ADAC (2007) 1:67–81
DOI 10.1007/s11634-006-0003-7

REGULAR ARTICLE

Least squares estimation of linear regression models for convex compact random sets

Gil González-Rodríguez · Ángela Blanco · Norberto Corral · Ana Colubi

Published online: 24 January 2007

Abstract Simple and multiple linear regression models are considered between variables whose "values" are convex compact random sets in $\mathbb{R}^p$, (that is, hypercubes, spheres, and so on). We analyze such models within a set-arithmetic approach. Contrary to what happens for random variables, the least squares optimal solutions for the basic affine transformation model do not produce suitable estimates for the linear regression model. First, we derive least squares estimators for the simple linear regression model and examine them from a theoretical perspective. Moreover, the multiple linear regression model is dealt with and a stepwise algorithm is developed in order to find the estimates in this case. The particular problem of the linear regression with interval-valued data is also considered and illustrated by means of a real-life example.

Keywords Linear regression model · Convex compact random sets · Support function · Set arithmetic approach · Point estimation · Least squares method · Interval-valued data · Set-valued data

1 Introduction

Classical simple regression analysis deals with real-valued, possibly multidimensional data and tries to estimate a linear relationship between an explanatory and a target variable. This scenario has been generalized, from different points of view, to the case where the variables take intervals as their values (see, e.g., Diamond 1990; Gil et al. 2001, 2002, 2006; Billard and Diday 2003;

G. González-Rodríguez (✉) · Á. Blanco · N. Corral · A. Colubi
Dpto. de Estadística e I.O., Universidad de Oviedo, 33007 Oviedo, Spain
e-mail: gil@uniovi.es

Lima Neto et al. 2004; De Carvalho et al. 2004; González-Rodríguez et al. 2006). In this paper, we formulate linear regression models for convex compact random sets in $\mathbb{R}^p$ by following a set arithmetic approach (other approaches are discussed in Gil et al. 2006). In fact, we present here a generalization of the models considered in Gil et al. (2001, 2002, 2006) and González-Rodríguez et al. (2006) for the particular case of interval data in $\mathbb{R}$ to the more general case of convex compact sets in $\mathbb{R}^p$.

Interval-valued random sets are useful in handling random experiments in which the interest is focused on essentially imprecise characteristics (like economical fluctuations, ranges, and so on) or it is associated with certain imprecisely known variables, like interval censoring times, or variables from which available data are grouped. A simple example is presented in Sect. 6 where the data intervals describe the daily ranges of pulse and blood pressure measurements of sampled patients.

Convex compact random sets are a natural generalization of interval-valued random sets which allows us to consider also random hypercubes, spheres and so on (see, for instance, Matheron 1975; Stoyan et al. 1987; Cressie 1993; Molchanov 2005).

In the context of interval data, Gil et al. (2002) have considered a kind of "descriptive simple linear regression" formulated by an affine transformation of the underlying random sets, and they have obtained the least squares optimal solutions based on a generalized operational metric. On the other hand, Gil et al. (2006) has considered the simple linear regression model (that will be analyzed in this paper) for the particular case of interval data, and they have shown that testing for linear independence in the underlying linear model reduces to testing for the covariances of certain real-valued variables which characterize the interval-valued random elements (concretely, the mid and the spread values).

The two preceding viewpoints for the regression are in fact quite different. In González-Rodríguez et al. (2006) it was shown by means of some examples and simulations that, contrary to what happens in the case of real-valued variables under the usual normality assumption, the least squares optimal solutions for the affine transformation model do not produce suitable estimates for the simple linear regression model. This is due to the special features of the arithmetic between interval-valued data. In this paper, these previous introductory studies are extended by considering simple and multiple linear regression models for convex compact set-valued data, and then analyzing the resulting estimators from a theoretical point of view.

After recalling some preliminary concepts in Sect. 2, we formulate the considered Simple Linear Regression Model in Sect. 3. The estimation problem is analyzed from a theoretical perspective in Sect. 4 and some illustrative examples are examined. Moreover, a stepwise algorithm to estimate the multiple linear regression problem is given in Sect. 5. In order to illustrate the results in a real-life situation, an example involving interval data is presented in Sect. 6. Finally, Sect. 7 concludes with some final remarks.

2 Preliminaries about random sets, metrics, and expectation

Let $\mathcal{K}_c(\mathbb{R}^p)$ be the class of the nonempty convex compact subsets of $\mathbb{R}^p$ endowed with the Minkowski addition and the product by a scalar, that is, for all sets $A, B \in \mathcal{K}_c(\mathbb{R}^p)$ and $\lambda \in \mathbb{R}$ we have

$$A + B = \{a + b \,|\, a \in A\,, b \in B\} \quad \text{and} \quad \lambda A = \{\lambda a \,|\, a \in A\}$$

It should be remarked that $[\mathcal{K}_c(\mathbb{R}^p), +, \cdot]$ is not a linear space due to the lack of a symmetric element w.r.t. the addition. For this reason, it is useful to consider the *Hukuhara difference* $A -_H B$, which is defined as the set difference C, provided that $C \in \mathcal{K}_c(\mathbb{R}^p)$, so that $A = B + C$. In the 1D case, if $A, B \in \mathcal{K}_c(\mathbb{R})$, we have that $A -_H B = [\inf A - \inf B, \sup A - \sup B]$ iff $\inf A - \inf B \leq \sup A - \sup B$.

The *support function* of a set $A \in \mathcal{K}_c(\mathbb{R}^p)$ is defined by $s_A(u) = \sup_{a \in A} \langle a, u \rangle$ for any $u \in \mathbb{S}^{p-1}$, where $\mathbb{S}^{p-1}$ is the unit sphere in $\mathbb{R}^p$ and $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^p$. The support function s_A is continuous and characterizes the convex set A. In fact the function $s : \mathcal{K}_c(\mathbb{R}^p) \to \mathcal{C}(\mathbb{S}^{p-1})$ defined by $s(A)(u) = s_A(u)$ for $A \in \mathcal{K}_c(\mathbb{R}^p)$, and $u \in \mathbb{S}^{p-1}$, embeds $\mathcal{K}_c(\mathbb{R}^p)$ onto a cone included in the class of all continuous functions $\mathcal{C}(\mathbb{S}^{p-1})$. The support function is semilinear, in the sense that, $s_{A+B} = s_A + s_B$ and $s_{\lambda A} = \lambda s_A$ and, furthermore, if $A -_H B$ exists, then $s_{A -_H B} = s_A - s_B$ for all $A, B \in \mathcal{K}_c(\mathbb{R}^p)$ and $\lambda > 0$ (see, for instance, Diamond and Kloeden 1994).

