

A Regression Model Based on Uncertain Set

Xiaona Li¹, Xiaosheng Wang¹, Sundaram Sampath², Mingchao Li¹ & Jiawei Wang³

¹School of Mathematics and Physics, Hebei University of Engineering, Handan, China

²Department of Statistics, University of Madras, Chennai, India

³School of Mathematics and Information Science, Shaanxi Normal University, Xi'an, China

Correspondence: Xiaosheng Wang, School of Mathematics and Physics, Hebei University of Engineering, Handan 056038, China.

Received: October 17, 2017

Accepted: November 6, 2017

Online Published: November 10, 2017

doi:10.20849/abr.v2i3.215

URL: <https://doi.org/10.20849/abr.v2i3.215>

Abstract

Traditional regression analysis is a method of statistical data analysis based on probability theory. Regression models play crucial roles in various branches of statistics including design of experiments, econometrics etc. In regression models, the dependent variable is assumed to be of stochastic nature where randomness enters via errors. Further, the independent variables are assumed to be of deterministic nature. The regression coefficients which explain the interdependency between the variables are assumed to be crisp quantities. Whenever, difficulty arises in expressing the values taken by the dependent variable in terms of crisp quantities, traditional regression models become irrelevant. This paper provides a framework for dealing with such situations on using the notion of uncertain sets of various forms. In this paper, a solution for this problem obtained via linear programming technique is introduced along with an illustrative example.

Keywords: uncertainty theory, uncertain set, parametric estimation, regression analysis

1. Introduction

Regression analysis is one of the major branches of statistical theory. Regression analysis aims at analyzing the extent of interdependency between variables by making use of the observed data. Further, models developed by using the experimental data are used for several tasks including forecasting and decision making.

In a regression model, some are called independent variables and others are named the dependent variables. The dependent variables change along with the changes of the independent variables. In the traditional regression model, the dependent variable is regarded as random variable because of the random disturbance term. Researchers have made significant contributions on the theoretical aspects as well as the applications of tools available in linear models in various fields. Regression analysis assumes the experimental data as well as the predicted values of the dependent variables are of crisp nature. And in the application of the model, it is needful to meet the following two conditions. On one hand, each independent repeated experiment is conducted under the same conditions; on the other hand, the experimental number is large enough. However, in many practical problems, the above conditions cannot be met and there also exists nondeterminacy in many systems which leads to the result that it contains language data in input or output. So under these circumstances, people began to pay attention to uncertain data such as fuzzy data to describe these phenomena.

The development of fuzzy set theory (Zadeh, 1965) created alternative directions for developing models which are similar to regression models using imprecise or vague data. In literature such models are referred to as fuzzy regression models. The notion of fuzzy regression analysis was initiated by Tanaka et al. (1982) and the linear model of fuzzy regression analysis was established. A fuzzy functional relationship was given between independent variables and dependent variables in fuzzy regression model. The input data and unknown coefficients may be crisp or fuzzy and the predicted values of dependent variables will be of fuzzy nature. Research in the direction of developing fuzzy linear regression models has received the attention of several researchers during the past three decades. Once the functional relationship is determined, the main work is to estimate the parameters. For different types of fuzzy numbers, Diamond (1988, 1997, 1987) and Korner (1998) used the least square approach in estimating the parameters involved in the model. And Su, Wang and Wang (2013) investigated parametric regression analyses which includes both linear and nonlinear of imprecise data by using the fuzzy evidential EM algorithm. Based on the possibilistic approach, verny et al. (2013) studied the

linear regression models for uncertain, indeterminate or interval data taking in to account the loss of information. Based on fuzzy neural networks, Ishibuchi et al. (2001) analyzed the fuzzy regression model by bringing in asymmetric fuzzy coefficients. Muller et al. (2014) introduced an algorithm for identification and estimation of relevant parameters in an optimized manner. Danesh et al. (2016) introduced the adaptive neuro-fuzzy inference system (ANFIS) which is used for fuzzy nonparametric regression function prediction where input and output are crisp and fuzzy respectively. Chen et al. (2016) have developed an approach to optimize the value of membership degree h of fuzzy linear regression whose coefficients are asymmetric triangular fuzzy data. In addition, Sclove (2014) improved the method of estimating the coefficients in an orthogonal linear regression model by point estimation which is more efficient than the ordinary.

Linear regression models are applicable only when distributional assumptions made in their construction via probability theory. They become irrelevant when randomness is replaced by impreciseness in the system being studied. Researchers working on fuzzy sets have developed alternative approaches to handle such situations. However, lack of laws of excluded middle and contradiction creates suspicion among practitioners in accepting the conclusions drawn from such fuzzy models. To overcome this problem, Liu (2007) initiated a new branch of study, namely, uncertainty theory in 2007 and made refinements on the same in 2010.

Uncertainty theory is established analogous to measure theory and now it has developed into a branch of axiomatic mathematics. Liu (2010) introduced uncertain statistics which is a methodology employed to collect and interpret experts' experimental data via uncertainty theory when no samples are obtainable. The problem of estimating parameters in uncertainty distributions with known functional form is one of the several research problems tackled in uncertain statistics. Towards addressing this problem, Liu (2007) introduced the concept of uncertainty distribution. Wang et al. (2012) considered Delphi method for estimating uncertainty distribution based on multiple domain experts data. A method based on the principle of least squares was suggested by Liu (2007) for estimating the parameters of uncertainty distribution. Later, a new method namely method of moments was suggested by Wang and Peng (2014). Besides, a statistical method called uncertain hypothesis testing was suggested by Wang et al. (2012) in order to test whether two uncertainty distributions are equal.

To describe the relationship between variables involved in uncertainty phenomena, Liu (2007) proposed an uncertain regression model based on uncertain variable. In the proposed model, uncertainty comes from the error component which is treated as an uncertain variable. It was assumed that the model receives as input crisp data and the output will be in terms of uncertain variables. Furthermore, an uncertain linear regression model was studied by Guo et al. (2017) and it was applied to predict China's GDP. Apart from these, for situations in which the data gathered from experts' knowledge is of imprecise form, Guo et al. (2011) came out with an uncertain regression model with an intrinsic error structure driven by uncertain canonical process. In the above-mentioned uncertain regression model, the observed values of variables are real numbers. However, in fact, the observations are with the characteristic of multiple memberships and maybe the sample size is small. Therefore, we can turn our attention to uncertain set which can be applied to describe uncertainty phenomenon. It is pertinent to note the difference between fuzzy and uncertain sets. While fuzzy sets use the concept of possibility measures, uncertain sets use the uncertain measure. It is to be highlighted that uncertain sets give room for accommodating several properties which include independence of uncertain sets as well. Research contributions related to uncertain sets are reviewed in the forthcoming section.

