Università degli Studi di Napoli
Federico II

Fuzzy Regression and PLS Path Modeling: a combined two-stage approach for multi-block analysis

Rosaria Romano

Tesi di Dottorato in
Statistica

XIX Ciclo

Dipartimento
di Matematica e Statistica
Università degli Studi di Napoli "Federico II"
via Cintia, Monte Sant'Angelo – 80126 Napoli

Fuzzy Regression and PLS Path Modeling: a combined two-stage approach for multi-block analysis

Napoli. 30 novembre 2006

Contents

List of figures **V**

List of tables **VII**

1 Introduction: Complexity and Uncertainty **1**

2 Different approaches to Uncertainty **7**

2.1 Imprecision, Vagueness, Uncertainty 7

2.2 Fuzzy Sets 12

2.2.1 Some noteworthy 15

2.2.2 Basic operations 17

2.3 Fuzzy numbers 19

2.4 Interval Data 21

2.5 Fuzzy numbers as a nested family of intervals 25

3 Linear Models for crisp and fuzzy data **29**

3.1 Modeling under uncertainty 29

3.2 Statistical Regression 30

3.3 Fuzzy Regression 34

3.3.1 Fuzzy possibilistic regression 35

3.3.2 Fuzzy Least Squares Regression 54

4 Structural Equation Models based on Fuzzy Regression 59
4.1 Introduction 59
4.2 Structural Equation Models 62
4.2.1 Partial Least Squares Path Modeling 64
4.2.2 Model Specification 65
4.2.3 Model estimation 69
4.2.4 Model Validation 72
4.3 Fuzzy and PLS path modeling: *a marriage of convenience* . . 73
4.4 Models comparison 75
4.4.1 Comparing PLS-PM models 78
4.4.2 Fuzzy clustering for fuzzy PLS-PM 80

5 Application in Customer Satisfaction Analysis 83
5.1 Introduction 83
5.2 Data description 84
5.3 FPR for Customer Satisfaction 85
5.4 Classification of SEM 88

6 Guidelines of future research 93
6.1 Insight of fuzzy possibilistic regression 93
6.2 Extensions of fuzzy approach to PLS-PM 94

Appendix 97
.1 Simplex algorithm 97

List of Figures

2.1 Example of fuzzy variable 14

3.1 Possibility distribution of a fuzzy parameter 36
3.2 Possibility distribution of fuzzy estimated output 37
3.3 Example of fuzzy regression 38
3.4 OLS confidence intervals (black lines) and fuzzy intervals (red lines) under the hypothesis $\beta = 0$ and independent predictors; 51
3.5 OLS confidence intervals (black lines) and fuzzy intervals (red lines) under the hypothesis $\beta = 0$ and dependent predictors; . 51
3.6 OLS confidence intervals (black lines) and fuzzy intervals (red lines) under the hypothesis $\beta = 0$ and independent predictors; 53
3.7 OLS confidence intervals (black lines) and fuzzy intervals (red lines) under the hypothesis $\beta = 0$ and dependent predictors; . 53

4.1 *Path model representation* 65
4.2 *Comparison between simple linear regressions* 78
4.3 *A simple model with a moderating effect* 79
4.4 *Detecting a moderating effect (d) through group comparison* . 79

5.1 **a)** *Expected quality* scree plot; **b)** *Perceived quality* scree plot . 86
5.2 Interval regression for *expected quality* 87
5.3 Interval regression for *perceived quality* 88
5.4 Generic local model 89

5.5 Pyramid 91

List of Tables

3.1 Experimental factors 50
3.2 FPR and FLSR 54

4.1 PLSPM Notation 67
4.2 Data matrix 80

5.1 Data matrix of fuzzy structural parameters 90

Chapter 1

Introduction: Complexity and Uncertainty

In the analysis of real world phenomena the first step consists in establishing a set of relevant variables able to describe the main aspects of interest. As result of this procedure a *system* is distinguished on the analyzed phenomena. In other words, the established variables form a *system*, which is an abstraction of the real phenomena.

The concept of *complexity* is strongly connected to the one of *system*. In fact, the quantity of information required to describe a system measures its degree of complexity.
In general statistical modeling the concept of *complexity* assumes a prominent rule. In this framework, *complexity* has many faces and there is not a unique definition. Specifically, complexity is defined differently as *kolmogorov Complexity* (Cover, Gacs & Gray 1989), *Shannon complexity* (Rissanen 1989) and *Stochastic complexity* (Rissanen 1989).
However, these are very technical definitions.
Complexity definition with a broader perspective is given in (Bozdogan 2004): *complexity of a system (of any type) is a measure of the degree of interdependency between the whole system and a simple enumerative composition of*

its subsystems or parts.

The notion of complexity may be best explained considering real world systems. For instance, a system may be physical, social, biological, economic or political, it is even characterized by a relatively big amount of variables interacting each other. In social science, system complexity increases since humanistic systems have to handle with human thinking and behavior.
In order to analyze such systems it is crucial to simplify them to an acceptable level of complexity.
There exist different strategies for simplifying a system. One way is to exclude some variables from the system. An important strategy of dealing complexity is to allow imprecision in the system description. Another important way for making complex systems manageable is to break them down into appropriate subsystems.
The present thesis focuses on the last two strategies. The first is not taken into account since, from a statistical point of view, all variables selected as relevant for the phenomenon under consideration are sources of information. Hence, it is preferred to use such variables for the analysis.

The strategy to allow imprecision for reducing complexity is appropriately expressed by the Zadeh's *principle of incompatibility* (Zadeh 1973): *as the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics.*
The main idea is that the traditional techniques for analyzing systems are not well suited for dealing with humanistic systems. In human thinking, the key elements are not numbers but classes of objects or concepts in which the membership of each element to the class is gradual (fuzzy) rather than sharp. For instance, the concept of "small number" does not correspond to an exact number. But it is possible to define the class "small numbers". In other words, the property *small* is a vague property, which leads to define the

class of objects having such property. On the other hand there is nor *first small number* nether *last small number*. There is just the class *small numbers* to which all objects belong to some degree. Such class of objects is what is called *fuzzy set*, which is the object of the *Fuzzy Set Theory* (Zadeh 1965). Hence, fuzzy theory with its approach to imprecision and its logic of *degree of membership* allows to deal with complexity inside real world phenomena.

On the other hand, the descriptive complexity of a system can also be reduced by breaking the system into its appropriate subsystem. This is a general principle behind Structural Equation Models (SEM) (Bollen 1989, Kaplan 2000).

The basic idea is that different subset of variables are expression of different *multidimensional* concepts, belonging to the same phenomenon. These concepts are named *latent variables* (LV) as they are not directly observable but measurable by means of a set of *manifest variables* (MV).

The aim of SEM is to study the system of relations between each LV and its MV, and among the different LV inside the system.

Considering one by one each part forming the whole system, and analyzing the relations among the different parts, the system complexity is reduced allowing a better description of the main system characteristics.

Thesis outline

In social science it is very common to analyze phenomena whose description requires the analysis of a complex structure of relations among the variables inside the system. In addition, in such frameworks there is an additional source of complexity arising from the influential human beings involvement. The present work focuses on modeling such complex systems. Specifically, a new strategy based on fuzzy set theory is proposed to analyze them. The strategy consists in introducing fuzzy models inside structural equation models. This allows to face system complexity both introducing an approach tolerant of imprecision and using a methodology well suited to link the different parts in which the system may be decomposed.

In this **first chapter** the concepts of *system* and *complexity* are explained. In particular, a definition of *complex system* and several strategies for handling such complexity are introduced.

The **second chapter** of the thesis provides an overview of the the different approaches to uncertainty. In particular, beside the well known probabilistic uncertainty, *fuzziness* and *imprecision* are introduced as additional sources of uncertainty.
Specifically, *fuzziness* is object of *Fuzzy Set Theory* (FST), whereas *imprecision* is object of *Interval Data Analysis* (IA). Hence, the basic concepts of fuzzy data and interval data are explained.
Strong focus is given to fuzzy approach, since it is well suited for analyzing real world phenomena. However, there is a strong relationship between FST and IA, as it will be shown.

In the **third chapter** the most widely used models for dealing *randomness* and *fuzzyness* are presented. Particularly, fuzzy linear regression models are extensively described.
Among different approaches to fuzzy modeling strong attention is given to the *Possibilistic Fuzzy Regression* (PFR).
The main characteristics of the model are extensively described and results of a detailed study of the relationships between statistical and fuzzy regression are presented. The study proves that FPR may be considered a valuable alternative to traditional regression in systems characterized by fuzzy uncertainty and in situations where statistical regression is not applicable as its strong assumptions are not satisfied.

The *core* of the thesis is in the **fourth chapter**. Here, the idea of introducing fuzzy models inside structural equation models is described.
First, Structural Equation Models (SEM) are extensively described. Then, the fuzzy approach to SEM is widely motivated. In fact, both PFR and SEM are *soft modeling* approaches well suited for analyzing *phenomena* where the

human judgment is influential, i.e. customer satisfaction and sensory analysis.
The proposed strategy has been used to face the crucial statistical problem of the models comparison.
The general approach is based on the comparison between model parameters. But this strategy could lead to unbiased results since information on the fit is not taken into account.
On the other hand, FPR for its own characteristics to embed residual information inside the model, permits to compare models avoiding unbiased results. In such a way, fuzzy SEM estimated for different groups of observations may be compared to each other.
This is a an important task in many application contexts. For instance, in marketing it is very common to apply the same model to different customer segments and successively comparing the results from different segments.

The **chapter fifth** shows an application of the proposed strategy to the *customer satisfaction analysis*.

Chapter 2

Different approaches to Uncertainty

2.1 Imprecision, Vagueness, Uncertainty

The first step of any statistical analysis is the codification of the information. Most relevant variables for describing the phenomenon under investigation are defined and successively measured over preselected statistical units. As well known in statistics, according to the nature of the modalities a variable (o *character*) is classified as *qualitative* or *quantitative*. Specifically, it is *quantitative* if the corresponding modalities are numerical values otherwise it is *quantitative*.
Focusing on the quantitative variables, it is very common to measure such variables in terms of single-values, i. e. their modalities are *precise values*. However, for many reasons *precise measures* are very hard to have.

A relevant source of imprecision can be found in the *data processing* phase, which consists in computing an estimate of a quantity for example the *weight*, based on the results of direct measurements. The outcomes of any *data processing* are never 100% accurate (Kreinovich, Lakejev, Rohn & Kahl 1997). In fact, given the actual value of the measured quantity, this differs from the

measurement result. If there is information on the error of the result of data processing, then it is known that the actual (unknown) value of the measured quantity falls into a specific interval of values.
On the other hand, there are some variables that for their own nature are better described by a pair of ordered value. Examples of these variables are the daily temperatures better registered in terms of minimum and maximum values Another example are the financial data expressed as interval whose endpoints are the opening and the closing daily prices at the stock market, respectively. Of course such information can be summarized by a single value, for instance the average, but this induces a loss of information.
There are also variables not directly measured but whose measure can be obtained as difference between two closely related variables. An example from the Consumer Analysis is the variable *satisfaction* which is measured as difference between the consumer's *expectations* and *perceptions* (Grassia, Lauro & Scepi 2004, Amato & Palumbo 2004).
In all these situations, as in many others, statistical units are better described by *interval values* rather than by *single values*.

It must be noticed that this *imprecision* in the value of measurement refers to lack of knowledge about the value of a parameter expressed as a tolerance interval. However, many times the range enclosing the actual value is well known but the interval has no sharp boundaries. This happens in any *decision making process*, where in addition to the results of measurements and observations, there are *expert estimates* formulated in terms of natural language, i.e. *very heavy*. This source of imprecision is named *vagueness*. The first philosophical papers on the *vagueness* rose only in the 20-th century (Russell 1923, Black 1937). However, the interest in *vagueness* increased when Zadeh founded *Fuzzy Set Theory* (FST) (Zadeh 1965). The basic idea was that the formalizations in traditional mathematical set theory were not satisfactory to handle concepts from the daily language used to classify and quantify information. For example, given the concept *the set of young per-*

sons, it is difficult or meaningless to specify a strong boundary between elements inside and outside the set. At this aim, Zadeh proposed to define sets using graded indicator functions called *membership functions*. These functions measure the magnitude of participation of each element to the *fuzzy set* by means of a scale. This graded approach is coherent with the general principle of the human mind, which naturally uses scales for describing vague concepts such as *very tall*, *quite young*, *too hot*, *etc*. In such a way, fuzzy sets provides a powerful representation of *fuzziness*, that means a meaningful representation of vague concepts expressed in natural language.

It is worth noticing that the word "fuzziness" has been used in a very different context by Sugeno (1977) : in situation where the uncertainty results from information deficiency rather than from the lack of sharp boundaries as in FST. This type of uncertainty named *ambiguity* is typical of relations *one-to many*. In other words, given two sets with distinct boundaries (crisp set) the graded approach is used to define the degree to which the evidence proves the membership of an object in either set. For example, a jury member for a criminal trial has to decide the degree to which the defendant is member of the set *guilty people* or of the set *innocent people*. There is no vagueness neither in the concept *guilty* nor in the concept *innocent*. The uncertainty is associated with several well-defined alternatives. Thus, the *fuzzy measures* assign a value to each crisp set of the universal set, signifying the degree of evidence that a particular element belongs in the set. Whilst, in FST a value is assigned to each element of the universal set signifying the degree of membership in a particular set with unsharp boundaries.

Since 1965, following Zadeh's paradigm (Zadeh 1973, Zadeh, Fu, Tanaka & Shimura 1975, Zadeh 1978) fuzzy set theory has been considerably developed by many researchers with a lot of papers (Gusev & Smirnova 1973, Gupta, N.Saridis & Gaines 1977, Kandel & Byatt 1978) and monographes (Negoita & Ralescu 1975, Dubois & Prade 1980, Klir & Yuan 1995).

There is a substantial ongoing misunderstanding between the probabilistic and the fuzzy approach to the uncertainty (Indahl 1998). In fact, very often the *membership values* are confused with *probabilities* and the *membership functions* with *probability distribution*.

Zadeh himself introducing the notion of a fuzzy set tried to avoid the confusion "*... such a framework provides a natural way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables*".

Uncertainty as well as *vagueness* form two complementary facets of a more general phenomenon called *indeterminacy* (Novák 2005). Specifically, the *uncertainty* phenomenon emerges due to the *lack of knowledge* about the *occurrence* of some event, i.e. during an experiment whose result is not known. A specific form of uncertainty is *randomness* which is uncertainty rising in connection with time. Before doing an experiment there is uncertainty about the result but there is no randomness after the experiment has been realized since the result is then well known. However, uncertainty is not only randomness, for example there is uncertainty in potentiality not referred to time (i.e. lack of knowledge) or with reference to the past (i.e. posterior Bayesian probability).

Stressing the difference between *vagueness* and *uncertainty*, it can be characterized as conflict between *actuality* and *potentiality*. A set is considered *actual* if all its elements are already existing; thought only a part of the set is physically present it is possible to assume they are at disposal as whole. Otherwise the set is *potential*, i.e. the events may occur or not. It follows that *vagueness* steams from the actualized non-sharply delineated set of objects, whereas *uncertainty* faces with still non-actualized grouping of objects. In fact, if there is not certainty about the existence of a specified object it makes no sense to speak about its degree of membership to the set.

The most widely used mathematical model of the uncertainty is the *probability theory* which provides probabilities as numerical measures of the likelihood that a particular event will occur. Probabilities measures have been studied at length, in fact the literature of *probability theory*, including textbooks at various level, is extremely abundant.

On the other hand, the study of the imprecision related to the value of the information is a subject of *Interval Analysis* (IA), a branch of mathematics. *Interval Methods* are of highest value to deal with problems in which the the uncertainty is rigorously bounded. The precursors of *Interval Analysis* include Sunaga 1958, Warmus 1956 and Young 1931. More recently *Interval Arithmetic* is thoroughly covered in two books by Moore (Moore 1966, Moore 1979).

It is derived from the above discussion that uncertainty, involved in any problem-solving situation, is a result of some information deficiency. Information may be incomplete, imprecise, vague, contradictory or deficient in some other way, and each one of these various information deficiencies result in different types of uncertainty.
Here, strong focus is given to three specific information deficiencies: *imprecision, vagueness, uncertainty.* Thus, depending on the nature of the information deficiency, the analysis of the system can be conducted using interval analysis, fuzzy theory, or a probabilistic approach, respectively. In interval analysis, the uncertain parameter is denoted by a simple range. In addition to the range, if a preference function is used to describe the desirability of using different values within the range, fuzzy theory can be used. On the other hand, if the uncertain parameter is described as a random variable following a specified probability distribution, the probabilistic approach can be used.
As in the reality the information occurs with more information deficiencies at the same time, it is convenient a combination of the different approaches. Many works in these direction has been proposed introducing concept as *random set* (Materón 1975, Miranda, Couso & Gil 2005), *interval-valued fuzzy set* (Dubois & Prade 2005) and *fuzzy random set* (Puri & Ralescu 1986, Krätschmer 2001).

2.2 Fuzzy Sets

Fuzzy sets are a generalization of conventional set theory introduced by Georg Cantor (1845 – 1918).

A conventional (*crisp*) set is a collection of objects which can be treated as a whole. Let Ω be a space of objects and ω the general element of Ω. A *crisp* set A, $A \subseteq \Omega$, is defined as a set of elements $\omega \in \Omega$, such that each element ω can either belong or not to the set A. Usually, a set is characterized by a function, called *characteristic function*, that defines which elements of Ω are members of the set an which are not:

$$\mu_A(\omega) = \begin{cases} 0 & \text{for } \omega \notin A \\ 1 & \text{for } \omega \in A \end{cases} \tag{2.1}$$

In other words, the *characteristic function* maps element of Ω to elements of the set $\{0, 1\}$:

$$\mu_A : \Omega \to \{0, 1\} \tag{2.2}$$

Classical set theory is based on the two-valued logic as the *characteristic function* assigns only the two values 0 and 1 to each element in the set. Fuzzy set theory is based on the multi-valued logic, thus the *characteristic function* assigns values within a specified range which indicate the degree of membership of each element in the given set. In this context, the *characteristic function* is called *membership function* and the unit interval $[0, 1]$ is the most common used range of values. Then, the *membership function* maps each element of a given set to a membership grade between 0 and 1:

$$\mu_A : \Omega \to [0, 1] \tag{2.3}$$

Formally, given the *universe of objects* Ω, whit ω as the generic element, a *fuzzy set* $\tilde{A}$ in Ω is defined as a set of ordered pairs:

$$\tilde{A} = \{(\omega, \mu_{\tilde{A}}(\omega))|\omega \in \Omega\}$$

For a generic element $\omega_0 \in \Omega$, the value $\mu_{\tilde{A}}(\omega_0)$ expresses the *membership degree*. The larger the value of $\mu_{\tilde{A}}(\omega)$, the higher the grade of membership of ω in $\tilde{A}$. If the *membership function* is permitted to have only the values 0 and 1 then the *fuzzy* set is reduced to a classical *crisp* set. The universal set Ω may consists of discrete (ordered and non ordered) objects or it can be a continuous space.
Obviously, the membership degrees are subjective measures which depend on individual differences in perceiving abstract concepts. In addition, the same concept may be interpreted differently according to different contexts. In other words, the specification of *membership functions* is subjective and context-dependent.

The universal set Ω is often called *linguistic variable*. This happens when Ω is a continuous space divided in several fuzzy sets representing linguistic concepts whose *membership functions* cover Ω in a more or less uniform way. An example is the set "age" shown in figure 2.1 ranging between the values 0 and 90 and divided in the fuzzy sets *young*, *middle aged* and *old*, respectively. Such a variable is defined *fuzzy variable* and the respective fuzzy sets are the *states* of the fuzzy variable.

A fuzzy set is a codification of the information which allows to represent vague concepts expressed in natural language. Although, fuzzy sets have a greater expressive power than the ordinary sets, they strictly depend on the definition of appropriate *membership functions*, not even easy to define. To face this problem it is possible to introduce such uncertainty in the same definition of *membership function*. In other words, the *membership function* can assign to each object $\omega \in \Omega$ a closed interval of values rather than one single real number. Fuzzy sets defined by this type of *membership functions*

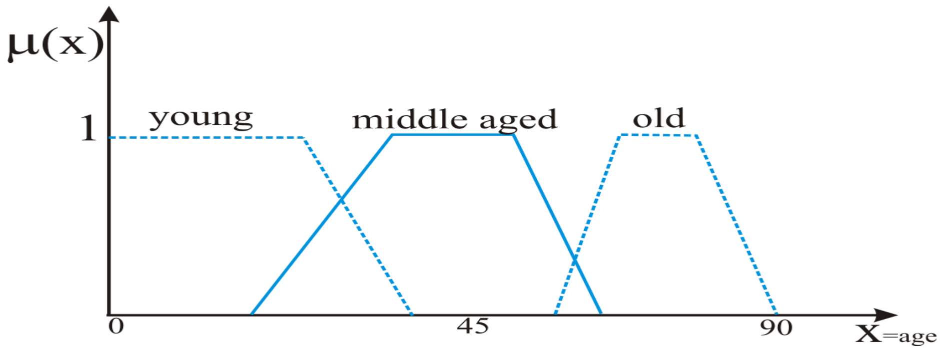

Figure 2.1: Example of fuzzy variable

are called *interval-valued fuzzy sets* and may be described as follows:

$$\mu_{\tilde{A}} : \Omega \rightarrow \varepsilon([0,1]) \tag{2.4}$$

where $\varepsilon([0,1])$ denotes the set of all closed intervals of real numbers in $[0,1]$.
A codification with more expressive power is obtained if the intervals in $\varepsilon([0,1])$ are fuzzy sets defined in $[0,1]$. That means each interval is an ordinary *fuzzy set* defined in $[0,1]$. This type of *fuzzy set*, so-called *fuzzy set of type* 2, is denoted as:

$$\mu_{\tilde{A}} : \Omega \rightarrow F([0,1]) \tag{2.5}$$

where $F([0,1])$ is called *fuzzy power set* of $[0,1]$.
Level 2 *fuzzy sets* are another generalization of the ordinary *fuzzy sets* in which the elements of the universal set are *fuzzy sets*:

$$\mu_{\tilde{A}} : F(\Omega) \rightarrow [0,1] \tag{2.6}$$

Of course a combination of *fuzzy sets of type* 2 and *level* 2 *fuzzy sets* it is also possible arising to generalized *fuzzy sets* whose *membership function* has the form:

$$\mu_{\tilde{A}} : F(\Omega) \rightarrow F([0,1]) \tag{2.7}$$

Finally, a *probabilistic set* $\tilde{A}$ is defined by a randomized *membership function* $\mu_{\tilde{A}}$ as follows:

$$\mu_{\tilde{A}} : \Omega \times \Theta \rightarrow [0,1] \tag{2.8}$$

where the *membership function* $\mu_{\tilde{A}}$ of ω in $\tilde{A}$ is a random variable built from the distribution p of ω assumed independent of $\tilde{A}$ (Hirota 1977).
Although such generalizations of *ordinary fuzzy sets* permit to embed the uncertainty in identifying proper *membership functions*, their processing is computationally more demanding.