In order to formalize least-squares estimation methods we need some distance measure between convex sets or intervals. For example, the least-squares estimation method in Gil et al. (2002) is based on the d_W*-distance* between intervals, W being a non-degenerate symmetric probability measure on the real Borel space $([0, 1]), \mathcal{B}_1([0, 1])$, which, for all $A, B \in \mathcal{K}_c(\mathbb{R})$, is defined by

$$d_W(A, B) = \sqrt{\int_{[0,1]} \left[f_A(\lambda) - f_B(\lambda)\right]^2 \, d\, W(\lambda)}$$

with $f_A(\lambda) = \lambda \sup A + (1 - \lambda) \inf A$ (see Bertoluzza et al. 1995).

In Gil et al. (2002) it is shown that the d_W-metric has a suitable intuitive meaning, in addition to very good operational properties. The usefulness of this metric is mainly due to two facts, namely,

- it is an L_2-type metric, which implies very good statistical properties in connection with least squares methods;
- it involves not only distances between extreme points (infima and suprema), but also distances between inner points in the intervals, except for $W(0) = W(1) = 0.5$, which involves only distances between the extremes.

The d_W-metric can be generalized (see Körner and Näther 2002) in the following way that preserves these useful properties:

$$D_K(A,B) = \sqrt{\int_{S^{p-1}} (s_A(u) - s_B(u))(s_A(v) - s_B(v))dK(u,v)}$$

for $A, B \in \mathcal{K}_c(\mathbb{R}^p)$ where K is a symmetric and positive definite kernel function. When considering the *metric* D_K in $\mathcal{K}_c(\mathbb{R}^p)$ the functional s establishes an isometry between the space $\mathcal{K}_c(\mathbb{R}^p)$ and a closed convex cone of the Hilbert space $[\mathcal{C}(\mathbb{S}^{p-1}), \|\cdot\|_K]$, where $\|\cdot\|_K$ denotes the L_2 distance w.r.t. the kernel K. Thus, if $<\cdot,\cdot>_K$ denotes the associated inner product, we have that

$$D_K^2(A,B) = < s_A - s_B, s_A - s_B >_K .$$

This metric will be used later on.

If $(\Omega, \mathcal{A}, P)$ is a probability space, a *convex compact random set* in $\mathbb{R}^p$ is usually defined as a Borel measurable mapping $X : \Omega \to \mathcal{K}_c(\mathbb{R}^p)$ with respect to the σ-field generated by the topology induced by the well-known Hausdorff metric d_H on $\mathcal{K}_c(\mathbb{R}^p)$. In this respect, it should be noted that the Hausdorff metric and D_K are topologically equivalent (see Diamond and Kloeden 1994; Körner and Näther 2002). When $p = 1$, the random sets above will be referred to as *interval-valued random sets*. An interval-valued random set X can be characterized by means of the random vector $(\text{mid}\, X, \text{spr}\, X)$ where $\text{mid}\, A = (\sup A + \inf A)/2$ and $\text{spr}\, A = (\sup A - \inf A)/2$ denote the *mid-point* (center) and the *spread* (radius) of the interval $A \in \mathcal{K}_c(\mathbb{R})$, respectively. Similarly, the *random rectangles* X parallel to the Cartesian axes in $\mathcal{K}_c(\mathbb{R}^2)$ can be parameterized by a 4D random vector $(\text{mid}_1 X, \text{mid}_2 X, \text{spr}_1 X, \text{spr}_2 X)$, where $(\text{mid}_1 A, \text{mid}_2 A)$ denotes the 2D center, and $(\text{spr}_1 A, \text{spr}_2 A)$ the 2D spread (the first coordinate being associated with the x-axis and the second with the y-axis) for $A \in \mathcal{K}_c(\mathbb{R}^2)$.

There are various ways to define the "expectation" of a random set. We will use the following Aumann approach: let $X : \Omega \to \mathcal{K}_c(\mathbb{R}^p)$ be a convex compact random set such that $E(|X|) < \infty$ [with $|X|(\omega) = \sup\{|x| \,|\, x \in X(\omega)\}$ for $\omega \in \Omega$]. Then, the *expected value of X in Kudō-Aumann's sense* (see, e.g., Aumann 1965) is given by the set of all the means of the integrable p-dimensional real-valued random vectors that are contained in X with probability one:

$$E^A[X] = \left\{ E(f) \,\middle|\, f : \Omega \to \mathbb{R}, f \in L^1, f \in X \, a.s.\,[P] \right\}$$

If X is an interval-valued random set, the expected value $E^A[X]$ can be written as $[E(\inf X), E(\sup X)]$. Moreover, if $E(|X|^2) < \infty$, the *variance* of X is defined by $\text{Var}(X) = E\left(\left[D_K\left(X, E^A[X]\right)\right]^2\right)$ (see, for instance, Lubiano et al. 2000; Körner and Näther 2002).

3 The affine transformation model and the simple linear regression model

Let $\{(x_i, y_i)\}_{i=1}^n$ be a data matrix with $(x_i, y_i) \in \mathcal{K}_c(\mathbb{R}^p) \times \mathcal{K}_c(\mathbb{R}^p)$ for all $i = 1, \ldots, n$. This matrix can be viewed as a realization of a random sample

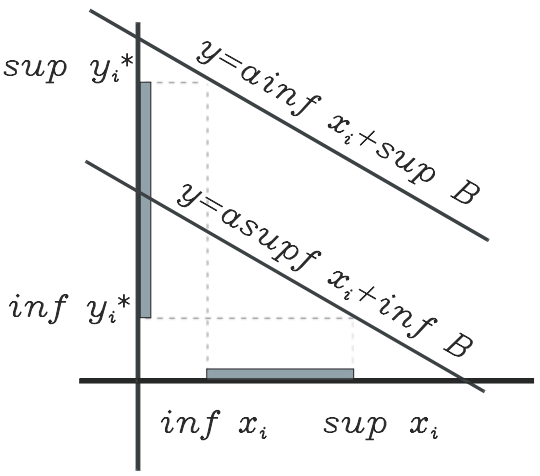

Fig. 1 Affine transformation $y_{i*} = ax_i + B$ for intervals with $a < 0$

$\{X_i, Y_i\}_{i=1}^n$ obtained from a random element $(X, Y) : \Omega \to \mathcal{K}_c(\mathbb{R}^p) \times \mathcal{K}_c(\mathbb{R}^p)$. In order to relate X and Y by some functional form, two interesting statistical-probabilistic models based on the arithmetic defined in Sect. 2 can be considered:

Affine transformation $(X, Y) : \Omega \to \mathcal{K}_c(\mathbb{R}^p) \times \mathcal{K}_c(\mathbb{R}^p)$ with $Y^* = aX + B$ with $a \in \mathbb{R}$ and a set $B \in \mathcal{K}_c(\mathbb{R}^p)$.

