Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2006) 287 – 313
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Fuzzy clusterwise linear regression analysis with symmetrical fuzzy output variable

Pierpaolo D'Urso*, Adriana Santoro

Dipartimento di Scienze Economiche, Gestionali e Sociali, Università degli Studi del Molise, Via De Sanctis, 86100 Campobasso, Italy

Available online 27 June 2006

Abstract

The traditional regression analysis is usually applied to homogeneous observations. However, there are several real situations where the observations are not homogeneous. In these cases, by utilizing the traditional regression, we have a loss of performance in fitting terms. Then, for improving the goodness of fit, it is more suitable to apply the so-called clusterwise regression analysis. The aim of clusterwise linear regression analysis is to embed the techniques of clustering into regression analysis. In this way, the clustering methods are utilized for overcoming the heterogeneity problem in regression analysis. Furthermore, by integrating cluster analysis into the regression framework, the regression parameters (regression analysis) and membership degrees (cluster analysis) can be estimated simultaneously by optimizing one single objective function. In this paper the clusterwise linear regression has been analyzed in a fuzzy framework. In particular, a fuzzy clusterwise linear regression model (FCWLR model) with symmetrical fuzzy output and crisp input variables for performing fuzzy cluster analysis within a fuzzy linear regression framework is suggested. For measuring the goodness of fit of the suggested FCWLR model with fuzzy output, a fitting index is proposed. In order to illustrate the usefulness of FCWLR model in practice, several applications to artificial and real datasets are shown.

Keywords: Heterogeneous data; Fuzzy clusterwise linear regression analysis; Crisp input; Symmetrical fuzzy output; Fuzzy clustering; Goodness of fit; Cluster validity

1. Introduction

In a statistical perspective, the regression analysis is utilized for studying the dependence relationship between a real phenomenon (dependent variable or output variable) and other (explanatory) real phenomena (explanatory variables or independent variables or input variables). The traditional regression analysis can be suitably utilized in the case of homogeneous observations. However, in many real cases, there are several situations where the observations are not homogeneous. In these cases, by utilizing the traditional regression, we have a loss of fitting performance of the regression model. In order to improve the goodness of fit, it is more suitable to utilize the so-called clusterwise regression analysis, in which we embed the techniques of clustering into regression analysis. In this way, the clustering methods are utilized for overcoming the heterogeneity problem in regression analysis. For explaining more clearly the aim and the real usefulness of the clusterwise regression analysis, we consider the following explicative example of clusterwise on a market segmentation problem in business, drawn by Lau et al. (1999): "The manager collects a sample of the sales and income data from a set of costumers. If the costumers have homogeneous income elasticity (i.e., the

* Corresponding author. Tel.: +39 0874 404407; fax: +39 0874 311124.
E-mail addresses: durso@unimol.it, pierpaolo.durso@uniroma1.it (P. D'Urso), santoro@unimol.it (A. Santoro).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2006.06.001

regression coefficient β), β can simply be estimated by regression of sales on income. In real business, costumers are heterogeneous and income elasticity will vary with customers of different clusters in the sample. The major tasks for the manager are: (i) use the income elasticity as the basis to divide customers into mutually exclusive segments, (ii) estimate the average income elasticity for each segment, (iii) identify the members of each segment. If we ignore the income elasticity differences among segments, the income elasticity estimated from the regression of sales on income will certainly be biased and inaccurate. In other words, if we want to model the parameter heterogeneity in the traditional regression, the appropriate statistical analysis will involve the simultaneous applications of the cluster analysis and regression model. One straightforward approach is the two stage method. In stage 1, we apply cluster analysis to the dataset to divide customers into segments. In stage 2, we perform regression for each segment to estimate the income elasticity. The problem is that the functions optimized in stages 1 and 2 are two different objective functions which are not necessarily related. A better formulation is to integrate the cluster analysis into regression framework, so that the income elasticities and segment membership parameters can be estimated simultaneously by optimizing one single objective function".

In the body of literature, there are many theoretical works on clusterwise regression analysis (see, for example, De Sarbo and Cron, 1988; De Sarbo et al., 1989; De Veaux, 1989; Hathaway and Bezdek, 1993; Hathaway et al., 1996; Hennig, 2000, 2003; Hong and Chao, 2002; Lau et al., 1999; Leşki, 2004; Preda and Saporta, 2005; Quandt and Ramsey, 1978; Shao and Wu, 2005; Spath, 1979; Yang and Ko, 1997; Van Aelest et al., 2006; Wedel and De Sarbo, 1995). Furthermore, the clusterwise regression analysis finds application in several fields, such as market segmentation and business, socio-economics, biology, engineering, and so on (see, for instance, Aurifeille and Quester, 2003; De Sarbo and Cron, 1988; Hosmer, 1974; Lau et al., 1999; Wedel and Steenkamp, 1991).

In this paper the clusterwise linear regression is analyzed in a fuzzy framework. In particular, we propose a fuzzy clusterwise linear regression model (FCWLR model) with symmetrical fuzzy output and crisp input variables for performing fuzzy cluster analysis within a fuzzy linear regression framework. We build our FCWLR model by considering, simultaneously, the Bezdek's approach to fuzzy cluster analysis (Bezdek, 1981) and the linear regression model with fuzzy output variable ($\tilde{Y}$) and crisp explanatory variables ($X_1, \ldots, X_k$) suggested by Coppi and D'Urso (2003):

$$\begin{cases} m_i = m_i^* + e_i, \quad m_i^* = \mathbf{x}_i'\mathbf{a}, \\ {}_{(-)}s_i = {}_{(-)}s_i^* + {}_{(-)}\varepsilon_i, \quad {}_{(-)}s_i = m_i - l_i, \quad {}_{(-)}s_i^* = m_i^* - l_i^*, \quad l_i^* = m_i^* b + d, \\ {}_{(+)}s_i = {}_{(+)}s_i^* + {}_{(+)}\varepsilon_i, \quad {}_{(+)}s_i = m_i + l_i, \quad {}_{(+)}s_i^* = m_i^* + l_i^*, \end{cases}$$

where $\mathbf{x}_i'$ is $(1 \times (k+1))$-vector containing the scalar 1 and the values of the k crisp input variables observed on the ith unit, m_i, m_i^* are, respectively, the ith *observed center* and the ith *interpolated center*, l_i, l_i^* are, respectively, the ith *observed spreads* and the ith *interpolated spreads*, $\mathbf{a}$ is $((k+1) \times 1)$-vector of regression parameters for m_i, b, d are the regression parameters for the other models, and e_i, ${}_{(-)}\varepsilon_i$, ${}_{(+)}\varepsilon_i$ are the residuals.

In matrix form, we can write the previous model as follows:

$$\begin{cases} \mathbf{m} = \mathbf{m}^* + \mathbf{e}, \quad \mathbf{m}^* = \mathbf{X}\mathbf{a}, \\ {}_{(-)}\mathbf{s} = {}_{(-)}\mathbf{s}^* + {}_{(-)}\boldsymbol{\epsilon}, \quad {}_{(-)}\mathbf{s} = \mathbf{m} - \mathbf{l}, \quad {}_{(-)}\mathbf{s}^* = \mathbf{m}^* - \mathbf{l}^*, \quad \mathbf{l}^* = \mathbf{m}^* b + \mathbf{1}d, \\ {}_{(+)}\mathbf{s} = {}_{(+)}\mathbf{s}^* + {}_{(+)}\boldsymbol{\epsilon}, \quad {}_{(+)}\mathbf{s} = \mathbf{m} + \mathbf{l}, \quad {}_{(+)}\mathbf{s}^* = \mathbf{m}^* + \mathbf{l}^*, \end{cases} \tag{1.1}$$

where $\mathbf{1}$ is $(n \times 1)$-vector of all 1's, $\mathbf{X}$ is $(n \times (k+1))$-matrix containing the vector $\mathbf{1}$ concatenated to k crisp input variables, $\mathbf{m}$, $\mathbf{m}^*$ are, respectively, $(n \times 1)$-vectors of *observed centers* and *interpolated centers*, $\mathbf{l}$, $\mathbf{l}^*$ are, respectively, $(n \times 1)$-vectors of *observed spreads* and *interpolated spreads*, $\mathbf{a}$ is $((k+1) \times 1)$-vector of regression parameters for $\mathbf{m}$, b, d are, respectively, the regression parameters for the other models, and $\mathbf{e}$, ${}_{(-)}\boldsymbol{\epsilon}$, ${}_{(+)}\boldsymbol{\epsilon}$ are, respectively, $(n \times 1)$-vectors of residuals.

Notice that, the above fuzzy regression model is based on three linear models. The first one interpolates the centers of the fuzzy observations, the second and third ones yield the *lower* and *upper bounds* (centers $\pm$ spreads), by building other linear models over the first one. The model is hence capable to take into account possible linear relations between the size of the spreads and the magnitude of the estimated centers. This is often the case in realistic applications, where dependence among centers and spreads is likely to occur (for instance, the uncertainty or fuzziness concerning a measurement may depend on its magnitude) (Coppi and D'Urso, 2003; D'Urso, 2003).

Furthermore, in order to test the performance of the proposed FCWLR we suggest a suitable fitting measure, i.e., the R^2 coefficient.

The structure of the paper is characterized in the following way. In Section 2, we define the fuzzy data, i.e., the symmetrical fuzzy data and in Section 3 we consider a particular distance measure between symmetrical fuzzy data. Successively, in Section 4, we propose a FCWLR model with symmetrical fuzzy output variable and crisp input variables. In particular, we formalize the model and solve the connected optimization problem; furthermore, for measuring the fitting of our model, we propose the R^2 coefficient and then prove the decomposition of the total deviation. In Section 5, for showing the applicative performances, our model is applied to several datasets. Some concluding remarks are considered in Section 6.

2. Fuzzy data

Models based on fuzzy data are diffusely used in several fields. "Sometimes, such models are used as simpler alternatives to probabilistic models (Laviolette et al., 1995). Other times they are, more appropriately, used to study data which, for their intrinsic nature, cannot be known or quantified exactly, and, hence, are correctly regarded as vague or fuzzy. A typical example of fuzzy data is a human judgment or a linguistic term. The concept of fuzzy number can be effectively used to describe formally this concept of vagueness associated with a subjective evaluation. Every time we are asked to quantify our sensations or our perceptions, we feel that our quantification has a degree of arbitrariness. But, when our information is analyzed through nonfuzzy techniques, it is regarded as it were exact, and the original fuzziness is not taken into account in the analysis. The aim of fuzzy techniques is to incorporate all the original vagueness of the data. Therefore, models based on fuzzy data use more information than models where the original vagueness of the data is ignored or arbitrarily canceled. Further, models based on fuzzy data are more general, because a crisp number can be regarded as a special fuzzy number having no fuzziness associated with it" (D'Urso and Gastaldi, 2002).

In several substantive applications, the most utilized class of fuzzy variable is the so-called *symmetrical* fuzzy variable. Usually, a symmetrical fuzzy variable is denoted by $\tilde{Y}=(m,l)$, where m denotes the *center* and l the *left* and *right spreads* (in fact, the spreads are symmetrical) with the following *membership function*:

$$\mu(\omega)=L\left(\frac{m-\omega}{l}\right),\quad m-l\leqslant\omega\leqslant m+l\quad (l>0), \tag{2.1}$$

where L is a decreasing "shape" function from $\Re^+$ to $[0, 1]$ with $L(0)=1$, $L(z)<1$ for all $z>0$, $L(z)>0$ for all $z<1$, $L(1)=0$ (or $L(z)>0$ for all z and $L(+\infty)=0$) $L(z)=L(-z)$, furthermore, the shape function L is symmetric (Zimmermann, 1996). Particular cases of (2.1) are

$$L\left(\frac{m-\omega}{l}\right)=\max\left\{0,1-\left|\frac{m-\omega}{l}\right|^q\right\},\quad L\left(\frac{m-\omega}{l}\right)=\exp\left(-\left|\frac{m-\omega}{l}\right|^q\right),$$
$$L\left(\frac{m-\omega}{l}\right)=1\bigg/\left(1+\left|\frac{m-\omega}{l}\right|^q\right),\quad l>0,\ q>0.$$

Following (2.1), we can define different types of (symmetrical) fuzzy data; the more utilized symmetric fuzzy variables are the symmetric triangular, normal, parabolic and square root fuzzy variables. Each case takes into account a different level of uncertainty around the centers of the (fuzzy) output data.

3. Distance measures between symmetrical fuzzy data

In literature, several topological measures have been generalized to the fuzzy framework. By restricting our interest to distance measures between fuzzy data we can consider, e.g., the following references: Bertoluzza et al. (1995), Coppi and D'Urso (2003), D'Urso and Giordani (2006), Diamond and Kloeden (1994), Kim and Kim (2004), Näther (2000), Yang and Ko (1996), Yang and Liu (1999), Tran and Duckstein (2002a,b). In particular, a squared Euclidean distance between a pair of symmetrical fuzzy variables $\tilde{Y}_i=(m_i,l_i)$, $\tilde{Y}_i^*=\left(m_i^*,l_i^*\right)$ can be defined in the following way (notice that we refer to variables $\tilde{Y}_i$ and $\tilde{Y}_i^*$ as observed on the ith unit) (Coppi and D'Urso, 2003):

$$\Delta_i^2=\left(m_i-m_i^*\right)^2+2\left((m_i-\lambda l_i)-\left(m_i^*-\lambda l_i^*\right)\right)^2=3\left(m_i-m_i^*\right)^2+2\lambda^2\left(l_i-l_i^*\right)^2, \tag{3.1}$$

where m_i, m_i^*, l_i, l_i^* are, respectively, the observed and interpolated centers, and the observed and interpolated (left and right) spreads of the two fuzzy variables, $\lambda=\int_0^1 L^{-1}(\omega)\,d\omega$. The parameter λ has a two-fold role: considering the variability of the membership function and decreasing the emphasis on the spreads. In fact, as it is reasonable to think, the weights of the centers are larger than the weights pertaining to the spreads. In particular, the stronger contribution of the centers is immediately perceivable by considering the squared Euclidean distance of the different particular cases (Coppi and D'Urso, 2003). E.g., the value of the parameter λ connected to the shape function of the symmetric triangular, normal, parabolic and square root fuzzy data are, respectively, $\frac{1}{2}$, $\frac{\sqrt{\pi}}{2}$, $\frac{2}{3}$, $\frac{1}{3}$.