It must be noticed that in such logic of overlapped subsets, two fundamental laws of classical set theory are broken:

- *Law of Contradiction*: a set and its complement must comprise the universal set

  $$A \cup \bar{A} = \Omega$$

- *Law of Excluded Middle*: an element can either be in its set or in its complement but never simultaneously in both

  $$A \cap \bar{A} = \emptyset$$

2.2.1 Some noteworthy

Fuzzy sets can be characterized by a family of *crisp sets*, called $\alpha-$cut. Given a *fuzzy set* $\tilde{A}$ defined on Ω and any number $\alpha \in [0,1]$ the $\alpha-$cut, ${}^{\alpha}\tilde{A}$, is defined as:

$$^{\alpha}\tilde{A} = \{\omega \in \Omega : \mu_{\tilde{A}}(\omega) \geq \alpha\} \tag{2.9}$$

In other words, the $\alpha-$cut of $\tilde{A}$ is a *crisp set* including all $\omega \in \Omega$ which are members of the *fuzzy set* $\tilde{A}$ with a membership degree greater or equal to a specified value of α. The $\alpha-$cut with the strict inequality $\mu_{\tilde{A}}(\omega) > \alpha$ is named *strong* $\alpha-$cut and defined as follows:

$$^{\alpha+}\tilde{A} = \{\omega \in \Omega : \mu_{\tilde{A}}(\omega) > \alpha\} \tag{2.10}$$

Every element in the universal set Ω is a member of the fuzzy set $\tilde{A}$ to some grade, maybe even zero. The set of all objects $\omega \in \Omega$ at which $\mu_{\tilde{A}}(\omega) > 0$ is defined *support* of the fuzzy set. Whereas, the *core* is the set of all objects $\omega \in \Omega$ at which $\mu_{\tilde{A}}(\omega) = 1$. In particular, if the *core* of a fuzzy set is nonempty then the fuzzy set is *normal*, that means it is always possible to find at least a point $\omega \in \Omega$ such that $\mu_{\tilde{A}}(\omega) = 1$. Differently, the *normality* of a fuzzy set can be derived considering the notion of *height* of a *fuzzy set*. Let $h(\tilde{A})$ be the *height* of a *fuzzy set*, it corresponds to the largest membership degree obtained by any $\omega \in \Omega$:

$$h(\tilde{A}) = \sup_{\omega \in \Omega} \tilde{A}(\omega) \tag{2.11}$$

A fuzzy set $\tilde{A}$ is defined *normal* iff

$$h(\tilde{A}) = 1 \tag{2.12}$$

otherwise if $h(\tilde{A}) < 1$ it is defined *subnormal fuzzy set*.

The *cardinality* of a *fuzzy set*, so-called Σcount (*sigma-count*), is a different concept defined as:

$$card(\tilde{A}) = |\tilde{A}| = \sum_{i=1}^{I} \mu_{\tilde{A}}(\omega_i) \tag{2.13}$$

For *fuzzy sets* defined on $\mathbb{R}$, a very important property is the *convexity*. A *fuzzy set* $\tilde{A}$ is convex if and only if for any $\omega_1, \omega_2 \in \Omega$ and any $\lambda \in [0, 1]$:

$$\mu_{\tilde{A}}(\lambda\omega_1 + (1-\lambda)\omega_2) \geq min\{\mu_{\tilde{A}}(\omega_1), \mu_{\tilde{A}}(\omega_2)\} \tag{2.14}$$

Alternatively, a *fuzzy set* is *convex* iff all its $\alpha-$cuts are convex in the classical set.

It is common in literature to denote a *fuzzy set* by means of the pairs elements-*membership function*:

$$\tilde{A} = \frac{\mu_{\tilde{A}}(\omega_1)}{\omega_1} + \frac{\mu_{\tilde{A}}(\omega_2)}{\omega_2} + \ldots + \frac{\mu_{\tilde{A}}(\omega_I)}{\omega_I} \tag{2.15}$$

where the plus sign is meant in the set-theoretic sense, i.e. a group of objects forming a set.
On the other hand, a *fuzzy set* defined in an interval of real numbers may be denoted as:

$$\tilde{A} = \int_{\omega} \tilde{A}(\omega)/\omega \tag{2.16}$$

2.2.2 Basic operations

In Zadeh's seminal paper on fuzzy sets (Zadeh 1965) only the standard operations of complement, union and intersection were introduced. They are here shortly described with some related examples.

The *complement (negation)* of a *fuzzy set* $\tilde{A}$ is the *fuzzy set* $\neg\tilde{A}$ whose *mem-*

bership function is given by

$$\mu_{\neg\tilde{A}}(\omega) = 1 - \mu_{\tilde{A}}(\omega) \tag{2.17}$$

The *union (disjunction)* of two *fuzzy sets* $\tilde{A}$ and $\tilde{B}$ is the *fuzzy set* $\tilde{C} = \tilde{A} \cup \tilde{B}$ whose *membership function* is given by

$$\mu_{\tilde{C}}(\omega) = max(\mu_{\tilde{A}}(\omega), \mu_{\tilde{B}}(\omega)) \tag{2.18}$$

i.e. the smallest *fuzzy set* containing both $\tilde{A}$ and $\tilde{B}$.
The *intersection (conjunction)* of two *fuzzy sets* $\tilde{A}$ and $\tilde{B}$ is the *fuzzy set* $\tilde{C} = \tilde{A} \cap \tilde{B}$ with the *membership function*

$$\mu_{\tilde{C}}(\omega) = min(\mu_{\tilde{A}}(\omega), \mu_{\tilde{B}}(\omega)) \tag{2.19}$$

i.e. the largest *fuzzy set* contained in both $\tilde{A}$ and $\tilde{B}$.
Given $\Omega = \{1, 2, \ldots, 10\}$, let $\tilde{A}$ ="small numbers"$= 1/1 + 1/2 + 0.8/3 + 0.5/4 + 0.3/5 + 0.1/6$ and $\tilde{B}$ ="large numbers"$= 0.1/5 + 0.2/6 + 0.5/7 + 0.8/8 + 1/9 + 1/10$ be two fuzzy sets in the form (2.15) then:

$$\begin{aligned}
\neg\tilde{A} &= 0.2/3 + 0.5/4 + 0.7/5 + 0.9/6 + 1/7 + 1/8 + 1/9 + 1/10 \\
\tilde{A} \cup \tilde{B} &= 1/1+1/2+0.8/3+0.5/4+0.3/5+0.2/6+0.5/7+0.8/8+1/9+1/10 \\
\tilde{A} \cap \tilde{B} &= 0.1/5 + 0.1/6
\end{aligned}$$

Since Zadeh's paper, most of the operations inside the classical set theory have been extended to fuzzy sets (Zimmermann 1996, Dubois & Prade 1980).

It often occurs that a *crisp* function is extended to act on *fuzzy sets*. Such a function represents what is called *fuzzified* function. A fundamental principle for fuzzifyng crisp functions is the so-called *extension principle*. Let $f : \Omega \rightarrow \Gamma$ and $\tilde{A}$ be a function (operation) and a *fuzzy set* in Ω, respectively. Then $\tilde{A}$ induces via f a fuzzy set $\tilde{B}$ in Γ with γ as the generic element, as follows:

$$\mu_{\tilde{B}}(\gamma) = \begin{cases} sup_{\gamma=f(\omega)}\mu_{\tilde{A}}(\omega) & \text{if } f^{-1}(\omega) = \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{2.20}$$

For example, let $\Omega = \{1, 2, 3, 4\}$ and $\Gamma = \{1, 2, \ldots, 6\}$. If $\tilde{A} = 0.1/1 + 0.2/2 + 0.7/3 + 1/4$, then $\tilde{B} = 0.1/3 + 0.2/4 + 0.7/5 + 1/6$.

2.3 Fuzzy numbers

A fuzzy set in the real line that satisfies both the conditions of *normality* (2.12) and *convexity* (2.14) is a *fuzzy number*.
It must be normal so that the statement "real number close to r" is fully satisfied by r itself, i.e. $\mu_{\tilde{A}}(r) = 1$. In addition, all its $\alpha-$cuts for $\alpha \neq 0$ must be closed intervals so that the arithmetic operations on *fuzzy sets* can be defined in terms of operations on closed intervals. On the other hand, if all its $\alpha-$cuts are closed intervals it follows that the *fuzzy number* is a convex *fuzzy set*.

A method for developing fuzzy arithmetic it to extend operations on real numbers to *operations on fuzzy sets* by means of the *extension principle*. Unfortunately, this approach is numerically inefficient (Fedrizzi & Kacprzyk 1992). Hence, in literature it is very common to use *fuzzy numbers* in the so-called *$L - R$ representation* (Dubois & Prade 1980):

$$\mu_{\tilde{A}}(\omega) = \begin{cases} L\left(\frac{c-\omega}{l}\right) & \alpha > 0, \forall\omega \leq c \\ R\left(\frac{\omega-c}{r}\right) & \beta > 0, \forall\omega \geq c \end{cases} \tag{2.21}$$

where $L(\omega) = L(-\omega)$; $L(0) = 1$; L is increasing in $[0, +\infty]$; and similarly function R. The value c is the mean value of $\tilde{A}$, whereas l and r are called *left* and *right spreads*, respectively. Notice that if the spreads are zero then $\tilde{A}$ is a nonfuzzy number, whilst $\tilde{A}$ becomes fuzzier as well as the spreads increase. A *fuzzy number* represented by the $L-R$ *representation* is symbolically denoted as $\tilde{A} = (c, l, r)_{LR}$. A fuzzy number $\tilde{A}$ with $l = r = w$ is called *symmetrical fuzzy number* and denoted as:

$$\tilde{A} = (c, w)_L \tag{2.22}$$

Arithmetic operations on $L - R$ fuzzy numbers may be defined in terms of the c, l, r values. For instance, the *addition* between two *fuzzy numbers* $\tilde{A} = (c_{\tilde{A}}, l_{\tilde{A}}, r_{\tilde{A}})$ and $\tilde{B} = (c_{\tilde{B}}, l_{\tilde{B}}, r_{\tilde{B}})$ is given by:

$$\begin{aligned} \tilde{A} + \tilde{B} &= (c_{\tilde{A}}, l_{\tilde{A}}, r_{\tilde{A}}) + (c_{\tilde{B}}, l_{\tilde{B}}, r_{\tilde{B}}) \\ &= (c_{\tilde{A}} + c_{\tilde{B}}, l_{\tilde{A}} + l_{\tilde{B}}, r_{\tilde{A}} + r_{\tilde{B}}) \end{aligned}$$

and similarly for the other operations.
In the *possibility theory* (Zadeh 1978), a branch of fuzzy set theory, fuzzy numbers are described by *possibility distributions*.
A *possibility distribution* $\pi_{\tilde{A}}(\omega)$ is a function satisfying the following conditions (Tanaka & Guo 1999):

- there exists an ω such that $\pi_{\tilde{A}}(\omega) = 1$ (normality (2.12))
- $\alpha-$cuts of fuzzy numbers are convex (2.14)
- $\pi_{\tilde{A}}(\omega)$ is piecewise continuous

Most widely used *possibility distributions* of fuzzy numbers are the *intervals*, and the *triangular fuzzy numbers*.
The *possibility distributions* of an interval denoted as $\tilde{A}_i = (c_i, w_i)_I$ is:

$$\pi_{\tilde{A}_i}(\omega) = \begin{cases} 1 & \{\omega | c_i - w_i \le \omega \le c_i + w_i\} \\ 0 & \text{otherwise} \end{cases} \tag{2.23}$$

where I stands for *interval*.
The *possibility distribution* of a triangular fuzzy number denoted as $\tilde{A}_i =$

$(c_i, l_i, r_i)_T$ is:

$$\pi_{\tilde{A}_i}(\omega) = \begin{cases} 0 & \omega \leq c_i - l_i \\ 1 - \left(\frac{c_i - \omega}{l_i}\right) & c_i - l_i \leq \omega \leq c_i \\ 1 - \left(\frac{\omega - c_i}{r_i}\right) & c_i \leq \omega \leq c_i + r_i \\ 0 & \omega \geq c_i + r_i \end{cases} \tag{2.24}$$

where T stands for *triangular*.
Particular fuzzy numbers are the *symmetrical fuzzy numbers* whose *possibility distribution* may be denoted as:

$$\pi_{\tilde{A}_i}(\omega) = max\left(0, 1 - \left|\frac{\omega - c_i}{r_i}\right|^q\right) \tag{2.25}$$

Specifically, (2.25) corresponds to *triangular* fuzzy numbers when $q = 0$, to *square root* fuzzy numbers when $q = 1/2$ and *parabolic* fuzzy numbers when $q = 2$. Considering (2.4) it is easy to show that (2.25) corresponds to *intervals* when $q = 0$.
It is worth noticing that fuzzy variables are associated to possibility distributions in the similar way that random variable are associated with probability distributions. Furthermore, *possibility distributions* are numerically equal to membership functions (Zadeh 1978).

2.4 Interval Data

A rigorous study of *interval data* is given by *Interval Analysis* (Alefeld & Herzerberger 1983). In this framework, an *interval value* is a bounded subset of real numbers $\mathbf{x} = [\underline{x}, \overline{x}]$, formally:

$$\mathbf{x} = \{x \in \mathbb{R} | \quad \underline{x} \leq x \leq \overline{x}\} \tag{2.26}$$

where $\underline{x}$ and $\overline{x}$ are called *lower* and *upper bound*, respectively. Alternatively, an *interval value* may by expressed in terms of *width* (or *radius*), x_w, and *center* (or *midpoint*), x_c:

$$\begin{aligned} x_w &= |\overline{x} - \underline{x}| \\ x_c &= \frac{1}{2}|\underline{x} + \overline{x}| \end{aligned} \tag{2.27}$$

If an interval has zero radius, i.e. $\underline{x} = \overline{x}$, it is a degenerate interval called a *point* or *thin interval*, containing a single point represented by:

$$\mathbf{x} \equiv [x, x] \tag{2.28}$$

Given $\mathbf{x} = [\underline{x}, \overline{x}]$ and $\mathbf{y} = [\underline{y}, \overline{y}]$, algebraic operations on intervals are defined in such a way that their results are again closed intervals embedding all possible real results. Formally, the result of $\mathbf{x} \diamond \mathbf{y}$ is again an interval $\mathbf{z}$ with property:

$$\mathbf{x} \diamond \mathbf{y} = \mathbf{z} = \{z = x \diamond y | x \in \mathbf{x}, y \in \mathbf{y}\} \tag{2.29}$$

where $\diamond$ belongs to the set $\{+, -, \times, \div\}$.
Arithmetic operations on intervals are expressed in terms of ordinary arithmetics on their bounds as follows:

$$\mathbf{x} + \mathbf{y} = [\underline{x} + \underline{y}, \overline{x} + \overline{y}] \tag{2.30}$$

$$\mathbf{x} - \mathbf{y} = [\underline{x} - \overline{y}, \overline{x} - \underline{y}] \tag{2.31}$$

$$\begin{aligned} \mathbf{x} \times \mathbf{y} = [&min([\underline{x} \times \underline{y}, \underline{x} \times \overline{y}, \overline{x} \times \underline{y}, \overline{x} \times \overline{y}), \\ &max(\underline{x} \times \underline{y}, \underline{x} \times \overline{y}, \overline{x} \times \underline{y}, \overline{x} \times \overline{y})] \end{aligned} \tag{2.32}$$

$$\begin{aligned} \mathbf{x} \div \mathbf{y} = [&min([\underline{x} \div \underline{y}, \underline{x} \div \overline{y}, \overline{x} \div \underline{y}, \overline{x} \div \overline{y}), \\ &max(\underline{x} \div \underline{y}, \underline{x} \div \overline{y}, \overline{x} \div \underline{y}, \overline{x} \div \overline{y})] \end{aligned} \tag{2.33}$$

with the extra condition $0 \notin \mathbf{y}$ for the division.

Examples of interval-valued arithmetic operations follow:

$$
\begin{aligned}
[2,5]+[1,3] &= [3,8] \\
[2,5]-[1,3] &= [-1,4] \\
[3,4]\times[2,2] &= [6,8] \\
[4,10]\div[1,2] &= [2,10]
\end{aligned}
$$

Arithmetic operations on intervals satisfy some useful properties. Given the intervals $\mathbf{a}=[\underline{a},\overline{a}]$, $\mathbf{b}=[\underline{b},\overline{b}]$, $\mathbf{c}=[\underline{c},\overline{c}]$, $\mathbf{0}=[0,0]$ and $\mathbf{1}=[1,1]$, the most important properties follow:

1. *commutativity*
   $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$
   $\mathbf{a}\times\mathbf{b}=\mathbf{b}\times\mathbf{a}$

2. *associativity*
   $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$
   $(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$

3. *identity*
   $\mathbf{a}=\mathbf{0}+\mathbf{a}=\mathbf{a}+\mathbf{0}$
   $\mathbf{a}=\mathbf{0}+\mathbf{a}=\mathbf{a}+\mathbf{0}$

Geometrically, interval is just a section of a real line, determined by its own bounds. The set of all intervals is commonly denoted by $\mathbb{IR}$.
An interval vector $\mathbf{X}$ is defined to be a vector with interval components:

$$\mathbf{x_n}=[\underline{x}_n,\overline{x}_n] \tag{2.34}$$

where $(n=1,\ldots,N)$ and the space of all N dimensional interval vectors is denoted by $\mathbb{IR}^N$:

Similarly an interval matrix $[\mathbf{X}]$ is a matrix with interval components:

$$\mathbf{x_{np}} = [\underline{x}_{np}, \overline{x}_{np}] \tag{2.35}$$

where $(p = 1, \ldots, P)$ and the space of all $N \times P$ matrices is denoted by $\mathbb{IR}^{N \times P}$.
Arithmetic operations on interval vectors and matrices are carried out according to the operations on $\mathbb{IR}$ in the same way that real vector and matrices operations are carried out according to real operations.
It is possible to determine distances between intervals using the Hausdorff metric. Let $\mathbf{a} = [a_c, a_w]$ and $\mathbf{b} = [b_c, b_w]$ be two generic unidimensional intervals, the simplest case of Hausdorff metric $d(\mathbf{a}; \mathbf{b})$ in $\mathbb{R}$ is shown below:

$$d(\mathbf{a}; \mathbf{b}) = |\mathbf{a}_c - \mathbf{b}_c| + |\mathbf{a}_w - \mathbf{b}_w| \tag{2.36}$$

It has been shown that 2.36 is a distance since it satisfy the following properties (Neumaier 1990):

i) $d(\mathbf{a}; \mathbf{b}) \geq 0$

ii) $d(\mathbf{a}; \mathbf{b}) = d(\mathbf{b}; \mathbf{a})$

iii) $d(\mathbf{a}; \mathbf{c}) \leq d(\mathbf{a}; \mathbf{b}) + d(\mathbf{b}; \mathbf{c})$

where C is a generic interval in $\mathbb{R}$.
The properties of the Hausdorff metric in $\mathbb{R}^P$ are widely discussed in (Braun, Mayberry, Powers & Schlicker 2003). It is worth noticing that intervals are represented as segments in $\mathbb{R}$, parallelograms in $\mathbb{R}^2$ and boxes in higher dimensional spaces. As a consequence, the generalization of the Hausdorff metric to $\mathbb{IR}^P$ is quite complex, and the complexity increases as well as P tends to be large. However, it is possible an easy generalization when the compact subsets in $\mathbb{IR}^P$ are restricted to some special cases (Palumbo & Irpino 2005).

It must be noticed that interval data treatment involves *NP*-hard problem solutions. A lot of papers over the years have faced the problem to find more feasible solutions (Kreinovich et al. 1997, Ferson, Ginzburg, Kreinovoch, Longpré & Aviles 2002). However, such problem not inevitably occurs if the solutions derive from square symmetric matrices. Luckily, this condition is satisfied by a lot of statistical methods, i.e. least squares regression and factorial analysis.

