An International Journal
computers & mathematics
with applications

PERGAMON Computers and Mathematics with Applications 45 (2003) 1849–1859
www.elsevier.com/locate/camwa

Linear Regression Analysis for Fuzzy Input and Output Data Using the Extension Principle

HSIEN-CHUNG WU
Department of Information Management
National Chi Nan University, Puli, Nantou 545, Taiwan, R.O.C.
hcwu@im.ncnu.edu.tw

(Received and accepted April 2002)

Abstract—The method for obtaining the fuzzy least squares estimators with the help of the extension principle in fuzzy sets theory is proposed. The membership functions of fuzzy least squares estimators will be constructed according to the usual least squares estimators. In order to obtain the membership value of any given value taken from the fuzzy least squares estimator, optimization problems have to be solved. We also provide the methodology for evaluating the predicted fuzzy output from the given fuzzy input data.

Keywords—Fuzzy numbers, Fuzzy least squares estimator, Optimization.

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 a similar reason. Therefore, the fuzzy sets theory is naturally 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 meters. The phrase "around 30 meters" can be regarded as a fuzzy number $\widetilde{30}$. This is the main concern of this paper.

Since Zadeh [1] 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.* [2] initiated this research topic. They also generalized their approaches to the more general models in [3–5].

In the approach of Tanaka *et al.* [2], they considered the L-R fuzzy data and minimized the index of fuzziness of the fuzzy linear regression model. Redden and Woodall [6] compared various fuzzy regression models and gave the difference between the approaches of fuzzy regression analysis and usual regression analysis. They also pointed out some weaknesses of the approaches proposed by Tanaka *et al.* Chang and Lee [7] also pointed out another weakness of the approaches proposed by Tanaka *et al.* Bárdossy [8] proposed many different measures of fuzziness which must be minimized with respect to some suggested constraints. Peters [9] introduced a

0898-1221/03/$ - see front matter Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX
PII: S0898-1221(03)00175-5

new fuzzy linear regression model based on Tanaka's approach by considering the fuzzy linear programming problem. Diamond [10] 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.* [11] generalized Diamond's approach by embedding the set of fuzzy numbers into a Banach space isometrically and isomorphically. Näther and Albrecht [12] and Körner and Näther [13] introduced the concept of random fuzzy sets (fuzzy random variables) into the linear regression model, and developed an estimation theory for the parameters. The other interesting references are also given in [14–26].

In this paper, we will construct the fuzzy least squares estimators using the extension principle in fuzzy sets theory which was introduced by Zadeh [27–29]. The membership functions of fuzzy least squares estimators will be constructed according to the usual least squares estimators with the help of the extension principle. In order to obtain the membership value of any given value taken from the fuzzy least squares estimator, optimization problems have to be solved. We also provide the methodology for evaluating the predicted fuzzy output from the given fuzzy input data.

In Section 2, we give some properties of fuzzy numbers. In Section 3, we give some useful results from the extension principle. In Section 4, we conduct the membership functions of fuzzy least squares estimators according to the usual least squares estimators with the help of the extension principle. In Section 5, we will develop the computational procedures to obtain the membership value of any given value taken from the fuzzy least squares estimators. We also provide an example to clarify the theoretical results, and show the possible applications in linear regression analysis for fuzzy data. In Section 6, the methodology for transacting the predicted fuzzy output from the given fuzzy input data is proposed.

2. FUZZY NUMBERS

Let X be a universal set. Then a fuzzy subset $\tilde{A}$ of X is defined by its membership function $\xi_{\tilde{A}} : X \to [0,1]$. We denote by $\tilde{A}_\alpha = \{x : \xi_{\tilde{A}}(x) \geq \alpha\}$ the α-level set of $\tilde{A}$, where $\tilde{A}_0$ is the closure of the set $\{x : \xi_{\tilde{A}}(x) \neq 0\}$. $\tilde{A}$ is called a normal fuzzy set if there exists an x such that $\xi_{\tilde{A}}(x) = 1$. $\tilde{A}$ is called a convex fuzzy set if $\xi_{\tilde{A}}(\lambda x + (1-\lambda)y) \geq \min\{\xi_{\tilde{A}}(x), \xi_{\tilde{A}}(y)\}$ for $\lambda \in [0,1]$. (That is, $\xi_{\tilde{A}}$ is a quasi-concave function.)

In this paper, the universal set X is assumed to be a real number system; that is, $X = \mathbf{R}$. Let f be a real-valued function defined on $\mathbf{R}$. f is said to be upper semicontinuous if $\{x : f(x) \geq \alpha\}$ is a closed set for each α. Or equivalently, f is upper semicontinuous at y if and only if $\forall \epsilon > 0$, $\exists \delta > 0$ such that $|x - y| < \delta$ implies $f(x) < f(y) + \epsilon$.

$\tilde{a}$ is called a fuzzy number if the following conditions are satisfied.

(i) $\tilde{a}$ is a normal and convex fuzzy set.
(ii) Its membership function $\xi_{\tilde{a}}$ is upper semicontinuous.
(iii) The α-level set $\tilde{a}_\alpha$ is bounded for each $\alpha \in [0,1]$.

From Zadeh [1], $\tilde{A}$ is a convex fuzzy set if and only if its α-level set $\tilde{A}_\alpha = \{x : \xi_{\tilde{A}}(x) \geq \alpha\}$ is a convex set for all α. Therefore, if $\tilde{a}$ is a fuzzy number, then the α-level set $\tilde{a}_\alpha$ is a compact (closed and bounded in $\mathbf{R}$) and convex set; that is, $\tilde{a}$ is a closed interval. The α-level set of $\tilde{a}$ is then denoted by $\tilde{a}_\alpha = [\tilde{a}_\alpha^L, \tilde{a}_\alpha^U]$. We also see that $\tilde{a}_\alpha^L$ and $\tilde{a}_\alpha^U$ are continuous with respect to α, since its membership function is upper semicontinuous. The following proposition is useful for further discussions.

PROPOSITION 2.1. RESOLUTION IDENTITY. *(See [27–29].) Let $\tilde{A}$ be a fuzzy set with membership function $\xi_{\tilde{A}}$ and the α-level set $\tilde{A}_\alpha = \{x : \xi_{\tilde{A}}(x) \geq \alpha\}$ be given. Then*

$$\xi_{\tilde{A}}(x) = \sup_{\alpha \in [0,1]} \alpha \cdot 1_{\tilde{A}_\alpha}(x).$$