Simple linear regression model $(X, Y) : \Omega \to \mathcal{K}_c(\mathbb{R}^p) \times \mathcal{K}_c(\mathbb{R}^p)$ where $Y = aX + \epsilon_X$, where $a \in \mathbb{R}$ and ϵ_X is a random set with a fixed expected value $E^A[\epsilon_X] = B \in \mathcal{K}_c(\mathbb{R}^p)$. This implies that $E^A[Y|x] = ax+B$ for any realization $x \in \mathcal{K}_c(\mathbb{R}^p)$ of X.

As an example in Fig. 1 we show how the affine transformation $y_i^* = ax_i + B$ operates on a interval x_i to get y_i^*. It should be noted that since $a < 0$, we obtain $\inf y_i^*$ by linear transformation of $\sup x_i$ and $\sup y_i^*$ by linear transformation of $\inf x_i$.

Remark 1 In the Simple Linear Regression model it is quite restrictive to assume that the errors ϵ_X are centered at $\{0\}$ (that is $E^A[\epsilon_X] = \{0\}$), in contrast to what is usual for random variables, because the lack of linearity of $\mathcal{K}_c(\mathbb{R}^p)$ would make the errors degenerate into real-valued random variables. For this reason, in order to consider set-valued errors, the independent term is included in the formalization of the possible errors. On the other hand, the model $Y = aX + \epsilon_X$ implies the existence of the Hukuhara difference $Y -_H aX$, which should be taken into account in order to estimate the parameter a. □

Remark 2 In the Affine Transformation Model the estimation problem consists in finding the scale parameter $a^* \in \mathbb{R}$ and a set $B^* \in \mathcal{K}_c(\mathbb{R}^p)$ such that $\{(x_i, a^*x_i + B^*)\}_{i=1}^n$ is as close to the data matrix $\{(x_i, y_i)\}_{i=1}^n$ as possible according to a given criterion, which in this paper will be based on the least squares approach (see Sect. 4).

On the other hand, in the Simple Linear Regression Model the estimation problem consists in finding values $\widehat{a} \in \mathbb{R}$ and a set $\widehat{B} \in \mathcal{K}_c(\mathbb{R}^p)$ such that $\{(x_i, \widehat{a}x_i + \widehat{B})\}_{i=1}^n$ is as close to the data matrix as possible and, simultaneously, $y_i = \widehat{a}x_i + \epsilon_i$ holds for a some $\epsilon_i \in \mathcal{K}_c(\mathbb{R}^p)$, that is, so that $y_i -_H \widehat{a}x_i$ exists for all $i = \{1, \ldots, n\}$.

The two viewpoints are different in the sense that,

- in the Affine Transformation problem, the affine function closest to the data matrix is to be determined, but data are not assumed to fulfil this affine transformation,
- in the Simple Linear Regression problem the aim is to find the best linear relationship and, under such an assumption, the collected data are supposed to satisfy the regression model.

The first one is a numerical problem related to a descriptive statistical approach, whereas the second one is an estimation problem related to an inferential approach. □

Gil et al. (2002) obtained the least squares optimal solution in the d_W sense for the interval-valued case and for the affine transformation model. In González-Rodríguez et al. (2006) it has been shown that, contrary to what happens in the case of random variables, the least squares optimal solutions for the affine transformation do **not** produce suitable estimates for the linear regression model, because we may obtain estimates a^* for which the difference $Y -_H a^*X$ does not make sense (i.e., there may exist observations for which $y_i -_H a^*x_i$ is not a set). For instance, if $p = 1$ and the estimate a^* is closer to 0 than the true value of a (i.e., $|a^*| \leq |a|$) then, under the linear model assumption we have that $a^*X \subset aX \subset Y$, whence $Y -_H a^*X$ is well-defined. On the contrary, if a^* is such that $|a^*| > |a|$, the last assertion is not always true.

Example 1 To illustrate the last comments, consider $(X, Y) : \Omega \to \mathcal{K}_c(\mathbb{R}) \times \mathcal{K}_c(\mathbb{R})$ so that $Y = X + \epsilon$, mid X and mid ϵ are independent random variables with $\mathcal{N}(0, 1)$ distribution and spr X and spr ϵ are independent random variables with χ^2_1 distribution . Three realizations (data) from X and ϵ have been simulated (see Table 1). For these data we have that spr $y_1 =$ spr $x_1 +$ spr $\epsilon_1 = 0.2863 + 0.0893 = 0.3736$. If the least squares solutions are computed by applying the algorithm in Gil et al. (2002), then we obtain $a^* = 2.2644$ and $B^* = [-0.3282, 0.4041]$, whence spr $(a^*x_1) = 2.2644 \cdot 0.2863 = 0.6478$. Since spr $(a^*x_1) >$ spr y_1, it is not possible to compute the Hukuhara difference of y_1 and a^*x_1, and hence it is not possible to find any error ϵ_1^* so that $y_1 = a^*x_1 + \epsilon_1^*$. □

To overcome these difficulties González-Rodríguez et al. (2006) proposed a restricted least squares method for the *interval-valued case* which was empirically shown to improve the unrestricted least squares method in terms of the mean squared error. In this paper, we generalize this method to the *case of convex compact sets* and investigate it from a theoretical point of view. Additionally, we also analyze a multiple linear regression model.

