(\underline{x}_0)).$$

Hence, the thesis follows.

$\square$

In this context there are not realistic parametric distributions, hence, we use an empirical procedure to estimate the (gh-)prediction error. One possible approach is the bootstrap technique, described in the following section.

3.1. *Bootstrap estimation*

The bootstrap technique is used in many contexts for assessing statistical accuracy. In this case is used to directly estimate the (gh-)prediction error. It is based on re-sampling the training set $\mathcal{T}$. We randomly draw B bootstrap samples with replacement from the training set, each sample has the same size n of the original training set $\mathcal{T}$. Then for each one we compute the (gh-)prediction error. For each bootstrap sample b we estimate $\widehat{\underline{f}}^b$. Each bootstrap sample is considered as a training set and the original training set as test set. For $b = 1, \cdots, B$

$$err_b^{boot} = \frac{1}{n}\left(\left\|\underline{Y}^m - \left(\mathbf{X}\widehat{\underline{a}}_m^{b\,'} + \underline{1}\,\widehat{b}_m^b\right)\right\|^2\right.$$

$$+ \left\|\left(\underline{Y}^m - \lambda g(\underline{Y}^l)\right) - \left(\mathbf{X}\widehat{\underline{a}}_m^{b\,'} + \underline{1}\,\widehat{b}_m^b - \lambda\left(\mathbf{X}\widehat{\underline{a}}_l^{b\,'} + \underline{1}\,\widehat{b}_l^b\right)\right)\right\|^2$$

$$\left. + \left\|(\underline{Y}^m + \rho h(\underline{Y}^r)) - \left(\mathbf{X}\widehat{\underline{a}}_m^{b\,'} + \underline{1}\,\widehat{b}_m^b + \rho\left(\mathbf{X}\widehat{\underline{a}}_r^{b\,'} + \underline{1}\,\widehat{b}_r^b\right)\right)\right\|^2\right),$$

$$(9)$$

where $\widehat{\underline{a}}_m$, $\widehat{\underline{a}}_l(k_1)$, $\widehat{\underline{a}}_r(k_2)$, $\widehat{b}_m$, $\widehat{b}_l(k_1)$, $\widehat{b}_r(k_2)$ are the parameters estimated on the basis of the b-th bootstrap sample (for more details, see Ferraro *et al.* [18], Ferraro and Giordani [20]) and $\|\cdot\|$ is the Euclidean norm. The bootstrap (gh-)prediction error is expressed as

$$\widehat{PE}_{boot}(k_1, k_2) = \frac{1}{B} \sum_{b=1}^{B} err_b^{boot}. \tag{10}$$

Hence a direct bootstrap estimate of the (gh-)prediction error is obtained by the following algorithm.

Algorithm

Step 1: Draw a bootstrap sample b from the training set and compute the estimates $\widehat{\underline{a}}_m^b$, $\widehat{\underline{a}}_l^b(k_1)$, $\widehat{\underline{a}}_r^b(k_2)$, $\widehat{b}_m^b$, $\widehat{b}_l^b(k_1)$, $\widehat{b}_r^b(k_2)$

Step 2: Compute the value of the error err_b^{boot}.

Step 3: Repeat Step 2 and Step 3 B times to get a set of B err_{boot}.

Step 4: Compute the bootstrap estimator $\widehat{PE}_{boot}(k_1, k_2)$, by means of a mean of the B replications of err_b^{boot}