4. FCWLR model with symmetrical fuzzy output variable and crisp input variables

4.1. The model

In a fuzzy framework, there are several real situations in which the fuzzy observations are not homogeneous. For this reason, it is very useful, in these cases, to utilize a FCWLR analysis, in which we embed the techniques of fuzzy clustering into fuzzy regression analysis. In this way, fuzzy clustering methods are utilized for overcoming the heterogeneity problem in fuzzy regression analysis.

In this section, we propose a new FCWLR model.

In our model, for the clustering framework, we follow the Bezdek's approach to fuzzy clustering and for the regression framework, we analyze the relationship between a symmetrical fuzzy output variable and a set of crisp input variables following the fuzzy regression approach suggested by Coppi and D'Urso (2003).

In general, our clusterwise regression analysis is characterized as follows:

- Since the data are heterogeneous, i.e., the data are partitioned in different clusters, we consider one regression model for each cluster. We indicate with r the number of clusters. In particular, we assume that the dependent variable is (symmetrical) fuzzy and the explanatory variables are crisp; then the r regression models are fuzzy. In this manner, we have that the fuzzy regression models represent the prototypes of the r clusters. Notice that, we consider linear fuzzy regression models, then we assume the clusters are linear and then, obviously, also the respective prototypes.
- We integrate the fuzzy regression analysis and the fuzzy clustering techniques in an unique framework. In fact, by means of our *FCWLR* model, we classify, in a fuzzy manner, the n units in r (linear) clusters (by computing the (fuzzy) membership degrees of the n units in each cluster) and, simultaneously, we estimate the parameters of the r fuzzy linear regression models (representing the r fuzzy linear prototypes) of the r clusters. Then, we solve, in a fuzzy framework, simultaneously, a clustering problem and a linear interpolation problem (following a regression approach). In fact, we have a fuzzy assignment of the units to the clusters (clustering component) and an estimation problem of a set of linear regression models with fuzzy dependent variable (regression component).
- The considered information has a double source of uncertainty (fuzziness): in fact, we have fuzziness in the clustering model (*theoretical fuzziness*) and with regard to the regression component, we have fuzziness in a part of the dataset (the dependent variable is fuzzy) (*empirical fuzziness*).

In order to formalize the suggested FCWLR model, we consider preliminarily the following notation.

We consider n units and for each of them we observe k crisp explanatory variables and a fuzzy symmetrical dependent variable. In a clusterwise framework, the classification of a set of units into r (linear) clusters is based on the estimation of the coefficients of each fuzzy linear regression model representing the linear prototype of each cluster (notice that each model has the same mathematical structure of model (1.1)). Then, for each cluster j, we can consider the regression parameters $\mathbf{a}_j$, b_j and d_j, $j=1,\ldots,r$. We can arrange those parameters, in a matrix $\mathbf{A}$,

$$\mathbf{A} = \left.\underbrace{\begin{bmatrix} \mathbf{a}_1 & \mathbf{0}_k & . & . & \mathbf{0}_k \\ \mathbf{0}_k & \mathbf{a}_2 & & . & \\ . & . & . & . & \\ . & . & . & . & \mathbf{0}_k \\ \mathbf{0}_k & . & . & \mathbf{0}_k & \mathbf{a}_r \end{bmatrix}}_{r}\right\} r\,(k+1)\,,$$

and in the vectors $\mathbf{b}=\left[b_1 \dots b_j \dots b_r\right]'$ and $\mathbf{d}=\left[d_1 \dots d_j \dots d_r\right]'$. Moreover, we put $\mathbf{B}=diag(\mathbf{b})$. Then, we obtain the following $(nr \times 1)$-vectors of interpolated centers and spreads: $\mathbf{m}^* = \left[\mathbf{m}_1^{*\prime} \dots \mathbf{m}_j^{*\prime} \dots \mathbf{m}_r^{*\prime}\right]'$, $\mathbf{l}^* = \left[\mathbf{l}_1^{*\prime} \dots \mathbf{l}_j^{*\prime} \dots \mathbf{l}_r^{*\prime}\right]'$, where $\mathbf{m}_j^*$ and $\mathbf{l}_j^*$ are the interpolated centers and spreads, respectively, for the jth cluster, i.e., $\mathbf{m}_j^*=\mathbf{X}\mathbf{a}_j$, $\mathbf{l}_j^*=\mathbf{m}_j^* b_j + d_j$, $j=1,\dots,r$.

Then, we can formalize the *FCWLR* model in the following way:

- We formalize, in a compact form, the fuzzy linear regression models, representing the fuzzy linear prototypes of the r (linear) clusters, as follows:

$$\begin{cases} (\mathbf{1}_{r\times 1}\otimes \mathbf{m}) = \mathbf{m}^* + \mathbf{e}_m, & \mathbf{m}^* = (\mathbf{I}_r \otimes \mathbf{X})\,\mathbf{A}\mathbf{1}_{r\times 1}, \\ \left(\mathbf{1}_{r\times 1}\otimes {}_{(-)}\mathbf{s}\right) = {}_{(-)}\mathbf{s}^* + {}_{(-)}\boldsymbol{\epsilon}, & \left(\mathbf{1}_{r\times 1}\otimes {}_{(-)}\mathbf{s}\right) = (\mathbf{1}_{r\times 1}\otimes \mathbf{m}) - (\mathbf{1}_{r\times 1}\otimes \mathbf{l}), \quad {}_{(-)}\mathbf{s}^* = \mathbf{m}^* - \mathbf{l}^*, \\ & \mathbf{l}^* = \mathbf{I}_{nr}\mathbf{m}^*\mathbf{B}\mathbf{1}_{r\times 1} + (\mathbf{d}\otimes \mathbf{1}_{n\times 1}) = (\mathbf{I}_r\otimes \mathbf{X})\,\mathbf{A}\mathbf{B}\mathbf{1}_{r\times 1} + (\mathbf{d}\otimes \mathbf{1}_{n\times 1}), \\ \left(\mathbf{1}_{r\times 1}\otimes {}_{(+)}\mathbf{s}\right) = {}_{(+)}\mathbf{s}^* + {}_{(+)}\boldsymbol{\epsilon}, & \left(\mathbf{1}_{r\times 1}\otimes {}_{(+)}\mathbf{s}\right) = (\mathbf{1}_{r\times 1}\otimes \mathbf{m}) + (\mathbf{1}_{r\times 1}\otimes \mathbf{l}), \quad {}_{(+)}\mathbf{s}^* = \mathbf{m}^* + \mathbf{l}^*, \\ & \mathbf{l}^* = \mathbf{I}_{nr}\mathbf{m}^*\mathbf{B}\mathbf{1}_{r\times 1} + (\mathbf{d}\otimes \mathbf{1}_{n\times 1}) = (\mathbf{I}_r\otimes \mathbf{X})\,\mathbf{A}\mathbf{B}\mathbf{1}_{r\times 1} + (\mathbf{d}\otimes \mathbf{1}_{n\times 1}), \end{cases} \tag{4.1.1}$$

  where we have the $(nr \times 1)$-vectors $(\mathbf{1}_r \otimes \mathbf{m})$, $(\mathbf{1}_r \otimes \mathbf{l})$, $(\mathbf{d} \otimes \mathbf{1}_n)$, $\left(\mathbf{1}_{r\times 1}\otimes {}_{(-)}\mathbf{s}\right)$, $\left(\mathbf{1}_{r\times 1}\otimes {}_{(+)}\mathbf{s}\right)$, the $(nr\times (k+1)\,r)$-matrix $(\mathbf{I}_r \otimes \mathbf{X})$, and the $(nr \times 1)$-vectors of residuals in the three sub-models: $\mathbf{e}_m$, ${}_{(-)}\boldsymbol{\epsilon}$, ${}_{(+)}\boldsymbol{\epsilon}$.
- Then, by considering the Bezdek's approach to fuzzy clustering, we have the following classification problem:

$$\begin{cases} \min & \psi\,(\mathbf{U}, \mathbf{A}, \mathbf{b}, \mathbf{d}; \mathbf{X}, \mathbf{m}, \mathbf{l}, r, \alpha) = \mathbf{1}'_{1\times nr}\mathbf{U}^{\alpha}\boldsymbol{\Delta}^2\mathbf{1}_{nr\times 1} \\ \text{s.t.} & \mathbf{P}\mathbf{U}\mathbf{1}_{nr\times 1} = \mathbf{1}_{n\times 1}, \quad \mathbf{U} \geqslant \mathbf{0}_{nr\times nr} \ \text{(i.e. all elements in } \mathbf{U} \text{ are nonnegative)}, \end{cases} \tag{4.1.2}$$

  where

- $\mathbf{U}$ is the matrix containing all the membership degrees of each unit in each cluster. Specifically, $\mathbf{U}$ is the $(nr \times nr)$-diagonal matrix containing the membership degrees of the ith unit in the jth cluster $\left(u_{ij},\ i=1,\dots,n,\ j=1,\dots,r\right)$. The membership degrees of the units $(i=1,\dots,n)$ in different clusters $(j=1,\dots,r)$ are arranged, from the first cluster to the last one. This definition enables us to obtain the matrix $\mathbf{U}^{\alpha}$ from the original matrix $\mathbf{U}$, as a power operation from ordinary scalar elements; then, we can derive the expression (4.1.2) directly with respect to $\mathbf{U}$.
- $\alpha > 1$ is a weighting exponent that controls the fuzziness of the clustering; i.e., the *fuzzification factor* α produces a balance between the membership degree close to 0 or 1 and those with intermediate values.
- $\boldsymbol{\Delta}^2$ is the $(nr \times nr)$-diagonal matrix of the squared Euclidean distances between the ith unit and the linear prototype of the jth cluster, i.e., the jth fuzzy linear regression model; i.e., in the diagonal of the matrix $\boldsymbol{\Delta}^2$ the distances are arranged as follows: starting from the first linear cluster, we have the distances between each unit (from the first until the last one) and the linear prototype of the cluster; successively, we have the distances between each unit and the second cluster and so on. The matrix $\boldsymbol{\Delta}^2$ is formalized in the following way:

$$\begin{aligned} \boldsymbol{\Delta}^2 = 3diag\Big\{&diag\Big\{3\left[(\mathbf{1}_{r\times 1}\otimes \mathbf{m}) - (\mathbf{I}_r\otimes \mathbf{X})\,\mathbf{A}\mathbf{1}_{r\times 1}\right]\left[(\mathbf{1}_{r\times 1}\otimes \mathbf{m}) - (\mathbf{I}_r\otimes \mathbf{X})\,\mathbf{A}\mathbf{1}_{r\times 1}\right]'\Big\}\Big\} \\ &+ 2\lambda^2 diag\Big\{diag\Big\{\left[(\mathbf{1}_{r\times 1}\otimes \mathbf{l}) - (\mathbf{I}_r\otimes \mathbf{X})\,\mathbf{A}\mathbf{B}\mathbf{1}_{r\times 1} - (\mathbf{d}\otimes \mathbf{1}_{n\times 1})\right]\left[(\mathbf{1}_{r\times 1}\otimes \mathbf{l}) \right. \\ &\left. - (\mathbf{I}_r\otimes \mathbf{X})\,\mathbf{A}\mathbf{B}\mathbf{1}_{r\times 1} - (\mathbf{d}\otimes \mathbf{1}_{n\times 1})\right]'\Big\}\Big\}. \end{aligned} \tag{4.1.3}$$

  (Notice that if $\{\cdot\}$ is a matrix, the operation "$diag\,\{\cdot\}$" is the vector of the diagonal elements of the matrix $\{\cdot\}$, while if $\{\cdot\}$ is a vector, $diag\,\{\cdot\}$ is the diagonal matrix whose principal diagonal is the vector $\{\cdot\}$. Furthermore, the operator $diag\,\{diag\,\{\cdot\}\}$ is applied to each squared matrix and indicates the diagonal matrix, whose diagonal elements are the diagonal elements of the matrix $\{\cdot\}$.)
- $\mathbf{P}\mathbf{U}\mathbf{1}_{nr\times 1}=\mathbf{1}_{n\times 1}$ is the usual normalization constraint on the membership degrees, where $\mathbf{P}$ is the $(n \times nr)$-matrix $\mathbf{P}=\mathbf{1}'_{r\times 1}\otimes \mathbf{I}_n$. In fact, the constraint $\mathbf{P}\mathbf{U}\mathbf{1}_{nr\times 1}=\mathbf{1}_{n\times 1}$ implies that the sum of all membership degrees, for each unit, is equal to one.

- The constraint $\mathbf{U} \geqslant \mathbf{0}_{nr\times nr}$, where $\mathbf{0}_{nr\times nr}$ is a $(nr \times nr)$-squared matrix with all components null, means that all membership degrees cannot be negative.

4.2. The optimization problem

For estimating the regression parameters and the membership degrees of our model (4.1.1) in a fuzzy clusterwise regression framework, by considering (4.1.2) we have the following constrained minimization problem (in a Lagrangian form):

$$\min : \varphi\left(\mathbf{U}, \mathbf{A}, \mathbf{b}, \mathbf{d}; \mathbf{X}, \mathbf{m}, \mathbf{l}, r, \alpha\right) = \mathbf{1}'_{nr\times 1}\mathbf{W}^{2\alpha}\boldsymbol{\Delta}^2\mathbf{1}_{nr\times 1} + \boldsymbol{\rho}'\left(\mathbf{P}\mathbf{W}^2\mathbf{1}_{nr\times 1} - \mathbf{1}_{n\times 1}\right), \tag{4.2.1}$$

where

- $\mathbf{W}^2 = \mathbf{U}$ is the $(nr \times nr)$-diagonal matrix containing all the membership degrees of each unit to each cluster. Notice that we put $\mathbf{W}^2$ in (4.2.1) for taking into account the constraint $u_{ij} > 0$, $i = 1, \ldots, n$, $j = 1, \ldots, r$.
- $\boldsymbol{\Delta}^2$ is the $(nr \times nr)$-diagonal matrix defined above.
- The term $\boldsymbol{\rho}'\left(\mathbf{P}\mathbf{W}^2\mathbf{1}_{nr\times 1} - \mathbf{1}_{n\times 1}\right)$ is added for taking into account the constraint $\mathbf{P}\mathbf{W}^2\mathbf{1}_{nr\times 1} = \mathbf{1}_{n\times 1}$, i.e., $\sum_{j=1}^{r} u_{ij} = 1$; $\boldsymbol{\rho}$ is a $(n \times 1)$-vector of Lagrangian multipliers and $\mathbf{P}$ is the $(n \times nr)$-matrix defined above.