2.5 Fuzzy numbers as a nested family of intervals

The bridge between *fuzzy set theory* and *interval analysis* is *fuzzy arithmetic* since *fuzzy arithmetic* is *interval arithmetic* on $\alpha-$cuts (Kaufmann & Gupta 1985). In fact, each *fuzzy number* can uniquely be represented as a nested family of intervals $[\Omega(\alpha)]$, $\alpha-$cuts of $[\Omega]$, corresponding to different value of α (Nguyen, Wang & Kreinovich 2003). In other words, A fuzzy number can be represented as a nested collection of $\alpha-$cuts, i.e. intervals corresponding to different thresholds of *membership value*.
There exists also a property stating that a *fuzzy set* can be represented in terms of special fuzzy sets ${}_{\alpha}\tilde{A}$ which are defined in terms of its $\alpha-$cuts. This property is usually referred to as *decomposition theorems of fuzzy sets*, which is formulated in different version.
Here, it is empirically discussed one of the basic *decomposition theorems of fuzzy sets*. Given the fuzzy set $\tilde{A} = .2/\omega_1 + .4/\omega_2 + .6/\omega_3 + .8/\omega_4 + .1/\omega_5$, it is

associated with five $\alpha-$cuts defined by the following characteristic functions:

$$\begin{aligned}
{}^{.02}\tilde{A} &= .1/\omega_1 + .1/\omega_2 + .1/\omega_3 + .1/\omega_4 + .1/\omega_5 \\
{}^{.04}\tilde{A} &= 0/\omega_1 + .1/\omega_2 + .1/\omega_3 + .1/\omega_4 + .1/\omega_5 \\
{}^{.06}\tilde{A} &= 0/\omega_1 + 0/\omega_2 + .1/\omega_3 + .1/\omega_4 + .1/\omega_5 \\
{}^{.08}\tilde{A} &= 0/\omega_1 + 0/\omega_2 + 0/\omega_3 + .1/\omega_4 + .1/\omega_5 \\
{}^{.01}\tilde{A} &= 0/\omega_1 + 0/\omega_2 + 0/\omega_3 + 0/\omega_4 + .1/\omega_5
\end{aligned}$$

Then each of the $\alpha-$cut is concerted to a special fuzzy set ${}_{\alpha}\tilde{A}$:

$$ {}_{\alpha}\tilde{A}(\omega) = \alpha_{\alpha}\tilde{A}(\omega) \tag{2.37}$$

obtaining

$$\begin{aligned}
{}_{.02}\tilde{A} &= .2/\omega_1 + .2/\omega_2 + .2/\omega_3 + .2/\omega_4 + .2/\omega_5 \\
{}_{.04}\tilde{A} &= 0/\omega_1 + .4/\omega_2 + .4/\omega_3 + .4/\omega_4 + .4/\omega_5 \\
{}_{.06}\tilde{A} &= 0/\omega_1 + 0/\omega_2 + .6/\omega_3 + .6/\omega_4 + .6/\omega_5 \\
{}_{.08}\tilde{A} &= 0/\omega_1 + 0/\omega_2 + 0/\omega_3 + 0/\omega_4 + .8/\omega_5 \\
{}_{.01}\tilde{A} &= 0/\omega_1 + 0/\omega_2 + 0/\omega_3 + 0/\omega_4 + 1/\omega_5
\end{aligned}$$

Now, it is easy to shown that the standard fuzzy union of these five special *fuzzy sets* is exactly the original $\tilde{A}$:

$$\tilde{A} = {}_{.02}\tilde{A} \cup {}_{.04}\tilde{A} \cup {}_{.06}\tilde{A} \cup {}_{.08}\tilde{A} \cup {}_{1}\tilde{A} \tag{2.38}$$

For the theoretically proof of this basic *decomposition theorems of fuzzy sets*, as well as other versions of the theorem see (Klir & Yuan 1995).
The property of a *fuzzy set* to be represented by its $\alpha-$cuts is extremely useful from a computational point of you. In fact, this allows to use all the interval arithmetic operations discussed in (2.4), instead of the fundamental

extension principle 2.20 which is definitely computationally more demanding.

Chapter 3

Linear Models for crisp and fuzzy data

3.1 Modeling under uncertainty

Modeling real world is a fundamental task in Statistics. Models are built for describing, understanding, estimating, reproducing and inspecting real phenomena (Piccolo 1998). A model is an exemplification of reality. The basic aim is to explain the complexity inside a system studying the relationships between variables observed over statistical units. First the data are observed, then hypothesis are formulated and a cause effect relation between variables is assumed. The specification of the functional relation between variables is crucial during model formalization. It is based both on a-priori knowledge and empirical results.
Since models are idealizations of reality, statistical relations cannot be deterministic. They would be unrealistic. Hence, deterministic models are usually extended to stochastic ones, by introducing a *measurement error*. This is in line with the *statistical paradigm* which states that real world phenomena consist of two well defined components: one possible to explain (deterministic component) and another one inexplicable (stochastic component). Thus,

statistical models are based upon probability theory in which imprecision is synonymous with *randomness* and inferential procedures are considered to face the uncertainty of the estimates.
For a long time, probability theory has been the only available tool for representing uncertainty.
In 1965 Zadeh introduced the *fuzzy paradigm* for dealing *vagueness* inside real world phenomena (see chapter 2). In 1970 Bellman and Zadeh stated that *imprecision is not only randomness*, and that *fuzziness is a major source of imprecision in many decision problem*. Since then, the field of uncertainty modeling is dramatically changed.
Besides traditional model, new models to face different sources of uncertainty have been proposed. Particularly, *fuzzy models* for analyzing phenomena in which the major source of uncertainty is *fuzziness* rather than *randomness*.

3.2 Statistical Regression

Specialized literature on regression analysis (Gujarati 2003) and more generally on linear and non linear models (Ryan 1997) offers many solutions to study the dependence between two sets of variables.
Let Y and $\{X_1, X_2, \ldots, X_P\}$ be a quantitative dependent variable and a set of P independent variables observed on N statistical units, respectively.
Regression Analysis studies the statistical dependence of Y with respect to the predictors $\{X_p\}$ $(p = 1, \ldots, P)$.
This requires the choice of a suitable model and the related parameters estimation. Given the generic model:

$$Y = f(X_1, \ldots, X_P; \beta) + \varepsilon$$

the aim of statistical regression is to find the set of unknown parameters so that $\hat{Y} = f(X_1, \ldots, X_P; \hat{\beta})$ is a good prediction of Y. The term ε indicates the deviation of Y from the model.

The most widely used regression model is the *(Multiple) Linear Regression Model* (MLRM), as well as the *Least Squares* (LS) is the most widespread estimation procedure.
In MLRM the dependent variable Y would be expressed as the weighted sum of the independent variables $\{X_1, X_2, \ldots, X_P\}$, with the unknown weights $\{\beta_1, \beta_2, \ldots, \beta_P\}$.
Formally:

$$y_n = \beta_0 + x_{n1}\beta_1 + \ldots + x_{nP}\beta_P + \varepsilon_n \tag{3.1}$$

where $(n = 1, \ldots, N)$ and β_0 is the unknown parameter related to intercept term.
In matrix form the model is expressed as:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{3.2}$$

where $\mathbf{y} = [y_1, \ldots, y_N]'$, $\beta = [\beta_0, \ldots, \beta_P]'$, $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_N]'$ and

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1P} \\ 1 & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \ldots & x_{NP} \end{bmatrix}$$

LS is based on the minimization of the sum of squared deviations:

$$\min_{\beta} = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} + \mathbf{X}\beta) \tag{3.3}$$

The optimal solution of the minimization problem (3.3) is the following vector:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3.4}$$

Substituting (3.2) in (3.4), this may be rewritten as:

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \tag{3.5}$$

Such solutions is totally dependent on the given data. Generalizing results to the whole population the following LS estimator is obtained:

$$B = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3.6}$$

Ordinary Least Squares (OLS) regression assumes that observations are generated under the following assumptions:

1. $E(\varepsilon) = 0$
   where $E(\varepsilon)$ is an expectation vector and 0 is the $N \times 1$ null vector

2. $V(\varepsilon) = \sigma^2\mathbf{I}$
   where $V(\varepsilon)$ is a covariance matrix, $\mathbf{I}$ is the $N \times N$ identity matrix and σ^2 is a variance

3. $\mathbf{X}$ is non stochastic
   that means it consists of fixed numbers

4. $\rho(\mathbf{X}) = P + 1$
   where $\rho(\mathbf{X})$ is the rank of $\mathbf{X}$,with 1 representing the intercept term and P is lower than the number of observations. This constraint guarantees no exact linear relationship among the variables, i.e. no multicollinearity.

Under the OLS assumptions the LS estimates are BLUE (*Best Linear Unbiased Estimator*), as stated by the famous Gauss-Markov theorem. Specifically, B are linear unbiased estimators since $E(B) = \beta$. This is easily shown, considering (3.4):

$$E(B) = E(\beta) + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\varepsilon) = \beta \tag{3.7}$$

Furthermore, let $V(B)$ be the variance of B:

$$V(B) = E((B-\beta)(B-\beta)') = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{3.8}$$

Thus, B is the *best* estimator because of minimum variance. In other words, $V(B) \le V(B^*)$, where B^* is any other linear unbiased estimator.
Since the term error ε is a random vector, thus B is a random vector too.
When the context is not exploratory it is necessary to take into account the uncertainty of the estimates. The B estimator properties and the sample distribution can be specified only assuming a specific probability distribution of the disturbances ε_n.
When $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ then:

$$B \sim N(\beta, s^2(\mathbf{X}'\mathbf{X})^{-1}) \tag{3.9}$$

$$\mathbf{y} \sim N(\mathbf{X}\beta, s^2\mathbf{I}) \tag{3.10}$$

where s^2 is the unbiased estimator of σ^2, which is unknown.
Under the normality assumptions OLS estimators are Best Unbiased Estimators (BUE). In fact, Rao has proved they have the uniformly minimum variance in the class of all unbiased estimators. Moreover, their sampling distribution is fully specified allowing to define confidence intervals for the regression parameters.
The OLS model is easy to implement, however its assumptions are too restrictive with respect to many real world phenomena. In particular, there are some fields where these assumptions are violated almost surely. Their violation affects the estimates because causes biased and inefficient estimators (Gujarati 2003).
Of particular interest is the case of "quasi" multi-collinearity: that is many explanatory variables are highly correlated. This does not violate any OLS assumption, however it has a dramatic impact on the variance of B. Although efficient and unbiased, the OLS estimators have large variance, mak-

ing estimation unuseful from a practical point of view. The effects of the quasi multi-collinearity are more evident when the sample size is small. C.H. Achen (1982) presenting the quasi multicollinarity effects stated that in the case of small samples "*No statistical answer can be given*". In fact, the generally proposed solution consists in removing correlated exploratory variables. This solution is unsatisfying in many applications fields where the user would keep all variables in the model.
Specialized literature provides viable alternatives to OLS estimators in case of quasi multi-collinearity. The most famouse are the Ridge Regression estimators (Hoerl & Kennard 1970), the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 1996) and Partial Least Squares Regression (PLSR) (Wold, Ruhe, Wold & Dunn 1984).

3.3 Fuzzy Regression

The generalization of linear regression to fuzzy data is a problem that involved and still interests many researchers from different scientific area. Fuzzy Regression (FR) is a powerful tool for analyzing systems characterized by fuzzy relations among the variables or by fuzzy variables themselves. There exist two different approaches to fuzzy regression.
The first one considers the *vagueness* of the relation between the dependent and the independent variables, which can be fuzzy or crisp. Thus the aim is to minimize such *fuzziness* and the solution to this optimization problem is obtained through an extensive use of the *mathematical programming*. This approach is called *Fuzzy Possibilistic Regression* (FPR) since it is based on possibility concepts (see chapter 2).
The second approach is more close to the traditional statistical approach. In fact, following the Least Squares line of thought, the aim is to minimize the distance between the observed and the estimated fuzzy data. This approach is referred as *Fuzzy Least Squares Regression* (FLSR).

A very exhaustive bibliography of the different contributions in both the research lines is given in (Taheri 2003), whereas a thorough review can be found in (Shapiro 2004).

3.3.1 Fuzzy possibilistic regression

In the early 80's, Tanaka proposed the first fuzzy linear regression model, moving from fuzzy sets theory and possibility theory (Tanaka, Uejima & Asai 1982).
According to Zadeh's ideas, he considered that real world phenomena were mostly characterized by fuzzy uncertainty. Differently from the *statistical paradigm* where uncertainty modeling is considered an additive element to the deterministic relation among the variables he considered fuzziness as reflected inside the model via fuzzy parameters. Thus, he formulated a regression model where the functional relation between dependent and independent variables is represented as fuzzy linear function whose parameters are given by fuzzy numbers.
As discussed in chapter 2, fuzzy numbers are characterized by a possibility distribution. Parameters in fuzzy regression are associated with a possibility distribution in the same way that parameters in statistical regression are associated with a probability distribution (3.9).
Tanaka's idea was that possibilistic methods could derive new estimators (fuzzy estimators) by dealing directly with models formulated in a possibility context. The extensively use of the possibility concepts led to define Tanaka's model as possibilistic fuzzy regression.
Tanaka proposed the first possibilistic linear regression using the following fuzzy linear model with crisp input and fuzzy parameters:

$$\tilde{y}_n = \tilde{\beta}_0 + \tilde{\beta}_1 x_{n1} + \ldots + \tilde{\beta}_p x_{np}, + \ldots + \tilde{\beta}_P x_{nP} \tag{3.11}$$

where the parameters are symmetric triangular fuzzy numbers denoted by $\tilde{\beta}_p = (c_p; w_p)_L$ with c_p and w_p as center and the spread.
Furthermore, they are assumed to be non-interactive fuzzy numbers, which means that the respective membership functions are determined independently. As showed in chapter (2), the possibility distribution of parameters as symmetric fuzzy numbers is given by:

$$\pi_{\tilde{\beta}_p}(b_p) = \begin{cases} 1 - \frac{|b_p - c_p|}{w_p} & c_p - w_p \leq b_p \leq c_p + w_p \\ 0 & otherwise \end{cases} \tag{3.12}$$

Using the LR representation proposed by Duboise and Prade (2.21), parameters possibility distribution may be written as:

$$L(z) = max(0, 1 - |z|) \tag{3.13}$$

A representation of such possibility distribution is given in figure (3.1), where α is the *degree of possibility* of b_p to the fuzzy set $\tilde{\beta}_p$.

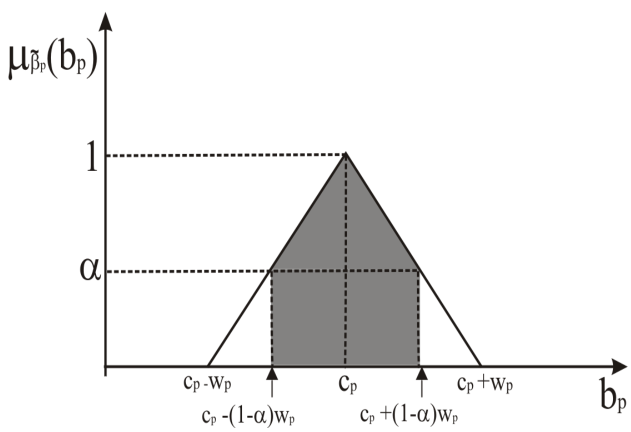

Figure 3.1: Possibility distribution of a fuzzy parameter

It is worth noticing that even if the input y_n is crisp, the theoretical value $\tilde{y}_n^*$

is fuzzy, as natural consequence of the fuzzy model (3.11). The possibility distribution of $\tilde{y}_n^*$ is obtained as:

$$\pi_{\tilde{y}_n^*}(y_n) = \begin{cases} 1 - \frac{|y_n - \mathbf{c'x_n}|}{\mathbf{w'x_n}} & \mathbf{x_n} \neq \mathbf{0} \\ 1 & \mathbf{x_n} = \mathbf{0}, y_n = 0 \\ 0 & otherwise \end{cases} \tag{3.14}$$

where $\mathbf{c'} = (c_0, c_1, \ldots, c_P)$, $\mathbf{w'} = (w_0, w_1, \ldots, w_P)$ and $\mathbf{x_n} = (1, x_{n1}, \ldots, x_{nP})'$. Fuzzy estimated output, whose possibility distribution is shown in figure (3.2) may be denoted also in terms of center and spread as $\tilde{y}_n^* = (\mathbf{c'x_n}; \mathbf{c'x_n})$.

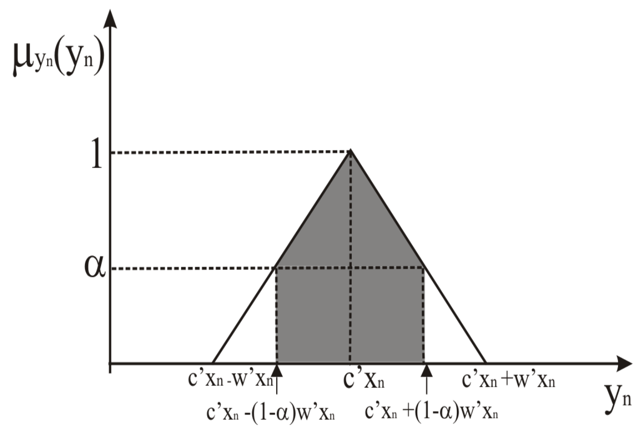

Figure 3.2: Possibility distribution of fuzzy estimated output

Differently from statistical regression, the deviations between data and linear model are assumed to depend on the vagueness of the parameters and not on the measurement errors. Then, the basic idea of Tanaka's approach was to minimize the uncertainty of the estimates minimizing the total spread of the fuzzy coefficients. Spread minimization must be pursued under the constraint of the inclusion of the whole given data set, satisfying a degree of belief α $(0 < \alpha < 1)$ defined by the decision maker. In contrast to the *frequentist interpretation* of probability, describing whether or not an event will occur on average, the notion of degree of possibility α is usually used to describe degrees of feasibility to which some conditions exist. In other words, the objective in fuzzy regression is to fit as much as possible the scatter plot, given a degree of fitness α chosen by the decision maker. An example of fuzzy regression model is given in figure (3.3), where it can be seen that the higher degree of fitness the wider fuzzy interval.

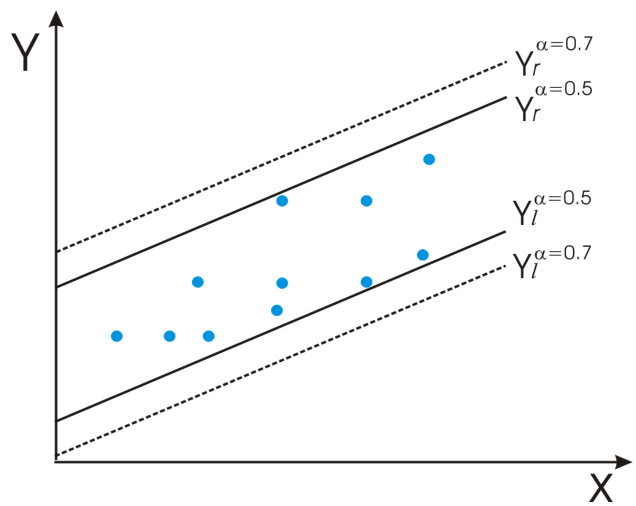

Figure 3.3: Example of fuzzy regression

The estimation problem is afforded by the Mathematical Programming, whose basic concept are explained in the following section.

Mathematical Programming for Possibilistic regression

In possibilistic regression, estimation procedure is pursued through an extensively use of Mathematical Programming (MP), which includes a wide range of powerful computer based optimization methods.
The aim of MP is to minimize (or maximize) a *real function* subject to constraints expressed as a set of inequalities (Boyd & Vandenberghe 2004).

Formally a mathematical programming problem is expressed as:

$$\begin{cases} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_l(\mathbf{x}) \leq \mathbf{b}_l \quad l = 1, \ldots, L \end{cases} \tag{3.15}$$

where the vector $x = (x_1, \ldots, x_M)$ is the *optimization (decision) variable* of the problem, the function $f_0 : \mathbf{R}^M \to \mathbf{R}$ is the *objective* function the function $f_i : \mathbf{R}^M \to \mathbf{R}, l = 1, \ldots, L$ are the (inequality) *constraint functions*, and the constants $b_1, \ldots, b_L$ are the limits, or bounds, for the constraints.
The formulation in (3.15) is the *standard form* of a MP. Any other MP problem such as maximization problems, or problems with constraints on alternative forms, may be always rewritten as (3.15).
There is some technical terminology associated with mathematical programming. Variables satisfying all the constraints simultaneously are said to form a *feasible solution* to the problem. The constraints set define the *feasible region* of the problem under consideration. A feasible solution that in addition optimizes the objective function is called an *optimal feasible solution*.
A MP in which the objective function and the constraints are all linear is defined *Linear Programming* (LP) problem. Formally:

$$\begin{cases} \min & \mathbf{c}'\mathbf{x} \\ \text{s.t.} & A\mathbf{x} \geq \mathbf{b} \end{cases} \tag{3.16}$$

where $A = \{a_{lm}\}$ is an $L \times M$ matrix.
A LP problem may be geometrically represented. Specifically, each constraint corresponds to an half-space and the constraints set give rise to a convex polyhedron, which corresponds to the feasible region. The convex polyhedron may be bounded (polytope) or unbounded.
Give the feasibility set:

$$\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^L \mid \mathbf{Ax} - \mathbf{b} \geq 0\} \tag{3.17}$$

and $\overline{\mathbf{x}} \in \mathcal{P}$, $\overline{\mathbf{x}}$ is an extreme point (vertex) of $\mathcal{P}$ *iff* there exists a submatrix $\mathbf{A}^* \in \mathbb{R}^{L \times L}$ of $\mathbf{A}$, $\mathbf{A}^*$ non singular such that $\mathbf{A}^*\overline{\mathbf{x}} = \mathbf{b}^*$.