The main objective of the paper is to introduce a regression model based on uncertain sets. Apart from this the present work considers and evaluates methods of estimation of parameters in the proposed model. Uncertain linear regression model based on symmetrical triangular, symmetrical trapezoidal and normal uncertain set are considered respectively. A procedure based on linear programming has been suggested for estimating the parameters in the uncertain linear models and appropriate performance evaluation criteria are also considered. Section organization of the paper is as follows. The second section of this paper gives a brief introduction to uncertain theory as well as uncertain statistics. Uncertain linear regression model using uncertain sets and the problem of estimating parameters involved in it are considered in the third section. An illustrative example is given in the fourth section and conclusions drawn from this study are given in the fifth section.

2. Preliminaries

In this section, we will introduce some fundamental definitions and theorems in uncertainty theory and uncertain statistics.

2.1 Uncertainty Theory

Definition 1. (Liu, 2007) Let Γ be a nonempty set, and L be a σ -algebra over Γ . Each element $\Lambda \in L$ is called

an event. A number $M\{\Lambda\}$ denotes the level that Λ will occur. Then M is called an uncertain measure if it satisfies the following axioms:

Axiom 1: (Normality Axiom) $M\{\Gamma\}=1$.

Axiom 2: (Duality Axiom) $M\{\Lambda\}+M\{\Lambda^c\}=1$ for any event Λ .

Axiom 3: (Subadditivity Axiom) For every countable sequence of events $\{\Lambda_i\}$, we have $M\{\bigcup_{i=1}^{\infty}\Lambda_i\}\leq\sum_{i=1}^{\infty}M\{\Lambda_i\}$.

Note that the triplet (Γ, L, M) is called an uncertainty space.

Besides, the product uncertain measure on the product σ -algebra L was defined by Liu (2009) as follows.

Axiom 4: (Product Axiom) Let (Γ_k, L_k, M_k) be uncertainty spaces for $k = 1, 2, \dots$. The product uncertain measure M is an uncertain measure satisfying $M\{\prod_{k=1}^{\infty}\Lambda_k\}=\bigwedge_{k=1}^{\infty}M_k\{\Lambda_k\}$, where Λ_k are arbitrarily chosen events from L_k .

The concept of uncertain variable ξ was introduced by Liu as a measurable function from an uncertainty space (Γ, L, M) to the set of real numbers. In order to describe an uncertain variable, uncertainty distribution was defined.

Definition 2. (Liu, 2007) An uncertain variable is a measure function ξ from an uncertainty space (Γ, L, M) to the set of real number. That is, for any Borel set B , the set $\{\xi \in B\} = \{\gamma \in \Gamma \mid \xi(\gamma) \in B\}$ is an event.

Definition 3. (Liu, 2007) The uncertainty distribution Φ of an uncertain variable ξ is $\Phi(x) = M\{\xi \leq x\}$, $\forall x \in R$.

Definition 4. (Liu, 2010) Let ξ be an uncertain variable with a regular uncertainty distribution $\Phi(x)$. Then $\Phi^{-1}(\alpha)$ is called the inverse uncertainty distribution of ξ .

2.2 Uncertain Set

In order to model the concepts whose boundaries are not sharp because of the ambiguous human language, uncertain set was proposed by Liu in 2010.

Definition 5. (Liu, 2010) An uncertain set is a function ξ from an uncertainty space (Γ, L, M) to a collection of sets of real numbers such that both $\{B \subset \xi\}$ and $\{\xi \subset B\}$ are events for any Borel set B of real numbers.

Definition 6. (Liu, 2012) An uncertain set ξ is said to have a membership function μ if for any Borel set B of real numbers, we have $M\{B \subset \xi\} = \inf_{x \in B} \mu(x)$ and $M\{\xi \subset B\} = \sup_{x \in B^c} \mu(x)$.

Definition 7. (Liu, 2007) Let $\xi_1, \xi_2, \dots, \xi_n$ be uncertain sets on the uncertainty space (Γ, L, M) , and let f be a measurable function. Then $\xi = f(\xi_1, \xi_2, \dots, \xi_n)$ is an uncertain set defined by $\xi(\gamma) = f(\xi_1(\gamma), \xi_2(\gamma), \dots, \xi_n(\gamma))$ for any $\gamma \in \Gamma$.

Definition 8. (Liu, 2007) An uncertain set ξ is called triangular if it has a membership function

$$\mu(x) = \begin{cases} x-a/b-a, & \text{if } a \leq x \leq b \\ x-c/b-c, & \text{if } b \leq x \leq c \end{cases}$$

denoted by (a, b, c) where a, b, c are real numbers with $a < b < c$.

ξ is called symmetrical triangular uncertain set if and only if $c - b = b - a$. Then $c - b$ or $b - a$ is named radius of ξ and b is the center of ξ .

Definition 9. (Liu, 2007) An uncertain set ζ is called trapezoidal if it has a membership function

$$\mu(x) = \begin{cases} x-a/b-a, & \text{if } a \leq x \leq b \\ 1, & \text{if } b \leq x \leq c \\ x-d/c-d, & \text{if } c \leq x \leq d \end{cases}$$

denoted by (a, b, c, d) where a, b, c, d are real numbers with $a < b < c < d$.

ζ is symmetrical trapezoidal uncertain set if and only if $d - c = b - a$. Then $d - c$ or $b - a$ is named radius of ζ .

Definition 10. (Guo, 2014) An uncertain set ζ is called normal if it has a membership function

$$\mu(x) = \exp\left(-\frac{(x-a)^2}{2b^2}\right)$$

denoted by $N(a, b)$ where a and b are real numbers with $b > 0$. Especially, a normal uncertain set $N(a, b)$ is called standard if the parameters $a = 0$ and $b = 1$; denoted by $N(0, 1)$. The parameter a is called the center of ζ .

