Qual Quant (2014) 48:3401–3414
DOI 10.1007/s11135-013-9963-9

A generalized maximum entropy (GME) approach for crisp-input/fuzzy-output regression model

Enrico Ciavolino · Antonio Calcagnì

Published online: 26 November 2013

Abstract In this paper we present a crisp-input/fuzzy-output regression model based on the rationale of generalized maximum entropy (GME) method. The approach can be used in several situations in which one have to handle with particular problems, such as small samples, ill-posed design matrix (e.g., due to the multicollinearity), estimation problems with inequality constraints, etc. After having described the GME-fuzzy regression model, we consider an economic case study in which the features provided from GME approach are evaluated. Moreover, we also perform a sensitivity analysis on the main results of the case study in order to better evaluate some features of the model. Finally, some critical points are discussed together with suggestions for further works.

Keywords Generalized maximum entropy method · Fuzzy regression model · Economic case study · Fuzzy statistics · Sensitivity analysis

1 Introduction

Uncertainty is an important attribute of reality. Traditionally, uncertainty has been modelled via the probability theory (Ross 2009; Verkuilen and Smithson 2006) although many empirical phenomena (such as decisions, rating, judgements or opinions) convey a special kind of information which cannot be described using the probability approach. Such real situations seems to convey another form of uncertainty called *fuzziness*. Broadly speaking, we can distinguish three kinds of empirical situations in which we deal with events whose properties

E. Ciavolino (✉)
Department of History, Society and Human Studies, University of Salento, Palazzo Parlangeli, Via V.M. Stampacchia 45, 73100 Lecce, Italy
e-mail: enrico.ciavolino@unisalento.it

A. Calcagnì
Department of Psychology and Cognitive Science, University of Trento, Corso Bettini 84 , 38068 Rovereto, Trento, Italy
e-mail: antonio.calcagni@unitn.it

to be measured can be (i) *vague*, (ii) *fuzzy* and (iii) *crisp*. From a measurement point of view, vague events cannot be quantified because of the lack of metrical information (they can properly be represented with qualitative models only) whereas, due to the presence of *testable information* (Jaynes 1968), crisp events can be well-quantified with metrical structures. Fuzzy events, instead, can suitably be represented with special formal structures based on the soft computing principles. In several cases researchers have to deal with this kind of empirical data, such us, for example, the case of *decision making under uncertainty* for economic and social sciences, economic investment decisions, government decisions, risk decisions, etc. In order to handle with this data, fuzzy set theory (FST) can be seen as good soft computing approach whose models and methods are able to extract useful information from the real data. To do this, several fuzzy statistical methods are available (Buckley 2004; Coppi et al. 2006; Taheri 2003; Nguyen and Wu 2006). In particular, fuzzy regression models have widely been studied using several perspectives (Kacprzyk and Fedrizzi 1992), such us, for example, the least squares approach (Celmiņš 1987; Diamond 1988; D'Urso and Gastaldi 2000; D'Urso 2003; Coppi 2008) and the possibilistic approach (Tanaka 1987).

In this article we describe a crisp-input/fuzzy-output regression model entirely based on the rationale of generalized maximum entropy (GME) method (Golan and Judge 1996; Ciavolino and Dahlgaard 2009) and the application of the method to an economic case study. To this end, the remainder of this paper is structured as follows. The second section is devoted to briefly describes the GME approach. The third section explains the main features of the proposed GME fuzzy regression model whereas the fourth section shows the economic case study as evidence for our statistical method. Finally, the fifth section concludes this article by providing final comments and suggestions about future works.

2 Generalized maximum entropy approach (GME): rationale and method

GME methods of estimation have been proposed for the first time by Golan and Judge (1996) as an extension of the well-known maximum entropy approach proposed by Jaynes (1957, 1982) and based on the principles of Shannon's theory of information (Shannon et al. 1949) [for a general introduction see Kapur (1989)]. These methods are based on a semi-parametric perspective whose features can usefully be employed in several statistical problems as we will show in the following sections. Firstly, in what follows, we describe its rationale.

2.1 Maximum entropy approach (ME)

It is well known as Entropy can be understand as a measure of information carried out from a probability distribution. Letting $\mathcal{X}$ be a random variable whose outcomes are $x_j \quad \forall j = 1, \ldots, m$, with mass probability p_j and $\sum_j^m p_j = 1$, the Shannon's entropy is defined as follows:

$$H(p) = -\sum_j^m p_j \cdot ln(p_j)$$

which is a convex function whose minimum is zero when P is a degenerated probability distribution (perfect certainty) while it assumes its maximum when P is a uniform distribution (perfect ignorance). Maximum entropy method of estimation has been proposed by Jaynes as a method to recover an unknown probability distribution from an empirical dataset by knowing a set of empirical knowledge (Jaynes 1957; Guiasu and Shenitzer 1985). Let us consider an empirical distribution x whose associated probability distribution P is unknown and $y = \sum_j^m x_j p_j$ be the empirical knowledge available (which will be called *consistency constraint*). The probability distribution P can be recovered by the formulation of a constrained-NLP problem as follows:

$$\text{Maximize:} \quad -\sum_j^m p_j \cdot ln(p_j) \quad \text{subject to:} \quad y = \sum_j^m x_j p_j \quad \text{and} \quad \sum_j^m p_j = 1$$

whose solutions can be founded, for example, using the Lagrangian multipliers method:

$$\mathcal{L} = -\sum_j^m p_j \cdot ln(p_j) + \lambda(y - \sum_j^m x_j p_j) + \tau(1 - \sum_j^m p_j)$$

where the first order conditions of the problem are:

$$\frac{\partial \mathcal{L}}{\partial p_j} = -ln(p_j) - \lambda - \tau x_{ij} = 0; \quad \frac{\partial \mathcal{L}}{\partial \lambda} = y - \sum_j^m x_j p_j = 0; \quad \frac{\partial \mathcal{L}}{\partial \tau} = 1 - \sum_j^m p_j = 0$$

whereas the final solution is:

$$\hat{p_j} = \frac{e^{-\lambda x_j}}{\sum_j^k e^{-\lambda x_j}}$$

The solution just described can be applied to solve ill-posed problems, such as the classical example of the *Jaynes's dice experiment*.

