SAPIENZA
UNIVERSITÀ DI ROMA

Dottorato di Ricerca in Statistica Metodologica

Tesi di Dottorato XXI Ciclo – 2005/2008

Dipartimento di Statistica, Probabilità e Statistiche Applicate

A linear regression model for LR fuzzy random variables: properties and inferential procedures

Maria Brigida Ferraro

Questo lavoro è il frutto del triennio del Dottorato di Ricerca in Statistica Metodologica presso il Dipartimento di Statistica, Probabilità e Statistiche Applicate, Sapienza Università di Roma. Ringrazio il Dipartimento e i due coordinatori del Dottorato, il Prof. Renato Coppi e il Prof. Fulvio De Santis.

Desidero ringraziare il Prof. Renato Coppi anche in qualità di supervisore di questa tesi ma soprattutto per avermi incoraggiato a intraprendere questa strada e per i suoi consigli e suggerimenti.

Un ringraziamento speciale va all'altro supervisore di questa tesi, la Dott.ssa Ana Colubi del "Departamento de Estadística e I.O. y Didáctica de la Matemática" dell'Università di Oviedo (Spagna), per avermi seguito in tutte le fasi di questo lavoro nel periodo Febbraio 2007 - Luglio 2008 trascorso presso il suo Dipartimento. Desidero ringraziarla per i suoi insegnamenti e per avermi sostenuto nei momenti più difficili.

Ringrazio l'European Centre for Soft Computing di Mieres (Spagna) per avermi ospitato nel periodo trascorso in Spagna e per gli strumenti messi a mia disposizione. In modo particolare desidero ringraziare il Dott. Gil González Rodrǵuez dell'unità "Intelligent Data Analysis and Graphical Models" per avermi aiutato nella realizzazione di questa tesi, per la sua infinita disponibilità e per il suo ottimismo.

Ringrazio tutto il "Departamento de Estadística e I.O. y Didáctica de la Matemática" dell'Università di Oviedo (Spagna), tutti i membri del gruppo SMIRE (Statistical Methods with Imprecise Random Elements), in particolare la Prof.ssa M. Angeles Gil per avermi dato la possibilità di un contratto di ricerca finanziato dal "Principado de Asturias" e per avermi accolto in maniera splendida.

Un ringraziamento speciale va a Paolo Giordani per i suoi consigli e per la sua amicizia, a Stefania, Serena e Alessia ("Quelle del XXI ciclo") per essere state sempre al mio fianco anche quando ero lontano, ad Ana Belén e Angela per avermi fatto sempre sentire una di loro e a tutti "Quelli della stanza 41".

Infine ringrazio la mia famiglia per non aver mai smesso di credere in me e per avermi sempre lasciato libera di fare le mie scelte.

Prologue

Different sources of uncertainty can affect statistical reasoning: randomness, imprecision, vagueness, partial ignorance, etc. In particular, in regression analysis the uncertainty is about: the relationship between response and explanatory variables; the randomness due to the data generation process; the imprecision of the observed values of the variables (see Coppi, 2008). In this work these three kinds of uncertainty are taken into account.

The classical techniques manage only the first two types of uncertainty (see, for instance, Casella & Berger, 2002), while recently the third one has started to be considered due to a practical demand. Actually, in several practical applications in public health, medical science, ecology, agriculture or economic problems, many useful variables are vague, and the researchers find it easier to capture the vagueness through more complex data than to discard the vagueness and obtain precise data. In addition it is often less expensive to obtain an imprecise observation than to look for precise measurements of the variable of interest (see, for instance, Heagerty & Lele, 1998).

In order to handle a typical kind of imprecision the so-called LR fuzzy sets are often used. Formally, they are a type of functional data determined by three values: the center, the left spread and the right spread. For example, in agriculture, quantitative soil data are unavailable over vast areas and imprecise measures, that can be modeled through LR fuzzy sets, are used (see Lagacherie *et al.*, 2000). Also in medical science symptoms, diagnosis and phenomena of disease may often lead to LR data (see, for instance, Di Lascio *et al.*, 2002). LR-type functional data may also arise in other contexts, like image processing or artificial intelligence (see, for instance, Sezgin & Sankur, 2004, Ranilla & Rodríguez-Muñiz, 2007).

In addition the LR fuzzy sets are a generalization of the intervals which are useful in many other contexts. For instance, epidemiological research often entails the analysis of failure times subject to grouping, and the analysis with interval-grouped data is numerically simple and statistically meaningful (see Pipper & Ritz, 2007, Gil *et al.*, 2007, Billard & Diday, 2003).

In the context of random experiments whose outcomes are not numbers (or vec-

tors in $\mathbb{R}^p$) but they are expressed in inexact terms, the concept of fuzzy random variable (FRV) arises. Kwakernaak (1978), Puri & Ralescu (1986) and Klement *et al.* (1986) have introduced the concept of FRV as an extension of both, random variables and random sets. In the first case (Kwakernaak, 1978) it is considered a random variable that can be perceived through a set of windows W_i to which each result can belong or not. The underlying crisp variable is called original (see, for instance, Kruse, 1982 and 1987). Some years later Puri & Ralescu (1986) have defined the concept of fuzzy random variable as an extension of random sets to handle random experiments whose results are purely fuzzy values. That is, the values are directly observed as fuzzy sets and there is not necessarily an underlying real-valued random variable imprecisely observed (see, for instance, Colubi, 2009).

Several regression studies involving fuzzy sets to model imprecise data have been developed (see, for instance, Diamond, 1988, Diamond & Körner, 1997, Körner & Näther, 1998, Wünsche & Näther, 2002, Krätschmer, 2004, Krätschmer, 2006, Näther, 2006, Coppi & D'Urso, 2003, D'Urso, 2003, Coppi *et al.*, 2006, González-Rodríguez *et al.*, 2009). In details, Diamond's model is one of the first regression studies based on the least squares approach, from which some works have taken inspiration. Among these, Körner & Näther (1998), Wünsche & Näther (2002), Krätschmer (2004, 2006), Näther (2006) and González-Rodríguez *et al.* (2009) have used a formalization by means of fuzzy random variables, but only in the last one a complete solution for the estimation problem has been obtained.

Coppi *et al.* (2006) have proposed a linear regression model with LR fuzzy response. The basic idea consists in modeling the centers of the response variable by means of a classical regression model, and simultaneously modeling the left and the right spread of the response through simple linear regressions on its estimated centers. In the study in Coppi *et al.* (2006) the authors impose a non-negativity condition to the numerical minimization problem to avoid negative estimated spreads. Unlike the previous models it has not been formalized through fuzzy random variables, and to look for this kind of formalization the model proposed in this work comes up. Furthermore the aim is not only the estimation, but also the analysis of the statistical properties of the estimators (consistency, unbiasedness), besides the construction of confidence intervals and of procedures for testing hypotheses on the regression parameters. Then we propose an alternative model to overcome the non-negativity condition, because the inferences for models with non-negativity restrictions are usually more complex and less efficient (see, for instance, Liew, 1976 and Gallant & Gerig, 1980). In order to avoid the non-negativity condition, appropriate transformations of the spreads of the response are introduced.

The work is organized in four chapters. In the first chapter some preliminary elements are introduced. The basic concepts of fuzzy sets theory are given: the

definition of a fuzzy set and the LR characterization, the arithmetic of fuzzy data and some distances between fuzzy sets, paying special attention to the Yang and Ko distance, D^2_{LR}. The concept of a fuzzy random variable, according to Puri & Ralescu (1986), and its population moments are defined. In the last part of the chapter some previous linear regression models with fuzzy data are briefly analyzed, in particular, the models introduced by Diamond (1988), González-Rodríguez *et al.*, (2009) and by Coppi *et al.* (2006) are reviewed.

Chapter 2 deals with a generalization of the Yang and Ko distance to $\mathbb{R}^3$, $D^2_{\lambda\rho}$, the correspondent scalar product $\langle\cdot,\cdot\rangle_{\lambda\rho}$ and a new definition of variance. It is proved that the space of LR fuzzy numbers is isometric to a closed convex cone of $\mathbb{R}^3$, endowed with the inner product $\langle\cdot,\cdot\rangle_{\lambda\rho}$. The concept of variance for fuzzy random variables based on D^2_{LR} is given, following ideas in Körner (1997) and Lubiano *et al.* (2000) for other metrics. Some properties of the variance are proved, in particular it is shown that it verifies the *Frèchet principle*. This is a necessary condition to employ appropriately the least squares criterion. Furthermore the covariance between two LR fuzzy variables is also defined. In the second part the estimation problem is discussed. The estimators are unbiased and strongly consistent. In order to illustrate the consistency of the estimators some simulation studies are presented and some empirical examples are given.

Chapter 3 contains the new linear regression model with LR fuzzy response and scalar predictors. It is formally described and the theoretical values of the parameters are expressed in terms of moments as usual. In order to measure the goodness-of-fit of the model, a determination coefficient is given. The main part of this chapter is focused on statistical inferences, in particular the estimation problem and hypothesis testing. By means of the least squares criterion, the estimators of the regression parameters are obtained. Their statistical properties are examined and the corresponding asymptotic distributions are established.

The absence of realistic parametric models for fuzzy random variables makes no sense to look for exact distributions for specific models as in the classical case (for instance, for the exponential family). Thus non-parametric techniques are employed.

In order to analyze the accuracy of the estimators, a bootstrap procedure is given. The results of simulation studies and real life applications are evaluated. To complete the statistical inferences on the regression parameters confidence intervals and hypothesis testing are defined and discussed.

As for the least squares estimators, some statistical properties and the asymptotic distribution of the estimator of the determination coefficient are analyzed. Then the linear independence test is given by means of the asymptotic approach and the bootstrap one. In addition, simulation studies are discussed to illustrate the empirical significance of the proposed test. The behaviour of the asymptotic test under local

alternatives (power analysis) is shown to be the expected one in linear regression models.

Chapter 4 is a generalization of Chapter 3. A multiple linear regression model with imprecise response is discussed. This model is formally different, due to the matrix notation which simplifies the extension of the results of the simple case. Only a brief outline of the procedure is described, due to the analogy with the previous chapter. Simulations and empirical results are presented in order to clarify the efficiency of the models.

Each chapter is closed by a final evaluation about its contributions and some open problems.

The last chapter is the epilogue. It contains concluding remarks and some future directions.

Contents

List of Figures

List of Tables

Chapter 1

Preliminaries

The present work is centered on the analysis of random experiments modeled by means of a probabilistic space $(\Omega, \mathcal{A}, P)$, for which the characteristic X observed on each $\omega \in \Omega$ is not precise and can be described using fuzzy sets. In this chapter some basic notions of fuzzy sets theory are given. It is organized in the following way. In the next section the definition of a fuzzy set is given and the LR subclass is discussed. Section 1.1.2 deals with the arithmetic of fuzzy data and in Section 1.2.2 some distances between fuzzy sets are introduced. In particular the interest is focused on the Yang and Ko distance, D^2_{LR}. In Section 1.2 the concept of a fuzzy random variable, according to Kwakernaak (1978) and to Puri & Ralescu (1986), is given. For what follows in the next chapters it is useful to define the expectation value and the variance of a fuzzy random variable (Section 1.2.1 and Section 1.2.4, respectively). A brief description of some previous linear regression models with fuzzy data is in Section 1.4. The first one has been introduced by Diamond (1988) and the second one by González-Rodríguez *et al.* (2009). The latter is the linear regression model with LR fuzzy response, introduced by Coppi *et al.* (2006), from which this work takes inspiration. Finally the reason to introduce a new regression model is discussed and some concluding remarks are presented (Section 1.5).

1.1 Description of the data: Fuzzy sets

In many practical situations there are some concepts that are vague, or imprecise. A classical set can not correctly represent these concepts. In the classical theory an element either belongs to a given set or it does not belong. In fact, each classical set A is represented by means of a characteristic function $c_A : X \to \{0, 1\}$, which associates with each $x \in X$ a number $c_A(x) = 1$ if x belongs to A and $c_A(x) = 0$ if

x does not belong to A. But for vague concepts this kind of representation is too rigid. To overcome this problem Lotfi A. Zadeh (1965) has introduced the fuzzy sets theory. The notion of fuzzy set is an extension of the classical one.

Each element can belong to a given set with a membership degree. The fuzzy set A of $\mathbb{R}^p$ is identified by 'a membership function' $A(x)$, i.e a mapping $A : \mathbb{R}^p \to [0, 1]$ so that $A(x)$ is the membership degree of x to the fuzzy set A.

For example, the concept of "tall people" is imprecise, it can not be represented by a single value. In a classical framework it can be forcedly represented by means of a classical set "people taller than 180 cm" (see Fig. 1.1 (a)). If John's height is 179.9 cm it means that John is not tall. It is obviously an artificial representation. To avoid this inconvenient the set of tall people can be considered by means of fuzzy sets (an example is shown in Fig. 1.1 (b)). It is evident that this is a more appropriate way to refer to a concept that is approximate.

Let $\mathcal{K}_c(\mathbb{R}^p)$ be the class of nonempty compact convex subsets of $\mathbb{R}^p$, the *class*

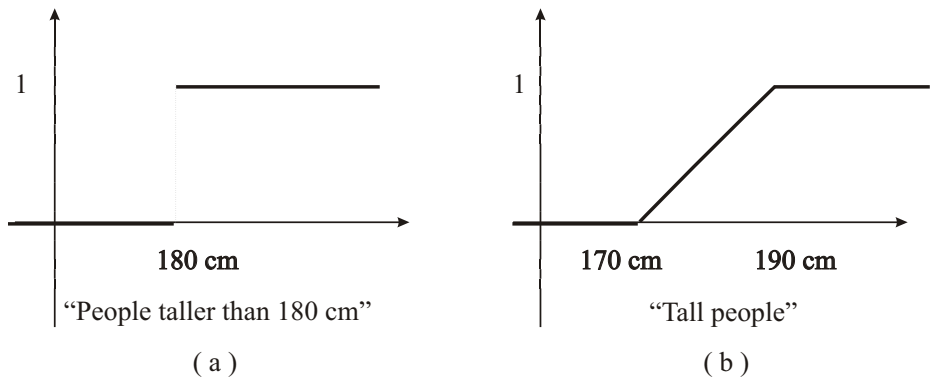

Figure 1.1: Representation of the set of tall people by means of a classical set (a) and a fuzzy set (b).

of fuzzy sets is $\mathcal{F}_c(\mathbb{R}^p) = \{A : \mathbb{R}^p \to [0, 1] | A_\alpha \in \mathcal{K}_c(\mathbb{R}^p)\}$, where A_α is the α-level of fuzzy set A, that is, $A_\alpha = \{x \in \mathbb{R}^p | A(x) \geq \alpha\}$, for $\alpha \in (0, 1]$, and $A_0 = cl(\{x \in \mathbb{R}^p | A(x) \geq 0\})$ (Zadeh, 1965).

When p is equal to 1 the compact convex sets will be intervals. An interval I can be characterized by means of the extremes $[\inf I, \sup I]$ or, alternatively, by means of the center $\operatorname{mid} \mathrm{I} = (\inf \mathrm{I} + \sup \mathrm{I})/2$ and the spread $\operatorname{spr} \mathrm{I} = (\sup \mathrm{I} - \inf \mathrm{I})/2$. In this case we will use the notation $[\operatorname{mid} \mathrm{I} \pm \operatorname{spr} \mathrm{I}]$. As a result we can represent each fuzzy datum $A \in \mathcal{F}_c(\mathbb{R})$ with the family of nested compact intervals $\{[\inf A_\alpha, \sup A_\alpha]\}_{\alpha \in [0,1]}$ or with the family $\{[\operatorname{mid} \mathrm{A}_\alpha \pm \operatorname{spr} \mathrm{A}_\alpha]\}_{\alpha \in [0,1]}$.

There are different kinds of fuzzy sets. In the next section a useful characteriza-

tion of a particular class is described.

1.1.1 LR fuzzy sets

A particular class of fuzzy sets very useful in practice is determined by 3 values: the center, the left spread and the right spread. This type of fuzzy datum is the LR *fuzzy number*. An LR fuzzy number A is characterized by the following membership function

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

where $A^m \in \mathbb{R}$ is the center, $A^l \in \mathbb{R}^+$ and $A^r \in \mathbb{R}^+$ are, respectively, the left and the right spread and, L and R are functions verifying the properties of the class of fuzzy sets $\mathcal{F}_c(\mathbb{R})$, such that $L(0) = R(0) = 1$ and $L(x) = R(x) = 0$, $\forall x \in \mathbb{R} \setminus [0, 1]$ (see Fig. 1.2). If $A^l = A^r$ the fuzzy number A is referred to as *symmetrical*.

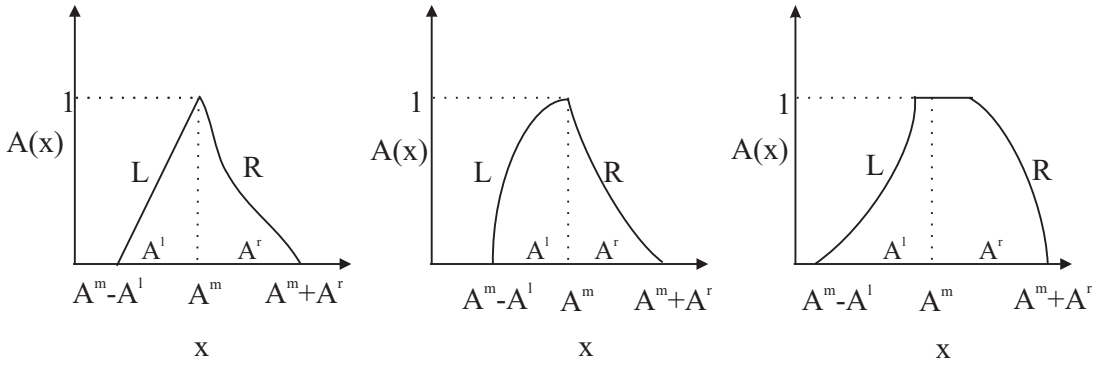

Figure 1.2: Examples of LR membership functions

The most used LR fuzzy numbers are the *triangular* ones, whose membership function is

$$T(x) = \begin{cases} 1 - \frac{T^m - x}{T^l} & T^m - T^l \leq x \leq T^m \\ 1 - \frac{x - T^m}{T^r} & T^m \leq x \leq T^m + T^r \\ 0 & \text{otherwise} \end{cases}$$

(see Fig. 1.3 (a)).

Remark 1.1.1 If the left and the right spread of a fuzzy number are null, the number is reduced to a classical one and it is referred to as *crisp*.

Remark 1.1.2 An interval I is a particular kind of LR fuzzy set where the membership function is the characteristic function 1_I, that is equal to 1, for all $x \in I$, and 0 otherwise (see Fig. 1.3 (b)).

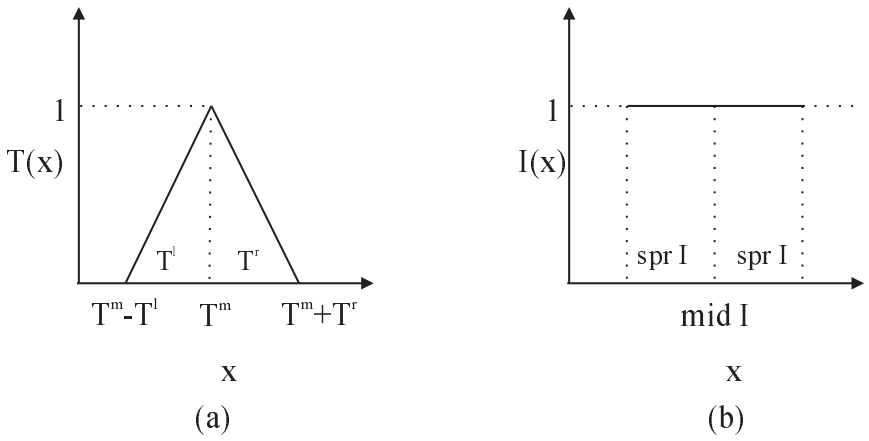

Figure 1.3: Functional representation of a triangular LR fuzzy number T (a) and an interval I (b).

LR fuzzy numbers are used in medical science (see, for instance, Di Lascio *et al.*, 2002), in agriculture (see Lagacherie *et al.*, 2000) or economic problems. This kind of LR-type functional data may also arise in other contexts, like image processing or artificial intelligence (see, for instance, Sezgin & Sankur 2004, Ranilla & Rodríguez-Muñiz 2007).

Example 1.1.1 (Di Lascio *et al.*, 2002) To analyze diabetic neuropathy, whose pathogenesis is not well-known, some patient's anagraphical and clinical data are considered. In particular, patient's hemoglicidic state, the amount of albumin in the urine, the values of systolic and diastolic pressure, the amount of insulin administrated to the patient, etc. are measured. The aim is to classify the patients on the basis of the severity of the symptoms. Each severity grade of symptoms can be represent by means of a label of linguistic variables. Thus, to model the uncertainty inherent to the clinical data and to represent the values of linguistic variables the class of LR fuzzy numbers has been used. In this way there is not relevant loss of information and the operations are very simple. For example, for the variable "diabetes age" the following labels are singled out: "very early", "early", "average", "late" (see Fig. 1.4). Each one is represented by means of an LR fuzzy number. The results obtained agree with most credited clinical analysis.

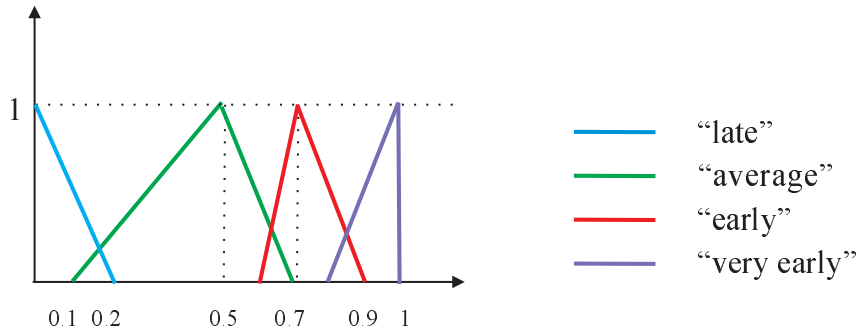

Figure 1.4: Representation of the labels singled out for the variable "diabetes age".

Example 1.1.2 (Lagacherie *et al.*, 2000) The estimation of crop yields is limited by the dimension of the areas, due to the difficulty of finding soil data on vast areas. So the spatial approach for the analysis takes into account imprecise soil data. Also in this case LR fuzzy numbers are used to model the imprecision.

1.1.2 Arithmetic of fuzzy data

We can define the sum and the product by scalars in the space $\mathcal{F}_c(\mathbb{R}^p)$ by means of Zadeh's extension principle

$$g(X_1, X_2, ..., X_p)(t) = \sup_{g(x_1,..,x_p)=t} \min\{X_1(x_1), ..., X_p(x_p)\}$$

which provides a general method for the extension of crisp continuous functions $g(x_1, ..., x_p)$ on $\mathbb{R}^p$ for fuzzy input X_1,...,X_p. If we consider $g(x_1, x_2) = x_1 + x_2$ and $g(x) = \lambda x$, for all λ in $\mathbb{R}^+$, these operations agree level-wise with the Minkowski sum and the product by a scalar on $\mathcal{K}_c(\mathbb{R}^p)$, for all $\alpha \in [0, 1]$, that is

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

whatever $A, B \in \mathcal{F}_c(\mathbb{R}^p)$ and $\lambda \in \mathbb{R}$.

Unfortunately neither $(\mathcal{K}_c(\mathbb{R}^p), +, \cdot)$ nor $(\mathcal{F}_c(\mathbb{R}^p), +, \cdot)$ are linear spaces because there is no inverse for the addition. For instance, $[1, 2] - [1, 2] = [-1, 1]$. For this reason we can use the *Hukuhara difference* $A -_H B$, which is defined (if it exists) as the element $C \in \mathcal{F}_c(\mathbb{R}^p)$ so that $A = B + C$. If A and B are in $\mathcal{K}_c(\mathbb{R})$ the difference $A -_H B = [\inf A - \inf B, \sup A - \sup B]$ exists if and only if $\operatorname{spr} \mathrm{B} \leq \operatorname{spr} \mathrm{A}$.

If A and $B \in \mathcal{F}_{LR}$, these operations can be alternatively determined by consid-

ering the fuzzy set $A + B$ in $\mathcal{F}_{LR}$ so that

$$\begin{cases} (A+B)^m = A^m + B^m \\ (A+B)^l = A^l + B^l \\ (A+B)^r = A^r + B^r \end{cases}$$

and γA, $\gamma \in \mathbb{R}$, is the fuzzy set so that

$$\begin{cases} (\gamma A)^m = \gamma A^m \\ (\gamma A)^l = \gamma A^l \\ (\gamma A)^r = \gamma A^r \end{cases} \text{if } \gamma \geq 0 \qquad \begin{cases} (\gamma A)^m = \gamma A^m \\ (\gamma A)^l = -\gamma A^l \\ (\gamma A)^r = -\gamma A^r \end{cases} \text{if } \gamma < 0$$

The space $(\mathcal{F}_{LR}, +, \cdot)$ is not closed because if $A \in \mathcal{F}_{LR}$, $-A \notin \mathcal{F}_{LR}$ except in the case of symmetrical LR fuzzy numbers. In this work this is not a problem, because only the product by positive scalars will be used. Anyway also in this case the Hukuhara difference $A -_H B$ can be introduced, and it exists if $A^l \geq B^l$ and $A^r \geq B^r$. It is given by

$$\begin{cases} (A -_H B)^m = A^m - B^m \\ (A -_H B)^l = A^l - B^l \\ (A -_H B)^r = A^r - B^r \end{cases}$$

These operations agree with the intuitive meaning of the sum and the product by positive scalars suitable for the kind of data that will be handled in this work. The imprecision is propagated by means of the Minkowski arithmetic, as shown in Example 1.1.3. Furthermore, to handle the above described situations it is natural to consider this arithmetic to define the average of imprecise values instead of other operations like the union or the intersection common in fuzzy logic.

Example 1.1.3 Consider the profit (X) of a company as interval-valued for each month in one year. Suppose that in April the company has gained a quantity of money (X_1) varying between 5000 and 6000 dollars and in May the profit (X_2) has varied from 4500 to 7000 dollars, that is, $X_1 = [5000, 6000]$ and $X_2 = [4500, 7000]$. If we are interested in considering the sum of the profit in both months, it is intuitive that the minimum is the sum of the minimum of each month and the maximum the sum of the maximum, that is, $X_1 + X_2 = [9500, 11000]$. If the profit of the same company in October (X_3) is the double of the profit in April, that is, $X_3 = 2X_1$, X_3 will, intuitively, vary from a minimum of 10000 dollars to a maximum of 12000.

1.2 Random models: Fuzzy random variables and characterization with LR

In practice there are random elements whose values are not numbers (or vectors in $\mathbb{R}^p$) but they are expressed in inexact terms. These random elements can be managed by means of the concept of fuzzy random variable (FRV).

Kwakernaak (1978), Puri & Ralescu (1986) and Klement et al. (1986) have introduced the concept of FRV from different points of view.

According to Kwakernaak (1978), a fuzzy random variable is an extension of a random variable. It is defined as follows: Let $(\Omega, \mathcal{A}, P)$ be a probability space and suppose that U is a random variable defined on this space. This random variable is perceived through a set of windows W_i, $i \in J$, with J a finite or countable set, each representing an interval in $\mathbb{R}$, such that $W_i \cap W_j = \emptyset$ for $i \neq j$, and $\bigcup_{i \in J} W_i = \mathbb{R}$. "Perceiving" the random variable through these windows means that for each ω we can only establish whether $U(\omega) \in W_i$, for some $i \in J$. Let the function $I_i : R \to [0,1]$ be the characteristic function of the set W_i and let S be the space of all piecewise continuous functions $C : \mathbb{R} \to [0,1]$. The perception of the random variable U is described through the mapping $X : \Omega \to S$, with $X(\omega) = I_i$ if and only if $U(\omega) \in W_i$. That is, not a real number $U(\omega)$ is associated with each $\omega \in \Omega$, as in the case of an ordinary random variable, but a characteristic function $X(\omega)$, which is an element of S. The above described map $X : \Omega \to S$ characterizes a special type of fuzzy random variable. The random variable U is an original of the perceived fuzzy random variable.

Remark 1.2.1 There may exist many originals corresponding to a given fuzzy variable.

In Kwakernaak (1978) a fuzzy random variable is defined as a map $\xi : S \to \mathcal{F}_c(\mathbb{R})$. The image of ω in $\mathcal{F}_c(\mathbb{R})$ under ξ is denoted as $\xi(\omega) = (\mathbb{R}, X_\omega, a_\omega)$, with $X_\omega \in S$ and $a_\omega : \mathbb{R} \to P$. The map $X : \Omega \to S$ has to fulfill some conditions, in particular, for each $\mu \in (0,1]$ both U_μ^* and U_μ^{**}, defined by

$$U_\mu^*(\omega) = \inf \{x \in \mathbb{R} | X_\omega(x) \geq \mu\}$$

and

$$U_\mu^{**}(\omega) = \sup \{x \in \mathbb{R} | X_\omega(x) \geq \mu\},$$

have to be finite real-valued random variables defined on $(\Omega, \mathcal{A}, P)$ satisfying, for each $\omega \in \Omega$, $X_\omega(U_\mu^*(\omega)) \geq \mu$ and $X_\omega(U_\mu^{**}(\omega)) \geq \mu$.

Finally, for each $\omega \in \Omega$ and each $x \in \mathbb{R}$, $a_\omega(x)$ is the statement

$$a_\omega(x) = (U \text{ takes on the value } x \text{ at the point } \omega)$$

where U is the original random variable of which ξ is a fuzzy perception.

The concept of FRV in Puri & Ralescu's sense arises to manage random experiments whose outcomes are not numbers but are expressed in inexact linguistic terms. A possible way of handling this kind of situations is by using the concepts of fuzzy sets and fuzzy functions found useful in many applications.

According to Puri and Ralescu (1986), an FRV is an extension of a random set. Let $(\Omega, \mathcal{A}, P)$ be a probability space, the mapping

$$X : \Omega \to \mathcal{F}_c(\mathbb{R}^p)$$

is an FRV on $\mathbb{R}^p$ if for all $\alpha \in [0,1]$ the $\alpha-$level mappings $X_\alpha : \Omega \to \mathcal{K}_c(\mathbb{R}^p)$, defined so that for all $\omega \in \Omega$

$$X_\alpha(\omega) = (X(\omega))_\alpha$$

are convex compact random sets.

The above definitions are carried out from different perspectives but they are formally the same. The first one considers an underlying original variable and the second one takes into account random variables whose values are purely fuzzy. Even if what will be introduced and analyzed in this work can be applied to both cases, the second definition is more appropriate, because the aim is handling FRVs themselves, not the underlying variables.

In the case of LR FRVs it is equivalent to require that

$$(X^m, X^l, X^r) : \Omega \to (\mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+)$$

be a random vector.

Example 1.2.1 An example of FRVs is introduced in Colubi (2009). In a recent study about the reforestation in a given area of Asturias (Spain), carried out in the INDUROT institute (University of Oviedo), the quality of the trees has been analyzed. This characteristic has not been assigned on the basis of an underlying real-valued magnitude, but rather on the basis of subjective judgements/perceptions, through the observation of the leaf structure, the root system, the relationship height/diameter, and so on. The experts used a fuzzy-valued scale to represent their perceptions, besides linguistic labels, because the usual categorical scale (very low, low, medium, high, very high) was not able to capture the perceptions. The considered support goes from 0 (absence of quality) to 100 (perfect quality). It is possible to have different values for the same linguistic label. Some fuzzy values are represented in Fig. 1.5. This variable has been observed on 238 trees. Thus $\Omega = \{$*sets of trees in a given area of Asturias*$\}$ endowed with the Borel σ-field. Since the observations were arbitrarily chosen, P is the uniform distribution

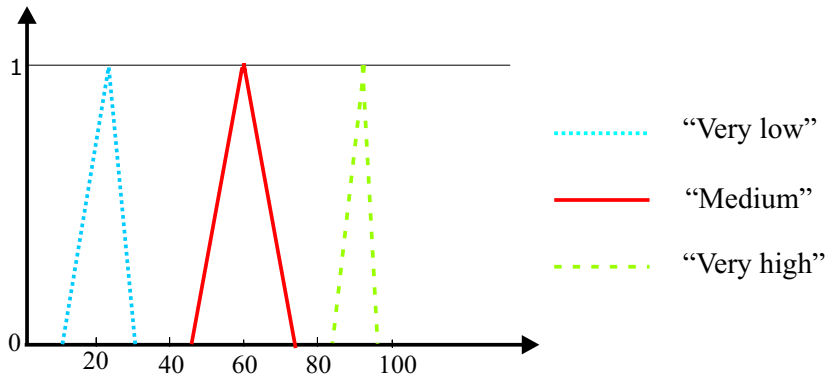

Figure 1.5: Values of the "quality" of three different trees

over Ω. For any $i \in \Omega$, several characteristics are to be observed. In particular, the quality, Y_i, has been considered as an LR triangular fuzzy variable ($\lambda = \rho = 1/2$) (see Table 1.1).

Table 1.1: Quality (Y^m, Y^l, Y^r), Height (X_1) and Diameter (X_2) of 238 trees in Asturias.

Y^m(center)	Y^l(left spread)	Y^r(right spread)	$X_1(cm)$	$X_2(cm)$
45	12.5	15	170	0.88
25	15	12.5	245	0.96
17.5	7.5	12.5	190	1.09
20	11.25	15	130	0.89
55	15	12.5	230	1.4
23.75	11.25	18.75	90	1.7
56.25	18.75	13.75	195	1.6
13.75	8.75	8.75	75	0.44
26.25	13.75	8.75	184	0.91
62.5	10	7.5	215	2.06
75	12.5	10	245	2.17
67.5	12.5	12.5	220	1.95
32.5	22.5	10	195	0.85
40	15	10	160	1.45
52.5	12.5	17.5	213	1.6
55	15	17.5	215	1.4
77.5	12.5	12.5	370	4
85	5	5	230	2.27
50	20	20	234	1.5
...	...	...	...	...

According to Puri & Ralescu (1986) an FRV X is *normal* if and only if

$$X = EX + \{\xi\},$$

where ξ is a classical Gaussian random vector on $\mathbb{R}^p$ with zero mean and $\{\xi\}$ is the fuzzy set with ξ as membership-one-singleton.

This is not a realistic model and, even if it is possible to make inference, it is useless because it does not model real elements (see Coppi, 2008).

There are not generalized models for FRVs that can be used in practice, for this reason in this work non-parametric techniques (asymptotic and bootstrap) will be employed.

1.2.1 Expected value and conditional expectation of an FRV

The expected value of an FRV is defined by means of the generalized Aumann integral (Aumann, 1965), that is the expected value of the FRV X is the unique fuzzy set $E(X)$ in $\mathcal{F}_c(\mathbb{R}^p)$, such that for all $\alpha \in [0,1]$,

$$(EX)_\alpha = \left\{ \int_\Omega f(\omega) dP(\omega) | f : \Omega \to \mathbb{R}, f \in L^1(\Omega, \mathcal{A}, P), f \in X_\alpha \, a.s.[P] \right\},$$

if $E\|X\| < \infty$ (Puri & Ralescu 1986), where $\|\cdot\|$ is the magnitude of a fuzzy set, defined as the distance from 0 (see Section 1.2.2). This expected value is coherent with the arithmetic used in this work, with respect to the strong law of large numbers. That is, if the sample mean is defined in terms of the Minkowski sum and product by a scalar, then the sample mean of FRVs independent and identically distributed converges almost surely to this expected value in terms of the strongest metrics (Colubi *et al.*, 1999).

In the $\mathcal{F}_c(\mathbb{R})$-valued case we have that

$$(EX)_\alpha = [E(\inf X_\alpha), E(\sup X_\alpha)],$$

for all $\alpha \in [0,1]$.