3.2. *Simulation study*

In order to check the adequacy of that approach a simulation study is computed. We consider an LR triangular fuzzy response variable $\widetilde{Y}$, a real explanatory variable X_1 and an LR triangular 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 Y^m as $X_1 + 1.2X_2^m + 0.3X_2^l + 0.5X_2^r + \varepsilon_m$, the left and the right spreads, Y^l and Y^r equal to $\exp\left(0.7X_1 + X_2^m + 0.4X_2^l + 0.3X_2^r + \varepsilon_l\right)$ and $\exp\left(-0.8X_1 + 1.3X_2^m + X_2^l + 0.4X_2^r + \varepsilon_r\right)$, respectively. Furthermore, we transform the spreads by means of a logarithmic transformation ($g = h = \log$). For different number of bootstrap replications B (100, 500, 800 and 1000) we analyze the empirical behaviour of the bootstrap estimator $\widehat{PE}_{boot}$. By using $N = 1000$ iterations of the algorithm we obtain an empirical distribution of the bootstrap estimator and we compute $\widehat{E}(\widehat{PE}_{boot}) = \sum_{j=1}^{N} \widehat{PE}_{boot_j}/N$ and $\widehat{MSE}(\widehat{PE}_{boot}) = \sum_{j=1}^{N} \left(\widehat{PE}_{boot_j} - PE\right)^2/N$ (see Table 1). It results that even for small sample sizes and a number of bootstrap replications equal to 100 the estimated mean of PE_{boot} is quite close to the theoretical value (.14) and the estimated mean squared error is close to 0. As expected, for increasing values of the sample sizes and of bootstrap replications the values are better.

Table 1. Estimated means and mean squared errors for $\widehat{PE}_{boot}$ for different sample sizes, n, and different number of bootstrap replications, B.

n	B	$\widehat{E}(\widehat{PE}_{boot})$	$\widehat{MSE}(\widehat{PE}_{boot})$
30	100	.1473	.0073
50	100	.1427	.0027
100	100	.1405	5.4294e-004
200	100	.1392	7.8762e-004
30	500	.1480	.0080
50	500	.1417	.0017
100	500	.1402	2.3574e-004
200	500	.1400	1.1857e-005
30	800	.1466	.0066
50	800	.1426	.0026
100	800	.1405	4.6414e-004
200	800	.1406	6.3362e-004
30	1000	.1484	.0020
50	1000	.1420	.0084
100	1000	.1411	.0011
200	1000	.1397	3.4988e-004

4. Inferential procedures and choice of the transforms

The choice of the transformation functions could affect the performance of the model. In details, in this section we consider a synthetic dataset 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

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

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.*, [19], and Ferraro and Giordani [20]), 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:

$$\begin{aligned}
Y^m &= a_m X + \varepsilon_m \\
Y^l &= \left[2\left(a_l X + \varepsilon_l \right) + 1 \right]^{\frac{1}{2}} \\
Y^r &= \left[-2\left(a_r X + \varepsilon_r \right) + 1 \right]^{-\frac{1}{2}}
\end{aligned} \tag{11}$$

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, $k_1 = 2$ 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 2. 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.

Table 2. Empirical percentages of rejection ($n = 50$).

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

5. Fitting the parameters of the Box-Cox transformations

As showed in Section 4 the inferential procedures related to model (2) could be influenced by the choice of the transforms. For this reason it is important to take into account this point. That is, it should be introduced a procedure for looking for the transformation parameters. Since our aim is to study the generalization performance of the model, the proposal is to get the transformations in the Box-Cox family that minimize the (gh-)prediction error, that is