4.3. The iterative least-squares estimates

We can compute the (local) optimal solutions of (4.2.1) by putting to zero the partial derivatives of $\varphi\left(\mathbf{U}, \mathbf{A}, \mathbf{b}, \mathbf{d}; \mathbf{X}, \mathbf{m}, \mathbf{l}, r, \alpha\right)$ w.r.t. each parameter.

The iterative solutions are given by the following proposition.

Proposition 1. *The iterative solutions of* (4.2.1) *are*

$$\mathbf{U} = diag\left\{diag\left\{\mathbf{P}'\left(\mathbf{P}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}\mathbf{P}'\right)^{-1}\mathbf{P}\right\}\right\}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}, \tag{4.3.1}$$

$$\begin{aligned}\mathbf{A} = {} & \left[(\mathbf{I}_r \otimes \mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r \otimes \mathbf{X}\right)\right]^{-1}(\mathbf{I}_r \otimes \mathbf{X})'\mathbf{U}^{\alpha}\Big[3\left(\mathbf{I}_r \otimes \mathbf{m}\right) + 2\lambda^2\left(\mathbf{I}_r \otimes \mathbf{l}\right)\mathbf{B} \\ & -2\lambda^2\left(\mathbf{I}_r \otimes \mathbf{1}_{n\times 1}\right) diag(\mathbf{d})\mathbf{B}\Big]\left[3\mathbf{I}_r + 2\lambda^2\mathbf{B}^2\right]^{-1},\end{aligned} \tag{4.3.2}$$

$$\mathbf{b} = diag\left(\left[\mathbf{A}'(\mathbf{I}_r \otimes \mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r \otimes \mathbf{X}\right)\mathbf{A}\right]^{-1}\left(\mathbf{A}'(\mathbf{I}_r \otimes \mathbf{X})'\mathbf{U}^{\alpha}\right)\left[\left(\mathbf{I}_r \otimes \mathbf{l}\right) - \left(\mathbf{I}_r \otimes \mathbf{1}_{n\times 1}\right) diag(\mathbf{d})\right]\right), \tag{4.3.3}$$

$$\mathbf{d} = \left(diag\left(\mathbf{U}^{\bullet^{(\alpha)'}}\mathbf{1}_{n\times 1}\right)\right)^{-1}\left[\mathbf{U}^{\bullet^{(\alpha)'}}\mathbf{l} - \mathbf{B}\mathbf{A}'\left((\mathbf{I}_r \otimes \mathbf{X})'\mathbf{U}^{\alpha}\mathbf{1}_{nr\times 1}\right)\right], \tag{4.3.4}$$

where $\mathbf{U}^{\bullet^{(\alpha)}}$ is the traditional matrix representation of the αth power of the membership degrees, i.e., the matrix of the αth power of the membership degrees arranged in n rows and r columns (i.e., the element on the ith and on the jth column is the membership degree of the ith unit to the jth cluster).

Proof. Firstly, we prove (4.3.1). By putting to zero the partial derivative of (4.2.1) w.r.t. $\mathbf{W}$, we have:

$$\begin{aligned}\frac{\partial\varphi}{\partial\mathbf{W}} &= 2\alpha\mathbf{W}^{2\alpha-1}\boldsymbol{\Delta}^2 + 2\left(\mathbf{I}_r \otimes diag\left(\boldsymbol{\rho}\right)\right)\mathbf{W} = \mathbf{0}_{nr\times nr} \\ &\Leftrightarrow \alpha\mathbf{W}^{2(\alpha-1)}\boldsymbol{\Delta}^2 = \left[-\left(\mathbf{I}_r \otimes diag\left(\boldsymbol{\rho}\right)\right)\right] \\ &\Leftrightarrow \mathbf{U} = \frac{1}{\alpha^{1/(\alpha-1)}}\left[-\left(\mathbf{I}_r \otimes diag\left(\boldsymbol{\rho}\right)\right)\right]^{1/(\alpha-1)}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)} \\ &\Leftrightarrow \mathbf{P}\mathbf{U}\mathbf{P}' = \frac{1}{\alpha^{1/(\alpha-1)}}\mathbf{P}\left[-\left(\mathbf{I}_r \otimes diag\left(\boldsymbol{\rho}\right)\right)\right]^{1/(\alpha-1)}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}\mathbf{P}'.\end{aligned}$$

Now, we can observe that $\mathbf{PUP}' = \mathbf{I}$, and, for the peculiarity of matrix $\mathbf{P}$, we can write

$$\mathbf{I}_n = \frac{1}{\alpha^{1/(\alpha-1)}}\mathbf{P}\left[-\left(\mathbf{I}_r \otimes diag\left(\boldsymbol{\rho}\right)\right)\right]^{1/(\alpha-1)}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}\mathbf{P}'$$

and

$$\begin{aligned}\left(\mathbf{PP}'\right)\mathbf{I}_n &= \frac{1}{\alpha^{1/(\alpha-1)}}\left(\mathbf{PP}'\right)\mathbf{P}\left[-\left(\mathbf{I}_r \otimes diag\left(\boldsymbol{\rho}\right)\right)\right]^{1/(\alpha-1)}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}\mathbf{P}'\\ &= \frac{1}{\alpha^{1/(\alpha-1)}}\mathbf{P}\left[-\left(\mathbf{I}_r \otimes diag\left(\boldsymbol{\rho}\right)\right)\right]^{1/(\alpha-1)}\mathbf{P}'\mathbf{P}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}\mathbf{P}'\\ &\Leftrightarrow \alpha^{1/(\alpha-1)}\left(\mathbf{PP}'\right)\left(\mathbf{P}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}\mathbf{P}'\right)^{-1} = \mathbf{P}\left[-\left(\mathbf{I}_r \otimes diag\left(\boldsymbol{\rho}\right)\right)\right]^{1/(\alpha-1)}\mathbf{P}' = \left(\mathbf{PP}'\right)diag\left(-\boldsymbol{\rho}\right)\\ &\Leftrightarrow diag\left(-\boldsymbol{\rho}\right) = \alpha^{1/(\alpha-1)}\left(\mathbf{P}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}\mathbf{P}'\right)^{-1}\\ &\Leftrightarrow diag\left\{\mathbf{P}'\left(diag\left(-\boldsymbol{\rho}\right)\right)\mathbf{P}\right\} = \left[-\left(\mathbf{I}_r \otimes diag\left(\boldsymbol{\rho}\right)\right)\right]^{1/(\alpha-1)} = \alpha^{1/(\alpha-1)}diag\left\{\mathbf{P}'\left(\mathbf{P}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}\mathbf{P}'\right)^{-1}\mathbf{P}\right\}\end{aligned}$$

and then, we have (4.3.1).

We remember that such matrix $\mathbf{U}$ is the matrix of membership degrees, arranged on the principal diagonal; then, from $\mathbf{U}$, we can easily obtain the membership matrix usually adopted in a fuzzy clustering framework, that we indicate with $\mathbf{U}^{\bullet}$. In fact, we can obtain the previous matrix by considering the following simple algebraic operation: $\mathbf{U}^{\bullet}=\mathbf{PU}\left(\mathbf{I}_r \otimes \mathbf{1}_{n\times 1}\right)$.

To prove (4.3.2)–(4.3.4), we put to 0 the derivatives of the objective function in (4.2.1) w.r.t. the parameters $\mathbf{A}$, $\mathbf{b}$, and $\mathbf{d}$.

By putting to $\mathbf{0}$ its partial derivative w.r.t. $\mathbf{A}$, we get:

$$\begin{aligned}\frac{\partial\varphi\left(\mathbf{U},\mathbf{A},\mathbf{b},\mathbf{d};\mathbf{X},r,\alpha\right)}{\partial\mathbf{A}} &= 3\left\{2(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r\otimes\mathbf{X}\right)\mathbf{A} - 2(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r\otimes\mathbf{m}\right)\right\}\\ &\quad + 2\lambda^2\Big\{2(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r\otimes\mathbf{X}\right)\mathbf{AB}^2 - 2(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r\otimes\mathbf{l}\right)\mathbf{B}\\ &\quad + 2(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r\otimes\mathbf{1}_{n\times 1}\right)diag(\mathbf{d})\mathbf{B}\Big\} = \mathbf{0}\\ &\Leftrightarrow \mathbf{A} = \left[(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r\otimes\mathbf{X}\right)\right]^{-1}(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\Big[3\left(\mathbf{I}_r\otimes\mathbf{m}\right) + 2\lambda^2\left(\mathbf{I}_r\otimes\mathbf{l}\right)\mathbf{B}\\ &\qquad -2\lambda^2\left(\mathbf{I}_r\otimes\mathbf{1}_{n\times 1}\right)diag(\mathbf{d})\mathbf{B}\Big]\left[3\mathbf{I}_r + 2\lambda^2\mathbf{B}^2\right]^{-1}.\end{aligned}$$

We can compact the obtained matrix $\mathbf{A}$ in a new $((k+1)\times r)$ matrix $\dot{\mathbf{A}}$, whose columns are the vectors of parameters $\mathbf{a}$ referred to each cluster. This can be done by means of the algebraic operation $\dot{\mathbf{A}} = \mathbf{HA}$, where $\mathbf{H} = \mathbf{1}'_{r\times 1}\otimes\mathbf{I}_k$.

By fixing to $\mathbf{0}$ the partial derivative of (4.2.1) w.r.t. $\mathbf{B}$, we have:

$$\begin{aligned}\frac{\partial\varphi\left(\mathbf{U},\mathbf{A},\mathbf{b},\mathbf{d};\mathbf{X},\mathbf{m},\mathbf{l},r,\alpha\right)}{\partial\mathbf{B}} &= 2\lambda^2\left\{2\mathbf{A}'(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r\otimes\mathbf{X}\right)\mathbf{AB} - 2\mathbf{A}'(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r\otimes\mathbf{l}\right)\right.\\ &\quad \left.+2\mathbf{A}'(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r\otimes\mathbf{1}_{n\times 1}\right)diag(\mathbf{d})\right\} = \mathbf{0}\\ &\Leftrightarrow \mathbf{B} = \left[\mathbf{A}'(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\left(\mathbf{I}_r\otimes\mathbf{X}\right)\mathbf{A}\right]^{-1}\left(\mathbf{A}'(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\right)\\ &\qquad\times\left\{\left[\left(\mathbf{I}_r\otimes\mathbf{l}\right) - \left(\mathbf{I}_r\otimes\mathbf{1}_{n\times 1}\right)diag(\mathbf{d})\right]\right\}\end{aligned}$$

and finally $\mathbf{b} = diag(\mathbf{B})$, that concludes the proof of (4.3.3).

Finally, by putting to $\mathbf{0}$ the partial derivative of (4.2.1) w.r.t. $\mathbf{d}$ we obtain:

$$\begin{aligned}\frac{\partial\varphi\left(\mathbf{U},\mathbf{A},\mathbf{b},\mathbf{d};\mathbf{X},\mathbf{m},\mathbf{l},r,\alpha\right)}{\partial\mathbf{d}} &= 2\lambda^2\left\{2diag\left(\mathbf{U}^{\bullet^{(\alpha)'}}\mathbf{1}_{n\times 1}\right)\mathbf{d} - 2\mathbf{U}^{\bullet^{(\alpha)'}}\mathbf{l} + 2\mathbf{BA}'(\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\mathbf{1}_{nr\times 1}\right\} = \mathbf{0}\\ &\Leftrightarrow \mathbf{d} = \left(diag\left(\mathbf{U}^{\bullet^{(\alpha)'}}\mathbf{1}_{n\times 1}\right)\right)^{-1}\left[\mathbf{U}^{\bullet^{(\alpha)'}}\mathbf{l} - \mathbf{BA}'\left((\mathbf{I}_r\otimes\mathbf{X})'\mathbf{U}^{\alpha}\mathbf{1}_{nr\times 1}\right)\right]\end{aligned}$$

that proves (4.3.4). □

Remark 1. If $l_i = 0$, $i = 1, \ldots, n$, then all output variables are crisp. In this situation, we have a particular case of our FCWLR model: the FCWLR model for crisp input and output variables. In fact, the linear regression model becomes:

$$(\mathbf{1}_{r\times 1} \otimes \mathbf{y}) = \mathbf{y}^* + \mathbf{e}, \quad \mathbf{y}^* = (\mathbf{I}_r \otimes \mathbf{X})\,\mathbf{B}\mathbf{1}_{r\times 1}, \tag{4.3.5}$$

where $\mathbf{y}$ is the vector pertaining to the (crisp) dependent variable, $\mathbf{X}$ is been defined above, $\mathbf{B}$ is the matrix of the regression coefficients $\boldsymbol{\beta}_j$, $j = 1, \ldots, r$, arranged as follows:

$$\mathbf{B} = \left.\underbrace{\begin{bmatrix} \boldsymbol{\beta}_1 & \mathbf{0}_k & \mathbf{0}_k & . & \mathbf{0}_k \\ \mathbf{0}_k & \boldsymbol{\beta}_2 & & . & \\ . & \mathbf{0}_k & . & . & \\ . & . & . & . & \\ \mathbf{0}_k & . & . & . & \boldsymbol{\beta}_r \end{bmatrix}}_{r}\right\} r\,(k+1)$$

and the vector $\mathbf{e}$ is a $(nr \times 1)$-vector of residuals.