Usually, in this empirical problems, the empirical knowledge available is not pure but noisy. Hence, in order to provide a good representative model, the consistency constraint has to be re-written adding a stochastic error term ϵ_i as follows:

$$y_i = \sum_j^m x_j p_j + \epsilon_i$$

This approach, called GME, can be directly explained through the example of regression model (Golan and Judge 1996), as we will show in the next subsection.

2.2 Generalized maximum entropy approach (GME) for the case of regression model

Let us consider the following regression model for the ith unit:

$$y_i = \alpha + \sum_j^m x_{ij} \beta_j + \epsilon_i$$

by using GME approach we can re-parametrized β_j (or α) and ϵ_i as a convex combination of expected value of a discrete random variable, particularly:

$$\beta_j = \sum_k^K z_{jk}^{\beta} p_{jk}^{\beta}$$

where z_{jk}^{β} is the generic element of the *support* $\boldsymbol{z_j^{\beta}}$, a symmetric vector around zero whose dimension is $K \times 1$ (with $2 \leq K \leq 7$), while p_{jk}^{β} is the generic element of the probability vector $\boldsymbol{p_j^{\beta}}$ associated to $\boldsymbol{z_j^{\beta}}$. The error term ϵ_i can also be written as:

$$\epsilon_i = \sum_h^H z_{ih}^{\epsilon} p_{ih}^{\epsilon}$$

where z_{is}^{ϵ} is the generic element of the support $\boldsymbol{z}_{\boldsymbol{i}}^{\boldsymbol{\epsilon}}$, a symmetric vector around zero, while p_{ih}^{ϵ} is the generic element of the probability vector $\boldsymbol{p}_{\boldsymbol{i}}^{\boldsymbol{\epsilon}}$ associated to $\boldsymbol{z}_{\boldsymbol{i}}^{\boldsymbol{\epsilon}}$ (with $2 \leq H \leq 7$). Finally, the GME regression model for the ith unit can be written as follows:

$$y_i = \left(\sum_k^K z_k^{\alpha}\ p_k^{\alpha}\right) + \sum_j^m x_{ij} \left(\sum_k^K\ z_{jk}^{\beta} p_{jk}^{\beta}\right) + \left(\sum_h^H z_{ih}^{\epsilon}\ p_{ih}^{\epsilon}\right)$$

Clearly, $\boldsymbol{z}^{\boldsymbol{\alpha}}$ and $\boldsymbol{p}^{\boldsymbol{\alpha}}$ are the GME-vectors associated to the intercept of the model, $\boldsymbol{z}_{\boldsymbol{j}}^{\boldsymbol{\beta}}$ and $\boldsymbol{p}_{\boldsymbol{j}}^{\boldsymbol{\beta}}$ are the vectors for the jth variable while $\boldsymbol{z}_{\boldsymbol{i}}^{\boldsymbol{\epsilon}}$ and $\boldsymbol{p}_{\boldsymbol{i}}^{\boldsymbol{\epsilon}}$ are defined for the ith component of error terms. In this framework, $\boldsymbol{z}_{\boldsymbol{j}}^{\boldsymbol{\beta}}$ and $\boldsymbol{z}_{\boldsymbol{i}}^{\boldsymbol{\epsilon}}$ have an important role in the probability estimation procedure together with the choice of their supports. These vectors represent the empirical knowledge available to the researchers. Therefore, their values can or directly reflect this information nor they must be chosen ad-hoc, for instance, by using the *three-sigma-rule* (Pukelsheim 1994) or a *sensitivity analysis*.

2.3 Goodness of fit index

The GME approach to regression analysis allow us to use the probability distributions of the regression parameters in order to define a suitable global fit index for the model, as follows:

$$R_{pseudo}^2 = 1 - \left(\frac{-\sum_j^m \sum_k^K\ p_{jk}^{\beta}\ ln(p_{jk}^{\beta}) - \sum_k^K\ p_k^{\alpha}\ ln(p_k^{\alpha})}{(m+1)\cdot ln(K)}\right)$$

that can be referred to Soofi's pseudo-R^2 (Soofi 1992). When $R_{pseudo}^2 = 0$ the model does not have a good fit while $R_{pseudo}^2 = 1$ indicates an excellent fit of the model. Note that the quantity in the bracket is called *Entropy ratio* and measures the reduction of the model's uncertainty related to the parameters estimated. In addition, one can define the following log-likelihood ratio statistic:

$$W = 2\ (m+1)\ ln(K)\cdot R_{pseudo}^2$$

that converges in distribution to a χ^2 with degrees of freedom equal to the number of constraints imposed for the null hypothesis that all parameters of the model are zero (Ciavolino and Dahlgaard 2009; Golan et al. 2000).

2.4 Some remarks

GME approach shows some advantages (Ciavolino and Al-Nasser 2009). In particular, it:

(a) does not require distributional error assumptions;
(b) is robust for a general class of error distributions;
(c) has an excellent work with small samples, when the number of observations is less than the number of variables, when the design matrix is affected by multicollinearity;
(d) allows to use inequality constraints in the estimation process;
(e) allows to employ the set of empirical knowledge about the phenomenon studied and to evaluate its impact on the parameters estimation procedure.

Furthermore, GME can easily be extended toward a more complex approach, called generalized cross-entropy (GCE) method, in which the estimation procedure is constrained under specific a-priori information represented by some distributional models like in the Bayesian context (Golan 2008; Kapur 1989).

3 Crisp-input/fuzzy-output regression model

The relation between crisp independent variables (input) and fuzzy dependent variables (output) plays a special role in the context of socio-economic data. For instance, in socio-economic studies, the relation between crisp (e.g., family income) and fuzzy quantities (e.g., quality of service, quality of teaching, etc) are often analyzed by the researchers (Chang and Yeh 2002; Chan et al. 1999; Benítez et al. 2007). Nevertheless, although some empirical contexts may require more complex representations (e.g., fuzzy-input/fuzzy-output models), in this article we wanted to develop a GME fuzzy regression model for the most simple case first. To this end, in this section we propose a crisp-input/fuzzy-output regression model based on the *generative approach* described for the first time by D'Urso and Gastaldi (2000) and D'Urso (2003). Unlike more complex approach such as the non-generative ones, this approach seems to provide a good compromise between model flexibility and model simplicity.