In case of LR FRVs, EX is the fuzzy set in $\mathcal{F}_{LR}$ whose center is EX^m and left and right spread, respectively, EX^l and EX^r.

Moreover Puri and Ralescu (1986) have introduced the concept of *the conditional expectation* of an FRV.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and Y an FRV on $\mathcal{F}_c(\mathbb{R}^p)$ with $E\|Y\| < \infty$. Consider a sub-σ-algebra $\mathcal{B} \subset \mathcal{A}$, the *conditional expectation* of Y with respect to $\mathcal{B}$ is the FRV $E(Y|\mathcal{B})$ such that $E(Y|\mathcal{B})$ is $\mathcal{B}$-measurable and for all $B \in \mathcal{B}$

$$\int_B E(Y|\mathcal{B}) dP = \int_B Y dP.$$

If $\mathcal{B} = \sigma(X)$ is induced by a further FRV X, it results

$$E(Y|\mathcal{B}) = E(Y|X).$$

1.2.2 Distances between fuzzy sets

To define a metric for the family $\mathcal{F}_c(\mathbb{R}^p)$, it is possible to consider a metric δ for the family $\mathcal{K}_c(\mathbb{R}^p)$, to apply it to the family of all corresponding α-cuts and to integrate with respect to α.

The best-known metric for compact convex sets A and B in $\mathbb{R}^p$ is the Hausdorff one, defined as

$$d_H(A,B) = \max\left\{\sup_{b\in B}\inf_{a\in A}\|a-b\|_p, \sup_{a\in A}\inf_{b\in B}\|a-b\|_p\right\}$$

where $\|\cdot\|_p$ denotes the usual Euclidean norm in $\mathbb{R}^p$.

This metric can be extended to the family $\mathcal{F}_c(\mathbb{R}^p)$, but it does not fulfill the *Frèchet principle* with respect to the Aumann expectation. So it is not useful in practice when we consider the least squares procedures (see Näther, 1997).

Example 1.2.2 (Näther, 1997) Let X be an interval-valued random set so that

$$\operatorname{mid} \mathrm{X} = \begin{cases} 2 & p_1 = \frac{2}{3} \\ 3 & p_2 = \frac{1}{3} \end{cases} \qquad \operatorname{spr} \mathrm{X} = \begin{cases} 1 & p_1 = \frac{2}{3} \\ 2 & p_2 = \frac{1}{3} \end{cases}$$

It is easy to check that the Aumann expectation is equal to $E(X) = [4/3 \pm 7/3]$, while if we consider the d_H metric the real interval U that minimize $E_{d_H}(X,U)$ is equal to $[2.2 \pm 1.2]$.

To overcome this problem an L^2-type metric can be employed.

Each compact convex set A in $\mathcal{K}_c(\mathbb{R}^p)$ can be represented by means of its support function

$$s_A(u) = \sup_{a\in A}\langle a,u\rangle, \qquad u \in \mathbb{S}^{p-1} \tag{1.1}$$

where $\langle\cdot,\cdot\rangle$ is the inner product in $\mathbb{R}^p$ and $\mathbb{S}^{p-1}$ is the $(p-1)$-dimensional unit sphere in $\mathbb{R}^{p-1}$. s_A uniquely determines A (see Diamond & Kloeden, 1994). The ρ^2 metric for A and B in $\mathcal{K}_c(\mathbb{R}^p)$ can be defined in terms of support functions as

$$\rho_2(A,B) = \left(p \int_{\mathbb{S}^{p-1}} (s_A(u) - s_B(u))^2 d\mu(u)\right)^{1/2}$$

where μ is the Lesbegue measure. For the intervals the sphere unit $\mathbb{S}^0 = \{-1, 1\}$ is taken into account, and it results $s_A(1) = \sup A$ and $s_A(-1) = -\inf A$, so

$$\rho_2(A, B) = \left(\frac{1}{2}(\sup A - \sup B)^2 + \frac{1}{2}(\inf A - \inf B)^2\right)^{1/2}.$$

By means of the family of α-cuts it is possible to extend the ρ_2 metric to the metric δ_2 in the space $\mathcal{F}_c(\mathbb{R}^p)$. Let X and Y be in $\mathcal{F}_c(\mathbb{R}^p)$, we obtain

$$\delta_2(X, Y) = \left(p \int_0^1 \int_{\mathbb{S}^{p-1}} (s_X(u, \alpha) - s_Y(u, \alpha))^2 d\mu(u) d\alpha\right)^{1/2},$$

where s is a mapping that generalize level-wise the support function (1.1) and it is defined (see Klement *et al.*, 1986) as

$$s : \mathcal{F}_c(\mathbb{R}^p) \to \mathcal{L}(\mathbb{S}^{p-1} \times [0, 1])$$

such that

$$s_A(a, \alpha) = \sup_{w \in A_\alpha} \langle a, w \rangle,$$

for any $a \in \mathbb{S}^{p-1}$ and $\alpha \in [0, 1]$, where $\mathcal{L}(\mathbb{S}^{p-1} \times [0, 1])$ is the class of the Lebesgue real-valued integrable function on $\mathbb{S}^{p-1} \times [0, 1]$. $\mathcal{F}_c(\mathbb{R}^p)$ can be isometrically embedded in a space of functions on $\mathbb{S}^{p-1} \times [0, 1]$ by means of the δ_2 metric through support functions (see Krätschmer 2002b).

This metric has interesting statistical properties, but as the Hausdorff distance it has some inconveniences from an intuitive point of view (see Bertoluzza *et al.*, 1995)

Example 1.2.3 (Bertoluzza *et al.*, 1995) Consider two pairs of intervals $A_1 = [0, 5]$, $B_1 = [6, 7]$ and $A_2 = [0, 5]$, $B_2 = [6, 10]$. It is easy to check that the Hausdorff metric assigns the same distance to A_1 and B_1 and to A_2 and B_2. But intuitively it seems that the distance between the second pair should be greater. If we consider now the pairs of intervals $C_1 = [-2, 2]$, $D_1 = [-1, 1]$ and $C_2 = [-2, 1]$, $D_2 = [-1, 2]$ the ρ^2 distance is the same for the two pairs. Also in this case it seems more intuitive to assign a greater value to the second one.

To avoid the inconveniences illustrated in Example 1.2.3 Bertoluzza *et al.* (1995) have introduced an L^2-type metric, taking into account a non-degenerate probability measure W and a weight measure φ. Let X and B in $\mathcal{F}_c(\mathbb{R})$, it is defined as

$$D_W^\varphi(A, B) = \left(\int_{[0,1]} \int [f_A(\alpha, \lambda) - f_B(\alpha, \lambda)]^2 dW(\lambda) d\varphi(\alpha)\right)^{\frac{1}{2}},$$

with $f_A(\alpha,\lambda) = \lambda \sup A_\alpha + (1-\lambda)\inf A_\alpha$ and W probability measure on the measurable space $([0,1],\mathcal{B}_{[0,1]})$.

The D_W^φ and δ_2 metric can be generalized by means of a family of metrics that depend on certain kernels (Korner & Nather 2002). The D_K-distance between A and $B \in \mathcal{F}_c(\mathbb{R}^p)$ is defined as

$$D_K(A,B) = \left(\int\limits_{(\mathbb{S}^{p-1})^2\times[0,1]^2} (s_A(u,\alpha) - s_B(u,\alpha))(s_A(v,\beta) - s_B(v,\beta)) dK(a,\alpha,b,\beta) \right)^{\frac{1}{2}},$$

where $K : \mathbb{S}^{p-1} \times \mathbb{S}^{p-1} \to \mathbb{R}$ is a certain symmetrical and positive kernel.

1.2.3 Yang and Ko distance between fuzzy sets

The previous distances are defined for general fuzzy sets in $\mathcal{F}_c(\mathbb{R}^p)$ and can be considerably simplified if we consider the particular case $\mathcal{F}_{LR}$. In addition, Yang and Ko (1996) have defined a distance D_{LR} between two LR fuzzy numbers A, $B \in \mathcal{F}_{LR}$ as follows

$$\begin{aligned} D_{LR}(A,B) = \big((A^m - B^m)^2 + ((A^m - \lambda A^l) - (B^m - \lambda B^l))^2 \\ +((A^m + \rho A^r) - (B^m + \rho B^r))^2\big)^{\frac{1}{2}}, \end{aligned} \tag{1.2}$$

where $\lambda = \int\limits_0^1 L^{-1}(\omega)d\omega$ and $\rho = \int\limits_0^1 R^{-1}(\omega)d\omega$ represent the influence of the shape of the membership function on the distance. In particular, λ (or ρ) less than 0.5 represents an imprecision decreasing rapidly; λ (or ρ) equal to 0.5 represents an imprecision decreasing linearly and λ (or ρ) greater than 0.5 represents an imprecision decreasing slowly. This is illustrated in the following example.

Example 1.2.4 Consider two LR fuzzy numbers $A = (2,1,1)$ and $B = (4,1.5,1.5)$, for different L_i, R_i functions, $i = 1,2,3$, (see Fig. 1.6). In the case $i = 1$ $\lambda = \rho = 0.2$, and $D^2_{L_1R_1}(A,B) = 4.02$; in the case $i = 2$ $\lambda = \rho = 0.5$ (triangular case) we get $D^2_{L_2R_2}(A,B) = 4.125$ and, finally, in the case $i = 3$ $\lambda = \rho = 0.8$, and $D^2_{L_3R_3}(A,B) = 4.32$. The obtained values show that as the imprecision increases, the value of the distance also increases.

1.2.4 Variance of an FRV

In literature there are different definitions of the variance of an FRV. Along this work a real-valued variance will be considered, because the aim is to use it for measuring

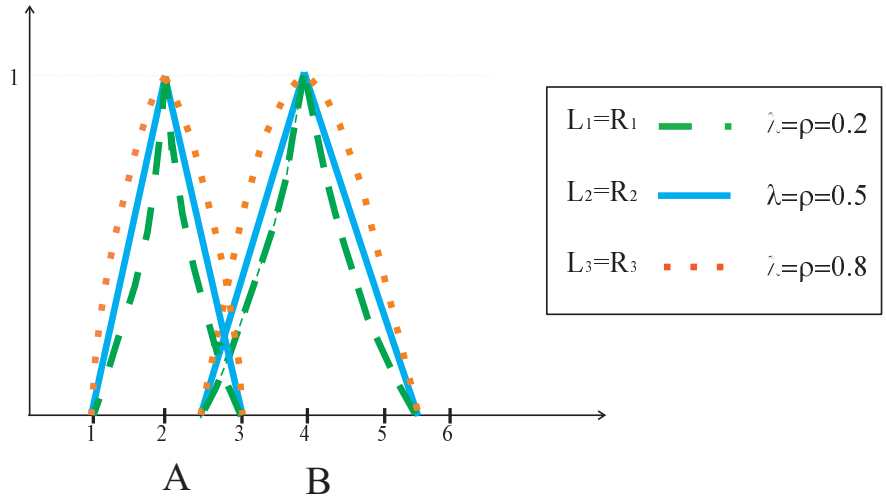

Figure 1.6: Examples of fuzzy numbers with different membership functions

the error due to the approximation of the values of a variable by means of the values predicted by the regression model.

Körner (1997) has defined the variance of an FRV X based on the distance δ_2 as

$$Var(X) = E\delta_2^2(X, EX),$$

if $E\|X\|_2^2 < \infty$.

Another definition of variance is given in Körner & Näther (2002). In details, by means of the D_K-distance, if $E\|X\|_K^2 < \infty$

$$Var(X) = ED_K^2(X, EX) = E\left(\langle s_X - s_{EX}, s_X - s_{EX}\rangle_K\right), \tag{1.3}$$

where $\|\cdot\|_K^2$ and $\langle\cdot,\cdot\rangle_K$ are, respectively, the norm and the inner product corresponding to the D_K-distance.

Independently a measure of quadratic dispersion with respect to the metric D_W^{φ} is given and analyzed in Lubiano *et al.* (2000).

Using the same idea in the next chapter a measure of the variance for FRVs with respect to the metric D_{LR} is introduced.

As for the variance also for the covariance there are different definitions. In particular, based on the variance in (1.3), expressed in terms of support functions, a covariance between two FRVs X and Y is defined as

$$Cov(X, Y) = E\left(\langle s_X - s_{EX}, s_Y - s_{EY}\rangle_K\right),$$

if $E\|X\|_K < \infty$, $E\|Y\|_K < \infty$ and $E\|X\|_K\|Y\|_K < \infty$ (Körner & Näther, 2002).

1.3 Basic statistical inference

Let $\{X_1, ..., X_n\}$ be a random sample obtained from an FRV X in $\mathcal{F}_c(\mathbb{R})$. As usual, the sample mean will be denoted as

$$\overline{X} = (X_1 + ... + X_n)/n,$$

and it will be considered as an estimator of the expected value of a FRV (see Lubiano & Gil, 1999). In the particular case of LR FRVs it results that $\overline{X}$ is an LR fuzzy number whose center is $\overline{X}^m = (X_1^m + ... + X_n^m)/n$ and whose left and right spread are, respectively, $\overline{X}^l = (X_1^l + ... + X_n^l)/n$ and $\overline{X}^r = (X_1^r + ... + X_n^r)/n$.

The sample variance is defined as

$$\hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n} D^2(X_i, \overline{X}),$$

where D is a generic metric between fuzzy elements. As estimator of the variance can be used

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n} D^2(X_i, \overline{X}),$$

that is an unbiased sample variance.

Analogously a sample covariance, based on a random sample $\{Y_i, X_i\}_{i=1,...,n}$, obtained from two FRVs X and Y, can be defined as

$$\hat{\sigma}_{XY} = \frac{1}{n}\sum_{i=1}^{n} \left(\langle s_{X_i} - s_{\overline{X}}, s_{Y_i} - s_{\overline{Y}}\rangle\right).$$

The estimation of the variance and some properties have been analyzed in Lubiano et al. (2000) and Körner (1997b). In the next chapter the statistical problem of the estimation of the variance with the D_{LR} distance is discussed.

1.4 Previous linear regression models

In this section some previous regression model in a fuzzy framework are presented. In particular the models introduced by Diamond (1988), González-Rodríguez *et al.* (2009) and Coppi *et al.* (2006) are briefly discussed. Diamond's model is one of the first fuzzy regression analyses by least squares approaches. The second one is a simple linear regression model between FRVs and the last one is devoted to the analysis of a regression model with LR response, from which the model in this work has taken inspiration.

1.4.1 Fuzzy least squares (Diamond, 1988)

Let X and Y be two triangular fuzzy random variables observed on n statistical units, whose centers, left and right spreads are, respectively, X^m, X^l, X^r and Y^m, Y^l, Y^r. It is assumed throughout that the explanatory variable X has positive support, that is, $X^m - X^l \geq 0$.

Two models have been considered:

(F1): $Y = a + bX$, where a, b are in $\mathbb{R}$

(F2): $Y = C + bX$, where b is in $\mathbb{R}$ and C is a triangular fuzzy number

It is clear that $(F2)$ is a generalization of $(F1)$ because if C is a triangular fuzzy number with null spreads, it is equal to a value a in $\mathbb{R}$.

The least squares optimization problem corresponding to the model $(F2)$ is

$$\min \sum_{i=1}^{n} d^2(Y_i, C + bX_i) \tag{1.4}$$

where d is a metric for LR fuzzy numbers, defined as

$$\begin{aligned} d(A,B) &= \Big(\big((A^m - A^l) - (B^m - B^l)\big)^2 + \big((A^m + A^r) - (B^m + B^r)\big)^2 \\ &\quad + (A^m - B^m)^2\Big)^{\frac{1}{2}} \end{aligned}$$

where A^m, A^l, A^r and B^m, B^l, B^r are, respectively, the center, the left and the right spread of the LR fuzzy numbers A and B.

Two cases have to be distinguished for the analysis, $b \geq 0$ and $b < 0$. If $b \geq 0$, it results

$$\begin{aligned} d^2(Y_i, C + bX_i) &= \big[(Y_i^m - C^m - bX_i^m) - (Y_i^l - C^l - bX_i^l)\big]^2 \\ &+ \big[(Y_i^m - C^m - bX_i^m) + (Y_i^r - C^r - bX_i^r)\big]^2 \\ &+ (Y_i^m - C^m - bX_i^m)^2 \end{aligned}$$

where C^m, C^l and C^r are, respectively, the center, the left and the right spread of the triangular fuzzy number C. If C is symmetrical, that is, $C^l = C^r$, if a solution to the minimization problem (1.4) exists for $b \geq 0$, it is given by the solutions C^{m*}, C^{l*}, b^* to the equations

$(\mathbf{S}^*)$:

$$C^m = \bar{Y}^m + \frac{(\bar{Y}^l - \bar{Y}^r)}{3} - b\left(\bar{X}^m + \frac{(\bar{X}^l - \bar{X}^r)}{3}\right) \tag{1.5}$$

$$C^l = C^r = \frac{(\bar{Y}^l + \bar{Y}^r)}{2} - b\left(\frac{(\bar{X}^l + \bar{X}^r)}{2}\right) \tag{1.6}$$

$$\begin{aligned} & nC^m\left[3\bar{X}^m + \left(\bar{X}^l - \bar{X}^r\right)\right] + nC^l\left(\bar{X}^l + \bar{X}^r\right) \\ & +b\sum_{i=1}^{n}\left[\left(X_i^m - X_i^l\right)^2 + (X_i^m - X_i^r)^2 + (X_i^m)^2\right] \\ & = \sum_{i=1}^{n}\left[\left(X_i^m - X_i^l\right)\left(Y_i^m + Y_i^r\right) + \left(X_i^m + X_i^r\right)\left(Y_i^m - Y_i^l\right) + \left(X_i^m\right)\left(Y_i^m\right)\right] \end{aligned} \tag{1.7}$$

If consider the case $b < 0$ and C symmetrical, a solution to the minimization problem (1.4) is given by

$(\mathbf{S}_*)$:

$$C^m = \bar{Y}^m + \frac{(\bar{Y}^l - \bar{Y}^r)}{3} - b\left(\bar{X}^m + \frac{(\bar{X}^l - \bar{X}^r)}{3}\right) \tag{1.8}$$

$$C^l = C^r = \frac{(\bar{Y}^l + \bar{Y}^r)}{2} + b\left(\frac{(\bar{X}^l + \bar{X}^r)}{2}\right) \tag{1.9}$$

$$\begin{aligned} & nC^m\left[3\bar{X}^m + \left(\bar{X}^l - \bar{X}^r\right)\right] - nC^l\left(\bar{X}^l + \bar{X}^r\right) \\ & +b\sum_{i=1}^{n}\left[\left(X_i^m - X_i^l\right)^2 + (X_i^m + X_i^r)^2 + (X_i^m)^2\right] \\ & = \sum_{i=1}^{n}\left[\left(X_i^m - X_i^l\right)\left(Y_i^m - Y_i^l\right) + \left(X_i^m + X_i^r\right)\left(Y_i^m + Y_i^r\right) + \left(X_i^m\right)\left(Y_i^m\right)\right] \end{aligned} \tag{1.10}$$

The fuzzy data set $\{Y_i, X_i\}_{i=1,\ldots,n}$ is said to be *coherent* if the following conditions are fulfilled:

1. $\sum_{i=1}^{n}\left[\left(X_i^l - \bar{X}^l\right)\left(X_i^r - \bar{X}^r\right)\right]\left[\left(Y_i^l - \bar{Y}^l\right)\left(Y_i^r - \bar{Y}^r\right)\right] \geq 0$
2. either $b_* \geq 0$ or $b^* \leq 0$

If $b_* \geq 0$, the data set is coherent positive and if $b^* \leq 0$ it is coherent negative.

Diamond (1988) has proved that the optimization problem (1.4) has a unique solution if the non-degenerate data set is coherent. If the data set is coherent positive, the least squares solution is given by the (S^*) system of equations, and if is coherent negative by the (S_*).

Remark 1.4.1 In Diamond (1988) a complete analytical expression for the estimators is not provided.

Remark 1.4.2 The regression models proposed by Diamond & Körner (1997), Körner & Näther (1998), Wünsche & Näther (2002) and Krätschmer (2004) are extensions or variations of the model briefly described in this section. In particular, Diamond & Körner (1997) have extended fuzzy linear models and least squares estimates to overcome and discuss the occurrence of negative spreads. They have introduced the quadratic optimization problem that can be solved by means of the Kuhn-Tucker theorem, but there is not an analytic expression for the solutions. Körner & Näther (1998) have introduced a linear regression with random fuzzy variables and have analyzed extended classical estimates, best linear estimates and least squares estimates. In the last case a quadratic problem is formalized but in presence of negative spreads, they are replaced by 0. Wünsche & Näther (2002) have presented some contributions to the theoretical regression problem with fuzzy random variables, but the solution is not complete because in $\mathcal{F}_c(\mathbb{R}^p)$ the regression function does not determine the model (see González-Rodríguez *et al.*, 2009). The problem is totally solved for a natural model in González-Rodríguez *et al.* (2009).

1.4.2 A simple linear regression model for FRVs (González-Rodríguez *et al.*, 2009)

Let Y and X be two FRVs, the simple linear regression model considered is

$$Y = aX + \varepsilon, \tag{1.11}$$

where $a \in \mathbb{R}$ and ε is an FRV with expected value $E\varepsilon = B \in \mathcal{F}_c(\mathbb{R}^p)$. The model (1.11) agrees with those in Diamond (1988), Diamond & Körner (1997), Körner & Näther (1998), Krätschmer (2004) in the sense of involving the same regression function under particular conditions.

The least squares problem consists in looking for $\hat{a} \in \mathbb{R}$ and $\hat{B} \in \mathcal{F}_c(\mathbb{R}^p)$ in order to

$$\min_{a \in A} \frac{1}{n} \sum_{i=1}^{n} D_K^2(Y_i, aX_i + B)$$

in $A = \{a^* \in \mathbb{R} | Y_i -_H a^* X_i \text{ exists for all } i = 1, ..., n\}$.

It results that either $A = \mathbb{R}$, or there exist a_0, $b_0 \in [0, 1)$, so that $A = [-a_0, b_0]$. The solutions of the above minimization problem are

$$\hat{B} = \bar{Y} -_H \hat{a}\bar{X}$$

and

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

where

$$\beta = \begin{cases} 0 & \text{if} \quad \hat{\sigma}_{XY} \leq 0 \\ \min\left\{1, \frac{b_0}{\hat{\sigma}_{XY}/\hat{\sigma}^2_X}\right\} & \text{if} \quad \hat{\sigma}_{XY} > 0 \end{cases}$$

and

$$\alpha = \begin{cases} 0 & \text{if} \quad \hat{\sigma}_{-XY} \leq 0 \\ \min\left\{1, \frac{a_0}{\hat{\sigma}_{-XY}/\hat{\sigma}^2_X}\right\} & \text{if} \quad \hat{\sigma}_{-XY} > 0 \end{cases}$$

Remark 1.4.3 If the explanatory variable X is not fuzzy, the fuzziness of the response variable depends only on the error term ε.

Remark 1.4.4 If the model is split into two models, that is it is written in terms of *mid* and *spr*,

$$mid\, Y = a \cdot mid\, X + mid\, \varepsilon \qquad spr\, Y = |a|\, spr\, X + spr\, \varepsilon$$

it entails the same regression coefficient a for both models, which limits its applicability in practice.

1.4.3 A linear regression model with LR fuzzy response (Coppi *et al.*, 2006)

Let $X_1, ..., X_m$ be m crisp quantitative explanatory variables and Y an LR fuzzy response variable, observed on n statistical units. For each unit i it results $Y_i \in \mathcal{F}_{LR}$, i.e. the observational space for the vector $\underline{Y}$ is $\mathcal{F}_{LR}$. The basic idea is modeling the centers of the LR response variable by means of a classical regression model, and simultaneously modeling the left and the right spreads of the response through simple linear regressions on its estimate centers, that is,

$$\begin{aligned} \underline{Y}^m &= \underline{\mu} + \underline{\varepsilon}, \\ \underline{Y}^m - \underline{Y}^l &= (\underline{\mu} - \underline{\delta}_L) + \underline{\varepsilon}_L \\ \underline{Y}^m - \underline{Y}^r &= (\underline{\mu} + \underline{\delta}_R) + \underline{\varepsilon}_R \end{aligned}$$

where $\underline{\varepsilon}, \underline{\varepsilon}_L, \underline{\varepsilon}_R$ are the vector of residuals and $\underline{\mu}, \underline{\delta}_L, \underline{\delta}_R$ are the vectors of the theoretical values of the response variable. These theoretical values are reparametrized in the following way:

$$\begin{aligned} \underline{\mu} &= \mathbf{F}\underline{\gamma}, \\ \underline{\delta}_L &= \eta_L\underline{\mu} + \underline{\xi}_L\underline{1} \\ \underline{\delta}_R &= \eta_R\underline{\mu} + \underline{\xi}_R\underline{1}, \end{aligned} \tag{1.12}$$

where $\mathbf{F}$ is a design matrix.

The optimization problem consists in looking for $\hat{\underline{\gamma}}$, $\hat{\eta}_L$, $\hat{\eta}_R$, $\hat{\underline{\xi}}_L$, $\hat{\underline{\xi}}_R$ in order to

$$\min D^2_{LR}(\underline{Y}, \underline{Y}^*)$$

where $D^2_{LR}(\underline{Y}, \underline{Y}^*)$ is a generalization of the Yang and Ko metric between the observed values $\underline{Y}$ and the theoretical ones $\underline{Y}^*$ with $\underline{\mu}$ as vector of centers and $\underline{\delta}_L$, $\underline{\delta}_R$, respectively, vector of left spreads and vector of right spreads. It results

$$\begin{aligned} D^2_{LR}(\underline{Y}, \underline{Y}^*) &= \|\underline{Y}^m - \underline{\mu}\|^2 + \|(\underline{Y}^m - \lambda\underline{Y}^l) - (\underline{\mu} - \lambda\underline{\delta}_L)\|^2 \\ &\quad + \|(\underline{Y}^m + \rho\underline{Y}^r) - (\underline{\mu} + \rho\underline{\delta}_R)\|^2 \\ &= 3(\underline{Y}^m - \underline{\mu})'(\underline{Y}^m - \underline{\mu}) - 2\lambda(\underline{Y}^m - \underline{\mu})'(\underline{Y}^l - \underline{\delta}_L) \\ &\quad + \lambda^2(\underline{Y}^l - \underline{\delta}_L)'(\underline{Y}^l - \underline{\delta}_L) \\ &\quad + 2\rho(\underline{Y}^m - \underline{\mu})'(\underline{Y}^r - \underline{\delta}_R) + \rho^2(\underline{Y}^r - \underline{\delta}_R)'(\underline{Y}^r - \underline{\delta}_R), \end{aligned}$$

where $\lambda = \int_0^1 L^{-1}(\omega)d\omega$ and $\rho = \int_0^1 R^{-1}(\omega)d\omega$.

Equating to zero the partial derivatives of D^2_{LR} with respect to the parameters $\underline{\gamma}$, η_L, η_R, ξ_L, ξ_R it is easy to check the following set of equations, on which an iterative solution can be based,

$$\begin{aligned} \eta_L &= \lambda^{-1}(\underline{\gamma}'\mathbf{F}'\mathbf{F}\underline{\gamma})^{-1}\left[\lambda\left(\underline{\gamma}'\mathbf{F}'\underline{Y}^l - \underline{\gamma}'\mathbf{F}'\underline{1}\xi_L\right) - \left(\underline{\gamma}'\mathbf{F}'\underline{Y}^m - \underline{\gamma}'\mathbf{F}'\mathbf{F}\underline{\gamma}\right)\right], \\ \eta_R &= \rho^{-1}(\underline{\gamma}'\mathbf{F}'\mathbf{F}\underline{\gamma})^{-1}\left[\rho\left(\underline{\gamma}'\mathbf{F}'\underline{Y}^r - \underline{\gamma}'\mathbf{F}'\underline{1}\xi_R\right) + \left(\underline{\gamma}'\mathbf{F}'\underline{Y}^m - \underline{\gamma}'\mathbf{F}'\mathbf{F}\underline{\gamma}\right)\right], \\ \xi_L &= (n\lambda)^{-1}\left[\lambda\underline{1}'\left(\underline{Y}^l - \mathbf{F}\underline{\gamma}\eta_L\right) - \underline{1}'\left(\underline{Y}^m - \mathbf{F}\underline{\gamma}\right)\right], \\ \xi_R &= (n\rho)^{-1}\left[\rho\underline{1}'\left(\underline{Y}^r - \mathbf{F}\underline{\gamma}\eta_R\right) + \underline{1}'\left(\underline{Y}^m - \mathbf{F}\underline{\gamma}\right)\right], \\ \underline{\gamma} &= [3 - \lambda\eta_L(2 - \lambda\eta_L) + \rho\eta_R(2 + \rho\eta_R)]^{-1}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}' \\ &\quad \times \left[3\underline{Y}^m - \lambda(\underline{Y}^m\eta_L + \underline{Y}^l - \underline{1}\xi_L) + \lambda^2(\underline{Y}^l\eta_L - \underline{1}\eta_L\xi_L)\right. \\ &\quad \left. + \rho(\underline{Y}^m\eta_R + \underline{Y}^r - \underline{1}\xi_R) + \rho^2(\underline{Y}^r\eta_R - \underline{1}\eta_R\xi_R)\right]. \end{aligned}$$

Since the evaluation of the regression coefficients, $\underline{\gamma}$, is crisp and the response variable is fuzzy while the explanatory ones are crisp, the authors have introduced an implicit fuzzy regression model, that is,

$$Y_i^* = \beta_1 f_{i1} + ... + \beta_p f_{ip}, \qquad i = 1, ..., n, \tag{1.13}$$

where β_k $(k = 1, ..., p)$ are LR fuzzy numbers whose centers are β_k^m and whose left and right spreads are, respectively, β_k^l and β_k^r. These fuzzy coefficients can be related to the parameters of the responses by means of

$$\begin{aligned} \underline{\mu} &= \mathbf{F}\underline{\beta}^m, \\ \underline{\delta}_L &= |\mathbf{F}|\,\underline{\beta}^l, \\ \underline{\delta}_R &= |\mathbf{F}|\,\underline{\beta}^r, \end{aligned} \tag{1.14}$$

where $|\mathbf{F}|$ denotes the matrix of the absolute values $|f_{ik}|$.

The above obtained iterative LS solutions may not verify the system (1.14), but these relationships may be exploited in order to obtain estimates of $\underline{\beta}^m$, $\underline{\beta}^l$, $\underline{\beta}^r$ which are compatible with the estimates $\widehat{\underline{\mu}}$, $\widehat{\underline{\delta}}_L$ and $\widehat{\underline{\delta}}_R$, that is

$$\begin{aligned} \widehat{\underline{\mu}} &= \mathbf{F}\underline{\beta}^m + \underline{\tau}^m, \\ \widehat{\underline{\delta}}_L &= |\mathbf{F}|\,\underline{\beta}^l + \underline{\tau}^l, \\ \widehat{\underline{\delta}}_R &= |\mathbf{F}|\,\underline{\beta}^r + \underline{\tau}^r, \end{aligned}$$

where $\underline{\tau}^m$, $\underline{\tau}^l$ and $\underline{\tau}^r$ are the vectors of residuals.

The ordinary least squares estimate of $\underline{\beta}^m$ is given by

$$\widehat{\underline{\beta}}^m = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\widehat{\underline{\mu}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{F}\widehat{\underline{\gamma}} = \widehat{\underline{\gamma}},$$

where $\widehat{\underline{\gamma}}$ is the least squares estimate obtained by means of model (1.12). The estimates of the spreads $\widehat{\underline{\beta}}^l$ and $\widehat{\underline{\beta}}^r$ may be got by means of a constrained least squares problem due to the non-negativity condition.

In this way the imprecision of the regression function has been assessed.

Concerning the uncertainty linked with the data generation process it is used a bootstrap procedure. Starting from the data it is possible to generate bootstrap samples. For each of these samples optimal parameters can be computed. Through the variations of the score across the bootstrap samples it is possible to estimate the probabilistic uncertainty.

Remark 1.4.5 The global minimum is not attained but only a local one, due to the use of an iterative algorithm.

Remark 1.4.6 Unlike the other models introduced in this section, Coppi *et al.* (2006) have not formalized a model based on FRVs, and to find this type of formalization the new regression model, presented in this work, comes up.

1.5 Concluding remarks

- Some basic concepts have been introduced, in order to handle random experiments for which the observed characteristic is imprecisely measured. Namely, the concept of fuzzy set has been introduced and illustrated in some environmental and medical applications. The arithmetics between fuzzy sets has been discussed and the concept of fuzzy random variable has been analyzed from different points of view.

- The main important regression models in literature have been introduced and their advantages and limitations in connection with the aim of the present work have been described.

- The model in this thesis is strongly connected with that in Coppi *et al.* (2006), although the formalization involving FRVs is closer to González-Rodríguez *et al.* (2009).

Chapter 2

An isometry for $\mathcal{F}_{LR}$ and a variance for LR fuzzy random variables

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

$$s(A) = (A^m, A,^l, A^r). \tag{2.1}$$

In what follows we use without distinction $A \in \mathcal{F}_{LR}$ or its s-representation. The function s is obviously *semi-linear*, because $s(A) + s(B) = s(A + B)$ and $\gamma s(A) = s(\gamma A)$, if $\gamma > 0$.

In the next chapter a regression model with LR fuzzy response Y is introduced. Each LR fuzzy random variable Y can be expressed as a random vector

$$(Y^m, Y^l, Y^r) : \Omega \to \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+.$$

Since the left and the right spread, Y^l and Y^r, of the response variable will be transformed by means of functions from $\mathbb{R}^+$ to $\mathbb{R}$, the response can be considered as a vector in $\mathbb{R}^3$. In view of the utilization we are going to make of this result, it is necessary to define an appropriate distance between elements of $\mathbb{R}^3$.

In Section 2.1 it is proved that the space of LR fuzzy numbers is isometric to a closed convex cone of $\mathbb{R}^3$, by means of $\langle \cdot, \cdot \rangle_{\lambda\rho}$. This is the inner product corresponding to a generalization of the Yang and Ko distance, $D^2_{\lambda\rho}$.

The operation $\langle A, B \rangle_{LR} = \langle s_A, s_B \rangle_{LR}$ is not exactly an inner product due to the lack of linearity, but due to its interesting properties it is used in Section 2.2. In this section the concept of variance for fuzzy random variables based on D^2_{LR} is given. The idea is the same followed in Körner (1997) and Lubiano *et al.* (2000) in terms of

other metrics. As in the classical theory, some properties of the variance are proved. In particular it is shown that it verifies the *Frèchet principle*, so the least squares criterion can be soundly applied. Furthermore the covariance is defined. Section 2.2.2 contains the estimation problem. It is proved that the estimators are unbiased and strongly consistent.

In order to illustrate the consistency of the estimators some simulation studies are presented and some empirical examples are given (Sections 2.3 and 2.4).

The last section is focused on final evaluation and open problems.

2.1 The isometry

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

Proposition 2.1.1 *Given* $a = (a_1, a_2, a_3)$, $b = (b_1, b_2, b_3) \in \mathbb{R}^3$ *and* λ, $\rho \in \mathbb{R}^+$, $(\mathbb{R}^3, D_{\lambda\rho})$ *is a metric space, where*

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

takes inspiration from the Yang-Ko distance. Moreover

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

is an inner product.