Definition 11. (Liu, 2013) The uncertain sets $\zeta_1, \zeta_2, \dots, \zeta_n$ are said to be independent if for any Borel sets B_1, B_2, \dots, B_n of real numbers, we have

$$M\left\{\bigcap_{i=1}^n (\zeta_i^* \subset B_i)\right\} = \bigwedge_{i=1}^n M\{\zeta_i^* \subset B_i\}$$

$$M\left\{\bigcap_{i=1}^n (\zeta_i^* \subset B_i)\right\} = \bigwedge_{i=1}^n M\{\zeta_i^* \subset B_i\}, \quad M\left\{\bigcup_{i=1}^n (\zeta_i^* \subset B_i)\right\} = \bigvee_{i=1}^n M\{\zeta_i^* \subset B_i\}$$

where ζ_i^* are arbitrarily chosen from $\{\zeta_i, \zeta_i^c\}, i = 1, 2, \dots, n$, respectively.

Definition 12. (Liu, 2012) Let ζ be an uncertain set with membership function μ . Then the set-valued function

$$\mu^{-1}(\alpha) = \{x \in R | \mu(x) \geq \alpha\}, \quad \forall \alpha \in [0, 1]$$

is called the inverse membership function of ζ . For each given α , the set $\mu^{-1}(\alpha)$ is also called the α -cut of μ .

Example 1. The triangular uncertain set $\zeta = (a, b, c)$ has an inverse membership function

$$\mu^{-1}(\alpha) = [(1 - \alpha) + \alpha b, \alpha b + (1 - \alpha)c].$$

Example 2. The trapezoidal uncertain set $\zeta = (a, b, c, d)$ has an inverse membership function

$$\mu^{-1}(\alpha) = [(1 - \alpha) + \alpha b, \alpha c + (1 - \alpha)d].$$

Example 3. The inverse membership functions of normal uncertain set $N(a, b)$ and the standard normal uncertain set $N(0, 1)$ are $\mu^{-1}(\alpha) = [a - b\sqrt{-2\ln\alpha}, a + b\sqrt{-2\ln\alpha}]$ and $\mu^{-1}(\alpha) = [-\sqrt{-2\ln\alpha}, \sqrt{-2\ln\alpha}]$, respectively.

Theorem 1. ([16]) Let $\zeta_1, \zeta_2, \dots, \zeta_n$ be independent uncertain sets with inverse membership function $\mu_1^{-1}, \mu_2^{-1}, \dots, \mu_n^{-1}$

and let f be a measurable function. Then $\xi = f(\zeta_1, \zeta_2, \dots, \zeta_n)$ has an inverse membership function

$$\lambda^{-1}(\alpha) = f(\mu_1^{-1}(\alpha), \mu_2^{-1}(\alpha), \dots, \mu_n^{-1}(\alpha)).$$

Theorem 2. (Liu, 2007) Let ζ be an uncertain set with inverse membership function $\mu^{-1}(\alpha)$. Then the membership function of ζ is determined by $\mu(x) = \sup\{\alpha \in [0, 1] | x \in \mu^{-1}(\alpha)\}$.

With the aim of measuring the uncertainty treated as uncertain set, the entropy for uncertain set is introduced by Liu (2011). Similarly, Peng and Li (2013) defined the radical entropy for uncertain set and discussed several important properties. At the same time, the way to calculate the radical entropy was proposed. Lu and Wang (2013) defined the triangular entropy for uncertain set and discussed its properties. Meanwhile, the computational formula of the triangular entropy was studied. Wang and Ha (2013) defined the quadratic entropy for uncertain set and investigated the relationship between quadratic entropy and Liu's entropy. Besides, quadratic cross entropy was also introduced to measure the difference between two uncertain sets. In addition,

Yao (2014) gave the definition of sine entropy and applied it to portfolio selection and clustering. The research of uncertain set has been utilized widely.

Theorem 3. (Zhao, 2008) For a special kind of nonlinear programming problems (1) where the objective function contains absolute value symbol

$$\begin{cases} \min \sum_{i=1}^n c_i x_i + \sum_{k=1}^K | \sum_{i=1}^n d_{ki} x_i | \\ \text{subject to:} \\ AX \leq b \\ X \geq 0 \end{cases} \tag{1}$$

it can be transformed to the following linear programming problem,

$$\begin{cases} \min \sum_{i=1}^n c_i x_i + \sum_{k=1}^K (u_k + v_k) \\ \text{subject to:} \\ AX \leq b \\ \sum_{i=1}^n d_{ki} x_i + u_k - v_k = 0 \\ X \geq 0, U \geq 0, V \geq 0 \end{cases} \tag{2}$$

where $U = (u_1, u_2, \dots, u_K)$, $V = (v_1, v_2, \dots, v_K)$.

3. Uncertain Regression Model and Parameter Estimation

3.1 Uncertain Regression Model

In real life situations, practitioners come across wide variety of data sets. Tools available in statistics are effectively employed in decision making process as long as there is no impreciseness in data sets. However, statistical tools do not perform well when uncertainty enters the system. Hence, alternative approaches become necessary. Sensing this necessity, different tools parallel to those available in statistics are being developed under uncertainty theory. Liu proposed an uncertain regression model which is explained below.

Let x be a vector of independent variables and y be a dependent variable. Assume the functional relationship between y and x can be expressed by the regression model

$$y=f(x|\beta) \tag{3}$$

where β is an unknown vector of parameters. If $f(x|\beta)$ is a linear function, i.e.

$$y(x_i)=\beta_0 x_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} \tag{4}$$

where p is the number of independent variables, then we obtain an uncertain linear regression model. In traditional linear regression model, we assume that the observed values of y are influenced by independent variables and random error. In other words, the dependent variable is a random variable with nondeterminacy. In uncertain linear regression model (4), the disturbance term and y are regarded as uncertain variables. This is the essential difference between uncertain regression model and traditional regression model.

As a matter of fact, in many practical problems, the observed values can not be expressed as a crisp quantity but it may be possible to identity a wide range of possible values for y . For example, the experts often give a range when predict the price of a stock. In such cases, if we continue to use some crisp values for y , the conclusion drawn from the model may not reflect the actual behavior of the system. To handle such situations, we have to consider the uncertain regression model based uncertain set, namely,

$$y=f(x|\beta) \tag{5}$$

where $x \in R$, β and y are uncertain sets.

Models of the form stated in (5) can be used in the place of conventional statistical model when decision making process involves uncertainty. In this paper, we mainly focus our attention on the linear regression model, namely,

$$y(x_i)=\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}, i=1,2,\dots,n,$$

where $x_{ji} \in R, j=1; 2; \dots; p$ and $y(x_i)$ are uncertain sets.