Table 1 Simulated data to estimate the model $E^A[Y|x] = x + [-1, 1]$

mid x_i	spr x_i	mid ϵ_i	spr ϵ_1
0.6561	0.2863	0.1238	0.0893
−0.0334	0.06533	1.0936	0.8334
−0.2719	0.5166	−1.7599	1.3875

4 Estimation in the simple linear regression model

Here we propose to restrict the Least Squares estimators for the unknown parameter a and the set B to satisfy the Linear Regression Model, at least over the set of observed data. The aim of this method can be formalized as follows: to look for $\widehat{a} \in \mathbb{R}$ and $\widehat{B} \in \mathcal{K}_c(\mathbb{R}^p)$ in order to

$$\begin{cases} \text{Minimize } \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} D_K^2(Y_i, aX_i + B) [\text{w.r.t. } a \in \mathbb{R} \text{ and } B \in \mathcal{K}_c(\mathbb{R}^p)] \\ \text{subject to the constraint that } Y_i -_H aX_i \text{ exists for all } i = 1, \ldots, n. \end{cases}$$

Let $A = \{a \in \mathbb{R} \mid Y_i -_H aX_i \text{ exists for all } i = 1, \ldots, n\}$ denote the set of all as which fulfil the constraint. It is possible to show that A is a nonempty, closed and convex subset. Actually either, $A = \mathbb{R}$, or there exist $a^0, b^0 \in [0, +\infty)$, so that $A = [-a^0, b^0]$. In addition, it is easy to verify that for a given sample $\{y_i, x_i\}_{i=1}^n$, $A = \mathbb{R}$ if, and only if, $\text{Card}(x_i) = 1$ for all $i = 1, \ldots, n$ [i.e., the sets $x_i \in K_c(\mathbb{R}^p)$ are all degenerated into real numbers].

As an example, if we consider the case in which X and Y are interval-valued random sets, then $y_i -_H ax_i$ exists if, and only if, $\text{spr}\, ax_i \leq \text{spr}\, x_i$. Thus, if $\text{spr}\, x_i > 0$ [that is, $\text{card}(x_i) > 1$], then $|a| \leq \text{spr}\, y_i / \text{spr}\, x_i$ for all $i = 1, \ldots, n$, whence $a^0 = b^0 = \min_{i=1\ldots n} \text{spr}\, y_i / \text{spr}\, x_i$.

From now on, $\overline{X}$ and $\overline{Y}$ will denote the sample means [that is, the sets $\overline{X} = (X_1 + \cdots + X_n)/n$ and $\overline{Y} = (Y_1 + \cdots + Y_n)/n$], $\hat{\sigma}_X^2$ will denote the sample variance [that is, $\hat{\sigma}_X = \overline{D_K(X, \overline{X})^2}$], and $\hat{\sigma}_{X,Y}^2$ will denote the sample covariance of the corresponding support functions (that is, $\hat{\sigma}_{X,Y} = \overline{< s_Y - s_{\overline{Y}}, s_X - s_{\overline{X}} >_K}$). The support function allows us to search the least squares estimator of parameter B in a way similar to that for the case of real-valued data.

Theorem 1 *Given $a \in A$, the minimum of $\frac{1}{n}\sum_{i=1}^n D_K^2(Y_i, aX_i + B)$ over $B \in \mathcal{K}_c(\mathbb{R}^p)$ is attained at $B(a) = \overline{Y} -_H a\overline{X}$.*

Proof Consider $a \in A$. Given that there exists $Y_i -_H aX_i$ for all $i = 1 \ldots n$, we have

$$\frac{1}{n}\sum_{i=1}^{n} D_K^2(Y_i, aX_i + B) = \frac{1}{n}\sum_{i=1}^{n} D_K^2(Y_i -_H aX_i, B).$$

The definition of the variance w.r.t. the D_K metric satisfies the Fréchet approach (see Körner and Näther 2002), which means that the set that minimizes the

square distance to a variable is precisely the mean, whence the expression is minimized at $B(a) = \overline{Y -_H aX} = \overline{Y} -_H a\overline{X}$. □

As a result of Theorem 1, the Least Squares minimization problem reduces to

$$\begin{cases} \text{Find } \widehat{a} \in \mathbb{R} \text{ minimizing } \phi(a) := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} D_K^2(Y_i -_H aX_i, \overline{Y} -_H a\overline{X}) \\ \text{subject to } a \in A. \end{cases}$$

Theorem 2 *Whenever $\hat{\sigma}_X^2 > 0$, the function $\phi(a) = \frac{1}{n}\sum_{i=1}^{n} D_K^2(Y_i -_H aX_i, \overline{Y} -_H a\overline{X})$ is minimized w.r.t $a \in A$ at the following value: a^* such that*

For $A = \mathbb{R}$ the optimal $a \in A$ is given by

$$a^* = \begin{cases} 0 & \text{if } \hat{\sigma}_{-X,Y} \leq 0 \text{ and } \hat{\sigma}_{X,Y} \leq 0 \\ \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2} & \text{if } \hat{\sigma}_{X,Y} > 0 \text{ and } \hat{\sigma}_{-X,Y} \leq \hat{\sigma}_{X,Y} \\ -\frac{\hat{\sigma}_{-X,Y}}{\hat{\sigma}_X^2} & \text{if } \hat{\sigma}_{-X,Y} > 0 \text{ and } \hat{\sigma}_{X,Y} \leq \hat{\sigma}_{-X,Y} \end{cases}$$

For $A = [-a^0, b^0]$ the optimal $a \in A$ is given by

$$a^* = \begin{cases} \beta\frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2} - \alpha\frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2} & \text{if } \alpha = 0 \text{ or } \beta = 0 \\ -\alpha\frac{\hat{\sigma}_{-X,Y}}{\hat{\sigma}_X^2} & \text{if } \frac{\hat{\sigma}_{-X,Y}^2}{\hat{\sigma}_{X,Y}^2} \geq \frac{2\beta-\beta^2}{2\alpha-\alpha^2} \quad \text{and} \quad \alpha \cdot \beta \neq 0 \\ \beta\frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2} & \text{if } \frac{\hat{\sigma}_{-X,Y}^2}{\hat{\sigma}_{X,Y}^2} \leq \frac{2\beta-\beta^2}{2\alpha-\alpha^2} \quad \text{and} \quad \alpha \cdot \beta \neq 0 \end{cases}$$

where

$$\alpha = \begin{cases} 0 & \text{if } \hat{\sigma}_{-X,Y} \leq 0 \\ \min\left\{1, \frac{a^0}{\hat{\sigma}_{-X,Y}/\hat{\sigma}_X^2}\right\} & \text{if } \hat{\sigma}_{-X,Y} > 0 \end{cases}$$

$$\beta = \begin{cases} 0 & \text{if } \hat{\sigma}_{X,Y} \leq 0 \\ \min\left\{1, \frac{b^0}{\hat{\sigma}_{X,Y}/\hat{\sigma}_X^2}\right\} & \text{if } \hat{\sigma}_{X,Y} > 0 \end{cases}$$

Proof The linearity of the inner product and the properties of the support function allow us to express $\phi(a)$ as

$$\phi(a) = \frac{1}{n}\sum_{i=1}^{n} < s_{Y_i} - s_{\overline{Y}}, s_{Y_i} - s_{\overline{Y}} >_K + \frac{1}{n}\sum_{i=1}^{n} < s_{aX_i} - s_{a\overline{X}}, s_{aX_i} - s_{a\overline{X}} >_K$$
$$-2\frac{1}{n}\sum_{i=1}^{n} < s_{Y_i} - s_{\overline{Y}}, s_{aX_i} - s_{a\overline{X}} >_K .$$

Since the homogeneity property of the support function w.r.t. the multiplication by a scalar is only satisfied for positive constants, positive and negative values of a will be considered separately. For the sake of simplicity, we will present our reasoning in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, and define $a^0 = b^0 = \infty$ if $A = \mathbb{R}$.