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

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

By Cauchy-Schwarz inequality, we obtain

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

Thus $D_{\lambda\rho}(a,b) \leq D_{\lambda\rho}(a,c) + D_{\lambda\rho}(c,b)$. It results that $D_{\lambda\rho}(a,b)$ is a metric in $\mathbb{R}^3$. Since the terms defining $\langle\cdot,\cdot\rangle_{\lambda\rho}$ are based on $\langle\cdot,\cdot\rangle_{\mathbb{R}}$, it is easy to check that

1. $\langle a,b\rangle_{\lambda\rho} = \langle b,a\rangle_{\lambda\rho}$
2. $\langle (a+c),b\rangle_{\lambda\rho} = \langle a,b\rangle_{\lambda\rho} + \langle c,b\rangle_{\lambda\rho}$
3. $\langle ka,b\rangle_{\lambda\rho} = k\langle a,b\rangle_{\lambda\rho}$

The thesis is proved.

□

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

Proposition 2.1.2 *Given the space $\mathcal{F}_{LR}$, consider $\lambda = \int_0^1 L^{-1}(\omega)d\omega$ and $\rho = \int_0^1 R^{-1}(\omega)d\omega$, then $\mathcal{F}_{LR}$ is isometric to a closed convex cone of $\mathbb{R}^3$ endowed with the inner product $\langle\cdot,\cdot\rangle_{\lambda\rho}$.*

Proof. As $\mathcal{S} = \{s(A) : A \in \mathcal{F}_{LR}\}$ is $\mathbb{R} \times [0,\infty) \times [0,\infty)$, $\mathcal{S}$ is clearly a closed convex cone, and the metric is preserved by definition.

□

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

2.2 The variance based on D_{LR}

As discussed in Chapter 1, the concept of variance for FRVs has been previously established in terms of several metrics (see Körner 1997a, 1997b and Lubiano *et al.* 2000). By following the same ideas, we can also consider it in the sense of the D_{LR} metric.

2.2.1 Definition and properties of the variance

The variance of an LR fuzzy random variable $X = (X^m, X^l, X^r)$ with $E\|X\|^2_{LR} < \infty$ is defined by

$$Var(X) = ED^2_{LR}(X, EX),$$

or, equivalently, in terms of support functions

$$Var(X) = E\left\langle s_X - s_{EX}, s_X - s_{EX}\right\rangle_{LR} \tag{2.2}$$

It can be easily checked that

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

Inspired by the expression (2.2) of the variance, we can also define the covariance as follows.

Definition 2.2.1 *The covariance between two LR fuzzy random variables* $X = (X^m, X^l, X^r)$ *and* $Y = (Y^m, Y^l, Y^r)$ *is defined by*

$$Cov(X, Y) = E\langle s_X - s_{EX}, s_Y - s_{EY}\rangle_{LR}.$$

In this case it is easy to prove that

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

The D_{LR}-variance satisfies the same suitable properties of the usual variance in $\mathbb{R}$, that is,

Proposition 2.2.1 *Let* X *and* Y *be LR fuzzy random variables,* $A \in \mathcal{F}_{LR}$ *and* $\gamma \in \mathbb{R}$. *Then*

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

Proof. By means of properties of the variance and the covariance for real-valued random variables it is easy to prove this proposition.

1. Since, for Z and W real-valued random variables, $Var(Z) = EZ^2 - (EZ)^2$ and $Cov(Z,W) = E(ZW) - EZEW$, we have

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

2. Since, for Z, W real-valued random variables and $\gamma \in \mathbb{R}$, $Var(\gamma Z) = \gamma^2 Var(Z)$ and $Cov(\gamma Z, \gamma W) = \gamma^2 Cov(Z,W)$,

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

3. Since, for real-valued random variables, the variance and the covariance are invariant with respect to translation, it follows that

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

4. Taking into account that $X + Y$ is an LR FRV whose s-representation is $(X^m + Y^m, X^l + Y^l, X^r + Y^r)$, it results

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

and since $X^m + Y^m$, $X^l + Y^l$ and $X^r + Y^r$ are sums of real-valued random variables

$$\begin{aligned} Var(X+Y) &= 3Var(X^m) + 3Var(Y^m) + 6Cov(X^m, Y^m) \\ &\quad +\lambda^2 Var(X^l) + \lambda^2 Var(Y^l) + 2\lambda^2 Cov(X^l, Y^l) \\ &\quad +\rho^2 Var(X^r) + \rho^2 Var(Y^r) + 2\rho^2 Cov(X^r, Y^r) \\ &\quad -2\lambda Cov(X^m, X^l) - 2\lambda Cov(X^m, Y^l) \\ &\quad -2\lambda Cov(Y^m, X^l) - 2\lambda Cov(Y^m, Y^l) \\ &\quad +2\rho Cov(X^m, X^r) + 2\rho Cov(X^m, Y^r) \\ &\quad +2\rho Cov(Y^m, X^r) + 2\rho Cov(Y^m, Y^r) \\ &= Var(X) + Var(Y) + 2Cov(X, Y). \end{aligned}$$

If X and Y are independent, $Cov(X, Y) = 0$, hence

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

5. Using the metric D^2_{LR} we have

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

and

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

By adding and subtracting in the term $(X^m - A^m)$ the expected value of X^m, it results

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

Analogously

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

and

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

As a consequence

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

Taking into account that

$$\begin{aligned}
D^2_{LR}(A, EX) &= 3(A^m - EX^m)^2 + \lambda^2(A^l - EX^l)^2 + \rho^2(A^r - EX^r)^2 \\
&\quad -2\lambda(A^m - EX^m)(A^l - EX^l) \\
&\quad +2\rho(A^m - EX^m)(A^r - EX^r),
\end{aligned}$$

and

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

we obtain the thesis

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

□

Property 5 of Proposition 2.2.1 shows that $E\left[D^2_{LR}(X, A)\right]$ is minimized for $A = EX$, that is, the Aumann expectation agrees with the Frèchet-expectation with respect to the Yang-Ko metric D^2_{LR}.

Proposition 2.2.2 *Let X and Y be LR fuzzy random variables. Then*

1. $Cov(X,Y) = E\langle s_X, s_Y\rangle_{LR} - \langle s_{EX}, s_{EY}\rangle_{LR}$,
2. $Var(X) = Cov(X,X)$,
3. $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$.

Remark 2.2.1 Due to the lack of linearity of $\mathcal{F}_{LR}$ the covariance does not have the same meaning or the properties of the covariance in $\mathbb{R}$. For example, if X is non-degenerate and symmetrical with respect to 0, $X = -X$, then $Cov(X,-X) = Cov(X,X) = Var(X) \neq 0$, that is, $Cov(X,-X) \neq -Cov(X,X)$, contrary to what happens in the real case.

2.2.2 Estimation of the variance and covariance

The estimation of the variance and some properties have been also discussed in Lubiano *et al.* (2000) and Körner (1997b). In this section we analyze the statistical problem of the estimation for the variance of FRVs based on D_{LR}.

Let X be an LR FRV with $E\|X\|^2_{LR} < \infty$, observed on n statistical units $\{X_i\}_{i=1,\ldots,n}$. Analogously to the classical case, the estimator of $Var(X)$ can be defined as follows

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} D^2_{LR}(X_i, \overline{X}),$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Note that $S_n^2 : \Omega^n \to \mathbb{R}$ is a real-valued random variable.

Proposition 2.2.3 *Let X be an LR FRV, $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} D^2_{LR}(X_i, \overline{X})$ is an unbiased and strongly consistent estimator of the variance $Var(X)$, that is*

$$ES_n^2 = Var(X), \quad \forall n \in N, \qquad and \qquad S_n^2 \overset{n\to\infty}{\longrightarrow} Var(X) \qquad a.s. - [P].$$

Proof. We start by proving the unbiasedness of the estimator.

$$\begin{aligned} E\frac{n-1}{n}S_n^2 &= E\frac{1}{n}\sum_{i=1}^{n} D^2_{LR}(X_i, \overline{X}) \\ &= E\frac{1}{n}\sum_{i=1}^{n}\left[3(X_i^m - \overline{X}^m)^2 + \lambda^2(X_i^l - \overline{X}^l)^2 + \rho^2(X_i^r - \overline{X}^r)^2\right] \\ &\quad + E\frac{1}{n}\sum_{i=1}^{n}\left[-2\lambda(X_i^m - \overline{X}^m)(X_i^l - \overline{X}^l) + 2\rho(X_i^m - \overline{X}^m)(X_i^r - \overline{X}^r)\right]. \end{aligned}$$

By adding and subtracting in the terms $(X_i^m - \overline{X}^m)$, $(X_i^l - \overline{X}^l)$ and $(X_i^r - \overline{X}^r)$, respectively, the expectation values of X^m, X^l and X^r, it results

$$\begin{aligned} E\frac{n-1}{n}S_n^2 &= E\frac{1}{n}\sum_{i=1}^{n}\left[3(X_i^m - EX^m + EX^m - \overline{X}^m)^2\right] \\ &+E\frac{1}{n}\sum_{i=1}^{n}\left[\lambda^2(X_i^l - EX^l + EX^l - \overline{X}^l)^2\right] \\ &+E\frac{1}{n}\sum_{i=1}^{n}\left[\rho^2(X_i^r - EX^r + EX^r - \overline{X}^r)^2\right] \\ &-E\frac{1}{n}\sum_{i=1}^{n}\left[2\lambda(X_i^m - EX^m + EX^m - \overline{X}^m)(X_i^l - EX^l + EX^l - \overline{X}^l)\right] \\ &+E\frac{1}{n}\sum_{i=1}^{n}\left[2\rho(X_i^m - EX^m + EX^m - \overline{X}^m)(X_i^r - EX^r + EX^r - \overline{X}^r)\right]. \end{aligned}$$

Through simple operations it is easy to check that

$$\begin{aligned} E\frac{n-1}{n}S_n^2 &= 3Var(X^m) + \lambda^2 Var(X^l) + \rho^2 Var(X^r) \\ &-2\lambda Cov(X^m, X^l) + 2\rho Cov(X^m, X^r) \\ &-3Var(\overline{X}^m) - \lambda^2 Var(\overline{X}^l) - \rho^2 Var(\overline{X}^r) \\ &+2\lambda Cov(\overline{X}^m, \overline{X}^l) - 2\rho Cov(\overline{X}^m, \overline{X}^r). \end{aligned}$$

Since the variance and the covariance of the sample means of real-valued random variables can be written in terms of the variance and the covariance of the given variables, it results

$$E\frac{n-1}{n}S_n^2 = Var(X) - Var(\overline{X}) = Var(X) - \frac{1}{n}Var(X).$$

It follows that

$$ES_n^2 = Var(X), \quad \forall n \in N.$$

Concerning the consistency of the estimator, starting from

$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left[3(X_i^m - \overline{X}^m)^2 + \lambda^2(X_i^l - \overline{X}^l)^2 + \rho^2(X_i^r - \overline{X}^r)^2\right] \\ &+\frac{1}{n-1}\sum_{i=1}^{n}\left[-2\lambda(X_i^m - \overline{X}^m)(X_i^l - \overline{X}^l) + 2\rho(X_i^m - \overline{X}^m)(X_i^r - \overline{X}^r)\right], \end{aligned}$$

we have that

$$S_n^2 = \frac{3}{n-1}\sum_{i=1}^{n}(X_i^m - \overline{X}^m)^2 + \frac{\lambda^2}{n-1}\sum_{i=1}^{n}(X_i^l - \overline{X}^l)^2 + \frac{\rho^2}{n-1}\sum_{i=1}^{n}(X_i^r - \overline{X}^r)^2$$

$$-\frac{2\lambda}{n-1}\sum_{i=1}^{n}(X_i^m-\overline{X}^m)(X_i^l-\overline{X}^l)+\frac{2\rho}{n-1}\sum_{i=1}^{n}(X_i^m-\overline{X}^m)(X_i^r-\overline{X}^r)$$
$$= \frac{3n}{n-1}\hat{\sigma}^2_{X^m}+\frac{\lambda^2 n}{n-1}\hat{\sigma}^2_{X^l}+\frac{\rho^2 n}{n-1}\hat{\sigma}^2_{X^r}-\frac{2\lambda n}{n-1}\hat{\sigma}_{X^mX^l}+\frac{2\rho n}{n-1}\hat{\sigma}_{X^mX^r}.$$

Since the sample variance of a real-valued random variable and the sample covariance between two real-valued random variables are, respectively, strongly consistent estimators of the variance and covariance, we obtain the consistency of the estimator S_n^2, that is

$$\begin{aligned} S_n^2 &\overset{n\to\infty}{\longrightarrow} 3Var(X^m)+\lambda^2 Var(X^l)+\rho^2 Var(X^r) \\ &\quad -2\lambda Cov(X^m,X^l)+2\rho Cov(X^m,X^r), \\ S_n^2 &\overset{n\to\infty}{\longrightarrow} Var(X) \qquad a.s.-[P]. \end{aligned}$$

□

In an analogous way, it is possible to determine an estimator for the covariance between two LR FRVs and to check some statistical properties.

Let Y and X be two LR FRVs, observed on n statistical units $\{Y_i, X_i\}_{i=1,\ldots,n}$, with $E\|Y\|^2_{LR}<\infty$ and $E\|X\|^2_{LR}<\infty$, the estimator of $Cov(X,Y)$ can be defined as

$$C_n=\frac{1}{n-1}\sum_{i=1}^{n}\langle s_X-s_{\overline{X}}, s_Y-s_{\overline{Y}}\rangle_{LR},$$

where $\overline{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$ and $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

Taking into account that C_n can be written in terms of sample covariances of real-valued random variables, it is easy to check the next proposition, by following reasoning similar to that in Proposition 2.2.3.

Proposition 2.2.4 *Let X and Y be two LR FRVs, $C_n=\frac{1}{n-1}\sum_{i=1}^{n}\langle s_X-s_{\overline{X}}, s_Y-s_{\overline{Y}}\rangle_{LR}$ is an unbiased and strongly consistent estimator of the covariance $Cov(X,Y)$, that is*

$$EC_n=Cov(X,Y) \qquad and \qquad C_n \overset{n\longrightarrow\infty}{\longrightarrow} Cov(X,Y) \qquad a.s.-[P].$$

2.3 Simulations

In order to illustrate the consistency of the estimator of the variance S_n^2 in an empirical way, we consider a simulated situation. An LR fuzzy random variable X has been generated by considering a real variable X^m normally distributed as

$N(0,1)$ and two real random variables X^l and X^r distributed as χ^2_1 and generating values of X based on the assumption of independence among the above three random variables. If we choose the triangular case, $\lambda = \rho = 1/2$, it is easy to check that the theoretical variance of the fuzzy variable X is equal to 4.

Table 2.1: D_{LR}-variance estimates in a simulated case.

n	S_n^2
100	3.6831
1000	3.9045
10000	3.9273
100000	3.9984

As shown in Table 2.1, the estimates of the variance for the simulated data are close to the theoretical value, as n increases. In particular, from $n = 1000$, they are quite accurate.

2.4 Empirical results

We consider the data introduced in Example 3.1.1. In this example we consider the quality of the trees, that is an LR fuzzy random variable. By means of the n statistical observations in Table 1.1, it is possible to estimate the variance of the quality. It is easy to check that the estimate is equal to 1068.1 (estimated standard deviation equal to 32.6816).

2.5 Final evaluation and open problems

In this chapter we have introduced and analyzed the concept of variance in the sense of the D_{LR}-metric by following the ideas in Körner (1997) and Lubiano *et al.* (2000).

This analysis is necessary for the subsequent chapters, since we will apply the least squares criterion to find the estimators of a given regression model involving LR FRVs. The properties verified for this variance make it suitable for analyzing the variability of involved LR FRVs, as usual in least squares problems.

As open problems concerning this chapter we propose to follow the idea in (Ramos, 2008) to establish the asymptotic distribution of the sample estimators

and to use them for developing confidence intervals and hypothesis testing procedures.

Chapter 3

A linear regression model with imprecise response

The problem of linear regression in a fuzzy framework has been developed in several studies, as described in Chapter 1. In the present work a new linear regression model for LR fuzzy responses and scalar predictors is given. It takes inspiration from the model introduced in Coppi *et al.* (2006).

In the next section the new population regression model is formally defined. In order to measure the degree of linear relationship in Section 3.2 a determination coefficient is given, defined by means of the metric $D^2_{\lambda\rho}$. The main part of this chapter is focused on statistical inferences. Section 3.3 contains the minimization problem and the procedure to get the least squares estimators. Throughout this section some statistical properties are proved, the asymptotic distribution of the estimators is determined and, to analyze the accuracy of the estimators, a bootstrap procedure is given. The model is employed on simulated data and in two real life situations. Sections 3.4 and 3.5, respectively, contain confidence regions and hypothesis testing on the regression parameters. Section 3.6 concerns the estimation of the determination coefficient and some statistical properties. In Section 3.7 a linear independence test is introduced. It is given by means of the asymptotic approach and the bootstrap one. To illustrate the empirical significance level of the test some simulation studies and empirical results are discussed. The last part is devoted to the study of the behavior of the power of the asymptotic test by means of a sequence of local alternatives.

3.1 The regression model

Consider a random experiment in which an LR fuzzy response variable Y and a real explanatory variable X are observed on n statistical units, $\{Y_i, X_i\}_{i=1,\ldots,n}$. Since Y is determined by (Y^m, Y^l, Y^r), the proposed regression model concerns the real-valued random variables in this tuple. The center Y^m can be related to the explanatory variable X through a classical regression model. However, as shown also in the model introduced in Coppi *et al.* (2006), described in the preliminary part of this work, the restriction of non-negativity satisfied by Y^l and Y^r entails some difficulties. One solution is to consider a model with the restriction of non-negativity but, when a variable has this kind of restriction, the errors of the model may be dependent on the explanatory variable, and the classical inferential methods are not efficient (see, for instance, Liew 1976, Gallant & Gerig 1980).

In contrast we propose modeling a transformation of the left spread and a transformation of the right spread of the response through simple linear regressions (on the explanatory variable X). This can be represented in the following way, letting $g : (0, +\infty) \longrightarrow \mathbb{R}$ and $h : (0, +\infty) \longrightarrow \mathbb{R}$ be invertible:

$$\begin{cases} Y^m = a_m X + b_m + \varepsilon_m \\ g(Y^l) = a_l X + b_l + \varepsilon_l \\ h(Y^r) = a_r X + b_r + \varepsilon_r \end{cases} \tag{3.1}$$

where ε_m, ε_l and ε_r are real-valued random variables with $E(\varepsilon_m|X) = E(\varepsilon_l|X) = E(\varepsilon_r|X) = 0$. The variance of the explanatory variable X will be denoted by σ_X^2 and Σ will stand for the covariance matrix of $(\varepsilon_m, \varepsilon_l, \varepsilon_r)$, whose variances are strictly positive and finite.

It is easy to check that the variables ε_m, ε_l and ε_r are uncorrelated with the variable X. For instance,

$$\begin{aligned} Cov(\varepsilon_m, X) &= E\left[(\varepsilon_m - E\varepsilon_m)(X - EX)\right] \\ &= E\left[E\left((\varepsilon_m - E\varepsilon_m)(X - EX)\right)|X\right] \\ &= E(X - EX)E\left[E\left(\varepsilon_m - E\varepsilon_m\right)|X\right], \end{aligned}$$

as the expected value of the variable ε_m given X is equal to 0, we get the result, that is $Cov(\varepsilon_m, X) = 0$. Analogously, it is possible to check the same result for the variables ε_l and ε_r.

The functions g and h transform the left and the right spread, that are positive variables, into variables that can take all the real values. This makes possible the use of the linear regressions.

Remark 3.1.1 In practice, particularly in the socio-economical domain, it is possible to have restrictions on the center Y^m or on the explanatory variable X. In this case we can transform these variables too. It results a non linear model.

Example 3.1.1 We consider the data introduced in Chapter 1 (see Table 1.1). In this example we consider only the dependence relationship of the quality of trees on the height. We will use the new linear regression model to analyze the part of the *quality*, Y, of the 238 trees explained by the *height*, X. In presence of constrained variables, a common approach consists in transforming the constrained variable into an unconstrained one by means of the logarithmic transformation (that is g=h=ln). We will use this approach in this example to transform the spreads into real variables without the restriction of non-negativity (see Examples 3.3.1, 4.3.1).

3.1.1 Theoretical values

In Proposition 3.1.1 we show that the population parameters can be expressed, as usual, in terms of some moments involving the considered random variables.

Proposition 3.1.1 *Let Y be an LR fuzzy random variable and X a real random variable satisfying the linear model (3.1), then we have that*

$$a_m = \frac{\sigma_{Y^m X}}{\sigma_X^2}, \quad a_l = \frac{\sigma_{g(Y^l)X}}{\sigma_X^2}, \quad a_r = \frac{\sigma_{h(Y^r)X}}{\sigma_X^2}, \quad b_m = E(Y^m|X) - \frac{\sigma_{Y^m X}}{\sigma_X^2} EX,$$

$$b_l = E[g(Y^l)|X] - \frac{\sigma_{g(Y^l)X}}{\sigma_X^2} EX, \quad b_r = E[h(Y^r)|X] - \frac{\sigma_{h(Y^r)X}}{\sigma_X^2} EX.$$

Proof. Under the assumptions in this proposition, we have that $Y^m = a_m X + b_m + \varepsilon_m$ and $EY^m = a_m EX + b_m + E\varepsilon_m$, hence

$$\begin{aligned} \sigma_{Y^m X} = Cov(Y^m, X) &= E\left[(Y^m - EY^m)(X - EX)\right] \\ &= E\left[(a_m X + \varepsilon_m - a_m EX)(X - EX)\right] \\ &= a_m E(X - EX)^2 + E\left[\varepsilon(X - EX)\right]. \end{aligned}$$

Since the variables ε_m and X are uncorrelated it follows that

$$\sigma_{Y^m X} = a_m Var(X) = a_m \sigma_X^2,$$

and as a result

$$a_m = \frac{\sigma_{Y^m X}}{\sigma_X^2}.$$

By means of opportune substitutions it is easy to check that

$$b_m = EY^m - \frac{\sigma_{Y^m X}}{\sigma_X^2} EX.$$

Analogously, following the same reasoning for $\sigma_{g(Y^l)X}$ and $\sigma_{h(Y^r)X}$,

$$a_l = \frac{\sigma_{g(Y^l)X}}{\sigma_X^2} \qquad b_l = Eg(Y^l) - \frac{\sigma_{g(Y^l)X}}{\sigma_X^2} EX$$
$$a_r = \frac{\sigma_{h(Y^r)X}}{\sigma_X^2} \qquad b_r = Eh(Y^r) - \frac{\sigma_{h(Y^r)X}}{\sigma_X^2} EX$$

□

3.2 Determination coefficient

In order to quantify the degree of linear relationship between the response variables and the explanatory ones in a regression model, it is possible to use the determination coefficient. The following proposition proves the decomposition of the total variation, and taking it into account it is possible to define the determination coefficient for the new regression model.

Proposition 3.2.1 *Let Y be an LR fuzzy random variable and X a random variable satisfying the linear model (3.1), by indicating $\widetilde{Y} = (Y^m, g(Y^l), h(Y^l))$, we obtain*

$$E\left[D^2_{\lambda\rho}(\widetilde{Y}, E\widetilde{Y})\right] = E\left[D^2_{\lambda\rho}(\widetilde{Y}, E(\widetilde{Y}|X))\right] + E\left[D^2_{\lambda\rho}(E(\widetilde{Y}|X), E\widetilde{Y})\right], \qquad (3.2)$$

that is the total variation of the response $\widetilde{Y}$ is equal to the sum of the variation that does not depend on the model and the variation explained by the model.

Proof. The total variation can be written as follows

$$\begin{aligned} E\left[D^2_{\lambda\rho}(\widetilde{Y}, E\widetilde{Y})\right] &= 3E[Y^m - EY^m]^2 + \lambda^2 E[g(Y^l) - Eg(Y^l)]^2 \\ &\quad +\rho^2 E[h(Y^r) - Eh(Y^r)]^2 \\ &\quad -2\lambda E[(Y^m - EY^m)(g(Y^l) - Eg(Y^l))] \\ &\quad +2\rho E[(Y^m - EY^m)(h(Y^r) - Eh(Y^r))]. \end{aligned}$$

Starting from the first term $E[Y^m - EY^m]^2$, that is the variance of the real random variable Y^m, we add and subtract the conditional expectation of Y^m given X, $E(Y^m|X)$, from the term $Y^m - EY^m$ and we get

$$\begin{aligned} E[Y^m - EY^m]^2 &= E[Y^m - E(Y^m|X) + E(Y^m|X) - EY^m]^2 \\ &= E[Y^m - E(Y^m|X)]^2 + E[E(Y^m|X) - EY^m]^2 \\ &\quad +2E[(Y^m - E(Y^m|X))(E(Y^m|X) - EY^m)]. \end{aligned}$$

Since the expectation of a real-valued random variable is equal to the expectation of the conditional expectation of the same variable it follows that

$$\begin{aligned} E[Y^m - EY^m]^2 &= E[Y^m - E(Y^m|X)]^2 + E[E(Y^m|X) - EY^m]^2 \\ &\quad +2E\{E[(Y^m - E(Y^m|X))(E(Y^m|X) - EY^m)|X]\}. \end{aligned}$$

Given X the expected value $E[(Y^m - E(Y^m|X))(E(Y^m|X) - EY^m)|X]$ is equal to $(E(Y^m|X) - EY^m)E[(Y^m - E(Y^m|X))|X]$ and, consequently, it is equal to 0. Hence

$$E[Y^m - EY^m]^2 = E[Y^m - E(Y^m|X)]^2 + E[E(Y^m|X) - EY^m]^2.$$

Analogously, using the real random variables $g(Y^l)$ and $h(Y^r)$, we get

$$\begin{aligned} E[g(Y^l) - Eg(Y^l)]^2 &= E[g(Y^l) - E(g(Y^l)|X)]^2 + E[E(g(Y^l)|X) - Eg(Y^l)]^2, \\ E[h(Y^r) - Eh(Y^r)]^2 &= E[h(Y^r) - E(h(Y^r)|X)]^2 + E[E(h(Y^r)|X) - Eh(Y^r)]^2. \end{aligned}$$

Following the same idea, by adding and subtracting $E(Y^m|X)$ from $Y^m - EY^m$ and $E(g(Y^l)|X)$ from $g(Y^l) - Eg(Y^l)$,

$$\begin{aligned} E[(Y^m - EY^m)g(Y^l) - Eg(Y^l)] &= E[(Y^m - E(Y^m|X) + E(Y^m|X) - EY^m)] \\ &\quad \times [(g(Y^l) - E(g(Y^l)|X) + E(g(Y^l)|X) - Eg(Y^l))] \\ &= E[(Y^m - E(Y^m|X))(g(Y^l) - E(g(Y^l)|X))] \\ &\quad +E[(E(Y^m|X) - EY^m)(E(g(Y^l)|X) - Eg(Y^l))] \\ &\quad +E[(Y^m - E(Y^m|X))(E(g(Y^l)|X) - Eg(Y^l))] \\ &\quad +E[(g(Y^l) - E(g(Y^l)|X))(E(Y^m|X) - EY^m)]. \end{aligned}$$

Trough simple passages it is easy to check that

$$\begin{aligned} E[(Y^m - EY^m)g(Y^l) - Eg(Y^l)] &= E[(Y^m - E(Y^m|X))(g(Y^l) - E(g(Y^l)|X))] \\ &\quad + E[(E(Y^m|X) - EY^m)(E(g(Y^l)|X) - Eg(Y^l))] \\ &\quad + E\{[E(Y^m|X) - EY^m]E[(g(Y^l) - E(g(Y^l)|X))|X]\} \\ &\quad + E\{[E(g(Y^l)|X) - Eg(Y^l)]E[(Y^m - E(Y^m|X))|X]\}, \end{aligned}$$

that is

$$\begin{aligned} E[(Y^m - EY^m)g(Y^l) - Eg(Y^l)] &= E[(Y^m - E(Y^m|X))(g(Y^l) - E(g(Y^l)|X))] \\ &\quad + E[(E(Y^m|X) - EY^m)(E(g(Y^l)|X) - Eg(Y^l))]. \end{aligned}$$

Analogously

$$\begin{aligned} E[(Y^m - EY^m)h(Y^r) - Eh(Y^r)] &= E[(Y^m - E(Y^m|X))(h(Y^r) - E(h(Y^r)|X))] \\ &\quad + E[(E(Y^m|X) - EY^m)(E(h(Y^r)|X) - Eh(Y^r))]. \end{aligned}$$

Since

$$\begin{aligned} E\left[D^2_{\lambda\rho}(\widetilde{Y}-E(\widetilde{Y}|X))\right] &= 3E[Y^m-E(Y^m|X)]^2 \\ &\quad +\lambda^2 E[g(Y^l)-E(g(Y^l)|X)]^2 \\ &\quad +\rho^2 E[h(Y^r)-E(h(Y^r)|X)]^2 \\ &\quad -2\lambda E[(Y^m-E(Y^m|X))(g(Y^l)-E(g(Y^l)|X))] \\ &\quad +2\rho E[(Y^m-E(Y^m|X))(h(Y^r)-E(h(Y^r)|X))] \end{aligned}$$

and

$$\begin{aligned} E\left[D^2_{\lambda\rho}(E(\widetilde{Y}|X)-E\widetilde{Y})\right] &= 3E[E(Y^m|X)-EY^m]^2 \\ &\quad +\lambda^2 E[E(g(Y^l)|X)-Eg(Y^l)]^2 \\ &\quad +\rho^2 E[E(h(Y^r)|X)-Eh(Y^r)]^2 \\ &\quad -2\lambda E[(E(Y^m|X)-EY^m)(E(g(Y^l)|X)-Eg(Y^l))] \\ &\quad +2\rho E[(E(Y^m|X)-EY^m)(E(h(Y^r)|X)-Eh(Y^r))], \end{aligned}$$

then, we get the thesis

$$E\left[D^2_{\lambda\rho}(\widetilde{Y}-E\widetilde{Y})\right]=E\left[D^2_{\lambda\rho}(\widetilde{Y}-E(\widetilde{Y}|X))\right]+E\left[D^2_{\lambda\rho}(E(\widetilde{Y}|X)-E\widetilde{Y})\right].$$

□

Starting from the above decomposition we can define a determination coefficient.

Definition 3.2.1 *Let Y be the LR FRV of the linear model (3.1), by indicating $\widetilde{Y}=(Y^m,g(Y^l),h(Y^l))$, the determination coefficient can be defined as follows*

$$R^2 = \frac{E\left[D^2_{\lambda\rho}(E(\widetilde{Y}|X),E\widetilde{Y})\right]}{E\left[D^2_{\lambda\rho}(\widetilde{Y},E\widetilde{Y})\right]}=1-\frac{E\left[D^2_{\lambda\rho}(\widetilde{Y},E(\widetilde{Y}|X))\right]}{E\left[D^2_{\lambda\rho}(\widetilde{Y},E\widetilde{Y})\right]}. \tag{3.3}$$

This coefficient represents approximately the part of the total variation of Y explained by the regression model and for this reason it can be used to quantify the degree of linear relationship. Furthermore, it takes values in the interval $[0,1]$. If $R^2=0$ it is the case of *linear independence*, that is, the regression model does not explain any variability of the imprecise response variable. When R^2 is equal to 1, it is the case of the best fit, that is, the regression model explains completely the variability of the response variable.

3.3 The estimation problem

In order to get the estimators of the regression parameters the Least Squares (LS) criterion will be used. It consists in minimizing the squared distance between the observed values and the values predicted by the model.

3.3.1 The minimization problem

The minimization problem is defined by means of the generalized Yang-Ko metric $D^2_{\lambda\rho}$. As it was previously mentioned, the use of the LS criterion is justified by the properties of the variance proved in Proposition 2.2.1, among which we find the *Frèchet principle*.

In this case the LS problem consists in looking for $\hat{a}_m$, $\hat{a}_l$, $\hat{a}_r$, $\hat{b}_m$, $\hat{b}_l$ e $\hat{b}_r$ in order to

$$\min \Delta^2_{\lambda\rho} = \min \sum_{i=1}^{n} D^2_{\lambda\rho}((Y_i^m, g(Y_i^l), h(Y_i^r)), ((Y^m)_i^*, g^*(Y_i^l), h^*(Y_i^r))) \tag{3.4}$$

where $(Y^m)_i^* = a_m X_i + b_m$, $g^*(Y_i^l) = a_l X_i + b_l$ and $h^*(Y_i^r) = a_r X_i + b_r$ are the predicted values.

The function to minimize becomes

$$\begin{aligned} \Delta^2_{\lambda\rho} &= \sum_{i=1}^{n} \left[3(Y_i^m - a_m X_i - b_m)^2\right] \\ &+ \sum_{i=1}^{n} \left[\lambda^2 (g(Y_i^l) - a_l X_i - b_l)^2 + \rho^2 (h(Y_i^r) - a_r X_i - b_r)^2\right] \\ &+ \sum_{i=1}^{n} \left[-2\lambda (Y_i^m - a_m X_i - b_m)(g(Y_i^l) - a_l X_i - b_l)\right] \\ &+ \sum_{i=1}^{n} \left[+2\rho (Y_i^m - a_m X_i - b_m)(h(Y_i^r) - a_r X_i - b_r)\right]. \end{aligned} \tag{3.5}$$

3.3.2 Least squares estimators

In Proposition 3.3.1 the optimization problem (3.4) is solved.

Proposition 3.3.1 *The solutions of the LS problem are*

$$\hat{a}_m = \frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}^2_X}, \qquad \hat{a}_l = \frac{\hat{\sigma}_{Xg(Y^l)}}{\hat{\sigma}^2_X}, \qquad \hat{a}_r = \frac{\hat{\sigma}_{Xh(Y^r)}}{\hat{\sigma}^2_X},$$

$$\hat{b}_m = \frac{\sum_{i=1}^{n} Y_i^m}{n} - \hat{a}_m \frac{\sum_{i=1}^{n} X_i}{n}, \quad \hat{b}_l = \frac{\sum_{i=1}^{n} g(Y_i^l)}{n} - \hat{a}_l \frac{\sum_{i=1}^{n} X_i}{n}, \quad \hat{b}_r = \frac{\sum_{i=1}^{n} h(Y_i^r)}{n} - \hat{a}_r \frac{\sum_{i=1}^{n} X_i}{n}.$$

Proof. Since the functions to minimize are continuous and convex, to solve the minimization problem we equate to zero the partial derivative with respect to the parameter to be estimated. As in classical regression, we start with the constant elements of the model, because their estimators are expressed in terms of the regression coefficients.