3.2 Parameter Estimation

Once the functional form of f is determined, next arises the problem of estimating the parameters β involved in the identified functional form.

This section presents a solution to the estimation of parameters $\beta_0, \beta_1, \dots, \beta_p$ assumed as uncertain sets. The proposed method makes use of the observed values (x_i, y_i) and it is based on linear programming method. This

work considers three types of uncertain sets, namely, symmetrical triangular, symmetrical trapezoidal and normal uncertain sets. In order to estimate the parameters, we first introduce the concept of fitted value.

Definition 13. Let $(x_i, y_i), i=1,2, \dots, n$ denote the i -th uncertain sample value, y_i denote the i -th observed value with inverse membership function $\mu_{y_i}^{-1}(\alpha)$ and $y(x_i)$ denote the i -th uncertain dependent variable with inverse membership function i.e.

$$\mu_{y(x_i)}^{-1}(\alpha) = \{x \mid \mu_{y(x_i)}(x) \geq \alpha\}$$

$$\mu_{y_i}^{-1}(\alpha) = \{x \mid \mu_{y_i}(x) \geq \alpha\}.$$

where $0 \leq \alpha \leq 1$. If α_i is the maximum among α which meets the condition $\mu_{y_i}^{-1}(\alpha) \subset \mu_{y(x_i)}^{-1}(\alpha)$ for each index i then α_i is called the fitted value of the i -th sample.

The subset notation $\mu_{y_i}^{-1}(\alpha) \subset \mu_{y(x_i)}^{-1}(\alpha)$ means that the membership information of y_i is contained in $y(x_i)$. Fitted value is an important index to describe the degree of congruence between the observed value y_i and the theoretical value $y(x_i)$. We generally restrict that all the α_i are greater than a specific value H ($0 \leq H \leq 1$). H is fixed by taking into account the opinion of domain experts. That is $\alpha_i \geq H, i=1,2, \dots, n$. It can be guaranteed that the degree of congruence between y_i and $y(x_i)$ is higher than level H for any i . Higher the value of α_i , better will be the quality of the corresponding fit.

3.2.1 Regression Model Based on Symmetrical Triangular Uncertain Set

Let $\beta_0, \beta_1, \dots, \beta_p$ be independent symmetrical triangular uncertain sets, where $\beta_j = (a_j, b_j, c_j), j=0,1, \dots, p$. Let (x_i, y_i) be observed value, where x_i is a vector made up of crisp numbers with $x_i \geq 0$ and $y_i = (m_i, p_i, n_i)$ is a symmetrical triangular uncertain set with an inverse membership function $\mu_{y_i}^{-1}(\alpha) = [(1-\alpha)m_i + \alpha p_i, \alpha p_i + (1-\alpha)n_i], i=1,2, \dots, n$.

Assume that the relationship between x_i and $y(x_i)$ can be expressed by the linear function

$$y(x_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, i=1,2, \dots, n.$$

It follows from Definition 7 that

$$y(x_i) = (a_0 + \sum_{j=1}^p a_j x_{ij}, b_0 + \sum_{j=1}^p b_j x_{ij}, c_0 + \sum_{j=1}^p c_j x_{ij}), i=1,2, \dots, n$$

is a symmetrical triangular uncertain set. By Theorem 1, $y(x_i)$ has an inverse membership function as follows.

$$\mu_{y(x_i)}^{-1}(\alpha) = [(1-\alpha)(a_0 + \sum_{j=1}^p a_j x_{ij}) + \alpha(b_0 + \sum_{j=1}^p b_j x_{ij}), \alpha(b_0 + \sum_{j=1}^p b_j x_{ij}) + (1-\alpha)(c_0 + \sum_{j=1}^p c_j x_{ij})].$$

The relationship among $\alpha_i, y(x_i)$ and y_i is shown in Figure 1. To meet the requirement on α_i , namely, $\mu_{y_i}^{-1}(\alpha) \subset \mu_{y(x_i)}^{-1}(\alpha)$, the two inequalities

$$(1-\alpha_i)m_i + \alpha_i p_i \geq (1-\alpha_i) \sum_{j=0}^p a_j x_{ij} + \alpha_i \sum_{j=0}^p b_j x_{ij} \text{ and } (1-\alpha_i)n_i + \alpha_i p_i \leq (1-\alpha_i) \sum_{j=0}^p c_j x_{ij} + \alpha_i \sum_{j=0}^p b_j x_{ij}$$

should be satisfied.

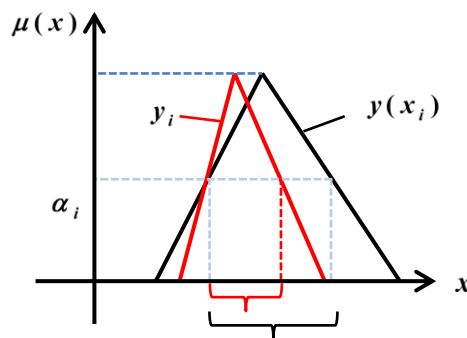


Figure 1. Parameter estimation of symmetrical triangular uncertain set

Generally speaking, the closer y_i and $y(x_i)$ better will be the fit. Hence, we aim for estimating the parameters in such a that y_i has the same membership function as $y(x_i)$. Because they are both symmetrical triangular uncertain sets, if the radius of y_i and the radius of $y(x_i)$ are same under the condition of they are pretty anastomotic. As a result of the above observations, parameter estimation problem can be transformed into the following LP model.

$$\left\{ \begin{array}{l} \min \sum_{i=0}^n [\sum_{j=0}^p (c_j - b_j)x_{ji} - (n_i - p_i)] \\ \text{subject to:} \\ (1 - \alpha_i)m_i + \alpha_i p_i \geq (1 - \alpha_i) \sum_{j=0}^p a_j x_{ji} + \alpha_i \sum_{j=0}^p b_j x_{ji} \\ (1 - \alpha_i)n_i + \alpha_i p_i \leq (1 - \alpha_i) \sum_{j=0}^p c_j x_{ji} + \alpha_i \sum_{j=0}^p b_j x_{ji} \\ x_{0i} = 1, 0 \leq H \leq \alpha_i \leq 1, i = 1, 2, \dots, n \\ a_j \leq b_j \leq c_j, j = 1, 2, \dots, p. \end{array} \right. \tag{6}$$

