Formulating Intuitionistic Fuzzy Regression Models: A Mathematical Programming Approach

Sheng-Hsing Nien

National Defense University, Department of Logistics Management

Liang-Hsuan Chen (✉ lhchen@mail.ncku.edu.tw)

National Cheng Kung University Department of Industrial and Information Management

<https://orcid.org/0000-0002-9974-3423>

Research Article

Keywords: Intuitionistic fuzzy , regression model , Intuitionistic fuzzy number, Mathematical programming , Distance criterion

Posted Date: July 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-681349/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Formulating Intuitionistic Fuzzy Regression Models: A Mathematical Programming Approach

Abstract

This study develops a mathematical programming approach to establish intuitionistic fuzzy regression models (IFRMs) by considering the randomness and fuzziness of intuitionistic fuzzy observations. In contrast to existing approaches, the IFRMs are established in terms of five ordinary regression models representing the components of the estimated triangular intuitionistic fuzzy response variable. The optimal parameters of the five ordinary regression models are determined by solving the proposed mathematical programming problem, which is linearized to make the resolution process efficient. Based on the concepts of randomness and fuzziness in the formulation processes, the proposed approach can improve on existing approaches' weaknesses with establishing IFRMs, such as the limitation of symmetrical triangular membership (or non-membership) functions, the determination of parameter signs in the model, and the wide spread of the estimated responses. In addition, some numerical explanatory variables included in the intuitionistic fuzzy observations are also allowed in the proposed approach, even though it was developed for intuitionistic fuzzy observations. In contrast to existing approaches, the proposed approach is general and flexible in applications. Comparisons show that the proposed approach outperforms existing approaches in terms of similarity and distance measures.

Keywords Intuitionistic fuzzy regression model Intuitionistic fuzzy number Mathematical programming Distance criterion.

1 Introduction

Regression analysis is a useful approach for establishing the relationship between a response variable and one or more explanatory variables. In real-world applications, observations usually cannot be represented as exact (or numerical) values due to a shortage of information or the decision-maker's subjective judgment (Chen and Hsueh 2007). Linguistic terms, defined and characterized by fuzzy set theories (Zadeh 1965), are thus often adopted to describe observations. For fuzzy observations, a number of researchers have established fuzzy regression models (FRMs) by developing various approaches to determine the optimal parameters in the models based on their performance criterion, such as Chen and Hsueh (2007), Choi and Buckley (2008), Kelkinnama and Taheri (2012), and others. These approaches have been applied to many fields by Chan and Engelke (2017), Spiliotis and Garrote (2021), Boukezzoula et al. (2021), and others. Although FRMs have numerous applications, some limitations have been found. For example, the fuzzy parameters in FRMs could lead to a wider spread of the estimated fuzzy response when the magnitude of the fuzzy (or crisp) explanatory variables increases (Kao and Chyu 2002). This significantly increases the unnecessary information in the estimated fuzzy response, possibly

influencing FRM performance in some applications. This has motivated studies to build up FRMs with crisp parameters by applying mathematical programming (Chen and Hsueh 2007), a least-squares method (Chen and Hsueh 2009), or least absolute deviation (Choi and Buckley 2008), or to adopt shape-preserving operations (Kelkinnama and Taheri 2012) for generating narrower spreads through fuzzy arithmetic. In addition, the signs of parameters in FRMs should be considered in the formulation processes since they influence the arithmetic operations; however, almost all approaches presumed them to be positive.

By extending fuzzy set theorems (Zadeh 1965) that use membership degrees to characterize the relationship grades, Atanassov (1986) proposed intuitionistic fuzzy sets (IFSs), in which non-membership degrees are included. Observations described by IFSs can contain more information than those described by fuzzy sets. Membership and non-membership degrees are usually used to express positive and negative information, respectively, in decision-making problems. IFSs have been widely studied and applied in various problems (Atanassov 1999), such as in Hesamian et al. (2020), Ho et al. (2019), Yolcu et al. (2020), and so on. To the best of our knowledge, only a few approaches have been proposed for developing intuitionistic fuzzy regression models (IFRMs). Parvathi et al. (2013)

extended the approaches for FRMs from Tanaka et al. (1982) to formulate an IFRM based on a linear programming problem to determine coefficients that are characterized as symmetrical triangular intuitionistic fuzzy numbers (TIFNs). In their study, observations contained numeric response and explanatory variables. By solving the linear programming problem, only the symmetrical TIFN parameter of the intercept term is obtained; the other parameters are always numerical due to the nature of their problem. The resulting parameters are used to determine the estimated responses, which are then defuzzified to delimit the upper and lower bounds of the estimated response variable. Arefi and Taheri (2015) proposed an IFRM based on a least-squares method where the response variables, explanatory variables, and parameters are all symmetrical TIFNs. They proposed a metric for calculating the squared errors between observed and estimated response variables to determine the symmetrical TIFN parameters. The explanatory variables and parameters were all assumed to be positive in their formulation processes; however, a negative parameter was produced in their example.

Recently, Chen and Nien (2020) extended Chen and Hsueh's (2007) approach to establish an IFRM based on the criterion of least absolute deviation. Although the problem of determining the parameter sign can be solved, their approach still suffers from the problem of producing wider spreads, since the intuitionistic fuzzy arithmetic is applied. Al-Qudaimi (2021) proposed an approach, named the Ishita approach, to build up an IFRM by formulating non-linear programming problems to determine the parameters of the IFRM. However, some unreasonable conditions are defined in the Ishita approach, since they violate the basic definitions of an IFN, leading to its resolution processes being problematic. In addition, the TIFN-parameters are also presumed to be positive in the IFRM formulation, while a negative slope TIFN-parameter is obtained in the example. This approach also has the problem that wider spreads could be produced.

Recognizing the above limitations of existing approaches for formulating IFRMs, such as producing wider spreads and the wrong setting of the sign of parameters, this study proposes mathematical programming approaches for establishing IFRMs that consider the randomness and fuzziness of TIFN observations, referring to the concepts of fuzzy random variables (Watada et al. 2009). Randomness is measured based on the most likely value (center) of a fuzzy number, and fuzziness is represented by the spreads of the fuzzy number. This study adopts these concepts to decompose the representations of intuitionistic fuzzy numbers (IFNs) into two critical parts to simplify the formulation processes. More importantly, the characteristics of IFNs can be considered in the IFRM formulations by individually expressing the five components of an IFN, i.e.,

the most likely value, the left/right spreads of the membership function, and those of the non-membership function. Taking randomness and fuzziness into account, a mathematical programming model that is restricted to five ordinary regression models representing the five components of the estimated TIFN response variable, respectively, is formulated to minimize the least absolute deviations to build up the optimal IFRM. The proposed approaches overcome the weaknesses of existing approaches, such as the determined IFRM having wider spreads and the problem of determining the sign of parameters, because the resulting IFRM is determined from the five ordinary regression models representing the five components of the estimated TIFN response variable, respectively, in the formulation processes.

The objective of this study is to develop a mathematical programming approach based on the least absolute deviations for formulating the optimal IFRM considering the randomness and fuzziness of intuitionistic fuzzy observations. The proposed approach overcomes the weaknesses of existing approaches by ensuring that the spreads of the resulting IFRM will not be wide and the problem of the sign of parameters does not require any concern. Furthermore, although the proposed approach focuses on the development of IFNs, it actually can deal with observations containing numerical, fuzzy, intuitionistic fuzzy, or mixed variable types. The rest of this paper is organized as follows. Some basic concepts of IFSs and related measures, such as the similarity measure, squared errors measure, and distance measure, are described briefly in Section 2. In Section 3, considering randomness and fuzziness, the optimal IFRM is formulated by solving the proposed mathematical programming model to minimize the least absolute deviations in terms of five ordinary regression models representing the five components of the estimated TIFN response variable, respectively. The proposed approaches are general, since the explanatory variables in the observations are not required to be symmetrical TIFNs; even numerical observations are allowed. Two examples are used to demonstrate performance via comparisons with existing approaches in Section 4. A dataset is used to demonstrate the applicability of the proposed model in Section 5. The comparison analyses with existing approaches are conducted, and the advantages of the proposed approach are explained in Section 6. Finally, Section 7 provides some conclusions.

2 Background

Intuitionistic fuzzy sets (IFSs), proposed by Atanassov (1986), are a generalization of fuzzy sets that includes the degrees of membership and non-membership.

Definition 1 (Atanassov 1986; Xu and Yager 2006): Let X denote a universe of discourse. An IFS in X is given by:

$$\tilde{\mathcal{A}} = \{(x, \mu_A(x), v_A(x)) \mid x \in X\} \quad (1)$$

where $\mu_A(x), v_A(x): X \rightarrow [0,1]$ are functions that satisfy $0 \leq \mu_A(x) + v_A(x) \leq 1$ for all $x \in X$. Referring to Fig. 1, the values of $\mu_A(x)$ and $v_A(x)$ represent the degrees of membership and non-membership, respectively. In addition, the degree of hesitancy is defined as $\pi_A(x) = 1 - \mu_A(x) - v_A(x)$ of element $x \in X$. If the non-membership, $v_A(x)$, is ignored, the IFS degenerates into a fuzzy set.