3.1 Fuzzy data

Traditionally, fuzzy data are represented by the *LR-fuzzy numbers* defined for the first time in Dubois et al. (1988). Generally, a LR-fuzzy number is a convex and normal fuzzy set with a unique core and by considering the analytic expression of its membership function, we can have triangular, trapezoidal, Gaussian, fuzzy numbers (Hanss 2005). In the following, we will only refer to triangular fuzzy numbers. By using two decreasing smooth functions (said *shape functions*):

$$L : \mathbb{R}^+ \to [0,1] \quad \text{and} \quad R : \mathbb{R}^+ \to [0,1]$$

whose properties are:

$$L(\upsilon) \begin{cases} = 0 & \text{if } \upsilon = 1 \\ = 1 & \text{if } \upsilon = 0 \\ > 0 & \text{if } \upsilon < 1 \\ < 1 & \text{if } \upsilon > 0 \end{cases} \qquad R(\upsilon) \begin{cases} = 0 & \text{if } \upsilon = 1 \\ = 1 & \text{if } \upsilon = 0 \\ > 0 & \text{if } \upsilon < 1 \\ < 1 & \text{if } \upsilon > 0 \end{cases} \qquad \forall \upsilon \in \mathbb{R}^+$$

a triangular fuzzy number can be defined through the following membership function:

$$\mu_{\tilde{a}}(x) = \begin{cases} L\left(\frac{c-x}{l}\right) & \text{if } x \le m \\ R\left(\frac{x-c}{r}\right) & \text{if } x \ge m \end{cases}$$

where c is the core, l and r are the left and right spreads respectively. Finally, a LR-fuzzy number $\tilde{a}$ with triangular membership function can be represented by the following triple:

$$\tilde{a} = (c, l, r)_{LR}$$

conveying the main information about fuzzy data. In Fig. 1 we can see a graphical representation of LR-fuzzy number.

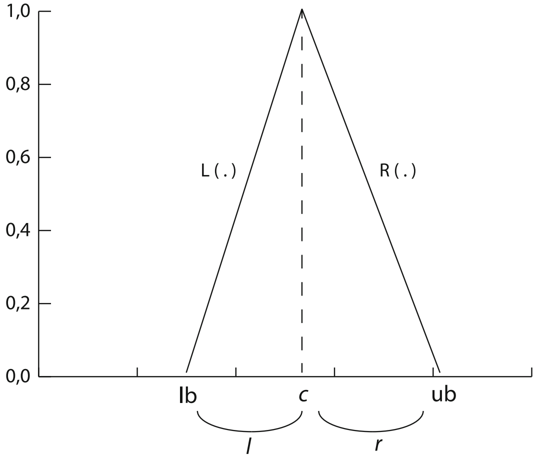

Fig. 1 A triangular fuzzy set (convex set)

3.2 Fuzzy regression model

Let $\boldsymbol{X}_{n,m}$ be a n (cases) x m (variables) matrix representing the set of independent variables and let

$$\tilde{\boldsymbol{Y}} = (\boldsymbol{c}, \boldsymbol{l}, \boldsymbol{r})$$

be a fuzzy matrix representing the dependent variable whose components are $n \times 1$ data vectors. The model representation for $\tilde{\boldsymbol{Y}}$ is the following:

$$\tilde{Y}^* = \begin{cases} c_i = \alpha^c + \sum_{j=1}^{m} x_{ij}\beta_j^c + \epsilon_i^c \quad \equiv \quad c_i = c_i^* + \epsilon_i^c \\ l_i = \alpha^l + c_i^*\beta^l + \epsilon_i^l \quad \equiv \quad l_i = l_i^* + \epsilon_i^l \\ r_i = \alpha^r + c_i^*\beta^r + \epsilon_i^r \quad \equiv \quad r_i = r_i^* + \epsilon_i^r \qquad \forall i = 1 \ldots n \end{cases} \tag{1}$$

where c_i^*, l_i^* and r_i^* are the estimated deterministic components of dependent fuzzy variable, β^c, β^l, β^r, α^c, α^l and α^r are the regression coefficients for the centers, left and right spreads respectively, whereas ϵ_i^c, ϵ_i^l and ϵ_i^r are the error terms of the model. The model adopts a *generative approach* in which c_i^* generates l_i^* and r_i^*. Therefore, the model is able to take into account possible relations between the centers and the spreads. This model representation is in line with a semi-confirmatory approach which assumes that our data are consistent with a generative hypothesis representing possible relations among the centers/modes of the spreads. More precisely, the model captures the dynamic of the spreads as a function of the magnitude of the (estimated) centers/modes. In other words, in some contexts it can be natural to think that the spread (fuzziness) in the measure of an empirical phenomenon is to some extent proportional to its intensity (D'Urso and Gastaldi 2000).

3.3 GME fuzzy regression model

The proposed GME Crisp-Input/Fuzzy-output regression model takes the main advantages of the GME estimation method by handling with ill-posed problems, multicollinearity in the predictors' matrix, small and fat datasets, definition of inequality constraints (such as the definition of positive spreads) or the use of prior information about the phenomenon.