The function to be minimized can be written as

$$\begin{aligned}
\Delta^2_{\lambda\rho} &= \sum_{i=1}^{n}\left[3\left((Y_i^m)^2 + a_m^2X_i^2 + b_m^2 - 2a_mY_i^mX_i - 2b_mY_i^m + 2a_mb_mX_i\right)\right] \quad (3.6)\\
&+ \sum_{i=1}^{n}\left[\lambda^2\left((g(Y_i^l))^2 + a_l^2X_i^2 + b_l^2 - 2a_lg(Y_i^l)X_i - 2b_lg(Y_i^l) + 2a_lb_lX_i\right)\right]\\
&+ \sum_{i=1}^{n}\left[\rho^2\left((h(Y_i^r))^2 + a_r^2X_i^2 + b_r^2 - 2a_rh(Y_i^r)X_i - 2b_rh(Y_i^r) + 2a_rb_rX_i\right)\right]\\
&- \sum_{i=1}^{n}\left[2\lambda\left(Y_i^mg(Y_i^l) - a_lY_i^mX_i - b_lY_i^m - a_mg(Y_i^l)X_i + a_ma_lX_i^2\right)\right]\\
&- \sum_{i=1}^{n}\left[2\lambda\left(a_mb_lX_i - b_mg(Y_i^l) + a_lb_mX_i + b_mb_l\right)\right]\\
&+ \sum_{i=1}^{n}\left[2\rho\left(Y_i^mh(Y_i^r) - a_rY_i^mX_i - b_rY_i^m - a_mh(Y_i^r)X_i + a_ma_rX_i^2\right)\right]\\
&+ \sum_{i=1}^{n}\left[2\rho\left(a_mb_rX_i - b_mh(Y_i^r) + a_rb_mX_i + b_mb_r\right)\right].
\end{aligned}$$

To estimate b_l we equate to zero the partial derivative of $\Delta^2_{\lambda\rho}$ with respect to b_l, that is

$$\begin{aligned}
\frac{\partial\Delta^2_{\lambda\rho}}{\partial b_l} = 0 \quad\Leftrightarrow\quad & 2\lambda^2nb_l - 2\lambda^2\sum_{i=1}^{n}g(Y_i^l) + 2\lambda^2a_l\sum_{i=1}^{n}X_i\\
& +2\lambda\sum_{i=1}^{n}Y_i^m - 2\lambda a_m\sum_{i=1}^{n}X_i - 2\lambda nb_m = 0\\
\Leftrightarrow\quad & b_l = \frac{\sum_{i=1}^{n}g(Y_i^l)}{n} - a_l\frac{\sum_{i=1}^{n}X_i}{n} - \frac{1}{\lambda}\frac{\sum_{i=1}^{n}Y_i^m}{n} + \frac{a_m}{\lambda}\frac{\sum_{i=1}^{n}X_i}{n} + \frac{b_m}{\lambda}. \quad (3.7)
\end{aligned}$$

By following the same procedure, and equating to zero the partial derivative of $\Delta^2_{\lambda\rho}$ with respect to b_r we have

$$\begin{aligned}
\frac{\partial\Delta^2_{\lambda\rho}}{\partial b_r} = 0 \quad\Leftrightarrow\quad & 2\rho^2nb_r - 2\rho^2\sum_{i=1}^{n}h(Y_i^r) + 2\rho^2a_r\sum_{i=1}^{n}X_i\\
& -2\rho\sum_{i=1}^{n}Y_i^m + 2\rho a_m\sum_{i=1}^{n}X_i + 2\rho nb_m = 0
\end{aligned}$$

$$\Leftrightarrow \quad b_r = \frac{\sum_{i=1}^{n} h(Y_i^r)}{n} - a_r \frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{\rho}\frac{\sum_{i=1}^{n} Y_i^m}{n} - \frac{a_m}{\rho}\frac{\sum_{i=1}^{n} X_i}{n} - \frac{b_m}{\rho}. \quad (3.8)$$

To estimate b_m we have to take into account that b_l and b_r obtained above are expressed as functions of b_m. Thus, by substituting (3.7) and (3.8) in (3.6), we obtain

$$\begin{aligned}
\Delta^2_{\lambda\rho} \;=\; & \sum_{i=1}^{n} \left[3\left((Y_i^m)^2 + a_m^2 X_i^2 + b_m^2 - 2a_m Y_i^m X_i - 2b_m Y_i^m + 2a_m b_m X_i\right)\right] \\
& + \lambda^2 \sum_{i=1}^{n} \left[(g(Y_i^l))^2 + a_l^2 X_i^2 + \left(\frac{\sum_{i=1}^{n} g(Y_i^l)}{n} - a_l \frac{\sum_{i=1}^{n} X_i}{n} - \frac{1}{\lambda}\frac{\sum_{i=1}^{n} Y_i^m}{n} + \frac{a_m}{\lambda}\frac{\sum_{i=1}^{n} X_i}{n} + \frac{b_m}{\lambda} \right)^2 \right] \\
& + \lambda^2 \sum_{i=1}^{n} \left[-2a_l g(Y_i^l) X_i - 2 \left(\frac{\sum_{i=1}^{n} g(Y_i^l)}{n} - a_l \frac{\sum_{i=1}^{n} X_i}{n} - \frac{1}{\lambda}\frac{\sum_{i=1}^{n} Y_i^m}{n} + \frac{a_m}{\lambda}\frac{\sum_{i=1}^{n} X_i}{n} + \frac{b_m}{\lambda} \right) g(Y_i^l) \right] \\
& + \lambda^2 \sum_{i=1}^{n} \left[+2a_l \left(\frac{\sum_{i=1}^{n} g(Y_i^l)}{n} - a_l \frac{\sum_{i=1}^{n} X_i}{n} - \frac{1}{\lambda}\frac{\sum_{i=1}^{n} Y_i^m}{n} + \frac{a_m}{\lambda}\frac{\sum_{i=1}^{n} X_i}{n} + \frac{b_m}{\lambda} \right) X_i \right] \\
& + \rho^2 \sum_{i=1}^{n} \left[(h(Y_i^r))^2 + a_r^2 X_i^2 + \left(\frac{\sum_{i=1}^{n} h(Y_i^r)}{n} - a_r \frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{\rho}\frac{\sum_{i=1}^{n} Y_i^m}{n} - \frac{a_m}{\rho}\frac{\sum_{i=1}^{n} X_i}{n} - \frac{b_m}{\rho} \right)^2 \right] \\
& + \rho^2 \sum_{i=1}^{n} \left[-2a_r h(Y_i^r) X_i - 2 \left(\frac{\sum_{i=1}^{n} h(Y_i^r)}{n} - a_r \frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{\rho}\frac{\sum_{i=1}^{n} Y_i^m}{n} - \frac{a_m}{\rho}\frac{\sum_{i=1}^{n} X_i}{n} - \frac{b_m}{\rho} \right) h(Y_i^r) \right] \\
& + \rho^2 \sum_{i=1}^{n} \left[+2a_r \left(\frac{\sum_{i=1}^{n} h(Y_i^r)}{n} - a_r \frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{\rho}\frac{\sum_{i=1}^{n} Y_i^m}{n} - \frac{a_m}{\rho}\frac{\sum_{i=1}^{n} X_i}{n} - \frac{b_m}{\rho} \right) X_i \right] \\
& -2\lambda \sum_{i=1}^{n} \left[Y_i^m g(Y_i^l) - a_l Y_i^m X_i - a_m g(Y_i^l) X_i + a_m a_l X_i^2 \right] \\
& -2\lambda \sum_{i=1}^{n} \left[- \left(\frac{\sum_{i=1}^{n} g(Y_i^l)}{n} - a_l \frac{\sum_{i=1}^{n} X_i}{n} - \frac{1}{\lambda}\frac{\sum_{i=1}^{n} Y_i^m}{n} + \frac{a_m}{\lambda}\frac{\sum_{i=1}^{n} X_i}{n} + \frac{b_m}{\lambda} \right) Y_i^m \right] \\
& -2\lambda \sum_{i=1}^{n} \left[a_m \left(\frac{\sum_{i=1}^{n} g(Y_i^l)}{n} - a_l \frac{\sum_{i=1}^{n} X_i}{n} - \frac{1}{\lambda}\frac{\sum_{i=1}^{n} Y_i^m}{n} + \frac{a_m}{\lambda}\frac{\sum_{i=1}^{n} X_i}{n} + \frac{b_m}{\lambda} \right) X_i - b_m g(Y_i^l) \right] \\
& -2\lambda \sum_{i=1}^{n} \left[+a_l b_m X_i + b_m \left(\frac{\sum_{i=1}^{n} g(Y_i^l)}{n} - a_l \frac{\sum_{i=1}^{n} X_i}{n} - \frac{1}{\lambda}\frac{\sum_{i=1}^{n} Y_i^m}{n} + \frac{a_m}{\lambda}\frac{\sum_{i=1}^{n} X_i}{n} + \frac{b_m}{\lambda} \right) \right]
\end{aligned}$$

$$
\begin{aligned}
&+2\rho\sum_{i=1}^{n}\left[Y_i^m h(Y_i^r) - a_r Y_i^m X_i - a_m h(Y_i^r) X_i + a_m a_r X_i^2\right] \\
&+2\rho\sum_{i=1}^{n}\left[-\left(\frac{\sum_{i=1}^{n} h(Y_i^r)}{n} - a_r\frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{\rho}\frac{\sum_{i=1}^{n} Y_i^m}{n} - \frac{a_m}{\rho}\frac{\sum_{i=1}^{n} X_i}{n} - \frac{b_m}{\rho}\right) Y_i^m\right] \\
&+2\rho\sum_{i=1}^{n}\left[a_m\left(\frac{\sum_{i=1}^{n} h(Y_i^r)}{n} - a_r\frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{\rho}\frac{\sum_{i=1}^{n} Y_i^m}{n} - \frac{a_m}{\rho}\frac{\sum_{i=1}^{n} X_i}{n} - \frac{b_m}{\rho}\right) X_i - b_m h(Y_i^r)\right] \\
&+2\rho\sum_{i=1}^{n}\left[+a_r b_m X_i + b_m\left(\frac{\sum_{i=1}^{n} h(Y_i^r)}{n} - a_r\frac{\sum_{i=1}^{n} X_i}{n} + \frac{1}{\rho}\frac{\sum_{i=1}^{n} Y_i^m}{n} - \frac{a_m}{\rho}\frac{\sum_{i=1}^{n} X_i}{n} - \frac{b_m}{\rho}\right)\right].
\end{aligned}
$$

By equating to zero the partial derivative of $\Delta^2_{\lambda\rho}$ with respect to b_m we get

$$
\begin{aligned}
\frac{\partial\Delta^2_{\lambda\rho}}{\partial b_m} = 0 \quad\Leftrightarrow\quad & +6nb_m - 6\sum_{i=1}^{n} Y_i^m + 6a_m\sum_{i=1}^{n} X_i \\
& +2nb_m + 2\lambda\sum_{i=1}^{n} g(Y_i^l) - 2\lambda a_l\sum_{i=1}^{n} X_i - 2\sum_{i=1}^{n} Y_i^m \\
& +2a_m\sum_{i=1}^{n} X_i - 2\lambda\sum_{i=1}^{n} g(Y_i^l) + 2\lambda a_l\sum_{i=1}^{n} X_i \\
& +2nb_m - 2\rho\sum_{i=1}^{n} h(Y_i^r) + 2\rho a_r\sum_{i=1}^{n} X_i - 2\sum_{i=1}^{n} Y_i^m \\
& +2a_m\sum_{i=1}^{n} X_i + 2\rho\sum_{i=1}^{n} h(Y_i^r) - 2\rho a_r\sum_{i=1}^{n} X_i \\
& +2\sum_{i=1}^{n} Y_i^m - 2a_m\sum_{i=1}^{n} X_i + 2\lambda\sum_{i=1}^{n} g(Y_i^l) - 2\lambda a_l\sum_{i=1}^{n} X_i \\
& -4nb_m - 2\lambda\sum_{i=1}^{n} g(Y_i^l) + 2\lambda a_l\sum_{i=1}^{n} X_i + 2\sum_{i=1}^{n} Y_i^m - 2a_m\sum_{i=1}^{n} X_i \\
& +2\sum_{i=1}^{n} Y_i^m - 2a_m\sum_{i=1}^{n} X_i - 2\rho\sum_{i=1}^{n} h(Y_i^r) + 2\rho a_r\sum_{i=1}^{n} X_i \\
& -4nb_m + 2\rho\sum_{i=1}^{n} h(Y_i^r) - 2\rho a_r\sum_{i=1}^{n} X_i + 2\sum_{i=1}^{n} Y_i^m - 2a_m\sum_{i=1}^{n} X_i = 0.
\end{aligned}
$$

As results we obtain the following solutions that depend on the parameters a_m, a_l and a_r

$$
b_m \;=\; \frac{\sum_{i=1}^{n} Y_i^m}{n} - a_m\frac{\sum_{i=1}^{n} X_i}{n}
$$

$$\begin{aligned} b_l &= \frac{\sum_{i=1}^{n} g(Y_i^l)}{n} - a_l \frac{\sum_{i=1}^{n} X_i}{n} \\ b_r &= \frac{\sum_{i=1}^{n} h(Y_i^r)}{n} - a_r \frac{\sum_{i=1}^{n} X_i}{n}. \end{aligned}$$

For this reason in the estimation of a_l, a_r and a_m we have to take into account the values obtained above.

If we consider the centered values

$$\widetilde{Y_i^m} = Y_i^m - \frac{\sum_{i=1}^{n} Y_i^m}{n} \qquad \widetilde{X_i} = X_i - \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\widetilde{g(Y_i^l)} = g(Y_i^l) - \frac{\sum_{i=1}^{n} g(Y_i^l)}{n} \qquad \widetilde{h(Y_i^r)} = h(Y_i^r) - \frac{\sum_{i=1}^{n} h(Y_i^r)}{n}$$

the objective function can be written as follows

$$\begin{aligned} \Delta^2_{\lambda\rho} &= \sum_{i=1}^{n} \left[3(\widetilde{Y_i^m} - a_m \widetilde{X_i})^2 \right] \\ &+ \sum_{i=1}^{n} \left[\lambda^2 (\widetilde{g(Y_i^l)} - a_l \widetilde{X_i})^2 + \rho^2 (\widetilde{h(Y_i^r)} - a_r \widetilde{X_i})^2 \right] \\ &+ \sum_{i=1}^{n} \left[-2\lambda (\widetilde{Y_i^m} - a_m \widetilde{X_i})(\widetilde{g(Y_i^l)} - a_l \widetilde{X_i}) \right] \\ &+ \sum_{i=1}^{n} \left[+2\rho (\widetilde{Y_i^m} - a_m \widetilde{X_i})(\widetilde{h(Y_i^r)} - a_r \widetilde{X_i}) \right]. \end{aligned} \tag{3.9}$$

By equating to zero the partial derivative of $\Delta^2_{\lambda\rho}$ with respect to a_l we obtain

$$\begin{aligned} \frac{\partial \Delta^2_{\lambda\rho}}{\partial a_l} = 0 \quad \Leftrightarrow \quad & -2\lambda^2 \sum_{i=1}^{n} \widetilde{X_i} (\widetilde{g(Y_i^l)} - a_l \widetilde{X_i}) \\ & +2\lambda \sum_{i=1}^{n} \widetilde{X_i} (\widetilde{Y_i^m} - a_m \widetilde{X_i}) = 0 \\ \Leftrightarrow \quad & a_l = \frac{\frac{1}{n} \sum_{i=1}^{n} \widetilde{X_i} \widetilde{g(Y_i^l)}}{\frac{1}{n} \sum_{i=1}^{n} \widetilde{X_i}^2} - \frac{1}{\lambda} \frac{\frac{1}{n} \sum_{i=1}^{n} \widetilde{Y_i^m} \widetilde{X_i}}{\frac{1}{n} \sum_{i=1}^{n} \widetilde{X_i}^2} + \frac{a_m}{\lambda} \\ & a_l = \frac{\hat{\sigma}_{X g(Y^l)}}{\hat{\sigma}^2_X} - \frac{1}{\lambda} \frac{\hat{\sigma}_{X Y^m}}{\hat{\sigma}^2_X} + \frac{a_m}{\lambda}. \end{aligned}$$

Analogously, by equating to zero the partial derivative of $\Delta^2_{\lambda\rho}$ with respect to a_r, we have

$$\frac{\partial \Delta^2_{\lambda\rho}}{\partial a_r} = 0 \quad \Leftrightarrow \quad -2\rho^2 \sum_{i=1}^{n} \widetilde{X_i} (\widetilde{h(Y_i^r)} - a_r \widetilde{X_i})$$

$$
\begin{aligned}
& -2\rho \sum_{i=1}^{n} \widetilde{X_i}(\widetilde{Y_i^m} - a_m \widetilde{X_i}) = 0 \\
\Leftrightarrow \quad & a_r = \frac{\frac{1}{n}\sum_{i=1}^{n} \widetilde{X_i}\widetilde{h(Y_i^r)}}{\frac{1}{n}\sum_{i=1}^{n} \widetilde{X_i}^2} - \frac{1}{\rho}\frac{\frac{1}{n}\sum_{i=1}^{n} \widetilde{Y_i^m}\widetilde{X_i}}{\frac{1}{n}\sum_{i=1}^{n} \widetilde{X_i}^2} + \frac{a_m}{\rho} \\
& a_r = \frac{\hat{\sigma}_{Xh(Y^r)}}{\hat{\sigma}_X^2} + \frac{1}{\rho}\frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2} - \frac{a_m}{\rho}.
\end{aligned}
$$

By substituting into (3.10) a_l and a_r obtained above, the objective function becomes

$$
\begin{aligned}
\Delta_{\lambda\rho}^2 \quad = \quad & \sum_{i=1}^{n} \left[3(\widetilde{Y_i^m} - a_m \widetilde{X_i})^2 \right] \\
+ \quad & \sum_{i=1}^{n} \left[\lambda^2 \left(\widetilde{g(Y_i^l)} - \widetilde{X_i} \left(\frac{\hat{\sigma}_{Xg(Y^l)}}{\hat{\sigma}_X^2} - \frac{1}{\lambda}\frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2} + \frac{a_m}{\lambda} \right) \right)^2 \right] \\
+ \quad & \sum_{i=1}^{n} \left[\rho^2 \left(\widetilde{h(Y_i^r)} - \widetilde{X_i} \left(\frac{\hat{\sigma}_{Xh(Y^r)}}{\hat{\sigma}_X^2} + \frac{1}{\rho}\frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2} - \frac{a_m}{\rho} \right) \right)^2 \right] \\
+ \quad & \sum_{i=1}^{n} \left[-2\lambda(\widetilde{Y_i^m} - a_m \widetilde{X_i}) \left(\widetilde{g(Y_i^l)} - \widetilde{X_i} \left(\frac{\hat{\sigma}_{Xg(Y^l)}}{\hat{\sigma}_X^2} - \frac{1}{\lambda}\frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2} + \frac{a_m}{\lambda} \right) \right) \right] \\
+ \quad & \sum_{i=1}^{n} \left[+2\rho(\widetilde{Y_i^m} - a_m \widetilde{X_i}) \left(\widetilde{h(Y_i^r)} - \widetilde{X_i} \left(\frac{\hat{\sigma}_{Xh(Y^r)}}{\hat{\sigma}_X^2} + \frac{1}{\rho}\frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2} - \frac{a_m}{\rho} \right) \right) \right].
\end{aligned}
$$

Finally, by equating to zero the partial derivative of $\Delta_{\lambda\rho}^2$ with respect to a_m we obtain the estimation of a_m, that is

$$
\begin{aligned}
\frac{\partial \Delta_{\lambda\rho}^2}{\partial a_m} = 0 \quad \Leftrightarrow \quad & -6 \sum_{i=1}^{n} \left[\widetilde{X_i}(\widetilde{Y_i^m} - a_m \widetilde{X_i}) \right] \\
& -2\lambda \sum_{i=1}^{n} \left[\widetilde{X_i} \left(\widetilde{g(Y_i^l)} - \widetilde{X_i}\frac{\hat{\sigma}_{Xg(Y^l)}}{\hat{\sigma}_X^2} + \frac{\widetilde{X_i}}{\lambda}\frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2} - a_m \frac{\widetilde{X_i}}{\lambda} \right) \right] \\
& +2\rho \sum_{i=1}^{n} \left[\widetilde{X_i} \left(\widetilde{h(Y_i^r)} - \widetilde{X_i}\frac{\hat{\sigma}_{Xh(Y^r)}}{\hat{\sigma}_X^2} - \frac{\widetilde{X_i}}{\rho}\frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2} + a_m \frac{\widetilde{X_i}}{\rho} \right) \right] \\
& +2\lambda \sum_{i=1}^{n} \left[\widetilde{X_i} \left(\widetilde{g(Y_i^l)} - \widetilde{X_i}\frac{\hat{\sigma}_{Xg(Y^l)}}{\hat{\sigma}_X^2} + \frac{\widetilde{X_i}}{\lambda}\frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2} - a_m \frac{\widetilde{X_i}}{\lambda} \right) \right] \\
& -2\rho \sum_{i=1}^{n} \left[\widetilde{X_i} \left(\widetilde{h(Y_i^r)} - \widetilde{X_i}\frac{\hat{\sigma}_{Xh(Y^r)}}{\hat{\sigma}_X^2} - \frac{\widetilde{X_i}}{\rho}\frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2} + a_m \frac{\widetilde{X_i}}{\rho} \right) \right] \\
\Leftrightarrow \quad & \hat{a}_m = \frac{\sum_{i=1}^{n} \widetilde{Y_i^m}\widetilde{X_i}}{\sum_{i=1}^{n} \widetilde{X_i}^2} = \frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2}.
\end{aligned}
$$

The solutions are

$$\hat{a}_m = \frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2}, \qquad \hat{a}_l = \frac{\hat{\sigma}_{Xg(Y^l)}}{\hat{\sigma}_X^2}, \qquad \hat{a}_r = \frac{\hat{\sigma}_{Xh(Y^r)}}{\hat{\sigma}_X^2},$$

$$\hat{b}_m = \frac{\sum_{i=1}^n Y_i^m}{n} - \hat{a}_m \frac{\sum_{i=1}^n X_i}{n}, \qquad \hat{b}_l = \frac{\sum_{i=1}^n g(Y_i^l)}{n} - \hat{a}_l \frac{\sum_{i=1}^n X_i}{n}, \qquad \hat{b}_r = \frac{\sum_{i=1}^n h(Y_i^r)}{n} - \hat{a}_r \frac{\sum_{i=1}^n X_i}{n}.$$

□

The LS estimators, analogously to the classical case of linear regression analysis, fulfill some algebraic properties that will be analyzed in the next chapter for the general case, and some statistical properties that will be introduced and proved in Proposition 3.3.2.

Proposition 3.3.2 *The estimators $\hat{a}_m$, $\hat{a}_l$, $\hat{a}_r$, $\hat{b}_m$, $\hat{b}_l$ and $\hat{b}_r$ are unbiased and strongly consistent.*

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

$$E(\hat{a}_m|X) = E\left(\frac{\hat{\sigma}_{XY^m}}{\hat{\sigma}_X^2}\Big|X\right) = E\left(\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i^m - \bar{Y}^m)}{\sum_{i=1}^n (X_i - \bar{X})^2}\Big|X\right).$$

Since $Y_i^m = a_m X_i + b_m + \varepsilon_{mi}$ and $\overline{Y}^m = a_m \overline{X} + b_m + \overline{\varepsilon}_m$ we obtain

$$\begin{aligned} E(\hat{a}_m|X) &= \frac{\sum_{i=1}^n \left((X_i - \overline{X}) E((a_m X_i + b_m + \varepsilon_{mi} - a_m \overline{X} - b_m - \overline{\varepsilon}_m)|X)\right)}{\sum_{i=1}^n (X_i - \overline{X})^2} \\ &= \frac{\sum_{i=1}^n \left((X_i - \overline{X}) a_m (X_i - \overline{X})\right)}{\sum_{i=1}^n (X_i - \overline{X})^2} + \frac{\sum_{i=1}^n \left(E(X_i - \overline{X}) E(\varepsilon_m - \overline{\varepsilon}_m|X)\right)}{\sum_{i=1}^n (X_i - \overline{X})^2}, \end{aligned}$$

and taking into account that the conditional expectation of ε_m given X is equal to 0 and $\overline{\varepsilon}_m$ is an unbiased estimator of $E\varepsilon_m$, we get

$$E(\hat{a}_m|X) = a_m.$$

Since $E(\hat{a}_m) = E\left(E(\hat{a}_m|X)\right)$, the thesis is proved. Analogously it is possible to check that $E(\hat{a}_l) = a_l$ and $E(\hat{a}_r) = a_r$.

Furthermore, $E(\hat{b}_m|X) = E\left(\overline{Y}^m - \hat{a}_m \overline{X}|X\right) = E\left(\overline{Y}^m|X\right) - E(\overline{X}) E(\hat{a}_m|X)$, it

follows that $E(\hat{b}_m) = E(E(\hat{b}_m|X)) = E\left(E\left(\overline{Y}^m|X\right)\right) - E(\overline{X})E\left(E(\hat{a}_m|X)\right)$, that is, $E(\hat{b}_m) = E\left(\overline{Y}^m\right) - E(\overline{X})E(\hat{a}_m)$. Taking into account that the sample means are unbiased estimators of the expectations, it is easy to check that

$$E(\hat{b}_m) = E\left(Y^m\right) - a_m E(X) = b_m,$$

and, by means of similar reasoning, $E(\hat{b}_l) = b_l$ and $E(\hat{b}_r) = b_r$.

In order to analyze the consistency of the estimators with respect to the population constants of the linear model, we have to study how these estimators behave as the size of random samples increases. Since in the real case the sample moments are consistent estimators of the population moments we get the thesis, i.e.,

- $\hat{a}_m = \frac{\hat{\sigma}_{xY^m}}{\hat{\sigma}_x^2} \overset{n\to\infty}{\longrightarrow} \frac{\sigma_{xY^m}}{\sigma_x^2} = a_m \qquad a.s. - [P]$
- $\hat{a}_l = \frac{\hat{\sigma}_{xg(Y^l)}}{\hat{\sigma}_x^2} \overset{n\to\infty}{\longrightarrow} \frac{\sigma_{xg(Y^l)}}{\sigma_x^2} = a_l \qquad a.s. - [P]$
- $\hat{a}_r = \frac{\hat{\sigma}_{xh(Y^r)}}{\hat{\sigma}_x^2} \overset{n\to\infty}{\longrightarrow} \frac{\sigma_{xh(Y^r)}}{\sigma_x^2} = a_r \qquad a.s. - [P]$
- $\hat{b}_m = \overline{Y}^m - \hat{a}_m\overline{X} \overset{n\to\infty}{\longrightarrow} EY^m - a_m EX = b_m \qquad a.s. - [P]$
- $\hat{b}_l = \overline{g(Y^l)} - \hat{a}_l\overline{X} \overset{n\to\infty}{\longrightarrow} Eg(Y^l) - a_l EX = b_l \qquad a.s. - [P]$
- $\hat{b}_r = \overline{h(Y^r)} - \hat{a}_r\overline{X} \overset{n\to\infty}{\longrightarrow} Eh(Y^r) - a_r EX = b_r \qquad a.s. - [P]$

□

In order to develop inferences it is useful to provide an approximation to the distribution of the estimators. A typical approximation is the asymptotic distribution of the estimators.

Proposition 3.3.3 *Under the assumptions of model (3.1), as $n \to \infty$,*

$$\sqrt{n}\begin{pmatrix} \hat{a}_m - a_m \\ \hat{a}_l - a_l \\ \hat{a}_r - a_r \end{pmatrix} \xrightarrow{D} N\left(\underline{0}', \frac{\Sigma}{\sigma_X^2}\right).$$

Proof. Starting from the expression of $\hat{a}_m$, $\hat{a}_l$ and $\hat{a}_r$ in terms of sample moments

$$\begin{pmatrix} \hat{a}_m \\ \hat{a}_l \\ \hat{a}_r \end{pmatrix} = \begin{pmatrix} \frac{\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})(Y_i^m-\overline{Y^m})}{\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2} \\ \frac{\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})(g(Y_i^l)-\overline{g(Y^l)})}{\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2} \\ \frac{\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})(h(Y_i^r)-\overline{h(Y^r)})}{\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2} \end{pmatrix},$$

and taking into account that $Y_i^m = a_m X_i + b_m + \varepsilon_{mi}$, $g(Y_i^l) = a_l X_i + b_l + \varepsilon_{li}$, $h(Y_i^r) = a_r X_i + b_r + \varepsilon_{ri}$ and $\overline{Y^m} = a_m \overline{X} + b_m + \overline{\varepsilon}_m$, $\overline{g(Y^l)} = a_l \overline{X} + b_l + \overline{\varepsilon}_l$, $\overline{h(Y^r)} = a_r \overline{X} + b_r + \overline{\varepsilon}_r$, it is easy to check that

$$\begin{pmatrix} \hat{a}_m \\ \hat{a}_l \\ \hat{a}_r \end{pmatrix} = \begin{pmatrix} a_m + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(\varepsilon_{mi}-\overline{\varepsilon}_m)}{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2} \\ a_l + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(\varepsilon_{li}-\overline{\varepsilon}_l)}{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2} \\ a_r + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(\varepsilon_{ri}-\overline{\varepsilon}_r)}{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2} \end{pmatrix}.$$

In this way, we have that,

$$\sqrt{n}\begin{pmatrix} \hat{a}_m - a_m \\ \hat{a}_l - a_l \\ \hat{a}_r - a_r \end{pmatrix} = \left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\overline{X})(\varepsilon_{mi}-\overline{\varepsilon}_m) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\overline{X})(\varepsilon_{li}-\overline{\varepsilon}_l) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\overline{X})(\varepsilon_{ri}-\overline{\varepsilon}_r) \end{pmatrix}$$

and then,

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\overline{X})(\varepsilon_{mi}-\overline{\varepsilon}_m) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\overline{X})(\varepsilon_{li}-\overline{\varepsilon}_l) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\overline{X})(\varepsilon_{ri}-\overline{\varepsilon}_r) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\varepsilon_{mi} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\varepsilon_{li} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\varepsilon_{ri} \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\varepsilon_{mi} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\varepsilon_{li} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\varepsilon_{ri} \end{pmatrix}$$
$$- \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_m \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_l \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_r \end{pmatrix} - \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_m \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_l \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_r \end{pmatrix}.$$

Furthermore,

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\varepsilon_{mi} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\varepsilon_{li} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\varepsilon_{ri} \end{pmatrix} \xrightarrow{a.s.} \underline{0}',$$

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_m \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_l \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_r \end{pmatrix} \xrightarrow{a.s.} \underline{0}',$$

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_m \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_l \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_r \end{pmatrix} \xrightarrow{a.s.} \underline{0}',$$

and $\left\{ \begin{pmatrix} (X_i-EX)\varepsilon_{mi} \\ (X_i-EX)\varepsilon_{li} \\ (X_i-EX)\varepsilon_{ri} \end{pmatrix} \right\}_{i=1,\ldots,n}$ is a sequence of random vectors i.i.d., centered at $\underline{0}'$, whose covariance matrix is $\sigma_X^2\Sigma$, so applying the Central Limit Theorem it results that

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\varepsilon_{mi} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\varepsilon_{li} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\varepsilon_{ri} \end{pmatrix} \xrightarrow{D} N(\underline{0}', \sigma_X^2\Sigma).$$

Hence

$$\sqrt{n}\begin{pmatrix} \hat{a}_m-a_m \\ \hat{a}_l-a_l \\ \hat{a}_r-a_r \end{pmatrix} \xrightarrow{D} N\left(\underline{0}', \frac{\Sigma}{\sigma_X^2}\right).$$

□

The accuracy of the estimators is analyzed by means of standard error (the square root of the variance). In presence of sampling model whose functional form has not been further specified, it is possible to use a bootstrap procedure to get an estimate of standard error (Efron & Tibshirani, 1993).

The bootstrap algorithm for estimating standard errors

Step 1 Draw B independent bootstrap samples $\{Y^{m*}_i, Y^{l*}_i, Y^{r*}_i, X^*_i\}_{i=1,\ldots,n}$ of size n with replacement from the original sample $\{Y^m_i, Y^l_i, Y^r_i, X_i\}_{i=1,\ldots,n}$.

Step 2 Compute the value of the bootstrap estimator corresponding to each bootstrap sample $\hat{T}^*_b$, $b = 1, ..., B$.

Step 2 Estimate the standard error $\widehat{se}_B$ by the sample standard deviation of the B replications,

$$\widehat{se}_B = \sqrt{\frac{\sum_{b=1}^{B}\left(\hat{T}^*_b - \hat{T}^*\right)^2}{B-1}},$$

where $\hat{T}^* = \sum_{b=1}^{B} \hat{T}_b^*/B$.

3.3.3 Simulations

In order to compare empirically the estimates obtained by means of the least squares procedure with the theoretical values, we consider a simulated situation. We have drawn a sample of 30 units in the following way. We have generated an explanatory variable X and three random variables ε_m, $_l$, ε_r normally distributed as $N(0,1)$. We have supposed that the parameters of the model are: $a_m = 2$, $a_l = 1.2$, $a_r = -2.6$, $b_m = 44$, $b_l = -12$ and $b_r = 12$. The response variables are obtained as

$$\begin{cases} Y_i^m = 2X_i + 44 + \varepsilon_{mi} \\ g(Y_i^l) = 1.2X_i - 12 + \varepsilon_{li} \\ h(Y_i^r) = -2.6X_i + 12 + \varepsilon_{ri} \end{cases} \tag{3.10}$$

for $i = 1, ..., 30$. The simulated data are shown in Table 3.1.

Table 3.1: Simulated data from Model (3.10).

Y_i^m	$g(Y_i^l)$	$h(Y_i^r)$	X_i	Y_i^m	$g(Y_i^l)$	$h(Y_i^r)$	X_i
41.6902	-13.4821	15.9795	-1.6466	47.7300	-10.3252	7.1249	1.7295
46.6195	-12.3565	11.8423	0.4287	46.9530	-11.0842	10.1590	0.7090
43.9529	-12.8048	13.3834	-0.7372	41.4519	-14.6555	14.8292	-0.7479
46.0416	-11.8437	9.6301	0.5649	45.0833	-10.0386	11.9874	0.2289
41.5585	-15.0749	14.7063	-1.3842	42.7554	-11.9408	10.9668	-0.2235
44.9902	-11.8320	11.0820	0.4603	41.9799	-12.3080	12.7149	-0.8533
43.7590	-11.7027	9.6178	0.6294	44.0890	-9.9866	11.6750	0.3456
44.3415	-11.8357	12.6159	0.3798	45.4786	-13.9330	10.8041	0.1098
41.9523	-13.5172	15.2089	-1.0133	42.5924	-14.1033	13.3146	-1.1330
43.5339	-14.0053	13.2235	-0.3472	40.5285	-12.6436	13.4170	-0.6831
43.8756	-10.3754	10.6996	0.4419	43.0834	-11.8056	12.3248	-0.2779
40.1549	-12.5841	16.4504	-1.5902	45.8631	-11.7674	9.1362	0.6548
43.1553	-12.9682	15.1674	-0.7014	39.9468	-13.1998	14.1360	-1.2484
40.6563	-14.0303	12.5639	-1.0776	42.5983	-13.9437	13.8443	-0.5975
45.2290	-10.5836	10.6871	1.0022	42.6108	-12.7679	11.3425	-0.4818

The estimated model is

$$\begin{cases} \widehat{Y^m} = 2.1652X + 43.9847 \\ \widehat{g(Y^l)} = 1.2170X + -12.1636 \\ \widehat{h(Y^r)} = -2.3047X + 11.8122 \end{cases} \tag{3.11}$$

By comparing (3.10) and (3.11), we observe that in this simulated case the estimates are quite good.

3.3.4 Empirical results

To illustrate the application of the regression model introduced in this work two examples are analyzed. The first one is referred to triangular fuzzy numbers and the second one concerns interval data.

Example 3.3.1 We consider the data of Example 3.1.1. For analyzing the part of the quality explained by the height of the trees we use the new regression model and we obtain the following estimated models

$$\begin{cases} \widehat{Y^m} &= 0.1558X + 18.7497 \\ \widehat{Y^l} &= \exp(-0.00017X + 2.5780) \\ \widehat{Y^r} &= \exp(-0.00067X + 2.6489) \end{cases} \tag{3.12}$$

The value of the estimated parameter $\hat{a}_m$ equal to 0.1558 represents a positive linear relationship between the response and the explanatory variable. In particular, the quality is expected to increase of about 0.16 for any additional cm of the height. The estimated spreads of the response variable, $\widehat{Y^l}$ and $\widehat{Y^r}$, represent the imprecision of the quality estimated by the new model. In Fig. 3.1 the extreme values of the 0-level and the single-value of the 1-level of the quality by the height are indicated, respectively, by means of the vertical segments and the dots, while the estimated centers and the estimated spreads are represented by the solid line and the dash line.

To evaluate the accuracy of these estimates we draw 800 bootstrap samples of size $n = 238$ with replacement from our data set. For each bootstrap replication we calculate the estimate of the parameters of the linear regression model. By means of the 800 replications of the estimation procedure we compute the estimate of the standard errors $\widehat{se}$ of the parameters and we check the value in Table 3.2.