First, if $0 \leq a \leq b^0$, we have that

$$\phi(a) = \hat{\sigma}_Y^2 + a^2 \hat{\sigma}_X^2 - 2a\hat{\sigma}_{X,Y}.$$

This function is continuous, differentiable and convex in a, and it is therefore easy to check that the minimum is attained at $a_1 = \beta \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2}$. Analogously, if $-a^0 \leq a \leq 0$, we have that

$$\phi(a) = \hat{\sigma}_Y^2 + a^2 \hat{\sigma}_{-X}^2 + 2a\hat{\sigma}_{-X,Y}.$$

Since the kernel K in the distance D_K is assumed to be symmetric, we can easily show that $\hat{\sigma}_{-X}^2 = \hat{\sigma}_X^2$ (nevertheless, there is no *general* relationship between $\hat{\sigma}_{-X,Y}$ and $\hat{\sigma}_{X,Y}$.) Thus, we obtain that the minimum of $\phi(a)$ in this case is attained at $a_2 = -\alpha \frac{\hat{\sigma}_{-X,Y}}{\hat{\sigma}_X^2}$. Simple computations allow us to express the different cases that arise in terms of the sample covariances to obtain that the minimum of $\phi(a)$ over $[-a^0, b^0]$ is attained at

$$a^* = \begin{cases} \beta \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2} - \alpha \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2} & \text{if } \alpha = 0 \text{ or } \beta = 0 \\ -\alpha \frac{\hat{\sigma}_{-X,Y}}{\hat{\sigma}_X^2} & \text{if } \frac{\hat{\sigma}_{-X,Y}^2}{\hat{\sigma}_{X,Y}^2} \geq \frac{2\beta - \beta^2}{2\alpha - \alpha^2} \quad \text{and} \quad \alpha \cdot \beta \neq 0 \\ \beta \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2} & \text{if } \frac{\hat{\sigma}_{-X,Y}^2}{\hat{\sigma}_{X,Y}^2} \leq \frac{2\beta - \beta^2}{2\alpha - \alpha^2} \quad \text{and} \quad \alpha \cdot \beta \neq 0 \end{cases}$$

□

Remark 3 The solution of the minimization problem is unique, but in the case where $(\hat{\sigma}_{-X,Y}^2)/(\hat{\sigma}_{X,Y}^2) = (2\beta - \beta^2)/(2\alpha - \alpha^2)$ and $\alpha \cdot \beta \neq 0$, when a double solution for a can be obtained (one being positive and the other one negative).

□

Remark 4 If A is symmetric, that is, $A = [-a^0, a^0]$ (which happens, for instance when the sampled sets $\{y_i, x_i\}_{i=1}^n$ are all of them either intervals, or rectangles parallel to the Cartesian axes), then the expression of a^* is simply given by,

$$a^* = \begin{cases} 0 & \text{if } \hat{\sigma}_{-X,Y} \leq 0 \text{ and } \hat{\sigma}_{X,Y} \leq 0 \\ \min\left\{a^0, \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2}\right\} & \text{if } \hat{\sigma}_{X,Y} > 0 \text{ and } \hat{\sigma}_{-X,Y} \leq \hat{\sigma}_{X,Y} \\ -\min\left\{a^0, \frac{\hat{\sigma}_{-X,Y}}{\hat{\sigma}_X^2}\right\} & \text{if } \hat{\sigma}_{-X,Y} > 0 \text{ and } \hat{\sigma}_{X,Y} \leq \hat{\sigma}_{-X,Y} \end{cases}$$

□

Table 2 Sample from a 2D random rectangle

$\text{mid}_1 Y$	$\text{mid}_2 Y$	$\text{spr}_1 Y$	$\text{spr}_2 Y$	$\text{mid}_1 X$	$\text{mid}_2 X$	$\text{spr}_1 X$	$\text{spr}_2 X$
1.0	2.0	1.0	1.0	0.0	1.0	1.0	2.0
0.0	2.0	0.5	0.5	1.0	1.0	0.5	1.0
2.0	2.0	0.5	0.5	1.0	1.0	0.5	1.0
1.0	2.0	1.0	1.0	2.0	1.0	1.0	2.0

Although a double solution is uncommon in practice, a situation in which this happens is illustrated by the following Example 2. Furthermore, after a slight change in a single data value, a second situation with a unique solution will result.

Example 2 Table 2 displays a sample from a 2D random rectangle parallel to the Cartesian axes $(X, Y) : \Omega \to \mathcal{K}_c(\mathbb{R}^2) \times \mathcal{K}_c(\mathbb{R}^2)$.

The least squares estimates for the model $Y = aX + \epsilon_X$, where $a \in \mathbb{R}$ and ϵ_X is a random set with a fixed, but unknown $E^A[\epsilon_X] = B \in \mathcal{K}_c(\mathbb{R}^2)$, can be computed as it is shown in Theorems 1 and 2. Since (X, Y) are random rectangles parallel to the Cartesian axes, properties of the Aumann expectation allow to guarantee that B must be also a rectangle parallel to the Cartesian axes [that is, B will be determined by $(\text{mid}_1 B, \text{mid}_2 B, \text{spr}_1 B, \text{spr}_2 B)$]. On the other hand, we can verify that $A = [-1, 1]$, that is, A is symmetric and $a^0 = 1$. Therefore the solution of the minimization problem is given by the expression in Remark 4. In this case $\hat{\sigma}_{-X,Y} = \hat{\sigma}_{X,Y} = 0.153 > 0$, whence a double solution for a and B is obtained, which yield two alternative regression functions $\hat{y} = 0.32x + (0.68, 1.68, 0.51, 0.28)$ and $\hat{y} = -0.32x + (1.32, 2.32, 0.51, 0.28)$, respectively. We observe that $\hat{a}$ is an interior point of A.