Therefore, in line with the features described in the Sect. 2, the equations for the GME fuzzy regression model are the following:

$$
\begin{aligned}
c_i &= \overbrace{\underbrace{\sum_k^K z_k^{\alpha^c} p_k^{\alpha^c}}_{\alpha^c} + \sum_j^m x_{ij} \underbrace{\sum_k^K z_{jk}^{\beta^c} p_{jk}^{\beta^c}}_{\beta_j^c}}^{c_i^*} + \underbrace{\sum_h^H z_{ih}^{\epsilon^c} p_{ih}^{\epsilon^c}}_{\epsilon_i^c} \\
l_i &= \underbrace{\sum_k^K z_k^{\alpha^l} p_k^{\alpha^l}}_{\alpha^l} + c_i^* \cdot \underbrace{\sum_k^K z_{jk}^{\beta^l} p_{jk}^{\beta^l}}_{\beta^l} + \underbrace{\sum_h^H z_{ih}^{\epsilon^l} p_{ih}^{\epsilon^l}}_{\epsilon_i^l} \\
r_i &= \underbrace{\sum_k^K z_k^{\alpha^r} p_k^{\alpha^r}}_{\alpha^r} + c_i^* \cdot \underbrace{\sum_k^K z_{jk}^{\beta^r} p_{jk}^{\beta^r}}_{\beta^r} + \underbrace{\sum_h^H z_{ih}^{\epsilon^r} p_{ih}^{\epsilon^r}}_{\epsilon_i^r}
\end{aligned}
\tag{2}
$$

The parameters of the model are obtained by recovering the unknown probability distributions $\boldsymbol{p^{\beta^c}}$, $\boldsymbol{p^{\beta^l}}$,$\boldsymbol{p^{\beta^r}}$, $\boldsymbol{p^{\alpha^c}}$, $\boldsymbol{p^{\alpha^l}}$, $\boldsymbol{p^{\alpha^r}}$, $\boldsymbol{p^{\epsilon^c}}$, $\boldsymbol{p^{\epsilon^l}}$ and $\boldsymbol{p^{\epsilon^r}}$ through the following optimization problem.

The GME *objective function* is:

$$
\begin{aligned}
&H(p^{\alpha^c}, p^{\alpha^l}, p^{\alpha^r}, p^{\beta^c}, p^{\beta^l}, p^{\beta^r}, p^{\epsilon^c}, p^{\epsilon^l}, p^{\epsilon^r}) \\
&= -\sum_k^K p_k^{\alpha^c} \cdot ln(p_k^{\alpha^c}) - \sum_k^K p_k^{\alpha^l} \cdot ln(p_k^{\alpha^l}) - \sum_k^K p_k^{\alpha^r} \cdot ln(p_k^{\alpha^r}) \\
&\quad - \sum_j^m \sum_k^K p_{jk}^{\beta^c} \cdot ln(p_{jk}^{\beta^c}) - \sum_j^m \sum_k^K p_{jk}^{\beta^l} \cdot ln(p_{jk}^{\beta^l}) - \sum_j^m \sum_k^K p_{jk}^{\beta^r} \cdot ln(p_{jk}^{\beta^r}) \\
&\quad - \sum_i^n \sum_h^H p_{ih}^{\epsilon^c} \cdot ln(p_{ih}^{\epsilon^c}) - \sum_i^n \sum_h^H p_{ih}^{\epsilon^l} \cdot ln(p_{ih}^{\epsilon^l}) - \sum_i^n \sum_h^H p_{ih}^{\epsilon^r} \cdot ln(p_{ih}^{\epsilon^r})
\end{aligned}
\tag{3}
$$

The *consistency constraints* are represented by the three equations described in equation 2 for the centers and left/right spreads whereas the *normalization constraints* on the probabilities are:

$$
\begin{aligned}
&\sum_k^K p_k^{\alpha^c} = 1 \qquad \sum_k^K p_k^{\alpha^l} = 1 \qquad \sum_k^K p_k^{\alpha^r} = 1 \\
&\sum_k^K p_{jk}^{\beta^c} = 1 \qquad \sum_k^K p_{jk}^{\beta^l} = 1 \qquad \sum_k^K p_{jk}^{\beta^r} = 1 \qquad \forall j = 1, \dots, m \\
&\sum_h^H p_{ih}^{\epsilon^c} = 1 \qquad \sum_h^H p_{ih}^{\epsilon^l} = 1 \qquad \sum_h^H p_{ih}^{\epsilon^r} = 1 \qquad \forall i = 1, \dots, n
\end{aligned}
\tag{4}
$$

A possible way to solve this problem is based on Lagrangian multipliers method whose *Lagrangian function* $\mathcal{L}$ is composed by the entropy function 3 together with the constraints 2 and 4. In particular:

$$\begin{aligned}
\mathcal{L} = {} & H(p^{\alpha^c}, p^{\alpha^l}, p^{\alpha^r}, p^{\beta^c}, p^{\beta^l}, p^{\beta^r}, p^{\epsilon^c}, p^{\epsilon^l}, p^{\epsilon^r}) \\
& + \sum_i^n \lambda_i^c \left(c_i - \sum_{k=1}^K z_k^{\alpha^c} p_k^{\alpha^c} - \sum_j^m x_{ij} \sum_k^K z_{jk}^{\beta^c} p_{jk}^{\beta^c} - \sum_h^H z_{ih}^{\epsilon^c} p_{ih}^{\epsilon^c} \right) \\
& + \sum_i^n \lambda_i^l \left(l_i - \sum_{k=1}^K z_k^{\alpha^l} p_k^{\alpha^l} - c_i^* \sum_k^K z_{jk}^{\beta^l} p_{jk}^{\beta^l} - \sum_h^H z_{ih}^{\epsilon^l} p_{ih}^{\epsilon^l} \right) \\
& + \sum_i^n \lambda_i^r \left(r_i - \sum_{k=1}^K z_k^{\alpha^r} p_k^{\alpha^r} - c_i^* \sum_k^K z_{jk}^{\beta^r} p_{jk}^{\beta^r} - \sum_h^H z_{ih}^{\epsilon^r} p_{ih}^{\epsilon^r} \right) \\
& + \mu^{\alpha^c} \left(\sum_{k=1}^K p_k^{\alpha^c} - 1 \right) + \mu^{\alpha^l} \left(\sum_{k=1}^K p_k^{\alpha^l} - 1 \right) + \mu^{\alpha^r} \left(\sum_{k=1}^K p_k^{\alpha^r} - 1 \right) \\
& + \sum_j^m \mu_j^{\beta^c} \left(\sum_{k=1}^K p_{jk}^{\beta^c} - 1 \right) + \sum_j^m \mu_j^{\beta^l} \left(\sum_{k=1}^K p_{jk}^{\beta^l} - 1 \right) + \sum_j^m \mu_j^{\beta^r} \left(\sum_{k=1}^K p_{jk}^{\beta^r} - 1 \right) \\
& + \sum_i^n \mu_i^{\epsilon^c} \left(\sum_{h=1}^H p_{ih}^{\epsilon^c} - 1 \right) + \sum_i^n \mu_i^{\epsilon^l} \left(\sum_{h=1}^H p_{ih}^{\epsilon^l} - 1 \right) + \sum_i^n \mu_i^{\epsilon^r} \left(\sum_{h=1}^H p_{ih}^{\epsilon^r} - 1 \right)
\end{aligned} \tag{5}$$