$$min_{k_1,k_2}\widehat{PE}_{boot}(k_1, k_2). \tag{12}$$

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

Fig. 1. Empirical percentages of rejection for increasing values of a_l in situation A and situation B
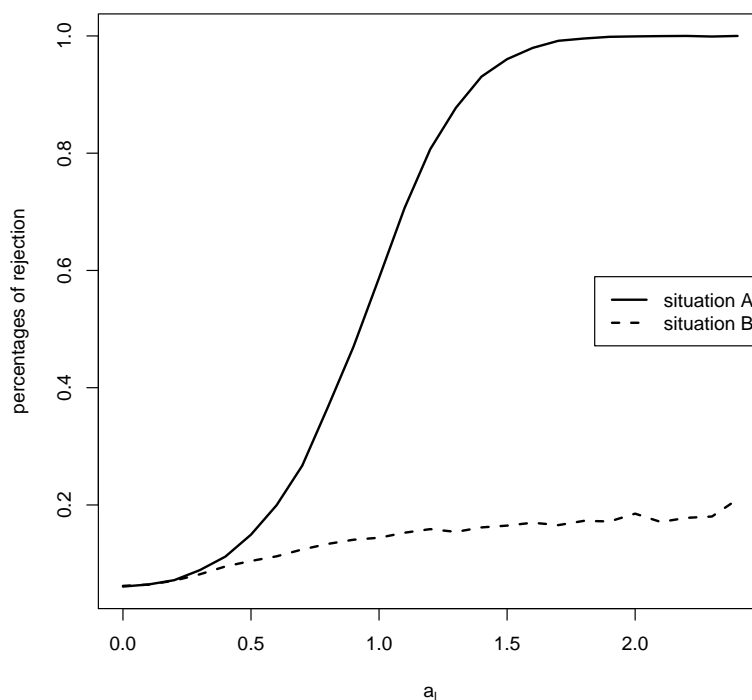
and k_2 we obtain the estimated regression parameters, $\widehat{a}_m, \widehat{b}_m, \widehat{a}_l(k_1), \widehat{b}_l(k_1), \widehat{a}_r(k_2)$ and $\widehat{b}_r(k_2)$, and the $(gh\text{-})$prediction error estimated by means of a bootstrap procedure $\widehat{PE}_{boot}(k_1, k_2)$ reported in a 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 [7]). The aim is checking the values in the grid/matrix that represent the minimum $(gh\text{-})$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 $\widehat{PE}_{boot}(k_1, k_2)$

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

5.1. *Simulation studies*

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 triangular 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:

$$
\begin{aligned}
Y^m &= X_1 + 1.2X_2^m + 0.3X_2^l + 0.5X_2^r + \varepsilon_m \\
Y^l &= \left[1.2\left(0.7X_1 + X_2^m + 0.4X_2^l + 0.3X_2^r + \varepsilon_l\right) + 1\right]^{\frac{1}{1.2}} \\
Y^r &= \left[-1\left(-0.8X_1 + 1.3X_2^m + X_2^l + 0.4X_2^r + \varepsilon_r\right) + 1\right]^{-1}
\end{aligned}
\tag{13}
$$

We draw $N = 100$ 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 (30, 50, 100 and 200) and different number of bootstrap replications B (100, 500, 800, 1000). In details, for each estimated parameter, we compute $\widehat{E}(\widehat{k}_j) = \sum_{j=1}^{N} \widehat{k}_j / N$ and $\widehat{MSE}(\widehat{k}_j) = \sum_{j=1}^{N} \left(\widehat{k}_j - k_j\right)^2 / N$.

As reported in Table 3, the estimated means tend to the real values of the parameters and the estimated mean squared errors tend to 0 as n increases.

6. Applications

6.1. *Quality of trees*

We consider 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 [9], and Ferraro *et al.* [19]). 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

Table 3. Estimated means and mean squared errors of $\widehat{k}_1$ and $\widehat{k}_2$ for different sample sizes, n, and different number of bootstrap replications, B.

n	B	$\widehat{E}(\widehat{k}_1)$	$\widehat{MSE}(\widehat{k}_1)$	$\widehat{E}(\widehat{k}_2)$	$\widehat{MSE}(\widehat{k}_2)$
30	100	1.1490	.3526	.9660	.2977
50	100	1.1750	.0899	.9910	.0708
100	100	1.1870	.0673	.9890	.0508
200	100	1.1890	.0469	.9880	.0354
30	500	1.1790	.0535	.9910	.0286
50	500	1.1860	.0980	.9790	.0765
100	500	1.1810	.00731	.9940	.0443
200	500	1.1840	.0463	.9950	.0260
30	800	1.1800	.1549	.9480	.3002
50	800	1.1740	.0890	.9800	.0648
100	800	1.1900	.0671	.9800	.0490
200	800	1.1900	.0436	.9830	-0376
30	1000	1.1730	.1393	.9430	.2994
50	1000	1.1740	.0901	.9850	.0654
100	1000	1.1800	.0566	.9970	.0411
200	1000	1.1880	.0496	.9870	.0365

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 proposed procedure for different values of the parameter k_1 and k_2 of the Box-Cox family we obtain the results in Table 4.