Hence two kinds of uncertainty have been taken into account: the imprecision of the estimated quality and the stochastic uncertainty of the regression model represented by the values in the third column of Table 3.2.

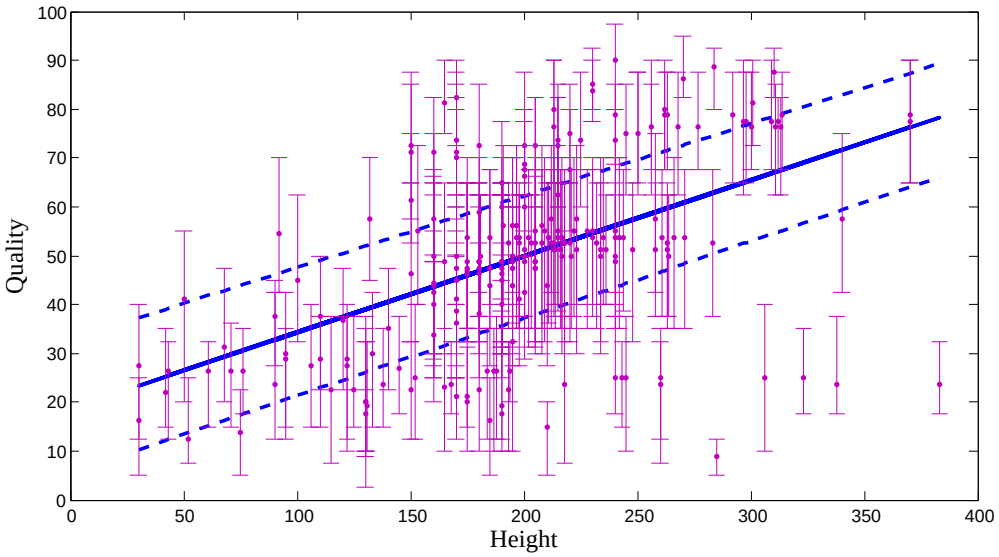

Figure 3.1: The observed extreme values of the 0-level and the single-value of the quality by the height of the trees, and the estimated linear regression models

Table 3.2: Estimation of the parameters of Model (3.12) and estimation of their standard errors.

Estimator	Estimated value	Estimate of standard error
$\hat{a}_m$	0.1558	0.0210
$\hat{a}_l$	-0.00017	0.0004
$\hat{a}_r$	-0.00067	0.0004
$\hat{b}_m$	18.7497	3.9745
$\hat{b}_l$	2.5780	0.0821
$\hat{b}_r$	2.6489	0.0839

Example 3.3.2 In this example we are interested in analyzing the dependence relationship of the Retail Trade Sales (in millions of dollars) of the U.S. in 2002 by kind of business on the number of employees (see http://www.census.gov/econ/www/). The Retail Trade Sales are intervals in the period: January 2002 through December 2002 (see Table 3.3). For each interval we consider the center and the spreads and we apply the new regression model in order to evaluate the dependence relationship. As in Example 3.3.1 we have transformed the spreads by means of the logarithmic transformation.

Table 3.3: The Retail Trade Sales and the Number of Employees of 22 kinds of Business in the U.S. in 2002.

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

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

$$\begin{cases} \widehat{Y^m} &= 0.0181X - 672.731 \\ \widehat{Y^l} &= \exp(0.000002482X + 5.9244) \\ \widehat{Y^r} &= \exp(0.000002482X + 5.9244) \end{cases} \tag{3.13}$$

The value 0.0181 indicates the strength of the relationship between the response and the explanatory variable, in particular, the retail trade sales are expected to increase of about 18.100 dollars for any additional employee.

Also in this case we evaluate the accuracy of the estimators by means of a bootstrap procedure with 800 replications.

As Table 3.4 shows, the intercept term $\hat{b}_m$ is affected by a high degree of uncer-

Table 3.4: Estimation of the parameters of Model (3.13) and estimation of their standard errors.

Estimator	Estimated value	Estimate of standard error
$\hat{a}_m$	0.0181	0.0015
$\hat{a}_l$	0.000002482	0.0000
$\hat{a}_r$	0.000002482	0.0000
$\hat{b}_m$	-672.731	412.0407
$\hat{b}_l$	5.9244	0.2151
$\hat{b}_r$	5.9244	0.2151

tainty, while the uncertainty of $\hat{a}_l$ and $\hat{a}_r$, which represent the relationship between the explanatory variable and the logarithmic transformation of the spreads of the response, is practically equal to 0. As Fig. 3.2 shows, the predicted values of the

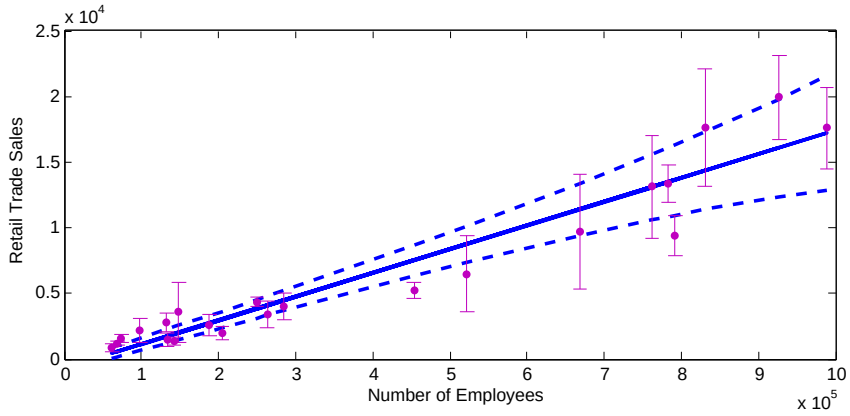

Figure 3.2: The observed interval retail trade sales by number of employees and the estimated linear regression models

spreads grow as the number of employees increases.

3.4 Confidence regions

As in classical Statistics, in this case it is useful to estimate the regression parameters not only by a single value but by a confidence interval too. These intervals represent the reliability of the estimates. How likely the interval is to contain the parameter is determined by the confidence level α. The aim of this section is to get asymptotic

confidence regions for the regression coefficients a_m, a_l and a_r. In particular, taking into account the asymptotic distribution of Proposition 3.3.3, it results

$$P\left(-c_{\alpha/2} \leq \sqrt{n}\begin{pmatrix} \hat{a}_m - a_m \\ \hat{a}_l - a_l \\ \hat{a}_r - a_r \end{pmatrix} \leq +c_{\alpha/2}\right) = 1-\alpha,$$

where the vector $c_{\alpha/2}$ defines a $\alpha/2$-quantile of a $N\left(\underline{0}', \frac{\Sigma}{\sigma_X^2}\right)$. As consequence it is easy to check that for the vector of parameters (a_m, a_l, a_r) the $100(1-\alpha)$ confidence interval is

$$\left[\begin{pmatrix} \hat{a}_m \\ \hat{a}_l \\ \hat{a}_r \end{pmatrix} - \frac{c_{\alpha/2}}{\sqrt{n}}, \begin{pmatrix} \hat{a}_m \\ \hat{a}_l \\ \hat{a}_r \end{pmatrix} + \frac{c_{\alpha/2}}{\sqrt{n}}\right]. \tag{3.14}$$

It follows that the probability that the random interval includes the theoretical parameters, as the sample size n increases, tends to $1-\alpha$.

Since $c_{\alpha/2}$ is not unique, it results not easy in practice to find confidence regions for $(a_m, a_l, a_r)'$. To face this inconvenience useful confidence bands can be found. The rule for constructing the bands may be providing a lower and an upper bound, $L(X)$ and $U(X)$, such that the probability that $[L(X), U(X)]$ contain the true vector of parameters, $(a_m, a_l, a_r)'$, is approximately equal to $1-\alpha$, that is,

$$P\left([L(X), U(X)] \supset (a_m, a_l, a_r)'\right) \approx 1-\alpha.$$

An example of confidence bands is

$$\left[\begin{pmatrix} \hat{a}_m \\ \hat{a}_l \\ \hat{a}_r \end{pmatrix} - \frac{\lambda}{\sqrt{n}}\begin{pmatrix} \frac{\Sigma_{11}}{\sigma_X^2} \\ \frac{\Sigma_{22}}{\sigma_X^2} \\ \frac{\Sigma_{33}}{\sigma_X^2} \end{pmatrix}, \begin{pmatrix} \hat{a}_m \\ \hat{a}_l \\ \hat{a}_r \end{pmatrix} + \frac{\lambda}{\sqrt{n}}\begin{pmatrix} \frac{\Sigma_{11}}{\sigma_X^2} \\ \frac{\Sigma_{22}}{\sigma_X^2} \\ \frac{\Sigma_{33}}{\sigma_X^2} \end{pmatrix}\right] \tag{3.15}$$

where λ is in $\mathbb{R}$, σ_X^2 is the variance of the explanatory variable and Σ_{11}, Σ_{22}, Σ_{33} the diagonal elements of the covariance matrix of the vector $(\varepsilon_m, \varepsilon_l, \varepsilon_r)$. The constant λ is chosen so that the probability that the (3.15) contains the theoretical vector of parameters is approximately equal to $1-\alpha$.

In the next sections the accuracy of these results is evaluated by means of simulation studies and applicative examples.

Remark 3.4.1 If we consider separately the regression parameters, that is the case of independence between the spreads and the center of the response variable, it is possible to obtain confidence intervals for each parameters. In this case smaller confidence intervals could be obtained.

Remark 3.4.2 The problem of confidence bands for the regression parameters could be analyzed also by means of a bootstrap technique.

3.4.1 Simulations

Consider a simulated situation in order to construct confidence regions for the regression parameters for different sample sizes. For each sample size n we have generated an explanatory variable X and three random variables ε_m, ε_l, ε_r normally distributed as $N(0,1)$ assuming stochastic independence among all of them. Suppose that the parameters of the model are: $a_m = 2$, $a_l = 1.2$, $a_r = -3.4$, $b_m = 10$, $b_l = 3.5$ and $b_r = 4.2$. The response variables are obtained as

$$\begin{cases} Y_i^m = 2X_i + 10 + \varepsilon_{mi} \\ g(Y_i^l) = 1.2X_i + 3.5 + \varepsilon_{li} \\ h(Y_i^r) = -3.4X_i + 4.2 + \varepsilon_{ri} \end{cases} \tag{3.16}$$

for $i = 1, .., n$.

The variables ε_m, ε_l, ε_r are independent and $c_{\alpha/2}$ in (3.14) is a $\alpha/2$-quantile of a $N\left(\underline{0}', I\right)$, where I is the identity matrix. Due to the data generation process the estimators $\hat{a}_m$, $\hat{a}_l$ and $\hat{a}_r$ are independent, so it is possible to consider separately the confidence interval for each parameter: a_m, a_l, a_r

$$\left[\hat{a}_m - \frac{t_{\alpha/2}}{\sqrt{n}}, \hat{a}_m + \frac{t_{\alpha/2}}{\sqrt{n}}\right], \quad \left[\hat{a}_l - \frac{t_{\alpha/2}}{\sqrt{n}}, \hat{a}_l + \frac{t_{\alpha/2}}{\sqrt{n}}\right], \quad \left[\hat{a}_r - \frac{t_{\alpha/2}}{\sqrt{n}}, \hat{a}_r + \frac{t_{\alpha/2}}{\sqrt{n}}\right]$$

where $t_{\alpha/2}$ is the $\alpha/2$-quantile of a $N(0,1)$. If $\alpha = 0.05$ it results $t_{\alpha/2} = 1.96$.

As shown in Table 3.5, as the sample size increases the estimates are closer to the theoretical value and the confidence regions are smaller.

Table 3.5: Estimates and Confidence Regions of the parameters of Model (3.16) for a simulation

n	$\hat{a}_m$	$\mathrm{CI}_{0.05}(\hat{a}_m)$	$\hat{a}_l$	$\mathrm{CI}_{0.05}(\hat{a}_l)$	$\hat{a}_r$	$\mathrm{CI}_{0.05}(\hat{a}_r)$
30	2.21	[1.852,2.568]	1.0106	[0.653,1.368]	-3.6879	[-4.046,-3.330]
50	1.8788	[1.602,2.156]	1.2779	[1.001,1.555]	-3.2004	[-3.478,-2.923]
100	2.0959	[1.900,2.292]	1.1307	[0.935,1.327]	-3.5721	[-3.768,-3.376]
200	2.0856	[1.947,2.224]	1.1520	[1.013,1,291]	-3.5119	[-3.651,-3.373]
500	2.0429	[1.955,2.131]	1.1751	[1.087,1.263]	-3.3324	[-3.420,-3.245]
1000	2.0112	[1.949,2.073]	1.1915	[1.129,1.254]	-3.3938	[-3.456,-3.332]
10000	1.9989	[1.979,2.018]	1.1951	[1.176,1.215]	-3.3952	[-3.415,-3.376]

Consider the simulated data set described above. The probability that each $100(1-\alpha)$-confidence interval contains the theoretical value should tend to $1-\alpha$, as the sample size n increases. This is indicated by the values of Table 3.6.

Table 3.6: Empirical confidence level of the confidence intervals.

n	$\text{Prob}(\text{CI}_{0.05}(\hat{a}_m) \supset a_m)$	$\text{Prob}(\text{CI}_{0.05}(\hat{a}_l) \supset a_l)$	$\text{Prob}(\text{CI}_{0.05}(\hat{a}_r) \supset a_r)$
50	0.9432	0.9433	0.9446
100	0.9459	0.9474	0.9469
200	0.9483	0.9489	0.9488
300	0.9487	0.9502	0.9499

3.4.2 Empirical results

To illustrate the results we have considered the data of Example 3.1.1. Confidence bands of the type (3.15) for the vector of parameters (a_m, a_l, a_r) have been constructed. The covariance matrix of the vector $(\varepsilon_m, \varepsilon_l, \varepsilon_r)$ has been replaced by the covariance matrix of the residuals $\widehat{\varepsilon_{mi}} = \widehat{Y_i^m} - Y_i^m, \widehat{\varepsilon_{li}} = \widehat{g(Y_i^l)} - g(Y_i^l), \widehat{\varepsilon_{ri}} = \widehat{h(Y_i^r)} - h(Y_i^r)$, and it results

$$\Sigma_{\widehat{\underline{\varepsilon}}} = \begin{pmatrix} 264.7 & 0 & -0.3 \\ 0 & 0.1 & 0 \\ -0.3 & 0 & 0.1 \end{pmatrix}.$$

The variance of the explanatory variable, σ_X^2, has been estimated by means of the sample variance $\widehat{\sigma_X}^2 = 3715.9$.

Through empirical trials a constant λ equal to 6300, that can be used to obtain a confidence band of level $\alpha = 0.05$, has been found, that is,

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

3.5 Hypothesis testing on the regression parameters

The parameters a_m, a_l and a_r of Model (3.1) represent the strength of the relationship between the response variables Y^m, $g(Y^l)$, $h(Y^r)$ and the explanatory one X.

Testing the explicative power of X consists in testing that the coefficients a_m, a_l and a_r are equal to 0. In general it is possible to test the null hypothesis

$$H_0 : \begin{pmatrix} a_m \\ a_l \\ a_r \end{pmatrix} = \begin{pmatrix} k_m \\ k_l \\ k_r \end{pmatrix} \tag{3.17}$$

against the alternative

$$H_1 : \begin{pmatrix} a_m \\ a_l \\ a_r \end{pmatrix} \neq \begin{pmatrix} k_m \\ k_l \\ k_r \end{pmatrix},$$

where k_m, k_l, and k_r are constant values in $\mathbb{R}$, on the basis of the available sample information. As test statistic we can use

$$T_n = V_n' V_n, \tag{3.18}$$

where

$$V_n = \sqrt{n} \begin{pmatrix} \hat{a}_m - k_m \\ \hat{a}_l - k_l \\ \hat{a}_r - k_r \end{pmatrix}.$$

In the next section, through an asymptotic approach, we will test H_0 against H_1.

3.5.1 Asymptotic approach

Since previously we have proved that, under H_0,

$$V_n \xrightarrow{D} N\left(\underline{0}', \frac{\Sigma}{\sigma_X^2}\right),$$

it follows that

$$T_n \xrightarrow{D} f_1(V),$$

where $V \sim N\left(\underline{0}', \frac{\Sigma}{\sigma_X^2}\right)$ and $f_1(A) = A'A$. Based on it, a rejection region for the null hypothesis (3.17) has been defined.

Proposition 3.5.1 *In testing the null hypothesis (3.17) at the nominal significance level α, H_0 should be rejected if*

$$T_n = V_n' V_n > c_\alpha,$$

where c_α is a α-quantile of the asymptotic distribution of T_n.

As usual this asymptotic test works suitably for samples with very large size, and for this reason we propose a bootstrap test in the next section.

3.5.2 Bootstrap approach

To get a bootstrap population fulfilling the null hypothesis, the new variables $Z^m = Y^m - \hat{a}_m X + k_m X$, $Z^l = g(Y^l) - \hat{a}_l X + k_l X$ and $Z^r = h(Y^r) - \hat{a}_r X + k_r X$ are considered. Then, a sample of size n with replacement $\left\{(X_i^*, Z^{m*}{}_i, Z^{l*}{}_i, Z^{r*}{}_i)\right\}_{i=1,\ldots,n}$ from the bootstrap population is drawn. The bootstrap statistic is

$$T_n^* = V_n^{*'} V_n^*,$$

where

$$V_n^* = \sqrt{n} \begin{pmatrix} \hat{a}_m^* - k_m \\ \hat{a}_l^* - k_l \\ \hat{a}_r^* - k_r \end{pmatrix}.$$

and

$$\hat{a}_m^* = \frac{\hat{\sigma}_{X^* Z^{m*}}}{\hat{\sigma}^2_{X^*}}, \qquad \hat{a}_l^* = \frac{\hat{\sigma}_{X^* Z^{l*}}}{\hat{\sigma}^2_{X^*}}, \qquad \hat{a}_r^* = \frac{\hat{\sigma}_{X^* Z^{r*}}}{\hat{\sigma}^2_{X^*}}.$$

Proposition 3.5.2 *Under the assumptions of model (3.1) and if $E(X^4) < \infty$, $E(\varepsilon_m^4) < \infty$, $E(\varepsilon_l^4) < \infty$ and $E(\varepsilon_r^4) < \infty$, the asymptotic distribution of the bootstrap statistic T_n^* is almost sure $f_1(V)$, where $V \sim N\left(\underline{0}', \frac{\Sigma}{\sigma_X^2}\right)$.*

Proof. Let $(X_i^*, \varepsilon_{mi}^*, \varepsilon_{li}^*, \varepsilon_{ri}^*)$ be a simple random sample from $(X_i, \varepsilon_{mi}, \varepsilon_{li}, \varepsilon_{ri})$. Since

$$\begin{aligned} Y_i^{m*} &= a_m X_i^* + b_m + \varepsilon_{mi}^* = \hat{a}_m X_i^* + \hat{b}_m + \hat{\varepsilon}_{mi}^*, \\ g(Y_i^{l*}) &= a_l X_i^* + b_l + \varepsilon_{li}^* = \hat{a}_l X_i^* + \hat{b}_l + \hat{\varepsilon}_{li}^*, \\ h(Y_i^{r*}) &= a_r X_i^* + b_r + \varepsilon_{ri}^* = \hat{a}_r X_i^* + \hat{b}_r + \hat{\varepsilon}_{ri}^*, \end{aligned}$$

and

$$\begin{aligned} \hat{\varepsilon}_{mi}^* &= Y_i^{m*} - \hat{a}_m X_i^* - \hat{b}_m = (a_m - \hat{a}_m) X_i^* + \left(b_m - \hat{b}_m\right) + \varepsilon_{mi}^*, \\ \hat{\varepsilon}_{li}^* &= g(Y_i^{l*}) - \hat{a}_l X_i^* - \hat{b}_l = (a_l - \hat{a}_l) X_i^* + \left(b_l - \hat{b}_l\right) + \varepsilon_{li}^*, \\ \hat{\varepsilon}_{ri}^* &= h(Y_i^{r*}) - \hat{a}_r X_i^* - \hat{b}_r = (a_r - \hat{a}_r) X_i^* + \left(b_r - \hat{b}_r\right) + \varepsilon_{ri}^*, \end{aligned}$$

it results that

$$\hat{a}_m^* = \frac{\hat{\sigma}_{X^* Z^{m*}}}{\hat{\sigma}^2_{X^*}} = \frac{\hat{\sigma}_{X^* Y^{m*}}}{\hat{\sigma}^2_{X^*}} - \hat{a}_m + k_m = \frac{\hat{\sigma}_{X^* \hat{\varepsilon}_m^*}}{\hat{\sigma}^2_{X^*}} + k_m,$$

and analogously $\hat{a}_l^* = \frac{\hat{\sigma}_{X^* \hat{\varepsilon}_l^*}}{\hat{\sigma}^2_{X^*}} + k_l$, $\hat{a}_r^* = \frac{\hat{\sigma}_{X^* \hat{\varepsilon}_r^*}}{\hat{\sigma}^2_{X^*}} + k_r$.
Hence

$$\sqrt{n} \begin{pmatrix} \hat{a}_m^* - k_m \\ \hat{a}_l^* - k_l \\ \hat{a}_r^* - k_r \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{n}\hat{\sigma}_{X^* \hat{\varepsilon}_m^*}}{\hat{\sigma}^2_{X^*}} \\ \frac{\sqrt{n}\hat{\sigma}_{X^* \hat{\varepsilon}_l^*}}{\hat{\sigma}^2_{X^*}} \\ \frac{\sqrt{n}\hat{\sigma}_{X^* \hat{\varepsilon}_r^*}}{\hat{\sigma}^2_{X^*}} \end{pmatrix}.$$

Since, as $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} \hat{a}_m - a_m \\ \hat{a}_l - a_l \\ \hat{a}_r - a_r \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{n}\hat{\sigma}_{X\varepsilon_m}}{\hat{\sigma}_X^2} \\ \frac{\sqrt{n}\hat{\sigma}_{X\varepsilon_l}}{\hat{\sigma}_X^2} \\ \frac{\sqrt{n}\hat{\sigma}_{X\varepsilon_r}}{\hat{\sigma}_X^2} \end{pmatrix} \xrightarrow{D} N\left(\underline{0}', \frac{\Sigma}{\sigma_X^2}\right),$$

if we prove that

1. $\hat{\sigma}_X^{2\,*} - \hat{\sigma}_X^2 \xrightarrow{P} 0$
2. $\sqrt{n}\begin{pmatrix} \hat{\sigma}_{X^*\hat{\varepsilon}_m^*} - \hat{\sigma}_{X\varepsilon_m} \\ \hat{\sigma}_{X^*\hat{\varepsilon}_l^*} - \hat{\sigma}_{X\varepsilon_l} \\ \hat{\sigma}_{X^*\hat{\varepsilon}_r^*} - \hat{\sigma}_{X\varepsilon_r} \end{pmatrix} \xrightarrow{P} \underline{0}',$

the thesis follows.

1. This part has been proved in Bickel & Freedman (1981).
2. In what follows we prove that $\sqrt{n}\left(\hat{\sigma}_{X^*\hat{\varepsilon}_m^*} - \hat{\sigma}_{X\varepsilon_m}\right) \xrightarrow{P} 0$.
   According to Bickel & Freedman (1981), we can fix A in the σ-field with $P(A) = 1$, so that for any $\omega \in A$ there exists a random vector $(\varepsilon_m^*, X^*, \varepsilon_m, X)$ with

$$\begin{aligned} (\varepsilon_m^*, X^*) &\sim \widehat{F}_n(\omega), \\ (\varepsilon_m, X) &\sim F, \end{aligned}$$

   where $\widehat{F}_n$ denotes the empirical distribution function of $(\varepsilon_{m1}, X_1), \ldots, (\varepsilon_{mn}, X_n)$ and F the theoretical distribution, and

$$E\left[\|(\varepsilon_m^*, X^*) - (\varepsilon_m, X)\|^4\right] \longrightarrow 0,$$

   that is,

$$\begin{aligned} &E\left[\left(\varepsilon_m^* - \varepsilon_m\right)^4\right] \longrightarrow 0, \\ &E\left[\left(X^* - X\right)^4\right] \longrightarrow 0, \\ &E\left[\left(\varepsilon_m^* - \varepsilon_m\right)^2 \left(X^* - X\right)^2\right] \longrightarrow 0. \end{aligned}$$

   We start from

$$E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left(X_i^* - \overline{X}^*\right)\left(\widehat{\varepsilon}_{mi}^* - \overline{\widehat{\varepsilon}}_m^*\right) - \left(X_i - \overline{X}\right)\left(\varepsilon_{mi} - \overline{\varepsilon}_m\right)\right]\right)^2\right]$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n} E\left(\left[\left(X_i^* - \overline{X}^*\right)\left(\widehat{\varepsilon}_{mi}^* - \overline{\widehat{\varepsilon}}_m^*\right) - \left(X_i - \overline{X}\right)\left(\varepsilon_{mi} - \overline{\varepsilon}_m\right)\right]^2\right) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\left[E\left(\left(X_i^* - \overline{X}^*\right)\left(\widehat{\varepsilon}_{mi}^* - \overline{\widehat{\varepsilon}}_m^*\right) - \left(X_i - \overline{X}\right)\left(\varepsilon_{mi} - \overline{\varepsilon}_m\right)\right)\right] \\
&\times \left[E\left(\left(X_j^* - \overline{X}^*\right)\left(\widehat{\varepsilon}_{mj}^* - \overline{\widehat{\varepsilon}}_m^*\right) - \left(X_j - \overline{X}\right)\left(\varepsilon_{mj} - \overline{\varepsilon}_m\right)\right)\right].
\end{aligned}
$$

Taking into account that

$$
\left(\hat{\varepsilon}_{mi}^* - \overline{\hat{\varepsilon}}_m^*\right) = \left(a_m - \hat{a}_m^{\omega}\right)\left(X_i^* - \overline{X}^*\right) + \left(b_m - \hat{b}_m^{\omega}\right) + \left(\varepsilon_{mi}^* - \overline{\varepsilon}_m^*\right),
$$

it follows that

$$
\begin{aligned}
E\Big(\left(X_i^* - \overline{X}^*\right)\left(\widehat{\varepsilon}_{mi}^* - \overline{\widehat{\varepsilon}}_m^*\right)\Big) &= E\Big(\left(a_m - \hat{a}_m^{\omega}\right)\left(X_i^* - \overline{X}^*\right)^2\Big) \\
&+ E\Big(\left(b_m - \hat{b}_m^{\omega}\right)\left(X_i^* - \overline{X}^*\right) + \left(\varepsilon_{mi}^* - \overline{\varepsilon}_m^*\right)\left(X_i^* - \overline{X}^*\right)\Big),
\end{aligned}
$$

and it is straightforward to derive

$$
E\left(\left(X_i^* - \overline{X}^*\right)\left(\widehat{\varepsilon}_{mi}^* - \overline{\widehat{\varepsilon}}_m^*\right)\right) = E\left(\left(X_i - \overline{X}\right)\left(\varepsilon_{mi} - \overline{\varepsilon}_m\right)\right) = 0.
$$

It results that

$$
E\left(\sqrt{n}\left(\hat{\sigma}_{X^*\hat{\varepsilon}_m^*} - \hat{\sigma}_{X\varepsilon_m}\right)\right)^2 = E\left(\left[\left(X_1^* - \overline{X}^*\right)\left(\widehat{\varepsilon}_{m1}^* - \overline{\widehat{\varepsilon}}_m^*\right) - \left(X_1 - \overline{X}\right)\left(\varepsilon_{m1} - \overline{\varepsilon}_m\right)\right]^2\right).
$$

If we prove that the above expression tends to 0, $\sqrt{n}\left(\hat{\sigma}_{X^*\hat{\varepsilon}_m^*} - \hat{\sigma}_{X\varepsilon_m}\right)$ converges to 0 in mean square, hence also in probability.
By means of simple calculations and by applying the Hölder's inequality it can be easily proved that it is enough to check that

$$
E\left(\left[\left(X_1^* - \overline{X}^*\right)\left(\varepsilon_{m1}^* - \overline{\varepsilon}_m^*\right) - \left(X_1 - \overline{X}\right)\left(\varepsilon_{m1} - \overline{\varepsilon}_m\right)\right]^2\right) \longrightarrow 0.
$$

In a similar way, by adding and subtracting $E(\varepsilon_m^*)$ in $(\varepsilon_{m1}^* - \overline{\varepsilon}_m^*)$, $E(X^*)$ in $\left(X_1^* - \overline{X}^*\right)$, $E(X)$ in $\left(X_1 - \overline{X}\right)$ and the $E(\varepsilon_m)$ in $(\varepsilon_{m1} - \overline{\varepsilon}_m)$, and using again the Hölder's inequality the proof is reduced to check if

$$
E\left(\left[\left(X_1^* - E(X^*)\right)\left(\varepsilon_{m1}^* - E(\varepsilon_m^*)\right) - \left(X_1 - E(X)\right)\left(\varepsilon_{m1} - E(\varepsilon_m)\right)\right]^2\right) \longrightarrow 0.
$$

Finally, using the conditions of the random vector $(\varepsilon_m^*, X^*, \varepsilon_m, X)$ it results that

$$
E\left(\left[\left(X_1^* - E(X^*)\right)\left(\varepsilon_{m1}^* - E(\varepsilon_m^*)\right) - \left(X_1 - E(X)\right)\left(\varepsilon_{m1} - E(\varepsilon_m)\right)\right]^2\right) \longrightarrow 0.
$$

Analogously it can be showed that

$$
\begin{aligned}
\sqrt{n}\left(\hat{\sigma}_{X^*\hat{\varepsilon}_l^*} - \hat{\sigma}_{X\varepsilon_l}\right) &\xrightarrow{P} 0 \\
\sqrt{n}\left(\hat{\sigma}_{X^*\hat{\varepsilon}_r^*} - \hat{\sigma}_{X\varepsilon_r}\right) &\xrightarrow{P} 0,
\end{aligned}
$$

so the second part of proposition is proved.

□

Proposition 3.5.3 *In testing the null hypothesis (3.17) at the nominal significance level α, H_0 should be rejected if*

$$T_n^* = V_n^{*'} V_n^* > c_\alpha,$$

where c_α is a α-quantile of the asymptotic distribution of T_n^*.

The application of the test in Proposition 3.5.3 is presented in the following algorithm.

Algorithm

Step 1: Compute the estimate values $\hat{a}_m$, $\hat{a}_l$ and $\hat{a}_r$ and the value of the statistic

$$T_n = V_n' V_n.$$

Step 2: Compute the bootstrap population

$$\left\{(X_i, Z_i^m, Z_i^l, Z_i^r)\right\}_{i=1,\ldots,n}, \tag{3.19}$$

where

$$\begin{aligned} Z_i^m &= Y_i^m - \hat{a}_m X_i + k_m X_i, \\ Z_i^l &= g(Y_i^l) - \hat{a}_l X_i + k_l X_i, \\ Z_i^r &= h(Y_i^r) - \hat{a}_r X_i + k_r X_i. \end{aligned}$$

Note that the bootstrap population (3.19) is defined with the aim of guaranteeing that the null hypothesis is fulfilled.

Step 3: Draw a sample of size n with replacement

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

from the bootstrap population.

Step 4: Compute the value of the bootstrap statistic

$$T_n^* = V_n^{*'} V_n^*.$$

Step 5: Repeat Steps 3 and 4 a large number B of times to get a set of B estimators, denoted by $\{T_1^*, ..., T_B^*\}$.

Step 6: Compute the bootstrap p-value as the proportion of values in $\{T_1^*, ..., T_B^*\}$ such that being greater than T_n.

Simulation studies

In order to illustrate the empirical significance of the bootstrap test proposed in Proposition 3.5.3, a simulated situation has been taken into account. For the simulations we have considered $B = 1000$ replications of the bootstrap estimator and we have carried out 10.000 iterations of the test at 3 different nominal significance levels $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$ for different sample sizes n. Two simulation cases are presented. The first one considers real random variables X, ε_m, ε_l and ε_r behaving as independent $N(0,1)$ random variables. The empirical percentages of rejection under $H_0 : (a_m, a_l, a_r)' = (1,1,1)'$ are represented in Table 3.7.

Table 3.7: Empirical percentages of rejection under $H_0 : (a_m, a_l, a_r)' = (1,1,1)'$ (case of normality).

$n \setminus \alpha \times 100$	1	5	10
30	0.48	4.28	9.52
50	0.86	4.96	10.55
100	0.8	5.05	10.59
200	0.98	5.01	10.57

In the second one we deal with the following real random variables: X, behaving as an $Unif(-3,10)$ random variable, ε_m, ε_l and ε_r behaving as independent $N(0,1)$ random variables. The empirical percentages of rejection under $H_0 : (a_m, a_l, a_r)' = (1,1,1)'$ are represented in Table 3.8.

Table 3.8: Empirical percentages of rejection under $H_0 : (a_m, a_l, a_r)' = (1,1,1)'$ (case of non-normality).

$n \setminus \alpha \times 100$	1	5	10
30	0.92	5.31	10.47
50	1.03	5.86	11.09
100	0.98	5.36	10.80
200	1.03	5.15	9.92

In both cases, by means of the application of the bootstrap procedure, also for very small sample sizes the empirical percentages of rejection are quite close to the nominal levels.

Empirical results

As in previous sections, two real life examples are considered to illustrate the bootstrap test introduced in Section 3.5.2. Taking into account the LR fuzzy data in Table 1.1, to test if the vector of regression parameters $(a_m, a_l, a_r)'$ is equal to $(1, 1, 1)'$, $B = 1000$ replications of the bootstrap statistic are used and a p-value equal to 0 is obtained. Hence the considered hypothesis should be rejected. In testing if the vector $(a_m, a_l, a_r)'$ is equal to a vector whose elements are approximately equal to the estimations of the parameters, that is $(0.16, -0.0002, -0.0007)'$, a p-value equal to 0.85 is obtained. Obviously the hypothesis tested should not be rejected. The second example is referred to the data in Table 3.3. In testing the null hypothesis that the vector of regression parameters $(a_m, a_l, a_r)'$ is equal to $(1, 1, 1)'$, it results a p-value equal to 0, hence the null hypothesis should be rejected. On the contrary the null hypothesis $H_0 : (a_m, a_l, a_r)' = (0.017, 0.000002, 0.000002)'$ should not be rejected, in fact a p-value equal to 0.901 is obtained.

3.5.3 Local alternatives

To show that the effectiveness of the asymptotic test on the regression parameters is the expected one for a linear regression model its power will be studied. To deal with this kind of study is often difficult. To overcome this problem it is possible to analyze the asymptotic power function under a sequence of *local alternatives*, concretely under a sequence of Pitman alternatives. That is, to consider a sequence of alternative hypotheses which converge to the null hypothesis when the sample size increases.

Proposition 3.5.4 *We consider the null hypothesis (3.17) against the alternative H_1 and we use the statistic (3.18) and the critical region $(T_n > k)$. Let H_n be the sequence of Pitman alternatives verifying*

$$\begin{pmatrix} a_m \\ a_l \\ a_r \end{pmatrix} = \begin{pmatrix} k_m \\ k_l \\ k_r \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} \delta_m \\ \delta_l \\ \delta_r \end{pmatrix},$$

where $|\underline{\delta}| > 0$.

1. *Under H_n, $T_n \xrightarrow{D} f_1(V)$, where $V \sim N\left(\underline{\delta}', \frac{\Sigma}{\sigma_X^2}\right)$.*
2. *If we consider the sequence of local alternatives for which $\underline{\delta} = \underline{\delta_n}$, with $|\underline{\delta_n}| \longrightarrow \infty$, then*

$$\lim_{n\to\infty} P_{H_n}(T_n > k) = 1.$$

Proof.