Definition 2 (Grzegorzewski 2003; Liu and Shi 2000): An IFN is an intuitionistic fuzzy subset of the real number R , and satisfies the following properties.

- (1) The intuitionistic fuzzy subset is on the real line.
- (2) A unique value $m \in X$ exists, and $\mu_A(m) = 1$ and $v_A(m) = 0$ are satisfied, where m is called the most likely value of $\tilde{\mathcal{A}}$.

- (3) Convex for the membership function $\mu_A(x)$, i.e.,

$$\mu_A(\lambda x_1 + (1 - \lambda)x_2) \geq \min(\mu_A(x_1), \mu_A(x_2)), \quad \text{where } x_1, x_2 \in R \text{ and } \lambda \in [0,1] \quad (2)$$

- (4) Concave for the non-membership function $v_A(x)$, i.e.,

$$v_A(\lambda x_1 + (1 - \lambda)x_2) \leq \max(v_A(x_1), v_A(x_2)), \quad \text{where } x_1, x_2 \in R \text{ and } \lambda \in [0,1] \quad (3)$$

Generally, the membership and non-membership functions of $\tilde{\mathcal{A}}$ can be characterized in the following forms:

$$\mu_A(x) = \begin{cases} 0 & , \text{ for } -\infty < x \leq m - s^{ML} \\ f_1(x) & , \text{ for } x \in [m - s^{ML}, m] \\ 1 & , \text{ for } x = m \\ h_1(x) & , \text{ for } x \in [m, m + s^{MR}] \\ 0 & , \text{ for } m + s^{MR} \leq x < \infty \end{cases} \quad (4)$$

$$v_A(x) = \begin{cases} 1 & , \text{ for } -\infty < x \leq m - s^{VL} \\ f_2(x) & , \text{ for } x \in [m - s^{VL}, m], 0 \leq f_1(x) + f_2(x) \leq 1 \\ 0 & , \text{ for } x = m \\ h_2(x) & , \text{ for } x \in [m, m + s^{VR}], 0 \leq h_1(x) + h_2(x) \leq 1 \\ 1 & , \text{ for } m + s^{VR} \leq x < \infty \end{cases} \quad (5)$$

where $f_1(x)$ and $h_2(x)$ are strictly increasing functions, and $h_1(x)$ and $f_2(x)$ are strictly decreasing functions. s^{ML} and s^{MR} are the left and right spreads of membership function $\mu_A(x)$, and s^{VL} and s^{VR} are the left and right spreads of non-membership function $v_A(x)$, respectively. Note that a spread of the non-membership function must be greater than or equal to the corresponding spread of the membership function, i.e., $s^{VL} \geq s^{ML}$ and $s^{VR} \geq s^{MR}$.

Definition 3 (Mahapatra and Roy 2009): A triangular intuitionistic fuzzy number (TIFN) $\tilde{\mathcal{A}}_{TIFN}^c = (m; s^{ML}, s^{MR}; s^{VL}, s^{VR})$ is characterized by the membership function $\mu_A(x)$ and non-membership function $v_A(x)$ shown in Eq. (6). Fig. 2 shows the two functions.

Fig. 1 Intuitionistic fuzzy set

Fig. 2 Triangular intuitionistic fuzzy set

Fig. 3 TIFN and its α -cuts.

$$\mu_A(x) = \begin{cases} \frac{x - (m - s^{ML})}{s^{ML}}, & m - s^{ML} \leq x \leq m \\ \frac{(m + s^{MR}) - x}{s^{MR}}, & m \leq x \leq m + s^{MR} \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad v_A(x) = \begin{cases} 1 - \frac{x - (m - s^{VL})}{s^{VL}}, & m - s^{VL} \leq x \leq m \\ 1 - \frac{(m + s^{VR}) - x}{s^{VR}}, & m \leq x \leq m + s^{VR} \\ 1, & \text{otherwise} \end{cases} \quad (6)$$

Definition 4 (Guha and Chakraborty 2010): The α -cut of an IFS $\tilde{\mathcal{A}}$ is defined as:

$$\tilde{\mathcal{A}}_\alpha^c = \{(x, \mu(x), v(x)) \mid \mu(x) \geq \alpha \text{ and } v(x) \leq 1 - \alpha, \alpha \in [0,1]\} \quad (7)$$

The inequality $v_A(x) \leq 1 - \alpha$ is equivalent to $1 - v_A(x) \geq \alpha$ and thus $\tilde{\mathcal{A}}_\alpha^c$ can be expressed as the crisp sets $\tilde{\mathcal{A}}_\mu^c(\alpha) = \{x: \mu_A(x) \geq \alpha\}$ and $\tilde{\mathcal{A}}_{1-v}^c(\alpha) = \{x: 1 - v_A(x) \geq \alpha\}$. Alternatively, $\tilde{\mathcal{A}}_\alpha^c$ can be represented by the following pair of intervals:

$$\tilde{A}_\alpha^c = \{[\tilde{A}_\mu^b(\alpha), \tilde{A}_\mu^b(\alpha)], [\tilde{A}_{1-\nu}^b(\alpha), \tilde{A}_{1-\nu}^b(\alpha)]\} \quad (8)$$

Fig. 3 shows that the two crisp sets $\{x: v_A(x) \leq 1-\alpha\}$ and $\tilde{A}_{1-\nu}^c(\alpha) = \{x: 1-v_A(x) \geq \alpha\}$ have the same intervals. To be compatible with a previous study (Arefi and Taheri 2015), the notation $\tilde{A}_{1-\nu}^c(\alpha)$ is adopted here.

Based on this definition, the α -cuts of a TIFN $\tilde{A}^c$ can be formulated in the following general form:

$$\tilde{A}_\alpha^c = \left\{ \left[m - (1-\alpha)s^{ML}, m + (1-\alpha)s^{MR} \right], \right. \\ \left. \left[m - (1-\alpha)s^{VL}, m + (1-\alpha)s^{VR} \right] \right\} \quad (9)$$

The two extreme cases are: $\tilde{A}_{\alpha=0}^c = \{[m-s^{ML}, m+s^{MR}], [m-s^{VL}, m+s^{VR}]\}$ and $\tilde{A}_{\alpha=1}^c = \{m\}$.

Definition 5 (Chakraborty et al. 2014): A TIFN $\tilde{A}^c = (m; s^{ML}, s^{MR}; s^{VL}, s^{VR})$ is positive if $(m-s^{VL}) > 0$; it is negative if $(m+s^{VR}) < 0$.

A useful measure of IFNs is necessary to evaluate the performance of an IFRM. For this purpose, Arefi and Taheri (2015) proposed two measures, namely a similarity measure based on the absolute difference of membership and non-membership degrees of two TIFNs and a squared difference measure, as defined below.

Definition 6 (Arefi and Taheri 2015): Suppose that $\tilde{A}^c = (m_A; s_A^{ML}, s_A^{MR}; s_A^{VL}, s_A^{VR})$ and $\tilde{B}^c = (m_B; s_B^{ML}, s_B^{MR}; s_B^{VL}, s_B^{VR})$ are two TIFNs. The similarity measure is expressed as:

$$SM(\tilde{A}^c, \tilde{B}^c) = 1 - \frac{1}{2} \left[\frac{\int_{-\infty}^{\infty} |\mu_A(x) - \mu_B(x)| dx}{\int_{-\infty}^{\infty} \mu_A(x) dx + \int_{-\infty}^{\infty} \mu_B(x) dx} + \frac{\int_{-\infty}^{\infty} |v_A(x) - v_B(x)| dx}{\int_{-\infty}^{\infty} (1-v_A(x)) dx + \int_{-\infty}^{\infty} (1-v_B(x)) dx} \right] \quad (10)$$

with the value of $0 \leq SM(\tilde{A}^c, \tilde{B}^c) \leq 1$. This index is measured in terms of the area of the average difference between the membership and non-membership functions of two IFNs. However, the degree of difference cannot be determined if the two IFNs have no intersections.

Definition 7 (Arefi and Taheri 2015): Suppose that $\tilde{A}^c = (m_A; s_A^{ML}, s_A^{MR}; s_A^{VL}, s_A^{VR})$ and $\tilde{B}^c = (m_B; s_B^{ML}, s_B^{MR}; s_B^{VL}, s_B^{VR})$ are two TIFNs. Then, the generalized squared difference measure, $d^2(\tilde{A}^c, \tilde{B}^c)$, can be expressed as:

$$d^2(\tilde{A}^c, \tilde{B}^c) = (m_A - m_B)^2 + \frac{1}{24} \left[(s_B^{ML} - s_A^{ML})^2 + (s_B^{MR} - s_A^{MR})^2 + (s_B^{VL} - s_A^{VL})^2 + (s_B^{VR} - s_A^{VR})^2 \right] \\ + \frac{1}{6} (m_A - m_B) \left[(s_B^{ML} - s_A^{ML}) - (s_B^{MR} - s_A^{MR}) + (s_B^{VL} - s_A^{VL}) - (s_B^{VR} - s_A^{VR}) \right] \quad (11)$$

Another distance measure based on the absolute difference of α -cuts, shown below, is also adopted in this study.