Consider now $x_1' = (2, 2, 1, 2)$ as a new data interval replacing $x_1 = (0, 1, 1, 2)$ in Table 2. Since only the center of x_1 has changed, the threshold a^0 is still 1. However, $\hat{\sigma}_{X',Y} = 0.278$ and $\hat{\sigma}_{-X',Y} = 0.028$, and then the solution is unique. Concretely, it is given by $\hat{y} = 0.41x' + (0.48, 1.48, 0.13, 0.41)$. The estimate $\hat{a}'$ is also an interior point of A. □

5 Estimation of the multiple linear regression model

In this section, we consider a Multiple Linear Regression Model and present a stepwise method for estimation. As an extension of the Simple Linear Regression Model presented in Sect. 3, we consider the next probabilistic model,

Multiple linear regression model $(X_1, \ldots, X_k, Y) : \Omega \to [\mathcal{K}_c(\mathbb{R}^p)]^{k+1}$

$$Y = \beta_1 X_1 + \ldots \beta_k X_k + \epsilon_{X_1 \ldots, X_k} = \beta^t \overrightarrow{X} + \epsilon_{\overrightarrow{X}},$$

where $\boldsymbol{\beta}^t = (\beta_1, \ldots, \beta_k)$, $\overrightarrow{X}^t = (X_1, \ldots, X_k)$, $\beta_j \in \mathbb{R}$ for all $j \in \{1 \ldots k\}$ and $\epsilon_{\overrightarrow{X}}$ is a random set with expected value $E^A[\epsilon_{\overrightarrow{X}}] = B \in \mathcal{K}_c(\mathbb{R}^p)$. This implies that

$E^A[Y|\overrightarrow{x}] = \beta_1 x_1 + \cdots + \beta_k x_k + B$ for any sample realization $\overrightarrow{x} \in [\mathcal{K}_c(\mathbb{R}^p)]^k$ of $\overrightarrow{X}$.

The aim is to estimate the unknown parameters of the Multiple Linear Regression Model on the basis of a random sample $\{\overrightarrow{X}_i, Y_i\}_{i=1}^n$ from $(\overrightarrow{X}^t, Y)$. Hence, the problem is to estimate $\boldsymbol{\beta} \in \mathbb{R}^k$ and $B \in \mathcal{K}_c(\mathbb{R}^p)$ so that $\{(\overrightarrow{X}_i, \widehat{\boldsymbol{\beta}}^t \overrightarrow{X}_i + \widehat{B})\}_{i=1}^n$ is as close to the data matrix $\{\overrightarrow{X}_i, Y_i\}_{i=1}^n$ as possible subject to the constraint $Y_i = \widehat{\boldsymbol{\beta}}^t \overrightarrow{X}_i + \epsilon_i$ for some $\epsilon_i \in [\mathcal{K}_c(\mathbb{R}^p)]$ $(i \in \{1, \ldots, n\})$.

We propose to estimate stepwise the parameters $\beta_j \in \mathbb{R}$ $(j \in \{1, \ldots, k\})$ which are significantly different from zero, that is, to include in the estimate model the set-valued independent variables X_j which contribute new linear information to approximate the response Y. In order to verify whether or not a coefficient $\beta_j \in \mathbb{R}$ is significantly different from zero, we propose to use a linear independence bootstrap test. The test will be applied to *simple regressions models* relating a independent variable X_j and a given response. A way of applying this test in practice is presented in the following auxiliary algorithm for a *generic* simple linear regression model $Y = aX + \epsilon_X$.

Auxiliary algorithm:

Linear independence bootstrap test for a = 0

Step 1 Compute the value of the statistic $T = 1 - \hat{\sigma}^2_{Y -_H \hat{a}X} / \hat{\sigma}^2_Y$.

Step 2 Obtain $(X_1, \epsilon_1^*), \ldots, (X_n, \epsilon_n^*)$, where $\{\epsilon_i^*\}_{i=1}^n$ is sampled from $\{Y_1 -_H \hat{a}X_1, \ldots, Y_n -_H \hat{a}X_n\}$ and compute the value

$$T^* = 1 - \hat{\sigma}^2_{\epsilon^* -_H \hat{a}^* X} / \hat{\sigma}^2_{\epsilon^*}.$$

Step 3 Repeat Step 2 a large number b of times and approximate the p-value for β_j as the proportion of values in $\{T_1^*, \ldots, T_b^*\}$ greater than T.

The stepwise estimation algorithm should evaluate the *new* contribution of the available variables at each step, in order to include a new variable X_j in the model if the corresponding β_j is significatively different from zero. Concretely, since the estimators introduced in Sect. 4 for the simple linear regression model guarantee the existence of the Hukuhara differences, we can compute the residuals and check at each step if there is any variable that contributes with new linear information to these residuals. To measure the significance of the linear contribution we will use the p-values of the linear independence test obtained by applying the auxiliary algorithm. In this respect, if $(Z_1, \ldots Z_n)^t \in [\mathcal{K}_c(\mathbb{R}^p)]^n$ stands for any of the sample residuals, we will denote by $\theta_j(Z)$ the p-value of the test of the null hypothesis $H_0 : \beta_j = 0$ on the basis of the sample information $\{((X_j)_i, Z_i)\}_{i=1}^n$.

Stepwise algorithm:

Estimation of the multiple linear model

Step 1 Fix the significance level $\alpha \in (0, 1)$ and set $\hat{\boldsymbol{\beta}}^t = (0, \ldots, 0)^t \in \mathbb{R}^k$ and $J = \{1, \ldots k\}$;

Step 2 For each $j \in J$, compute $\theta_j(Y -_H \hat{\boldsymbol{\beta}}^t \overrightarrow{X})$ by means of the **Auxiliary Algorithm** and set
$j^* = \arg\min_{j \in J} \theta_j(Y -_H \hat{\boldsymbol{\beta}}^t \overrightarrow{X})$;

Step 3 IF $\theta_{j^*}(Y -_H \hat{\boldsymbol{\beta}}^t \overrightarrow{X}) < \alpha$, THEN
- Compute the estimate of the coefficient a in the simple linear regression model for $(X_{j^*}, Y -_H \hat{\boldsymbol{\beta}}^t \overrightarrow{X})$ (see Sect. 4);
- Set $\hat{\boldsymbol{\beta}}_{j^*} = a$ and $J = J \setminus \{j^*\}$
- GO TO Step 2

ELSE the estimate of $(\boldsymbol{\beta}^t, B)$ is $(\hat{\boldsymbol{\beta}}^t, \overline{Y -_H \hat{\boldsymbol{\beta}}^t \overrightarrow{X}})$ STOP.

