Expert Systems With Applications 71 (2017) 230–239

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Fuzzy semi-parametric partially linear model with fuzzy inputs and fuzzy outputs

G. Hesamian[a,*], M.G. Akbari[b], M. Asadollahi[b]

[a] Department of Statistics, Payame Noor University, Tehran 19395-3697, Iran
[b] Department of Statistics, University of Birjand, Birjand 615–97175, Iran

ARTICLE INFO

Article history:
Received 19 July 2016
Revised 21 November 2016
Accepted 22 November 2016
Available online 26 November 2016

Keywords:
Fuzzy semi-parametric partially linear
Kernel method
Least absolute deviation
Optimal bandwidth
Fuzzy input
Fuzzy output
Non-fuzzy coefficient
Goodness-of-fit measure

ABSTRACT

A large number of accounting studies have focused on parametric or non-parametric forms of fuzzy regression relationships between dependent and independent variables. Notably, semi-parametric partially linear model as a powerful tool to incorporate statistical parametric and non-parametric regression analyses has gained attentions in many real-life applications recently. However, fuzzy data find application in many real studies. This study is an investigation of semi-parametric partially linear model for such cases to improve the conventional fuzzy linear regression models with fuzzy inputs, fuzzy outputs, fuzzy smooth function and non-fuzzy coefficients. For this purpose, a hybrid procedure is suggested based on curve fitting methods and least absolutes deviations to estimate the fuzzy smooth function and fuzzy coefficients. The proposed method is also examined to be compared with a common fuzzy linear regression model via a simulation data set and some real fuzzy data sets. It is shown that the proposed fuzzy regression model performs more convenient and efficient results in regard to six goodness-of-fit criteria which concludes that the proposed model could be a rational substituted model of some common fuzzy regression models in many practical studies of fuzzy regression model in expert and intelligent systems.

1. Introduction

Regression analysis is a statistical tool for investigating the relationships between independent (predictor or explanatory) variables and a dependent (response or outcome) variable. These methods are mainly divided into two general methods, so-called parametric and non-parametric methods. Parametric regression models are very useful for describing the relationship between the response and dependent variables. However, such methods may sometimes be applied at the risk of introducing modeling biases. Non-parametric regression models assume no prior model structure and may provide useful insight for further parametric fitting. However, being an entirely non-parametric approach, it usually causes some drawbacks such as the curse of dimensionality, difficulty of interpretation, and lack of extrapolation capability. However, semi-parametric partially linear models are more advantageous and are used in many applications since both the parametric and non-parametric components can simultaneously exist in the model. Partially linear models assume that the relationship between the response variables and the covariates can be represented as

$$y_i = \beta_1 x_i + f(t_i) + \epsilon_i, i = 1, 2, \dots, n,$$

where x_is are covariates, t_i's are increasing scalar covariates, the function f is unknown, and the model's errors ϵ_i's are independent random variables with mean zero. The models allow easier interpretation of the effect of each variable and may be preferable to a completely non-parametric model because of the well-known reason, "curse of dimensionality". On the other hand, partially linear models are more flexible than the standard linear model because they combine both parametric and non-parametric components; and when it is believed that the response variable Y depends on random variable X in a linear way, it is nonlinearly related to another scalar variable t. Several methods have been proposed to consider partially linear models, mainly including local linear technique (Hamilton & Truong, 1997), kernel method (Speckman, 1988) and profile least-square method (Green, Jennison, & Seheult, 1985; Opsomer & Ruppert, 1999).

Notably, to study of the relationship between a set of independent variables and one or more dependent variables, the data sometimes cannot be recorded precisely due to some unexpected situations. For instance, consider studying the functional relationship between atmospheric concentration of carbon monoxide (CO) and a set of meteorological variables in a city. For such a case, the

* Corresponding author.
E-mail addresses: gh.hesamian@pnu.ac.ir (G. Hesamian), g_-z_-akbari@birjand.ac.ir (M.G. Akbari), m.asadollahi@birjand.ac.ir (M. Asadollahi).