Thorem 1. *If a feasible region S is bounded, then the problem has an optimal solution at an extreme point of S.*

Since the objective function is also linear, all local optima are automatically global optima. The linear objective function implies that an optimal solution can only occur at a boundary point of the feasible region, unless the objective function is constant.
To find a vertex it is necessary to solve system of constraints, i.e. working with matrix A. The number of constraints defines the amount of effort, that means LP can handle many more decision variables than constraints.
Solving a linear program can result in three possible situations, as stated by the following *Fundamental theorem of Linear Programming*:

Thorem 2. *If the feasible region does not contain any line, then one and only one of the following statements holds: i)the problem is infeasible; ii) the problem is unbounded from below; iii) the problem has optimal solutions and there exists an extreme point of the feasible region which is optimal.*

The LP problem is infeasible if there are no values of the decision variables that simultaneously satisfy all the constraints. This occurs when the constraints contradict each other (for instance, $x < 2$ and $x > 2$) then the feasible region is empty since there no solutions at all.
The problem has an unbounded solution, if maximizing an objective function its value may be increased indefinitely without violating any of the constraints or if minimizing, the value of the objective function may be decreased indefinitely. In this case there is no optimal solution since solutions with arbitrarily high values of the objective function can be constructed.
The problem has at least one finite optimal solution and often it has multiple optimal solutions. Geometrically, the optimum is always attained at a vertex of the polyhedron. If there are multiple solutions they cover an edge or face

of the polyhedron, or even the entire polyhedron (if the objective function were constant).

The simplex method for solving linear programs, which will be discussed in Appendix, provides an efficient procedure for constructing an optimal solution, if one exists, or for determining whether the problem is infeasible or unbounded (Chvátal 1983, Gill, Murray & Wright 1981).

The simplex algorithm, so named because of the geometry of the feasible set, More recently, new procedures called *interior-point algorithms* have been proposed for solving *linear programming problems*.

Thanaka's model: estimation and output interpretation

Parameters in FPR are estimated according to the following Linear Programming (LP) problem, where the objective function aims at minimizing the spread parameters and the constraints guarantee that observed data fall inside fuzzy interval:

$$\text{minimize} \sum_{n=1}^{N} \sum_{p=0}^{P} w_p |x_{np}| \tag{3.18}$$

subject to the following constraints:

$$\left(c_0 + \sum_{p=1}^{P} c_p x_{np} \right) + (1-\alpha) \left(w_0 + \sum_{p=1}^{P} w_p |x_{np}| \right) \geq y_n \tag{3.19}$$

$$\left(c_0 + \sum_{p=1}^{P} c_p x_{np} \right) - (1-\alpha) \left(w_0 + \sum_{p=1}^{P} w_p |x_{np}| \right) \leq y_n \tag{3.20}$$

where $x_{n0} = 1$ $(n = 1, \ldots, N)$, $w_p \geq 0$ and $c_p \in R$ $(p = 1, \ldots, P)$. The LP problem has always feasible solutions. In fact, as discussed in (3.3.1), the constraints (3.19 and 3.20) do not contradict each other. The fuzzy parameters w_p are determined as the optimal solution solution of this LP problem. The uniqueness of such solution should be further investigated.

There are no restrictive assumptions on the model. Increasing the α–coefficient expands the fuzzy intervals as well as increasing the confidence level in statistical regression expands the confidence interval width.
The degree of possibility α is a precise but subjective measure that depends on the context. A rule of thumb for choosing the value of α is given in (Tanaka & Guo 1999): *If there are enough data, the possibility shown from these data is sufficient. Thus, $\alpha = 0$ is recommendable. If the given data are considered to include only half, comparing with the ideal number of data, $\alpha = 0.5$ is recommendable.*

Wang and Tsaur (2000) provided a proper interpretation of the fuzzy regression interval. The basic idea was to find a representative value of the fuzzy interval among the infinite values enclosed inside the interval boundaries. Let $\underline{y}_n$ and $\overline{y}_n$ be the lower and the upper bound of the estimated value $\tilde{y}_n^*$. The authors proved that in models with symmetric coefficients the mean value of $\tilde{y}_n^*$ given by:

$$y_n^m = \frac{\underline{y}_n + \overline{y}_n}{2}$$

is equal to the the value occurring with the higher possibility level $\tilde{y}_n^1$ ($\alpha = 1$). In other words, $\tilde{y}_n^1$ is the best representative value of the fuzzy interval and, more generally, the regression line $\tilde{Y}^1$ has the best ability to interpret the given data. Starting from this results the following quantities were defined:

- *Total Sum of Squares* (SST)
  a measure of the total variation of y_n in $\tilde{y}_n$

$$SST = \sum_{n=1}^{N} \left(y_n - \underline{y}_n \right)^2 + \sum_{n=1}^{N} \left(\underline{y}_n - y_n \right)^2 \qquad (3.21)$$

- *Regression Sum of Squares* (SSR)

a measure of the variation of $\tilde{y}_n^1$ in $\tilde{y}_n^*$

$$SSR = \sum_{n=1}^{N} \left(\tilde{y}_n^1 - \underline{y}_n\right)^2 + \sum_{n=1}^{N} \left(\underline{y}_n - \tilde{y}_n^1\right)^2 \tag{3.22}$$

- *Error Sum of Squares* (SSE)
  an estimate of the difference when $\tilde{y}_n^1$ is used to estimate y_n

$$SSE = \sum_{n=1}^{N} \left(y_n - \underline{y}_n\right)^2 + \sum_{n=1}^{N} \left(\underline{y}_n - y_n\right)^2 \tag{3.23}$$

Thus, using 3.21 and 3.23 an index of confidence is built, similar to the traditional R^2 in Statistics. The index is defined as IC=SSR/SST and gives a measure of the variation of Y between $\underline{Y}$ and $\overline{Y}$. The higher the IC the better is the $\tilde{Y}^1$ used to represent the given data. The partial version of the index IC is then used by the authors for a variables selection procedure. It should be mentioned that IC increases as well increases the the value of α. However increasing α the spread coefficients become wider. This trade-off should be considered in choosing the α possibility level of the estimates.

Many contributions have compared statistical regression and fuzzy regression (Kim, Moskowitz & Koksalan 1996, Kim & Chen 1997, Romano & Palumbo 2006*b*).
Kim et *al* exhaustively analyzed the two approaches both conceptually and empirically.
From a conceptual point of view the basic assumptions, the estimation procedure and the usage of each one of the two models are considered. As it is well known, statistical regression makes rigid assumption on the error terms which are due both to relevant omitted variables and random measurement errors. Differently, in FPR there are no restrictive assumption on the errors. It is assumed that errors, reflected in the spread of the fuzzy parameters, are due to the indefiniteness of the system.

Parameter estimation both in FPR and LS regression aims at minimizing the difference between the model and the given data. In classical statistical regression (LS) the objective function is the minimization of the squared residuals. The normality assumption guarantees BLUE estimators and confidence intervals for the coefficients. In fuzzy regression the objective function is the minimization of the spread parameters. The constraints guarantee the estimated values from the model include the observed values for a certain $\alpha-$level, where $0 < \alpha < 1$.

The usage of statistical and fuzzy regression is very different. In fact, the first is more focused on predictions whereas the second focuses on the given data. This is clarified looking at the confidence and possibility intervals. As a matter of fact, a 95% confidence interval for the regression coefficient means that if many independent samples are taken from the same population and a 95% confidence interval for the regression coefficient is built for each one of them, then 95% of intervals will contain the true value of the coefficient. This means that in building confidence interval strong focus is given to the predictions. In fuzzy regression, a 0.95 possibility or fuzzy interval for the regression coefficient indicates the narrowest interval obtained when each observation has a membership value of at least $0.95 - cut$ to its fuzzy interval. This means that fuzzy regression is more focused on the given data rather than on predictions. Another important difference between statistical and fuzzy regression is that as the sample size increases the spread of confidence interval decreases, whereas the spread of fuzzy interval increases. From a statistical point of view this happens since more information leads to more accurate estimates. From a possibilistic point of view, each observation represents a portion of the possibility to be explained by the model. Hence, as the number of observations increases more possibility need to be explained, increasing the spread of fuzzy interval: *fuzzy model explains possibility by sacrificing precision.* The authors through a simulation study compare the performance of the two approaches.

The main conclusions of the simulation study are given:

- statistical regression is superior to fuzzy in terms of predictive capabilities;
- descriptive performance depends on various factor associated with the data set (size, quality) and proper specificity of the model (aptness of the model, heteroscedasticity, autocorrelation, non randomness of error terms). Fuzzy linear regression performance becomes relatively better when:
    - the size of the data set diminishes;
    - the aptness of the regression model deteriorates.

Romano and Palumbo (2006) compare the classical statistical linear regression model with the fuzzy regression model through an empirical study on simulated data. They point out that fuzzy estimators are unbiased and not affected by quasi multi-collinearity. The simulation study will be extensively described in (3.3.1).

Possibilistic regression: criticism and improvements

Although Tanaka's basic model is still one of the most widely used fuzzy model due to its simplicity, it presents several shortcomings.
Diamond (1988) pointed out how this approach is far from the statistical regression line of thought. In (1994) Peters as well as Redden and Woodall criticized the extremely sensitiveness of the model to the outliers. Sakawa and Yano (1992) stressed that Tanaka's models does not take into account fuzzy explanatory variables. Kao and Chyu (2002) showed that spread of the estimated response becomes wider when the independent variables increase their magnitude and/or when more data are included in the model.
Many researchers over the years have built new models and proposed some solutions to improve Tanaka's basic model. Tanaka himself generalized the

basic model (3.11) to coefficients with an exponential possibility distribution, which is similar to a normal distribution in probability theory (Tanaka & Guo 1999). Exponential possibility distribution permits to consider interactive possibility distributions avoiding the typical LP problem of crisp coefficients. The same problem has been solved by formulating the Tanaka's basic model estimation procedure in terms of a Quadratic Programming problem instead of Linear Programming one.
Furthermore, Tanaka extended the basic model (3.11) to the case of fuzzy dependent variable. Let $\tilde{y}_n = (y_n; e_n)_L$ be the fuzzy value of the dependent variable, where y_n and e_n are the center and the spread, respectively. The estimation procedure aims to minimize the fuzziness of the system, minimizing the spread of the parameters as well as in (3.18). The additional source of fuzziness embed in the spreads e_n is taken into account by the constraints of the following optimization problem:

$$\text{minimize} \sum_{n=1}^{N} \sum_{p=0}^{P} w_p |x_{np}| \tag{3.24}$$

subject to:

$$\left(c_0 + \sum_{p=1}^{P} c_p x_{np}\right) + (1-\alpha)\left(w_0 + \sum_{p=1}^{P} w_p |x_{np}|\right) \geq y_n + (1-\alpha)e_n \tag{3.25}$$

$$\left(c_0 + \sum_{p=1}^{P} c_p x_{np}\right) - (1-\alpha)\left(w_0 + \sum_{p=1}^{P} w_p |x_{np}|\right) \leq y_n - (1-\alpha)e_n \tag{3.26}$$

where $x_{n0} = 1$ $(n = 1, \dots, N)$, $w_p \geq 0$ and $c_p \in R$ $(p = 1, \dots, P)$.
In 2000, Wang and Tsaur proposed a new version of Tanaka model for crisp input (independent variables) and fuzzy output (dependent variable)(Wang & Tsaur 2000*b*). The main aim is to solve the problem of too wide ranges in estimation which leads this model to be not very useful for practical application. The idea is to minimize the fuzziness of the system minimizing

the spread of both parameters and dependent variable. Therefore (3.24) is rewritten as follows:

$$\text{minimize} \sum_{n=1}^{N} (\mathbf{w}'|\mathbf{x_n}| - e_n) \tag{3.27}$$

subject to the constraints (3.25-3.26). Moreover, the authors proposed a new model coherently with the concept of Tanaka's approach. The model consists in a new optimization problem whose objective function aims at minimizing both the spread parameters and the errors in central values. The solution is obtained by the following Quadratic Programming (QP) problem:

$$\text{minimize} \sum_{n=1}^{N} (w'|\mathbf{x_n}| - e_n)^2 \tag{3.28}$$

$$\left(c_0 + \sum_{p=1}^{P} c_p x_{np}\right) + \left(w_0 + \sum_{p=1}^{P} w_p |x_{np}|\right) \lesssim y_n + e_n \tag{3.29}$$

$$\left(c_0 + \sum_{p=1}^{P} c_p x_{np}\right) - \left(w_0 + \sum_{p=1}^{P} w_p |x_{np}|\right) \gtrsim y_n - e_n \tag{3.30}$$

where $x_{n0} = 1$ $(n = 1, \ldots, N)$, $w_p; c_p \in R$ $(p = 1, \ldots, P)$.
In (3.29-3.30) α is selected equal to 0 and the constraints $(\leq; \geq)$ are relaxed into $(\lesssim; \gtrsim)$, respectively, for a more efficient solution due to a wider feasible region. Furthermore, the constraint of positive spread coefficients is relaxed. This avoids the problem of conflicting trends, i.e. spread size dependent on the independent variables magnitude.
Moving from Peter's paper (Peters 1994), Chen (2001) proposed a solution to handle the outliers problem in case of fuzzy dependent variable. The main idea is to assign a pre-defined k value which discriminates the potential outliers. The restriction leads to an additional constraint in the optimization

problem 3.24. The new optimization problem is given by:

$$\text{minimize} \sum_{n=1}^{N} \sum_{p=0}^{P} w_p |x_{np}| \tag{3.31}$$

subject to the following constraints:

$$\begin{aligned}
\left(c_0 + \sum_{p=1}^{P} c_p x_{np}\right) + \quad & (1-\alpha)\left(w_0 + \textstyle\sum_{p=1}^{P} w_p |x_{np}|\right) \geq y_n + (1-\alpha) e_n \\
\left(c_0 + \sum_{p=1}^{P} c_p x_{np}\right) - \quad & (1-\alpha)\left(w_0 + \textstyle\sum_{p=1}^{P} w_p |x_{np}|\right) \leq y_n - (1-\alpha) e_n \\
& \left(w_0 + \textstyle\sum_{p=1}^{P} w_p |x_{np}|\right) - e_n \leq k
\end{aligned}$$

where $x_{n0} = 1$ $(n = 1, \ldots, N)$, $w_p; c_p \in R$ $(p = 1, \ldots, P)$.
The third constraint analyzes the difference between the spread of the estimated and observed data, respectively. If their difference is higher than k the problem (3.31) has no feasible solution. This means that the presence of outliers is detected. The author gives different criteria for choosing the value of k, some of them follows:

$$\begin{aligned}
k &= \max_n \{e_n\}_{n=1}^{N} \\
k &= \min_n \{e_n\}_{n=1}^{N} \\
k &= e^m = \sum_{n=1}^{N} \frac{e_n}{N} \\
k &= 3 s_{e_n} = 3 \sqrt{\sum_{n=1}^{N} \frac{(e_n - e^m)^2}{N-1}}
\end{aligned}$$

Smaller values of k lead to restrictive requirements as well as larger values lead to consider the whole data as abnormal.
More recently, a model for dealing with fuzzy independent variables has been

proposed (Nasrabadi & Nasrabadi 2004).
Let $\tilde{x}_p = (x_p, r_p)$ be the p-th independent fuzzy variable. The model is formalized as follows:

$$\tilde{y}_n = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{x}_{n1} + \ldots + \tilde{\beta}_p \tilde{x}_{np}, + \ldots + \tilde{\beta}_P \tilde{x}_{nP} \tag{3.32}$$

The estimation procedure leads to the following QP problem:

$$\text{minimize} \sum_{n=1}^{N} (\mathbf{w}'|\mathbf{r_n}| - e_n)^2 \tag{3.33}$$

where $\mathbf{r}_n = (r_0, r_1, \ldots, r_N)\prime$. The problem 3.33 is subject to the constraints (**??**) but with spread parameters unrestricted in sign.

A simulation study

In order to compare the performance of the ordinary regression and the fuzzy regression models, a simulation study is carried out.
The main aim is to compare confidence intervals and fuzzy intervals for regression coefficients.
Data are simulated under the following hypotheses:

1. The true parameter values of the model are assumed to be know. The model has been simulated both under the null hypothesis of coefficients equal to zero and equal to 0.5. Thus, the regression model from which the data are generated is:

$$y_n = \beta_o + \beta_1 x_{n1} + \beta_2 x_{n2} + \epsilon_n \tag{3.34}$$

   where $\epsilon_n \sim N(0, 0.25)$, $\beta = [0, 0, 0]'$ and $\beta = [0.5, 0.5, 0.5]'$ under the null hypothesis of coefficients equal to zero and equal to 0.5, respectively.

2. Explanatory variables are assumed to be dependent or independent. In

the first case, $[X_1, X_2] \sim N$ having mean vector $\mu' = [0, 0]$ and extra diagonal terms of the var–cov matrix $\sigma_{1,2}$ and $\sigma_{2,1}$ equal to 0.75 and 0.75, respectively. In the second case, $[X_1, X_2] \sim N$ having mean vector $\mu' = [0, 0]$ and extra diagonal terms of the var–cov matrix $\sigma_{1,2}$ and $\sigma_{2,1}$ both equal to 0.

3. As test data, 3000 set of 25 statistical units have been used.

4. The confidence and the possibility coefficients have been set equal to 0.95 and 0.75, respectively.

The experimental factors are summarized in table (3.1).

Table 3.1: Experimental factors

$\beta = [b_0, b_1, b_2]'$	$[X_1, X_2] \sim N(0, \sigma^2)$	*replicates*	N
$\beta = [0, 0, 0]'$	$\mu' = [0, 0]; cov(X_1, X_2) = 0$	3000	25
$\beta = [0.5, 0.5, 0.5]'$	$\mu' = [0, 0]; cov(X_1, X_2) = 0$	3000	25
$\beta = [0, 0, 0]'$	$\mu' = [0, 0]; cov(X_1, X_2) = 0.75$	3000	25
$\beta = [0.5, 0.5, 0.5]'$	$\mu' = [0, 0]; cov(X_1, X_2) = 0.75$	3000	25

Results only for one regression coefficient are shown. Anyway, same results are obtained for the other coefficient.
Figure (3.4) shows results under the null hypothesis of coefficients equal to zero and no correlation between predictors.
Figure (3.5) shows results under the null hypothesis of coefficients equal to zero and correlated predictors.
It should be noted that in both cases intervals of Tanaka's model are narrowest.
Analogous results are obtained under the null hypothesis of coefficient equal to 0.5. Obviously the OLS intervals are narrower when there is no collinearity among the predictors.
Figures (3.4-3.5) show also that fuzzy estimators are unbiased as well as OLS

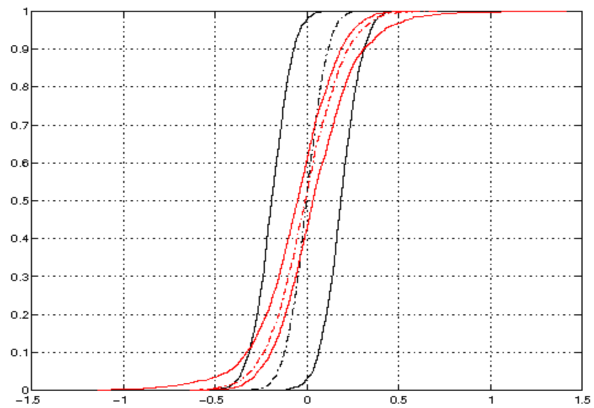

Figure 3.4: OLS confidence intervals (black lines) and fuzzy intervals (red lines) under the hypothesis $\beta = 0$ and independent predictors;

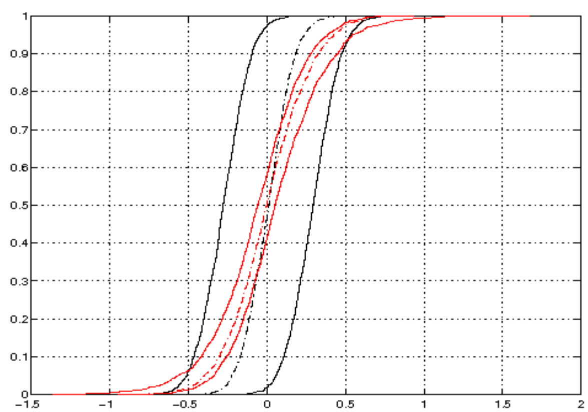

Figure 3.5: OLS confidence intervals (black lines) and fuzzy intervals (red lines) under the hypothesis $\beta = 0$ and dependent predictors;

estimators. In fact, analytically $B_{ols} = 0.0014$ and $B_{fuzzy} = 0.0034$ under the null hypothesis of coefficients equal to one.

Empirical results underline a better performance of the fuzzy regression in presence of high correlations among the predictors. It follows that fuzzy regression may be considered a good alternative to OLS regression in case of multicollinearity

The same simulation study is performed on the Tanaka's model whose estimation parameters is obtained as solution of the following QP problem:

$$\min\,(|\mathbf{X}|\mathbf{w})^T(|\mathbf{X}|\mathbf{w}) \Longleftrightarrow \min\sum_{n=1}^{N}(|\mathbf{x}_n|\cdot\mathbf{w})^2 \tag{3.35}$$

subject to the following constraints:

$$\begin{cases} \mathbf{Xc}-|\mathbf{X}|\mathbf{w} & \leq \mathbf{y} \\ \mathbf{Xc}+|\mathbf{X}|\mathbf{w} & \geq \mathbf{y} \\ \alpha & \geq 0 \end{cases} \tag{3.36}$$

where $\mathbf{X}$ is a $(N \times P)$ data matrix with $N > P$, $\mathbf{c}$ and $\mathbf{w}$ are two (unknown) $((P+1)\times 1)$ parameter vectors.

As discussed in (3.3.1), the QP (Coleman & Li 1996) has been introduced in order to solve the problem of crisp parameters, i.e. no information about the fuzziness inside the system is taken into account by the model.