1. By subtracting and adding the vector $(a_m, a_l, a_r)'$ in V_n as follows

$$V_n = \sqrt{n} \begin{pmatrix} \hat{a}_m - a_m + a_m - k_m \\ \hat{a}_l - a_l + a_l - k_l \\ \hat{a}_r - a_r + a_r - k_r \end{pmatrix}$$

it results

$$\begin{aligned} V_n &= \sqrt{n} \begin{pmatrix} \hat{a}_m - a_m \\ \hat{a}_l - a_l \\ \hat{a}_r - a_r \end{pmatrix} + \sqrt{n} \frac{1}{\sqrt{n}} \begin{pmatrix} \delta_m \\ \delta_l \\ \delta_r \end{pmatrix} \\ &= V_n^0 + \begin{pmatrix} \delta_m \\ \delta_l \\ \delta_r \end{pmatrix} \end{aligned}$$

Since $V_n^0 \xrightarrow{D} \left(N\left(\underline{0}', \frac{\Sigma}{\sigma_X^2} \right) \right)$,

$$V_n \xrightarrow{D} \left(N\left(\underline{\delta}', \frac{\Sigma}{\sigma_X^2} \right) \right)$$

and the thesis follows.

2. If $\underline{\delta} = \underline{\delta_n}$, with $|\underline{\delta_n}| \longrightarrow \infty$, we obtain that $V_n \to \infty$ and $T_n \to \infty$, hence

$$\lim_{n \to \infty} P_{H_n}(T_n > k) = 1.$$

□

Remark 3.5.1 Proposition 3.5.4 (1) establishes the maximum speed at which the vector of parameters $(a_m, a_l, a_r)'$ can tend to $(k_m, k_l, k_r)'$ as n increases, so that the power of the test is greater than the significance level. That is, the speed at which the test is able to detect that the null hypothesis is not true. In addition in (2) it is stated that for smaller speeds the asymptotic test will always detect that H_0 is not true. In particular, for any fixed alternative hypothesis the consistency of the test is established.

Remark 3.5.2 Proposition 3.5.4 indicates that the behaviour under local alternatives of the asymptotic test is the same as in classical linear regression studies. However, it should be noted that the explicit expression of the power (Proposition 3.5.4 (1)) is not relevant from a practical point of view because, as the asymptotic distribution is too far from the sampling distribution, a bootstrap test will be used.

Power of the test

In order to obtain a graphical representation of the power function of the test

$$H_0 : \begin{pmatrix} a_m \\ a_l \\ a_r \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

against the alternative

$$H_1 : \begin{pmatrix} a_m \\ a_l \\ a_r \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

we have fixed a sample size equal to 500 and we have simulated different situations in which the null hypothesis is not fulfilled. By means of a bootstrap procedure we have computed the p-value under each of the these different situations. As usual, we have considered $B = 1000$ replications of the bootstrap estimator and we have carried out 10.000 iterations of the test at a nominal significance level α equal to 0.05. The result is shown in Fig. 3.3.

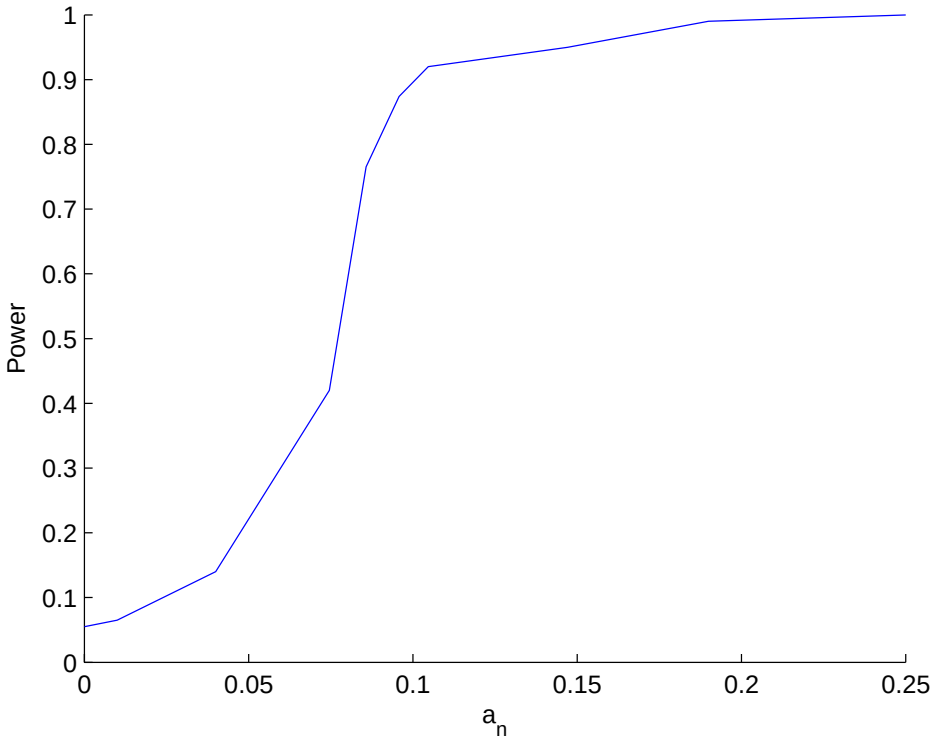

Figure 3.3: The graphical representation of the power of the test

3.6 Estimation of the determination coefficient

Definition 3.6.1 *Let Y be an LR fuzzy random variable and X a random variable satisfying the linear model (3.1), observed on n statistical units, $\{Y_i, X_i\}_{i=1,...,n}$. We can define*

- *the total sum of squares (SST)*

$$SST = \sum_{i=1}^{n} D^2_{\lambda\rho}(\widetilde{Y}_i, \overline{\widetilde{Y}})$$

- *the residual sum of squares (SSE)*

$$SSE = \sum_{i=1}^{n} D^2_{\lambda\rho}(\widetilde{Y}_i, \widehat{\widetilde{Y}})$$

- *the regression sum of squares (SSR)*

$$SSR = \sum_{i=1}^{n} D^2_{\lambda\rho}(\widehat{\widetilde{Y}}_i, \overline{\widetilde{Y}})$$

where, for $i = 1, ..., n$,

$$\begin{aligned}
\widetilde{Y}_i &= (Y_i^m, g(Y_i^l), h(Y_i^r)) = (a_m X_i + b_m + \varepsilon_m, a_l X_i + b_l + \varepsilon_l, a_r X_i + b_r + \varepsilon_r) \\
\widehat{\widetilde{Y}}_i &= (\widehat{Y}_i^m, \widehat{g(Y_i^l)}, \widehat{h(Y_i^r)}) = (\widehat{a}_m X_i + \widehat{b}_m, \widehat{a}_l X_i + \widehat{b}_l, \widehat{a}_r X_i + \widehat{b}_r) \\
\overline{\widetilde{Y}} &= (\overline{Y}^m, \overline{g(Y^l)}, \overline{h(Y^r)}) = (a_m \overline{X} + b_m + \overline{\varepsilon_m}, a_l \overline{X} + b_l + \overline{\varepsilon_l}, a_r \overline{X} + b_r + \overline{\varepsilon_r})
\end{aligned}$$

Proposition 3.6.1 *The total sum of squares, SST, is equal to the sum of the residual sum of squares, SSE, and the regression sum of squares, SSR, that is*

$$SST = SSE + SSR. \tag{3.20}$$

Proof. The total sum of squares can be written as follows

$$\begin{aligned}
\sum_{i=1}^{n} D^2_{\lambda\rho}(\widetilde{Y}_i, \overline{\widetilde{Y}}) &= 3\sum_{i=1}^{n}(Y_i^m - \overline{Y}^m)^2 + \lambda^2 \sum_{i=1}^{n}(g(Y_i^l) - \overline{g(Y^l)})^2 \\
&\quad + \rho^2 \sum_{i=1}^{n}(h(Y_i^r) - \overline{h(Y^r)})^2 - 2\lambda \sum_{i=1}^{n}(Y_i^m - \overline{Y}^m)(g(Y_i^l) - \overline{g(Y^l)}) \\
&\quad + 2\rho \sum_{i=1}^{n}(Y_i^m - \overline{Y}^m)(h(Y_i^r) - \overline{h(Y^r)}).
\end{aligned}$$

By subtracting and adding $\widehat{Y}_i^m$ to the term $(Y_i^m - \overline{Y}^m)$ we get

$$\begin{aligned}\sum_{i=1}^{n}(Y_i^m - \overline{Y}^m)^2 &= \sum_{i=1}^{n}(Y_i^m - \widehat{Y}_i^m + \widehat{Y}_i^m - \overline{Y}^m)^2 \\ &= \sum_{i=1}^{n}(Y_i^m - \widehat{Y}_i^m)^2 + \sum_{i=1}^{n}(\widehat{Y}_i^m - \overline{Y}^m)^2 \\ &\quad -2\sum_{i=1}^{n}(Y_i^m - \widehat{Y}_i^m)(\widehat{Y}_i^m - \overline{Y}^m).\end{aligned}$$

Now we prove that the last term of the sum is equal to 0.

$$\begin{aligned}\sum_{i=1}^{n}(Y_i^m - \widehat{Y}_i^m)(\widehat{Y}_i^m - \overline{Y}^m) &= \sum_{i=1}^{n}(Y_i^m - \widehat{Y}_i^m)\widehat{Y}_i^m - \overline{Y}^m\sum_{i=1}^{n}(Y_i^m - \widehat{Y}_i^m) \\ &= \sum_{i=1}^{n}(Y_i^m - \widehat{a}_m X_i - \widehat{b}_m)(\widehat{a}_m X_i + \widehat{b}_m) \\ &\quad -\overline{Y}^m(\sum_{i=1}^{n} Y_i^m - \widehat{a}_m\sum_{i=1}^{n} X_i - n\widehat{b}_m).\end{aligned}$$

Since $\widehat{b}_m = \frac{\sum_{i=1}^{n} Y_i^m}{n} - \widehat{a}_m\frac{\sum_{i=1}^{n} X_i}{n}$, we have $\overline{Y}^m(\sum_{i=1}^{n} Y_i^m - \widehat{a}_m\sum_{i=1}^{n} X_i - n\widehat{b}_m) = 0$ and it follows

$$\begin{aligned}\sum_{i=1}^{n}(Y_i^m - \widehat{Y}_i^m)(\widehat{Y}_i^m - \overline{Y}^m) &= \sum_{i=1}^{n}(Y_i^m - \widehat{a}_m X_i - \widehat{b}_m)(\widehat{a}_m X_i + \widehat{b}_m) \\ &= \widehat{a}_m\sum_{i=1}^{n} Y_i^m X_i - \widehat{a}_m^2\sum_{i=1}^{n} X_i^2 - \widehat{a}_m\widehat{b}_m\sum_{i=1}^{n} X_i \\ &\quad +\widehat{b}_m\sum_{i=1}^{n} Y_i^m - \widehat{a}_m\widehat{b}_m\sum_{i=1}^{n} X_i - n\widehat{b}_m^2 \\ &= \widehat{a}_m\sum_{i=1}^{n} Y_i^m X_i - \widehat{a}_m^2\sum_{i=1}^{n} X_i^2 - \widehat{a}_m\overline{Y}^m\sum_{i=1}^{n} X_i \\ &\quad +\widehat{a}_m^2\overline{X}\sum_{i=1}^{n} X_i + \overline{Y}^m\sum_{i=1}^{n} Y_i^m \\ &\quad -\widehat{a}_m\overline{X}\sum_{i=1}^{n} Y_i^m - \widehat{a}_m\overline{Y}^m\sum_{i=1}^{n} X_i + \widehat{a}_m^2\overline{X}\sum_{i=1}^{n} X_i \\ &\quad -n(\overline{Y}^m)^2 - n\widehat{a}_m^2\overline{X}^2 + 2n\widehat{a}_m\overline{Y}^m\overline{X}.\end{aligned}$$

Since $\sum_{i=1}^{n} Y_i^m X_i - n\overline{XY}^m = n\widehat{\sigma}_{Y^m X}$, $\sum_{i=1}^{n} X_i^2 - n\overline{X}^2 = n\widehat{\sigma}_X^2$ and $\widehat{a}_m = \frac{\widehat{\sigma}_{xY^m}}{\widehat{\sigma}_x^2}$, we obtain

$$n\widehat{a}_m\widehat{\sigma}_{Y^m X} - n\widehat{a}_m^2\widehat{\sigma}_X^2 = n\frac{\widehat{\sigma}_{xY^m}}{\widehat{\sigma}_x^2} - n\frac{\widehat{\sigma}_{xY^m}}{\widehat{\sigma}_x^2} = 0.$$

Analogously it can be proved that

$$\begin{aligned}
\sum_{i=1}^{n}(g(Y_i^l)-\overline{g(Y^l)})^2 &= \sum_{i=1}^{n}(g(Y_i^l)-\widehat{g(Y_i^l)})^2+\sum_{i=1}^{n}(\widehat{g(Y_i^l)}-\overline{g(Y^l)})^2\\
\sum_{i=1}^{n}(h(Y_i^r)-\overline{h(Y^r)})^2 &= \sum_{i=1}^{n}(h(Y_i^r)-\widehat{h(Y_i^r)})^2+\sum_{i=1}^{n}(\widehat{h(Y_i^r)}-\overline{h(Y^r)})^2\\
\sum_{i=1}^{n}(Y_i^m-\overline{Y}^m)(g(Y_i^l)-\overline{g(Y^l)}) &= \sum_{i=1}^{n}(Y_i^m-\widehat{Y}_i^m)(g(Y_i^l)-\widehat{g(Y_i^l)})\\
&\quad+\sum_{i=1}^{n}(\widehat{Y}_i^m-\overline{Y}^m)(\widehat{g(Y_i^l)})-\overline{g(Y^l)})\\
\sum_{i=1}^{n}(Y_i^m-\overline{Y}^m)(h(Y_i^r)-\overline{h(Y^r)}) &= \sum_{i=1}^{n}(Y_i^m-\widehat{Y}_i^m)(h(Y_i^r)-\widehat{h(Y_i^r)})\\
&\quad+\sum_{i=1}^{n}(\widehat{Y}_i^m-\overline{Y}^m)(\widehat{h(Y_i^r)}-\overline{h(Y^r)}).
\end{aligned}$$

The residual sum squares is equal to

$$\begin{aligned}
\sum_{i=1}^{n}D^2_{\lambda\rho}(\widetilde{Y}_i,\widehat{\widetilde{Y}}_i) &= 3\sum_{i=1}^{n}(Y_i^m-\widehat{Y_i^m})^2+\lambda^2\sum_{i=1}^{n}(g(Y_i^l)-\widehat{g(Y_i^l)})^2\\
&\quad+\rho^2\sum_{i=1}^{n}(h(Y_i^r)-\widehat{h(Y_i^r)})^2-2\lambda\sum_{i=1}^{n}(Y_i^m-\widehat{Y}_i^m)(g(Y_i^l)-\widehat{g(Y_i^l)})\\
&\quad+2\rho\sum_{i=1}^{n}(Y_i^m-\widehat{Y}_i^m)(h(Y_i^r)-\widehat{h(Y_i^r)}),
\end{aligned}$$

and the regression sum of squares is equal to

$$\begin{aligned}
\sum_{i=1}^{n}D^2_{\lambda\rho}(\widehat{\widetilde{Y}}_i,\overline{\widetilde{Y}}) &= 3\sum_{i=1}^{n}(\widehat{Y}_i^m-\overline{Y}^m)^2+\lambda^2\sum_{i=1}^{n}(\widehat{g(Y_i^l)}-\overline{g(Y^l)})^2\\
&\quad+\rho^2\sum_{i=1}^{n}(\widehat{h(Y_i^r)}-\overline{h(Y^r)})^2-2\lambda\sum_{i=1}^{n}(\widehat{Y}_i^m-\overline{Y}^m)(\widehat{g(Y_i^l)}-\overline{g(Y^l)})\\
&\quad+2\rho\sum_{i=1}^{n}(\widehat{Y}_i^m-\overline{Y}^m)(\widehat{h(Y_i^r)}-\overline{h(Y^r)}),
\end{aligned}\tag{3.21}$$

as consequence,

$$SST=SSE+SSR.$$

□

Proposition 3.6.2 *Let Y be an LR fuzzy random variable and X a random variable satisfying the linear model (3.1), observed on n statistical units, $\{Y_i,X_i\}_{i=1,\dots,n}$. The*

estimator of the determination coefficient R^2 *is*

$$\widehat{R}^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}.$$

Proposition 3.6.3 *The estimator* $\widehat{R}^2$ *is strongly consistent, concretely we have that*

1. $\frac{SSR}{n} \xrightarrow{a.s.} E\left[D^2_{\lambda\rho}(E(\widetilde{Y}|X) - E\widetilde{Y})\right]$,
2. $\frac{SST}{n} \xrightarrow{a.s.} E\left[D^2_{\lambda\rho}(\widetilde{Y} - E\widetilde{Y})\right] = \sigma^2_{\widetilde{Y}}$

Proof. To prove the consistency of $\hat{R}^2$ we have to study how this estimator behaves as the size of random samples increases.

1. Starting from the expression (3.21) and taking into account that

$$\begin{aligned}
\widehat{Y_i^m} - \overline{Y^m} &= \hat{a}_m X_i + \hat{b}_m - \overline{Y^m} = \hat{a}_m(X_i - \overline{X}) \\
\widehat{g(Y_i^l)} - \overline{g(Y^l)} &= \hat{a}_l X_i + \hat{b}_l - \overline{g(Y^l)} = \hat{a}_l(X_i - \overline{X}) \\
\widehat{h(Y_i^r)} - \overline{h(Y^r)} &= \hat{a}_r X_i + \hat{b}_r - \overline{h(Y^r)} = \hat{a}_r(X_i - \overline{X})
\end{aligned}$$

it follows

$$\begin{aligned}
\frac{SSR}{n} = \frac{1}{n}\Bigg[& 3\sum_{i=1}^n \hat{a}_m^2(X_i - \overline{X})^2 + \lambda^2 \sum_{i=1}^n \hat{a}_l^2(X_i - \overline{X})^2 + \rho^2 \sum_{i=1}^n \hat{a}_r^2(X_i - \overline{X})^2 \\
& -2\lambda \sum_{i=1}^n \hat{a}_m\hat{a}_l(X_i - \overline{X})^2 + 2\rho \sum_{i=1}^n \hat{a}_m\hat{a}_r(X_i - \overline{X})^2 \Bigg],
\end{aligned}$$

that is

$$\frac{SSR}{n} = 3\hat{a}_m^2\hat{\sigma}_X^2 + \lambda^2\hat{a}_l^2\hat{\sigma}_X^2 + \rho^2\hat{a}_r^2\hat{\sigma}_X^2 - 2\lambda\hat{a}_m\hat{a}_l\hat{\sigma}_X^2 + 2\rho\hat{a}_m\hat{a}_r\hat{\sigma}_X^2. \qquad (3.22)$$

Since the sample moments are strongly consistent estimators of the respective population moments and the estimators of the regression parameters are strongly consistent too, it results

$$\frac{SSR}{n} \xrightarrow{a.s.} 3a_m^2\sigma_X^2 + \lambda^2 a_l^2\sigma_X^2 + \rho^2 a_r^2\sigma_X^2 - 2\lambda a_m a_l\sigma_X^2 + 2\rho a_m a_r\sigma_X^2.$$

Taking into account that

$$\begin{aligned}
E\left[D^2_{\lambda\rho}(E(\widetilde{Y}|X) - E\widetilde{Y})\right] = & 3E[E(Y^m|X) - EY^m]^2 \\
& + \lambda^2 E[E(g(Y^l)|X) - Eg(Y^l)]^2 \\
& + \rho^2 E[E(h(Y^r)|X) - Eh(Y^r)]^2 \\
& - 2\lambda E[(E(Y^m|X) - EY^m)(E(g(Y^l)|X) - Eg(Y^l))] \\
& + 2\rho E[(E(Y^m|X) - EY^m)(E(h(Y^r)|X) - Eh(Y^r))],
\end{aligned}$$

where $E(Y^m|X) = a_m X + b_m$, $E(g(Y^l)|X) = a_l X + b_l$, $E(h(Y^r)|X) = a_r X + b_r$ and $EY^m = a_m EX + b_m$, $Eg(Y^l) = a_l EX + b_l$, $Eh(Y^r) = a_r EX + b_r$, it is easy to check that

$$E\left[D^2_{\lambda\rho}(E(\widetilde{Y}|X) - E\widetilde{Y})\right] = 3a_m^2\sigma_X^2 + \lambda^2 a_l^2\sigma_X^2 + \rho^2 a_r^2\sigma_X^2 - 2\lambda a_m a_l \sigma_X^2 + 2\rho a_m a_r \sigma_X^2,$$

hence

$$\frac{SSR}{n} \xrightarrow{a.s.} E\left[D^2_{\lambda\rho}(E(\widetilde{Y}|X) - E\widetilde{Y})\right].$$

2. Furthermore

$$\frac{SST}{n} = 3\hat{\sigma}^2_{Y^m} + \lambda^2\hat{\sigma}^2_{g(Y^l)} + \rho^2\hat{\sigma}^2_{h(Y^r)} - 2\lambda\hat{\sigma}_{Y^m g(Y^l)} + 2\rho\hat{\sigma}_{Y^m h(Y^r)},$$

as consequence

$$\frac{SST}{n} \xrightarrow{a.s.} 3\sigma^2_{Y^m} + \lambda^2\sigma^2_{g(Y^l)} + \rho^2\sigma^2_{h(Y^r)} - 2\lambda\sigma_{Y^m g(Y^l)} + 2\rho\sigma_{Y^m h(Y^r)},$$

that is,

$$\frac{SST}{n} \xrightarrow{a.s.} E\left[D^2_{\lambda\rho}(\widetilde{Y} - E\widetilde{Y})\right] = \sigma^2_{\widetilde{Y}}.$$

Thus, $\hat{R}^2 = \frac{SSR}{SST}$ is a strongly consistent estimator of the determination coefficient.

□

3.6.1 Simulations

In order to compare the empirical behaviour of the estimator of the determination coefficient, $\hat{R}^2$, with the theoretical value, and to evaluate the accuracy of the estimation, a simulation study is considered. In particular, we have generated an explanatory variable X normally distributed as $N(0,1)$ and an LR fuzzy response Y in the following way: the center Y^m normally distributed as $N(0,1)$, the left and the right spread as χ^2_1. A logarithmic transformation has been used for both spreads. The variables have been generated independently, for this reason the coefficient R^2 is equal to 0. Starting from $n = 30$ for this simulated situation we calculate the value of $\hat{R}^2$. The results are represented in Table 3.9. As the sample size n increases, the estimated values are closer to 0. It illustrates the consistency of $\hat{R}^2$.

Table 3.9: Estimated values $\hat{R}^2$ for samples of different size.

n	$\hat{R}^2$	n	$\hat{R}^2$
30	0.0724	800	0.0007358
50	0.0264	1000	0.000076408
100	0.0032	2000	0.000050197
500	0.0007597	3000	0.000035863

3.6.2 Empirical results

As in the case of the least squares estimators of the regression parameters, we can illustrate the results exposed in the previous sections by means of empirical examples. In particular, if we consider the data of Example 3.3.1 we get $\widehat{R}^2 = 0.2539$, that is, approximately 25.39% of the total variation of the quality is explained by the regression model with the height of the trees as explicative variable.

For the interval data introduced in Example 3.3.2 we get $\widehat{R}^2 = 0.9146$. Approximately almost 92% of the total variation of the Retail Trade Sales of the U.S. in 2002 is explained by the number of employees.

3.7 Linear independence test

A linear regression model is used to analyze the relationship between random variables and to predict a variable based on one or more explanatory variables. If the determination coefficient, that measures the goodness-of-fit of a linear regression model, is equal to zero, there is linear independence. Hence it is necessary to test the linear independence because, in case of lack of linear relationship between random variables, it has no sense to employ the linear regression model with explanatory/predictive purposes.

The goal of this section is to test the null hypothesis

$$H_0 : R^2 = 0 \tag{3.23}$$

against the alternative hypothesis

$$H_1 : R^2 > 0$$

on the basis of the available sample information. For testing the null hypothesis we propose as a test statistic,

$$T_n = n\widehat{R}^2 = n\frac{SSR}{SST}.$$

In the next sections this problem will be analyzed by means of two different approaches: the asymptotic approach and the bootstrap one.

3.7.1 Asymptotic approach

It is simple to derive an asymptotic distribution of the test statistic T_n under the null hypothesis (3.23).

Proposition 3.7.1 *Under the assumptions of model (3.1) and the hypothesis of linear independence (3.23), as* $n \to \infty$

$$T_n = n\widehat{R}^2 \xrightarrow{D} \frac{f_2(W)}{\sigma^2_{\widetilde{Y}}},$$

where $W \sim N(\underline{0}', \Sigma)$ *and* $f_2 : \mathbb{R}^3 \to \mathbb{R}$ *is a mapping that associates to each vector* (a, b, c) *in* $\mathbb{R}^3$ *a value* $f_2(a, b, c) = 3a^2 + \lambda^2 b^2 + \rho^2 c^2 - 2\lambda ab + 2\rho ac$.

Proof. To get the asymptotic distribution of $n\hat{R}^2$ under the null hypothesis, we consider the expression (3.22). It is easy to check that

$$SSR = \frac{n(3\hat{\sigma}^2_{Y^m X} + \lambda^2\hat{\sigma}^2_{g(Y^l)X} + \rho^2\hat{\sigma}^2_{h(Y^r)X} - 2\lambda\hat{\sigma}_{Y^m X}\hat{\sigma}_{g(Y^l)X} + 2\rho\hat{\sigma}_{Y^m X}\hat{\sigma}_{h(Y^r)X})}{\hat{\sigma}^2_X},$$

that is

$$SSR = f_2\left(\frac{\sqrt{n}\hat{\sigma}_{Y^m X}}{\sqrt{\hat{\sigma}^2_X}}, \frac{\sqrt{n}\hat{\sigma}_{g(Y^l)X}}{\sqrt{\hat{\sigma}^2_X}}, \frac{\sqrt{n}\hat{\sigma}_{h(Y^r)X}}{\sqrt{\hat{\sigma}^2_X}}\right).$$

Under the null hypothesis of linear independence $(a_m, a_l, a_r)'$ is equal to $\underline{0}'$. Since $Y^m_i = b_m + \varepsilon_{mi}$, $g(Y^l_i) = b_l + \varepsilon_{li}$, $h(Y^r_i) = b_r + \varepsilon_{ri}$ and $\overline{Y^m} = b_m + \overline{\varepsilon_m}$, $\overline{g(Y^l)} = b_l + \overline{\varepsilon_l}$, $\overline{h(Y^r)} = b_r + \overline{\varepsilon_r}$, it follows that

$$\sqrt{n}\begin{pmatrix} \hat{\sigma}_{Y^m X} \\ \hat{\sigma}_{g(Y^l)X} \\ \hat{\sigma}_{h(Y^r)X} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \overline{X})(\varepsilon_{mi} - \overline{\varepsilon_m}) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \overline{X})(\varepsilon_{li} - \overline{\varepsilon_l})) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \overline{X})(\varepsilon_{ri} - \overline{\varepsilon_r}) \end{pmatrix}.$$

Taking into account that $(X_i - \overline{X}) = (X_i - EX) + (EX - \overline{X})$, we check that

$$\sqrt{n}\begin{pmatrix} \hat{\sigma}_{Y^m X} \\ \hat{\sigma}_{g(Y^l)X} \\ \hat{\sigma}_{h(Y^r)X} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - EX)\varepsilon_{mi} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - EX)\varepsilon_{li} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - EX)\varepsilon_{ri} \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n (EX - \overline{X})\varepsilon_{mi} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n (EX - \overline{X})\varepsilon_{li} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n (EX - \overline{X})\varepsilon_{ri} \end{pmatrix}$$

$$-\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_m \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_l \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_r \end{pmatrix} - \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_m \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_l \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_r \end{pmatrix}.$$

Furthermore,

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\varepsilon_{mi} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\varepsilon_{li} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\varepsilon_{ri} \end{pmatrix} \xrightarrow{a.s.} \underline{0}',$$

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_m \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_l \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-EX)\overline{\varepsilon}_r \end{pmatrix} \xrightarrow{a.s.} \underline{0}',$$

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_m \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_l \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(EX-\overline{X})\overline{\varepsilon}_r \end{pmatrix} \xrightarrow{a.s.} \underline{0}',$$

and $\left\{ \begin{pmatrix} (X_i-EX)\varepsilon_{mi} \\ (X_i-EX)\varepsilon_{li} \\ (X_i-EX)\varepsilon_{ri} \end{pmatrix} \right\}_{i=1,\ldots,n}$ is a sequence of random vectors i.i.d., centered at $\underline{0}'$, whose covariance matrix is $\sigma_X^2\Sigma$, so applying the Central Limit Theorem it results that

$$\begin{pmatrix} \sqrt{n}\hat{\sigma}_{Y^mX} \\ \sqrt{n}\hat{\sigma}_{g(Y^l)X} \\ \sqrt{n}\hat{\sigma}_{h(Y^r)X} \end{pmatrix} \xrightarrow{D} N(\underline{0}', \sigma_X^2\Sigma).$$

Hence $SSR \xrightarrow{D} f_2(W)$.
Besides, Proposition 3.6.3 (2) ensures that

$$\frac{1}{n}SST \longrightarrow \sigma_{\widetilde{Y}}^2.$$

Consequently applying Slutsky's theorem we can assure that

$$n\widehat{R}^2 \xrightarrow{D} \frac{f_2(W)}{\sigma_{\widetilde{Y}}^2}.$$

□

By means of the large sample theory it is possible to define a rejection region for the null hypothesis.

Proposition 3.7.2 *In testing the null hypothesis of linear independence at the nominal significance level α, H_0 should be rejected if*

$$T_n > c_\alpha,$$

where c_α is a α-quantile of the asymptotic distribution of T_n.

As in the case of hypothesis testing on the regression parameters, this asymptotic test works suitably for samples with very large size, and for this reason we propose also for the linear independence test a bootstrap approach.

3.7.2 Bootstrap test of linear independence

In order to obtain a bootstrap population fulfilling the hypothesis of linear independence, the residual variables can be considered. That is, the new variables $Z^m = Y^m - \hat{a}_m X$, $Z^l = g(Y^l) - \hat{a}_l X$ and $Z^r = h(Y^r) - \hat{a}_r X$ are considered. Then, a sample of size n with replacement $\left\{(X^*_i, Z^{m*}_i, Z^{l*}_i, Z^{r*}_i)\right\}_{i=1,\dots,n}$ from the bootstrap population is drawn. The bootstrap statistic is

$$T^*_n = n\frac{\sum_{i=1}^{n} D^2_{\lambda\rho}(\widetilde{\widehat{Z^*}}_i, \widetilde{\overline{Z^*}})}{\sigma^2_{\widetilde{Y}}}$$

where $\widetilde{Z^*}_i = (Z^{m*}_i, Z^{l*}_i, Z^{r*}_i)$.

Proposition 3.7.3 *Under the assumptions of model (3.1) and if $E(X^4) < \infty$, $E(\varepsilon^4_m) < \infty$, $E(\varepsilon^4_l) < \infty$ and $E(\varepsilon^4_r) < \infty$, as $n \to \infty$,*

$$T^*_n \xrightarrow{D} \frac{f_2(W)}{\sigma^2_{\widetilde{Y}}},$$

where $W \sim N(\underline{0}', \Sigma)$ and f_2 is the same function defined in Proposition 3.7.1.

Proof. Since the bootstrap statistic T^*_n can be expressed as

$$\frac{f_2\left(\frac{\sqrt{n}\hat{\sigma}_{Z^{m*}X^*}}{\sqrt{\hat{\sigma}^2_{X^*}}}, \frac{\sqrt{n}\hat{\sigma}_{Z^{l*}X^*}}{\sqrt{\hat{\sigma}^2_{X^*}}}, \frac{\sqrt{n}\hat{\sigma}_{Z^{r*}X^*}}{\sqrt{\hat{\sigma}^2_{X^*}}}\right)}{\sigma^2_{\widetilde{Y}}},$$

for what showed in the proof of Proposition 3.5.2 it results that

$$\hat{\sigma}_{X^*Z^{m*}} = \hat{\sigma}_{X^*Y^{m*}} - \hat{a}_m\hat{\sigma}^2_{X^*} = \hat{a}_m\hat{\sigma}^2_{X^*} + \hat{\sigma}_{X^*\hat{\varepsilon}^*_m} - \hat{a}_m\hat{\sigma}^2_{X^*},$$

and analogously

$$\hat{\sigma}_{X^*Z^{l*}} = \hat{\sigma}_{X^*Y^{l*}} - \hat{a}_l\hat{\sigma}^2_{X^*} = \hat{a}_l\hat{\sigma}^2_{X^*} + \hat{\sigma}_{X^*\hat{\varepsilon}^*_l} - \hat{a}_l\hat{\sigma}^2_{X^*},$$
$$\hat{\sigma}_{X^*Z^{r*}} = \hat{\sigma}_{X^*Y^{r*}} - \hat{a}_r\hat{\sigma}^2_{X^*} = \hat{a}_r\hat{\sigma}^2_{X^*} + \hat{\sigma}_{X^*\hat{\varepsilon}^*_r} - \hat{a}_r\hat{\sigma}^2_{X^*}.$$

Hence

$$T^*_n = \frac{f_2\left(\frac{\sqrt{n}\hat{\sigma}_{X^*\hat{\varepsilon}^*_m}}{\sqrt{\hat{\sigma}^2_{X^*}}}, \frac{\sqrt{n}\hat{\sigma}_{X^*\hat{\varepsilon}^*_l}}{\sqrt{\hat{\sigma}^2_{X^*}}}, \frac{\sqrt{n}\hat{\sigma}_{X^*\hat{\varepsilon}^*_r}}{\sqrt{\hat{\sigma}^2_{X^*}}}\right)}{\sigma^2_{\widetilde{Y}}}$$

Taking into account that for Proposition 3.7.1, as $n \to \infty$,

$$\begin{pmatrix} \sqrt{n}\hat{\sigma}_{X\varepsilon_m} \\ \sqrt{n}\hat{\sigma}_{X\varepsilon_l} \\ \sqrt{n}\hat{\sigma}_{X\varepsilon_r} \end{pmatrix} \xrightarrow{D} N\left(\underline{0}', \sigma^2_X\Sigma\right),$$

and, as previously proved (see Proposition 3.5.2)

1. $\hat{\sigma}^{2\,*}_X - \hat{\sigma}^2_X \xrightarrow{P} 0$
2. $\sqrt{n}\begin{pmatrix} \hat{\sigma}_{X^*\hat{\varepsilon}^*_m} - \hat{\sigma}_{X\varepsilon_m} \\ \hat{\sigma}_{X^*\hat{\varepsilon}^*_l} - \hat{\sigma}_{X\varepsilon_l} \\ \hat{\sigma}_{X^*\hat{\varepsilon}^*_r} - \hat{\sigma}_{X\varepsilon_r} \end{pmatrix} \xrightarrow{P} \underline{0}',$

the thesis follows.

□

Proposition 3.7.4 *In testing the null hypothesis of linear independence at the nominal significance level α, H_0 should be rejected if*

$$T^*_n = n\frac{\sum_{i=1}^{n} D^2_{\lambda\rho}(\widehat{\widetilde{Z^*}}_i, \overline{\widetilde{Z^*}})}{\sigma^2_{\widetilde{Y}}} > c_\alpha,$$

where c_α is a α-quantile of the asymptotic distribution of T^*_n.

The application of the test in Proposition 3.7.4 is presented in the following algorithm.