Then, the FCWLR model for crisp input and output variables is formalized as follows:

$$\min : \varphi\,(\mathbf{U}, \mathbf{A}, \mathbf{B}; \mathbf{X}, \mathbf{y}, r, \alpha) = \mathbf{1}'_{1\times nr}\mathbf{W}^{2\alpha}\boldsymbol{\Delta}^2\mathbf{1}_{nr\times 1} + \boldsymbol{\rho}'\left(\mathbf{P}\mathbf{W}^2\mathbf{1}_{nr\times 1} - \mathbf{1}_{n\times 1}\right), \tag{4.3.6}$$

where α and the matrices $\mathbf{W}$ and $\mathbf{P}$ are been defined above and $\boldsymbol{\Delta}$ is the following diagonal matrix of distances:

$$\boldsymbol{\Delta}^2 = diag\left\{diag\left\{\left[(\mathbf{1}_{r\times 1} \otimes \mathbf{y}) - (\mathbf{I}_r \otimes \mathbf{X})\,\mathbf{B}\mathbf{1}_{r\times 1}\right]\left[(\mathbf{1}_{r\times 1} \otimes \mathbf{y}) - (\mathbf{I}_r \otimes \mathbf{X})\,\mathbf{A}\mathbf{1}_{r\times 1}\right]'\right\}\right\}. \tag{4.3.7}$$

For the matrix $\mathbf{U}$, we have the same solution of the fuzzy case:

$$\mathbf{U} = diag\left\{diag\left\{\mathbf{P}'\left(\mathbf{P}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}\mathbf{P}'\right)^{-1}\mathbf{P}\right\}\right\}\left(\boldsymbol{\Delta}^2\right)^{-1/(\alpha-1)}. \tag{4.3.8}$$

Instead, the regression parameters vector, for each cluster, that minimizes (4.3.6) becomes

$$\mathbf{B} = \left[(\mathbf{I}_r \otimes \mathbf{X})'\mathbf{U}^{\alpha}\,(\mathbf{I}_r \otimes \mathbf{X})\right]^{-1}(\mathbf{I}_r \otimes \mathbf{X})'\mathbf{U}^{\alpha}\,(\mathbf{I}_r \otimes \mathbf{y})\,. \tag{4.3.9}$$

Obviously, the (local) optimal solutions of (4.3.6) are obtained by iterating formulas (4.3.8) and (4.3.9) until the convergence of the iterative process.

4.4. The algorithm

We can note that (4.2.1) has no closed solutions, but the solutions can be obtained in an iterative manner. The steps of the algorithm FCWLR model with fuzzy output variable and crisp input variables are shown in the following.

Algorithm

1. We fix α and r and put initial values of $\mathbf{U}$, $\mathbf{b}$ and $\mathbf{d}$.
2. We compute the parameters $\mathbf{A}$, $\mathbf{b}$, $\mathbf{d}$, by utilizing the formulas (4.3.2)–(4.3.4).
3. We compute the new matrix $\mathbf{U}$ by means of formula (4.3.1).
4. By assuming that $\mathbf{U}^{(\tau)}$ represents the membership degrees matrix at the τth iteration we compare $\mathbf{U}^{(\tau)}$ with $\mathbf{U}^{(\tau+1)}$ using a convenient matrix norm: if $\left\|\mathbf{U}^{(\tau+1)} - \mathbf{U}^{(\tau)}\right\| < \upsilon$ (where υ is a small positive number fixed by the researcher) stop; otherwise, return to step 2.

We observe that the convergence of the algorithm depends on the initial values assigned to $\mathbf{U}$. Moreover, the iterative procedure does not guarantee the attainment of a global minimum, but only a local one. For these reasons, it is suitable to repeat the iterative procedure by considering several starting points.

4.5. The goodness of fit and the cluster validity

4.5.1. A fitting index: the R^2 coefficient

In this section, we define an useful goodness of fit index for the FCWLR model with fuzzy output variable and crisp input variables, that is a fuzzy extension of the coefficient of determination for the traditional regression analysis. In D'Urso and Santoro (2006) and Coppi et al. (2006, this issue) we defined such an index in the framework of fuzzy regression analysis, respectively, with LL (symmetric) and LR (asymmetric) fuzzy response variable. Here, we set up a similar index in the case of clusterwise regression analysis for symmetric fuzzy response variables. To define this index, we investigate, preliminarily, some properties of the FCWLR model (4.1.1).

To this purpose, we define, for each cluster j, the following diagonal $(n \times n)$-matrix:

$$\mathbf{U}^{\alpha}_{[j]} = \mathbf{U}^{\alpha}_{(n(j-1)+1):(nj)}.$$

This matrix contains on the principal diagonal the membership degrees of each unit in the jth cluster $(j = 1, \ldots, r)$.

Proposition 2. *The weighted sum (and, hence, the weighted mean) of the n center-residuals* $(\mathbf{m}-\mathbf{m}^*_j)$, *and the weighted sum (and, hence, the weighted mean) of the n spread-residuals* $(\mathbf{l}-\mathbf{l}^*_j)$, $j = 1, 2, \ldots, r$ *are null, i.e.*,

$$\mathbf{1}'\mathbf{U}^{\alpha}_{[j]}(\mathbf{m}-\mathbf{m}^*_j) = 0, \tag{4.5.1}$$

$$\mathbf{1}'\mathbf{U}^{\alpha}_{[j]}(\mathbf{l}-\mathbf{l}^*_j) = 0. \tag{4.5.2}$$

Proof. The objective function (4.2.1), in the part referred to the jth cluster, can be written as

$$\varphi_j = 3\left\|\mathbf{m}-\mathbf{m}^*_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}} + 2\lambda^2\left\|\mathbf{l}-\mathbf{l}^*_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}}, \quad j = 1, \ldots, r, \tag{4.5.3}$$

where $\|\cdot\|^2_{\mathbf{U}^{\alpha}_j}$ is the squared norm in the metrics $\mathbf{U}^{\alpha}_j$.

We can rewrite (4.5.3) as follows:

$$\varphi_j = \left\|\begin{matrix} 3\left(\mathbf{m}-\mathbf{m}^*_j\right) \\ 2\lambda^2\left(\mathbf{l}-\mathbf{l}^*_j\right)\end{matrix}\right\|^2_{\mathbf{U}^{\alpha}_{[j]}} = \left\|\begin{pmatrix} 3\mathbf{m} \\ 2\lambda^2\left(\mathbf{l}-\mathbf{1}d_j\right)\end{pmatrix} - \begin{pmatrix} 3\mathbf{X} \\ 2\lambda^2\mathbf{X}b_j\end{pmatrix}\mathbf{a}_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}} = \left\|\mathbf{y}-\mathbf{Z}\mathbf{a}_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}}. \tag{4.5.4}$$

In this way, in (4.5.4) we transform the fuzzy regression problem in an ordinary (generalized) regression problem between $\mathbf{y}$ (the vector of the fuzzy output variable) and $\mathbf{Z}$ (the matrix of the crisp input ones). It is well known that, for the optimal vector $\mathbf{a}_j$, we have

$$\mathbf{1}'\mathbf{U}^*_{[j]}\mathbf{y} = \mathbf{1}'\mathbf{U}^*_{[j]}\mathbf{Z}\mathbf{a}_j. \tag{4.5.5}$$

By putting $\mathbf{y}$ and $\mathbf{Z}$ into (4.5.4), we get for each j $(j = 1, \ldots, r)$:

$$\begin{aligned}&\mathbf{1}'\mathbf{U}^{\alpha}_{[j]}3\mathbf{m} + \mathbf{1}'\mathbf{U}^{\alpha}_{[j]}2\lambda^2\mathbf{l} - \mathbf{1}'\mathbf{U}^{\alpha}_{[j]}2\lambda^2\mathbf{1}d_j = \mathbf{1}'\mathbf{U}^{\alpha}_{[j]}3\mathbf{X}\mathbf{a}_j + \mathbf{1}'\mathbf{U}^{\alpha}_{[j]}2\lambda^2\mathbf{X}\mathbf{a}_j b_j \\ &\qquad\Leftrightarrow 3\left[\mathbf{1}'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{m}-\mathbf{m}^*_j\right)\right] = -2\lambda^2\left[\mathbf{1}'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{l}-\mathbf{l}^*_j\right)\right].\end{aligned} \tag{4.5.6}$$

Then, by remembering the first order condition with respect to d_j:

$$\begin{aligned}&\frac{\partial\varphi_j}{\partial d_j} = 2\lambda^2\left(-2\mathbf{l}'\mathbf{U}^{\alpha}_{[j]}\mathbf{1} + 2\mathbf{a}'_j\mathbf{X}'\mathbf{U}^{\alpha}_{[j]}\mathbf{1}b_j + 2\mathbf{1}'\mathbf{U}^{\alpha}_{[j]}\mathbf{1}d_j\right) = 0 \\ &\quad\Leftrightarrow \mathbf{1}'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{l}-\mathbf{l}^*_j\right) = 0, \quad (j = 1, \ldots, r),\end{aligned} \tag{4.5.7}$$

we obtain (4.5.2) and, by substituting (4.5.2) into (4.5.6), we have

$$0 = 3\left[\mathbf{1}'\mathbf{U}^*_{[j]}\left(\mathbf{m}-\mathbf{m}^*_j\right)\right], \quad j = 1, \ldots, r \tag{4.5.8}$$

which concludes the proof of Proposition 2. □

Proposition 3. *For each cluster, the errors in the model of centers, are orthogonal in the metrics* $\mathbf{U}_{[j]}^{\alpha}$ *with the interpolated centers*:

$$\left(\mathbf{m}-\mathbf{m}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{m}_j^*=0 \quad (j=1,\ldots,r). \tag{4.5.9}$$

Proof. First order conditions of φ_j w.r.t. $\mathbf{a}_j$ and b_j imply $(j=1,\ldots,r)$:

$$\begin{aligned}\frac{\partial\varphi_j}{\partial\mathbf{a}_j}&=3\left(-2\mathbf{X}'\mathbf{U}_{[j]}^{\alpha}\mathbf{m}+2\mathbf{X}'\mathbf{U}_{[j]}^{\alpha}\mathbf{X}\mathbf{a}_j\right)\\&\quad+2\lambda^2\left(-2\mathbf{X}'\mathbf{U}_{[j]}^{\alpha}\mathbf{l}b_j+2\mathbf{X}'\mathbf{U}_{[j]}^{\alpha}\mathbf{X}\mathbf{a}_jb_j^2+2\mathbf{X}'\mathbf{U}_{[j]}^{\alpha}\mathbf{1}b_jd_j\right)=\mathbf{0},\end{aligned} \tag{4.5.10}$$

$$\frac{\partial\varphi_j}{\partial b_j}=2\lambda^2\left(-2\mathbf{l}'\mathbf{U}_j^{\alpha}\mathbf{X}\mathbf{a}_j+2\mathbf{a}_j'\mathbf{X}'\mathbf{U}_j^{\alpha}\mathbf{X}\mathbf{a}_jb_j+2\mathbf{a}_j'\mathbf{X}'\mathbf{U}_j^{\alpha}\mathbf{1}d_j\right)=0. \tag{4.5.11}$$

From (4.5.10), we obtain:

$$\mathbf{X}'\mathbf{U}_{[j]}^{\alpha}\left(3\left(\mathbf{m}-\mathbf{m}_j^*\right)+2\lambda^2b_j\left(\mathbf{l}-\mathbf{l}_j^*\right)\right)=\mathbf{0}. \tag{4.5.12}$$

From (4.5.11) we have

$$\mathbf{m}_j^*\mathbf{U}_{[j]}^{\alpha}\left(\mathbf{l}-\mathbf{l}_j^*\right)=0. \tag{4.5.13}$$

From (4.5.12) we get

$$\mathbf{a}_j'\mathbf{X}'\mathbf{U}_{[j]}^{\alpha}\left(3\left(\mathbf{m}-\mathbf{m}_j^*\right)+2\lambda^2b_j\left(\mathbf{l}-\mathbf{l}_j^*\right)\right)=0. \tag{4.5.14}$$

Then, by substituting (4.5.13) into (4.5.14) we obtain:

$$\mathbf{a}_j'\mathbf{X}'\mathbf{U}_{[j]}^{\alpha}\left(3\left(\mathbf{m}-\mathbf{m}_j^*\right)\right)=0$$

that concludes proof of (4.5.9). □

Proposition 4. *For each cluster* j $(j=1,\ldots,r)$, *the residuals in the model of spreads* $\left(\mathbf{l}-\mathbf{l}_j^*\right)$ *are orthogonal in the metrics* $\mathbf{U}_j^{\alpha}$ *with the interpolated spreads*:

$$\left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{l}_j^*=0. \tag{4.5.15}$$

Proof. We have that:

$$\left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{l}_j^*=\left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\left(\mathbf{X}\mathbf{a}_jb_j+\mathbf{1}d_j\right).$$

From (4.5.2) and (4.5.13), we know that:

$$\left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{m}_j^*=0 \;\Leftrightarrow\; \left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{X}\mathbf{a}_j \;\Leftrightarrow\; \left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{X}\mathbf{a}_jb_j=0$$

and

$$\left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{1}=0 \;\Leftrightarrow\; \left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{1}d_j=0$$

and then we obtain (4.5.15). □

Definition 1. By considering the model (4.1.1), the mean of the centers and the spreads in the jth cluster are, respectively:

$$\tilde{\bar{m}}_j = \frac{\mathbf{m}'\mathbf{U}^{\alpha}_{[j]}\mathbf{1}}{\mathbf{1}'\mathbf{U}^{\alpha}_{[j]}\mathbf{1}} \quad \text{and} \quad \tilde{\bar{l}}_j = \frac{\mathbf{l}'\mathbf{U}^{\alpha}_{[j]}\mathbf{1}}{\mathbf{1}'\mathbf{U}^{\alpha}_{[j]}\mathbf{1}},$$

where $\mathbf{U}_j$ is the diagonal matrix of the membership degrees of each unit to the jth cluster.