Definition 8 (Patriota 2013): Suppose that $\tilde{A}^c = (m_A; s_A^{ML},$

$s_A^{MR}; s_A^{VL}, s_A^{VR})$ and $\tilde{B}^c = (m_B; s_B^{ML}, s_B^{MR}; s_B^{VL}, s_B^{VR})$ are two TIFNs. Then, the distance measure, $D_p(\tilde{A}^c, \tilde{B}^c)$, can be expressed as:

$$D_p(\tilde{A}^c, \tilde{B}^c) = \frac{1}{4} \left(\int_0^1 |\tilde{A}_\mu^b(\alpha) - \tilde{B}_\mu^b(\alpha)|^p d\alpha + \int_0^1 |\tilde{A}_{1-\nu}^b(\alpha) - \tilde{B}_{1-\nu}^b(\alpha)|^p d\alpha + \int_0^1 |\tilde{A}_{1-\nu}^b(\alpha) - \tilde{B}_{1-\nu}^b(\alpha)|^p d\alpha + \int_0^1 |\tilde{A}_{1-\nu}^b(\alpha) - \tilde{B}_{1-\nu}^b(\alpha)|^p d\alpha \right)^{1/p} \quad (12)$$

where p is a parameter and $1 \leq p \leq \infty$. The measure $D_p(\tilde{A}^c, \tilde{B}^c)$ is determined by calculating the integral of the average absolute difference of all α -cuts based on the parameter p , which is usually set to 1. Furthermore, the three measures mentioned above, i.e., the similarity measure, squared difference measure, and distance measure, all satisfy their respective axioms, as proven in previous studies (Arefi and Taheri 2015; Grzegorzewski 2003).

3 Intuitionistic fuzzy regression modeling

The formulation processes of an IFRM are introduced in this section. An IFRM usually suffers from the problems of determining the sign of parameters in the formulation processes and producing wider spreads in the estimations, thus influencing IFRM performance. To avoid these problems, the proposed approach considers that randomness and fuzziness are the two critical factors that affect IFRM performance. To deal with randomness and fuzziness, the formulation of IFRMs is carried out by directly determining the five components of the estimated TIFN response, where each component is formulated as an ordinary regression function of the corresponding ones of the observed TIFN explanatory variables. Mathematical programming problems are formulated to determine the optimal numerical parameters of each component of the estimated TIFN response to minimize the total difference between the observed and estimated TIFN responses.

Regarding the randomness and fuzziness of a fuzzy number (Fong and Hong 2010), the most likely value can be expressed as a random variable and the left and right spreads can be used to express fuzziness. In this study, for a TIFN $\tilde{A}^c = (m_A; s_A^{ML}, s_A^{MR}; s_A^{VL}, s_A^{VR})$, randomness and fuzziness are represented by the most likely value, m_A , and the spreads by s^{ML} , s^{MR} , s^{VL} , and s^{VR} , respectively. To evaluate the differences between the observed and estimated TIFN responses in terms of randomness and fuzziness, the α -cuts of TIFNs are adopted, since they can easily be used to measure the distance between two membership (non-membership) functions at each α level.

Let a TIFN observation set be denoted as $(\tilde{Y}_i^c, \tilde{X}_{li}^c, \tilde{X}_{2i}^c, \tilde{X}_{ji}^c, \tilde{X}_{pi}^c)$, where $\tilde{Y}_i^c = (m_{y_i}; s_{y_i}^{ML}, s_{y_i}^{MR}; s_{y_i}^{VL}, s_{y_i}^{VR})$ is the TIFN response variable and $\tilde{X}_{ji}^c = (m_{x_{ji}}; s_{x_{ji}}^{ML},$

$s_{x_{ji}}^{MR}, s_{x_{ji}}^{VL}, s_{x_{ji}}^{VR}$) represents the TIFN explanatory variables.

Let the estimated TIFN response variable be $\hat{Y}_i^c = (\hat{m}_{y_i}; \hat{s}_{y_i}^{ML}, \hat{s}_{y_i}^{MR}, \hat{s}_{y_i}^{VL}, \hat{s}_{y_i}^{VR})$. Based on a previous study that presented a distance measure for fuzzy numbers (Chen and Hsueh 2007), the difference between the i th observed and estimated TIFN responses, D_i , can be determined from the average deviations of r corresponding α -cuts as:

$$D_i = \frac{1}{4r} \sum_{k=1}^r \left(\left| \left(\hat{Y}_i^c \right)_\mu^L(\alpha_k) - \left(\hat{Y}_i^c \right)_\mu^L(\alpha_k) \right| + \left| \left(\hat{Y}_i^c \right)_\mu^R(\alpha_k) - \left(\hat{Y}_i^c \right)_\mu^R(\alpha_k) \right| + \left| \left(\hat{Y}_i^c \right)_{1-\nu}^L(\alpha_k) - \left(\hat{Y}_i^c \right)_{1-\nu}^L(\alpha_k) \right| + \left| \left(\hat{Y}_i^c \right)_{1-\nu}^R(\alpha_k) - \left(\hat{Y}_i^c \right)_{1-\nu}^R(\alpha_k) \right| \right) \quad (13)$$

where $\alpha_k = (k-1)/(r-1)$. Based on Definition 4, the first part of D_i can be reformulated as:

$$\begin{aligned} & \sum_{k=1}^r \left| \left(\hat{Y}_i^c \right)_\mu^L(\alpha_k) - \left(\hat{Y}_i^c \right)_\mu^L(\alpha_k) \right| \\ &= \sum_{k=1}^r \left| (m_{y_i} - (1-\alpha_k)s_{y_i}^{ML}) - (\hat{m}_{y_i} - (1-\alpha_k)\hat{s}_{y_i}^{ML}) \right| \\ &= \sum_{k=1}^r \left| (m_{y_i} - \hat{m}_{y_i}) - (1-\alpha_k)(s_{y_i}^{ML} - \hat{s}_{y_i}^{ML}) \right| \end{aligned} \quad (14)$$

The above equation is derived from the absolute differences of the most likely value and the left spread of membership functions between the observed and estimated TIFN responses, i.e., the differences of the randomness $(m_{y_i} - \hat{m}_{y_i})$ and fuzziness $(s_{y_i}^{ML} - \hat{s}_{y_i}^{ML})$. Similarly, the remaining parts of D_i can be reformulated as the differences of the randomness $(m_{y_i} - \hat{m}_{y_i})$ and the corresponding fuzziness, i.e., $(s_{y_i}^{MR} - \hat{s}_{y_i}^{MR})$, $(s_{y_i}^{VL} - \hat{s}_{y_i}^{VL})$, and $(s_{y_i}^{VR} - \hat{s}_{y_i}^{VR})$. Two TIFNs are thus equal if there are no differences in randomness and fuzziness between them. This also indicates that an IFRM with better explanatory power will produce a smaller difference compared with the observed responses. To establish such an IFRM, each component of the estimated TIFN responses should be able to be explained by the corresponding ones in the observed TIFN explanatory variables. To this end, each component of the estimated TIFN responses can be formulated by an ordinary regression model as follows:

$$\hat{m}_{y_i} = \hat{b}_0^m + \hat{b}_1^m m_{x_{1i}} + \hat{b}_2^m m_{x_{2i}} + L + \hat{b}_p^m m_{x_{pi}} \quad (15)$$

$$\hat{s}_{y_i}^{ML} = \hat{b}_0^{ML} + \hat{b}_1^{ML} s_{x_{1i}}^{ML} + \hat{b}_2^{ML} s_{x_{2i}}^{ML} + L + \hat{b}_p^{ML} s_{x_{pi}}^{ML} \quad (16)$$

$$\hat{s}_{y_i}^{MR} = \hat{b}_0^{MR} + \hat{b}_1^{MR} s_{x_{1i}}^{MR} + \hat{b}_2^{MR} s_{x_{2i}}^{MR} + L + \hat{b}_p^{MR} s_{x_{pi}}^{MR} \quad (17)$$

$$\hat{s}_{y_i}^{VL} = \hat{b}_0^{VL} + \hat{b}_1^{VL} s_{x_{1i}}^{VL} + \hat{b}_2^{VL} s_{x_{2i}}^{VL} + L + \hat{b}_p^{VL} s_{x_{pi}}^{VL} \quad (18)$$

$$\hat{s}_{y_i}^{VR} = \hat{b}_0^{VR} + \hat{b}_1^{VR} s_{x_{1i}}^{VR} + \hat{b}_2^{VR} s_{x_{2i}}^{VR} + L + \hat{b}_p^{VR} s_{x_{pi}}^{VR} \quad (19)$$

where $\hat{\mathbf{b}}_j = [\hat{b}_j^m, \hat{b}_j^{ML}, \hat{b}_j^{MR}, \hat{b}_j^{VL}, \hat{b}_j^{VR}]$ is the numerical estimated parameter vector for the j th TIFN explanatory variable.