Simulation results under the condition $\beta_1 = 0$ and independence/dependence of X_1 X_2 are shown in figure (3.6) and (3.7), respectively. Empirical results show that FPR presents same properties both using a LP or a QP approach. This simulation study point out two important information:

- LS and FPR produce unbiased estimates under the same hypothesis: $\mathrm{E}(\epsilon) = \mathbf{0}$

- FPR estimators are not affected by quasi multi-collinearity

Such conclusions, combined to the ones from Kim et *al.* introduced in (3.3.1), leads to consider FPR model a very useful tool for analyzing system where:

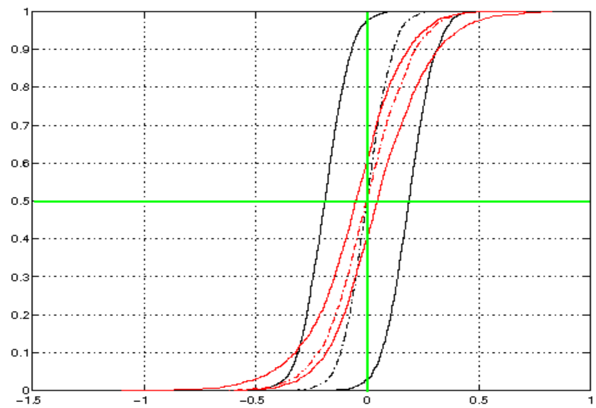

Figure 3.6: OLS confidence intervals (black lines) and fuzzy intervals (red lines) under the hypothesis $\beta = 0$ and independent predictors;

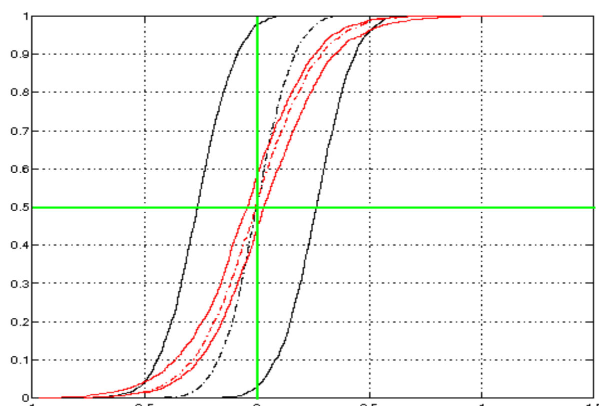

Figure 3.7: OLS confidence intervals (black lines) and fuzzy intervals (red lines) under the hypothesis $\beta = 0$ and dependent predictors;

a) the data set is too small; b) the normality assumption of the error term is not easy to verify; c) the linearity assumption behind the model is inappropriate; d) there is vagueness in the relationship between the dependent and

independent variables and/or vagueness in the same variables (fuzzy variables); e) predictors are highly correlated; f) number of predictors higher than number of observations.
It is worth mentioning that all these conditions are very common in real world phenomena.

3.3.2 Fuzzy Least Squares Regression

The most strong opposition to the possibilistic regression from the statistical community is that such approach is very different from the classical one. This led many authors to propose new fuzzy models more in line with the traditional statistical regression. The results of these contributions has produced models with a better explanatory power consisting in fuzzy interval narrower than those from possibilistic regression. However, these new models based on the classical LS approach are computationally more demanding and requires often restrictive assumptions.
Following the LS approach, FLSR aims to minimize the distance between observed and estimated fuzzy data. Many criteria for measuring such distance have been proposed over the years.
In table (3.2) are summarized the main characteristics of FPR and FLSR:

Table 3.2: FPR and FLSR

hypotheses	
soft modelling approach	hard modelling approach
fit	
minimum fuzziness criterion	least squares criterion
numerical approach	
MP problem	distance between fuzzy numbers
computational complexity	
$O(N^2P)$	$O(N^2P^4)$
where N is the number of samples and P the number of variables	

In 1987, Celmins proposed the first FLSR based on the *compatibility mea-*

sure. Let $\mu_{\tilde{A}}(\omega)$ and $\mu_{\tilde{B}}(\omega)$ be the membership functions of two quantities $\tilde{A}$ and $\tilde{B}$. If $\mu_{\tilde{A}}(\omega)$ and $\mu_{\tilde{B}}(\omega)$ are normalized triangular membership functions, the *compatibility measure* between $\tilde{B}$ and $\tilde{C}$ is expressed as:

$$\gamma(\tilde{A}, \tilde{B}) = \max_{\omega} min\{\mu_{\tilde{A}}(\omega), \mu_{\tilde{B}}(\omega)\} \tag{3.37}$$

where $0 \leq \gamma \leq 1$.
If there is no overlapping between $\tilde{A}$ and $\tilde{B}$ then $\gamma = 0$. On the other hand, $\gamma = 1$ when the centers overlap. The measure γ assumes the same meaning of the α level in (3.3.1). Thus, the idea of this approach is to maximize the overall compatibility between data and model. This objective may be reformulated in a minimization problem with the following objective function:

$$W = \sum_{n=1}^{N} (1 - \gamma_n)^2 \tag{3.38}$$

Consider the following model for crisp data with a single independent variable:

$$\tilde{Y} = \tilde{\beta}_0 + \tilde{\beta}_1 X \tag{3.39}$$

where the coefficients are symmetric triangular fuzzy numbers denoted by $\tilde{\beta}_p = (c_p; w_p)_L$. The final formula for the FLSR using (3.38) is given by:

$$\tilde{Y} = c_0 + c_1 X \pm \sqrt{w_0^2 + 2w_{01}^2 X + w_1^2 X^2} \tag{3.40}$$

The first part of the equation (3.40) corresponds to the centerline of the fuzzy regression obtained as weighted LS regression. The second part specifies the upper and lower boundary of the regression model. The term w_{01} is the fuzzy concordance between $\tilde{\beta}_0$ and $\tilde{\beta}_1$, a concept similar to the *covariance* in statistics.
In 1988, Diamond proposed a new FLSR more in line with the traditional

LS regression. In fact, the objective function is an L^2 metric between fuzzy numbers. Considering the model (3.39), the aim is to minimize the distance between the data and the model as follows:

$$D = \min_{\tilde{\beta}_0, \tilde{\beta}_1} \sum_{n=1}^{N} d(\tilde{\beta}_0 + \tilde{\beta}_1 x_n, y_n)^2 \tag{3.41}$$

It is possible to expand the (3.41) as follows:

$$\begin{aligned} D &= (c_0 + w_0 + c_1 x_{n1} + w_1 x_{n1} - y_n - e_n)^2 + (c_0 + c_1 x_{n1} + -y_n - e_n)^2 \\ &\quad + (c_0 - w_0 + c_1 x_{n1} - w_1 x_{n1} - y_n - e_n)^2 \end{aligned} \tag{3.42}$$

The parameter estimates are obtained from the derivatives associated with (3.42) being set equal to zero. This approach gives error estimates of residuals. However it requires restrictive assumptions, i.e. same trends between center and spread and positive values of c_0, c_1, w_0, w_1.
Successively, Diamond's approach has been improved by many authors.
Ming et *al.* (1997) proposed a new metric for dealing all fuzzy numbers represented by single maxima piecewise continuous functions. The model involved fuzzy dependent and independent fuzzy variable.
Savic and Pedrycz proposed a model which is a combination of the possibilistic and the LS approach. The model consists in a double step procedure. In the first step an ordinary LS regression procedure is implemented on the centers of the fuzzy parameters. In the second step a possibilistic regression is performed to find the spread parameters, using the centers estimated in the first step.
Wang and Tsaur (2000) proposed a modified version of Diamond's model relaxing its restrictive assumptions.
More recently, Xu and Li (2001) have proposed a new model taking into account a new distance defined on a fuzzy number space.
Following the idea of the Savic and Pedrycz, D'Urso and Gastaldi (2000) have

proposed a new FLSR. The model, named *doubly linear adaptive model*, is a model for crisp input and symmetric triangular fuzzy output. It follows the classical approach minimizing the Euclidean distance between the observed and the estimated fuzzy output. Based on the assumption that the spreads are proportional to the respective centers, first a linear model for the centers is built and after a model for the spreads is derived:

$$\begin{cases} \mathbf{c} = \mathbf{c}^* + \varepsilon_c & \text{where } \mathbf{c}^* = \mathbf{X}\mathbf{a} \\ \mathbf{w} = \mathbf{w}^* + \varepsilon_w & \text{where } \mathbf{w}^* = \mathbf{c}^* b + \mathbf{1} d \end{cases} \tag{3.43}$$

where $\mathbf{X}$ is the matrix including the independent variables, $\mathbf{c}$ and $\mathbf{c}^*$ are the observed and estimated centers, $\mathbf{w}$ and $\mathbf{w}^*$ are the observed and the estimated spreads, and $\mathbf{a}$, $\mathbf{b}$, $\mathbf{d}$ are the parameters of the respective models. It should be mentioned that the model does not provide fuzzy parameters. However, in a recent work devoting to generalize their procedure to LR fuzzy numbers the authors set up a procedure for estimating also the spreads of the regression coefficients (Coppi, D'Urso, Giordani & Santoro 2006).

Chapter 4

Structural Equation Models based on Fuzzy Regression

4.1 Introduction

The analysis of a socio-economic system would take into account many complex relationships. As discussed in chapter (3), *Regression Analysis* is known as one of the most widely used statistical methods for analyzing the dependence between two sets of variables. However, the complexity of many real world *phenomena* makes single equation models ineffective to analyze and describe dependence structures that are in the data. As a matter of fact, the multiple regression equation is additive by definition. Thus, only direct relationships between the independent variables and the dependent variable are allowed. This strongly limits the variables to have no indirect effects on each other, as instead permitted in *path analysis* (Tukey 1964, Alwin & Hauser 1975). Path models are a logical extension of regression models as they involve the analysis of simultaneous multiple regression equations. More specifically, a path model is a relational model with direct and indirect effects between observed variables. It does not represent a tool for specifying a model, but it just estimates the effects among the variables once the model has been specified by the researcher. When the variables inside the path are latent variables whose measure is inferred by a set of observed indicators, path analysis is termed *structural equation modeling*.

Structural Equation Models (SEM) (Bollen 1989, Kaplan 2000) combine the idea behind path analysis with the basic principles of confirmatory factor analysis. Factor analysis (Thurstone 1931) presumes that a number of factors smaller than the number of observed variables are responsible for the shared variance-covariance amongst the observed variables. Hence, SEM receive from confirmatory factor analysis the idea that different subsets or blocks of variables are expression of different concepts. These concepts are named *latent variables* (LV) as they are not directly observable but measurable by means of a set of *manifest variables* (MV). On the other hand, such blocks of variables are linked to each other through the existing relations among the respective LVs, as well as in path analysis single observed variables are connected among them. Roughly speaking, path models are used for defining relations among variables while confirmatory factor analysis for creating latent variables.

In the SEM framework, the literature presents two dominant approaches: covariance-based SEM (Jöreskog 1970) and *partial least squares*-based SEM (Wold 1982). The present work mainly refers to SEM-PLS, alternatively defined PLS Path Modeling (PLS-PM) (Tenenhaus, Vinzi, Chatelin & Lauro 2005). However, differences between the two approaches will be discussed.

In the last years, SEM have become a reference technique for analyzing real world phenomena. However, as fully explained in chapters (2) and (3), such systems are mostly characterized by fuzzy uncertainty. Moreover, fuzzy possibilistic regression and PLS path modeling share several characteristics, yielding the idea to combine them into a new strategy of analysis based on a "fuzzy approach to PLS-PM". Such a strategy regards a two-stage procedure for multi-block analysis combining *fuzzy linear regression* and PLS path modeling.

The new methodology is applied to face the model comparison problem. In other words, when according to one specific characteristic, a data set is *a priori* divided into homogeneous groups, the same model or SEM may be replicated and estimated according to each group. This approach offers more efficient estimates and detailed information but it is not obvious how to compare then the different models.
Various approaches have been proposed for the model comparison and the related literature is quite wide. However, it is possible to distinguish two dominant approaches. The first focuses on the comparison of different models for the same data set, based on the goodness of fit indexes (Myung & Pitt 2003). The second concerns the comparison of the same model for different data set, and the most significative contributions were in the following frameworks: time series analysis (Piccolo 1990); Bayesian statistics (O'Hagan 1995); SEM (Lee & Song 2001, Chin & Dibbern 2007).
In the following strong focus is given to the presence of multi-group structure data, implying the estimation of the same model for the different groups.
In the regression analysis framework, statistical methods for comparing models are mostly based on the comparison of the estimated model parameters (Clogg, Petkova & Haritou 1995). It is quite intuitive to understand differences among several statistical populations or samples looking at the respective model parameters. However, the same model fits differently with respect to the different populations so that the model parameters are expression of a different amount of information. It is worth noticing that the whole information is given both by the part explained by the model - through the model parameters - and the part indicated as residual. These two kinds of information are called *structural information* and *residual information*, respectively.
In this framework, the idea is to face the model parameters comparison by introducing the possibilistic fuzzy regression (3.11) in the PLS-PM context.
As extensively explained in chapter (3), the estimation of *fuzzy* parameters, instead of single-valued (crisp) parameters, permits to gather both the

structural and the residual information. Thus, the proposed model comparison is accomplished taking into account the fuzzy parameters and introducing a suitable classification approach based on fuzzy variables (Romano & Palumbo 2006*a*).

4.2 Structural Equation Models

Structural equation modeling is a rather general methodology in the contexts of regression analysis, path analysis and factor analysis. The basic aim is to specify and estimate a pattern of linear relationships among variables. Such network of causal relations is a more realistic representation of real-world phenomena than simple linear models.
The development of SEM may be traced bach to the 1970s, when two seminal papers were published approaching SEM from two different perspectives.
Essentially developed in a social domain, structural equations were firstly introduced by Jöreskog (Jöreskog 1970) as confirmatory models to asses cause-effect relations among two or more set of variables, based on maximum likelihood (ML) estimation method (SEM-ML). This method, known as LISREL (*LInear Structural RELations*), has been for many years the only estimation method for SEM. The term LISREL was initially used for the software implementing the methodology. However, it had such a rapid development that the methodology and software have been associated to each other.
In 1975, Wold finalized a *soft modeling* approach to the analysis of the relations among several blocks of variables observed on the same statistical units. This method, known as *PLS approach* to SEM (SEM-PLS), was developed as a flexible technique for handling a huge amount of data characterized by missing values, strongly correlated variables and small sample size as compared to the number of variables.
Several authors have compared the two approaches over the years; see, for example, (Jöreskog & Wold 1982), (Fornell & Bookstein 1982), (Djkstra 1983).

The two approaches differ in the objectives of the analysis, the statistical assumptions, the estimation procedures and related output.
According to their respective objectives, LISREL may be defined as a *causal model* whereas PLS-PM as a *predictive model*. The aim of LISREL is that the model *a priori* specified by the researcher is corroborated by the data. Hence, parameter estimates are obtained so that the implied covariance structure C estimated by the model is as close as possible to the empirical covariance matrix S observed for the manifest variables. On the other hand, the aim of PLS-PM is to achieve the best set of predictions available for a given data set. Hence, it is a variance-based approach aiming to maximize the explained variance of each latent variable inside the model.
Different objectives imply differences in the estimation procedures. In LISREL, the estimation process involves the selection of a particular *fitting function* for minimizing the distance between C and S. Maximum Likelihood (ML) estimates are mainly used but such method requires the multivariate-normal distribution. When there are substantial deviations from the normality assumption, alternative procedures such as the Unweighted Least Square (ULS) may be used. On the other side, PLS-PM performs an iterative sequence of interdependent OLS regressions, analyzing one block at a time. The parameters are estimated so that residual variances of all the dependent variables (both manifest and latent) in the model are minimized. This means that PLS approach is less affected by small sample size and less influenced by deviations from multivariate normal distributions.
Different estimation procedures and statistical assumptions involve different output. LISREL provides several global model fit indices and examines the fit of all the parameters in the model. In PLS approach there is no overall fitting function, but non-parametric procedures as *jackknife* and *bootstrap* are used to test the significance of the estimates.
The common root to all the differences between covariance-based approach and PLS-approach is the scientific aim at the basis of the model. The basic

aim of LISREL is to describe the causal mechanism inside a system since it focuses on the relations among the variables rather than on the individuals. The objective is to confirm a theoretical model which can be inferred to the population from which the sample has been drawn. Hence, the attention is on the covariance matrix, whereas the multinormality assumption allows the ML estimates having very nice statistical properties. However, such assumption is rarely met in social science where the data are mostly collected as ordinal variables. Furthermore, LISREL presents some problems related to the non-convergence of the algorithm, the factor scores indeterminacy and improper solutions, i.e. solutions outside the admissible parameter space. PLS-PM is a more *data oriented* approach. Here, the focus is on fixed observed individuals and the estimation procedures aim to optimize the prediction of the factor scores. This is a more flexible approach with no measurement, distributional, or sample size assumptions.
Of course, there is no best model but the existing differences between the two approaches make each of them more or less appropriate for certain type of analysis. Thus, the choice of the model to be used should be based on the research objectives and the limitations imposed by the sample size and distribution assumptions.

4.2.1 Partial Least Squares Path Modeling

The PLS approach to SEM has been proposed as an alternative estimation procedure to the ML approach, mostly focused on the detection and estimation of direct effects among the variables observed on fixed individuals.
In Wold's seminal paper (Wold 1975) the main principles of *partial least squares*, for the *principal component analysis* (Wold 1966), were extended to situations with more blocks of variables. The first presentation of the PLS path modeling is given in (Wold 1979), and the algorithm is described in (Wold 1982, 1985) . An extensive review on PLS approach to SEM is given in (Chin 1998). Later, (Tenenhaus et al. 2005) have shown some PLS-PM

extensions focusing on the statistical aspects and the relations between PLS-PM and multiple table analysis.

PLS Path Modeling aims to estimate the relationships among H blocks of variables, which are expression of unobservable constructs. Specifically, PLS-PM estimates the network of relations among the manifest variables and their own latent variables, and among the latent variables inside the model. The model can be explained following three basic steps: *model specification*, *model estimation*, *model validation*.

4.2.2 Model Specification

Model specification is sitting down with all of the relevant theory which the model is based on. In this phase the researcher defines a model specifying the pattern of relations among the variables inside the system.

Formally, let us assume P variables observed on N units ($n = 1, \ldots, N$). The resulting data x_{np} are collected in a partitioned table $\mathbf{X} = [\mathbf{X_1}, \mathbf{X_h}, \ldots, \mathbf{X_H}]$, where $\mathbf{X}_h$ is the generic block.

A path diagram (fig. 4.1) gives a graphical representation of the whole model, including all the connections among the variables inside the system.

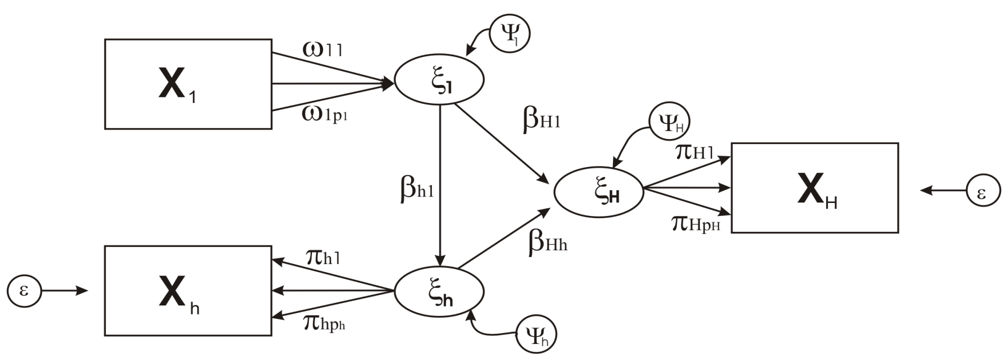

Figure 4.1: *Path model representation*

Path models adhere to certain common drawing conventions. Specifically, ellipses represent latent variables, rectangles refer to manifest variables and error terms are represented by circles. Arrows showing causations among the variables start from the independent variables pointing to the dependent ones.
In the PLS-PM literature, authors are used to distinguish causal relationships between the latent variables constitute the *structural model*, while relations between latent variables and related manifest variables define the *measurement model*. Thus, in the first step of in PLS-PM the researcher specifies the set of observed variables for each latent variable in the measurement model, and the relationships amongst the different latent variables in the structural model. Alternatively, structural and measurement model may be termed *inner model* and *outer model*, respectively. The notation used for the model is given in table (4.1). Matrices are denoted by bold upper case characters; vectors are always column vectors and denoted with bold lower case characters, where the subscript indicates to which block the vector belongs; scalars are denoted with normal lower case characters. Each block has the same number of objects, while the number of variables may be different for each block. The size of each matrix or vector is given.

The measurement model can be *reflective* or *formative* according to the linkage between the latent and the manifest variables.
In the *reflective model* each manifest variable reflects its latent variable, thus it is related to the latent variable by a simple regression:

$$\mathbf{x}_{hj} = \lambda_{hj}\xi_h + \varepsilon_{hj} \tag{4.1}$$

The error term ε_{hj} represents the imprecision in the measurement process. An example of reflective block whose items are drawn with an arrow leading away from the latent construct is given in figure (4.1) by ξ_h. The assumption behind this model is that the residual ε_{hj} has a zero mean and is uncorrelated

Table 4.1: PLSPM Notation

Symbol	**Meaning**
N $(n = 1, \dots, N)$	*Number of observations*
H $(h = 1, \dots, H)$	*Number of latent variables*
P $(p = 1, \dots, P)$	*Number of indicators*
P_h $(j = 1, \dots, P_h)$	*Number of indicators of the generic block*
$\mathbf{X}$ $(N \times P)$	*Matrix of indicators*
$\mathbf{X}_h$ $(N \times P_h)$	*Generic block of indicators*
$\mathbf{x}_{hj}$ $(N \times 1)$	*Generic j-th variable of the h-th block*
ξ_h $(N \times 1)$	*Latent variables*
$\beta_{hh'}$	*Structural parameters relating latent variables*
ε_{jh}	*Measurement errors in reflective model*
δ_h	*Measurement errors in formative model*
ψ_h	*Measurement errors in structural model*
ω_{hj}	*Weights in formative model*
π_{hj}	*Loadings in reflective model*

with the latent variable of the same block:

$$E(\mathbf{x}_{hj}|\xi_h) = \lambda_{hj}\xi_h \tag{4.2}$$

This assumption defined *predictor specification* assures desirable estimation properties in LS modeling. Furthermore, as the *reflective* block reflects the construct, it should be *unidimensional*. Hence, the set of indicators are assumed to measure the same unique underlying concept. There exist several tools for checking unidimensionality of a block:

a) *Cronbach's alpha*: a block is considered unidimensional if this index is larger than 0.7

$$\alpha = \frac{\sum_{j \neq j'} cor(\mathbf{x}_{hj}, \mathbf{x}_{hj'})}{p + \sum_{j \neq j'} cor(\mathbf{x}_{hj}, \mathbf{x}_{hj'})} \times \frac{p_h}{p_h - 1} \tag{4.3}$$

b) *Dillon-Goldstein's rho* (or *Jöreskog's*): a block is considered unidimensional if this index is larger than 0.7

$$\rho = \frac{(\sum_j \lambda_{hj})^2}{(\sum_j \lambda_{hj})^2 + (\sum_j 1 - \lambda_{hj}^2)} \tag{4.4}$$

d) *Principal component analysis of a block*: a block is considered unidimensional if the first eigenvalue of the correlation matrix is higher than 1, while the others are smaller.