Definition 2. By considering the model (4.1.1), for each cluster, we can define:

- the *fuzzy total deviance*:

$$SST_j = dev_{\tilde{Y}_j} = 3\left\|\mathbf{m} - \mathbf{1}\tilde{\bar{m}}_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}} + 2\lambda^2\left\|\mathbf{l} - \mathbf{1}\tilde{\bar{l}}_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}}, \quad j = 1, \ldots, r, \tag{4.5.16}$$

- the *fuzzy regression deviance*:

$$SSR_j = 3\left\|\mathbf{m}^*_j - \mathbf{1}\tilde{\bar{m}}_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}} + 2\lambda^2\left\|\mathbf{l}^*_j - \mathbf{1}\tilde{\bar{l}}_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}}, \quad j = 1, \ldots, r, \tag{4.5.17}$$

- the *fuzzy errors deviance*:

$$SSE_j = 3\left\|\mathbf{m} - \mathbf{m}^*_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}} + 2\lambda^2\left\|\mathbf{l} - \mathbf{l}^*_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}}, \quad j = 1, \ldots, r. \tag{4.5.18}$$

SSR_j is the deviance of $\tilde{Y}$ in the jth cluster, explained by the jth fuzzy linear regression model; SSE_j is the residual deviance, not explained by the model. They are analogous to the deviances defined for the fuzzy regression model, but they are computed in the metrics $\mathbf{U}^{\alpha}_j$, while in the fuzzy linear regression model they are computed in the classic Euclidean norm.

In Proposition 5 we prove the decomposition of (4.5.16) in the two components (4.5.17) and (4.5.18).

Proposition 5. *For each cluster $j = 1, \ldots, r$, the total deviance of $\tilde{Y}$ can be decomposed into two parts*: *the deviance explained by the jth model and the deviance of errors*:

$$SST_j = SSR_j + SSE_j, \quad j = 1, \ldots, r. \tag{4.5.19}$$

Proof. Developing the expression of SST_j, we obtain:

$$\begin{aligned}
SST_j &= 3\left\|\mathbf{m} - \mathbf{1}\tilde{\bar{m}}_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}} + 2\lambda^2\left\|\mathbf{l} - \mathbf{1}\tilde{\bar{l}}_j\right\|^2_{\mathbf{U}^{\alpha}_{[j]}} \\
&= 3\left(\mathbf{m} - \mathbf{m}^*_j\right)'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{m} - \mathbf{m}^*_j\right) + 2\lambda^2\left(\mathbf{l} - \mathbf{l}^*_j\right)'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{l} - \mathbf{l}^*_j\right) \\
&\quad + 3\left(\mathbf{m}^*_j - \mathbf{1}\tilde{\bar{m}}_j\right)'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{m}^*_j - \mathbf{1}\tilde{\bar{m}}_j\right) + 2\lambda^2\left(\mathbf{l}^*_j - \mathbf{1}\tilde{\bar{l}}_j\right)'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{l}^*_j - \mathbf{1}\tilde{\bar{l}}_j\right) \\
&\quad + 2\left[3\left(\mathbf{m} - \mathbf{m}^*_j\right)'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{m}^*_j - \mathbf{1}\tilde{\bar{m}}_j\right) + 2\lambda^2\left(\mathbf{l} - \mathbf{l}^*_j\right)'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{l}^*_j - \mathbf{1}\tilde{\bar{l}}_j\right)\right]
\end{aligned}$$

and then

$$SST_j = SSE_j + SSR_j + 2\left[3\left(\mathbf{m} - \mathbf{m}^*_j\right)'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{m}^*_j - \mathbf{1}\tilde{\bar{m}}_j\right) + 2\lambda^2\left(\mathbf{l} - \mathbf{l}^*_j\right)'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{l}^*_j - \mathbf{1}\tilde{\bar{l}}_j\right)\right]. \tag{4.5.20}$$

To prove (4.5.19), we have to show that the following term is null:

$$3\left(\mathbf{m} - \mathbf{m}^*_j\right)'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{m}^*_j - \mathbf{1}\tilde{\bar{m}}_j\right) + 2\lambda^2\left(\mathbf{l} - \mathbf{l}^*_j\right)'\mathbf{U}^{\alpha}_{[j]}\left(\mathbf{l}^*_j - \mathbf{1}\tilde{\bar{l}}_j\right). \tag{4.5.21}$$

If we write (4.5.21) as follows:

$$3\left(\left(\mathbf{m}-\mathbf{m}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{m}_j^*-\left(\mathbf{m}-\mathbf{m}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{1}\tilde{\bar{m}}_j\right)+2\lambda^2\left(\left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{l}_j^*-\left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{1}\tilde{\bar{l}}_j\right), \tag{4.5.22}$$

we immediately prove (4.5.19), since in the Propositions 2, 3 and 4, we proved that

$$\left(\mathbf{m}-\mathbf{m}_j^*\right)'\mathbf{U}_{\|j\|}^{\alpha}\mathbf{m}_j^*=0, \tag{4.5.23}$$

$$\left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{l}_j^*=0, \tag{4.5.24}$$

$$\left(\mathbf{m}-\mathbf{m}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{1}=0\ \left(\text{and then } \left(\mathbf{m}-\mathbf{m}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{1}\tilde{\bar{m}}_j=0\right), \tag{4.5.25}$$

$$\left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{[j]}^{\alpha}\mathbf{1}=0\ \left(\text{and then } \left(\mathbf{l}-\mathbf{l}_j^*\right)'\mathbf{U}_{j}^{\alpha}\mathbf{1}\tilde{\bar{l}}_j=0\right).\qquad\square \tag{4.5.26}$$

Now, we can define the following goodness of fit measure for the FCWLR model with fuzzy output.

Definition 3. By considering the model (4.1.1), the fit of the jth fuzzy linear regression model can be measured by means of the *fuzzy coefficient of determination*:

$$R_j^2=1-\frac{3\left\|\mathbf{m}-\mathbf{m}_j^*\right\|^2_{\mathbf{U}_{[j]}^{\alpha}}+2\lambda^2\left\|\mathbf{l}-\mathbf{l}_j^*\right\|^2_{\mathbf{U}_{[j]}^{\alpha}}}{3\left\|\mathbf{m}-\mathbf{1}\tilde{\bar{m}}_j\right\|^2_{\mathbf{U}_{[j]}^{\alpha}}+2\lambda^2\left\|\mathbf{l}-\mathbf{1}\tilde{\bar{l}}_j\right\|^2_{\mathbf{U}_{[j]}^{\alpha}}},\quad (j=1,\ldots,r). \tag{4.5.27}$$

Notice that by Proposition 5, we have that R_j^2 varies in the range [0, 1], where 0 corresponds to the horizontal hyperplanes, i.e., in this case the model does not explain anything of the variability of $\tilde{Y}$; the value 1 corresponds to the perfect fit case: in this case, each observation is placed on the interpolating hyperplane, and the model explains all the variability of $\tilde{Y}$.

Remark 2. Notice that the previous goodness of fit measure is able to investigate the fitting performance of the FCWLR model for fuzzy output variable and crisp input variables. By this fitting measure we can derive, as particular case, the fuzzy determination coefficient for the FCWLR model for crisp output and input variables (4.3.5) (see Remark 1):

$$R_j^2=1-\frac{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_j\right)'\mathbf{U}_{[j]}^{\alpha}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_j\right)}{\left(\mathbf{y}-\mathbf{1}\tilde{\bar{y}}_j\right)'\mathbf{U}_{[j]}^{\alpha}\left(\mathbf{y}-\mathbf{1}\tilde{\bar{y}}_j\right)},\quad (j=1,\ldots,r), \tag{4.5.28}$$

where $\tilde{\bar{y}}$ is the fuzzy mean of $\mathbf{y}$ in the jth cluster:

$$\tilde{\bar{y}}=\frac{\mathbf{1}'_{n\times 1}\mathbf{U}_{[j]}^{\alpha}\mathbf{y}}{\mathbf{1}'_{n\times 1}\mathbf{U}_{[j]}^{\alpha}\mathbf{1}_{n\times 1}}. \tag{4.5.29}$$

4.5.2. Cluster validity

Analogously to the fuzzy cluster analysis, also in the fuzzy linear clusterwise regression analysis a crucial topic is the choice of α (fuzziness factor) and r (number of clusters). Some cluster validity measures have been suggested for fuzzy clusterwise regression with crisp data (e.g., Kung and Lin, 2003) and fuzzy data (e.g., Yang and Ko, 1997). For our FCWLR model, we can utilize, for instance, the well-known *partition coefficient* (see Bezdek, 1981):

$$PC=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r}u_{ij}^2. \tag{4.5.30}$$

As known, the coefficient *PC* varies between $1/r$ and 1: the value $1/r$ corresponds to the maximum fuzziness, while the value 1 corresponds to the crisp case.

4.6. Final remarks

Summing up, our *FCWLR* model has the following advantages:

1. By means of the *FCWLR* model, we overcome the heterogeneity problem in fuzzy regression analysis. We solve this problem, by integrating in a single framework the fuzzy clustering and the fuzzy linear regression, so that the regression parameters and the membership degrees can be computed simultaneously by minimizing a single objective function. In this way, following a fuzzy approach, we pursue, simultaneously, two aims: classification of a set of units and interpolation of a dataset.
2. By considering a fuzzy approach to the clusterwise regression problem:
   - we take into account the *fuzziness* (*vagueness*) embodied in the assignment process of the objects to different classes and connected to the particular configuration of the data by means of the *membership degree* of each object in different clusters; it "measures" the uncertainty in the assignment process;
   - we consider the *fuzziness* connected to the *imprecision* embodied in the data (*human perception* expressed in judgments on certain observational situations, *interval valued data*, *vague measurements*). In particular, we assume that the output data of the regression model are symmetrical fuzzy (see Section 2). To this purpose, we remark that our fuzzy model is capable to take into account possible linear relations between the size of the spreads and the magnitude of the estimated centers.
3. Our *FCWLR* model can be seen as a generalization of a fuzzy linear clustering model, when all data are crisp; furthermore, it represents a generalization of the fuzzy linear regression model based on the least-squares approach suggested by Coppi and D'Urso (2003), when the observations are homogeneous (in this case $r = 1$).
4. We can evaluate the goodness of fit of *FCWLR* model by utilizing the suggested fitting measure suitably defined for our model and investigate the cluster validity by applying usefully, in the fuzzy clusterwise regression framework, some criteria suggested in literature for fuzzy clustering.

5. Applicative examples

As to the practical utilization of *FCWLR* model for symmetrical fuzzy output in this field of data analysis, several potential examples might be mentioned, ranging from marketing segmentation to biology and engineering problems.

In the following, for showing the applicative performances of our model, we illustrate different applicative examples. In particular, in Section 5.1, we utilize our model for interpolating three modified versions of the Yang–Ko dataset (1997). In Section 5.2 we apply our model to a completely crisp dataset (Cohen, 1980; De Veaux, 1989).

5.1. Modified Yang–Ko dataset

5.1.1. Experiment 1

We apply the FCWLR model in order to analyze a modified version of Yang–Ko data (1997); i.e., we consider a symmetrical version of the Yang–Ko dataset. The dataset contains 40 observations concerning a crisp independent variable X and a symmetrical fuzzy dependent variable $\tilde{Y}_i = (m_i, l_i)$ (see Table 1). Without loss in generality, we assume that the fuzzy output variable is triangular ($\lambda = 0.5$).

We represent the data in Fig. 1a: the signs "*" indicate the centers of the fuzzy dependent variable $\tilde{Y}_i$; the signs "o" indicate the lower and upper bounds of $\tilde{Y}_i$.

By observing Fig. 1a, we can see that, for about half of the observations, the centers have a decreasing linear relation with X, while, for the other observations, we have an increasing linear relation between centers and X. This situation suggests us that a unique fuzzy linear regression model is not able to interpolate properly the considered dataset. In fact, by interpolating the data by means of a single fuzzy linear regression model ($r = 1$) we get a very bad fitting ($R^2 = 0.0234$) (see Fig. 1b). Conversely, by adopting a FCWLR model with two clusters, and by putting $\alpha = 2$, we obtain a very good fitting (see Fig. 1c).

Table 1
Modified Yank–Ko data

i	$\tilde{Y}_i = (m_i, l_i)$	X_i
1	(2.20, 0.72)	18.11
2	(2.23, 1.02)	2.45
3	(3.20, 0.65)	15.42
4	(3.24, 0.67)	4.68
5	(4.18, 0.82)	16.83
6	(4.32, 0.45)	8.77
7	(4.56, 0.56)	7.71
8	(5.15, 0.55)	15.14
9	(5.71, 1.45)	9.72
10	(5.98, 0.63)	8.15
11	(6.12, 1.67)	13.57
12	(6.15, 0.86)	14.95
13	(6.55, 1.32)	10.25
14	(6.77, 1.56)	10.35
15	(6.87, 0.68)	16.87
16	(7.15, 1.03)	12.86
17	(7.45, 0.59)	11.35
18	(8.15, 0.73)	13.97
19	(8.23, 0.82)	13.32
20	(8.69, 1.63)	10.84
21	(9.56, 0.59)	12.70
22	(10.05, 0.56)	10.66
23	(10.37, 0.87)	11.87
24	(10.78, 1.61)	11.86
25	(10.92, 0.58)	15.87
26	(11.23, 0.76)	16.85
27	(11.49, 0.71)	10.77
28	(11.64, 1.22)	13.85
29	(12.87, 0.63)	9.13
30	(13.55, 0.82)	17.94
31	(13.96, 0.53)	8.44
32	(14.44, 0.87)	19.17
33	(14.55, 0.65)	20.13
34	(14.60, 0.88)	5.08
35	(14.75, 0.76)	7.99
36	(14.89, 0.48)	18.38
37	(15.90, 0.65)	8.80
38	(18.25, 0.62)	5.71
39	(18.33, 0.88)	3.68
40	(19.57, 0.59)	2.13

In fact, by applying the FCWLR model with two clusters ($r = 2$) we obtain the following estimated parameters:

$$\mathbf{A} = \begin{bmatrix} 22.85 & -0.88 \\ -1.11 & 0.78 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -0.0074 \\ -0.0172 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 0.87 \\ 1.02 \end{bmatrix}$$

and the following determination coefficients within each cluster: $R_1^2 = 0.9677$, $R_2^2 = 0.9749$, that indicate a very good fitting performance of the selected model.