To minimize the overall difference between the observed and estimated TIFN responses, the optimal estimated parameters of TIFN explanatory variables should be determined based on Eqs. (15) to (19) using the corresponding mathematical programming models with the least absolute deviations, shown in Eqs. (20) to (24), respectively.

$$\begin{cases} \min \sum_{i=1}^n |m_{y_i} - \hat{m}_{y_i}| \\ \text{s.t. } \hat{m}_{y_i} = \hat{b}_0^m + \hat{b}_1^m m_{x_{1i}} + \hat{b}_2^m m_{x_{2i}} + L + \hat{b}_p^m m_{x_{pi}} \end{cases} \quad (20)$$

$$\begin{cases} \min \sum_{i=1}^n |s_{y_i}^{ML} - \hat{s}_{y_i}^{ML}| \\ \text{s.t. } \hat{s}_{y_i}^{ML} = \hat{b}_0^{ML} + \hat{b}_1^{ML} s_{x_{1i}}^{ML} + \hat{b}_2^{ML} s_{x_{2i}}^{ML} + L + \hat{b}_p^{ML} s_{x_{pi}}^{ML} \end{cases} \quad (21)$$

$$\begin{cases} \min \sum_{i=1}^n |s_{y_i}^{MR} - \hat{s}_{y_i}^{MR}| \\ \text{s.t. } \hat{s}_{y_i}^{MR} = \hat{b}_0^{MR} + \hat{b}_1^{MR} s_{x_{1i}}^{MR} + \hat{b}_2^{MR} s_{x_{2i}}^{MR} + L + \hat{b}_p^{MR} s_{x_{pi}}^{MR} \end{cases} \quad (22)$$

$$\begin{cases} \min \sum_{i=1}^n |s_{y_i}^{VL} - \hat{s}_{y_i}^{VL}| \\ \text{s.t. } \hat{s}_{y_i}^{VL} = \hat{b}_0^{VL} + \hat{b}_1^{VL} s_{x_{1i}}^{VL} + \hat{b}_2^{VL} s_{x_{2i}}^{VL} + L + \hat{b}_p^{VL} s_{x_{pi}}^{VL} \end{cases} \quad (23)$$

$$\begin{cases} \min \sum_{i=1}^n |s_{y_i}^{VR} - \hat{s}_{y_i}^{VR}| \\ \text{s.t. } \hat{s}_{y_i}^{VR} = \hat{b}_0^{VR} + \hat{b}_1^{VR} s_{x_{1i}}^{VR} + \hat{b}_2^{VR} s_{x_{2i}}^{VR} + L + \hat{b}_p^{VR} s_{x_{pi}}^{VR} \end{cases} \quad (24)$$

The determination of these parameters should guarantee that the estimated responses satisfy the definition of a TIFN. That is, the spreads should all be greater than or equal to zero, and $\hat{s}_{y_i}^{ML} \leq \hat{s}_{y_i}^{VL}$ and $\hat{s}_{y_i}^{MR} \leq \hat{s}_{y_i}^{VR}$. This suggests that Eqs. (20) to (24) cannot be independently solved. The aggregated model that contains the above constraints is formulated as Eq. (25).

$$\begin{aligned} & \min \sum_{i=1}^n \left\{ |m_{y_i} - \hat{m}_{y_i}| + |s_{y_i}^{ML} - \hat{s}_{y_i}^{ML}| \right. \\ & \quad \left. + |s_{y_i}^{MR} - \hat{s}_{y_i}^{MR}| + |s_{y_i}^{VL} - \hat{s}_{y_i}^{VL}| + |s_{y_i}^{VR} - \hat{s}_{y_i}^{VR}| \right\} \\ & \text{s.t. } \hat{m}_{y_i} = \hat{b}_0^m + \hat{b}_1^m m_{x_{1i}} + \hat{b}_2^m m_{x_{2i}} + L + \hat{b}_p^m m_{x_{pi}} \\ & \quad \hat{s}_{y_i}^{ML} = \hat{b}_0^{ML} + \hat{b}_1^{ML} s_{x_{1i}}^{ML} + \hat{b}_2^{ML} s_{x_{2i}}^{ML} + L + \hat{b}_p^{ML} s_{x_{pi}}^{ML} \\ & \quad \hat{s}_{y_i}^{MR} = \hat{b}_0^{MR} + \hat{b}_1^{MR} s_{x_{1i}}^{MR} + \hat{b}_2^{MR} s_{x_{2i}}^{MR} + L + \hat{b}_p^{MR} s_{x_{pi}}^{MR} \\ & \quad \hat{s}_{y_i}^{VL} = \hat{b}_0^{VL} + \hat{b}_1^{VL} s_{x_{1i}}^{VL} + \hat{b}_2^{VL} s_{x_{2i}}^{VL} + L + \hat{b}_p^{VL} s_{x_{pi}}^{VL} \\ & \quad \hat{s}_{y_i}^{VR} = \hat{b}_0^{VR} + \hat{b}_1^{VR} s_{x_{1i}}^{VR} + \hat{b}_2^{VR} s_{x_{2i}}^{VR} + L + \hat{b}_p^{VR} s_{x_{pi}}^{VR} \\ & \quad \hat{s}_{y_i}^{ML}, \hat{s}_{y_i}^{VL}, \hat{s}_{y_i}^{MR}, \hat{s}_{y_i}^{VR} \geq 0 \\ & \quad \hat{s}_{y_i}^{ML} \leq \hat{s}_{y_i}^{VL}, \hat{s}_{y_i}^{MR} \leq \hat{s}_{y_i}^{VR}, \text{ for } i=1, L, n \end{aligned} \quad (25)$$

After solving model (25), the optimal estimated parameters for the components of the estimated TIFN responses can be obtained to minimize the absolute deviations with the corresponding ones of the observed TIFN responses. In the proposed formulation processes, the IFRM is constructed by determining the optimal components of the estimated TIFN responses using the corresponding ones of the observed TIFN explanatory variables. That is, the

optimal IFRM is established in terms of five ordinary regression models, considering randomness and fuzziness. The product operator of TIFNs is avoided in the model formulations to prevent the generation of wider spreads. In addition, the crisp parameters in model (25) are formulated for the components of the estimated TIFN responses to avoid the problem of the signs of the parameters in IFRMs, so as to decrease the impact on forecasting performance. Particularly, even if the observation set contains some numerical explanatory variables in which spreads are missing, the proposed approach can still be applied to establish the optimal IFRM.

Furthermore, the objective function in model (25) contains absolute deviations, formulated as a nonlinear function. To increase computational efficiency, some linear techniques (Kelkinnama and Taheri 2012) can be adopted to reformulate the model. For example, for two non-negative deviation variables, M_i^1 and M_i^2 , $M_i^1 = \max\{m_{y_i} - \hat{m}_{y_i}, 0\}$ and $M_i^2 = \max\{\hat{m}_{y_i} - m_{y_i}, 0\}$ are called the positive and negative deviation variables, respectively, which can be used to express an absolute deviation. It is clear that $M_i^1 + M_i^2$ is equivalent to $|m_{y_i} - \hat{m}_{y_i}|$, and $M_i^1 - M_i^2$ is equivalent to $m_{y_i} - \hat{m}_{y_i}$. Therefore, the sub-model (20) can be rewritten as:

$$\begin{aligned} \min \sum_{i=1}^n M_i^1 + M_i^2 \\ \text{s.t. } M_i^1 - M_i^2 = m_{y_i} - \hat{m}_{y_i} = m_{y_i} - (\hat{b}_0^m + \hat{b}_1^m m_{x_{i1}} + L + \hat{b}_p^m m_{x_{pi}}) \end{aligned} \quad (26)$$

$$M_i^1, M_i^2 \geq 0$$

Similarly, other non-zero positive and negative deviation variables can also be defined, such as $ML_i^1 = \max\{s_{y_i}^{ML} - \hat{s}_{y_i}^{ML}, 0\}$, $ML_i^2 = \max\{\hat{s}_{y_i}^{ML} - s_{y_i}^{ML}, 0\}$, $MR_i^1 = \max\{s_{y_i}^{MR} - \hat{s}_{y_i}^{MR}, 0\}$, $MR_i^2 = \max\{\hat{s}_{y_i}^{MR} - s_{y_i}^{MR}, 0\}$, $VL_i^1 = \max\{s_{y_i}^{VL} - \hat{s}_{y_i}^{VL}, 0\}$, $VL_i^2 = \max\{\hat{s}_{y_i}^{VL} - s_{y_i}^{VL}, 0\}$, $VR_i^1 = \max\{s_{y_i}^{VR} - \hat{s}_{y_i}^{VR}, 0\}$, and $VR_i^2 = \max\{\hat{s}_{y_i}^{VR} - s_{y_i}^{VR}, 0\}$. Then, the sub-models (21) to (24) can be modified as (27) to (30), respectively, as follows:

$$\begin{aligned} \min \sum_{i=1}^n ML_i^1 + ML_i^2 \\ \text{s.t. } ML_i^1 - ML_i^2 = s_{y_i}^{ML} - (\hat{b}_0^{ML} + \hat{b}_1^{ML} s_{x_{i1}} + L + \hat{b}_p^{ML} s_{x_{pi}}) \end{aligned} \quad (27)$$

$$ML_i^1, ML_i^2 \geq 0$$

$$\begin{aligned} \min \sum_{i=1}^n MR_i^1 + MR_i^2 \\ \text{s.t. } MR_i^1 - MR_i^2 = s_{y_i}^{MR} - (\hat{b}_0^{MR} + \hat{b}_1^{MR} s_{x_{i1}} + L + \hat{b}_p^{MR} s_{x_{pi}}) \end{aligned} \quad (28)$$

$$MR_i^1, MR_i^2 \geq 0$$

$$\begin{aligned} \min \sum_{i=1}^n VL_i^1 + VL_i^2 \\ \text{s.t. } VL_i^1 - VL_i^2 = s_{y_i}^{VL} - (\hat{b}_0^{VL} + \hat{b}_1^{VL} s_{x_{i1}} + L + \hat{b}_p^{VL} s_{x_{pi}}) \end{aligned} \quad (29)$$

$$VL_i^1, VL_i^2 \geq 0$$

$$\begin{aligned} \min \sum_{i=1}^n VR_i^1 + VR_i^2 \\ \text{s.t. } VR_i^1 - VR_i^2 = s_{y_i}^{VR} - (\hat{b}_0^{VR} + \hat{b}_1^{VR} s_{x_{i1}} + L + \hat{b}_p^{VR} s_{x_{pi}}) \end{aligned} \quad (30)$$

$$VR_i^1, VR_i^2 \geq 0$$

Therefore, the model (25) will become:

$$\begin{aligned} \min \sum_{i=1}^n (M_i^1 + M_i^2 + ML_i^1 + ML_i^2 + MR_i^1 + MR_i^2 + VL_i^1 + VL_i^2 + VR_i^1 + VR_i^2) \\ \text{s.t. } M_i^1 - M_i^2 = m_{y_i} - (\hat{b}_0^m + \hat{b}_1^m m_{x_{i1}} + L + \hat{b}_p^m m_{x_{pi}}) \\ ML_i^1 - ML_i^2 = s_{y_i}^{ML} - (\hat{b}_0^{ML} + \hat{b}_1^{ML} s_{x_{i1}} + L + \hat{b}_p^{ML} s_{x_{pi}}) \\ MR_i^1 - MR_i^2 = s_{y_i}^{MR} - (\hat{b}_0^{MR} + \hat{b}_1^{MR} s_{x_{i1}} + L + \hat{b}_p^{MR} s_{x_{pi}}) \\ VL_i^1 - VL_i^2 = s_{y_i}^{VL} - (\hat{b}_0^{VL} + \hat{b}_1^{VL} s_{x_{i1}} + L + \hat{b}_p^{VL} s_{x_{pi}}) \\ VR_i^1 - VR_i^2 = s_{y_i}^{VR} - (\hat{b}_0^{VR} + \hat{b}_1^{VR} s_{x_{i1}} + L + \hat{b}_p^{VR} s_{x_{pi}}) \\ M_i^1, M_i^2, ML_i^1, ML_i^2, MR_i^1, MR_i^2, VL_i^1, VL_i^2, VR_i^1, VR_i^2 \geq 0 \\ \hat{s}_{y_i}^{ML}, \hat{s}_{y_i}^{VL}, \hat{s}_{y_i}^{MR}, \hat{s}_{y_i}^{VR} \geq 0 \end{aligned} \quad (31)$$

$$\hat{s}_{y_i}^{ML} \leq \hat{s}_{y_i}^{VL}, \hat{s}_{y_i}^{MR} \leq \hat{s}_{y_i}^{VR}, \text{ for } i=1, L, n$$

The reformulated linear model (31) is a general linear programming model, and can thus be easily solved to determine the optimal parameters as $\hat{\mathbf{b}}_j = [\hat{b}_j^m, \hat{b}_j^{ML}, \hat{b}_j^{MR}, \hat{b}_j^{VL}, \hat{b}_j^{VR}]$, $j=0, 1, \dots, p$, using commercial software, such as LINGO (Anderson et al. 2017). To formulate and solve this model, only the most likely value and left/right spread of the membership/non-membership function from the IFN dataset are required. In addition, when the number of TIFN explanatory variables in an intuitionistic fuzzy dataset increases, only the numbers of corresponding components in the first five constraints increase. The solving process is quite efficient, even for datasets with a large quantity of observations and/or explanatory variables, because model (31) is linear.

4 Demonstration and comparisons

This study builds an IFRM in terms of five ordinary linear regression models based on the linear relations between components in the estimated TIFN response variable and the corresponding ones in the TIFN explanatory variables, considering randomness and fuzziness. For demonstration and comparisons, examples from Al-Qudaimi (2021) and Arefi and Taheri (2015) are adopted in this section.

Example 1 The example from Al-Qudaimi (2021) contains four patterns with one explanatory variable. Let $\mathcal{V}$ be the volume (given in liters) of one mole of methane gas under a constant pressure of one atm, and $\mathcal{T}$ be the temperature in Celsius ($^{\circ}\text{C}$). They are represented as symmetrical TIFNs and listed in Table 1. Al-Qudaimi (2021)

employed the Ishita approach to establish the IFRM, $\hat{Y}_{AI}^{\%}$, as:

$$\hat{Y}_{AI}^{\%} = (23.2910; 0.0081, 0.0081; 0.9890, 0.9890) \oplus (0.0991; 0, 0; 0, 0) \otimes \hat{X}^{\%} \quad (32)$$

Applying the proposed approach, the following optimal parameter vectors can be determined as $\hat{\mathbf{b}}_0 = [22.3762, 0.0111, 0.0111, 1, 1]$ and $\hat{\mathbf{b}}_1 = [0.0826, -0.0112, -0.0112, -0.0112, -0.0112]$. It is noted that parameters for the right and left spreads of the membership/non-membership function are the same because the observations are symmetrical TIFNs. As such, the five regression models representing the five components of the estimated TIFN response variable, $\hat{Y}_{CN}^{\%}$, can be formulated as:

$$\begin{aligned} \hat{m}_y &= 22.3762 + 0.0826m_x \\ \hat{s}_y^{ML} &= \hat{s}_y^{MR} = 0.111 - 0.0112s_x^{ML} = 0.111 - 0.0112s_x^{MR} \\ \hat{s}_y^{VL} &= \hat{s}_y^{VR} = 1 - 0.0112s_x^{VL} = 1 - 0.0112s_x^{VR} \end{aligned} \quad (33)$$

Table 2 shows performance comparisons between the IFRMs (32) and (33) in terms of similarity (10) and distance measures (11) and (12). Obviously, the proposed IFRM (33) outperforms model (32) based on the three measures.

Example 2 This example adopts a dataset with 24 symmetrical TIFN observations (Arefi and Taheri 2015), as listed in Table 3. For each observation, the response variable, namely the cation exchange capacity ($\hat{Y}^{\%}$), and two explanatory variables, namely the sand content percentage ($\hat{X}_1^{\%}$) and organic matter content ($\hat{X}_2^{\%}$), are used.

Arefi and Taheri (2015) developed an IFRM approach based on the criterion of the least squared errors using the observations with symmetrical TIFN response and explanatory variables. To demonstrate their approach, the dataset, listed as Table 3, is used in their example, in which the spreads of the TIFN response and explanatory variables of the observations are all defined as a proportion of the corresponding mean value, such as $s_y^{ML} = s_y^{MR} = 0.05m_{y_i}$, $s_y^{VL} = s_y^{VR} = 0.1m_{y_i}$, $s_{x_j}^{ML} = s_{x_j}^{MR} = 0.1m_{x_j}$, and $s_{x_j}^{VL} = s_{x_j}^{VR} = 0.15m_{x_j}$. The IFRM, $\hat{Y}_{LSM}^{\%}$, obtained by Arefi and Taheri (2015) is:

Table 1 Dataset for Example 1 (Al-Qudaimi 2021)

No.	$\hat{Y}^{\%}$	$\hat{X}^{\%}$
1	(22; 0.011, 0.011; 0.989, 0.989)	(-3; 0.011, 0.011; 0.989, 0.989)
2	(21; 0.198, 0.198; 0.9802, 0.9802)	(-23; 0.0842, 0.0842; 0.9158, 0.9158)
3	(18; 0.0081, 0.0081; 0.9919, 0.9919)	(-53; 0.194, 0.194; 0.8051, 0.8051)
4	(9; 0.0045, 0.0045; 0.9955, 0.9955)	(-162; 0.5931, 0.5931; 0.4069, 0.4069)

$$\begin{aligned} \hat{Y}_{LSM}^{\%} &= (21.9811; 0.8933, 0.8933; 1.9882, 1.9882) \\ &\oplus (-0.222, 0.0117, 0.0117; 0.0118, 0.0118) \otimes \hat{X}_1^{\%} \\ &\oplus (2.4701) \otimes \hat{X}_2^{\%} \end{aligned} \quad (34)$$

