Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 128 (2016) 13–25
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Original articles

A fuzzy regression approach using Bernstein polynomials for the spreads: Computational aspects and applications to economic models

Antonio Francisco Roldán López de Hierro[a,*], Juan Martínez-Moreno[b], Concepción Aguilar Peña[c], Concepción Roldán López de Hierro[d]

[a] *Department of Quantitative Methods for Economics and Business, University of Granada, Granada, Spain*
[b] *Department of Mathematics, University of Jaén, Jaén, Spain*
[c] *Department of Statistics and Operations Research, University of Jaén, Jaén, Spain*
[d] *Department of Statistics and Operations Research, University of Granada, Granada, Spain*

Received 26 January 2014; received in revised form 2 September 2015; accepted 21 March 2016
Available online 6 April 2016

Abstract

An important research topic in applied statistics consists of determining the relationship between several variables through a regression function. Recently, *fuzzy regression analysis* has become important to deal with fuzzy data and vague information captured from the real world. When we are modeling relationships between an imprecise response and several real exploratory variables, one of the main difficulties is to guarantee the condition of non-negativity of the spreads. In this paper, due to their ease of implementation, continuous differentiability, and theoretical properties, Bernstein polynomials are used to develop a fuzzy regression procedure which guarantees this condition. We demonstrate the applicability and effectiveness of our method through the analysis of real data and comparisons with existing methodologies.

Keywords: Fuzzy regression; Fuzzy number; Bernstein polynomials; Goodness-of-fit

1. Introduction

Regression analysis is a powerful statistical tool which has many applications in different areas, such as Engineering, Environmental Sciences, Finance and Economics, Medicine, Biology, Psychology, etc. We realize that, in real problems, experimental data can usually be affected by an imprecision or uncertainty degree, and existing techniques must be extended in order to manage such type of information. In addition to this, in some research areas, it is less expensive to obtain imprecision data than precise measurements of a variable. Therefore, a very interesting issue is the modeling of these real problems and fuzzy theory [12,16,26,34,40,44] is a helpful tool to reach this goal. Currently,

* Corresponding author.
E-mail addresses: aroldan@ugr.es (A.F. Roldán López de Hierro), jmmoreno@ujaen.es (J. Martínez-Moreno), caguilar@ujaen.es (C. Aguilar Peña), iroldan@ugr.es (C. Roldán López de Hierro).

http://dx.doi.org/10.1016/j.matcom.2016.03.012

there is an increasing interest in dealing with fuzzy data and, in this context, fuzzy regression plays an important role. Fuzzy approaches have also shown to be relevant to many real-life problems. For instance, Rodger [29] developed a fuzzy statistical model based on nearest neighbor, regression and fuzzy logic implemented to a neural network to solve the problem of predicting demand for natural gas for the purpose of energy cost savings in public buildings.

It can be considered that the fuzzy regression appears in the history with the contributions of Tanaka, Uegima and Asai in 1982 (see [38]). In the first proposed fuzzy regression model, which is often referred to as possibilistic regression, the parameters were defined as fuzzy numbers and the problem was formulated and solved by means of the use of linear programming. The fuzzy data analysis described by Tanaka [35] in 1987 is a sort of fuzzy interval analysis which is an extension of interval analysis. In 1988, Tanaka and Watada [39] gave a new interpretation of fuzzy linear regression and developed a new method by which interval analysis can be done in fuzzy numbers and, in 1989, Tanaka, Hayashi and Watada [36] proposed different expressions of the criterion to be optimized and different formulations of the constraints to be satisfy. These models of possibilistic linear regression analysis formulated by Tanaka et al. in [38,35,36,39] assumed that possibilistic parameters are non-interactive, i.e., the joint possibilistic distribution of parameters is defined by minimum operators. Then Tanaka and Ishibuchi [37] extended their approach for obtaining interactive fuzzy parameters.

The original idea of the professors Tanaka and Asai and their colleagues gave a new prospective direction to fuzzy data analysis and provided a new and innovative approach to data analysis with potential extensions and applications. Since models of possibilistic linear regression analysis were formulated in [35–39], a lot of studies on extensions, generalizations and applications have been developed in this research topic (see, for instance, [2,3,5,11,17–19,24,25, 32,33,43]).

In general, the aim of this topic is to obtain a fuzzy regression function that can cover all the data given as samples. The complete specification of the model highly depends on the nature of input–output data and the problem sometimes becomes more complex and difficulties can be found in searching for optimal solutions, for non-linear problems specially. This has led most researchers to solve the problem in some particular cases. Although some papers are thus devoted to fuzzy input–fuzzy output data, most commonly, a mixed approach which considers crisp input and fuzzy output data is considered. The assumptions of symmetrical triangular fuzzy data and that the model is restricted to the linear case are very frequently used but has some limitations.

Among the developed methodologies, we are interested in those which interpret their optimal solutions as a regression function consisting in a model for centers and other models for spreads. These processes can handle a wider variety of models but a natural problem arises: the models for the spreads can only take non-negative values. At this point starts a large discussion in the context of fuzzy regression analysis. Inequality constrained least-squares estimation is considered a non-efficient method (see, for instance, [11,22]), and if we consider the iterative least squares estimation procedure proposed by Coppi et al. [3], which imposes non-negative conditions for the numerical minimization problem, the method becomes more complex and also less efficient. Another solution (see [11]) consists in transforming the constrained variable into an unconstrained one by means of logarithmical transformations of the non-negative values of the random sample to estimate a regression function, and later the inverse transformation is used in the final model. However, this solution can have several problems due to the natural behavior of the exponential function and because it is not always possible to fit a linear model for the transformed spreads nor any other model as we showed in [32].

The main aim of the present paper is to provide the researcher with a viable and simple way for analyzing regression relationships when the problem of non-negativity spreads appears. In this paper we present a new fuzzy methodology based on the good properties that Bernstein polynomials verify, which have intensively been studied in recent times, mainly in the framework of approximation theory (see [10] for a comprehensive retrospective). They are also attractive in statistical problems due to their simple probabilistic nature (see, for instance, [4,27,41]). The best advantages of this new procedure are the following: (1) the non-negativity condition over the estimated spread functions (when other papers proposed complex and less efficient methods) is easily guaranteed; (2) we could consider linear and non-linear regression models; (3) models for spreads are polynomials, so they can be easily calculated on any computer (few time and resources are necessary); (4) when the degree of the polynomial spread functions increases, the fuzzy regression model better explains the initial data (but even when the degree is low, this method lets us obtain better results than other methodologies); (5) a large variety of fuzzy numbers can be considered as output data taking into account the following property: it is always possible to consider an approximation operator which produces a trapezoidal fuzzy number that could be the closest to the given original fuzzy numbers among all trapezoidal fuzzy numbers having an expected interval identical to the original (see [15]).