Furthermore, our model provides the membership degrees shown in Table 2 (i.e., the partition index $PC = 0.8997$). To this purpose, by seeing Table 2 and Fig. 1c, we can notice that our model selects in a correct manner the configuration structure of the data. In fact, the membership degrees of each unit in the clusters are very crisp, except for four units (units 15, 16, 20 and 21) (marked in Table 2 and Fig. 1c). Indeed, as one would expect, for these four units the membership degrees are suitably fuzzy.

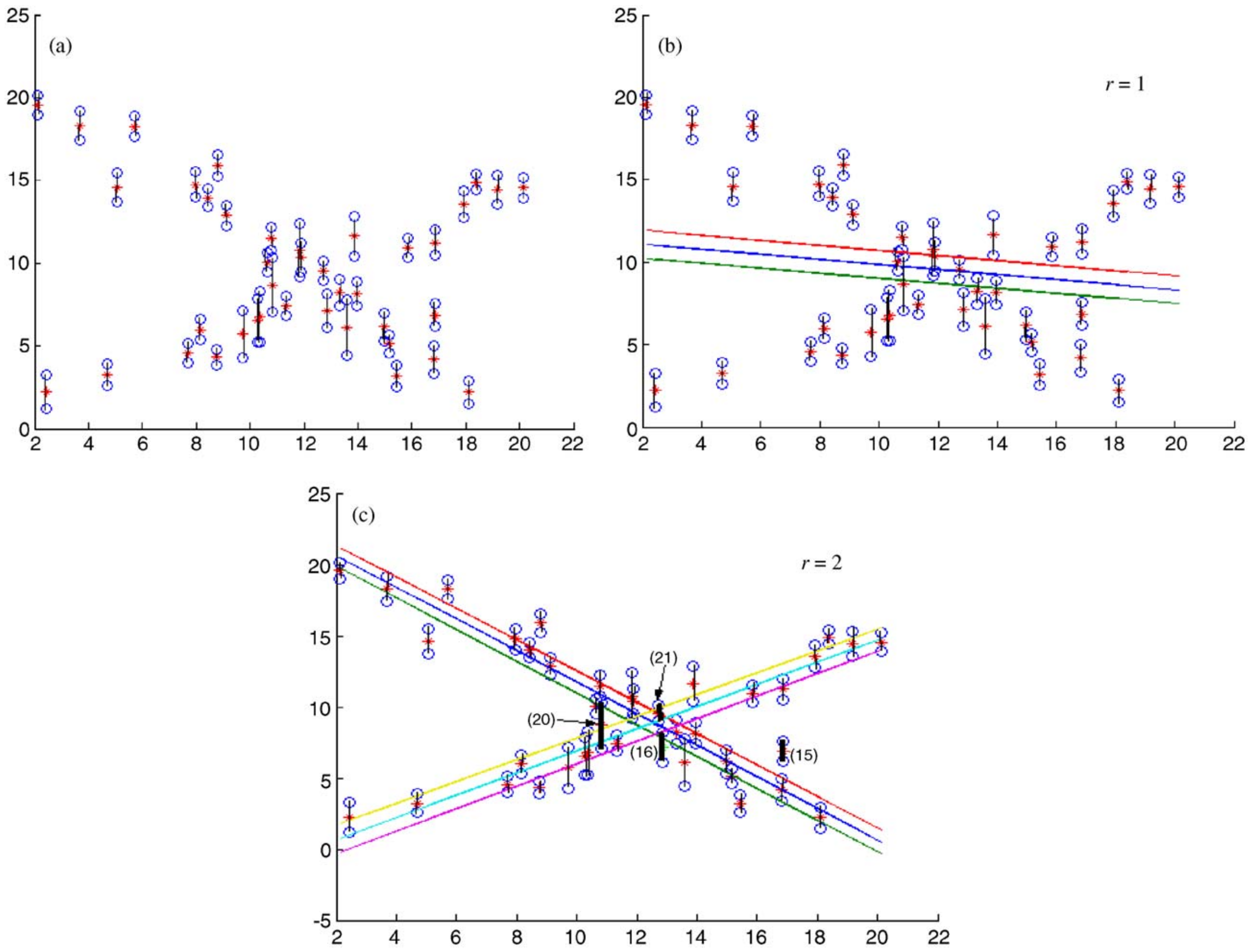

Fig. 1. Scatterplot of the modified Yang–Ko data (a), interpolation of data by using one fuzzy linear regression model ($r = 1$) (b), and interpolation of data by using the FCWLR model with two clusters ($r = 2$) (c).

Concluding, we have that our model interpolates correctly the complex structure of the data capturing, suitably, the empirical information connected to the specific cluster configuration of the data and to the linear structure of the clusters.

5.1.2. Experiment 2

In this example, we add to the modified Yang–Ko dataset (analyzed in the previous section) other 20 observations, that follow a distinct linear model different from the other 40 observations. In this way, we apply FCWLR model for showing how our model can be suitably utilized also in situations in which we have more that two clusters. In particular, for extending the dataset analyzed in Experiment 1, we integrate the original 40 observations of the independent variable X with other 20 observations, extracted from a realization of a uniform r.v. on the interval [0, 20]. Then, we suppose that these 20 observations hold the following symmetrical fuzzy model: $\mathbf{m} = \mathbf{Xa} + N(0, 3)$, $\mathbf{l} = \mathbf{Xa} + \mathbf{1}d + N(0, 1)$ where $\mathbf{a} = [15\ \ 1.5]'$, $b = 0.01$, $d = 2$.

In Table 3, we list the obtained values, i.e., the values of the fuzzy dependent variable and the values of the crisp explanatory variable for the units 41–60 (the values for the units 1–40, obviously, are those listed in Table 1). In Fig. 2a, we show the graphical representation of the complete dataset (all 60 units).

By putting $\lambda = 0.5$ (the fuzzy data are triangular) and $\alpha = 2$, we apply our FCWLR model for $r = 1, 2, 3$. The obtained results (interpolation lines, estimated coefficients, goodness of fit coefficients, partition coefficients and membership degrees of selected partitions) are shown in Fig. 2, Tables 4–6.

Table 2
Membership degrees

i	u_{i1}	u_{i2}
1	0.9974	0.0026
2	0.0045	0.9955
3	0.9064	0.0936
4	0.0012	0.9988
5	1.0000	0.0000
6	0.0335	0.9665
7	0.0036	0.9964
8	0.9755	0.0245
9	0.0242	0.9758
10	0.0045	0.9955
11	0.8135	0.1865
12	0.9994	0.0006
13	0.0135	0.9865
14	0.0112	0.9888
15	0.7947	0.2053
16	0.6553	0.3447
17	0.0330	0.9670
18	0.8424	0.1576
19	0.9848	0.0152
20	0.2267	0.7733
21	0.3314	0.6686
22	0.8787	0.1213
23	0.8950	0.1050
24	0.8225	0.1775
25	0.0097	0.9903
26	0.0198	0.9802
27	0.9787	0.0213
28	0.1499	0.8501
29	0.9994	0.0006
30	0.0019	0.9981
31	0.9966	0.0034
32	0.0010	0.9990
33	0.0003	0.9997
34	0.9510	0.0490
35	0.9935	0.0065
36	0.0138	0.9862
37	0.9257	0.0743
38	0.9862	0.0138
39	0.9993	0.0007
40	0.9976	0.0024

Preliminarily, we observe (see Fig. 2a) that the added observations follow a fuzzy linear regression model very different from the other observations; then we expect that, conversely to the dataset of Experiment 1, two clusters are not sufficient to explain the variability of the new dataset. In particular, we have the following results:

- The model with $r = 1$ is very bad; in fact it does not capture the variability of the dataset (see Fig. 2b and Table 5).
- By applying the FCWLR model with $r = 2$ (see the estimated coefficients in Table 4) we obtain a fuzzy regression line, quite horizontal, that does not fit very well the observed fuzzy points; on the contrary, the other fuzzy regression line is very close to the observed fuzzy points (see Fig. 2c and Table 5). Notice that the first fuzzy regression line crosses the original dataset about in the middle, without capturing their variability, while the second one interpolates almost perfectly the added observations. Then, we can conclude that this model is not satisfactory, since it fits well only a part of our data.
- As we can observe in Fig. 2d, the FCWLR model with $r = 3$ (see estimated coefficients in Table 4) fits very well all the data; this is confirmed by the high values of the estimated goodness of fit indices shown in Table 5. Notice

Table 3
Modified Yank–Ko data (simulated observations 41–60)

i	$\tilde{Y}_i = (m_i, l_i)$	X_i
41	(42.20, 2.73)	19.00
42	(16.94, 0.88)	4.62
43	(33.58, 3.05)	12.14
44	(30.44, 3.92)	9.72
45	(38.30, 1.73)	17.83
46	(41.44, 3.24)	15.24
47	(32.26, 3.54)	9.13
48	(15.44, 0.56)	0.37
49	(40.62, 0.96)	16.43
50	(28.86, 2.86)	8.89
51	(32.90, 1.93)	12.31
52	(40.94, 3.08)	15.84
53	(40.89, 3.24)	18.44
54	(43.70, 3.08)	14.76
55	(19.88, 3.49)	3.53
56	(27.51, 2.94)	8.11
57	(46.26, 3.62)	18.71
58	(42.68, 1.22)	18.34
59	(27.02, 2.25)	8.21
60	(39.31, 2.26)	17.87

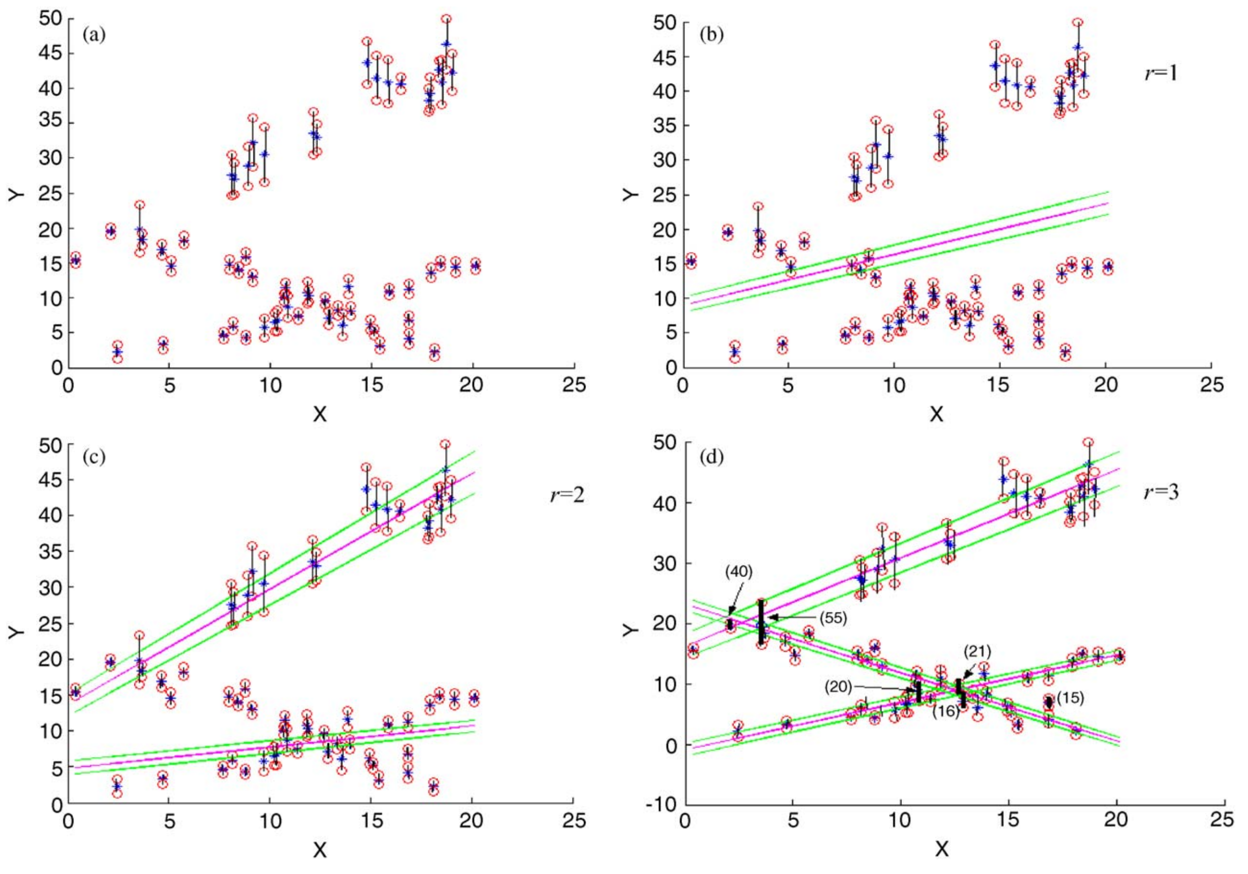

Fig. 2. Scatterplot of the modified Yang–Ko data (a), interpolation of data by using one fuzzy linear regression model ($r = 1$) (b), interpolation of data by using the FCWLR model with two clusters ($r = 2$) (c), and with three clusters ($r = 3$) (d).