Using the same dataset, Chen and Nien (2020) established the IFRM, $\hat{Y}_{LAD}^{\%}$, as:

$$\begin{aligned} \hat{Y}_{LAD}^{\%} &= (21.0663; 0, 0.1043; 0.3537, 0.6842) \\ &\oplus (-0.1969) \otimes \hat{X}_1^{\%} \oplus (2.6922) \otimes \hat{X}_2^{\%} \end{aligned} \quad (35)$$

Applying the proposed approach, the optimal parameter vectors can be obtained as:

$$\begin{aligned} \hat{\mathbf{b}}_0 &= [21.0663, 1.0553, 1.0553, 2.1066, 2.1066], \\ \hat{\mathbf{b}}_1 &= [-0.1982, -0.0091, -0.0091, -0.1321, -0.1321], \text{ and} \\ \hat{\mathbf{b}}_2 &= [2.6922, 1.3461, 1.3461, 1.7947, 1.7947]. \end{aligned}$$

As is the case in the previous example, the resulting optimal parameters for the right and left spreads of the membership/non-membership function are also identical, since observations in this example are also symmetrical TIFNs. Then, the IFRM, $\hat{Y}_{CN}^{\%}$, can be established as follows:

$$\begin{aligned} \hat{m}_y &= 21.0663 - 0.1982m_{x_1} + 2.6922m_{x_2} \\ \hat{s}_y^{ML} &= 1.0553 - 0.0991s_{x_1}^{ML} + 1.3461s_{x_2}^{ML} \\ \hat{s}_y^{MR} &= 1.0553 - 0.0991s_{x_1}^{MR} + 1.3461s_{x_2}^{MR} \\ \hat{s}_y^{VL} &= 2.1066 - 0.1321s_{x_1}^{VL} + 1.7947s_{x_2}^{VL} \\ \hat{s}_y^{VR} &= 2.1066 - 0.1321s_{x_1}^{VR} + 1.7947s_{x_2}^{VR} \end{aligned} \quad (36)$$

The results of performance comparisons among the IFRM (34) from Arefi and Taheri's (2015) approach, the IFRM (35) from Chen and Nien's (2020) approach, and the IFRM (36) based on the proposed formulation, in terms of the three performance measures, are listed in Table 4. The proposed IFRM (36) outperforms models (34) and (35). The total similarity SM from (36) is 11.8053, which is higher than

Table 2 Performance comparisons of Example 1 in terms of three measures

No.	$\hat{Y}_{AI}^{\%}$ (Al-Qudaimi 2021)			$\hat{Y}_{CN}^{\%}$ (Proposed approach)		
	SM	d^2	$D_{p=1}$	SM	d^2	$D_{p=1}$
1	0	1.0674	0.9937	0.4372	0.0165	0.1284
2	0.0944	0.0803	0.2206	0.2698	0.2764	0.5229
3	0.0271	0.0791	0.2099	0.9639	0.0001	0.0004
4	0	3.1811	1.7632	1	0	0
sum	0.1215	4.4079	3.1874	2.6709	0.2929	0.6518

Table 3 Dataset for Example 2 (Arefi and Taheri 2015)

No.	$Y\%$	$X_1\%$	$X_2\%$
1	(16.50; 0.83, 0.83; 1.65, 1.65)	(35.00; 3.50, 3.50; 5.25, 5.25)	(0.88; 0.09, 0.09; 0.13, 0.13)
2	(18.60; 0.93, 0.93; 1.86, 1.86)	(37.00; 3.70, 3.70; 5.55, 5.55)	(1.13; 0.11, 0.11; 0.17, 0.17)
3	(19.30; 0.97, 0.97; 1.93, 1.93)	(27.00; 2.70, 2.70; 4.05, 4.05)	(1.31; 0.13, 0.13; 0.20, 0.20)
4	(20.30; 1.02, 1.02; 2.03, 2.03)	(29.00; 2.90, 2.90; 4.35, 4.35)	(1.98; 0.20, 0.20; 0.30, 0.30)
5	(17.30; 0.87, 0.87; 1.73, 1.73)	(38.00; 3.80, 3.80; 5.70, 5.70)	(1.02; 0.10, 0.10; 0.15, 0.15)
6	(20.40; 1.02, 1.02; 2.04, 2.04)	(32.00; 3.20, 3.20; 4.80, 4.80)	(1.29; 0.13, 0.13; 0.19, 0.19)
7	(19.30; 0.97, 0.97; 1.93, 1.93)	(29.00; 2.90, 2.90; 4.35, 4.35)	(1.52; 0.15, 0.15; 0.23, 0.23)
8	(21.90; 1.10, 1.10; 2.19, 2.19)	(18.00; 1.80, 1.80; 2.70, 2.70)	(1.33; 0.13, 0.13; 0.20, 0.20)
9	(15.90; 0.80, 0.80; 1.59, 1.59)	(40.00; 4.00, 4.00; 6.00, 6.00)	(1.71; 0.17, 0.17; 0.26, 0.26)
10	(18.30; 0.92, 0.92; 1.83, 1.83)	(28.00; 2.80, 2.80; 4.20, 4.20)	(2.00; 0.20, 0.20; 0.30, 0.30)
11	(22.60; 1.13, 1.13; 2.26, 2.26)	(13.00; 1.30, 1.30; 1.95, 1.95)	(1.68; 0.17, 0.17; 0.25, 0.25)
12	(23.70; 1.19, 1.19; 2.37, 2.37)	(19.00; 1.90, 1.90; 2.85, 2.85)	(2.15; 0.22, 0.22; 0.32, 0.32)
13	(24.40; 1.22, 1.22; 2.44, 2.44)	(31.00; 3.10, 3.10; 4.65, 4.65)	(3.52; 0.35, 0.35; 0.53, 0.53)
14	(21.80; 1.09, 1.09; 2.18, 2.18)	(31.00; 3.10, 3.10; 4.65, 4.65)	(2.33; 0.23, 0.23; 0.35, 0.35)
15	(23.80; 1.19, 1.19; 2.38, 2.38)	(17.00; 1.70, 1.70; 2.55, 2.55)	(1.71; 0.17, 0.17; 0.26, 0.26)
16	(20.80; 1.04, 1.04; 2.08, 2.08)	(14.00; 1.40, 1.40; 2.10, 2.10)	(1.14; 0.11, 0.11; 0.17, 0.17)
17	(17.50; 0.88, 0.88; 1.75, 1.75)	(19.00; 1.90, 1.90; 2.85, 2.85)	(0.99; 0.10, 0.10; 0.15, 0.15)
18	(17.80; 0.89, 0.89; 1.78, 1.78)	(28.00; 2.80, 2.80; 4.20, 4.20)	(1.14; 0.11, 0.11; 0.17, 0.17)
19	(20.20; 1.01, 1.01; 2.02, 2.02)	(26.00; 2.60, 2.60; 3.90, 3.90)	(1.46; 0.15, 0.15; 0.22, 0.22)
20	(20.00; 1.00, 1.00; 2.00, 2.00)	(32.00; 3.20, 3.20; 4.80, 4.80)	(1.81; 0.18, 0.18; 0.27, 0.27)
21	(22.80; 1.14, 1.14; 2.28, 2.28)	(10.00; 1.00, 1.00; 1.50, 1.50)	(1.38; 0.14, 0.14; 0.21, 0.21)
22	(19.10; 0.96, 0.96; 1.91, 1.91)	(38.00; 3.80, 3.80; 5.70, 5.70)	(0.84; 0.08, 0.08; 0.13, 0.13)
23	(12.10; 0.61, 0.61; 1.21, 1.21)	(49.00; 4.90, 4.90; 7.35, 7.35)	(1.48; 0.15, 0.15; 0.22, 0.22)
24	(12.80; 0.64, 0.64; 1.28, 1.28)	(42.00; 4.20, 4.20; 6.30, 6.30)	(1.08; 0.11, 0.11; 0.16, 0.16)

11.3919 from (34) by 3.62%, and higher than 11.3640 from (35) by 3.88%. In addition, the proposed approach's distance measure d^2 is lower than that for Arefi and Taheri's (2015) approach by 9.7607 (13.94%), and lower than that for Chen and Nien's (2020) approach by 0.5144 (8.47%). As for the distance measure, $D_{p=1}$, the proposed approach is lower than Arefi and Taheri's (2015) approach by 8.5502 (23.44%), and lower than Chen and Nien's (2020) approach by 0.6351 (2.22%).

The four kinds of TIFN response, $\hat{Y}\%$, $\hat{Y}_{LSM}\%$, $\hat{Y}_{LAD}\%$, and $\hat{Y}_{CN}\%$ can also be graphically compared. As shown in Fig. 4, the most likely values of $\hat{Y}_{LSM}\%$, $\hat{Y}_{LAD}\%$ and $\hat{Y}_{CN}\%$ for all estimated TIFN observations are almost equivalent; however, the $\hat{Y}_{LSM}\%$ produced the widest spreads in three approaches, and $\hat{Y}_{CN}\%$ has the smallest spreads among three IFRMs. The results in Table 4 and Fig. 4 show that the proposed approach significantly outperforms the existing approach in terms of the similarity measure and the two distance measures.

