ELSEVIER
Computational Statistics & Data Analysis 42 (2003) 47–72
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data

Pierpaolo D'Urso

Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università degli Studi di Roma "La Sapienza", P.le Aldo Moro, 5, I-00185 Roma, Italy

Received 1 August 2001; received in revised form 1 February 2002

Abstract

In order to estimate fuzzy regression models, possibilistic and least-squares procedures can be considered. By taking into account a least-squares approach, regression models with crisp or fuzzy inputs and crisp or fuzzy output are suggested. In particular, for these fuzzy regression models, unconstrained and constrained (with inequality restrictions) least-squares estimation procedures are developed. Furthermore, for the various models presented, explanatory examples are shown and some concluding remarks are also included. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Fuzzy/crisp input data; Fuzzy/crisp output data; Linear regression analysis; Unconstrained and constrained least-squares estimates

1. Introduction

The fuzzy regression analysis has been introduced by Tanaka et al. (1980,1982). Successively, several works have been published by different authors (see, for instance, Kacprzyk and Fedrizzi, 1992; Sakawa and Yano, 1992; Chang and Lee, 1996; Näther, 1994; Diamond and Körner, 1997; Ming et al., 1997; Kim and Bishu, 1998; Tanaka and Lee, 1998; Wang and Tsaur, 2000; D'Urso and Gastaldi, 2000,2001,2002).

The fuzzy regression models can be classified in two classes. The first class is based on the *possibilistic* concepts (see, for instance, Tanaka et al., 1980,1982; Chang et al., 1996; Tanaka and Lee, 1998; Chen, 2001; Lee and Chen, 2001) and the second one on the *least-squares* approach (Celminš, 1987a,b; Diamond, 1988; Diamond and

E-mail address: pierpaolo.durso@uniroma1.it (P. D'Urso).

PII: S0167-9473(02)00117-2

Kloeden, 1994; Chang and Lee, 1996; Körner and Näther, 1998; D'Urso and Gastaldi, 2000,2001,2002; Chang, 2001; Ruoning and Li, 2001; Wu, 2002). In both approaches, the notion of "best fit" incorporates the optimisation of a functional associated with the problem. In particular, in the *possibilistic approach*, "this functional takes the form of a measure of the spreads of the estimated output, either as a weighted linear sum involving the estimated coefficients in linear regression, or as quadratic form in the case of exponential possibilistic regression" (Diamond and Tanaka, 1998). In the *least-squares approach*, "the functional to be minimized is an L_2 distance between the observed and estimated outputs. This reduces to a class of quadratic optimization problems and constrained quadratic optimization" (Diamond and Tanaka, 1998).

It should be underlined that the previous distinction has a typological character. It is not implied that *hybrid* criteria (Chang, 2001; Chang and Ayyub, 2001) cannot be devised.

We notice that, in a recent paper, Chang and Ayyub (2001) have carried out a comparative analysis between some methods related to the two previous approaches.

In this work, we consider different fuzzy regression models that can be included in the least-squares approach in which the functional relationship is crisp and the possible structures of the data are *crisp input-fuzzy output*, *fuzzy input-crisp output* and *fuzzy input-fuzzy output*. We prefer the least-squares approach because the possibilistic approach (in which linear programming methods are utilized to estimate the regression coefficients) presents some difficulties in the estimation procedures. For instance, as noticed by Chang and Ayyub (2001), "As the number of data sets increase, the number of constraints [of the linear programming method] increases proportionally. This increase might result in computational difficulties using LP [Linear Programming] software or computers". Other inconveniences are remarked by these Authors in the same paper.

The least-squares approach "involves ideas of goodness-of-fit and this requires a notion of distance (Diamond and Kloeden, 1994) between the fuzzy values predicted by a parametric model and the fuzzy data that is actually observed. The principal advantage of these techniques is that the residual gives some idea of the accuracy of the model" (Diamond and Tanaka, 1998). Following this approach, several proposals have been made. We refer to the previous references on least-squares fuzzy regression models or to Diamond and Tanaka (1998), who analyse systematically the more significant contributions within this approach.

The models proposed in the least-squares approach analyse, essentially, two situations: fuzzy inputs-fuzzy output and crisp inputs-fuzzy output. In the next sections, we study in detail all the possible combinations (fuzzy/crisp inputs-fuzzy/crisp output), following the formulation suggested by D'Urso and Gastaldi (2000) for the fuzzy regression problem. In particular, after having defined the so-called LR_1 and LR_2 fuzzy data (Section 2), in Sections 3, 5 and 7, respectively, linear regression analysis for crisp inputs and LR_1 fuzzy output, for LR_1 fuzzy inputs and crisp output and for LR_1 fuzzy inputs and output are proposed. Linear regression models for LR_2 fuzzy output are suggested in Section 9. In Sections 4, 6, 8, 10 applications of the different fuzzy regression models are carried out. Finally, Section 11 contains detailed concluding remarks and some possible future perspectives.