By equating to zero the first partial derivatives of the above functional (first order conditions), we obtain the following parametrized solutions:

$$\begin{aligned}
\hat{p}_k{}^{\alpha^c} &= \frac{e^{z_k^{\alpha^c} \cdot (\sum_i \lambda_i^c + \sum_i \lambda_i^l \sum_j p_{jk}^{\beta^l} + \sum_i \lambda_i^r \sum_j p_{jk}^{\beta^r})}}{\sum_k e^{z_k^{\alpha^c} \cdot (\sum_i \lambda_i^c + \sum_i \lambda_i^l \sum_j p_{jk}^{\beta^l} + \sum_i \lambda_i^r \sum_j p_{jk}^{\beta^r})} \equiv \Omega^{\alpha^c}} \\
\hat{p}_k{}^{\alpha^l} &= \frac{e^{\sum_i \lambda_i^l z_k^{\alpha^c}}}{\sum_k e^{\sum_i \lambda_i^l z_k^{\alpha^c}} \equiv \Omega^{\alpha^l}} \\
\hat{p}_k{}^{\alpha^r} &= \frac{e^{\sum_i \lambda_i^r z_k^{\alpha^r}}}{\sum_k e^{\sum_i \lambda_i^r z_k^{\alpha^r}} \equiv \Omega^{\alpha^r}}
\end{aligned} \tag{6}$$

$$\begin{aligned}
\hat{p}_{jk}^{\beta^c} &= \frac{e^{\sum_i^c x_{ij} z_{jk}^{\beta^c} \cdot (\lambda_i^c + \lambda_i^l z_{jk}^{\beta^l} p_{jk}^{\beta^l} + \lambda_i^r z_{jk}^{\beta^r} p_{jk}^{\beta^r})}}{\sum_k e^{\sum_i^c x_{ij} z_{jk}^{\beta^c} \cdot (\lambda_i^c + \lambda_i^l z_{jk}^{\beta^l} p_{jk}^{\beta^l} + \lambda_i^r z_{jk}^{\beta^r} p_{jk}^{\beta^r})} \equiv \Omega^{\beta^c}} \\
\hat{p}_{jk}^{\beta^l} &= \frac{e^{\sum_i \lambda_i^l z_{jk}^{\beta^l} \cdot (z_k^{\alpha^c} p_k^{\alpha^c} + x_{ij} z_{jk}^{\beta^c} p_{jk}^{\beta^c})}}{\sum_k e^{\sum_i \lambda_i^l z_{jk}^{\beta^l} \cdot (z_k^{\alpha^c} p_k^{\alpha^c} + x_{ij} z_{jk}^{\beta^c} p_{jk}^{\beta^c})} \equiv \Omega^{\beta^l}} \\
\hat{p}_{jk}^{\beta^r} &= \frac{e^{\sum_i \lambda_i^r z_{jk}^{\beta^r} \cdot (z_k^{\alpha^r} p_k^{\alpha^r} + x_{ij} z_{jk}^{\beta^r} p_{jk}^{\beta^r})}}{\sum_k e^{\sum_i \lambda_i^r z_{jk}^{\beta^r} \cdot (z_k^{\alpha^r} p_k^{\alpha^r} + x_{ij} z_{jk}^{\beta^r} p_{jk}^{\beta^r})} \equiv \Omega^{\beta^r}} \qquad \forall j = 1, \ldots, m
\end{aligned} \tag{7}$$

$$
\begin{aligned}
\hat{p}_{ih}^{\epsilon^c} &= \frac{e^{\lambda_i^c z_{ih}^{\epsilon^c}}}{\sum_h e^{\lambda_i^c z_{ih}^{\epsilon^c}} \equiv \Omega^{\epsilon^c}} \\
\hat{p}_{ih}^{\epsilon^l} &= \frac{e^{\lambda_i^l z_{ih}^{\epsilon^l}}}{\sum_h e^{\lambda_i^l z_{ih}^{\epsilon^l}} \equiv \Omega^{\epsilon^l}} \\
\hat{p}_{ih}^{\epsilon^r} &= \frac{e^{\lambda_i^r z_{ih}^{\epsilon^r}}}{\sum_h e^{\lambda_i^r z_{ih}^{\epsilon^r}} \equiv \Omega^{\epsilon^r}} \qquad \forall i = 1, \ldots, n
\end{aligned}
\tag{8}
$$

Note that each denominator $\Omega^{(\cdot)}$ of the above formulas represent the *normalization factor* for the probability distributions.

4 An economic case study: employment and unemployment in OECD countries

We studied a known econometric model related to the study of unemployment and employment (Overman and Puga 2002; Patacchini and Zenou 2007; Bernardini Papalia and Ciavolino 2011). Data was drawn from OECD dataset (OECD 2011, 2013) and referred to European OECD countries for the period 2010–2011. The unemployment rate for 2010 and the employment rate for 2011 were used as independent variables whereas the unemployment rate for 2011 was employed as dependent variable. The crisp dependent variable was fuzzified into a triangular fuzzy variables with three levels of definition (low, medium and high unemployment). To do this, we used an application developed in Python which allows to represent crisp variables into fuzzy variables by preserving the original information stored in the data. Such procedure is based on a Mamdani fuzzy system allowing to synthesize the information stored into the histogram of the dependent variable by means of a suitable fuzzy set which is able to reproduce the main characteristics of the histogram representation. This procedure has been widely adopted in the FST literature (Medasani et al. 1998; Nieradka and Butkiewicz 2007; Cheng and Chen 1997) and more details are available in Ciavolino et al. (2013). Thus, the final optimized triangular fuzzy sets for unemployment 2011 were the following: *low* (1, 3, 4), *middle* (5, 8, 9), *high* (10, 12, 15). Final dataset is represented in Table 1.