As reported in Table 4, $k_1 = -1.1$ and $k_2 = .9$ are the optimal parameters of Box-Cox family and the corresponding estimated $(gh\text{-})$prediction error is equal to 805.2947. The estimated model with the optimal parameters is

$$\begin{cases} \widehat{Y}^m = 0.1374X_1 + 1.7937X_2 + 19.6094, \\[2ex] \dfrac{\left(\widehat{Y}^l\right)^{-1.1} - 1}{-1.1} = 0.0001X_1 - 0.0099X_2 + 0.8506, \\[2ex] \dfrac{\left(\widehat{Y}^r\right)^{0.9} - 1}{0.9} = 0.0079X_1 - 1.3209X_2 + 10.5099, \end{cases} \quad (14)$$

hence, $\widehat{Y}^l = [-1.1(0.0001X_1 - 0.0099X_2 + 0.8506) + 1]^{-\frac{1}{1.1}}$ and $\widehat{Y}^r =$

Table 4. Bootstrap estimation of the $(gh\text{-})$prediction error for different values of the parameters k_1 and k_2.

$k_1 \backslash k_2$	-2	-1.9	$\cdots$	0.8	**0.9**	1.0	$\cdots$	1.9	2
-2	807.5	807	$\cdots$	805.9	806.7	806.1	$\cdots$	1123.7	1363.0
-1.9	807.3	807.4	$\cdots$	806.2	805.8	806.5	$\cdots$	1123.9	1364.2
-1.8	807.9	807.9	$\cdots$	806.5	806.1	807.3	$\cdots$	1123.2	1363.5
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
-1.2	807.6	807.3	$\cdots$	806.4	805.8	807.3	$\cdots$	1122.4	1363.3
-1.1	808	807.8	$\cdots$	806.3	**805.3**	805.7	$\cdots$	1123	1364.6
-1	806.5	807.2	$\cdots$	806	806.2	807.4	$\cdots$	1123	1363.1
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
1.8	1054.5	1055.5	$\cdots$	1052.5	1053.6	1053.9	$\cdots$	1370.3	1610.7
1.9	1228.6	1228.3	$\cdots$	1226.9	1227.7	1228.6	$\cdots$	1544.5	1784.7
2	1525.9	1525.8	$\cdots$	1524.8	1525.9	1524.9	$\cdots$	1842.9	2083.2

$$[0.9\,(0.0079X_1 - 1.3209X_2 + 10.5099) + 1]^{\frac{1}{0.9}}.$$

6.2. *Students' satisfaction*

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 and Giordani [20]). 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.

As shown in Table 5, by means of the introduced fitting parameters procedure it results that the optimal parameters k_1 and k_2 are, respectively, 1.2 and -0.8 and the corresponding estimated $(gh\text{-})$prediction error is 114.7735. In this case the

Table 5. Bootstrap estimation of the (gh-)prediction error for different values of the parameters k_1 and k_2.

$k_1 \setminus k_2$	-2	-1.9	$\cdots$	-0.9	**-0.8**	-0.7	$\cdots$	1.9	2
-2	120.4	118.7	$\cdots$	120.3	121.0	120.8	$\cdots$	365.3	522.4
-1.9	121.5	119.6	$\cdots$	120.2	120.7	119.3	$\cdots$	364.0	519.9
-1.8	120.7	120.7	$\cdots$	119.5	119.4	119.7	$\cdots$	372.6	520.4
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
1.1	116.9	115.8	$\cdots$	117.8	116.9	115.8	$\cdots$	361.9	516.4
1.2	114.8	115.7	$\cdots$	116.1	**114.8**	117.0	$\cdots$	359.2	515.5
1.3	117.1	118.4	$\cdots$	116.8	117.2	115.4	$\cdots$	362.9	516.3
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
.	.	.	$\cdots$	.	.	.	$\cdots$	.	.
1.8	172.8	170.3	$\cdots$	172.0	170.5	172.8	$\cdots$	414.3	567.2
1.9	215.2	217.9	$\cdots$	218.7	216.1	216.5	$\cdots$	460.3	613.1
2	292.1	294.8	$\cdots$	294.5	291.3	295.6	$\cdots$	539.9	687.5