In the *formative model*, each MV or each sub-blocks of MV's represents different dimensions of the underlying concept. The observed variables are not assumed to be correlated, thus there is no need for checking the block unidimensionality. In other words, the latent variable is a linear function of its manifest variables:

$$\xi_h = \sum_{j=1}^{P_h} \pi_{hj} \mathbf{x}_{hj} + \delta_h \tag{4.5}$$

The error term δ_h represents the fraction of the corresponding latent variable not accounted for by the manifest variables. The assumption behind this model is the following *predictor specification*:

$$E(\xi_h | \mathbf{x}_{hj}) = \sum_{j=1}^{P_h} \pi_{hj} \mathbf{x}_{hj} \tag{4.6}$$

An example of formative block in figure (4.1) is represented by ξ_1.
The basic algorithm has been successively extended (Lohmöller 1989) to support the MIMIC model which is a combination of the reflective and formative model.
In both *reflective* and *formative* model, the latent variables are estimated as weighted aggregates of their own MV's:

$$\xi = \omega \mathbf{X} \tag{4.7}$$

where ω is the vector of regression coefficients or loadings scaled so as to have latent variables with unitary variance.
The path coefficients (β) then come from a regular regression between the estimated latent variables.
The *structural model* describes the causations among the latent variables:

$$\xi_h = \beta_{h0} + \sum_{h'} \beta_{hh'}\xi_{h'} + \psi_h \tag{4.8}$$

where ξ_h and $\xi_{h'}$ are adjacent latent variables and $h, h' \in [1, \ldots, H]$ vary according to the model complexity. Each latent variable may be independent, dependent or both. A latent variable which is independent in the model is defined *exogenous*, whereas a dependent latent variable is called endogenous. The only constraint is to have no loop in the model, which is the main characteristic in the so-called *recursive models*.

4.2.3 Model estimation

Model estimation involves estimating each parameter specified in the model. An iterative procedure allows to estimate the latent variable scores (ξ), the outer weights ($\mathbf{w}$). The estimation procedure is named *partial* since it solves blocks one at time by means of alternating single and multiple linear regressions. The path coefficients (β) come afterwards from a regular regression between the estimated latent variables. As discussed in chapter (4.2) SEM-ML and SEM-PLS aim at different objectives. In SEM-ML, the aim is to minimize the residual covariance matrix $E(\psi\psi') = \Psi$ by reproducing the observed covariances. In SEM-PLS the aim is to minimize the trace (sum of diagonal elements=*variances*) of $E(\psi\psi') = \Psi$, of $E(\varepsilon\varepsilon') = \Theta_\varepsilon$ in case of reflective model, and $E(\delta\delta') = \Theta_\delta$ in case of formative model. That shows that PLS-PM is a variance-based model opposed to the covariance-based model. In PLS-PM algorithm the estimation of the latent variable scores are obtained through the alternation of *outer* and *inner* estimation, iterating till convergence. The procedure starts by choosing arbitrary weights w_{hj}. In the external estimation, the latent variable is estimated as a linear combination of its own MVs:

$$\mathbf{v}_h \propto \sum_{j=1}^{P_h} w_{hj}\mathbf{x}_{hj} = \mathbf{x}_h\mathbf{w}_h \tag{4.9}$$

where $\mathbf{v}_h$ is the standardized outer estimation of the latent variable ξ_h and the symbol $\propto$ means that the left side of the equation corresponds to the standardized right side. In the internal estimation, the latent variable is estimated by considering its links with the other adjacent latent variables:

$$\mathbf{z}_h \propto \sum e_{hh'}\mathbf{v}_{h'} \tag{4.10}$$

where the inner weights are equal to the signs of the correlations between $\mathbf{v}_h$ and the $\mathbf{v}_{h'}$'s connected with $\mathbf{v}_h$. These first two steps allow to update the outer weights w_{hj}. There are two options for updating the outer weights:

- *Mode A*: the weight is the regression coefficient of $\mathbf{z}_h$ in the simple regression of $\mathbf{x}_{jh}$ on the inner estimate $\mathbf{z}_h$, which corresponds to the covariance as $\mathbf{z}_h$ is standardized:

$$w_{jh} = cor(\mathbf{x}_{jh}\mathbf{z}_h) \tag{4.11}$$

- *Mode B*: the vector $\mathbf{w}_h$ of the weights w_{hj} is the regression coefficient vector in the multiple regression of $\mathbf{z}_h$ on its centered MVs:

$$\mathbf{w}_h = (\mathbf{X}_h'\mathbf{X}_h)^{-1}\mathbf{X}_h'\mathbf{z}_h \tag{4.12}$$

The choice of the external mode estimation depends on the nature of the model. For a *reflective model* the *Mode A* is more appropriate, while *Mode B* is better for the *formative model*. Furthermore, *Mode A* is suggested for en-

dogenous latent variables, while *Mode B* for the exogenous ones. Is is worth noticing, that *Mode B* is affected by multicollinearity. In such a situation, PLS regression may be used as a valuable alternative to OLS regression. The algorithm is iterated till convergence, which is demonstrated to be convergent for one and two-block models. However, for multi-block models, convergence is always verified in practice. After convergence, structural (or path) coefficients are estimated through an OLS multiple regression among the estimated latent variables (4.8). Wold's original algorithm has been further developed (Lohmöller 1987, Lohmöller 1989). In particular, new options for computing both inner and outer weights have been implemented together with a specific treatment for missing data (Tenenhaus et al. 2005).
Here, a schematic description of the original PLS-PM Wold's algorithm is given:

Algorithm 1 PLS Path Moldeling Wold's algorithm

Input: $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_h, \ldots, \mathbf{X}_H]$ standardized MV's;
Output: β_h, $\mathbf{w}_h$, ξ_h;

for all $h = 1, \ldots, H$ **do**
  initialize $\mathbf{w}_h$
  $\mathbf{v}_h \propto \sum_j w_{hj}\mathbf{x}_{hj} = \mathbf{x}_h\mathbf{w}_h$
  $e_{hh'} = sign[cor(\mathbf{v}_h, \mathbf{v}_{h'})]$
  $\mathbf{z}_h \propto \sum e_{hh'}\mathbf{v}_{h'}$
  update $\mathbf{w}_h : w_{jh} = cor(\mathbf{x}_{jh}, \mathbf{z}_h)$ or $\mathbf{w}_h = (\mathbf{X}_h'\mathbf{X}_h)^{-1}\mathbf{X}_h'\mathbf{z}_h)$
  if $\mathbf{v}_h = \mathbf{z}_h$ **then**
    $\xi_h \propto \mathbf{X}_h\mathbf{w}_h$
    $\beta_h = (\Xi'\Xi)^{-1}\Xi\xi_h$, where Ξ includes LVs connected to ξ_h
  else
    repeat till convergence
  end if
end for

4.2.4 Model Validation

Model validation involves determining how well the data fit the theoretical implied model. As described above, PLS approach lacks of a global optimization criterion so that there is no *global fitting function*. Furthermore, it is a variance-based model strongly oriented to prediction. Thus, model validation focuses on the model predictive capability. According to PLS-PM structure, each part of the model needs to be validated: *measurement model*, *structural model* and each structural equation inside the model.
The quality of the *measurement model* for each block is measured by means of the *communality* index measure:

$$Com_h = \frac{1}{P_h} \sum_{j=1}^{P_h} cor^2(\mathbf{x}_{jh}, \mathbf{v}_h) \tag{4.13}$$

This index measures how much variability of the MVs is explained by their own latent variable. That means how well the MVs describe the related LV. It is possible to measure the quality of the whole measurement model by means of the *average communality*, which is the average of all $cor^2(\mathbf{x}_{jh}, \mathbf{v}_h)$:

$$\overline{Com} = \frac{1}{P} \sum_{h=1}^{H} p_h Com_h \tag{4.14}$$

The quality of the *structural model* for each endogenous block is explained by mean of the *redundancy* index measure:

$$Red_h = Com_h \times R^2(\mathbf{v}_h, \{\mathbf{v}'_{h'}s \quad \text{explaining} \quad \mathbf{v}_h\}) \tag{4.15}$$

The *redundancy* index measures the portion of variability of the MVs connected to an endogenous latent variable explained by the latent variables indirectly connected to the block.
The quality of each structural equation is measured by a simple evaluation

of the R^2 fit index.
As aforementioned, there is no overall fit index in PLS-PM. Nevertheless, a global criterion of goodness-of-fit has been proposed (Amato, Vinzi & Tenenhaus 2005). Such index, called GoF, is the geometric mean of the average *communality* and the average R^2:

$$Gof = \sqrt{\overline{communality} \times \overline{R^2}} \tag{4.16}$$

As PLS-PM is a *soft modeling* approach with no distributional assumptions, it is possible to estimate the significance of the parameters based on cross-validation methods like jack-knife and bootstrap. It is also possible to build cross-validated version of both the *communality* and the *redundancy* fit indexes by means of a *blindfolding* procedure.

4.3 Fuzzy and PLS path modeling: *a marriage of convenience*

This section describes the original contribution of the present thesis in this chapter. The basic idea is to combine *fuzzy regression* and *PLS path modeling* through a two-stage approach for multi-block analysis. Advantages of the approach will be presented and discussed in chapter (5), dedicated to the application on a real dataset.
It is useful to remark that there exist many different fuzzy regression models, as widely discussed in chapter (3). Basically, fuzzy regression models may be classified in *Fuzzy Possibilistic Regression* (FPR) and *Fuzzy Least Squares Regression* (FLSR). The approach used in this context is the first one. More specifically, the basic FPR model for single-valued data (3.11). It is characterized by an estimation procedure based on *optimization* techniques, providing fuzzy/interval regression coefficients. There are no distributional assumptions. The estimates are obtained as solutions of a minimization problem

with the constraint that the estimated values include the observed values for a certain α-level ($0 < \alpha < 1$), called *possibility level*. The fuzziness of the coefficients represents the imprecision in estimating parameters. Hence, the objective function to be minimized is the sum of the spread coefficients. In other words, the fuzzy/interval coefficients defined in terms of minimum and maximum values measure the uncertainty in estimates, similarly to interval confidence in classical inference. It is just a different approach to uncertainty (see, chapter 2). Increasing the α-coefficient expands the fuzzy intervals as well as increasing the confidence level in statistical regression expands the confidence interval width. Notice that a *fuzzy data* boils down to an *interval data* when $\alpha = 0$, i.e. there is no information on the imprecision distribution. Furthermore, once the α *possibility level* has been selected, the corresponding fuzzy data is algebraically treated as an interval data since *fuzzy arithmetic* is *interval arithmetic* on $\alpha-$cuts (see, 2.5).
In chapter (3), FPR has been widely discussed. Strong focus is given on the comparison between FPR and OLS regression. It has pointed out that FPR may be considered a viable alternative to OLS regression when the data set is insufficient to support statistical regression analysis (strong assumption with respect to the distributions, sample size, multi-collinearity), and the human knowledge is the main source of uncertainty (Kim et al. 1996, Romano & Palumbo 2006*b*). Thus, the proposal is to introduce FPR inside PLS-PM in order to have a more flexible approach, combining the advantages of both the methodologies.
PLS-PM and FPR present many similar characteristics so that a combination between these two methodologies seems to be very appropriate. They are well suited methodologies for analyzing *phenomena* where the human judgment is influential. For instance in *consumer analysis*, where consumers give their opinions on a certain number of products and/or services. In this framework, such as in many other decision processes the major source of uncertainty is fuzziness rather than randomness (Zadeh 1973). In addition,

both PLS-PM and FPR are "soft modeling" approaches, that means there are no constraints on distributions and measurement scales.
Specifically, FPR joints PLS-PM in its final step, allowing for a *fuzzy structural model* but a still *crisp measurement model*. This connection implies a two stage estimation procedure:

- *stage* 1: latent variables are estimated according to the PLS-PM estimation procedure;

- *stage* 2: FPR on the estimated latent variables is performed so that the following *fuzzy structural model* is obtained:

$$\xi_h = \tilde{\beta}_{h0} + \sum_{h'} \tilde{\beta}_{hh'} \xi_{h'} \tag{4.17}$$

  where $\tilde{\beta}_{hh'}$ refers to the generic *fuzzy path coefficient* and h and h' vary as described in section (4.2.2).

It is worth noticing that the *structural model* from this procedure is different with respect to the traditional *structural model* presented in section (4.2.2). Here the path coefficients are fuzzy numbers and there is no error term, as a natural consequence of a FPR. As aforementioned, the error term is reflected in the model via fuzzy parameters. By definition the FPR identifies a family of lines covering the whole scatter plot, this implies that computing measures of fitness would be meaningless.

4.4 Models comparison

The analysis of complex systems, characterized by particularly heterogeneous statistical populations, leads to split the whole population into more homo-

Algorithm 2 Two stage PLS-PM & FPR algorithm

Input: $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_h, \ldots, \mathbf{X}_H]$ standardized MV's;
Output: $\tilde{\beta}_h$, $\mathbf{w}_h$, ξ_h;

for all $h = 1, \ldots, H$ **do**
    initialize $\mathbf{w}_h$
    $\mathbf{v}_h \propto \sum_j w_{hj}\mathbf{x}_{hj} = \mathbf{x}_h\mathbf{w}_h$
    $e_{hh'} = sign[cor(\mathbf{v}_h, \mathbf{v}_{h'})]$
    $\mathbf{z}_h \propto \sum e_{hh'}\mathbf{v}_{h'}$
    update $\mathbf{w}_h : w_{jh} = cor(\mathbf{x}_{jh}, \mathbf{z}_h)$ or $\mathbf{w}_h = (\mathbf{X}'_h\mathbf{X}_h)^{-1}\mathbf{X}'_h\mathbf{z}_h$
    if $\mathbf{v}_h = \mathbf{z}_h$ **then**
        $\xi_h \propto \mathbf{X}_h\mathbf{w}_h$
        $\xi_h = \tilde{\beta}_{h0} + \sum_{h'} \tilde{\beta}_{hh'}\xi_{h'} + \psi_h$
        fuzzy parameters $\tilde{\beta}_h$ = minimize $(|\Xi|\varpi)'(|\Xi|\varpi)$ subject to:

$$\begin{cases} \boldsymbol{\Xi} c - |\boldsymbol{\Xi}|\varpi & \leq \xi_{\mathbf{h}} \\ \boldsymbol{\Xi} c + |\boldsymbol{\Xi}|\varpi & \geq \xi_{\mathbf{h}} \\ \varpi & \geq 0 \end{cases}$$

c and ϖ vectors of centers and spreads of $\tilde{\beta}_h$
    else
        repeat till convergence
    end if
end for

geneous groups or segments.
Formally, let us assume the N units ($n = 1, \ldots, N$) are divided into G groups. Different groups may include a different number of observations N_g. The resulting data x^g_{np} are collected in a multiple table, where $n_g = 1, \ldots, N_g$, $p = 1, \ldots, P$ and the generic term $g = 1, \ldots, G$ indicates the table corresponding to the g–*th* group.
Like in classical inferential problems, sampling from heterogeneous populations, the stratified sampling is preferred to the random sampling. The segments are based on some predetermined criteria such as geographic location, size or any demographic characteristic. It is important the segments are as heterogeneous as possible according to the predetermined criterion. For instance, let the satisfaction of a hotel guests has to be determined. Knowing that the business clientele behaves quite differently from the leisure guests, you might want to separate them into different groups or strata.
Population segmentation leads to estimate the same model as many times as the segments identified into the target population. Several approaches have been proposed to compare the sub-populations. As discussed in section (4.1), the comparison will be based on the estimated parameters (Clogg et al. 1995). However, estimated models assess the relation structures in different proportions; in fact the residual component can vary with respect to the different models. It is important to stress that comparing models in such a way could lead to biased results. Consider the simple linear regression analysis. Specifically, consider two models with equal parameters (slope and location). Such models should be considered statistically equivalent, considering the approach based on the parameters comparison. Figure (4.2) shows similar linear regressions in case [a] and [b]. However, such models could have a different fit, as shown in figure (4.2) case [c].
In the analysis of a statistical model one should always, in one way or another, take into account the goodness of fit, above all in comparing different models among them. The proposal is then to use the FPR. The estimation

of fuzzy parameters, instead of single-valued (crisp) parameters, permits to gather both the structural and the residual information (Tanaka et al. 1982). In fact, FPR embeds the residual in the model via fuzzy parameters allowing a full comparison among the models.

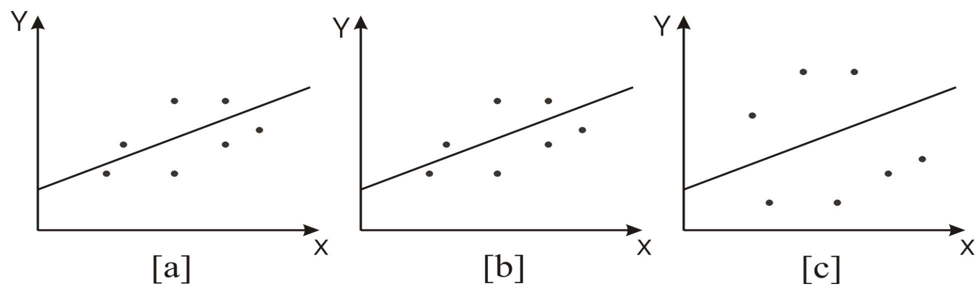

Figure 4.2: *Comparison between simple linear regressions*

4.4.1 Comparing PLS-PM models

In the specific framework of SEM, Hensler and Fassott (2007) consider the model comparison problem as a special case of *moderating effects*. Moderating effects (also called interaction effects) arise when some variables influence a direct effect between the latent variables inside the model. In particular if the *moderator variable* is categorical, it becomes a grouping variable involving group comparisons, i. e. comparisons of model estimates for different group of observations. A simple model with a moderating effect (d) is shown in figure (4.3), where it is symbolized by an arrow pointing to the direct relationship (b) between two latent variables.

Once the observations are grouped according to the moderator variable, the strategy is then to estimate local models with direct effects for each group and looking for differences in path coefficients across groups. At this aim,

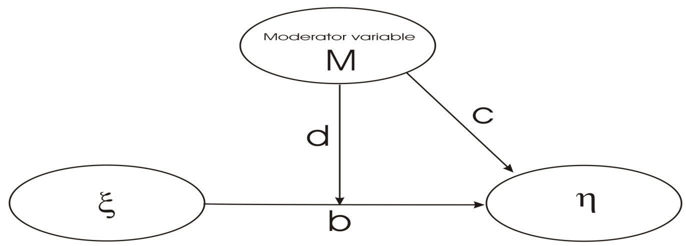

Figure 4.3: *A simple model with a moderating effect*

non-parametric approaches may be used to test for different path coefficients among groups (Chin & Dibbern 2007).

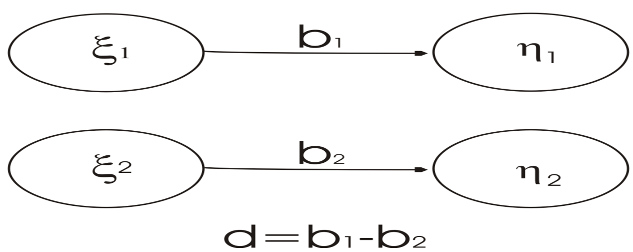

Figure 4.4: *Detecting a moderating effect (d) through group comparison*

If there is no difference between the parameters from the different models then there is no reason for considering local models. In other words, a global model is effective for the whole population. If there exist differences between parameters, these are evaluated as differences between local models.
It should be noticed how the moderating effects only concern the *structural models*. That is equivalent to assume the differences among *measurement models* to be not significant.
Under this hypothesis, in this work, model differences are gathered comparing

the *fuzzy structural parameters* in terms of distances.
The strategy consists in three basic steps:

a) estimate local fuzzy structural models for each group

$$\xi_h^g = \tilde{\beta}_{h0}^g + \sum_{h'} \tilde{\beta}_{hh'}^g \xi_{h'}^g$$

b) gather model differences comparing the related fuzzy path coefficients

Table 4.2: Data matrix

	$\tilde{\beta}_{h,1}$	$\dots$	$\tilde{\beta}_{H,1}$	$\dots$	$\tilde{\beta}_{H,h}$
mod(1)	$\tilde{b}_{h,1}$	$\dots$	$\tilde{b}_{H,1}$	$\dots$	$\tilde{b}_{H,h}$
$\dots$	$\dots$		$\dots$		$\dots$
mod(g)	$\tilde{b}_{h,1}$	$\dots$	$\tilde{b}_{H,1}$	$\dots$	$\tilde{b}_{H,h}$
$\dots$	$\dots$		$\dots$		$\dots$
mod(G)	$\tilde{b}_{h,1}$	$\dots$	$\tilde{b}_{H,1}$	$\dots$	$\tilde{b}_{H,h}$

c) clustering for fuzzy/interval data to produce clusters of models.