Note that, $x1$ refers to *Unemployment rate* for 2010, $x2$ refers to *Employment rate* for 2011 whereas c, l and r are the centers and spreads for the fuzzy dependent variable *Unemployment rate* for 2011. The support of the regression coefficients have been chosen through a *sensitivity analysis* whereas the supports for the error terms have been computed by the 3-σ-rule where σ is the empirical standard deviation of the independent variables.

4.1 Sensitivity analysis for β coefficients

In order to evaluate the parameters space for the regression coefficients and to measure their sensibility, we performed a sensitivity analysis for each fuzzy regression coefficients, as also suggested by Golan et al. (1996) and Ciavolino and Dahlgaard (2009). The set of parameters to be evaluated was choose to vary between [−10 −5 0 5 10] and [−200 −100 0 100 200] (20 steps of analysis). Table 2 describes the results of the sensitivity analysis for only significant values of the original space analysis. For each support we reported the values for regression parameters and R^2_{pseudo}. Figure 2 shows the graphical pattern for each parameters analyzed. According to the features of sensitivity analysis we chose the support whose parameters with a stable R^2_{pseudo}. As we can notice by the Fig. 2, the stability of the R2's pattern was reached after few steps. With a conservative attitude we chose the support [−120 −60 0 60 120].

Table 1 Dataset for the study of unemployment rate for the year 2011

	x1	x2	c	l	r
Austria	2.00	13.00	2.00	1.00	2.00
Belgium	5.00	6.00	2.00	1.00	2.00
Czech Republic	4.00	8.00	2.00	1.00	2.00
Denmark	4.00	13.00	2.00	1.00	2.00
Estonia	12.00	8.00	6.00	1.00	3.00
Finland	5.00	11.00	2.00	1.00	2.00
France	6.00	7.00	6.00	1.00	3.00
Germany	4.00	13.00	2.00	1.00	2.00
Greece	9.00	1.00	11.00	1.00	4.00
Hungary	7.00	1.00	6.00	1.00	3.00
Ireland	10.00	4.00	6.00	2.00	3.00
Italy	5.00	2.00	2.00	1.00	2.00
Luxembourg	2.00	7.00	2.00	1.00	2.00
Netherlands	2.00	15.00	2.00	1.00	2.00
Norway	1.00	15.00	2.00	1.00	2.00
Poland	6.00	4.00	6.00	1.00	3.00
Portugal	7.00	7.00	6.00	1.00	3.00
Slovak Republic	10.00	4.00	6.00	1.00	3.00
Slovenia	4.00	7.00	2.00	2.00	2.00
Spain	15.00	3.00	11.00	1.00	4.00
Sweden	5.00	14.00	2.00	1.00	2.00
United Kingdom	5.00	12.00	2.00	1.00	2.00

Table 2 Results of sensitivity analysis

Vector of fixed points	α^c	β^c_{x1}	β^c_{x2}	α^l	β^l	α^r	β^r	R^2_{pseudo}
[−20 −10 0 10 20]	2.168	0.557	−0.173	1.087	0.019	1.553	0.232	0.819
[−40 −20 0 20 40]	2.325	0.545	−0.182	1.085	0.019	1.549	0.233	0.818
[−60 −30 0 30 60]	2.356	0.542	−0.184	1.085	0.019	1.548	0.233	0.818
[−80 −40 0 40 80]	2.367	0.542	−0.184	1.085	0.019	1.548	0.233	0.818
[−100 −50 0 50 100]	2.373	0.541	−0.185	1.085	0.019	1.548	0.233	0.818
[−120 −60 0 60 120]	2.375	0.541	−0.185	1.085	0.019	1.547	0.233	0.818

4.2 Results

In Table 3 we report the estimated values for the fuzzy components of unemployment rate 2011, the regression coefficients are described in Table 4 whereas the fit indices are shown in Table 5.

Unemployment rate 2011 seems to be positively related to the unemployment rate for the previous year ($\beta^c_{x1} = 0.54$) and inversely related to employment rate ($\beta^c_{x2} = -0.18$). Left spread does not show a relation with the independent variables ($\beta^l = 0.02$) whereas the right spread seems to be moderately related to the predictor variables ($\beta^r = 0.23$). The model had a good global fit: about of 82 % of total information was reproduced by the model estimated together with a low degree of model's entropy.

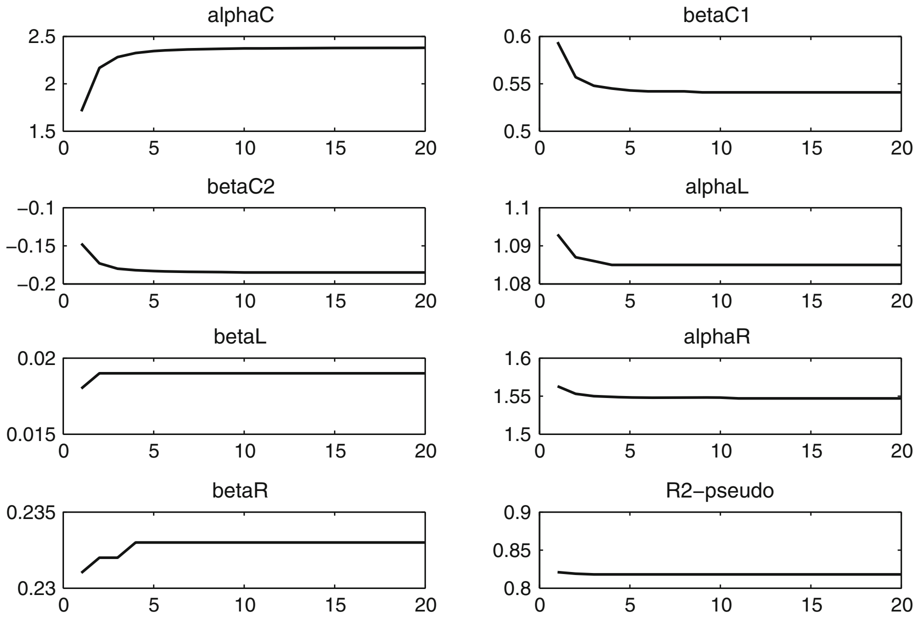

Fig. 2 Sensitivity analysis: pattern of stability for the regression parameters. On the *horizontal axis* are represented the 20 steps for the parameter space considered whereas on the *vertical* one are represented the range of the values for each parameter

Moreover, in order to evaluate the regression coefficients obtained we performed a bootstrap procedure with 5,000 re-sampled (see Table 6). From the t-test values and the confidence intervals obtained, the significant parameters are β^c_{x1}, α^l, α^r and β^r.