The model above can not be solved because of the unknown parameter α_i . Besides, it usually meets the condition $\alpha_i \geq H$. If model (6) is established while $\alpha_i = H$ it can be guaranteed that $\alpha_i \geq H$ for each index i . Hence it can be transformed into the following LP model.

$$\left\{ \begin{array}{l} \min \sum_{i=0}^n [\sum_{j=0}^p (c_j - b_j)x_{ji}] \\ \text{subject to:} \\ (1 - H)m_i + Hp_i \geq (1 - H) \sum_{j=0}^p a_j x_{ji} + H \sum_{j=0}^p b_j x_{ji} \\ (1 - H)n_i + Hp_i \leq (1 - H) \sum_{j=0}^p c_j x_{ji} + H \sum_{j=0}^p b_j x_{ji} \\ x_{0i} = 1, 0 \leq H \leq 1, i = 1, 2, \dots, n \\ a_j \leq b_j \leq c_j, j = 1, 2, \dots, p. \end{array} \right. \tag{7}$$

The solution of the above linear programming model gives the estimated value $\tilde{\beta}_j = (\tilde{a}_j, \tilde{b}_j, \tilde{c}_j)$.

3.2.2 Regression Model Based on Symmetrical Trapezoidal Uncertain Set

Let $\beta_0, \beta_1, \dots, \beta_p$ be independent symmetrical trapezoidal uncertain sets, where $\beta_j = (a_j, b_j, c_j, d_j), j = 0, 1, \dots, p$. Let (x_i, y_i) be observed value, where x_i is a vector made up of crisp numbers with $x_i \geq 0$ and $y_i = (p_i, q_i, s_i, t_i)$ is a symmetrical trapezoidal uncertain set with an inverse membership function

$$\mu_{y_i}^{-1}(\alpha) = [(1 - \alpha)p_i + \alpha q_i, \alpha s_i + (1 - \alpha)t_i], i = 1, 2, \dots, n.$$

Assume that the relationship between x_i and $y(x_i)$ can be expressed by the linear function

$$y(x_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, i = 1, 2, \dots, n.$$

It follows from Definition 7 that $y(x_i) = (a_0 + \sum_{j=1}^p a_j x_{ji}, b_0 + \sum_{j=1}^p b_j x_{ji}, c_0 + \sum_{j=1}^p c_j x_{ji}, d_0 + \sum_{j=1}^p d_j x_{ji}), i = 1, 2, \dots, n$ is a symmetrical trapezoidal uncertain set. By Theorem 1, $y(x_i)$ has an inverse membership function as follows.

$$\mu_{y(x_i)}^{-1}(\alpha) = [(1 - \alpha)(a_0 + \sum_{j=1}^p a_j x_{ji}) + \alpha(b_0 + \sum_{j=1}^p b_j x_{ji}), \alpha(c_0 + \sum_{j=1}^p c_j x_{ji}) + (1 - \alpha)(d_0 + \sum_{j=1}^p d_j x_{ji})].$$

The relationship among $\alpha_i, y(x_i)$ and y_i is shown in Figure 2. To meet the requirement on α_i , namely,

$\mu_{y_i}^{-1}(\alpha) \subset \mu_{y(x_i)}^{-1}(\alpha)$, the two inequalities

$$(1 - \alpha_i)p_i + \alpha_i q_i \geq (1 - \alpha_i) \sum_{j=0}^p a_j x_{ji} + \alpha_i \sum_{j=0}^p b_j x_{ji} \text{ and } (1 - \alpha_i)t_i + \alpha_i s_i \leq (1 - \alpha_i) \sum_{j=0}^p d_j x_{ji} + \alpha_i \sum_{j=0}^p c_j x_{ji}$$

should be satisfied.

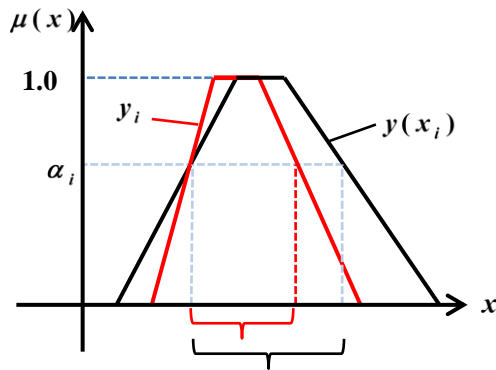


Figure 2. Parameter estimation of symmetrical trapezoidal uncertain set

Generally speaking, the closer y_i and $y(x_i)$ better will be the fit. Hence, we aim for estimating the parameters in such a that y_i has the same membership function as $y(x_i)$. Because they are both symmetrical trapezoidal uncertain sets, if their toplines and baselines are same under the condition $\alpha_i \geq H$, they are pretty anastomotic. As a result of the above observations, parameter estimation problem can be transformed into the following nonlinear programming (NLP) model.

$$\left\{ \begin{array}{l} \min \sum_{i=0}^n (|\sum_{j=0}^p (c_j - b_j)x_{ji} - (s_i - q_i)| + \sum_{j=0}^p (d_j - a_j)x_{ji} - (t_i - p_i)) \\ \text{subject to:} \\ (1 - \alpha_i)p_i + \alpha_i q_i \geq (1 - \alpha_i) \sum_{j=0}^p a_j x_{ji} + \alpha_i \sum_{j=0}^p b_j x_{ji} \\ (1 - \alpha_i)t_i + \alpha_i s_i \leq (1 - \alpha_i) \sum_{j=0}^p d_j x_{ji} + \alpha_i \sum_{j=0}^p c_j x_{ji} \\ x_{0i} = 1, 0 \leq H \leq \alpha_i \leq 1, i = 1, 2, \dots, n \\ a_j \leq b_j \leq c_j \leq d_j, j = 1, 2, \dots, p. \end{array} \right. \quad (8)$$

The model above can not be solved because of the unknown parameter α_i . Besides, it usually meets the condition $\alpha_i \geq H$. If model (8) is established while $\alpha_i = H$ it can be guaranteed that $\alpha_i \geq H$ for each index i . Hence it can be transformed into the following model.