The G estimated fuzzy structural models are characterized by fuzzy path coefficients. That means there are no residual terms, because in the fuzzy model the error terms are embedded in the parameters themselves. This peculiarity confers to the G fuzzy structural models the same explicative power, making the comparison, that is based on the estimated fuzzy path coefficients, meaningful.

4.4.2 Fuzzy clustering for fuzzy PLS-PM

In order to compare the groups with respect to the identified fuzzy structural parameters this section introduces a fuzzy classification algorithm for interval data. The choice of a fuzzy classification algorithm is consistent with the whole strategy approach, that is based on imprecise data.

In particular the HIPYR procedure (Brito 2000) has been used to obtain a data clustering. The procedure is implemented in the SODAS© software (Symbolic Official Data Analysis System, rel. 2.5) and is designed to cluster symbolic datasets (Bock & Diday 2000). It is worth noticing that fuzzy data and interval data can also be defined as special cases of symbolic data, only characterized by continuous interval valued variables (Bock & Diday 2000).
HIPYR clustering procedure determines the $G(G-1)/2$ distances between models by the Hausdorff metric in $\mathbb{R}^p$ (see, 2.36) (Neumaier 1990). This allows to appreciate the differences into the $\mathbb{R}^p$ parameters space.
HIPYR algorithm provides both hierarchical and pyramidal classification structures: hierarchical classification leads to disjunctive partition, pyramidal classification determines fuzzy clusters. In the following, only pyramidal clustering will be taken into account.
The HIPYR algorithm is a *bottom up* procedure that allows to cluster a set of objects $E = \{1, \dots, N\}$ characterized by P symbolic variables. The *pyramidal model* is a generalization of the hierarchical model in which non-disjoint classes are allowed at each given level:

1. a pyramid is a family $\{P = A, B, \dots\}$ of non-empty subsets or classes $A, B, \dots \subseteq E = \{1, \dots, N\}$ such that:

    a) the set of objects belongs to P
    $E \in P$

    b) all N singletons belong to P
    $[\{1\}, \{2\}, \dots, \{n\}, \dots, \{N\}] \in P$

    c) the intersection of two classes may be empty or belongs to P
    $A \cap B = \emptyset$ or $A \cap B \in P$

    d) there exists a linear order $\leq$ on E such that each class A of P is an *interval* of $(E, \leq)$
    $A = [\alpha, \beta] := \{k | k \in E, \alpha \leq k \leq \beta\}, [\alpha, \beta] \in E$

2. a pyramid is defined *indexed pyramid* or *pyramidal dendrogram* (P, h) if for each class $A \in P$ an index $h(A) \geq 0$ is defined so that $h(A) \geq h(B)$ if $A \subset B$

3. an *indexed pyramid* is *indexed in the broad sense* if for all $A, B \in P$ with $A \subsetneqq B$ and $h(A) = h(B)$ imply the existence of two classes $C, D \in P$ with $A = C \cup D$ and $C \neq A, D \neq A$

Here follows a concise and schematic description of the procedure.
In the initial step, there are singleton clusters C_n $(n = 1, \ldots, N)$. At each step a new cluster C_{n+1} is formed by merging already constructed clusters. The clusters are merged together if: a) they have not been aggregated twice in former steps; b) there exists a total linear order $\leq$ on E so that $C_{n+1} = C_1 \cup C_2$ is an interval with respect to $\leq$.
Among all possible pairs (C_1, C_2) satisfying such conditions, it is chosen the one with the smallest value of $G(s)$, where $G(s)$ is a numerical criterion named *generality degree*. The algorithm stops when $C_{n+1} = E$.

Chapter 5

Application in Customer Satisfaction Analysis

5.1 Introduction

The present chapter shows some applications of the methodologies described in the previous chapters on data from *customer satisfaction analysis*. As is well known, *customer satisfaction* (CS) is a variable hard to *define* and *measure*.

Besides the established definition in *absolute* terms, some researchers have proposed to alternatively define the CS in *relative* terms. For instance, the CS may be defined as difference between *consumer's expectations* and *perceptions* on quality of attributes characterizing products/services (Zeithaml, Parasuraman & Berry 1991). However, the analysis of such differences does not take into account the two-dimensional nature of CS. A new codification under the perspective of *Interval Analysis* as been proposed firstly by (Lauro, Esposito, Vinzi & Scepi 2001) and successively by (Grassia et al. 2004) and (Amato & Palumbo 2004). Specifically, the two components *consumer's expectations* and *consumer's perceptions*, respectively, are combined into a unique numerical structure: the *interval data*. Within this framework, the first application consists in combining *fuzzy possibilistic regression* and *interval codification* for analyzing the satisfaction in its own nature of *interval data* (Romano & Palumbo 2005).

Another problem is how to effectively measure the CS. One of the most widely used approaches is the latent variable approach based on structural equation modeling. Here, the CS is considered a *latent concept* measured by means of a set of indicators. Thus, a satisfaction index is obtained through a deep analysis of the causality relations between the latent variables that are the components of the customer's satisfaction and the the manifest variables representing the customer's answers to the questions concerning their satisfaction. In particular, the PLS approach to SEM is preferred for measuring CS as it is more oriented to predict the latent variables, rather than the ML approach more oriented to confirm the theory of the customer's decision process. In this context, an application of the methodology presented in chapter (4) based on the fuzzy approach to PLS-PM is shown. Introducing FPR inside PLS-PM allows to take into account both the vagueness connected with the use of linguistic terms in describing the real world and the imprecision in measuring the empirical phenomena. In addition such a methodology permits to face the problem of model comparison in presence of multi-group data structure.

5.2 Data description

The present application is based on a data set used to estimate the customer satisfaction of a service industry. The data contains 23 variables observed on 366 units. The variables, assessed on continuous scales anchored in 1 and 10, are grouped in 6 blocks: *perceived quality* (7 manifest variables), *expectations* (4 manifest variables), *perceived value* (3 manifest variables), *satisfaction index* (3 manifest variables), *image* (3 manifest variables) and *loyalty* (3 manifest variables). According to the moderator variable *sector of* *activity*, the statistical units are groped into 8 classes (labels are indicated in round brackets): hospital (Hosp), health authority (HeAu), school (Scho), university (Univ), local administration (LoAd), Red Cross (RCro), social security (SoSe), public administration (PuAd). Variable names cannot be revealed because confidential constraints on the data. The final data table is a matrix (336×23) partitioned in 6 columns (variables blocks) and 8 rows (classes).

5.3 FPR for Customer Satisfaction

The basic aim of this analysis is to apply fuzzy possibilistic regression for the CS estimation. For this application only a small sample of 50 units is considered. The main assumption of the model is that the CS depends on the *expected* and *perceived quality*. Thus, the first step of the analysis is to estimate these latent variables starting from their own manifest variables. The estimation is pursued by means of a *Principal Component Analysis* on the two respective blocks of indicators. Each block is unidimensional, as the variance explained by the first principal components of the two blocks is 85.2% and 50.6%, respectively (see figure 5.1). Therefore, the estimated latent variable of each block corresponds to the first principal component of the related block. The second step is the codification of the dependent variable *overall satisfaction*. Specifically, using the interval codification proposed by Amato and Palumbo (2004), the interval valued variable *satisfaction interval* is built synthesizing the two components *perceived satisfaction* (P) and *expected satisfaction* (E). The authors consider the *perceived satisfaction* as midpoint of the *satisfaction interval* ($P - E$), whereas the spread depends on both the *satisfaction interval* itself and the maximum observed gap ($\max([L_u - E; E - L_l])$):

$$r = \frac{|E - P|}{\max[L_u - E; E - L_l]}$$

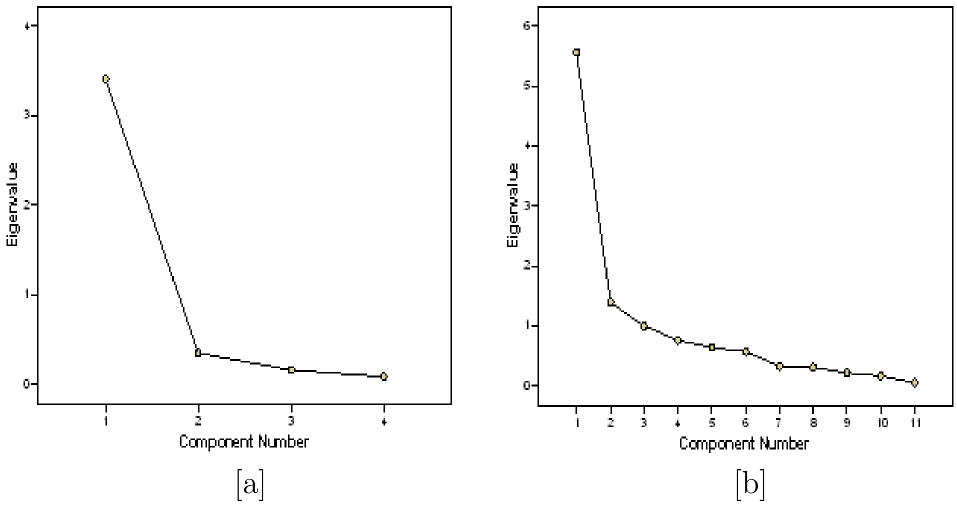

Figure 5.1: **a)** *Expected quality* scree plot; **b)** *Perceived quality* scree plot

where L_u and L_l correspond to the upper and the lower values of the scale. A FPR for crisp input and fuzzy output is performed. The estimation of the *satisfaction interval* is obtained as solution of the *linear programming problem* in (3.24) with constraints (3.25 and 3.26). The estimated model with a *possibility level* selected to $\alpha = 0.5$ follows:

$$Y = \{0.59; 2.47\} + \{0.23; 0.10\}X_1 + \{0.60; 0.05\}X_2 \tag{5.1}$$

where X_1 and X_2 are the principal component *expected* and *perceived quality*, respectively. The coefficients are symmetrical triangular fuzzy numbers expressed in terms of center and spread. Considering the center coefficients, the model shows that the *satisfaction interval* strongly depends on the *perceived quality*. On the other hand, the vagueness of the system is mstly explained by the *expected quality*, whose coefficient has a wider spread. Such a result, highlights how the customer is more ambiguous in expressing his expectations rather than its perceptions. These results are consistent with the model. In fact, it is more plausible to consider that the CS depends on the perceptions rather than on the expectations, and that assessment imprecision is higher when the customer expresses his expectations rather than his perceptions. Figures (5.2) and (5.3) show the observed *satisfaction intervals* and the regression intervals. The symbols "•" and "×" indicate the

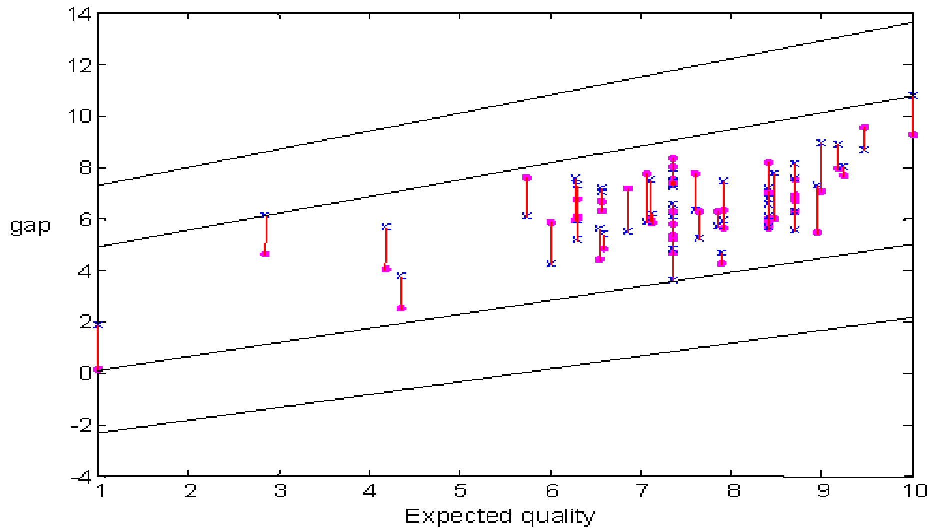

Figure 5.2: Interval regression for *expected quality*

expected and the *perceived satisfaction*, respectively. Such a representation allows to identify positive and negative intervals. In fact, the customer is satisfied if his perceptions are higher than his expectations ($P - E > 0$), otherwise he is unsatisfied ($P - E < 0$). The customer is indifferent when he has same expectations and perceptions ($P - E = 0$). Both figures represent two intervals regression. The tighter one corresponds to $\alpha = 0$, while the wider interval correspond to $\alpha = 0.5$. As discussed in section (3.3.1), the degree of possibility α is a precise but subjective measure that depends on the context. Furthermore, the α–coefficient expands the fuzzy intervals as well as increasing the confidence level in statistical regression expands the confidence interval width. The representation of the results allows to immediately perceive the intensity of the relation between the CS and its drivers

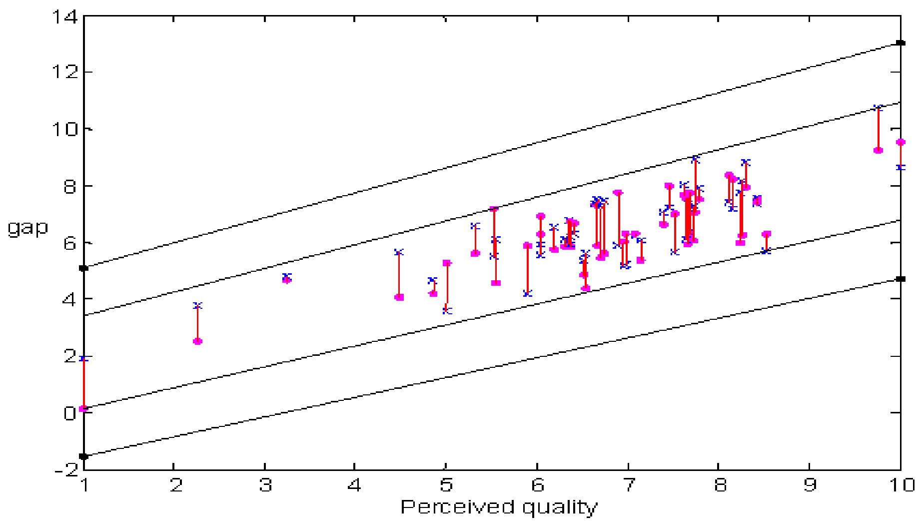

Figure 5.3: Interval regression for *perceived quality*

looking at the slope of the intervals regression. On the other hand, spread and versus of the intervals show the satisfaction level in *relative* terms.
This application shows the potentiality of FPR in *Customer Satisfaction Analysis*. However, a full evaluation of CS needs a deep analysis of the whole pattern of relations among the satisfaction drivers. At this aim, PLS path modeling is a well suited methodology. An example of fuzzy approach to PLS-PM is given in the next section.

5.4 Classification of SEM

A typical application for SEM is the estimation of the *customer satisfaction*. Within this framework, a widely adopted model is the one specified for the European Customer Satisfaction Index (ECSI) (Tenenhaus et al., 2005), where the satisfaction is estimated using the PLS approach.
ECSI model allows to estimate latent variables from their respective mani-

fest variables and to build individual indexes of satisfaction (CSI). The global model contains the following latent variables: *perceived quality*, *expectations*, *perceived value*, *satisfaction index*, *image*, *loyalty* and *complaints*. In particular, the *customer satisfaction* is explained by the drivers *perceived quality*, *expectations*, *perceived value* and *image*.
The multi-group data structure suggests significantly different models. Specifically, 8 local models have been estimated according to the procedure introduced in section 4.3. The generic model is represented in figure (5.4).

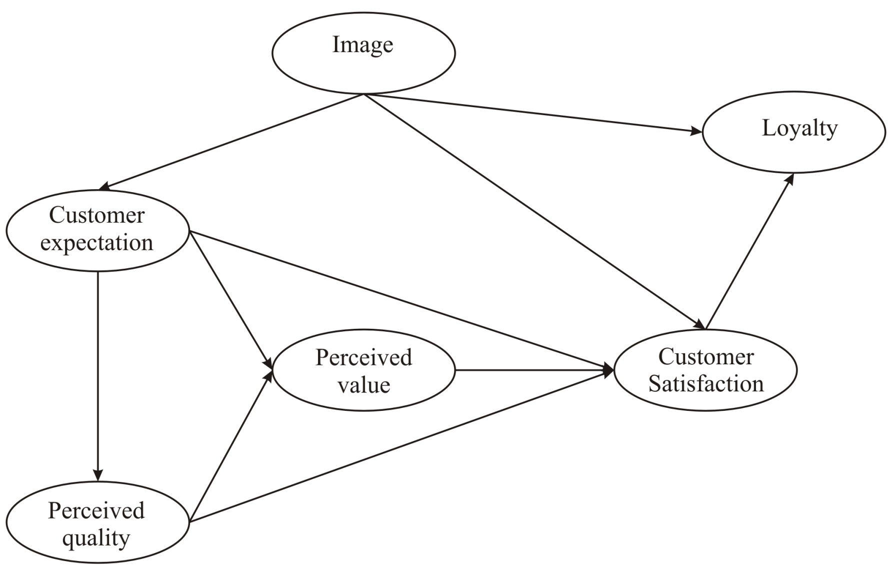

Figure 5.4: Generic local model

Differently from the classical PLS-PM, the approach proposed in this work produces fuzzy path coefficients. The five fuzzy structural equations corresponding to figure (5.4) may be written as follows:

$$
\begin{aligned}
\text{customer expectation} &= \tilde{\beta}_{21}\text{image} \\
\text{perceived quality} &= \tilde{\beta}_{32}\text{customer expectation} \\
\text{perceived value} &= \tilde{\beta}_{42}\text{customer expectation} + \tilde{\beta}_{43}\text{perceived quality} \\
\text{CSI} &= \tilde{\beta}_{51}\text{image} + \tilde{\beta}_{52}\text{customer expectation} \\
&+ \tilde{\beta}_{53}\text{perceived quality} + \tilde{\beta}_{54}\text{perceived value} \\
\text{loyalty} &= \tilde{\beta}_{61}\text{image} + \tilde{\beta}_{65}\text{CSI}
\end{aligned}
$$

The eight estimated models are compared on the basis of their fuzzy path coefficients (see table 5.1).

Table 5.1: Data matrix of fuzzy structural parameters

	$\tilde{\beta}_{2,1}$	$\tilde{\beta}_{3,2}$	$\tilde{\beta}_{4,2}$	$\tilde{\beta}_{4,3}$	$\tilde{\beta}_{5,1}$	$\tilde{\beta}_{5,2}$	$\tilde{\beta}_{5,3}$	$\tilde{\beta}_{5,4}$	$\tilde{\beta}_{6,1}$	$\tilde{\beta}_{6,5}$
Hosp	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
HeAu	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Scho	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Univ	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
LoAd	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
RCro	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
SoSe	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
PuAd	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

In other words, the distances among the different local models are considered for a pyramidal classification procedure. The results are shown in figure (5.5). Looking at the figure and taking into account the analytical results, it appears quite clear that there are two well separated groups. However, the fuzzy clustering does not produce disjoint clusters so that it is more consistent with the whole methodology. Here are the obtained fuzzy clusters:

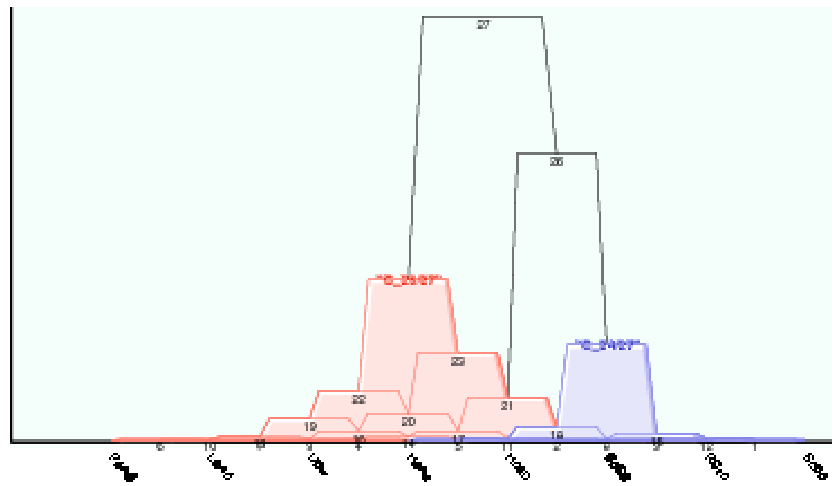

Figure 5.5: Pyramid

Cluster 1 [Scho, Hosp, SoSe, RCro, HeAu]

Cluster 2 [Hosp, SoSe, RCro, HeAu, Univ, LoAd, PaAd]

Notice that Cluster 1 is represented by the blue shadowed part of the pyramid in figure (5.5), whereas the Cluster 2 refers to the red shadowed part of the pyramid. As a fuzzy clustering has been adopted, there is a cluster overlapping: Hosp, SoSe, RCro, HeAu.
Results are consistent with other studies performed on the same data. It is quite evident that Cluster 1 is mainly characterized by public bodies having autonomy of expenditure. On the contrary Cluster 2 is mainly characterized by Local and Central administrations. Notice that these two later groups only appear in the Cluster 2.
This application has shown the prospective combination of Partial Least Squares Path-Modeling (PLS-PM) and Fuzzy Possibilistic Regression (FPR).