4.3 Comments and suggestions

By considering an economic point-of-view, these results seems to confirm the so-called *added workers effect* that can be explained by the following facts:

- As unemployment increases, the likelihood of being fired also increases and this stimulates job search activity by previously inactive individuals (Lundberg 1985).
- In the event of the payment of unemployment benefits, this can act an incentive for workers to improve the quality of their job search, by enabling them to refuse the early jobs firms offer them in the event these jobs are not in line with their skills or expectations. As a consequence of the increasing number of individuals searching a job, the unemployment rate at least in a short term perspective increases.
- Being unemployed causes, among other consequences, a loss of psychological well-being due to the loss of self esteem and esteem in society as a whole. Being employed is therefore supposedly preferable to the state of unemployment. This means that labour is not a pure dis-utility and hence that people prefers job search to inactivity (Spencer 2004). As a result, particularly in a condition of high unemployment, the increase of labour supply increases the unemployment rate.

In sum, unemployment for 2011 increased its values by increasing of the unemployment for 2010 and by decreasing of the employment for 2011, moreover it seemed to be a decreasing

Table 3 Estimated components for unemployment rate 2011

	c^*	l^*	r^*
Austria	1.06	1.11	1.79
Belgium	3.97	1.16	2.47
Czech Republic	3.06	1.14	2.26
Denmark	2.14	1.13	2.04
Estonia	7.39	1.23	3.27
Finland	3.05	1.14	2.26
France	4.33	1.17	2.55
Germany	2.14	1.13	2.04
Greece	7.06	1.22	3.19
Hungary	5.98	1.20	2.94
Ireland	7.05	1.22	3.19
Italy	4.71	1.18	2.64
Luxembourg	2.17	1.13	2.05
Netherlands	0.68	1.10	1.71
Norway	0.14	1.09	1.58
Poland	4.88	1.18	2.68
Portugal	4.87	1.18	2.68
Slovak Republic	7.05	1.22	3.19
Slovenia	3.25	1.15	2.30
Spain	9.93	1.28	3.86
Sweden	2.49	1.13	2.13
United Kingdom	2.86	1.14	2.21

Table 4 Regression coefficients for the study of unemployment rate 2011

α^c	β^c_{x1}	β^c_{x2}	α^l	β^l	α^r	β^r
2.37	0.54	−0.18	1.08	0.02	1.55	0.23

Table 5 Model statistics

R^2_{pseudo}	Entropy ratio
0.82	0.18

Table 6 Bootstrap results for the study of unemployment rate 2011

	Mean	Standard error	T test value	Bootstrap confidence intervals
α^c	2.08	2.71	0.77	[−1.8447 to 6.4429]
β^c_{x1}	0.57	0.28	2.03	[0.1638 to 0.997]
β^c_{x2}	−0.17	0.17	−0.95	[−0.4653 to 0.0845]
α^l	0.97	0.38	2.54	[0.6248 to 1.2267]
β^l	0.04	0.10	0.35	[0 to 0.1553]
α^r	1.51	0.17	8.82	[1.4943 to 1.5817]
β^r	0.24	0.06	4.30	[0.2235 to 0.2506]

trend for the unemployment predicted albeit this result is associated with a wide total spread (see for example the case of Spain in Table 3). From a decision-making point of view, seems to exist a considerable uncertainty about the classification of some countries (particularly, Spain, Estonia, Greece, Slovak Republic) in the low or middle level of the fuzzy unemployment scale.

5 Conclusion and further remarks

In this article we proposed a novel fuzzy regression model based on the rationale of the GME approach. Firstly, we discussed the main features of this approach of estimation, afterwards we described the GME fuzzy regression model. The proposed regression model was based on the well-known *generative fuzzy regression models* (D'Urso 2003) because of their simplicity and flexibility. We also describe two goodness of fit indices for the model, based on the principles of the entropy measure. In order to evaluate the features of our regression model, we described an economic case study based on a well-known economic model about employment and unemployment in OECD area (period 2010–2011). The results together with some suggestions about the case study were also described. In addition, in order to evaluate the stability of the results, we also performed a sensitivity analysis on the support vectors of the model's parameters. The parameters estimated were stable by varying their support space from $[-10\ -5\ 0\ 5\ 10]$ to $[-200\ -100\ 0\ 100\ 200]$. A bootstrap procedure was also carried out in order to evaluate the significance of the estimations obtained. Therefore, the results for the case study suggested us how the unemployment for 2011 is directly related to the unemployment for the previous year and it is inversely to the employment rate for the same year. Moreover, these results seems to confirm the so-called *added workers effect* theory.

In sum, this contribution would describe a novel approach to fuzzy regression model by taking the advantages of the GME estimator. In line with the results obtained, further works could extend the GME approach for more complex fuzzy regression models (e.g., fuzzy-input/fuzzy-output) as well as the develop of a comparative study with the traditional method of estimations.

Further studies should include simulation experiments in order to evaluate the impact of the number of independent variables, multicollinearity, missing data and outliers on the global performance of the model. Moreover, other developments could take into account models capable to interact with prior information or beliefs of the researches, by considering, cross-entropy based models (CE models) and weighted-entropy based models (WE models).