Algorithm

Step 1: Compute the estimate values $\hat{a}_m$, $\hat{a}_l$ and $\hat{a}_r$ and the value of the statistic

$$T_n = n\hat{R}^2 = n\frac{\sum_{i=1}^{n} D^2_{\lambda\rho}(\widehat{\widetilde{Y}}_i, \overline{\widetilde{Y}})}{\sum_{i=1}^{n} D^2_{\lambda\rho}(\widetilde{Y}_i, \overline{\widetilde{Y}})}$$

Step 2: Compute the bootstrap population

$$\left\{(X_i, Z_i^m, Z_i^l, Z_i^r)\right\}_{i=1,\ldots,n}, \tag{3.24}$$

where

$$\begin{aligned} Z_i^m &= Y_i^m - \hat{a}_m X_i, \\ Z_i^l &= g(Y_i^l) - \hat{a}_l X_i, \\ Z_i^r &= h(Y_i^r) - \hat{a}_r X_i. \end{aligned}$$

Step 3: Draw a sample of size n with replacement

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

from the bootstrap population.

Step 4: Compute the value of the bootstrap statistic

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

where $\widetilde{Z^*}_i = (Z^{m*}{}_i, Z^{l^*}{}_i, Z^{r*}{}_i)$.

Step 5: Repeat Steps 3 and 4 a large number B of times to get a set of B estimators, denoted by $\{T_1^*, ..., T_B^*\}$.

Step 6: Compute the bootstrap p-value as the proportion of values in $\{T_1^*, ..., T_B^*\}$ being greater than T_n.

Simulation studies

Simulations are considered to illustrate the empirical significance of the bootstrap test proposed in Proposition 3.7.4. For the simulations we have considered $B = 1000$ replications of the bootstrap estimator and we have carried out 10.000 iterations of

the test at 3 different nominal significance levels $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$ for different sample sizes n. Two simulation cases are presented. The first one considers real random variables X, ε_m, ε_l and ε_r behaving as independent $N(0,1)$ random variables. The empirical percentages of rejection under H_0 are represented in Table 3.10.

Table 3.10: Empirical percentages of rejection under the hypothesis of linear independence (case of normality).

$n \setminus \alpha \times 100$	1	5	10
30	1.81	6.75	11.93
50	1.71	6.20	11.20
100	1.27	5.64	10.90
200	1.27	5.55	10.57
300	1.05	5.06	10.08

By means of the application of the bootstrap procedure for $n > 200$ the empirical percentages of rejection are quite close to the nominal level.

In the second one we deal with the following real random variables: X, behaving as an $Unif(-20, 35)$ random variable, ε_m, behaving as an $Unif(0, 8)$, ε_l behaving as an $Unif(-12, 22)$ and ε_r behaving as an $Unif(-1, 5)$. The empirical percentages of rejection under H_0 are represented in Table 3.11.

Table 3.11: Empirical percentages of rejection under the hypothesis of linear independence (case of non-normality).

$n \setminus \alpha \times 100$	1	5	10
30	1.84	5.44	11.11
50	1.53	5.30	11.07
100	1.30	5.29	10.69
200	1.22	5.21	9.96
300	1.06	5.03	10.02

It results that for $n \geq 200$ the empirical percentages of rejection are quite close to the three nominal levels.

Remark 3.7.1 In future works it could be interesting to standardize the test statistic T_n in order to obtain empirical percentages of rejection quite close to the nominal levels also for very small sample sizes.

Empirical results

In order to employ the bootstrap test defined in Section 3.7.2 two real life examples are considered. The first one considers the LR fuzzy data in Table 1.1 and the second one is referred to the data in Table 3.3. For the simulations $B = 1000$ replications of the bootstrap estimator are used. For both examples the p-value is equal to 0. In both cases, the observed significance level is smaller than 0.01, that is, without hesitation the null hypothesis of linear independence is rejected.

3.7.3 Local alternatives

To study the power of the test, as in Section 3.5.3, a sequence of local alternatives is taken into account.

Proposition 3.7.5 *We consider the null hypothesis (3.23) of the linear independence test against the alternative H_1 and we use the statistic T_n and the critical region $(T_n > k)$. Let H_n be the sequence of Pitman alternatives verifying*

$$\begin{pmatrix} a_m \\ a_l \\ a_r \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} \delta_m \\ \delta_l \\ \delta_r \end{pmatrix},$$

where $|\underline{\delta}| > 0$. Then

1. *Under H_n, $T_n \xrightarrow{D} \frac{f_2(W)}{\sigma^2_{\tilde{Y}}}$, where $W \sim N(\underline{\delta}' \sqrt{\sigma^2_X}, \Sigma)$.*
2. *If we consider the sequence of local alternatives for which $\underline{\delta} = \underline{\delta_n}$, with $|\underline{\delta_n}| \longrightarrow \infty$, then*

$$\lim_{n\to\infty} P_{H_n}(T_n > k) = 1.$$

Proof.

1. As in the previous section, we use the test statistic $T_n = n\hat{R}^2 = n\frac{SSR}{SST}$ and

$$SSR = f_2\left(\frac{\sqrt{n}\hat{\sigma}_{Y^m X}}{\sqrt{\hat{\sigma}^2_X}}, \frac{\sqrt{n}\hat{\sigma}_{g(Y^l)X}}{\sqrt{\hat{\sigma}^2_X}}, \frac{\sqrt{n}\hat{\sigma}_{h(Y^r)X}}{\sqrt{\hat{\sigma}^2_X}}\right).$$

As, under the local alternatives H_n, $Y_i^m = (\delta_m/\sqrt{n})X_i + b_m + \varepsilon_{mi}$, $g(Y_i^l) = (\delta_l/\sqrt{n})X_i + b_l + \varepsilon_{li}$, $h(Y_i^r) = (\delta_r/\sqrt{n})X_i + b_r + \varepsilon_{ri}$ and $\overline{Y^m} = (\delta_m/\sqrt{n})\overline{X} +$

$b_m + \overline{\varepsilon_m}$, $\overline{g(Y^l)} = (\delta_l/\sqrt{n})\overline{X} + b_l + \overline{\varepsilon_l}$, $\overline{h(Y^r)} = (\delta_r/\sqrt{n})\overline{X} + b_r + \overline{\varepsilon_r}$, it results that

$$\sqrt{n}\begin{pmatrix} \hat{\sigma}_{Y^m X} \\ \hat{\sigma}_{g(Y^l)X} \\ \hat{\sigma}_{h(Y^r)X} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\delta_m}{\sqrt{n}}(X_i - \overline{X}) + (\varepsilon_i^m - \overline{\varepsilon^m})\right)(X_i - \overline{X}) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\delta_l}{\sqrt{n}}(X_i - \overline{X}) + (\varepsilon_i^l - \overline{\varepsilon^l})\right)(X_i - \overline{X}) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\delta_r}{\sqrt{n}}(X_i - \overline{X}) + (\varepsilon_i^r - \overline{\varepsilon^r})\right)(X_i - \overline{X}) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\delta_m(X_i - \overline{X})^2 \\ \frac{1}{n}\sum_{i=1}^{n}\delta_l(X_i - \overline{X})^2 \\ \frac{1}{n}\sum_{i=1}^{n}\delta_r(X_i - \overline{X})^2 \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \overline{X})(\varepsilon_i^m - \overline{\varepsilon^m}) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \overline{X})(\varepsilon_i^l - \overline{\varepsilon^l}) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \overline{X})(\varepsilon_i^r - \overline{\varepsilon^r}) \end{pmatrix}$$

Since

$$\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\delta_m(X_i - \overline{X})^2 \\ \frac{1}{n}\sum_{i=1}^{n}\delta_l(X_i - \overline{X})^2 \\ \frac{1}{n}\sum_{i=1}^{n}\delta_r(X_i - \overline{X})^2 \end{pmatrix} \longrightarrow \hat{\sigma}_X^2 \begin{pmatrix} \delta_m \\ \delta_l \\ \delta_r \end{pmatrix} \qquad a.s. - [P]$$

and, as previously proved,

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \overline{X})(\varepsilon_i^m - \overline{\varepsilon^m}) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \overline{X})(\varepsilon_i^l - \overline{\varepsilon^l}) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \overline{X})(\varepsilon_i^r - \overline{\varepsilon^r}) \end{pmatrix} \xrightarrow{D} N\left(\underline{0}', \sigma_X^2\Sigma\right),$$

as consequence

$$\sqrt{n}\begin{pmatrix} \hat{\sigma}_{Y^m X} \\ \hat{\sigma}_{g(Y^l)X} \\ \hat{\sigma}_{h(Y^r)X} \end{pmatrix} \xrightarrow{D} N\left(\underline{\delta}'\sigma_X^2, \sigma_X^2\Sigma\right).$$

It results that

$$SSR \xrightarrow{D} f_2(W),$$

where $W \sim N\left(\underline{\delta}'\sqrt{\sigma_X^2}, \Sigma\right)$, and the thesis follows.

2. If $\underline{\delta} = \underline{\delta_n}$, with $|\underline{\delta_n}| \longrightarrow \infty$, we obtain that $SSR \to \infty$ and $T_n \to \infty$, hence

$$\lim_{n\to\infty} P_{H_n}(T_n > k) = 1.$$

□

Remark 3.7.2 As for Proposition 3.5.4, the explicit expression of the power (Proposition 3.7.5 (1)) is not relevant from a practical point of view because, as the asymptotic distribution is too far from the sampling distribution, a bootstrap test will be used.

3.8 Final evaluation and open problems

In this chapter we have carried out a wide statistical analysis concerning a regression model to express an imprecise response as a function of a crisp explanatory variable. Namely

- The least squares estimators have been found, and some confidence regions and testing procedures have been developed on the basis of their asymptotic distributions.
- Some bootstrap techniques have been considered in order to improve the empirical results for small/moderate sample sizes and we have shown by means of some simulations their suitability in practice.
- A determination coefficient has been defined and an estimator has been analyzed. In addition a test to check the goodness-of-fit of the model has been developed on the basis of this estimator.
- Some analysis of power of the tests through local alternatives has been developed and some simulations to illustrate the empirical behaviour in this respect have been shown.
- All the results have been applied to some real-case studies with illustrative purpose.

As open problems concerning this chapter, we consider interesting

- The analysis of an appropriate family of functions g and h to transform the spreads of the LR response variables and the introduction of semi-parametric models.
- The study of non-linear models in which the explanatory variables are transformed due to the restrictions that they have to satisfy.

Chapter 4

A multiple linear regression model with imprecise response

In this chapter a multiple linear regression model with imprecise response is discussed. This model is a generalization of model (3.1), but it is formally different, because of the matrix notation. This formalization makes it possible to extend the results of the simple case. Only the outline of the procedure is described, due to the analogy with the previous chapter.

The chapter is organized as follows. In Section 4.1 the population multiple regression model is formally defined and described. Section 4.2 contains the definition of a multiple determination coefficient to measure the degree of linear dependence. Section 4.3 deals with the estimation problem. In details, the least squares estimators of the regression parameters are checked and some algebraic and statistical properties are proved. To employ the new model in Sections 4.3.1 and 4.3.2, respectively, a simulated situation and empirical results will be considered. The study of confidence regions and hypothesis testing on the regression parameters is analyzed in Section 4.4. In Section 4.5 it is proved the decomposition of the total sum of squares into the sum of the residual sum of squares and the regression sum of squares. Taking it into account an estimator of the determination coefficient is proposed. Section 4.6 contains a linear independence test. For all the last three sections there are simulations and empirical examples to illustrate the accuracy. Final evaluations and open problems are the last part of this chapter.

4.1 The regression model

Consider a random experiment in which an LR fuzzy response variable Y and p real explanatory variables $X_1, X_2, ..., X_p$ are observed on n statistical units, $\{Y_i, \underline{X}_i\}_{i=1,...,n}$, where $\underline{X}_i = (X_{1i}, X_{2i}, ..., X_{pi})$, or in a compact form $(\underline{Y}, \mathbf{X})$, where $\underline{Y}$ is the $1 \times n$-vector of the observations of Y and $\mathbf{X}$ is the $n \times p$-matrix of the observations of $\underline{X}$. The model (3.1) generalized to this multiple case is

$$\begin{cases} Y^m = \underline{X}\,\underline{a}_m^{'} + b_m + \varepsilon_m \\ g(Y^l) = \underline{X}\,\underline{a}_l^{'} + b_l + \varepsilon_l \\ h(Y^r) = \underline{X}\,\underline{a}_r^{'} + b_r + \varepsilon_r \end{cases} \tag{4.1}$$

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

As in the simple case, it is easy to check that the variables ε_m, ε_l and ε_r are uncorrelated with the explanatory variables.

4.1.1 Theoretical values

In Proposition 4.1.1 the expression of the population parameters in terms of moments, analogous to Proposition 3.1.1, is shown

Proposition 4.1.1 *Let Y be an LR fuzzy random variable and $\underline{X}$ the vector of the p real random variables satisfying the linear model (4.1), then we have that*

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

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

Proof. Under the assumptions in this proposition, by following the same reasoning of the proof of Proposition 3.1.1, it is easy to get the thesis. □

4.2 Multiple determination coefficient

As in the simple case, the decomposition of the total variation of the response in the variation that does not depend on the model and the variation explained by the model remains valid, that is,

Proposition 4.2.1 *Let Y be an LR fuzzy random variable and $\underline{X}$ a vector of real random variables satisfying the linear model (4.1) so that the errors are uncorrelated with $\underline{X}$, by indicating $\widetilde{Y} = (Y^m, g(Y^l), h(Y^l))$, we obtain*

$$E\left[D^2_{\lambda\rho}(\widetilde{Y}, E\widetilde{Y})\right] = E\left[D^2_{\lambda\rho}(\widetilde{Y}, E(\widetilde{Y}|\underline{X}))\right] + E\left[D^2_{\lambda\rho}(E(\widetilde{Y}|\underline{X}), E\widetilde{Y})\right]. \tag{4.2}$$

A multiple determination coefficient, analogously to the (3.3), is introduced.

Definition 4.2.1 *Let Y be the LR FRV of the linear model (4.1), by indicating $\widetilde{Y} = (Y^m, g(Y^l), h(Y^l))$, the determination coefficient can be defined as follows*

$$R^2 = \frac{E\left[D^2_{\lambda\rho}(E(\widetilde{Y}|\underline{X}), E\widetilde{Y})\right]}{E\left[D^2_{\lambda\rho}(\widetilde{Y}, E\widetilde{Y})\right]} = 1 - \frac{E\left[D^2_{\lambda\rho}(\widetilde{Y}, E(\widetilde{Y}|\underline{X}))\right]}{E\left[D^2_{\lambda\rho}(\widetilde{Y}, E\widetilde{Y})\right]}. \tag{4.3}$$

This coefficient measures the degree of linear relationship. As in the simple case it takes values in $[0, 1]$. In particular, $R^2 = 0$ indicates linear independence and when R^2 reaches the value 1, it indicates that the variability of the response is completely explained by the model.

4.3 The estimation problem

As in Section 3.3, the estimators of the population parameters of the multiple model will be based on the LS criterion. In this case, using the Yang-Ko metric $D^2_{\lambda\rho}$ written in vector terms, the LS problem consists in looking for $\underline{\hat{a}}_m$, $\underline{\hat{a}}_l$, $\underline{\hat{a}}_r$, $\hat{b}_m$, $\hat{b}_l$ and $\hat{b}_r$ in order to

$$\min \Delta^2_{\lambda\rho} = \min D^2_{\lambda\rho}((\underline{Y}^{m'}, g(\underline{Y}^l)', h(\underline{Y}^r)'), ((\underline{Y}^m)^{*'}, g^*(\underline{Y}^l)', h^*(\underline{Y}^r)') \tag{4.4}$$

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

The function to minimize

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

becomes

$$\begin{aligned}\Delta^2_{\lambda\rho} &= 3\left(\underline{Y}^{m'} - \mathbf{X}\underline{a}'_m - \underline{1}'b_m\right)'\left(\underline{Y}^{m'} - \mathbf{X}\underline{a}'_m - \underline{1}'b_m\right) \\ &+ \lambda^2\left(g(\underline{Y}^l)' - \mathbf{X}\underline{a}'_l - \underline{1}'b_l\right)'\left(g(\underline{Y}^l)' - \mathbf{X}\underline{a}'_l - \underline{1}'b_l\right) \\ &+ \rho^2\left(h(\underline{Y}^r)' - \mathbf{X}\underline{a}'_r - \underline{1}'b_r\right)'\left(h(\underline{Y}^r)' - \mathbf{X}\underline{a}'_r - \underline{1}'b_r\right) \\ &- 2\lambda\left(\underline{Y}^{m'} - \mathbf{X}\underline{a}'_m - \underline{1}'b_m\right)'\left(g(\underline{Y}^l)' - \mathbf{X}\underline{a}'_l - \underline{1}'b_l\right) \\ &+ 2\rho\left(\underline{Y}^{m'} - \mathbf{X}\underline{a}'_m - \underline{1}'b_m\right)'\left(h(\underline{Y}^r)' - \mathbf{X}\underline{a}'_r - \underline{1}'b_r\right).\end{aligned} \tag{4.5}$$

Analogously to Proposition 3.3.1, it holds

Proposition 4.3.1 *The solutions of the LS problem are*

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

where

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

are the centered values of the response and

$$\widetilde{\boldsymbol{X}} = \boldsymbol{X} - \underline{1}'(\underline{1}\,\underline{1}')^{-1}\underline{1}\boldsymbol{X}$$

the centered matrix of the explanatory variables.

Proof. By means of the same procedure used in Proposition 3.3.1 it is possible to get the least squares estimators in this multiple case.

□

The estimated values, obtained from the LS criterion, fulfill some algebraic properties.

Proposition 4.3.2 *For model (4.1) and the LS estimators of Proposition 4.3.1 we have*

(i) *The sums of the residual values are equal to* 0*, that is*

$$\underline{1}\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = 0,$$
$$\underline{1}\left(g(\underline{Y}^l) - \widehat{g(\underline{Y}^l)}\right)' = 0,$$
$$\underline{1}\left(h(\underline{Y}^r) - \widehat{h(\underline{Y}^r)}\right)' = 0.$$

(ii) *The residuals* $\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)'$, $\left(g(\underline{Y}^l) - \widehat{g(\underline{Y}^l)}\right)'$ *and* $\left(h(\underline{Y}^r) - \widehat{h(\underline{Y}^r)}\right)'$ *are uncorrelated with the matrix of the explanatory variables,* $\boldsymbol{X}$*, that is*

$$\boldsymbol{X}'\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = \underline{0}',$$
$$\boldsymbol{X}'\left(g(\underline{Y}^l) - \widehat{g(\underline{Y}^l)}\right)' = \underline{0}',$$
$$\boldsymbol{X}'\left(h(\underline{Y}^r) - \widehat{h(\underline{Y}^r)}\right)' = \underline{0}',$$

where $\underline{0}$ *is the* $1 \times p$ *null vector.*

(iii) *The residuals* $\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)'$, $\left(g(\underline{Y}^l) - \widehat{g(\underline{Y}^l)}\right)'$ *and* $\left(h(\underline{Y}^r) - \widehat{h(\underline{Y}^r)}\right)'$ *are uncorrelated, respectively, with the predicted values* $\widehat{\underline{Y}^m}'$, $\widehat{g(\underline{Y}^l)}'$ *and* $\widehat{h(\underline{Y}^r)}'$*, that is*

$$\widehat{\underline{Y}^m}\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = 0,$$
$$\widehat{g(\underline{Y}^l)}\left(g(\underline{Y}^l) - \widehat{g(\underline{Y}^l)}\right)' = 0,$$
$$\widehat{h(\underline{Y}^r)}\left(h(\underline{Y}^r) - \widehat{h(\underline{Y}^r)}\right)' = 0.$$

Proof. For each property, it will be only proved the first equality, because the other ones may be obtained analogously.

(i) Since $\widehat{\underline{Y}^m}' = \mathbf{X}\widehat{\underline{a}}_m' + \underline{1}'\widehat{b}_m$, it follows

$$\underline{1}\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = \underline{1}\left(\underline{Y}^{m\prime} - \mathbf{X}\widehat{\underline{a}}_m' - \underline{1}'\widehat{b}_m\right),$$

and taking into account that $\hat{b}_m = (\underline{1}\,\underline{1}')^{-1}\underline{1}\,\underline{Y}^{m\prime} - (\underline{1}\,\underline{1}')^{-1}\underline{1}\mathbf{X}\hat{\underline{a}}_m'$,

$$\underline{1}\,\underline{Y}^{m\prime} - \underline{1}\mathbf{X}\widehat{\underline{a}}_m' - (\underline{1}\,\underline{1}')(\underline{1}\,\underline{1}')^{-1}\underline{1}\,\underline{Y}^{m\prime} + (\underline{1}\,\underline{1}')(\underline{1}\,\underline{1}')^{-1}\underline{1}\mathbf{X}\hat{\underline{a}}_m' = 0.$$

(ii) Consider the centered matrix of explanatory variables $\widetilde{\mathbf{X}} = \mathbf{X} - \underline{1}'(\underline{1}\,\underline{1}')^{-1}\underline{1}\mathbf{X}$. It results that

$$\begin{aligned}\widetilde{\mathbf{X}}'\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' &= \widetilde{\mathbf{X}}'\underline{Y}^{m\prime} - \widetilde{\mathbf{X}}'\mathbf{X}\widehat{\underline{a}}_m' - \widetilde{\mathbf{X}}'\underline{1}'\widehat{b}_m \\ &= \widetilde{\mathbf{X}}'\underline{Y}^{m\prime} - \widetilde{\mathbf{X}}'\mathbf{X}\widehat{\underline{a}}_m' - \widetilde{\mathbf{X}}'\underline{1}'(\underline{1}\,\underline{1}')^{-1}\underline{1}\,\underline{Y}^{m\prime} + \widetilde{\mathbf{X}}'\underline{1}'(\underline{1}\,\underline{1}')^{-1}\underline{1}\mathbf{X}\widehat{\underline{a}}_m' \\ &= \widetilde{\mathbf{X}}'\widetilde{\underline{Y}^m}' - \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\underline{a}}_m'.\end{aligned}$$

Taking into account that $\widehat{\underline{a}}_m' = (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\widetilde{\underline{Y}^m}'$, we have

$$\widetilde{\mathbf{X}}'\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = \underline{0}'. \tag{4.6}$$

From (4.6)

$$\widetilde{\mathbf{X}}'\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = (\mathbf{X} - \underline{1}'(\underline{1}\,\underline{1}')^{-1}\underline{1}\mathbf{X})'\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = \underline{0}',$$

that is

$$\mathbf{X}'\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' - \mathbf{X}'\underline{1}'(\underline{1}\,\underline{1}')^{-1}\underline{1}\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = \underline{0}'.$$

Since $\underline{1}\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = 0$, it follows

$$\mathbf{X}'\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = \underline{0}'.$$

(iii) By means of the previous property it can be easily proved that

$$\begin{aligned}\widehat{\underline{Y}^m}\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' &= \left(\mathbf{X}\widehat{\underline{a}}_m' + \underline{1}'\widehat{b}_m\right)'\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' \\ &= \widehat{\underline{a}}_m\mathbf{X}'\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' + \widehat{b}_m\underline{1}\left(\underline{Y}^m - \widehat{\underline{Y}^m}\right)' = 0\end{aligned}$$

□

Analogously to the classical case of linear regression analysis, it is easy to check some statistical properties of the LS estimators.

Proposition 4.3.3 *The estimators $\widehat{\underline{a}}_m$, $\widehat{\underline{a}}_l$, $\widehat{\underline{a}}_r$, $\widehat{b}_m$, $\widehat{b}_l$ and $\widehat{b}_r$ are unbiased, strongly consistent and as $n \to \infty$*

$$\sqrt{n}\begin{pmatrix}\widehat{\underline{a}}_m' - \underline{a}_m' \\ \widehat{\underline{a}}_l' - \underline{a}_l' \\ \widehat{\underline{a}}_r' - \underline{a}_r'\end{pmatrix} \xrightarrow{D} N\left(\underline{0}', \frac{\Sigma}{\Sigma_{\underline{X}}}\right). \tag{4.7}$$

4.3.1 Simulations

In order to compare the estimates obtained by means of the least squares procedure with the theoretical values, we consider a simulated situation. A sample of 30 units is drawn. Three explanatory variables X_1, X_2, X_3 and three random variables ε_m, ε_l, ε_r have been generated normally distributed as $N(0,1)$. The theoretical values taken into account are: $\underline{a}_m = (3, -1.2, 16)$, $\underline{a}_l = (4, 1.2, 2.5)$, $\underline{a}_r = (-2, 8.2, -3)$, $b_m = 3.5$, $b_l = -12$ and $b_r = 4$. The response variables are obtained as

$$\begin{cases} Y_i^m = 3X_{1i} - 1.2X_{2i} + 16X_{3i} + 3.5 + \varepsilon_{mi} \\ g(Y_i^l) = 4X_{1i} + 1.2X_{2i} + 2.5X_{3i} - 12 + \varepsilon_{li} \\ h(Y_i^r) = -2X_{1i} + 8.2X_{2i} - 3X_{3i} + 4 + \varepsilon_{ri} \end{cases} \tag{4.8}$$

for $i = 1, .., 30$. The simulated data are shown in Table 4.1.

The estimated model is

$$\begin{cases} \widehat{Y^m} = 2.9865X_1 - 0.9708X_2 + 15.7021X_3 + 3.4252 \\ \widehat{g(Y^l)} = 3.8906X_1 + 1.5853X_2 + 2.5552X_3 - 11.8661 \\ \widehat{h(Y^r)} = -1.8671X_1 + 8.2256X_2 - 2.9442X_3 + 4.1558. \end{cases} \tag{4.9}$$

By comparing (4.8) and (4.9) we observe that the estimates for the parameters are quite good.

4.3.2 Empirical results

In order to illustrate the application of the multiple regression model introduced in this chapter, the following examples are analyzed. The first one is referred to triangular fuzzy numbers and the second one to interval data.

Example 4.3.1 We have considered the variables introduced in Example 3.1.1. In this example the aim is the analysis of the linear dependence relationship of the quality of the trees on two explanatory variables: X_1=height, X_2=diameter (see Table 1.1). The new multiple linear regression model is employed in order to analyze the problem. The spreads of the LR fuzzy response are transformed by means of the logarithmic transformation (that is g=h=ln). Through the LS procedure we obtain the following estimated models

$$\begin{cases} \widehat{Y^m} &= 0.1374X_1 + 1.7937X_2 + 19.6085 \\ \widehat{Y^l} &= \exp(0.0011X_1 - 0.1211X_2 + 2.52) \\ \widehat{Y^r} &= \exp(0.0008X_1 - 0.1471X_2 + 2.5785). \end{cases} \tag{4.10}$$

Table 4.1: Simulated data of Model (4.8)

Y_i^m	$g(Y_i^l)$	$h(Y_i^r)$	X_{1i}	X_{2i}	X_{3i}
-22.9847	-14.0403	22.7427	-0.5624	1.4596	-1.5755
1.7450	-8.5442	4.3603	1.0325	0.1697	-0.2518
14.7005	-16.2417	-3.3158	-1.2142	-0.9437	0.8747
-8.7315	-16.2936	0.6981	-0.7561	-0.7908	-0.6906
-0.1140	-18.4398	-1.0862	-1.0776	-0.9397	-0.2208
-31.3596	-15.5840	18.8256	0.3290	1.1919	-2.1974
27.5393	-10.8309	-2.7892	-0.8023	-0.3406	1.6818
18.7749	-10.5795	0.2480	-0.1462	-0.2526	1.0198
-1.1178	-9.0317	11.3000	0.5634	0.7689	-0.3822
10.7308	-7.7804	4.0573	0.5036	0.1849	0.4481
15.8898	-16.7836	-3.7861	-1.2006	-0.8845	0.9588
1.4936	-10.1194	8.9047	0.0898	0.5249	-0.0802
18.9339	-7.3200	7.8560	0.0387	0.8186	0.9623
32.2690	-9.6969	-10.5664	-0.3333	-1.1227	1.8858
-6.8706	-18.5083	4.6519	-1.2025	-0.2775	-0.3736
15.7910	-12.0064	11.1557	-0.8798	1.0485	0.9991
12.0485	-8.5603	-0.0211	0.7162	-0.0489	0.4810
-10.5228	-11.7908	23.5207	-0.2029	2.0821	-0.6985
0.2772	-11.8025	-4.9949	1.2013	-0.9967	-0.5234
-26.0635	-16.9350	20.7473	-0.9023	1.2604	-1.6059
-22.1592	-17.1332	-0.3016	0.0043	-1.2305	-1.5791
14.9907	-12.3651	-2.9761	-0.5082	-0.6562	0.8346
1.5722	-3.4690	14.5982	1.6858	1.5196	-0.3200
-18.8227	-17.7980	13.9320	-0.4883	0.5654	-1.3268
-3.1167	-15.4804	1.1831	-0.5143	-0.5197	-0.4171
-7.0028	-7.3808	-0.5069	2.0778	-0.5634	-1.2143
-46.7499	-20.6266	19.8188	-0.5760	0.7923	-2.9551
-10.1549	-19.3151	9.5865	-1.6785	0.0184	-0.5455
10.5697	-16.2964	5.3440	-1.5899	-0.0593	0.7201
10.8596	-8.2852	-4.6447	0.9365	-0.7835	0.2258

As in the simple case, we use a bootstrap procedure to estimate the standard errors $\widehat{se}$ of the parameters. In particular we draw 800 bootstrap samples of size $n = 238$ with replacement from our data set. For each bootstrap replication we calculate

the estimate of the parameters of the linear regression model. By means of the 800 replications of the estimation procedure we compute $\widehat{se}$. The estimated parameters and the estimates of their standard errors are represented in Table 4.2.

Table 4.2: Estimation of the parameters of Model (4.10) and estimation of their standard errors.

Estimator	Estimated value	Estimate of standard error
$\hat{a}_{m1}$	0.1374	0.0016
$\hat{a}_{m2}$	1.7937	0.0001
$\hat{a}_{l1}$	0.0011	0.0813
$\hat{a}_{l2}$	-0.1211	0.0007
$\hat{a}_{r1}$	0.0008	0.0814
$\hat{a}_{r2}$	-0.1471	0.0007
$\hat{b}_m$	19.6085	0.0000085
$\hat{b}_l$	2.52	0.00039437
$\hat{b}_r$	2.5785	0.00040506

There is a strong influence of the diameter on the quality of the tree ($\hat{a}_{m2} = 1.7937$), in particular, for any additional cm of the diameter the quality is expected to increase of about 1.8, while it is expected to increase of 0.14 for any additional cm of the height of the tree. The estimates of the standard errors are all close to zero.

As for the simple case, the estimated spreads represent the imprecision of the response variable while the estimates of standard error the stochastic uncertainty due to the data generation process.

Example 4.3.2 (http://www.census.gov/econ/www/). Consider the data related to the Retail Trade Sales (in millions of dollars) of the U.S. in 2002 by kind of business (see Table 3.3). As in Example 3.3.2, since the Retail Trade Sales are intervals, for each one we consider the center and the spreads. In Table 4.3 for each kind of business the Number of Employees (X_1) and the Establishments (X_2) are reported. These variables are referred to as explanatory in a multiple regression model where the Retail Trade Sale is the imprecise response.
By means of the least squares estimation the following predicted values are obtained

$$\begin{cases} \widehat{Y^m} &= 0.01817X_1 - 0.112X_2 - 559.849 \\ \widehat{Y^l} &= \exp(0.0045X_1 - 0.0188X_2 + 375.02211) \\ \widehat{Y^r} &= \exp(0.0045X_1 - 0.0188X_2 + 375.02211) \end{cases} \tag{4.11}$$

Table 4.3: Number of Employees (X_1) and Establishments (X_2) of 22 kinds of Business in the U.S. in 2002.

Kind of Business	Number of Employees	Establishments
Automotive parts, acc., and tire stores	453468	57698
Furniture stores	249807	28244
Home furnishings stores	285222	36960
Household appliance stores	69168	10330
Computer and software stores	73935	10134
Building mat. and supplies dealers	988707	67190
Hardware stores	142881	15103
Beer, wine, and liquor stores	133035	28957
Pharmacies and drug stores	783392	40234
Gasoline stations	926792	121446
Men's clothing stores	62223	9437
Family clothing stores	522164	24539
Shoe stores	205067	28499
Jewelry stores	148752	28625
Sporting goods stores	188091	22239
Book stores	133484	10860
Discount dept. stores	762309	5650
Department stores	668459	3705
Warehouse clubs and superstores	830845	2912
All other gen. merchandise stores	263116	28456
Miscellaneous store retailers	792361	129464
Fuel dealers	98574	11079

The value $a_{m1} = 0.01817$ indicates that the retail trade sales are expected to increase of about 18.170 dollars for any additional employee, while for any additional establishment the retail trade sales increase of 112.000 dollars.

As usual, the accuracy of the estimators is analyzed by means of a bootstrap procedure with 800 replications. The results are illustrated in Table 4.4.

Table 4.4: Estimation of the parameters of Model (4.11) and estimation of their standard errors.

Estimator	Estimated value	Estimate of standard error
$\widehat{\underline{a}}_m$	(0.01817,-0.112)	(0.0208,0.0012)
$\widehat{\underline{a}}_l$	(0.0045,-0.188)	(0.0507,0.0041)
$\widehat{\underline{a}}_r$	(0.0045,-0.188)	(0.0507,0.0041)
$\hat{b}_m$	-559.849	0.00000005442
$\hat{b}_l$	375.0211	0.00000014294
$\hat{b}_r$	375.0211	0.00000014294

4.4 Confidence regions and hypothesis testing on the regression parameters

In addition to the estimation of the regression parameters, as in the simple case, the confidence regions and the hypothesis test are introduced. Starting from the asymptotic distribution (4.7) it is easily obtained the following $100(1-\alpha)$ confidence region for the parameters $(\underline{a}_m^{'}, \underline{a}_l^{'}, \underline{a}_r^{'})^{'}$

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

where $c_{\alpha/2}$ is a $\alpha/2$-quantile of a $N\left(\underline{0}^{'}, \frac{\Sigma}{\Sigma_{\underline{X}}}\right)$.

In order to test the null hypothesis

$$H_0 : \begin{pmatrix} \underline{a}_m^{'} \\ \underline{a}_l^{'} \\ \underline{a}_r^{'} \end{pmatrix} = \begin{pmatrix} \underline{k}_m^{'} \\ \underline{k}_l^{'} \\ \underline{k}_r^{'} \end{pmatrix} \tag{4.12}$$

against the alternative

$$H_1 : \begin{pmatrix} \underline{a}_m^{'} \\ \underline{a}_l^{'} \\ \underline{a}_r^{'} \end{pmatrix} \neq \begin{pmatrix} \underline{k}_m^{'} \\ \underline{k}_l^{'} \\ \underline{k}_r^{'} \end{pmatrix},$$

where $\underline{k}_m$, $\underline{k}_l$, and $\underline{k}_r$ are vectors of constant values in $\mathbb{R}$, the test statistic $T_n = V_n' V_n$, where

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

can be used. As in the simple case it is possible to define a rejection region for the null hypothesis, that is

Proposition 4.4.1 *In testing the null hypothesis (4.12) at the nominal significance level α, H_0 should be rejected if*

$$T_n > c_\alpha,$$

where c_α is a α-quantile of the asymptotic distribution of T_n, that is $f_1(V)$ ($V \sim N\left(\underline{0}', \frac{\Sigma}{\Sigma_{\underline{X}}}\right)$ and $f_1(A) = A'A$).