estimated model with the optimal parameters is

$$
\begin{cases}
\widehat{Y}^m = 1.0796X_1^m + 0.1253X_1^l - 0.0660X_1^r - 0.1697X_2^m - 0.8870X_2^l + 0.6601X_2^r \\
\qquad -1.1203X_3 + 34.0597, \\[2mm]
\dfrac{\left(\widehat{Y}^l\right)^{1.2} - 1}{1.2} = 0.1887X_1^m + 0.3669X_1^l + 0.2137X_1^r - 0.0336X_2^m + 0.4324X_2^l \\
\qquad + 0.0770X_2^r + 0.0464X_3 - 13.7152, \\[2mm]
\dfrac{\left(\widehat{Y}^r\right)^{-0.8} - 1}{-0.8} = 0.0003X_1^m + 0.0055X_1^l - 0.0026X_1^r - 0.0013X_2^m + 0.0052X_2^l \\
\qquad + 0.0024X_2^r + 0.0049X_3 + 0.8589,
\end{cases}
\tag{15}
$$

hence,

$$
\widehat{Y}^l = \left[1.2\left(0.1887X_1^m + 0.3669X_1^l + 0.2137X_1^r - 0.0336X_2^m + 0.4324X_2^l + 0.0770X_2^r + 0.0464X_3 - 13.7152\right) + 1\right]^{\frac{1}{1.2}}
$$

and

$$
\widehat{Y}^r = \left[-0.8\left(0.0003X_1^m + 0.0055X_1^l - 0.0026X_1^r - 0.0013X_2^m + 0.0052X_2^l + 0.0024X_2^r + 0.0049X_3 + 0.8589\right) + 1\right]^{-\frac{1}{0.8}}.
$$

7. Concluding remarks

In this paper we have analyzed the performance of a linear regression model with LR fuzzy elements. We have introduced a (gh-)prediction error to check the model

quality and we have analyzed a bootstrap approach for estimating it. Since transforms of the left and the right spread of the fuzzy response are used and the choice of those functions could affect the inferential results, a computational procedure for fitting the best parameters of the link functions family (Box-Cox) is addressed. The procedure is based on a grid search method for choosing the parameters minimizing the (gh-)prediction error. The adequacy of the proposal has been checked by means of simulation and real case studies, and the obtained results are those expected in this context.

References

1. S. An, H. Shi, Q. Hu, X. Li and J. Dang, "Fuzzy rough regression with application to wind speed prediction, *Information Sciences* **282** (2014) 388–400.
2. B.F. Arnold and P. Stahlecker, "A surprising property of uniformly best linear affine estimation in linear regression when prior information is fuzzy", *Journal of Statistical Planning and Inference* **140** (2010) 954–960.
3. A. Atkinson and M. Riani, *Robust diagnostic regression analysis* (Springer, New York, 2000).
4. L. Billard and E. Diday, "Regression analysis for interval-valued data", in *Data Analysis, Classification and Related Methods*, eds. H.A.L. Kiers, J.-P. Rasson, P.J.F. Groenen and M. Schader (Springer-Verlag, Heidelberg, 2000) pp. 369–374.
5. G. González-Rodríguez, A. Blanco-Fernández, N. Corral and A. Colubi, "Least squares estimation of linear regression models for convex compact random sets", *Advances in Data Analysis and Classification* **1** (2007) 67–81.
6. G.E.P. Box and D.R. Cox, "An analysis of transformations", *Journal of the Royal Statistical Association* **26** (1964) 211–252.
7. R.J. Carroll, "Prediction and power transformations when the choice of power is restricted to a finite set", *Journal of the American Statistical Association* **77** (1982) 908–915.
8. M.E.G. V. Cattaneo and A. Wiencierz, "On the implementation of LIR: the case of simple linear regression with interval data", *Computational Statistics* **29** (2014) 743–767
9. A. Colubi, "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.
10. R. Coppi and P. D'Urso, "Regression Analysis with Fuzzy Informational Paradigm: A Least-Squares Approach Using Membership Function Information", *International Journal of Pure and Applied Mathematics*, **8** (2003) 279–306.
11. R. Coppi, P. D'Urso, P. Giordani and A. Santoro, "Least Squares Estimation of a Linear Regression Model with LR Fuzzy Response", *Computational Statistics and Data Analysis* **51** (2006) 267–286.
12. P. D'Urso, "Linear Regression Analysis for Fuzzy/Crisp Input and Fuzzy/Crisp Output Data", *Computational Statistics and Data Analysis* **42** (2003) 47–72.
13. P. D'Urso and T. Gastaldi, "A Least-Squares Approach to Fuzzy Linear Regression Analysis", *Computational Statistics and Data Analysis* **34** (2000) 427–440.
14. P. D'Urso and P. Giordani, "A robust fuzzy k -means clustering model for interval valued data", *Computational Statistics* **21** (2006) 251–269.
15. P. D'Urso, R. Massari and A. Santoro, "Robust fuzzy regression analysis", *Information Sciences* **181** (2011) 4154–4174.