5 Application

A dataset (Freund and Wilson 2003) is adopted in this section to demonstrate the applicability of the proposed models. The collected dataset contains 47 observations for estimating the weight of a tree. A number of trees were measured before

they were harvested based on four indices, namely the diameter at breast height (DBH) (X_1), height (X_2), age (X_3), and specific gravity of the wood (GRAV) (X_4). Subsequently, the trees were harvested and their weights (Y) were measured. Although the first two indices (DBH and height) can be easily measured, they are not precise since the trees usually have irregular shapes. GRAV is roughly estimated, and tree age can be exactly known.

Suppose that three indices and the weight in the dataset have measurement errors. Let the four indices be explanatory variables, and the weight be the response variable. For demonstration, TIFNs are used to characterize the three explanatory variables, $\hat{X}_1\%$, $\hat{X}_2\%$, and $\hat{X}_4\%$, and the response variable, $\hat{Y}\%$. The TIFNs are defined as follows. Let the original data be the most likely value (mean value) of the TIFN, and the left and right membership spreads of $\hat{X}_1\%$, $\hat{X}_2\%$, $\hat{X}_4\%$, and $\hat{Y}\%$ be randomly distributed from 8% to 10%, 8% - 10%, 10% - 12%, and 8% - 10% of the corresponding mean value, respectively. Similarly, let the left and right non-membership spreads be randomly distributed from 10% to 12%, 10% - 12%, 12% - 14%, and 10% - 12% of the corresponding mean value, respectively. The fuzzified data are listed in Table 5. Applying the proposed approach, the IFRM is built in terms of five ordinary regression models as:

Table 4 Performance comparisons of Example 2 in terms of three measures

No.	$\hat{Y}_{LSM}^{\%}$ (Arefi and Taheri 2015)			$\hat{Y}_{LAD}^{\%}$ (Chen and Nien 2020)			$\hat{Y}_{CN}^{\%}$ (Proposed approach)		
	SM	d^2	$D_{p=1}$	SM	d^2	$D_{p=1}$	SM	d^2	$D_{p=1}$
1	0.5608	0.6140	0.9881	0.9212	0.0119	0.1028	1.0000	0.0000	0.0000
2	0.2900	4.7794	2.0880	0.1748	3.5224	1.7459	0.1174	3.3278	1.8233
3	0.6541	0.3787	0.7722	0.9690	0.0132	0.0391	0.9561	0.0033	0.0571
4	0.6336	0.5027	0.9033	0.7283	0.0964	0.4227	0.7616	0.1229	0.3503
5	0.4487	2.2008	1.5992	0.4253	1.1581	0.9396	0.3205	1.0365	1.0176
6	0.2275	5.9011	2.3396	0.1208	5.0699	2.1275	0.0923	4.8527	2.2017
7	0.6333	0.4485	0.8068	0.8697	0.0064	0.1843	0.9153	0.0125	0.1119
8	0.6780	0.5456	0.7760	0.5028	0.7865	0.7543	0.5187	0.6729	0.8199
9	0.3912	2.9830	1.8886	0.1259	3.2841	1.9229	0.1021	3.4029	1.8437
10	0.2066	6.3307	2.4026	0.0463	6.5403	2.6741	0.0565	6.7792	2.6023
11	0.6819	0.5197	0.7176	0.6612	0.1892	0.4655	0.7487	0.1709	0.4132
12	0.6881	0.5945	0.8531	0.6662	0.4474	0.5443	0.6512	0.3730	0.6104
13	0.6193	1.0048	1.2657	0.8918	0.0313	0.2016	1.0000	0.0000	0.0000
14	0.5654	1.4272	1.3492	0.6946	0.4343	0.5301	0.6281	0.3648	0.6037
15	0.4448	2.0048	1.3707	0.2642	2.4260	1.4338	0.2889	2.2484	1.4987
16	0.5656	0.8963	0.9068	0.5286	0.3241	0.6243	0.6449	0.3153	0.5612
17	0.1034	7.5727	2.7066	0.0168	5.8791	2.5328	0.0584	6.0907	2.4666
18	0.5544	1.0316	1.1712	0.4173	0.5482	0.8588	0.4680	0.6201	0.7871
19	0.6550	0.4914	0.8955	0.7947	0.1653	0.2846	0.7536	0.1262	0.3551
20	0.5831	0.9669	1.2095	0.7854	0.2110	0.3276	0.7213	0.1616	0.4018
21	0.7904	0.1894	0.4609	0.7281	0.0942	0.3570	1.0000	0.0000	0.0000
22	0.0848	12.6771	3.4838	0.0201	11.2529	3.2242	0.0014	10.9156	3.3022
23	0.1657	8.5320	2.8745	0.0035	10.3513	3.3259	0.0000	10.5158	3.2411
24	0.1664	7.4196	2.6411	0.0073	7.9225	2.9318	0.0000	8.1386	2.8513
sum	11.3919	70.0124	36.4706	11.3640	60.7661	28.5555	11.8053	60.2517	27.9204