The Stepwise Algorithm will be illustrated by means of a simulated situation in the following example.

Example 3 In order to compare the estimates obtained by applying the Stepwise Algorithm with the exact values in a concrete Multiple Linear Model, a sample of $n = 10$ data of $Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon$ has been obtained, where $\beta_1 = 1$, $\beta_2 = 3$, and $(X_1, X_2, \epsilon) : \Omega \to [\mathcal{K}_c(\mathbb{R}^2)]^3$ is a 3D random rectangle parallel to the Cartesian axes whose characterizing random vectors are distributed as follows:

- random centers: $\text{mid}_1 X_1$, $\text{mid}_2 X_1$, $\text{mid}_1 X_2$, $\text{mid}_2 X_2$, $\text{mid}_1 \epsilon$, $\text{mid}_2 \epsilon$ normally distributed as $\mathcal{N}(0,1)$.
- random spreads: $\text{spr}_1 X_1$, $\text{spr}_1 \epsilon$, $\text{spr}_2 \epsilon$ distributed as χ^2_1, $\text{spr}_2 X_1$, $\text{spr}_1 X_2$ distributed as χ^2_2, and $\text{spr}_2 X_2$ as χ^2_3.

The simulated data are gathered in Table 3. To estimate the model $Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon$, the Stepwise Algorithm has been applied. In Step 1 the significance level α has been fixed to be equal to 0.05. In the first iteration the independence bootstrap test in the Auxiliary Algorithm leads to a p-value equal to 0.008 for (X_1, Y), and equal to zero for (X_2, Y). Consequently, the first variable to be chosen is X_2. Since the corresponding p-value is lower than α, X_2 is included in the estimated model with a parameter $\hat{\beta}_2 = 2.944$. In the second iteration the linear independence bootstrap test in the Auxiliary Algorithm for $(X_1, Y -_H \hat{\beta}_2 X_2)$ leads to a p-value $= 0 < \alpha$, therefore X_1 is also included in the estimated model with $\hat{\beta}_1 = 1.062$. Thus the estimate $\hat{B} = \overline{Y -_H \hat{\boldsymbol{\beta}}^t \overrightarrow{X}}$ is the rectangle parallel to the Cartesian axes given by the vector $(\text{mid}_1 B, \text{mid}_2 B, \text{spr}_1 B, \text{spr}_2 B) = (0.201, 0.428, 1.270, 0.937)$.

The exact model was $E^A[Y|(x_1, x_2)] = x_1 + 3x_2 + (0, 0, 1, 1)$ and the estimate from 10 data $\hat{y} = 1.062x_1 + 2.99x_2 + (0.201, 0.428, 1.270, 0.937)$, which seems to indicate that the estimates for β_1, β_2 are quite good and the estimated B is reasonable. □

6 An illustrative real-life example with interval data

The following example illustrates the methods developed in this paper in a real-life case. Data have been supplied by the Department of Nephrology of

Table 3 $n = 10$ simulated triples of data intervals for the model $Y = a_1X_1 + a_2X_2 + \epsilon$

$\text{mid}_1 Y$	$\text{mid}_2 Y$	$\text{spr}_1 Y$	$\text{spr}_2 Y$	$\text{mid}_1 X_1$	$\text{mid}_2 X_1$	$\text{spr}_1 X_1$	$\text{spr}_2 X_1$	$\text{mid}_1 X_2$	$\text{mid}_2 X_2$	$\text{spr}_1 X_2$	$\text{spr}_2 X_2$
1.463	4.601	1.235	14.170	0.771	1.340	0.445	0.618	0.611	0.710	0.065	4.472
0.580	−2.916	2.060	23.470	−0.189	1.429	0.024	0.373	0.275	−1.423	0.599	7.698
1.716	4.991	8.401	11.759	0.927	0.073	0.653	0.810	−0.064	1.385	0.461	3.207
−5.735	1.302	5.470	11.685	−1.252	0.547	0.020	2.837	−1.857	−0.059	1.761	2.744
0.666	−0.372	1.864	17.100	1.151	−1.057	0.029	0.420	0.107	0.224	0.578	5.144
−2.067	−0.606	36.004	5.164	−0.267	−1.261	1.844	1.942	−0.475	0.151	10.898	1.001
11.916	−2.992	2.652	12.361	0.857	−0.972	0.001	5.512	3.869	−0.348	0.398	1.985
−1.984	−0.551	1.800	10.967	−0.985	1.526	0.442	0.815	−0.601	−0.693	0.452	2.454
3.659	−0.791	14.608	4.984	1.197	0.103	0.644	0.227	0.700	−0.459	4.361	1.428
−1.782	4.212	6.738	22.738	−0.818	1.864	1.429	3.507	−0.886	0.098	1.576	6.184

the Hospital Valle del Nalón in Langreo (Asturias, Spain) and has been previously considered in Gil et al. (2002, 2006) to illustrate different aspect regarding the linear regression between interval data.

Data in Table 4 correspond to the "range of the pulse rate over a day", X, the "range of systolic blood pressure over the same day", Y, and the "range of diastolic blood pressure over the same day", Z, observed in a sample of $n = 59$ patients (suffering from different types of illness) from a population of 3,000 patients who are hospitalized per year.

Values of X, Y and Z are obtained from several registers of the pulse rate, systolic blood pressure and diastolic blood pressure of each patient measured at different time points (usually 60–70) over a concrete day. Blood pressures and rate pulse data are often collected by taking into account simply the lowest and highest registers during a day (actually, some devices used for this purpose even record and memorize only these extreme values during a day); in these cases, the whole registers for a day and the associated variation can distort the information on the characteristic which is considered to be relevant: the range.

The aim is to estimate the relationship that relates the diastolic blood pressure Y as a linear function of the pulse rate X_1 and the systolic blood pressure X_2, that is, to estimate $\beta_1, \beta_2 \in \mathbb{R}$ and $B \in \mathcal{K}_c(\mathbb{R})$ in the model $Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon_{X_1,X_2}$ with $E^A([\epsilon_{X_1,X_2}] = B$.