$$\left\{ \begin{array}{l} \min \sum_{i=0}^n (|\sum_{j=0}^p (c_j - b_j)x_{ji} - (s_i - q_i)| + \sum_{j=0}^p (d_j - a_j)x_{ji} - (t_i - p_i)) \\ \text{subject to:} \\ (1 - H)p_i + Hq_i \geq (1 - H) \sum_{j=0}^p a_j x_{ji} + H \sum_{j=0}^p b_j x_{ji} \\ (1 - H)t_i + Hs_i \leq (1 - H) \sum_{j=0}^p d_j x_{ji} + H \sum_{j=0}^p c_j x_{ji} \\ x_{0i} = 1, 0 \leq H \leq 1, i = 1, 2, \dots, n \\ a_j \leq b_j \leq c_j \leq d_j, j = 1, 2, \dots, p. \end{array} \right. \quad (9)$$

It follows from Theorem 3 that the nonlinear programming model (9) is equivalent to the following LP model.

$$\left\{ \begin{array}{l} \min \sum_{i=0}^n (\sum_{j=0}^p (u_j + v_j) + \sum_{j=0}^p (d_j - a_j)x_{ji} - (t_i - p_i)) \\ \text{subject to:} \\ (1 - \alpha_i)p_i + \alpha_i q_i \geq (1 - \alpha_i) \sum_{j=0}^p a_j x_{ji} + \alpha_i \sum_{j=0}^p b_j x_{ji} \\ (1 - \alpha_i)t_i + \alpha_i s_i \leq (1 - \alpha_i) \sum_{j=0}^p d_j x_{ji} + \alpha_i \sum_{j=0}^p c_j x_{ji} \\ \sum_{j=0}^p (c_j - b_j)x_{ji} - (s_j - q_j) + u_j - v_j = 0 \\ x_{0i} = 1, 0 \leq H \leq \alpha_i \leq 1, i = 1, 2, \dots, n \\ a_j \leq b_j \leq c_j \leq d_j, j = 1, 2, \dots, p. \end{array} \right. \tag{10}$$

The solution of the above linear programming model gives the estimated value $\tilde{\beta}_j = (\tilde{a}_j, \tilde{b}_j, \tilde{c}_j, \tilde{d}_j)$.

3.2.3 Regression Model Based on Normal Uncertain Set

Let $\beta_0, \beta_1, \dots, \beta_p$ be independent normal uncertain sets, where $\beta_j = N(a_j, b_j), j = 0, 1, \dots, p$. Let (x_i, y_i) be observed value, where x_i is a vector made up of crisp numbers with $x_i \geq 0$ and $y_i = N(m_i, n_i)$ is a normal uncertain set with an inverse membership function $\mu_{y_i}^{-1}(\alpha) = [m_i - n_i \sqrt{-2 \ln \alpha}, m_i + n_i \sqrt{-2 \ln \alpha}], i = 1, 2, \dots, n$.

Assume that the relationship between x_i and $y(x_i)$ can be expressed by the linear function

$$y(x_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, i = 1, 2, \dots, n.$$

It follows from Definition 7 that $y(x_i) = N(a_0 + \sum_{j=1}^p a_j x_{ji}, b_0 + \sum_{j=1}^p b_j x_{ji}), i = 1, 2, \dots, n$ is a normal uncertain set. By Theorem 1, $y(x_i)$ has an inverse membership function as follows.

$$\mu_{y(x_i)}^{-1}(\alpha) = [(a_0 + \sum_{j=1}^p a_j x_{ji}) - (b_0 + \sum_{j=1}^p b_j x_{ji}) \sqrt{-2 \ln \alpha}, (a_0 + \sum_{j=1}^p a_j x_{ji}) + (b_0 + \sum_{j=1}^p b_j x_{ji}) \sqrt{-2 \ln \alpha}].$$

The relationship among $\alpha_i, y(x_i)$ and y_i is shown in Figure 3. To meet the requirement on α_i , namely, $\mu_{y_i}^{-1}(\alpha) \subset \mu_{y(x_i)}^{-1}(\alpha)$, the two inequalities

$$m_i - n_i \sqrt{-2 \ln \alpha_i} \geq (a_0 + \sum_{j=1}^p a_j x_{ji}) - (b_0 + \sum_{j=1}^p b_j x_{ji}) \sqrt{-2 \ln \alpha_i} \quad \text{and} \quad m_i + n_i \sqrt{-2 \ln \alpha_i} \leq (a_0 + \sum_{j=1}^p a_j x_{ji}) + (b_0 + \sum_{j=1}^p b_j x_{ji}) \sqrt{-2 \ln \alpha_i}$$

should be satisfied.

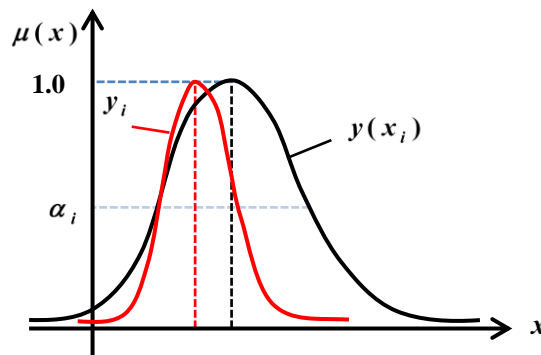


Figure 3. Parameter estimation of normal uncertain set

In generally, the closer y_i and $y(x_i)$ better will be the fit. Hence, we aim for estimating the parameters in such a

that y_i has the same membership function as $y(x_i)$. Because they are both normal uncertain sets, if the centers and dispersion degrees of membership functions are same under the condition of $\alpha_i \geq H$, they are pretty anastomotic. As a result of the above observations, parameter estimation problem can be transformed into the following nonlinear programming model.

$$\left\{ \begin{array}{l} \min \sum_{i=0}^n (|\sum_{j=0}^p a_j x_{ji} - m_i|) + (\sum_{j=0}^p b_j x_{ji} - n_i) \\ \text{subject to:} \\ m_i - n_i \sqrt{-2 \ln \alpha_i} \geq (a_0 + \sum_{j=1}^p a_j x_{ji}) - (b_0 + \sum_{j=1}^p b_j x_{ji}) \sqrt{-2 \ln \alpha_i} \\ m_i + n_i \sqrt{-2 \ln \alpha_i} \leq (a_0 + \sum_{j=1}^p a_j x_{ji}) + (b_0 + \sum_{j=1}^p b_j x_{ji}) \sqrt{-2 \ln \alpha_i} \\ x_{0i} = 1, 0 \leq H \leq \alpha_i \leq 1, i = 1, 2, \dots, n \\ b_j \geq 0, j = 0, 1, 2, \dots, p. \end{array} \right. \quad (11)$$