Analogously to Section 3.5.2 also in this case a bootstrap approach can be developed. Thus, the new variables $Z^m = Y^m - \underline{X}\,\underline{\widehat{a}}_m^{'} + \underline{X}\,\underline{k}_m^{'}$, $Z^l = g(Y^l) - \underline{X}\,\underline{\widehat{a}}_l^{'} + \underline{X}\,\underline{k}_l^{'}$ and $Z^r = h(Y^r) - \underline{X}\,\underline{\widehat{a}}_r^{'} + \underline{X}\,\underline{k}_r^{'}$ are considered, in order to obtain a bootstrap population satisfying the null hypothesis (4.12). A sample of size n with replacement $\left\{(\underline{X}_i^*, Z^{m*}{}_i, Z^{l*}{}_i, Z^{r*}{}_i)\right\}_{i=1,\ldots,n}$ from the bootstrap population is drawn and, as bootstrap statistic, $T_n^* = V_n^{*'} V_n^*$, where

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

and

$$\begin{aligned} \underline{\widehat{a}}_m^{*\,'} &= (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\widetilde{\underline{Z}^m}', \\ \underline{\widehat{a}}_l^{*'} &= (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\widetilde{\underline{Z}^l}', \\ \underline{\widehat{a}}_r^{*'} &= (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\widetilde{\underline{Z}^r}', \end{aligned}$$

($\widetilde{\underline{Z}^m}$, $\widetilde{\underline{Z}^l}$, $\widetilde{\underline{Z}^r}$ are the centered vector of the bootstrap variables and $\widetilde{\mathbf{X}}$ is the centered matrix) is used. It can be easily proved that, as $n \to \infty$

$$T_n^* \xrightarrow{D} f_1(V), \tag{4.13}$$

where $V \sim N\left(\underline{0}', \frac{\Sigma}{\Sigma_{\underline{X}}}\right)$, and analogously to Proposition 3.5.3, it follows

Proposition 4.4.2 *In testing the null hypothesis (4.12) at the nominal significance level α, H_0 should be rejected if*

$$T_n^* > c_\alpha,$$

where c_α is a α-quantile of the asymptotic distribution of T_n^*.

By means of the following algorithm, as usual, the test in Proposition 4.4.2 can be employed.

Algorithm

Step 1: Compute the estimate vectors $\widehat{\underline{a}}_m$, $\widehat{\underline{a}}_l$ and $\widehat{\underline{a}}_r$ and the value of the statistic

$$T_n = V_n' V_n$$

Step 2: Compute the bootstrap population

$$\left\{(\underline{X}_i, Z_i^m, Z_i^l, Z_i^r)\right\}_{i=1,\ldots,n}, \tag{4.14}$$

where

$$\begin{aligned} Z_i^m &= Y_i^m - \underline{X}_i \widehat{\underline{a}}_m^{'} + \underline{X}_i \underline{k}_m^{'}, \\ Z_i^l &= g(Y_i^l) - \underline{X}_i \widehat{\underline{a}}_l^{'} + \underline{X}_i \underline{k}_l^{'}, \\ Z_i^r &= h(Y_i^r) - \underline{X}_i \widehat{\underline{a}}_r^{'} + \underline{X}_i \underline{k}_r^{'}. \end{aligned}$$

Step 3: Draw a sample of size n with replacement

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

from the bootstrap population (4.14).

Step 4: Compute the value of the bootstrap statistic

$$T_n^* = V_n^{*'} V_n^*$$

Step 5: Repeat Steps 3 and 4 a large number B of times to get a set of B estimators, denoted by $\{T_1^*, ..., T_B^*\}$.

Step 6: Compute the bootstrap p-value as the proportion of values in $\{T_1^*, ..., T_B^*\}$ such that being greater than T_n.

As for the simple case in Section 3.5.3, it is possible to analyze the asymptotic power function under a sequence of *local alternatives*

Proposition 4.4.3 *Consider the null hypothesis (4.12) against the alternative H_1, the statistic T_n and the critical region $(T_n > k)$. Let H_n be the sequence of Pitman alternatives verifying*

$$\begin{pmatrix} \underline{a}_m^{'} \\ \underline{a}_l^{'} \\ \underline{a}_r^{'} \end{pmatrix} = \begin{pmatrix} \underline{k}_m^{'} \\ \underline{k}_l^{'} \\ \underline{k}_r^{'} \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} \underline{\delta}_m^{'} \\ \underline{\delta}_l^{'} \\ \underline{\delta}_r^{'} \end{pmatrix},$$

where $|\underline{\delta}| > 0$. Then

1. *Under* H_n, $T_n \xrightarrow{D} f_1(V)$, *where* $V \sim N\left(\underline{\delta}', \frac{\Sigma}{\Sigma_{\underline{X}}}\right)$;
2. *If we consider the sequence of local alternatives for which* $\underline{\delta} = \underline{\delta_n}$, *with* $|\underline{\delta_n}| \longrightarrow \infty$,

$$\lim_{n\to\infty} P_{H_n}(T_n > k) = 1.$$

4.4.1 Simulations

In order to illustrate the empirical significance of the bootstrap test proposed in Proposition 4.4.2, a simulated situation has been taken into account. For the simulations we have considered $B = 1000$ replications of the bootstrap estimator and we have carried out 10.000 iterations of the test at 3 different nominal significance levels $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$ for different sample sizes n. Two simulation cases are presented. The first one considers real random variables X_1, X_2, ε_m, ε_l and ε_r behaving as independent $N(0,1)$ random variables. The empirical percentages of rejection under $H_0 : (\underline{a}_m', \underline{a}_l', \underline{a}_r')' = (\underline{1}', \underline{1}', \underline{1}')'$ are represented in Table 4.5. In this case for $\alpha = 0.05$ and $\alpha = 0.1$, it results that for $n \geq 100$ the empirical percentages of rejection are quite close to the nominal levels.

Table 4.5: Empirical percentages of rejection under $H_0 : (\underline{a}_m', \underline{a}_l', \underline{a}_r')' = (\underline{1}', \underline{1}', \underline{1}')'$ (case of normality).

$n \setminus \alpha \times 100$	1	5	10
30	0.33	2.77	7.97
50	0.56	3.78	9.08
100	0.61	4.78	9.92
200	0.8	4.92	9.95

In the second one we deal with the following real random variables: X_1, behaving as an $Unif(-2,3)$ random variable, X_2, behaving as an $Unif(1,6)$ random variable, ε_m, ε_l and ε_r behaving as independent $N(0,1)$ random variables. The empirical percentages of rejection under $H_0 : (\underline{a}_m', \underline{a}_l', \underline{a}_r')' = (\underline{1}', \underline{1}', \underline{1}')'$ are represented in Table 4.6. By applying the bootstrap procedure, for $n \geq 100$ the empirical percentages of rejection are quite close tho the the three nominal levels.

Empirical results

In order to illustrate the bootstrap test introduced in Section 3.5.2 a real life example is considered . Taking into account the LR fuzzy data in Table 1.1, to test

Table 4.6: Empirical percentages of rejection under $H_0 : (\underline{a}_m', \underline{a}_l', \underline{a}_r')' = (\underline{1}', \underline{1}', \underline{1}')'$ (case of non-normality).

$n \setminus \alpha \times 100$	1	5	10
30	0.5	4.14	9.9
50	0.72	4.42	9.86
100	0.97	5.1	10.11
200	1.04	5.17	10.04

if the vector of regression parameters $(\underline{a}_m', \underline{a}_l', \underline{a}_r')'$ is equal to $(\underline{3}', \underline{3}', \underline{3}')'$, $B = 1000$ replications of the bootstrap statistic are used and a p-value equal to 0.02 is obtained. Hence the considered hypothesis should be rejected. In testing if the vector $(a_{m1}, a_{m2}, a_{l1}, a_{l2}, a_{r1}, a_{r2})'$ is equal to a vector whose elements are approximately equal to the estimations of the parameters, that is $(0.14, 1.8, 0.02, -0.1, 0.001, -0.2)'$, a p-value equal to 0.993 is obtained. Obviously the hypothesis tested should not be rejected.

4.5 Estimation of the multiple determination coefficient

As in the previous chapter, we can define an estimator for the multiple determination coefficient. Using the same scheme, the next proposition proves the decomposition of the total sum of squares and on this basis we can define the estimator.

Proposition 4.5.1 *Let Y be an LR fuzzy random variable and $\underline{X}$ a vector of real random variables satisfying the linear model (4.1) observed on n statistical units, $\{Y_i, \underline{X}_i\}_{i=1,\ldots,n}$. The total sum of squares, SST, is equal to the sum of the residual sum of squares, SSE, and the regression sum of squares, SSR, that is*

$$SST = SSE + SSR. \tag{4.15}$$

In details,

(i) *the total sum of squares (SST) is*

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

(ii) *the residual sum of squares (SSE) is*

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

(iii) *the regression sum of squares (SSR) is*

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

where $\widehat{\underline{Y}^{m}}'$, $\widehat{g(\underline{Y}^{l})}'$, $\widehat{h(\underline{Y}^{r})}'$ *are the vectors of the estimated values, that is,*

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

and $\overline{\underline{Y}^{m}} = (\underline{1}\,\underline{1}')^{-1}\underline{1}\underline{Y}^{m\prime}$, $\overline{g(\underline{Y}^{l})} = (\underline{1}\,\underline{1}')^{-1}\underline{1}g(\underline{Y}^{l})'$, $\overline{h(\underline{Y}^{r})} = (\underline{1}\,\underline{1}')^{-1}\underline{1}h(\underline{Y}^{r})'$ *are the vectors of the sample means of the response variables.*

Proposition 4.5.2 *Let Y be an LR fuzzy random variable and $\underline{X}$ a vector of real random variables satisfying the linear model (4.1), observed on n statistical units, $\{Y_i, \underline{X}_i\}_{i=1,\ldots,n}$. The estimator of the determination coefficient R^2 is*

$$\widehat{R}^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}.$$

Proposition 4.5.3 *The estimator $\widehat{R}^2$ is strongly consistent.*

4.5.1 Simulations

In order to illustrate the accuracy of the estimator of the multiple determination coefficient, a simulation study is considered. Three explanatory variables X_1, X_2 and X_3 have been generated normally distributed as $N(0,1)$ and an LR fuzzy response Y has been generated in the following way: the center Y^m normally distributed as $N(0,1)$, the left and the right spread as χ^2_1. A logarithmic transformation has been used for both spreads. It follows that the multiple determination coefficient R^2 is null because the variables have been independently generated. Taking into account different sample sizes, the idea is to calculate the estimate of R^2 for each sample size and to show that the estimated values are closer to 0, as the sample size n increases.

The results presented in Table 4.7 show the consistency of the estimator.

Table 4.7: Estimated values $\hat{R}^2$ (multiple) for samples of different size.

n	$\hat{R}^2$	n	$\hat{R}^2$
30	0.0518	800	0.0048
50	0.0373	1000	0.0011
100	0.0266	2000	0.00074932
500	0.0082	3000	0.0003759

4.5.2 Empirical results

The estimator of the determination coefficient, $\widehat{R}^2$, referred to Example 4.3.1 is equal to 0.2567. This value indicates that approximately almost 25.67% of the total variation is explained by the multiple regression model taken into account. Obviously as the number of explanatory variables increases the determination coefficient referred to the model improves. In details, compared with the simple model of Example 3.3.1 in Chapter 3, this multiple case explains approximately only 0.28% more of the total variation.

Taking into account the data set of Example 4.3.2, $\widehat{R}^2 = 0.9175$ is put up. Approximately almost 92% of the total variation of the retail trade sale is explained by means of the multiple model with Number of Employees and Estabishments as explanatory variables. To insert the variable Estabishments in the model entails an increment of $\widehat{R}^2$ of 0.0019%.

4.6 Linear independence test

In this section a linear independence test is introduced. To test the null hypothesis $H_0 : R^2 = 0$ against the alternative $H_1 : R^2 > 0$, the test statistic $T_n = n\widehat{R}^2$ is used. Taking into account that, under the assumption of model (4.1) and under the null hypothesis of linear independence, as $n \to \infty$

$$n\widehat{R}^2 \xrightarrow{D} \frac{f_2(W)}{\sigma^2_{\widetilde{Y}}}, \tag{4.16}$$

where $W \sim N(\underline{0}', \Sigma)$, analogously to Proposition 3.7.2, it follows the next asymptotic procedure.

Proposition 4.6.1 *In testing the null hypothesis of linear independence at the nominal significance level α, H_0 should be rejected if*

$$T_n > c_\alpha,$$

where c_α is a α-quantile of the asymptotic distribution of T_n, that is $\frac{f_2(W)}{\sigma^2_{\widetilde{Y}}}$, where $W \sim N(\underline{0}', \Sigma)$ and f_2 is the function introduced in Proposition 3.7.1.

Following the same idea of Section 3.7.2 a more efficient bootstrap approach can be developed. Thus, the residual variables $Z^m = Y^m - \underline{X}\,\widehat{\underline{a}}_m'$, $Z^l = g(Y^l) - \underline{X}\,\widehat{\underline{a}}_l'$ and $Z^r = h(Y^r) - \underline{X}\,\widehat{\underline{a}}_r'$ are used, in order to obtain a bootstrap population. A sample of size n with replacement $\left\{(\underline{X}^*_i, {Z^m}^*_i, {Z^l}^*_i, {Z^r}^*_i)\right\}_{i=1,\ldots,n}$ from the bootstrap population is drawn and, as bootstrap statistic,

$$T^*_n = n\frac{\sum_{i=1}^{n} D^2_{\lambda\rho}(\widehat{\widetilde{Z^*}}_i, \overline{\widetilde{Z^*}})}{\sigma^2_{\widetilde{Y}}}$$

($\widetilde{Z^*}_i = ({Z^m}^*_i, {Z^l}^*_i, {Z^r}^*_i)$) is used. It is easy to check the same results of Propositions 3.7.3 and 3.7.4, that are an asymptotic distribution of the bootstrap statistic and a bootstrap test for the linear independence. In particular, as $n \to \infty$

$$T^*_n \xrightarrow{D} \frac{f_2(W)}{\sigma^2_{\widetilde{Y}}}, \tag{4.17}$$

($W \sim N(\underline{0}', \Sigma)$ and f_2 is the function of Proposition 3.7.1) and analogously to Proposition 3.7.4, it follows

Proposition 4.6.2 *In testing the null hypothesis of linear independence at the nominal significance level α, H_0 should be rejected if*

$$T^*_n > c_\alpha,$$

where c_α is a α-quantile of the asymptotic distribution of T^*_n.

The application of the test in Proposition 4.6.2 is presented in the following algorithm.

Algorithm

Step 1: Compute the estimate vectors $\widehat{\underline{a}}_m$, $\widehat{\underline{a}}_l$ and $\widehat{\underline{a}}_r$ and the value of the statistic

$$T_n = n\hat{R}^2 = n\frac{\sum_{i=1}^{n} D^2_{\lambda\rho}(\widehat{\widetilde{Y}}_i, \overline{\widetilde{Y}})}{\sum_{i=1}^{n} D^2_{\lambda\rho}(\widetilde{Y}_i, \overline{\widetilde{Y}})}$$

Step 2: Compute the bootstrap population

$$\left\{(\underline{X}_i, Z^m_i, Z^l_i, Z^r_i)\right\}_{i=1,\ldots,n}, \tag{4.18}$$

where

$$\begin{aligned} Z_i^m &= Y_i^m - \underline{X}_i \widehat{\underline{a}}_m^{'}, \\ Z_i^l &= g(Y_i^l) - \underline{X}_i \widehat{\underline{a}}_l^{'}, \\ Z_i^r &= h(Y_i^r) - \underline{X}_i \widehat{\underline{a}}_r^{'}. \end{aligned}$$

Step 3: Draw a sample of size n with replacement

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

from the bootstrap population (4.18).

Step 4: Compute the value of the bootstrap statistic

$$T_n^* = n \frac{\sum_{i=1}^{n} D_{\lambda\rho}^2(\widetilde{\widehat{Z^*}}_i, \overline{\widetilde{Z^*}})}{\sigma_{\widetilde{Y}}^2}$$

where $\widetilde{Z^*}_i = (Z^{m*}{}_i, Z^{l*}{}_i, Z^{r*}{}_i)$.

Step 5: Repeat Steps 3 and 4 a large number B of times to get a set of B estimators, denoted by $\{T_1^*, ..., T_B^*\}$.

Step 6: Compute the bootstrap p-value as the proportion of values in $\{T_1^*, ..., T_B^*\}$ being greater than T_n.

4.6.1 Simulations

As usual we have used simulations in order to illustrate the empirical significance of the bootstrap test. We have carried out 10.000 iterations of the test at 3 different nominal significance levels $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$ for different sample sizes n. $B = 1000$ bootstrap replications have been considered.

Two simulation cases are presented. The first one is referred to real random variables X_1, X_2, X_3, ε_m, ε_l and ε_r behaving as a $N(0,1)$ random variable. The empirical percentages of rejection under H_0 are represented in Table 4.8. The conclusions are better, as the sample size n increases, that is, the empirical percentages of rejection are closer to the nominal levels.

In the second simulation case the following variables have been considered: X_1, X_2, ε_m, ε_l and ε_r behaving, respectively, as an $Unif(0,10)$, an $Unif(-2,2)$, an $Unif(-1,4)$, an $Unif(5,9)$ and an $Unif(0,6)$ independent random variables.

Table 4.8: Empirical percentages of rejection under the hypothesis of linear independence (case of normality).

$n \setminus \alpha \times 100$	1	5	10
50	1.40	5.83	11.03
100	1.36	5.48	10.83
200	1.27	5.32	10.67
300	1.09	5.09	10.14

Table 4.9: Empirical percentages of rejection under the hypothesis of linear independence (case of non-normality).

$n \setminus \alpha \times 100$	1	5	10
50	1.2	5.58	10.79
100	1.17	5.6	10.37
200	1.27	5	9.76
300	0.94	5.37	10.49

4.6.2 Empirical results

As in Section 3.7.2 the bootstrap test defined in Section 4.6 has been employed on two real life examples. The first one considers the LR fuzzy data in Table 1.1 and the second one is referred to the data in Table 4.3. For the simulations $B = 1000$ replications of the bootstrap estimator are used. Also in this multiple case for both examples the p-value is equal to 0. In both cases the null hypothesis of linear independence should be rejected.

4.6.3 Local alternatives

A study about the power function of the linear independence test is also presented for the multidimensional case. Due to the difficulties of this kind of analysis, a sequence of *local alternatives* is used for verifying how sensible the test is under small deviations from null hypothesis.

Proposition 4.6.3 *Consider the null hypothesis $H_0 : R^2 = 0$ of the linear independence test against the alternative H_1. Let T_n be the test statistic and $(T_n > k)$ the*

critical region. Let H_n be the sequence of Pitman alternatives verifying

$$\begin{pmatrix} \underline{a}_m^{'} \\ \underline{a}_l^{'} \\ \underline{a}_r^{'} \end{pmatrix} = \begin{pmatrix} \underline{0}^{'} \\ \underline{0}^{'} \\ \underline{0}^{'} \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} \underline{\delta}_m^{'} \\ \underline{\delta}_l^{'} \\ \underline{\delta}_r^{'} \end{pmatrix},$$

where $|\underline{\delta}| > 0$. Then

1. *Under H_n, $T_n \xrightarrow{D} \frac{f_2(W)}{\sigma^2_{\tilde{Y}}}$, where $W \sim N(\underline{\delta}^{'} (\Sigma_{\underline{X}})^{1/2}, \Sigma)$;*
2. *If If we consider the sequence of local alternatives for which $\underline{\delta} = \underline{\delta_n}$, with $|\underline{\delta_n}| \longrightarrow \infty$,*

$$\lim_{n\to\infty} P_{H_n}(T_n > k) = 1.$$

4.7 Final evaluation and open problems

In this chapter we have carried out a wide statistical analysis concerning a multiple regression model to express an imprecise response as a function of crisp explanatory variables. Since this model is a multidimensional extension of the model proposed in Chapter 3, final evaluation and open problems are similar. In addition, as open problems pertaining to this chapter, it could be interesting

- The analysis of multicollinearity problem that may be faced with the application of Principal Component Analysis. The idea consists in replacing the original variables by a set of uncorrelated artificial variables (principal components).
- The study of a selection procedure to obtain the appropriate number of explanatory variables to be used, based on the goodness-of-fit coefficient, by following the same reasoning proposed in D'Urso & Santoro (2006).

Epilogue

In this work a regression analysis to model statistical relationships between imprecise and real elements has been developed. In particular a linear regression model for LR fuzzy response and scalar predictors has been introduced and analyzed. In a classical framework one of the main difficulties is related to the condition of non-negativity of the spreads. The introduction of suitable functions g and h that transform the spreads into real numbers and of an appropriate metric, $D_{\lambda\rho}$, has made it possible to solve the problem.

In this work three kinds of uncertainty are taken into account: the relationship between response and explanatory variables; the relationship between the observed data and the universe of possible data (randomness due to the generation of the data); the observed value of the variables (imprecision). The first one has been handled by means of linear regression models, the second and the third ones by considering fuzzy random variables.

Some basic concepts have been introduced, in order to handle random experiments for which the observed characteristic is imprecise on the results. It has been also discussed the adequacy for the practical situations with which we treat and its coherence. Some regression models in a fuzzy framework have been introduced. In particular, the model proposed by Diamond (1988), that is one of the first works with fuzzy elements using the least squares criterion, and the model introduced by González-Rodríguez *et al.* (2009) that considers fuzzy random variables and it is the first work that presents a complete solution, have been briefly described. Finally the model analyzed by Coppi *et al.* (2006) that proposes a linear regression model with LR fuzzy response, from which this work has taken inspiration. The authors have taken into account the three kinds of uncertainty but they handle the randomness by means of a bootstrap procedure. This model has not been formalized based on fuzzy random variables, and to develop this formalization the new regression model, presented in this work, has come up.

The concept of variance in the sense of the $D_{\lambda\rho}$-metric by following the ideas in Körner (1997) and Lubiano *et al.* (2000) has been defined and some properties, necessary to apply the least squares criterion, have been proved. In further research an

asymptotic distribution of the sample variance could be determined for developing confidence intervals and hypothesis testing procedures.

A wide statistical analysis concerning a regression model to express an imprecise response as a function of a real explanatory variable has been carried out. In details, the least squares estimators have been found, and some confidence intervals and testing procedures have been developed on the basis of their asymptotic distributions. Some bootstrap techniques have been considered in order to improve the empirical results for small/moderate sample sizes and we have shown by means of some simulations their suitability in practice. A determination coefficient has been defined and an estimator has been analyzed. In addition a test to check the goodness-of-fit of the model has been developed on the basis of this estimator. Some analysis of power of the tests through local alternatives has showed that the test is asymptotically consistent.

All these analysis have been also developed in the multiple case, that is simply a multidimensional extension of the simple case.

For future works several open problems can be indicated. In particular, it could be interesting to find an appropriate family of functions g and h to transform the spreads of the LR response variables and to introduce semi-parametric models. Since also the explanatory variables in some cases have to fulfill some conditions, in order to face this restriction non-linear models could be introduced. Concerning the multiple regression model, it could be interesting to analyze the problem of multicollinearity, for example by means a preliminary Principal Component Analysis, and to study a selection procedure to obtain the appropriate number of explanatory variables to be used, based on the goodness-of-fit coefficient.

Bibliography

Arstein, Z. & Vitale R.A. (1975). A strong law of large numbers for random compact sets. *Annals of Probability.* **5**, 879–882.

Aumann, R.J. (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. *Mathware & Soft Computing* **2**, 71–84.

Bickel, P.J. & Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *The annals of statistics.* **9**, 1196–1217.

Billard, L. & Diday, E. (2000). Regression analysis for interval-valued data, in Kiers, H.A.L., Rasson, J.P., Groenen, P.J.F., Schader, M., (Eds.): *Data analysis, classification and related methods*, Springer Verlag, Berlin, 369–374.

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

Casella, G. & Berger, R.L. (2002). Statistical Inference. *Duxbury Advanced Series, Pacific Grove.*

Celminš, A. (1987). Least squares model fitting to fuzzy vector data. *Fuzzy sets and systems.* **22**, 245–269.

Celminš, A. (1987). Multidimensional least-squares fitting of fuzzy models. *Math. Model.* **9**, 669–690.

Chang, P.T. & Lee, E.S. (1994). Fuzzy linear regression with spreads unrestricted in sign. *Computers and Mathematics with Applications.* **28**, 61–70.

Chang, P.T. & Lee, E.S. (1996). A generalized fuzzy least-squares regression. *Fuzzy Sets and Systems.* **82**, 289–298.

Chang, Y.H. (2001). Hybrid fuzzy least-squares regression analysis and its reliability measures. *Fuzzy Sets and Systems.* **119**, 225–246.

Chang, Y.T. & Ayyub, B.M. (2001). Fuzzy regression methods–a comparative assessment. *Fuzzy sets and systems.* **119**, 187–203.

Cheng, K.F. & Chen, L.C. (2004). Testing goodness-of-fit of a logistic regression model with case-control data. *Journal of Statistical Planning and Inference.* **124**, 409–422.

Colubi, A., López-Díaz, M., Domínguez-Menchero, J.S., Gil. M.A. (1999). A generalized strong law of large numbers. *Probability Theory and Related Fields.* **114**, 401–417.

Colubi, A., Domínguez-Menchero, J.S., López-Díaz, M., Ralescu D.A. (2001). On the formalization of fuzzy random variables. *Information Science.* **133**, 3–6.

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

Coppi, R. (2003). The fuzzy approach to multivariate statistical analysis. *Technical Report n. 11, Department of Statistics, Probability and Applied Statistics, University of Rome "La Sapienza".*

Coppi, R. (2008). Management of uncertainty in statistical reasoning: the case of regression analysis. *International Journal of Approximate Reasoning.* **47**, 284–305.

Coppi, R. & D'Urso, P. (2003). Regression analysis with fuzzy informational paradigm: a least-squares approach using membership function information. *Int. J. Pure Appl. Math.* **8**, 279–306.

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

Coppi, R., Gil, M.A., Kiers, H.A.L. (2006). The fuzzy approach to statistical analysis. *Comp. Stat. Data Anal.* **51**, 1–14.

Cox, D.R. & Hinkley, D. (1974). Theoretical Statistics. *Chapman and Hall, London.*

D'Urso, P. (2003). Linear regression analysis for fuzzy/crisp input and fuzzy /crisp output data. *Comp. Stat. Data Anal.* **42**, 47–72.

D'Urso, P. & Gastaldi, T. (2000). A least-squares approach to fuzzy linear regression analysis. *Comp. Stat. Data Anal.* **34**, 427–440.

D'Urso, P. & Gastaldi, T. (2002). An orderwise polynomial regression procedure for fuzzy data. *Fuzzy sets and systems.* **130**, 1–19.

D'Urso, P. & Gastaldi, T. (2003). Fitting of fuzzy linear regression models with multivariate response. *Int. Math. J.* **3**, 655–664.

D'Urso, P. & Santoro, A. (2006). Goodness of fit and variable selection in the fuzzy multiple linear regression. *Fuzzy Sets and Systems.* **157**, 2627–2647.

Dette, H. & Neumeyer, N. (2001). Nonparametric analysis of covariance. *The Annals of Statistics* **29**, 1361–1400.

Diamond, P. (1988). Fuzzy least squares. *Inform. Sci.* **46**, 141–157.

Diamond, P. & Kloeden, P. (1990). Metric spaces of fuzzy sets. *Fuzzy sets and systems.* **35**, 241–249.

Diamond, P. & Kloeden, P. (1994). Metric spaces of fuzzy sets: Theory and applications. *World Scientific, Singapore.*

Diamond, P., and Körner, R. (1997). Extended fuzzy linear models and least squares estimates. *Computers and Mathematics with Applications.* **33**, 15–32.

Di Lascio, L., Ginolfi, L., Albunia, A., Galardi, G., meschi, F. (2002). A fuzzy-based methodology for the analysis of diabetic neuropathy. *Fuzzy sets and systems.* **129**, 203–228.

Dubois, D. & Prade, H. (1978). Operations on fuzzy numbers. *Int. J. Syst. Sci.* **9**, 613–626.

Dubois, D. & Prade, H. (1980). *Fuzzy sets and systems: theory and applications.* Academic Press, New York.

Dubois, D. & Prade, H. (1988). Possibility theory. *Plenum Press., New York.*

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

Filmoser, P. & Viertl, R. (2004). Testing hypotheses with fuzzy data: The fuzzy p-value. *Metrika.* **59**, 21–29.

Frèchet, M. (1948). Les èlèments alèatoires de natures quelconque dans un àspace distanciè. *Ann. Inst. H. Poincarè.* **10**, 215–310.

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

Gil, M. A., Corral, N., Gil, P. (1988). The minimum inaccuracy estimates in χ^2 tests for goodness of fit with fuzzy observations. *Journal of Statistical Planning and Inference.* **19**, 95–115.

Gil, M. A., Lubiano, M. A., Montenegro, M., Lòpez, M. T. (2002). Least squares fitting of an affine function and strength of association for interval-valued data. *Metrika.* **56**, 97–111.

Gil, M. A., López-Díaz, M., Ralescu, D.A. (2006). Overview on the development of a fuzzy random variable. *Fuzzy Sets and Systems.* **157**, 2546–2557.

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

Gil, M. A., González-Rodríguez, G., Colubi, A., Montenegro, M. (2007). Testing linear independence in linear models with interval-valued data. *Computational Statistics & Data Analysis.* **51**, 3002–3015.

Giné, E., Zinn, J. (1990). Bootstrapping general empirical measures. *The Annals of Probability.* **18**, 851–869.

González-Rodríguez, G., Montenegro, M., Colubi, A., Gil, M.A. (2006). Bootstrap techniques and fuzzy random variables: synergy in hypothesis testing with fuzzy data. *Fuzzy Sets and Systems.* **157**, 2608–2613.

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

Grzegorzewski, P. (2000). Testing statistical hypotheses with vague data. *Fuzzy Sets and Systems.* **112**, 501–510.

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

Hukuhara, M. (1967). Integration des applications mesurables dont la valeur est un compact convexe. *Funk. Ekvacioj.* 205–223.

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

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

Körner, R. (1997b). On the variance of fuzzy random variables. *Fuzzy sets and systems.* **92**, 83–93.

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

Körner, R. (2000). An asymptotic α-test for the expectation of random fuzzy variables. *Journal of Statistical Planning and Inference.* **83**, 331–346.

Körner, R. & Näther, W. (2002). On the variance of random fuzzy variables. *In: C. Bertoluzza, M.A. Gil, and D.A. Ralescu (Eds), Statistical Modeling, Analysis and Management of Fuzzy Data.* 22–39.

Krätschmer, V. (2001). A unified approach to fuzzy random variables. *Fuzzy sets and systems.* **123**, 1–9.

Krätschmer, V. (2002a). Limit theorems for fuzzy-random variables. *Fuzzy sets and systems.* **126**, 253–263.

Krätschmer, V. (2002b). Some complete metrics on spaces of fuzzy subsets. *Fuzzy sets and systems.* **130**, 357–365.

Krätschmer, V. (2004). Least-squares estimation in linear regression models with vague concepts. *In: Lopez-Diaz, M., Gil, M.A., Grzgorzewski, P., Hryniewicz, P., Lawry, J. (Eds), Soft Methodology and Random Information Systems. springer, Heidelberg.* 407–414.

Krätschmer, V. (2006a). Least-squares estimation in linear regression models with vague concepts. *Fuzzy sets and systems.* **157**, 2579–2592.

Krätschmer, V. (2006b). Strong consistency of least-squares estimation in linear regression models with vague concepts. *Journal of Multivariate Analysis.* **97**, 633–654.

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

Kruse, R. (1982). The strong law of large numbers for fuzzy random variables. *Inform. Sci.* **28**, 233–241.

Kruse, R. & Meyer, K.D. (1987). Statistics with vague data. *Reidel, Dordrecht, Boston.*

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

Kwakernaak, H. (1979). Fuzzy random variables-II. Algorithms and examples for discrete case. *Inform. Sci.* **17**, 253–278.

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

Laviolette, M., Seaman, J.W., Barrett, J.D., Woodall, W.H. (1995). A probabilistic and statistical view of fuzzy methods. *Technometrics.* **37**, 249–292 (with discussion).

Lee, H.T & Chen, S.H. (2001). Fuzzy regression model with fuzzy input and output data for manpower forecasting. *Fuzzy Sets and Systems.* **119**, 205–213.

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

López-Díaz, M. & Gil. M.A. (1997). Constructive definitions of fuzzy random variables. *Statistics & Probability Letters.* **36**, 135–143.

Lubiano, M. A. & Gil, M.A. (1999). Estimating the expected value of fuzzy random variables in random samplings from finite populations. *Statistical Papers.* **40**, 277–295.

Lubiano, M. A., Gil, M.A., López-Díaz, M., López, M.T. (2000). The $\vec{\lambda}$-mean squared dispersion associated with a fuzzy random variable. *Fuzzy sets and systems.* **111**, 307–317.

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

Molchanov, I. (2005). Theory of random sets. *Probability and Its Applications, Springer.*

Montenegro, M., Colubi, A., Casals, M.R., Gil, M.A. (2004). asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika.* **59**, 31–49.

Näther, W. (1997). Linear statistical inference for random fuzzy data. *Statistics.* **29**, 221–240.

Näther, W. (2000). On random fuzzy variables of second order and their application to linear statistical inference with fuzzy data. *Metrika.* **51**, 201–221.

Näther, W. (2001). Random fuzzy variables of second order and applications to statistical inference. *Information Sciences* **133**, 69–88.

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

Nguyen, H.T. (1997). Fuzzy sets and probability. *Fuzzy Sets and Systems* **90**, 129–132.

Nguyen, H.T. (2005). Fuzzy and random. *Fuzzy Sets and Systems* **156**, 349–356.

Nikitin, Y. (1995). Asymptotic efficiency of nonparametric tests. *Cambridge University Press.*

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

Puri, M.L. & Ralescu, D.A. (1985). The concept of normality for fuzzy random variables. *Ann. Probab.* **13**, 1373–1379.

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

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

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

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

Tanaka, H., Uejima, S., Asai, K. (1982). Linear regression analysis with fuzzy model. *IEEE Trans. Systems Man Cybernet.* **12**, 903–907.

Viertl, R. (1996). Statistical methods for non-precise data. *CRC Press, Boca Raton, New York, London, Tokyo.*

Viertl, R. (2006). Univariate statistical analysis with fuzzy data. *Computational Statistics & Data Analysis*. **51**, 133–147.

Vitale, R.A. (1985). L_p Metrics for compact, convex sets. *Journal of Approximation Theory*. **45**, 280–287.

Walley, P. (1997). A bounded derivative model for prior ignorance about a real-valued parameter. *Scandinavian Journal of Statistics*. **24**, 463–483.

Wu, H.C. (1999). Probability density functions of fuzzy random variables. *Fuzzy sets and systems*. **105**, 139–158.

Wünsche, A. & Näther, W. (2002). Least-squares fuzzy regression with fuzzy random variables. *Fuzzy sets and systems*. **130**, 43–50.

Xu, R. & Li, C. (2001). Multidimensional least-squares fitting with a fuzzy model. *Fuzzy sets and systems*. **119**, 215–223.

Yang, M.S. & Ko, C.H. (1996). On a class of fuzzy c-numbers clustering procedures for fuzzy data. *Fuzzy sets and systems.* **84**, 49–60.

Yang, M.S. & Lin, T.Z. (2002). Fuzzy least-squares linear regression analysis for fuzzy input-output data. *Fuzzy Sets and Systems.* **126**, 389–399.

Yang, M.S. & Liu, H.H. (2003). Fuzzy least-squares algorithms for interactive fuzzy linear regression models *Fuzzy Sets and Systems.* **135**, 305–316.

Yen, K.K, Ghoshray, S., Roig, G. (1999). A linear model using triangular fuzzy number coefficients. *Fuzzy Sets and Systems.* **106**, 167–177.

Zadeh, L.A. (1965). Fuzzy sets. *Information and control.* **8**, 338–353.

Zadeh, L.A. (1975a). The concept of a linguistic variable and its applications to approximate reasoning, Part 1. *Information Science.* **8**, 199–249.

Zadeh, L.A. (1975b). The concept of a linguistic variable and its applications to approximate reasoning, Part 2. *Information Science.* **8**, 301–357.

Zadeh, L.A. (1975c). The concept of a linguistic variable and its applications to approximate reasoning, Part 3. *Information Science.* **9**, 43–80.

Zimmermann, H.J. (2001). *Fuzzy set theory and its applications*. Kluwer, Boston.