In the first step of the Stepwise Algorithm, the significance level has been fixed again to be $\alpha = 0.05$, in the first iteration the independence bootstrap test in the Auxiliary Algorithm leads to a p-value $= 0.115$ for (X_1, Y) and a p-value $= 0$ for (X_2, Y). Consequently, the first variable to be chosen is X_2. Since the corresponding p-value is lower than α, X_2 is included in the estimate model with a parameter $\hat{\beta}_2 = 0.0077$. In the second iteration the linear independence bootstrap test in the Auxiliary Algorithm for $(X_1, Y -_H \hat{\beta}_2 X_2)$ leads to a p-value $= 0.074 > \alpha$, therefore X_1 is not included in the estimated model. We find that $\hat{B} = \overline{Y -_H \hat{\boldsymbol{\beta}}^t \overrightarrow{X}}$ is the interval in $\mathbb{R}^1$ with center 2.049 and spread 0.978, that is $[1.071, 3.027]$. As a consequence, the estimated model is $\hat{y} = 0.077x_2 + [1.071, 3.027]$.

Table 4 Data on the ranges of pulse rate (X_1), systolic (X_2) and diastolic (Y) blood pressure for $n = 59$ patients

X	Y	Z	X	Y	Z
58–90	118–173	63–102	52–78	119–212	47–93
47–68	104–161	71–118	55–84	122–178	73–105
32–114	131–186	58–113	61–101	127–189	74–125
61–110	105–157	62–118	65–92	113–213	52–112
62–89	120–179	59–94	38–66	141–205	69–133
63–119	101–194	48–116	48–73	99–169	53–109
51–95	109–174	60–119	59–98	126–191	60–98
49–78	128–210	76–125	59–87	99–201	55–121
43–67	94–145	47–104	49–82	88–221	37–94
55–102	148–201	88–130	48–77	113–183	55–85
64–107	111–192	52–96	56–133	94–176	56–121
54–84	116–201	74–133	37–75	102–156	50–94
47–95	102–167	39–84	61–94	103–159	52–95
56–90	104–161	55–98	44–110	102–185	63–118
44–108	106–167	45–95	46–83	111–199	57–113
63–109	112–162	62–116	52–98	130–180	64–121
62–95	136–201	67–122	56–84	103–161	55–97
48–107	90–177	52–104	54–92	125–192	59–101
26–109	116–168	58–109	53–120	97–182	54–104
61–108	98–157	50–111	49–88	124–226	57–101
54–78	98–160	47–108	75–124	120–180	59–90
53–103	97–154	60–107	58–99	100–161	54–104
47–86	87–150	47–86	59–78	159–214	99–127
70–132	141–256	77–158	55–89	138–221	70–118
63–115	108–147	62–107	55–80	87–152	50–95
47–83	115–196	65–117	70–105	120–188	53–105
56–103	99–172	42–86	40–80	95–166	54–100
71–121	113–176	57–95	56–97	92–173	45–107
68–91	114–186	46–103	37–86	83–140	45–91
62–100	145–210	100–136			

7 Concluding remarks and open problems

In this paper least squares estimators for linear regression models between convex compact random sets are found, on the basis of a set-arithmetic approach. The statistical properties of these estimators and of the corresponding goodness-of-fit measures (such as consistency, bias, asymptotic distributions, and so on) are still to be derived. The results generalize those for interval-valued random sets and insofar also provide a contribution to Symbolic Data Analysis (see, for instance, Bock and Diday 2000; Billard and Diday 2003).

References

Aumann RJ (1965) Integrals of set-valued functions. J Math Anal Appl 12:1–12

Bertoluzza C, Corral N, Salas A (1995) On a new class of distances between fuzzy numbers. Mathw Soft Comput 2:71–84

Billard L, Diday E (2003) From the statistics of data to the statistics of knowledge: symbolic data analysis. J Am Stat Assoc 98:470–487

Bock HH, Diday E (2000) Analysis of symbolic data. Exploratory methods for extracting statistical information from complex data. Studies in classification, data analysis, and knowledge organization. Springer, Heidelberg

Cressie NAC (1993) Statistics for spatial data. Wiley, New York

De Carvalho FAT, Lima Neto EA, Tenorio CP (2004) A new method to fit a linear regression model for interval-valued data. In: Biundo S, Frühwirth T, Palm G (eds) Lectures Notes on artificial intelligence no. 3238. Proceedings of the 27th German conference on artificial intelligence (KI 2004, Ulm, Germany). Springer, Heidelberg, 295–306

Diamond P (1990) Least squares fitting of compact set-valued data. J Math Anal Appl 14:531–544

Diamond P, Kloeden P (1994) Metric spaces of fuzzy sets. World Scientific, Singapore

Gil MA, López-García MT, Lubiano MA, Montenegro M (2001) Regression and correlation analyses of a linear relation between random intervals. Test 10:183–201

Gil MA, Lubiano MA, Montenegro M, López-García MT (2002) Least squares fitting of an affine function and strength of association for interval-valued data. Metrika 56:97–111

Gil MA, González-Rodríguez G, Colubi A, Montenegro M (2006) Testing linear independence in linear models with interval-valued data. Comput Stat Data Anal (in press, doi:10.1016/j.csda.2006.01.015)

González-Rodríguez G, Colubi A, Coppi R, Giordani P (2006) On the estimation of linear models with interval-valued data. 17th conference of IASC-ERS (COMPSTAT'2006), Physica-Verlag, Heidelberg, pp 697–704

Körner R, Näther W (2002) On the variance of random fuzzy variables. In: Bertoluzza C, Gil MA, Ralescu DA (eds) Statistical modeling, analysis and management of fuzzy data. Physica-Verlag, Heidelberg, pp 22–39

Lima Neto EA, De Carvalho FAT, Tenorio CP (2004) Univariate and multivariate linear regression methods to predict interval-valued features. In: Webb GI, Xinghuo (eds) Lecture Notes on artificial intelligence no. 3339. Proceedings of 17th Australian joint conference on artificial intelligence (AI 2004, Cairns, Australia). Springer, Heidelberg, pp 526–537

Lubiano MA, Gil MA, López-Díaz M, López-García MT (2000) The $\overrightarrow{\lambda}$-mean squared dispersion associated with a fuzzy random variable. Fuzzy Sets Syst 111:307–317

Matheron G (1975) Random sets and integral geometry. Wiley, New York

Molchanov I (2005) Probability and its applications. Springer, London

Stoyan D, Kendall WS, Mecke J (1987) Stochastic geometry and its applications. Wiley, Chichester