Theoretical and empirical results are very encouraging and suggest several further research directions.
The proposed strategy should be compared with alternative and similar approaches. However, the proposed method has some specific peculiarities that would make difficult comparisons with more traditional approaches.
There are some important and profitable aspects of the procedure it is worth noticing and that make the approach particularly suitable in some application domains. For example, let us think to sensory analysis and customer satisfaction: where any distributional hypothesis is satisfied and, at the same time, it is of interest to make comparisons among models estimated with respect to different populations or estimated on the same population over the time.
As regards the model comparison, the next step ahead will be devoted to extend the FPR to the measurement model. This will permit to consider the whole path model in the comparison phase. Another important research line is the validation of the proposed methodology.

Chapter 6

Guidelines of future research

There are many aspects of the proposed methodologies that require further developments. They involve both the FPR as well as the combined two-stage approach for multi-block analysis.
As regards the *fuzzy possibilistic regression*, future research lines concern the theoretical proofs of the main presented properties, the development of procedures for selecting the best set of variables as well as new methods for handling outliers.
On the other hand, it would be useful to extend the two steps algorithm combining FPR and PLS-PM to the whole model. At the same time, there is a need of a proper interpretation of the related results by means of appropriate goodness of fit indexes. Another important extension of the proposed methodology is devoted to consider also fuzzy input.

6.1 Insight of fuzzy possibilistic regression

Although the optimization problem implemented to determine the fuzzy parameters in FPR has always feasible solutions there is no proof of their uniqueness. Empirical results have shown the existence of a unique optimal solution. However, there is a need of a theoretical proof of such conclusion.

Probably, the proof is easier for the *quadratic programming problem*, rather than for the *linear* one, since in a quadratic objective function the *local minimum* always corresponds to the *global minimum*.
Another important research line which is now on-going concerns the study of the fuzzy estimators properties. Empirically such estimators are shown to be unbiased and robust to multi-collinearity, anyway further analysis are required.
A very important aspect to be improved in FPR is the interpretation of the results. There are already some contributions in this direction, but they focus on FPR for crisp input. Thus, it would be useful a proper extension to FPR for fuzzy input/output.
Finally, *ad hoc* procedures for selecting the best subset of variables, as well as new methods for handling outliers would be useful to improve the performance of FPR.

6.2 Extensions of fuzzy approach to PLS-PM

The fuzzy approach to PLS-PM is a very innovative methodology so that it provides many research perspectives.
First of all, this is a partial fuzzy approach since it provides a fuzzy *structural model* but a crisp *measurement model*. Thus the first research line regards the extension of the fuzzy approach to the whole model in order to have a *Fuzzy PLS Path Modeling* (FPL-PM). This leads to introduce FPR in any step of the PLS-PM algorithm. According to each step different FPR should be used: FPR for crisp input, for crisp input and fuzzy output, and for fuzzy input/output.
Another important extension of such a methodology, very useful for application purposes, is the fuzzy/interval data codification. In other words, the data to be processed may be *fuzzy data*, rather than single valued data. As it has been widely discussed, fuzzy approach is mostly appropriate in con-

text where human judgment is influential. This aspect has been outlined by the first application presented in chapter (5). However, in this application customers' preferences were expressed in terms of single values. In order to better capture the imprecision in expressing subjective preferences it would be interesting to adopt a fuzzy data coding. In other words, the consumer's assessment of a given product/service may be expressed in linguistic terms such as worst, poor, fair, good, best, each one of them associated to a fuzzy number.

On the other hand, interval coding may be used for synthesizing information in multi-way data. For instance, in Sensory Analysis a panel of assessors score blocks of sensory attributes for profiling products, thus yielding a three-way table crossing assessors, attributes and products. In this context, it is important to synthesize the scores into a global assessment to investigate differences between products. A number of different techniques have already been proposed to find a *consensus profile* for all the assessors. The simplest averaging over the assessors and more complex techniques as Three-way Factor Analysis, Generalized Procrustes Analysis and Generalized Canonical Analysis. The former approach does not take into account the variability between assessor's scores, whereas the latter provides results not always easy to interpret.

In this framework, an *ad hoc* interval coding may be used to collapse the tables over the assessors into a two-way table partitioned by the attributes. Then a fuzzy PLS path modeling would provide two sets of synthesized assessments: the overall latent scores for each product and the partial latent scores for the different blocks of attributes.

Appendix

.1 Simplex algorithm

The simplex algorithm for solving the LP problem in (3.16) is described:

Algorithm 3 Simplex

Input: A, $\mathbf{b}$;
Output: $\mathbf{x}$;

initialize B (basis matrix) and x^0 (basic feasible solution)
compute the vector $\bar{\mathbf{c}}$
if $\bar{\mathbf{c}} \geq 0$ **then**
    x^0 is optimal
else
    choose m for which $\mathbf{c}_m < 0$
    compute $u = B^{-1}A_m$
    if $u \leq 0$ **then**
        $\theta^* = \infty$ and the LP is unbounded
    else
        $\theta^* = \min_{l=1,\ldots,L:u_l \geq 0} \frac{x^0_{B_l}}{u_l}$
    end if
end if
choose g such that $\frac{x^0_{B_g}}{u_g}$ and form a new basis replacing $A_{B(g)}$ with A_m
the new basic variables are $x^1_m = \theta^*$ and $x^1_{B_l} = x^0_{B_l} - \theta^* u_l$ if $l \neq g$

Bibliography

Achen, C. H. (1982), *Interpreting and Using Regression*, Sage Publications, California.

Alefeld, G. & Herzerberger, J. (1983), *Introduction to Interval computation*, Academic Press, New York.

Alwin, D. F. & Hauser, R. M. (1975), 'The decomposition of effects in path', *American Sociological Review* **40**, 36–47.

Amato, S. & Palumbo, F. (2004), 'Multidimensional gap analysis', *Statistica Applicata* **16**(3), 265–283.

Amato, S., Vinzi, V. E. & Tenenhaus, M. (2005), A global goodness-of-fit index for pls structural equation modeling, Technical report, HEC School of Managment, France.

Bellman, R. E. & Zadeh, L. A. (1970), 'Decision making in a fuzzy envoirment', *Managment Science* **17**(B), 141–164.

Black, M. (1937), 'Vagueness: An excercise in logical analysis', *Philosophy of Science* **4**, 427–455.

Bock, H. & Diday, E. (2000), *Analysis of Symbolic Data, Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer, Heidelberg.

Bollen, K. A. (1989), *Structural equations with latent variables*, Wiley, New York.

Boyd, S. & Vandenberghe, L. (2004), *Convex Optimization*, Cambridge University Press, United Kingdom.

Bozdogan, H. (2004), *Intelligent Statistical Data Mining with Information Complexity and Genetic Algorithms*, Statistical Data Mining and Knowledge Discovery, CRC Press, chapter 2, pp. 15–56.

Braun, D., Mayberry, J., Powers, A. & Schlicker, S. (2003), The geometry of the Hausdorff metric. Available at: `http://faculty.gvsu.edu/schlicks/Hausdorff_ paper.pdf`.

Brito, P. (2000), *Hierarchical and Pyramidal Clustering with Complete Symbolic Objects*, Vol. Analysis of Symbolic Data of *Studies in Classification, Data Analysis and Knowledge Organization*, Springer, Heidelberg, chapter 11.3, pp. 312–341.

Celmins, A. (1987), 'Least squares model fitting to fuzzy vector data', *Fuzzy sets and Systems* **22**(3), 245–269.

Chen, Y. S. (2001), 'Outlier detection and confidence interval modification in fuzzy regression', *Fuzzy Sets and Systems* **119**, 259–272.

Chin, W. W. (1998), The partial least squares approach for structural equation modeling, *in* G. A. Marcoulides, ed., 'Modern Methods for Business Research', Lawrence Erlbaum Associates, London, pp. 295–236.

Chin, W. W. & Dibbern, J. (2007), A permutation based procedure for multi–group pls analysis: Results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between germany and the usa, *in* V. E. Vinzi, W. W. Chin, J. Henseler & H. Wang, eds, 'Handbook of Partial Least Squares - concepts, methods and applications', Springer.

Chvátal, V. (1983), *Linear Programming*, W. H. Freeman and Company, New York.

Clogg, C. C., Petkova, E. & Haritou, A. (1995), 'Statistical methods for comparing regression coefficients between models', *The American Journal of Sociology* **100**(5), 1261–1293.

Coleman, T. F. & Li, Y. (1996), 'A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables', *SIAM Journal on Optimization* **6**(4), 1040–1058.

Coppi, R., D'Urso, P., Giordani, P. & Santoro, A. (2006), 'Least squares estimation of a linear regression model with lr fuzzy response', *Computational Statistics and Data Analysis* **51**(1), 267–286.

Cover, T. M., Gacs, P. & Gray, R. M. (1989), 'Kolnogorov's contribution to information theory and algorithmic complexity', *Ann. Prob.* **17**, 840–865.

Diamond, P. (1988), 'Fuzzy least squares', *Information Science* **46**(3), 141–157.

Djkstra, T. (1983), 'Some comments on maximum likelihood and partial least squares methods', *Journal of Econometrics* **22**, 67–90.

Dubois, D. & Prade, H. (1980), *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York.

Dubois, D. & Prade, H. (2005), Interval-valued fuzzy sets, possibility theory and imprecise probability, *in* 'Proc.of International Conference in Fuzzy Logic and Technology (EUSFLAT'05)', Barcellona.

Fedrizzi, M. & Kacprzyk, J. (1992), *Brief inrtoduction to fuzzy sets and fuzzy arithmetic*, Vol. Fuzzy Regression Analysis of *Studies in Fuzziness*, Omnitech Press and Physica Verlag, chapter 1, pp. 14–20.

Ferson, S., Ginzburg, L., Kreinovoch, V., Longpré, L. & Aviles, M. (2002), 'Computing variance for interval data is np-hard', *ACM SIGACT News* **33**(2), 108–118.

Fornell, C. & Bookstein, F. L. (1982), 'Two structural equation models: Lisrel and pls appliead to consumer exit -voice theory', *Journal of Marketing Research* **XIX**, 440–452.

Gill, P. E., Murray, W. & Wright, M. H. (1981), *Practical Optimization*, Academic Press, New York.

Grassia, M. G., Lauro, N. C. & Scepi, G. (2004), L'analisi dei dati ad intervallo nell'ambito della qualità, *in* C. Davino & N. C. Lauro, eds, 'Data Mining e Analisi Simbolica', Franco Angeli, Milano, pp. 65–79.

Gujarati, D. N. (2003), *Basic Econometrics*, McGraw-Hill, New York.

Gupta, M. M., N.Saridis, G. & Gaines, B. R. (1977), *Fuzzy Automata and Decision Processes*, North-Holland Publ., Amsterdam.

Gusev, L. A. & Smirnova, I. M. (1973), 'Fuzzy sets: Theory and applications (a survey)', *Autom. Remote Control (USSR)* **5**, 66–85.

Hensler, J. & Fassott, G. (2007), Testing moderating effects in pls path models: An illustration of available procedure, *in* V. E. Vinzi, W. W. Chin, J. Henseler & H. Wang, eds, 'Handbook of Partial Least Squares - concepts, methods and applications', Springer.

Hirota, K. (1977), Concepts of probabilistic sets, *in* 'Proc. IEEE Conf. Decision Control', New Orleans, pp. 1361–1366.

Hoerl, A. E. & Kennard, R. W. (1970), 'Ridge regression: biased estimation of non-orthogonal components', *Technometrics* **12**, 55–67.

Indahl, U. G. (1998), Pattern Recognition and Feature Extraction- A Synthesis of Probabilistic and Fuzzy approaches, Phd thesis, Univesity of Oslo, Oslo.

Jöreskog, K. (1970), 'A general method for analysis of covariance structure', *Biometrika* **57**, 239–251.

Jöreskog, K. & Wold, H. (1982), The ml and pls techniques for modeling with latent variables: historical and comparative aspects, *in* K. Joreskog & H. Wold, eds, 'Systems Under Indirect Observation', Vol. Part I, North-Holland, Amsterdam, pp. 263–270.

Kandel, A. & Byatt, W. J. (1978), 'Fuzzy sets, fuzzy algebra and fuzzy statistics', *Proc. IEEE* **68**, 1619–1639.

Kao, C. & Chyu, C. L. (2002), 'A fuzzy linear regression model with better explanatory power', *Fuzzy Sets and Systems* **126**, 401–409.

Kaplan, D. (2000), *Structural Equation Modeling: Foundations and Extensions*, Sage Publications Inc., Thousands Oaks, California.

Kaufmann, A. & Gupta, M. M. (1985), *Introduction to Fuzzy Arithmetic-Theory and Applications*, Van Norstrand, Reinhold.

Kim, K. J. & Chen, H.-R. (1997), 'A comparison of fuzzy and nonparametric linear regression', *Computers Operations Research* **24**(6), 505–519.

Kim, K. J., Moskowitz, H. & Koksalan, D. (1996), 'Fuzzy versus statistical linear regression', *European J. Oper. Res.* **92**, 417–434.

Klir, G. J. & Yuan, B. (1995), *Fuzzy Sets and Fuzzy Logic, Theory and Applications*, Prentice Hall PTR, New Jersey.

Kreinovich, V., Lakejev, A., Rohn, J. & Kahl, P. (1997), *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht.

Krätschmer, V. (2001), 'A unified approach to fuzzy random variables', *Fuzzy Sets and Systems* **123**, 1–9.

Lauro, C. N., Esposito, V., Vinzi & Scepi, G. (2001), Viasualizzazione e sintesi della customer satisfaction in termini di analisi dei dati simbolici, *in* 'Processi e metodi statistici di valutazione', Vol. sessioni plenarie, specializzate, satellite, Società Italiana di Statistica, Roma, pp. 129–136.

Lee, S. Y. & Song, X. Y. (2001), 'Hypothesis testing and model comparison in two-level structural equation models', *Multivariate Behavioral Research* **36**(4), 639–655.

Lohmöller, J. (1987), Lvpls program manual, version 1.8, Technical report, Zentralarchiv für Empirische Sozialforschung, Universität Zu Köln, Köln.

Lohmöller, J. (1989), *Latent variable path modeling with partial least squares*, Physica-Verlag, Heildelberg.

Materón, G. (1975), *Random Sets and Integral Geometry*, Wiley, New York.

Ming, M., Friedman, M. & Kandel, A. (1997), 'General fuzzy least squares', *Fuzzy Sets and Systems* **88**, 107–118.

Miranda, E., Couso, I. & Gil, P. (2005), 'Random intervals as a model for imprecise information', *Fuzzy Sets and Systems* **154**, 386–412.

Moore, R. E. (1966), *Interval Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

Moore, R. E. (1979), *Methods and Applications of Interval Analysis*, Society for Industrial and Applied Mathematics, Philadelphia.

Myung, J. & Pitt, M. A. (2003), Model comparison methods, *in* 'Numerical Computer Methods', Vol. D, L. Brand and M. L. Johnson.

Nasrabadi, M. M. & Nasrabadi, E. (2004), 'A mathematical-programming approach to fuzzy linear regression analysis', *Applied Mathematics and Computation* **155**, 873–881.

Negoita, C. V. & Ralescu, D. A. (1975), *Applications of Fuzzy Sets to Systems Analysis*, Birkhaeuser, Basel.

Neumaier, A. (1990), *Interval methods for systems of Equations*, Cambridge University Press, Cambridge.

Nguyen, H. T., Wang, T. & Kreinovich, V. (2003), Towards foundations of processing imprecise data: From traditional statistical techniques of processing crisp data to statistical processing of fuzzy data, *in* 'Proceedings of the International Conference on Fuzzy Information Processing: Theories and Applications FIP'2003', Vol. II, Beijing, China, pp. 895–900.

Novák, V. (2005), 'Are fuzzy sets a reasonable tool for modeling vague phenomena?', *Fuzzy Sets ans Systems* **156**, 341–348.

O'Hagan, A. (1995), 'Fractional bayes factor for model comparison', *Journal of the Royal Statistcal Society B* **57**(1), 99–138.

Palumbo, F. & Irpino, A. (2005), Multidimensional interval-data: Metrics and factorial analysis, *in* 'Proceedings ASMDA', Brest. invited paper.

Peters, G. (1994), 'Fuzzy linear regression with fuzzy intervals', *Fuzzy Sets Systems* **63**(1), 45–55.

Piccolo, D. (1990), 'A distance measure for classifying arima models', *Journal of Time Series Analysis* **11**(2), 153–164.

Piccolo, D. (1998), *Statistica*, Il Mulino.

Puri, M. L. & Ralescu, D. A. (1986), 'Fuzzy random variables', *J. Math Anal. Appl.* **114**, 409–422.

Redden, D. T. & Woodall, W. H. (1994), 'Properties of certain fuzzy linear regression methods', *Fuzzy Sets and Systems* **64**, 361–375.

Rissanen, J. (1989), *Stochastic Complexity in Statistical Inquiry*, World Scientific Publishing Company, Teaneck, New York.

Romano, R. & Palumbo, F. (2005), 'Un modello di regressione fuzzy per la valutazione della soddisfazione', *Statistica e Applicazioni* **3**(1 Special Issue), 117–124.

Romano, R. & Palumbo, F. (2006*a*), Classification of structural equation models based on fuzzy regression., *in* V. E. Vinzi, C. Lauro, A. Braveman, H. A. Kiers & M. G. Schimek, eds, 'Knowledge Extarction and Modeling', Tilapia, Anacapri, pp. 67–68.

Romano, R. & Palumbo, F. (2006*b*), Fuzzy regression and least squares regression: the relationship between two different fitting criteria, *in* 'Abstracts of the SIS2006 Conference', Vol. 2, CLUEP, Torino, pp. 693–696.

Russell, B. (1923), 'Vagueness', *Australian Journal of Psycology and Philosophy* **1**(2), 84–92.

Ryan, T. (1997), *Modern Regression Methods*, Wiley, New York.

Sakawa & Yano (1992), 'Multiobjective fuzzy linear regression analysis for fuzzy input-output data', *Fuzzy Sets and Systems* **47**, 173–181.

Shapiro, A. (2004), 'Fuzzy regression and the term structure of interest rates revisited', *AFIR* .

Sugeno, M. (1977), *Fuzzy measures and fuzzy integrals: A survey*, Vol. Fuzzy Automata and Decision Processes, North-Holland Publ., Amsterdam, pp. 89–102.

Sunaga, T. (1958), 'Theory of interval algebra and its application to numerical analysis', *RAAG Mem.* **2**, 28–46.

Taheri, S. M. (2003), 'Trends in fuzzy statistics', *Austrian Journal of Statistics* **32**(3), 239–257.

Tanaka, H. & Guo, P. (1999), *Possibilistic Data Analysis for Operations Research*, Physica-Verlag, New York.

Tanaka, H., Uejima, S. & Asai, K. (1982), 'Linear regression analysis with fuzzy model', *IEEE Transactions Systems Man Cybernet* **12**, 903–907.

Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M. & Lauro, C. (2005), 'Pls path modeling', *Computational Statistics and Data Analysis* **48**, 159–205.

Thurstone, L. L. (1931), *The theory of multiple factors*, Edwards Brothers, Ann Arbor, MI.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society* **58**(B), 267–288.

Tukey, J. W. (1964), Causation, regression and path analysis, *in* 'Statistics and Mathematics in Biology', Hafner Publishing Company, New York.

Wang, H. F. & Tsaur, R. C. (2000*a*), 'Insight of a fuzzy regression model', *Fuzzy Sets and System* **112**, 355–369.

Wang, H. F. & Tsaur, R. C. (2000*b*), 'Resolution of fuzzy regression model', *Fuzzy Sets and Systems* **126**, 637–650.

Warmus, M. (1956), 'Calculus of approximations', *Bull. Sci. L'Acad. Pol. Sci.* **III-15**(5), 253–259.

Wold, H. (1966), Estimation of principal component and related models by iterative least squares, *in* P. R. Krishnaiah, ed., 'Multivariate Analysis', Academic Press, New York, pp. 391–420.

Wold, H. (1975), Modelling in complex situations with soft infromation, *in* 'Third World Congress of Econometric Society', Toronto, Canada.

Wold, H. (1979), Model construction and evaluation when theoretical knowledge is scarce: An example of the use of partial least squares, Technical report, Cahier 79.06 du Département d' Économétrie, Faculté des sciences économiques et sociales, Université de Genève, Genève.

Wold, H. (1982), Soft modeling: the basic design and some extensions, *in* K. G. Joreskog & H. Wold, eds, 'Systems under Indirect Observation', Vol. Part II, North-Holland, Amsterdam, pp. 1–54.

Wold, H. (1985), Partial least squares, *in* S. Kotz & N. L. Johnson, eds, 'Encyclopedia of Statistical Sciences', Vol. 6, Wiley, New York, pp. 581–591.

Wold, S., Ruhe, A., Wold, H. & Dunn, W. (1984), 'The collinearity problem in linear regression. the pls approach to generalised inverses', *Journal of Scientific Statistical Computing, SIAM* **5**, 735–743.

Xu, R. & Li, C. (2001), 'Multidimensional least-squares fitting with a fuzzy model', *Fuzzy Sets and System* **119**, 215–223.

Young, R. C. (1931), 'The algebra of many-valued quantities', *Math. Ann.* **31**, 260–290.

Zadeh, L. (1965), 'Fuzzy sets', *Information and Control* **8**, 338–353.

Zadeh, L. A. (1973), 'Outline of a new approach to the analysis of complex systems and decision processes', *IEEE Trans. Systems Man and Cybernet* **1**, 28–44.

Zadeh, L. A. (1978), 'Fuzzy sets as a basis for a theory of possibility', *Fuzzy Sets and Systems* **1**, 3–28.

Zadeh, L., Fu, K. S., Tanaka, K. & Shimura, M. (1975), *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, Academic Press, New York.

Zeithaml, V., Parasuraman, A. & Berry, L. (1991), *Servire Qualità*, McGraw-Hill.

Zimmermann, H. J. (1996), *Fuzzy Set Theory and its Applications*, Kluwer Academic Press, Dordrecht.