The model above can not be solved because of the unknown parameter α_i . If model (11) is established while $\alpha_i = H$ it can be guaranteed that $\alpha_i \geq H$ for each index i . Hence it can be transformed into the following model.

$$\left\{ \begin{array}{l} \min \sum_{i=0}^n (|\sum_{j=0}^p a_j x_{ji} - m_i|) + (\sum_{j=0}^p b_j x_{ji} - n_i) \\ \text{subject to:} \\ m_i - n_i \sqrt{-2 \ln H} \geq (a_0 + \sum_{j=1}^p a_j x_{ji}) - (b_0 + \sum_{j=1}^p b_j x_{ji}) \sqrt{-2 \ln H} \\ m_i + n_i \sqrt{-2 \ln H} \leq (a_0 + \sum_{j=1}^p a_j x_{ji}) + (b_0 + \sum_{j=1}^p b_j x_{ji}) \sqrt{-2 \ln H} \\ x_{0i} = 1, 0 \leq H \leq 1, i = 1, 2, \dots, n \\ b_j \geq 0, j = 0, 1, 2, \dots, p. \end{array} \right. \quad (12)$$

It follows from Theorem 3 that the NLP model (12) is equivalent to the following LP model.

$$\left\{ \begin{array}{l} \min \sum_{i=0}^n (u_j + v_j) + (\sum_{j=0}^p b_j x_{ji} - n_i) \\ \text{subject to:} \\ m_i - n_i \sqrt{-2 \ln H} \geq (a_0 + \sum_{j=1}^p a_j x_{ji}) - (b_0 + \sum_{j=1}^p b_j x_{ji}) \sqrt{-2 \ln H} \\ m_i + n_i \sqrt{-2 \ln H} \leq (a_0 + \sum_{j=1}^p a_j x_{ji}) + (b_0 + \sum_{j=1}^p b_j x_{ji}) \sqrt{-2 \ln H} \\ \sum_{j=0}^p a_j x_{ji} - m_i + u_j - v_j = 0, u_j \geq 0, v_j \geq 0 \\ x_{0i} = 1, 0 \leq H \leq 1, i = 1, 2, \dots, n \\ b_j \geq 0, j = 0, 1, 2, \dots, p. \end{array} \right. \quad (13)$$

The solution of the above linear programming model gives the estimated value $\tilde{\beta}_j = N(\tilde{a}_j, \tilde{b}_j)$.

3.3 Evaluation Criteria

Once the uncertain regression model is fitted using the available data, next arises the question of assessing the quality of fit. Towards this, two measures meant for assessing the quality of fit are presented below.

(1) The relative deviation between the centers of fitted value and observed data y_i :

$$\Lambda_1 = \frac{|y_i - \sum_{j=0}^p b_j |x_{ji}|}{y_i} \quad i = 1, 2, \dots, n.$$

(2) The ratio between the widths of fitted value and observed data y_i :

$$\Lambda_1 = \frac{\sum_{j=0}^P (c_j - b_j) |x_{ji}|}{y_i} \quad i = 1, 2, \dots, n.$$

Generally speaking, if Λ_1 and Λ_2 are both within 30%, the fitted uncertain regression model is considered to be acceptable.

4. Numerical Experiment

In this section, an illustrative example is given in support of the estimation procedure suggested in the previous section for the uncertain regression model. Consider the relationship between the heat released by some cement during solidification denoted by y_i and two kinds of chemical compositions denoted by x_1 and x_2 . The industry experience has shown that y_i is not a fixed value but assumes values over an identified interval of values. Hence it is reasonable to treat heat as an uncertain set. The observed values are furnished in Table 1.

Table 1

i	x_{1i}	x_{2i}	$y_i = (m_i, p_i, n_i)$
1	7	26	(70.5, 78.5, 86.5)
2	1	29	(89.3, 74.3, 79.3)
3	11	56	(97.3, 104.3, 111.3)
4	11	31	(80.6, 87.6, 94.6)
5	7	52	(88.9, 95.9, 102.9)
6	11	55	(100.2, 109.2, 118.2)
7	3	71	(96.7, 102.7, 108.7)
8	1	31	(68.5, 72.5, 76.5)
9	2	54	(87.1, 93.1, 99.1)
10	21	47	(107.9, 115.9, 123.9)

It is known from the past experience that there exists a linear function relationship between y and x , where $x = (1, x_1, x_2)^T$. That is to say, y can be expressed by x and parameter $\beta = (\beta_0, \beta_1, \beta_2)$ via the uncertain linear regression model, i.e. $y(x_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$, where $\beta_j = (a_j, b_j, c_j)$ is a symmetrical triangular uncertain set, $j = 0, 1, 2, i = 1, 2, \dots, 10$.

In order to estimate the unknown parameters β_j , the LP model is built as below based on the discussion made in Section 3.

$$\left\{ \begin{array}{l} \min \sum_{i=0}^{10} [\sum_{j=0}^2 (c_j - b_j) x_{ji}] \\ \text{subject to:} \\ (1-H)m_i + Hp_i \geq (1-H) \sum_{j=0}^2 a_j x_{ji} + H \sum_{j=0}^2 b_j x_{ji} \\ (1-H)n_i + Hp_i \leq (1-H) \sum_{j=0}^2 c_j x_{ji} + H \sum_{j=0}^2 b_j x_{ji} \\ x_{0i} = 1, 0 \leq H \leq 1, i = 1, 2, \dots, 10 \\ a_j \leq b_j \leq c_j, j = 0, 1, 2. \end{array} \right. \quad (14)$$

The above LP problem which corresponds to the model explained in (14) has been solved by taking $H = 0.4$ and the following solution is obtained.

$$\begin{cases} a_0 = 44.2966, a_1 = 1.4132, a_2 = 0.6273 \\ b_0 = 48.0616, b_1 = 1.4302, b_2 = 0.7857 \\ c_0 = 51.8270, c_1 = 1.4471, c_2 = 0.9440 \end{cases}$$

The solution leads to the estimates $\hat{\beta}_0 = (44.2966, 48.0618, 51.8270)$, $\hat{\beta}_1 = (1.4132, 1.4302, 1.4471)$, $\hat{\beta}_2 = (0.6273, 0.7857, 0.9440)$. Then the uncertain linear regression model is

$$y(x) = (44.2966, 48.0618, 51.8270) + (1.4132, 1.4302, 1.4471)x_1 + (0.6273, 0.7857, 0.9440)x_2.$$

As one could observe, the values of Λ_1 and Λ_2 as shown in Table 2 are less than 30%. This supports the appropriateness of the fitted model.