http://dx.doi.org/10.1016/j.eswa.2016.11.032
0957-4174/

carbon monoxide concentration usually report by fuzzy numbers instead of exact numbers. Fuzzy set theory seems to have suitable tools for modeling such data and provide appropriate statistical methods based on these data. The fuzzy regression analysis has been introduced by Tanaka, Hayashi, and Watada (1989) and then fuzzy regression methods have been successfully applied to many real applications under fuzzy data. Many methods followed parametric regression models based on fuzzy data which are mainly divided into two approaches including the possibilistic and fuzzy least-squares methods (for comprehensive review, see Chachi, Taheri, & Arghami, 2014). In addition, Poleshchuk and Komarov (2014) proposed a non linear regression model with interval type-2 fuzzy inputs and outputs based on the least squares estimation technique using an aggregation tool called weighted intervals. Arefi and Taheri (2015) proposed a linear regression model based on the least squares method when both explanatory variable(s) and response variable, as well as the parameters of model are assumed to be Atanassov's intuitionistic fuzzy numbers. Some other methods developed non-parametric approaches into regression analysis in a fuzzy environment (Cheng & Lee, 1999; Petit-Renaud & Denoeuxa, 2004; Wang, Zhang, & Mei, 2007). For more on other approaches to regression analysis including, non-parametric methods and some recent works see (Chachi et al., 2014; Cheng & Lee, 1999; Petit-Renaud & Denoeuxa, 2004; Wang et al., 2007).

However, as it is mentioned above, a semi-parametric partially linear model may be more flexible than the corresponding linear regression. Notably, to the best of the authors' knowledge, semi-parametric partially linear model technique has not yet been developed in applications on intelligence systems. Therefore, the main contribution of the present work is to investigate the effect of predicting fuzzy outputs by adding a fuzzy smooth function to the existing fuzzy linear regression models with fuzzy inputs and fuzzy outputs. In this regards, the present work is going to extend a common used regression technique, so-called "semi-parametric partially linear model" to fuzzy environment with fuzzy inputs, fuzzy outputs, fuzzy coefficients and fuzzy smooth function. A common two-step procedures is suggested to estimate non-fuzzy coefficients and fuzzy smooth function. For this purpose, both classical curve fitting including kernel methods and least absolute deviation are employed to estimate the components of the fuzzy regression model. An algorithm based on minimization of the cross-validation criterion is also applied to select the optimal bandwidth to estimate the fuzzy smooth function. To make a comparative study, a simulation analysis is performed between the proposed semi-parametric partially linear model and a common fuzzy parametric linear regression model and the effectiveness and advantages of the proposed methods are examined by comparing the proposed method in terms of some well-known goodness-of-fit criteria. For practical reasons, we will illustrate the proposed methods based on some real world data set. The numerical and comparative results showed that the proposed model is able to provide sufficiently accurate results in fuzzy regression analysis.

The rest of the present paper is organized as follows: Section 2 recalls some necessary concepts related to fuzzy numbers and distance measures between fuzzy numbers. In Section 3, the proposed semi-parametric partially linear model is introduced. The method of estimating the components of the model will be illustrated in this section through a common two-steps procedures. An algorithm based on the minimization of the cross-validation criterion is also applied in this section to determine the optimal bandwidth based on some distances among fuzzy numbers. Furthermore, a simulation study is performed to assess the effectiveness and performance of the proposed method with respect to some common methods of fuzzy linear regression in Section 4 using some performance criteria. Moreover, some real data sets are employed to explain the proposed fuzzy semi-parametric partially linear model in the comparison study. Finally, the main contributions of this paper is summarized in Section 5 and some concluding remarks concerning the possible developments pertaining to the present work are also made to be considered in further research.

2. Preliminary

This section briefly reviews several concepts and terminologies related to fuzzy numbers and a distance between fuzzy numbers used throughout the paper. Let $\mathbb{X}$ be a universal set. A fuzzy set of $\mathbb{X}$ is a mapping $\widetilde{A}:\mathbb{X}\to[0,1]$, which assigns a degree of membership $0\le\widetilde{A}(x)\le 1$ to each $x\in\mathbb{X}$. For each $\alpha\in(0,1]$, the subset $\left\{x\in\mathbb{X}\mid\widetilde{A}(x)\ge\alpha\right\}$ is called the α-cut of $\widetilde{A}$ and is denoted by $\widetilde{A}[\alpha]$. The set $\widetilde{A}[0]$ is also defined as equal to the closure of $\left\{x\in\mathbb{R}\mid\widetilde{A}(x)>0\right\}$. Let $\mathbb{R}$ be the set of all real numbers. A fuzzy set $\widetilde{A}$ of $\mathbb{R}$ is called a fuzzy number if it satisfies the following two conditions:

(1) For each $\alpha\in[0,1]$, the set $\widetilde{A}[\alpha]$ is a compact interval, which will be denoted by $[\widetilde{A}^L[\alpha],\widetilde{A}^U[\alpha]]$. Here, $\widetilde{A}^L[\alpha]=\inf\left\{x\in\mathbb{R}\mid\widetilde{A}(x)\ge\alpha\right\}$ and $\widetilde{A}^U[\alpha]=\sup\left\{x\in\mathbb{R}\mid\widetilde{A}(x)\ge\alpha\right\}$.
(2) There is a unique real number $x^*=x^*_{\widetilde{A}}\in\mathbb{R}$, such that $\widetilde{A}(x^*)=1$, i.e. $\widetilde{A}[1]$ is a singleton set.

The set of all fuzzy numbers with continuous membership functions is denoted by $\mathcal{F}(\mathbb{R})$. Notably, the most commonly used type of fuzzy numbers in $\mathcal{F}(\mathbb{R})$ are the so-called *LR*-fuzzy numbers denoted by $\widetilde{A}=(a;l_a,r_a)_{LR}$, $l_a, r_a>0$. The membership function of a *LR*-fuzzy number $\widetilde{A}$ is defined by:

$$\widetilde{A}(x)=\begin{cases}L\left(\dfrac{a-x}{l_a}\right), & x\le a,\\ R\left(\dfrac{x-a}{r_a}\right), & x>a,\end{cases}$$

where L and R are strictly decreasing functions from [0, 1] to [0, 1] satisfying $L(0)=R(0)=1$ and $L(1)=R(1)=0$. A special type of *LR*-fuzzy number is the triangular fuzzy number (**TFN**) with the shape functions $L(x)=R(x)=\max\{0,1-|x|\}$, $x\in\mathbb{R}$ which is denoted by $\widetilde{A}=(a;l_a,r_a)_T$. If $l_a=r_a$, then $\widetilde{A}$ is called a symmetric triangular fuzzy number.

Algebraic operations on fuzzy numbers that we use in this paper are defined based on the extension principle as follows.

Definition 2.1 (Lee, 2005)**.** Let $\widetilde{A}=(a;l_a,r_a)_{LR}$ and $\widetilde{B}=(b;l_b,r_b)_{LR}$ be two *LR*-type fuzzy numbers and $\lambda\in\mathbb{R}$. Then:

(1) $A\oplus B=(a+b;l_a+l_b,r_a+r_a)_{LR}$.
(2) $\lambda\otimes\widetilde{A}=(\lambda a;\lambda l_a,\lambda r_a)_{LR}$ if $\lambda\ge 0$ and $\lambda\otimes A=(\lambda a;-\lambda r_a,-\lambda l_a)_{RL}$ if $\lambda<0$.

Definition 2.2 (Hukuhara, 1967)**.** Let $\widetilde{A}$ and $\widetilde{B}$ be two *LR*-fuzzy numbers. If there exists a fuzzy number $\widetilde{C}$ such that $\widetilde{C}\oplus\widetilde{B}=\widetilde{A}$, then $\widetilde{C}$ is called Hukuhara difference of $\widetilde{A}$ and $\widetilde{B}$ and is denoted by $\widetilde{A}\ominus_H\widetilde{B}$.

The methods of measuring the distance between fuzzy numbers have an important role in decision making in many real applications like data mining, pattern recognition, multivariate data analysis and so on. Several distance measures on the space of fuzzy numbers have been suggested by many authors (see, for example, Bertoluzza, Corral, & Salas, 1995; Chen & Hsieh, 2000; Kaufmann & Gupta, 1991; Trutschnig, Gonzalez-Rodriguez, Colubi, & Gil, 2009). However, the absolute error distance is applied in this paper due to its good interpretation in many statistical applications which make it popular from the practical point of view (for more, see Heilpern, 1997). Let $\widetilde{A}$ and $\widetilde{B}$ be two *LR*-fuzzy numbers in $\mathcal{F}(\mathbb{R})$.