References

Benítez, J.M., Martín, J.C., Román, C.: Using fuzzy number for measuring quality of service in the hotel industry. Tour. Manag. **28**(2), 544–555 (2007)

Bernardini Papalia, R., Ciavolino, E.: Gme estimation of spatial structural equations models. J. Classif. **28**, 126–141 (2011)

Buckley, J.: Fuzzy Statistics, vol. 149. Springer, Heidelberg (2004)

Celmiņš, A.: Least squares model fitting to fuzzy vector data. Fuzzy Sets Syst. **22**(3), 245–269 (1987)

Chan, L., Kao, H., Wu, M.: Rating the importance of customer needs in quality function deployment by fuzzy and entropy methods. Int. J. Prod. Res. **37**(11), 2499–2518 (1999)

Chang, Y.H., Yeh, C.H.: A survey analysis of service quality for domestic airlines. Eur. J. Oper. Res. **139**(1), 166–177 (2002)

Cheng, H., Chen, J.: Automatically determine the membership function based on the maximum entropy principle. Inf. Sci. **96**(3), 163–182 (1997)

Ciavolino, E., Al-Nasser, A.: Comparing generalised maximum entropy and partial least squares methods for structural equation models. J. Nonparametr. Stat. **21**(8), 1017–1036 (2009)
Ciavolino, E., Dahlgaard, J.: Simultaneous equation model based onthe generalized maximum entropy for studying the effect of management factors on enterprise performance. J. Appl. Stat. **36**(7), 801–815 (2009)
Ciavolino, E., Salvatore, S., Calcagnì, A.: A fuzzy set theory based computational model to represent the quality of inter-rater agreement. Qual. Quant. **47**, 1–16 (2013)
Coppi, R.: Management of uncertainty in statistical reasoning: the case of aegression analysis. Int. J. Approx. Reason. **47**(3), 284–305 (2008)
Coppi, R., Gil, M.A., Kiers, H.A.L.: The fuzzy approach to statistical analysis. Comput. Stat. Data Anal. **51**, 1–14 (2006)
Diamond, P.: Fuzzy least squares. Inf. Sci. **46**(3), 141–157 (1988)
Dubois, D., Prade, H., Harding, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum, New York (1988)
D'Urso, P.: Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. Comput. Stat. Data Anal. **42**(1–2), 47–72 (2003)
D'Urso, P., Gastaldi, T.: A least-squares approach to fuzzy linear regression analysis. Comput. Stat. Data Anal. **34**(4), 427–440 (2000)
Golan, A.: Information and Entropy Econometrics–A Review and Synthesis, vol. 2. Now Publishers Inc., Norwell (2008)
Golan, A., Judge, G.: Maximum Entropy Econometrics Robust Estimation with Limited Data. Wiley, Chichester (1996)
Golan, A., Judge, G., Karp, L.: A maximum entropy approach to estimation and inference in dynamic models or counting fish in the sea using maximum entropy. J. Econ. Dyn. Control **20**(4), 559–582 (1996)
Golan, A., Karp, L., Perloff, J.: Estimating coke's and pepsi's price and advertising strategies. J. Bus. Econ. Stat. **18**(4), 398–409 (2000)
Guiasu, S., Shenitzer, A.: The principle of maximum entropy. Math. Intell. **7**(1), 42–48 (1985)
Hanss, M.: Applied Fuzzy Arithmetic. Springer, Berlin (2005)
Jaynes, E.: Information theory and statistical mechanics. Phys. Rev. **106**(4), 620 (1957)
Jaynes, E.: Prior probabilities. IEEE Trans. Syst. Sci. Cybern. **4**(3), 227–241 (1968)
Jaynes, E.: On the rationale of maximum-entropy methods. Proc. IEEE **70**(9), 939–952 (1982)
Kacprzyk, J., Fedrizzi, M.: Fuzzy Regression Analysis, vol. 1. Physica, Heidelberg (1992)
Kapur, J.: Maximum-Entropy Models in Science and Engineering. Wiley, New York (1989)
Lundberg, S.: The added worker effect. J. Labor. Econ. **3**, 11–37 (1985)
Medasani, S., Kim, J., Krishnapuram, R.: An overview of membership function generation techniques for pattern recognition. Int. J. Approx. Reason. **19**(3), 391–417 (1998)
Nguyen, H., Wu, B.: Fundamentals of Statistics with Fuzzy Data. Springer, Berlin (2006)
Nieradka, G., Butkiewicz, B.: A Method for Automatic Membership Function Estimation Based on Fuzzy Measures. Foundations of Fuzzy Logic and Soft Computing. Springer, Berlin (2007)
OECD: Employment rate. http://dx.doi.org/10.1787/emp-table-2011-1-en./content/table/emp-table-2011-1-en (2011).
OECD: Unemployment rates: as a percentage of labour force. http://dx.doi.org/10.1787/factbook-2013-table148-en./content/table/factbook-2013-table148-en (2013).
Overman, H., Puga, D.: Unemployment clusters across europe's regions and countries. Econ. Policy **17**(34), 115–148 (2002)
Patacchini, E., Zenou, Y.: Spatial dependence in local unemployment rates. J. Econ. Geogr. **7**(2), 169–191 (2007)
Pukelsheim, F.: The three sigma rule. Am. Stat. **48**(2), 88–91 (1994)
Ross, T.: Fuzzy Logic with Engineering Applications. Wiley, New York (2009)
Shannon, C., Weaver, W., Blahut, R., Hajek, B.: The Mathematical Theory of Communication, vol. 117. University of Illinois press, Urbana (1949)
Soofi, E.: A generalizable formulation of conditional logit with diagnostics. J. Am. Stat. Assoc. **87**(419), 812–816 (1992)
Spencer, D.: Deconstructing the labour supply curve. Metroeconomica **55**(4), 442–458 (2004)
Taheri, S.: Trends in fuzzy statistics. Austrian J. Stat. **32**(3), 239–257 (2003)
Tanaka, H.: Fuzzy data analysis by possibilistic linear models. Fuzzy Sets Syst. **24**(3), 363–375 (1987)
Verkuilen, J., Smithson, M.: Fuzzy Set Theory: Applications in the Social Sciences, vol. 147. Sage, London (2006)