Table 4
Estimated coefficients for the FCWLR models with two ($r = 2$) and three clusters ($r = 3$)

Estimated coefficients for the FCWLR model with 2 cluster ($r = 2$)	Estimated coefficients for the FCWLR model with 3 cluster ($r = 3$)
$\mathbf{A} = \begin{bmatrix} 4.83 & 13.62 \\ 0.29 & 1.61 \end{bmatrix}$,	$\mathbf{A} = \begin{bmatrix} -0.88 & 23.08 & 16.23 \\ 0.78 & -1.13 & 1.46 \end{bmatrix}$
$\mathbf{b} = \begin{bmatrix} -0.04 \\ 0.04 \end{bmatrix}$, $\mathbf{d} = \begin{bmatrix} 1.21 \\ 0.88 \end{bmatrix}$	$\mathbf{b} = \begin{bmatrix} -0.02 \\ 0.02 \\ 0.03 \end{bmatrix}$, $\mathbf{d} = \begin{bmatrix} 1.02 \\ 0.68 \\ 1.56 \end{bmatrix}$

Table 5
Goodness of fit and partition coefficients

FCWLR model ($r = 1$)	FCWLR model ($r = 2$)	FCWLR model ($r = 3$)
$R^2 = 0.0754$	$R_1^2 = 0.54$	$R_1^2 = 0.98$
	$R_2^2 = 0.98$	$R_2^2 = 0.98$
	$PC = 0.9$	$R_3^2 = 0.98$
		$PC = 0.9$

that we obtain, for the first two clusters, approximately the same linear clusters obtained in the original dataset, analyzed in Experiment 1. I.e., the first cluster is exactly identical to the second one of Experiment 1, while the second cluster is similar to the first one of Experiment 1 (see Fig. 2d and Table 4). Furthermore, by seeing Table 6, we can observe that approximately all observations have crisp membership degrees in the three clusters, except for the observations marked in Table 6 and in Fig. 2d. In fact, the units belonging in a fuzzy manner to both first and second linear clusters are the same of those obtained in Experiment 1 (i.e., the units with fuzzy membership degrees in the first and second cluster are: 15, 16, 20, 21). The other two units, (40) and (55), have fuzzy membership degrees in the second and third linear clusters, since their position is between the two fuzzy regression lines. Instead, the first and third clusters are logically not overlapped, because they are characterized by two parallel fuzzy regression lines, with no points in the middle, except those ones strongly belonging to the second model. The computed partition coefficient for the defined fuzzy partition is indicated in Table 5.

5.1.3. Experiment 3

For showing the performance of FCWLR model with two explanatory variables, in the third application, we modify the initial Yang–Ko dataset (Experiment 1) by adding a new input variable (i.e., a column in the matrix $\mathbf{X}$). The new variable is generated as follows; the units 1, 3, 5, 8, 11, 12, 15–19, 16, 18, 19, 22–24, 27, 29, 31, 35–40 (in Experiment 1, we found that approximately they belong strongly all to the same cluster), are defined by means of $X_{i2} = m_i * 2 + N(0, 3)$. The other values of X_{i2} are obtained from $X_{i2} = m_i * (-1.5) + N(0, 3)$. The obtained values of this explanatory variable (X_2) are shown in Table 7 (the values of $\tilde{Y}$ and X_1, obviously, are those listed in Table 1).

In Table 8, we illustrate the goodness of fit and the partition coefficients of FCWLR model with $r = 1, 2, 3, 4$.

By inspecting Table 8, we choose the FCWLR model with two linear clusters ($r = 2$); in fact this model is the most parsimonious model with a very high goodness of fit, and the best PC.

The estimated coefficients of the chosen model are:

$$\mathbf{A} = \begin{bmatrix} 20.47 & -0.91 \\ -1.09 & 0.79 \\ 0.07 & 0.01 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -0.01 \\ -0.02 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 0.95 \\ 0.97 \end{bmatrix}.$$

The fuzzy partition obtained by utilizing the FCWLR model with two linear clusters is indicated in Table 9.

5.2. Tone perception data

In this section, in order to show the peformances of FCWLR model and their applicability in particular cases, i.e., empirical situation in which the data are crisp, we analyze a crisp dataset drawn by an experiment of Cohen (1980) and

Table 6
Membership matrix

i	u_{i1}	u_{i2}	u_{i3}
1	0.0017	0.9982	0.0001
2	0.9910	0.0044	0.0046
3	0.0893	0.9063	0.0044
4	0.9983	0.0011	0.0006
5	0.0001	0.9999	0.0000
6	0.9628	0.0330	0.0042
7	0.9958	0.0035	0.0006
8	0.0217	0.9776	0.0007
9	0.9745	0.0239	0.0016
10	0.9950	0.0044	0.0006
11	0.1818	0.8156	0.0026
12	0.0003	0.9997	0.0000
13	0.9861	0.0133	0.0005
14	0.9885	0.0111	0.0004
15	0.2139	0.7808	0.0054
16	0.3404	0.6579	0.0017
17	0.9667	0.0329	0.0004
18	0.1669	0.8324	0.0007
19	0.0186	0.9813	0.0000
20	0.7763	0.2217	0.0019
21	0.6747	0.3249	0.0003
22	0.1290	0.8691	0.0019
23	0.1024	0.8968	0.0008
24	0.1738	0.8242	0.0020
25	0.9900	0.0096	0.0004
26	0.9794	0.0195	0.0011
27	0.0194	0.9799	0.0007
28	0.8487	0.1471	0.0042
29	0.0004	0.9995	0.0001
30	0.9979	0.0019	0.0003
31	0.0027	0.9965	0.0009
32	0.9989	0.0009	0.0002
33	0.9996	0.0003	0.0001
34	0.0496	0.8699	0.0804
35	0.0052	0.9921	0.0027
36	0.9839	0.0134	0.0027
37	0.0684	0.8928	0.0388
38	0.0112	0.9284	0.0603
39	0.0013	0.9666	0.0321
40	0.0009	0.2537	0.7454
41	0.0036	0.0018	0.9946
42	0.0042	0.9728	0.0230
43	0.0003	0.0003	0.9995
44	0.0007	0.0012	0.9981
45	0.0232	0.0119	0.9649
46	0.0096	0.0070	0.9834
47	0.0110	0.0196	0.9695
48	0.0078	0.0386	0.9535
49	0.0008	0.0005	0.9986
50	0.0003	0.0006	0.9991
51	0.0028	0.0029	0.9943
52	0.0031	0.0021	0.9948
53	0.0064	0.0033	0.9903
54	0.0306	0.0241	0.9453
55	0.0030	0.6034	0.3936
56	0.0007	0.0019	0.9974
57	0.0073	0.0039	0.9888
58	0.0005	0.0003	0.9992
59	0.0029	0.0077	0.9894
60	0.0125	0.0065	0.9810

Table 7
Modified Yang–Ko data (simulated values of X_2)

i	X_{i2}
1	6.3058
2	−3.3600
3	8.0536
4	−1.0306
5	8.6180
6	−8.0477
7	−6.5297
8	11.6861
9	−6.5237
10	−16.0638
11	10.3462
12	5.3244
13	−9.0400
14	−6.5147
15	13.4003
16	15.4377
17	17.7326
18	9.9387
19	14.5260
20	−12.1924
21	−15.9636
22	19.4538
23	25.1630
24	21.4447
25	−16.4139
26	−16.8475
27	23.0026
28	−16.2703
29	27.5008
30	−25.3170
31	29.3604
32	−20.9307
33	−25.5948
34	−22.9415
35	36.4279
36	31.3539
37	31.7646
38	39.2394
39	36.8278
40	35.8188

Table 8
Goodness of fit and partition coefficients

FCWLR model ($r=1$)	FCWLR model ($r=2$)	FCWLR model ($r=3$)	FCWLR model ($r=4$)
$R^2=0.1140$	$R_1^2=0.9559$	$R_1^2=0.9443$	$R_1^2=0.9763$
	$R_2^2=0.9455$	$R_2^2=0.9645$	$R_2^2=0.9786$
	$PC=0.9235$	$R_3^2=0.9890$	$R_3^2=0.9881$
		$PC=0.8215$	$R_4^2=0.9763$
			$PC=0.8662$

Table 9
Membership degrees

i	u_{i1}	u_{i2}
1	0.9985	0.0015
2	0.0064	0.9936
3	0.9284	0.0716
4	0.0015	0.9985
5	0.9998	0.0002
6	0.0535	0.9465
7	0.0052	0.9948
8	0.9808	0.0192
9	0.0467	0.9533
10	0.0146	0.9854
11	0.8808	0.1192
12	0.9929	0.0071
13	0.0339	0.9661
14	0.0301	0.9699
15	0.8334	0.1666
16	0.7235	0.2765
17	0.0718	0.9282
18	0.7729	0.2271
19	0.9771	0.0229
20	0.9352	0.0648
21	0.0429	0.9571
22	0.8393	0.1607
23	0.9932	0.0068
24	0.8503	0.1497
25	0.0048	0.9952
26	0.0120	0.9880
27	0.9906	0.0094
28	0.0775	0.9225
29	0.9987	0.0013
30	0.0020	0.9980
31	0.9998	0.0002
32	0.0009	0.9991
33	0.0001	0.9999
34	0.9957	0.0043
35	0.9999	0.0001
36	0.0086	0.9914
37	0.9498	0.0502
38	0.9957	0.0043
39	0.9970	0.0030
40	0.9963	0.0037

studied also by De Veaux (1989). A pure fundamental tone was played to a trained musician. Electronically generated overtones are added, determined by a stretching ratio (X), corresponding to the harmonic pattern usually heard in traditional definite pitched instruments. The musician is asked to tune an adjustable tone to the octave above the fundamental tone. Tuned (Y) gives the ratio of the adjusted tone to the fundamental, i.e., tuned would be the correct tuning for all stretch-ratio values. The analyzed data (Cohen, 1980) belong to 150 trials with the same musician. In the original study, there were four further musicians. Thus, we have one independent variable (stretch-ratio) and one dependent variable (tuned) (see Fig. 3a); both variables are crisp: in this way we adapt our model to a particular case, i.e., spreads $= 0$. This situation enables us to show how our model is useful whenever we handle heterogeneous data, independently of the type of variables. The analyzed dataset is represented in Fig. 3a. We start by interpolating the observed data with a simple regression model: $\mathbf{y} = \mathbf{X}\mathbf{b}$, where $\mathbf{X}$ is the (150×2)-matrix containing the unit column and the column of the observed variable X and $\mathbf{y}$ is the vector of the observed variable Y. Notice that, since we have only crisp variables, the distance (3.1) becomes the *classic Euclidean distance*, and we have to adopt the traditional regression

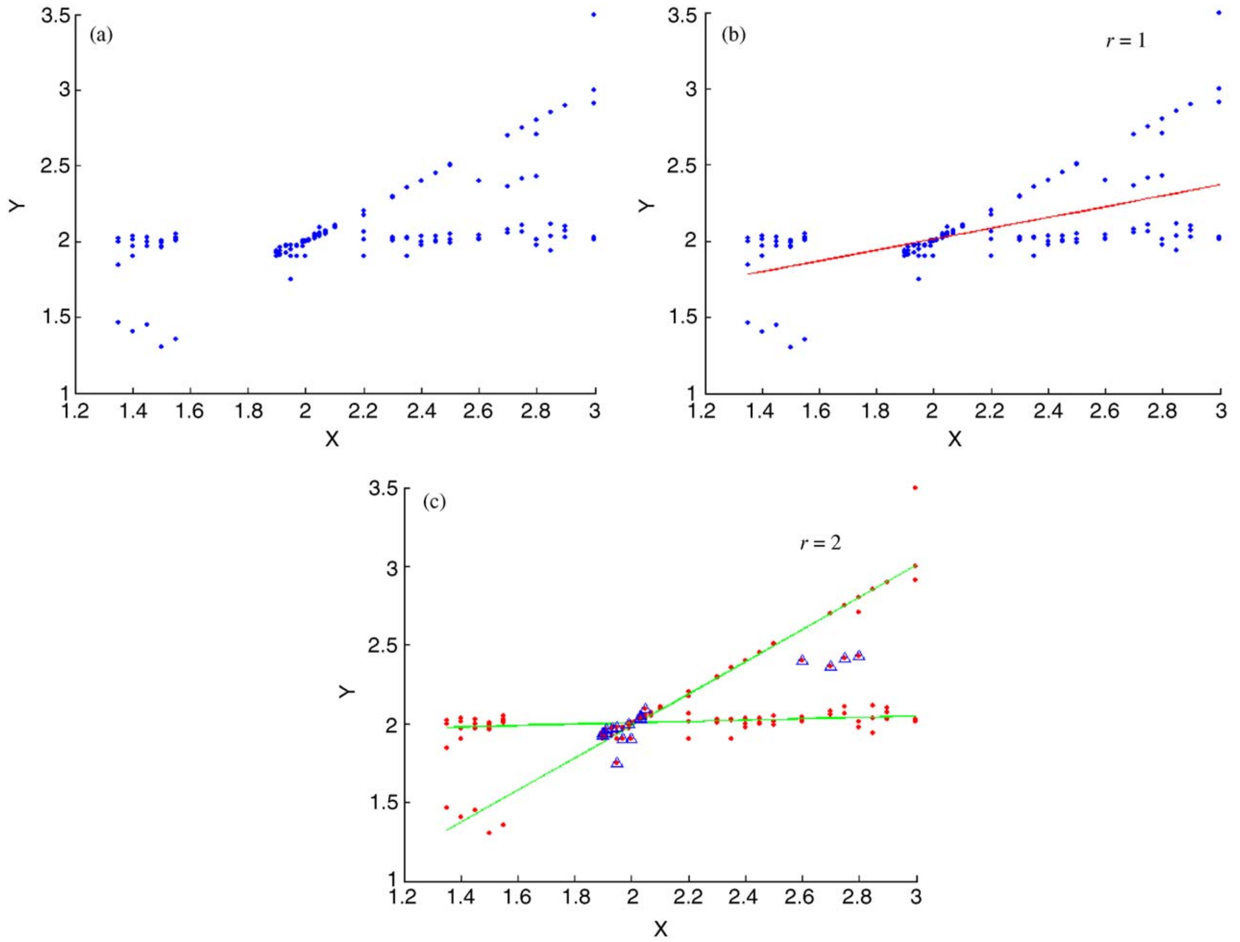

Fig. 3. Scatterplot of the harmonic tone perception data (Cohen, 1980) (a), interpolation of data by using one fuzzy linear regression model ($r = 1$) (b), and interpolation of data by using the FCWLR model with two clusters ($r = 2$) (c).