Table 2

i	1	2	3	4	5	6	7	8	9	10
Λ_1	1.7834*10 ⁻⁵	0.0272	0.0334	0.0062	0.0315	0.0200	0.0529	0.0186	0.0026	0.0075
		2	9	9	9	8	4	0	9	6
Λ_2	0.1019	0.1127	0.1229	0.1012	0.1264	0.1160	0.1467	0.1199	0.1327	0.0998

5. Conclusion

In this paper, a linear regression model based on uncertain sets that can be used for investigating the relationship between variables involved in an uncertain situation. The proposed model assumes the independent variables are of crisp nature and the dependent variable as well as regression coefficients are uncertain sets. Estimation procedures meant for handling situation where one encounters any one of the three uncertain sets, namely, symmetrical triangular, symmetrical trapezoidal and normal uncertain sets have been discussed in detail. An illustrative example has been given to add strength to the proposed model as well as the method of estimation.

Acknowledgements

This work was supported by Hebei Natural Science Foundation (Grant No. G2013402063), Hebei Social Science Foundation (Grant No. HB14JY068) and the Foundation of Hebei Education Department (Grant No. ZD2017016).

References

- Cerny, M., Antoch, J., & Hladik, M. (2013). On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data. *Information Sciences*, 244(7), 26-47.
- Chen, F., Chen, Y., Zhou, J., & Liu, Y. (2016). Optimizing h value for fuzzy linear regression with asymmetric triangular fuzzy coefficients. *Engineering Applications of Artificial Intelligence*, 47, 16-24.
- Danesh, S., Farnoosh, R., & Razzaghnia, T. (2016). Fuzzy nonparametric regression based on an adaptive neuro-fuzzy Inference system. *Neurocomputing*, 173, 1450-1460.
- Diamond, P. (1987). Least squares fitting of several fuzzy variables. *In Proceedings of the 2nd IFSA Congress, Tokyo*.
- Diamond, P. (1988). Fuzzy least squares. *Journal of Information Science*, 46(3), 141-157.
- Diamond, P., & Korner, R. (1997). Extended fuzzy linear models and least squares estimates. *Computers and Mathematics with Applications*, 33(9), 15-32.
- Guo, H., Wang, X., & Gao, Z. (2017). Uncertain linear regression model and its application. *Journal of Intelligent Manufacturing*, 28(3), 559-564.
- Guo, H., Wang, X., & Wang, L. The Normal Uncertain Set and its Application. Retrieved from <http://orsc.edu.cn/online/140910.pdf>
- Guo, R., Cui, Y., & Guo, D. (2011). An uncertain regression model. *Grey Systems Theory and Application*, 1(3), 202-215.
- Ishibuchi, H., & Nii, M. (2001). Fuzzy regression using asymmetric fuzzy coefficients and fuzzified neural networks. *Fuzzy Sets and Systems*, 119(2), 273-290.

- Korner, R., & Nather, W. (1998). Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates. *Journal of Information Science*, 109(1-4), 95-118.
- Liu, B. (2007). *Uncertain Theory* (2nd ed.). Springer-Verlag, Berlin.
- Liu, B. (2009). Some research problems in uncertainty theory. *Journal of Uncertain Systems*, 3(1), 3-10.
- Liu, B. (2010). *A branch of mathematics for modeling human uncertainty*. Springer-Verlag, Berlin.
- Liu, B. (2010). Uncertain set theory and uncertain inference rule with application to uncertain control. *Jouanal of Uncertain Systems*, 4(2), 83-98.
- Liu, B. (2011). Uncertain logic for modeling human language. *Journal of Uncertain Systems*, 5(1), 3-20.
- Liu, B. (2012). Membership functions and operational law of uncertain sets. *Fuzzy Optimization and Decision Making*, 11(4), 387-410.
- Liu, B. (2013). A new definition of independence of uncertain sets. *Fuzzy Optimization and Decision Making*, 12(4), 451-461.
- Lu, R., & Wang X. (2013). Triangular entropy of uncertain sets. *Information: An International Interdisciplinary Journal*, 16(2), 1249-1256.
- Muller, D., Esche, E., Diana, C. L. C., & Wozny, G. (2014). An algorithm for the identification and estimation of relevant Parameters for optimization under uncertainty. *Computers and Chemical Engineering*, 71(71), 94-103.
- Peng, J., & Li, S. (2013). Radical entropies of uncertain sets. *Information: An International Interdisciplinary Journal*, 16(2), 895-902.
- Sclove, S. (2014). Improved estimators for coefficients in linear regression. *Journal of the American Statistical Association*, 63(63), 596-606.
- Su, Z., Wang, Y., & Wang, P. (2013). Parametric regression analysis of imprecise and uncertain data in the fuzzy belief function framework. *International Journal of Approximate Reasoning*, 54(8), 1217-1242.
- Tanaka, H., Uejima, S., & Asai, K. (1982). Fuzzy linear regression model. *IEEE Transactions on Systems Man and Cybernetics*, 12(6), 2933-2938.
- Wang, X., Gao, Z., & Guo, H. (2012). Delphi method for estimating uncertainty distributions. *Information: An International Interdisciplinary Journal*, 15(2), 449-460.
- Wang, X., Gao, Z., & Guo, H. (2012). Uncertain hypothesis testing for two experts empirical data. *Mathematical and Computer Modelling*, 55(3-4), 1478-1482
- Wang, X., & Ha, M. (2013). Quadratic entropy of uncertain sets. *Fuzzy Optimization and Decision Making*, 12(1), 99-109.
- Wang, X., & Peng, Z. (2014). Method of moments for estimating uncertainty distribution. *Journal of Uncertainty Analysis and Applications*, 2(1), 1-10.
- Yao, K. (2014) Sine entropy of uncertain set and its applications. *Applied Soft Computing*, 22(5), 432-442.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(65), 338-353.
- Zhao, Y. (2008). Solution of special nonlinear programming problems. *Journal of Jiangnan University*, 36(3). (in Chinese).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).