ELSEVIER Computational Statistics & Data Analysis 42 (2003) 203–217

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Fuzzy estimates of regression parameters in linear regression models for imprecise input and output data

Hsien-Chung Wu

Department of Information Management, National Chi Nan University, Puli, Nantou 545, Taiwan

Received 1 January 2002; accepted 2 March 2002

Abstract

The method for obtaining the fuzzy estimates of regression parameters with the help of "Resolution Identity" in fuzzy sets theory is proposed. The α-level least-squares estimates can be obtained from the usual linear regression model by using the α-level real-valued data of the corresponding fuzzy input and output data. The membership functions of fuzzy estimates of regression parameters will be constructed according to the form of "Resolution Identity" based on the α-level least-squares estimates. In order to obtain the membership degree of any given value taken from the fuzzy estimate, optimization problems have to be solved. Two computational procedures are also provided to solve the optimization problems. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Linear regression analysis; Fuzzy numbers; Fuzzy estimate; Optimization; Confidence degree

1. Introduction

In the real world, the data sometimes cannot be recorded or collected precisely. For instance, the water level of a river cannot be measured in an exact way because of the fluctuation, and the temperature in a room is also not able to be measured precisely because of the similar reason. Therefore, the fuzzy sets theory is naturally to be an appropriate tool in modeling the statistical models when the fuzzy data have been observed. The more appropriate way to describe the water level is to say that the water level is around 30 m. The phrase "around 30 m" can be regarded as a fuzzy number $\widetilde{30}$. This is the main concern of this paper.

E-mail address: hcwu@im.ncnu.edu.tw (H.-C. Wu).

PII: S0167-9473(02)00116-0

Since Zadeh (1965) introduced the concept of fuzzy sets, the applications of considering fuzzy data to the regression models have been proposed in the literature. Tanaka et al. (1982) initiated this research topic. They also generalized their approaches to the more general models in Tanaka and Warada (1988), Tanaka et al. (1989), Tanaka and Ishibuchi (1991). The collection of papers edited by Kacprzyk and Fedrizzi (1992) gave an insightful survey.

In the approach of Tanaka et al. (1982), they considered the *L-R* fuzzy data and minimized the index of fuzziness of the fuzzy linear regression model. Yager (1982) used a linguistic variable to represent imprecise information for the regression models. Moskowitz and Kim (1993) also proposed a method to assess the H-value in a fuzzy linear regression model proposed by Tanaka et al. (1982). Redden and Woodall (1994) compared various fuzzy regression models and gave the differences between the approaches of fuzzy regression analysis and usual regression analysis. They also pointed out some weakness of the approaches proposed by Tanaka et al. Chang and Lee (1994) also pointed out another weakness of the approaches proposed by Tanaka et al. Wang and Tsaur (2000) also proposed a new model to improve the predictability of Tanaka's model. Bárdossy (1990) proposed many different measures of fuzziness which must be minimized with respect to some suggested constraints. Peters (1994) introduced a new fuzzy linear regression model based on Tanaka's approach by considering the fuzzy linear programming problem. Diamond (1988) introduced a metric on the set of fuzzy numbers by invoking the Hausdorff metric on the compact α-level sets, and used this metric to define a least-squares criterion function as in the usual sense, which must be minimized. Ma et al. (1997) generalized Diamond's approach by embedding the set of fuzzy numbers into a Banach space isometrically and isomorphically. Näther (1997,2000), Näther and Albrecht (1990) and Körner and Näther (1998) introduced the concept of random fuzzy sets (fuzzy random variables) into the linear regression model, and developed an estimation theory for the parameters. Chang and Ayyub (2001) gave the differences between the fuzzy regression and ordinary regression analysis and also Kim et al. (1996) compared both fuzzy regression and statistical regression conceptually and empirically. Chang (2001) proposed a method for hybrid fuzzy least-squares regression by defining the weighted fuzzy-arithmetic and using the well-accepted least-squares fitting criterion. Celminš (1987,1991) proposed a methodology for the fitting of differentiable fuzzy model function by minimizing a least-squares objective function. Chang and Lee (1996) proposed a fuzzy regression technique based on the least-squares approach to estimate the modal value and the spreads of *L-R* fuzzy number. Dunyak and Wunsch (2000) described a method for nonlinear fuzzy regression using a special training technique for fuzzy number neutral networks. D'Urso and Gastaldi (2000) proposed a doubly linear adaptive fuzzy regression model based on a core regression model and a spread regression model. D'Urso (2002) also developed the unconstrained and constrained least-squares estimation procedures. Jajuga (1986) calculated the linear fuzzy regression coefficients using a generalized version of the least-squares method by considering the fuzzy classification of a set of observations and obtaining the homogeneous classes of observations. Kim and Bishu (1998) used a criterion of minimizing the difference of the membership degrees between the observed and estimated fuzzy numbers. Sakawa and Yano (1992) introduced three