Table 10
Estimated coefficients for the FCWLR models with two clusters ($r = 2$), goodness of fit and partition coefficient

Estimated coefficients	Goodness of fit	Partition coefficient
$\mathbf{B} = [\boldsymbol{\beta}_1 \ \ \boldsymbol{\beta}_2] = \begin{bmatrix} -0.048 & 1.919 \\ 1.017 & 0.044 \end{bmatrix}$	$R_1^2 = 0.9675$	$PC = 0.8949$
	$R_2^2 = 0.9183$	

models; then, we estimate the regression tparameters by means of the ordinary least-squares estimates: $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$. In our case, we obtain $\hat{\boldsymbol{\beta}} = [1.31 \ \ 0.35]'$. To this purpose, in Fig. 3b we represent the observed and interpolated points. As we can see in Fig. 3b, the estimated regression line does not fit very well the observed data. This evidence is confirmed by the computed determination coefficient: $R^2 = 0.34$ (notice that, in this case, since we have only crisp variables, we use the traditional determination coefficient: $R^2 = 1 - \left(\left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\,(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) / \left((\mathbf{y} - \mathbf{1}_{150\times 1}\bar{y})'\,(\mathbf{y} - \mathbf{1}_{150\times 1}\bar{y})\right)\right)$. We can try to increase the model fit by adopting FCWLR model (4.3.5) for crisp variables. Notice that, in this case, in the FCWLR model the clustering framework is fuzzy (i.e., $u_{ij} \in [0, 1]$ $(i = 1, \ldots, n, \ \ j = 1, \ldots, r)$), whereas the obtained r regression lines (linear prototypes) are crisp. In the following, we illustrate the results of the FCWLR model with 2 clusters ($r = 2$) to our data. In particular, the output of the FCWLR algorithm is summarized in Table 10 (estimated

Table 11
Membership degrees

i	u_{i1}	u_{i2}
1	0.9353	0.0647
2	0.9970	0.0030
3	0.9977	0.0023
4	0.9372	0.0628
5	0.9280	0.0720
6	0.9770	0.0230
7	0.9592	0.0408
8	0.9753	0.0247
9	0.9593	0.0407
10	0.8802	0.1198
11	0.5795	0.4205
12	0.2808	0.7192
13	0.0114	0.9886
14	0.6516	0.3484
15	0.6166	0.3834
16	0.9705	0.0295
17	0.9572	0.0428
18	0.9972	0.0028
19	0.9987	0.0013
20	0.9977	0.0023
21	0.9996	0.0004
22	0.9995	0.0005
23	0.9997	0.0003
24	0.7745	0.2255
25	1.0000	0.0000
26	1.0000	0.0000
27	1.0000	0.0000
28	1.0000	0.0000
29	1.0000	0.0000
30	1.0000	0.0000
31	0.0037	0.9963
32	0.0026	0.9974
33	0.0056	0.9944
34	0.0019	0.9981
35	0.0063	0.9937
36	0.7446	0.2554
37	0.6744	0.3256
38	0.1591	0.8409
39	0.8996	0.1004
40	0.6776	0.3224
41	0.0015	0.9985
42	0.6158	0.3842
43	0.0793	0.9207
44	0.9734	0.0266
45	0.9824	0.0176
46	1.0000	0.0000
47	0.9922	0.0078
48	0.1412	0.8588
49	0.0005	0.9995
50	0.0001	0.9999
51	0.0040	0.9960
52	0.0002	0.9998
53	0.9994	0.0006
54	0.0002	0.9998
55	1.0000	0.0000
56	0.5550	0.4450
57	0.9796	0.0204
58	0.0001	0.9999

Table 11 (*Contd*)

i	u_{i1}	u_{i2}
59	0.0008	0.9992
60	0.8962	0.1038
61	0.0607	0.9393
62	0.0003	0.9997
63	0.0005	0.9995
64	0.0009	0.9991
65	0.0127	0.9873
66	0.9614	0.0386
67	0.9482	0.0518
68	0.9834	0.0166
69	0.6543	0.3457
70	0.8063	0.1937
71	0.2193	0.7807
72	0.0564	0.9436
73	0.0032	0.9968
74	0.7539	0.2461
75	0.8784	0.1216
76	0.9296	0.0704
77	0.9940	0.0060
78	0.9862	0.0138
79	1.0000	0.0000
80	0.0668	0.9332
81	0.0037	0.9963
82	0.0019	0.9981
83	0.0052	0.9948
84	0.0018	0.9982
85	0.4851	0.5149
86	0.0100	0.9900
87	0.0059	0.9941
88	0.0137	0.9863
89	0.0006	0.9994
90	0.0005	0.9995
91	0.0032	0.9968
92	0.0070	0.9930
93	0.0049	0.9951
94	0.0003	0.9997
95	0.0022	0.9978
96	0.9735	0.0265
97	0.9686	0.0314
98	0.9897	0.0103
99	0.9313	0.0687
100	0.6031	0.3969
101	0.9761	0.0239
102	0.1435	0.8565
103	0.0285	0.9715
104	0.7046	0.2954
105	0.9953	0.0047
106	0.9503	0.0497
107	0.9572	0.0428
108	0.0008	0.9992
109	0.0004	0.9996
110	0.0000	1.0000
111	0.0145	0.9855
112	0.0037	0.9963
113	0.0016	0.9984
114	0.0010	0.9990
115	0.0006	0.9994
116	0.0009	0.9991
117	0.0012	0.9988

Table 11 (*Contd*)

i	u_{i1}	u_{i2}
118	0.0001	0.9999
119	0.0005	0.9995
120	0.0014	0.9986
121	0.0008	0.9992
122	0.0201	0.9799
123	0.0006	0.9994
124	0.0019	0.9981
125	0.0019	0.9981
126	0.5419	0.4581
127	0.3025	0.6975
128	0.3145	0.6855
129	0.3337	0.6663
130	0.7939	0.2061
131	0.1734	0.8266
132	0.0001	0.9999
133	0.0259	0.9741
134	0.6575	0.3425
135	0.6937	0.3063
136	0.9397	0.0603
137	0.9902	0.0098
138	0.1335	0.8665
139	0.0022	0.9978
140	0.0001	0.9999
141	0.0012	0.9988
142	0.0007	0.9993
143	0.0016	0.9984
144	0.0003	0.9997
145	0.0037	0.9963
146	0.0014	0.9986
147	0.5132	0.4868
148	0.0096	0.9904
149	0.0039	0.9961
150	0.9884	0.0116

coefficients for the FCWLR models with two clusters ($r = 2$), goodness of fit and partition coefficients) and Table 11 (matrix of the membership degrees). Notice that, $\mathbf{B}$ is the matrix whose columns are the regression coefficient vectors of each cluster. The interpolated linear regression lines are shown in Fig. 3c.

From Table 10 and Fig. 3c, we can conclude that the FCWLR model with two clusters fits very well the observed data: it presents very high within clusters fits and a high partition coefficient. Then, it is not useful to compute other models with higher number of clusters; instead, we can say that the true number of clusters in the observed dataset is 2. Notice that, in Table 11, we mark the units with fuzzy membership degrees. They are also marked with the symbol "Δ" in Fig. 3c. This fuzziness is an obvious consequence of the position of these units in the reference space: their distance from the two regression lines (prototypes) is quite the same. On the contrary, as one would expect, the units positioned near one of the two linear models have crisp membership degrees.

6. Conclusions

In this paper, we have suggested a fuzzy clusterwise linear regression model (*FCWLR* model) for symmetrical fuzzy output. Furthermore, in order to measure the fitting of our model we have proposed the R^2 coefficient and then proved the decomposition of the total deviation. For showing the applicative performances, our model has been applied to different datasets.

Interesting questions for future research include:

1. Simulation study in order to analyze in depth the computational performances of our suggested fuzzy clusterwise linear regression analysis.

2. Fuzzy clusterwise linear regression models based on different types of entropy regularization (for instance, by considering the Shannon entropy measure, the Rényi entropy, etc.).
3. Cluster-validity criteria for fuzzy clusterwise regression.
4. The extension of fuzzy clusterwise regression for fuzzy variables with mixed membership functions.
5. The construction of fuzzy clusterwise regression techniques for dealing with fuzzy random response variables. In this case, the uncertainty to be processed in the model will include both fuzziness and randomness.
6. The extension of the suggested fuzzy clusterwise regression analysis for datasets completely fuzzy (i.e., for fuzzy response variable and fuzzy explanatory variables).
7. The utilization in the estimation procedures of other types of distance measures (e.g., Bertoluzza et al., 1995; Tran and Duckstein, 2002a,b) between fuzzy variables.
8. Fuzzy clusterwise nonlinear regression analysis (i.e., by considering fuzzy polynomial regression models (see, for instance, D'Urso and Gastaldi, 2002).

Acknowledgment

We would like to express our gratitude to the co-editor and the referees whose comments improved the quality of the paper.

References

Aurifeille, J.-M., Quester, P.G., 2003. Predicting business ethical tolerance in international markets: a concomitant clusterwise regression analysis. Internat. Business Rev. 12, 253–272.

Bertoluzza, C., Corral, N., Salas, A., 1995. On a new class of distance between fuzzy numbers. Mathware Soft Comput. 2, 71–84.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

Cohen, E.A., 1980. Inharmonic tone perception, Unpublished Ph.D. Dissertation, Stanford University.

Coppi, R., D'Urso, P., 2003. Regression analysis with fuzzy informational paradigm: a least-squares approach using membership function information. Internat. J. Pure Appl. Math. 8 (3), 279–306.

Coppi, R., D'Urso, P., Giordani, P., Santoro, A., 2006b. Least squares estimation of a linear regression model with LR fuzzy response, Comput. Stat. Data An., this issue, 10.1016/j.csda.2006.04.036.

D'Urso, P., 2003. Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. Comput. Statist. Data Anal. 42 (1–2), 47–72.

D'Urso, P., Gastaldi, T., 2002. An "orderwise" polynomial regression procedure for fuzzy data. Fuzzy Sets and Systems 130 (1), 1–19.

D'Urso, P., Giordani, P., 2006. A weighted fuzzy c-means clustering model for fuzzy data. Comput. Statist. Data Anal. 50 (6), 1496–1523.

D'Urso, P., Santoro, A., 2006. Goodness of fit and variable selection in the fuzzy multiple linear regression. Fuzzy Sets and Systems, in press.

De Sarbo, W.S., Cron, W.L., 1988. A maximum likelihood methodology for clusterwise linear regression. J. Classification 5 (2), 249–282.

De Sarbo, W.S., Oliver, R.L., Rangaswamy, A., 1989. A simulated annealing methodology for clusterwise linear regression. Psychometrika 54 (4), 707–736.

De Veaux, R.D., 1989. Mixtures of linear regressions. Comput. Statist. Data Anal. 8, 227–245.

Diamond, P., Kloeden, P., 1994. Metrics Spaces of Fuzzy Sets. Theory and Applications. World Scientific, Singapore.

Hathaway, R.J., Bezdek, J.C., 1993. Switching regression models and fuzzy clustering. IEEE Trans. Fuzzy Systems 1 (3), 195–204.

Hathaway, R.J., Bezdek, J.C., Pedrycz, W., 1996. A parametric model for fusing heterogeneous fuzzy data. IEEE Trans. Fuzzy Systems 4 (3), 1277–1282.

Hennig, C., 2000. Identifiability of models for clusterwise linear regression. J. Classification 17, 273–296.

Hennig, C., 2003. Clusters, outliers, and regression: fixed point clusters. J. Multivariate Anal. 86, 183–212.

Hosmer, D.W., 1974. Maximum likelihood estimates of parameters of a mixture of two regression lines. Commun. Stat. 3 (10), 995–1005.

Hong, Y.-Y., Chao, Z.-T., 2002. Development of energy loss formula for distribution systems using FCN algorithm and cluster-wise fuzzy regression. IEEE Trans. Power Delivery 17 (3), 794–799.

Kim, D.S., Kim, Y.K., 2004. Some properties of a new metric on the space of fuzzy numbers. Fuzzy Sets and Systems 145, 395–410.

Kung, C.C., Lin, J.C., 2003. A novel cluster validity criterion for fuzzy c-regression model clustering algorithm. IEEE International Conference on Fuzzy Systems, pp. 1368–1373.

Lau, K., Leung, P., Tse, K., 1999. A mathematical programming approach to clusterwise regression model and its extensions. European J. Oper. Res. 111, 640–652.

Laviolette, M., Seaman, J.W., Barrett, J.D., Woodall, W.H., 1995. A probabilistic and statistical view of fuzzy methods, with discussion. Technometrics 37 (3), 249–292.

Leşki, J.M., 2004. ε-insensive fuzzy c-regression models: introduction to ε-insensive fuzzy modelling. IEEE Trans. Systems Man Cybernet.—Part B: Cybernet. 34 (1), 4–15.

Näther, W., 2000. On random fuzzy variables of second order and their application to linear statistical inference with fuzzy data. Metrika 51, 201–221.

Preda, C., Saporta, G., 2005. Clusterwise PLS regression on a stochastic process. Comput. Statist. Data Anal. 49, 99–108.

Quandt, R.E., Ramsey, J.B., 1978. Estimating mixtures of normal distributions and switching regressions (with discussion). J. Amer. Statist. Assoc. 73, 730–752.

Shao, Q., Wu, Y., 2005. A consistent procedure for determining the number of clusters in regression clustering. J. Statist. Plann. Inference 135, 461–476.

Spath, H., 1979. Clusterwise linear regression. Computing 22 (4), 367–373.

Tran, L.T., Duckstein, L., 2002a. Comparison of fuzzy numbers using a fuzzy distance measure. Fuzzy Sets and Systems 130 (3), 331–341.

Tran, L.T., Duckstein, L., 2002b. Multiobjective fuzzy regression with central tendency and possibilistic properties. Fuzzy Sets and Systems 130 (1), 21–31.

Van Aelest, S., Wang, X.S., Zamar, R.H., Zhu, R., 2006. Linear grouping using orthogonal regression. Comput. Statist. Data Anal. 50 (5), 1287–1312.

Wedel, M., De Sarbo, W.S., 1995. A mixture likelihood approcah for generalized linear models. J. Classif. 12, 21–56.

Wedel, M., Steenkamp, J.-B.E.M., 1991. A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. J. Marketing Res. 28, 385–396.

Yang, M.S., Ko, C.H., 1996. On a class of fuzzy c-numbers clustering procedures for fuzzy data. Fuzzy Sets and Systems 84, 49–60.

Yang, M.S., Ko, C.H., 1997. On cluster-wise fuzzy regression analysis. IEEE Trans. Systems Man Cybernet.—Part B 27 (1), 1–13.

Yang, M.S., Liu, H.H., 1999. Fuzzy clustering procedures for conical fuzzy vector data. Fuzzy Sets and Systems 106, 189–200.

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