16. P. D'Urso and A. Santoro, "Goodness of Fit and Variable Selection in the Fuzzy Multiple Linear Regression", *Fuzzy Sets and Systems* **157** (2006) 2627–2647.

17. L.J. Edwards and S.A. Hamilton, "Errors-in-variables and the Box-Cox transformation", *Computational Statistics and Data Analysis* **20** (1995) 131–140.

18. M.B. Ferraro, R. Coppi, G. González-Rodríguez and A. Colubi, "A linear regression model for imprecise response", *International Journal of Approximate Reasoning* **51** (2010) 759–770.

19. M.B. Ferraro, A. Colubi, G. González-Rodríguez and R. Coppi, "A determination coefficient for a linear regression model with imprecise response", *Environmetrics* **22** (2011) 487–596.

20. M.B. Ferraro and P. Giordani, "A multiple linear regression model for imprecise information", *Metrika* **75** (2012), 1049–1068.

21. A.M. Foster, L. Tian and L.J. Wei, "Estimation for the Box-Cox transformation model without assuming parametric error distribution", *Journal of the American Statistical Association* **96** (2001) 1097–1101.

22. P. Giordani, "Lasso-constrained regression analysis for interval-valued data", *Advances in Data Analysis and Classification* **9** (2015) 5–19.

23. T. Hamasaki and S.Y. Kim, "Box and Cox power-transformation to confined and censored nonnormal responses in regression", *Computational Statistics and Data Analysis* **51** (2007) 3788–3799.

24. T. Hastie, R.J. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2009).

25. W.-L. Hung, "Using statistical viewpoint in developing correlation of intuitionistic fuzzy sets", *Int. J. Unc. Fuzz. Knowl. Based Syst.* **9** (2001) 509.

26. R. Körner and W. Näther, "Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates", *Information Sciences* **109** (1998) 95–118.

27. A. Marazzi and V.J. Yohau, "Robust Box-Cox transformations based on minimum residual autocorrelation", *Computational Statistics and Data Analysis* **50** (2006) 2572–2768.

28. M.L. Puri and D.A. Ralescu, "Fuzzy random variables", *Journal of Mathematical Analysis and Applications* **114** (1986) 409–422.

29. A. Scallan, R. Gilchrist and M Green, "Fitting parametric link functions in generalised linear models", *Computational Statistics and Data Analysis* **2** (1984) 37–49.

30. B. Sinova, S. De La Rosa De Sáa, M. A. Lubiano and M. A. Gil, "An Overview on the Statistical Central Tendency for Fuzzy Data Sets", *Int. J. Unc. Fuzz. Knowl. Based Syst.* **23** (2015) 105.

31. C.-M. Sun and B. Wu, "New statistical approaches for fuzzy data", *Int. J. Unc. Fuzz. Knowl. Based Syst.* **15** (2007) 89.

32. M.S. Yang and 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.