Fig. 4 Graphical comparison of four kinds of TIFN responses

Table 5 TIFN observations in dataset

No.	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$	$\mathcal{Y}$
1	(5.70; 0.53, 0.54; 0.64, 0.60)	(34.00; 2.82, 3.25; 3.69, 3.56)	10	(0.41; 0.04, 0.05; 0.06, 0.06)	(174.00; 16.84, 14.69; 18.24, 19.98)
2	(8.10; 0.76, 0.68; 0.91, 0.94)	(68.00; 5.44, 6.39; 7.70, 7.51)	17	(0.50; 0.04, 0.05; 0.05, 0.06)	(745.00; 68.24, 70.63; 84.04, 89.40)
3	(8.30; 0.70, 0.74; 0.89, 0.83)	(70.00; 6.72, 6.72; 7.70, 8.01)	17	(0.45; 0.04, 0.04; 0.05, 0.05)	(814.00; 76.84, 71.31; 91.17, 85.63)
4	(7.00; 0.68, 0.58; 0.82, 0.72)	(54.00; 4.47, 5.29; 5.44, 6.35)	17	(0.44; 0.04, 0.04; 0.05, 0.05)	(408.00; 34.27, 36.07; 42.27, 44.72)
5	(6.20; 0.51, 0.54; 0.67, 0.64)	(37.00; 3.43, 3.55; 3.85, 4.37)	12	(0.35; 0.03, 0.03; 0.04, 0.04)	(226.00; 22.33, 20.43; 23.41, 23.32)
6	(11.40; 0.93, 0.95; 1.21, 1.27)	(79.00; 6.86, 7.49; 9.39, 8.47)	27	(0.43; 0.04, 0.04; 0.05, 0.04)	(1675.00; 142.04, 134.00; 175.54, 192.29)
7	(11.60; 0.98, 1.09; 1.29, 1.29)	(70.00; 7.00, 5.68; 8.20, 8.06)	26	(0.50; 0.04, 0.04; 0.06, 0.05)	(1491.00; 129.42, 138.96; 164.61, 154.47)
8	(4.50; 0.36, 0.41; 0.53, 0.45)	(37.00; 3.39, 3.09; 3.74, 4.35)	12	(0.38; 0.04, 0.04; 0.04, 0.04)	(121.00; 9.83, 10.70; 12.54, 12.25)
9	(3.50; 0.31, 0.29; 0.36, 0.41)	(32.00; 2.57, 2.70; 3.64, 3.33)	15	(0.42; 0.04, 0.04; 0.04, 0.05)	(58.00; 4.76, 5.50; 6.89, 6.43)
10	(6.20; 0.53, 0.53; 0.67, 0.66)	(45.00; 3.74, 4.43; 5.38, 4.52)	15	(0.45; 0.04, 0.04; 0.05, 0.05)	(278.00; 24.24, 24.24; 30.36, 31.47)
11	(5.70; 0.52, 0.53; 0.61, 0.68)	(48.00; 4.72, 4.22; 4.82, 5.16)	20	(0.47; 0.04, 0.04; 0.05, 0.05)	(220.00; 20.59, 21.12; 24.64, 23.23)
12	(6.00; 0.56, 0.55; 0.67, 0.67)	(57.00; 5.45, 5.13; 6.68, 6.27)	20	(0.45; 0.04, 0.04; 0.05, 0.05)	(342.00; 30.78, 27.50; 36.94, 39.67)
13	(5.60; 0.54, 0.47; 0.58, 0.67)	(40.00; 3.31, 3.54; 4.13, 4.38)	20	(0.44; 0.04, 0.04; 0.05, 0.05)	(209.00; 17.05, 19.65; 21.40, 21.15)
14	(4.00; 0.35, 0.33; 0.48, 0.42)	(44.00; 4.19, 3.82; 5.23, 4.44)	27	(0.39; 0.03, 0.03; 0.04, 0.05)	(84.00; 7.83, 6.82; 9.68, 8.50)
15	(6.70; 0.65, 0.66; 0.80, 0.67)	(52.00; 4.91, 4.62; 5.60, 5.62)	21	(0.42; 0.04, 0.04; 0.05, 0.05)	(313.00; 25.42, 26.92; 35.06, 33.05)
16	(4.00; 0.35, 0.39; 0.47, 0.45)	(38.00; 3.72, 3.22; 4.41, 3.98)	27	(0.50; 0.05, 0.04; 0.05, 0.06)	(60.00; 5.38, 5.30; 7.13, 6.02)
17	(12.10; 1.20, 1.19; 1.39, 1.29)	(74.00; 6.16, 6.99; 7.96, 7.43)	27	(0.48; 0.04, 0.04; 0.06, 0.05)	(1692.00; 154.99, 161.76; 188.15, 187.47)
18	(4.50; 0.36, 0.45; 0.51, 0.49)	(37.00; 3.66, 3.15; 3.94, 4.06)	12	(0.38; 0.04, 0.03; 0.04, 0.04)	(74.00; 6.63, 6.72; 7.99, 8.38)
19	(8.60; 0.80, 0.69; 1.03, 1.00)	(60.00; 4.94, 4.90; 7.20, 7.13)	23	(0.50; 0.05, 0.04; 0.05, 0.06)	(515.00; 48.20, 44.08; 52.94, 54.80)
20	(9.30; 0.84, 0.84; 1.06, 0.94)	(63.00; 6.02, 5.12; 6.33, 6.65)	18	(0.46; 0.04, 0.04; 0.05, 0.05)	(766.00; 69.25, 70.47; 82.12, 82.12)
21	(6.50; 0.62, 0.64; 0.69, 0.70)	(57.00; 4.88, 5.43; 6.73, 5.97)	18	(0.47; 0.04, 0.04; 0.05, 0.05)	(345.00; 28.70, 30.36; 38.64, 36.71)
22	(5.60; 0.56, 0.48; 0.58, 0.63)	(46.00; 4.51, 4.08; 5.32, 4.97)	12	(0.41; 0.04, 0.04; 0.05, 0.04)	(210.00; 19.32, 20.58; 22.18, 22.85)
23	(4.30; 0.36, 0.43; 0.50, 0.46)	(41.00; 3.36, 4.07; 4.69, 4.84)	12	(0.38; 0.04, 0.04; 0.04, 0.04)	(100.00; 8.84, 8.48; 10.64, 11.32)
24	(4.50; 0.41, 0.42; 0.47, 0.53)	(42.00; 4.20, 4.10; 4.77, 4.47)	12	(0.46; 0.04, 0.04; 0.05, 0.05)	(122.00; 10.30, 10.20; 13.08, 13.47)
25	(7.70; 0.62, 0.75; 0.87, 0.92)	(64.00; 5.32, 5.12; 6.94, 7.45)	19	(0.48; 0.05, 0.05; 0.05, 0.05)	(539.00; 44.20, 47.86; 63.60, 60.58)
26	(8.80; 0.86, 0.82; 0.98, 0.91)	(70.00; 6.27, 6.58; 7.56, 7.73)	22	(0.50; 0.05, 0.05; 0.05, 0.05)	(815.00; 67.16, 77.91; 95.19, 92.26)
27	(5.00; 0.50, 0.41; 0.51, 0.58)	(53.00; 4.32, 5.19; 5.32, 5.70)	23	(0.49; 0.05, 0.05; 0.06, 0.05)	(194.00; 16.06, 16.22; 23.05, 22.19)
28	(5.40; 0.54, 0.47; 0.59, 0.55)	(61.00; 5.44, 5.78; 6.61, 7.22)	23	(0.49; 0.04, 0.04; 0.05, 0.05)	(280.00; 22.74, 24.86; 28.22, 33.04)
29	(6.00; 0.52, 0.56; 0.64, 0.69)	(56.00; 5.53, 4.75; 5.60, 5.85)	23	(0.44; 0.04, 0.04; 0.05, 0.05)	(296.00; 26.17, 29.01; 31.26, 33.74)
30	(7.40; 0.60, 0.73; 0.79, 0.77)	(52.00; 4.16, 4.89; 6.24, 5.78)	14	(0.47; 0.04, 0.05; 0.06, 0.06)	(462.00; 38.07, 41.76; 52.48, 51.93)
31	(5.60; 0.53, 0.52; 0.64, 0.64)	(48.00; 4.78, 4.38; 5.32, 5.51)	19	(0.44; 0.04, 0.04; 0.05, 0.04)	(200.00; 19.28, 16.24; 23.28, 21.84)
32	(5.50; 0.46, 0.48; 0.61, 0.64)	(50.00; 5.00, 4.06; 5.32, 5.68)	19	(0.51; 0.04, 0.05; 0.06, 0.05)	(229.00; 22.63, 20.98; 24.46, 24.09)
33	(4.30; 0.36, 0.38; 0.49, 0.50)	(50.00; 4.46, 4.64; 5.30, 6.00)	19	(0.41; 0.04, 0.04; 0.05, 0.05)	(125.00; 12.10, 12.50; 14.00, 14.35)
34	(4.20; 0.39, 0.37; 0.46, 0.46)	(31.00; 2.62, 2.57; 3.22, 3.58)	10	(0.41; 0.03, 0.03; 0.05, 0.04)	(84.00; 8.03, 7.69; 9.98, 9.17)
35	(3.70; 0.32, 0.33; 0.38, 0.39)	(27.00; 2.53, 2.38; 3.16, 3.05)	10	(0.42; 0.04, 0.04; 0.05, 0.04)	(70.00; 6.24, 6.66; 8.20, 8.15)
36	(6.10; 0.60, 0.54; 0.61, 0.62)	(39.00; 3.87, 3.26; 4.59, 4.29)	10	(0.47; 0.05, 0.04; 0.05, 0.05)	(224.00; 22.13, 20.52; 22.49, 26.79)
37	(3.90; 0.32, 0.34; 0.39, 0.44)	(35.00; 3.11, 3.19; 3.56, 3.72)	19	(0.43; 0.04, 0.04; 0.05, 0.05)	(99.00; 8.87, 8.40; 10.06, 11.64)
38	(5.20; 0.51, 0.45; 0.60, 0.52)	(48.00; 4.03, 3.90; 5.43, 5.76)	13	(0.44; 0.04, 0.04; 0.05, 0.04)	(200.00; 16.08, 18.40; 22.56, 21.68)
39	(5.60; 0.48, 0.50; 0.62, 0.60)	(47.00; 3.80, 4.15; 5.19, 5.08)	13	(0.47; 0.04, 0.04; 0.05, 0.05)	(214.00; 20.12, 19.26; 23.28, 23.28)
40	(7.80; 0.73, 0.69; 0.90, 0.87)	(69.00; 6.84, 5.80; 7.51, 7.04)	13	(0.47; 0.04, 0.05; 0.05, 0.05)	(712.00; 67.78, 68.07; 78.89, 74.62)
41	(6.10; 0.60, 0.56; 0.62, 0.71)	(49.00; 4.49, 4.76; 5.57, 5.19)	13	(0.46; 0.04, 0.04; 0.05, 0.05)	(297.00; 28.87, 25.30; 33.86, 33.86)
42	(6.10; 0.61, 0.59; 0.66, 0.72)	(44.00; 3.57, 3.73; 5.28, 4.68)	13	(0.45; 0.04, 0.04; 0.05, 0.05)	(238.00; 23.13, 23.42; 25.80, 26.18)
43	(4.00; 0.36, 0.37; 0.45, 0.45)	(34.00; 3.20, 3.02; 3.98, 3.41)	13	(0.42; 0.04, 0.04; 0.04, 0.05)	(89.00; 7.69, 7.97; 9.72, 10.64)
44	(4.00; 0.38, 0.37; 0.43, 0.42)	(38.00; 3.53, 3.24; 3.97, 4.15)	13	(0.41; 0.04, 0.04; 0.04, 0.04)	(76.00; 6.93, 7.05; 7.66, 7.60)
45	(8.00; 0.76, 0.68; 0.87, 0.86)	(61.00; 5.27, 5.93; 6.59, 6.88)	13	(0.51; 0.05, 0.05; 0.05, 0.06)	(614.00; 51.82, 54.28; 67.29, 71.96)
46	(5.20; 0.51, 0.45; 0.58, 0.52)	(47.00; 4.32, 3.84; 5.19, 5.21)	13	(0.43; 0.04, 0.04; 0.05, 0.05)	(194.00; 17.30, 17.85; 22.35, 22.12)
47	(3.70; 0.32, 0.31; 0.41, 0.42)	(33.00; 2.67, 2.89; 3.67, 3.87)	13	(0.39; 0.03, 0.04; 0.04, 0.04)	(66.00; 5.39, 5.41; 6.86, 6.60)

$$\begin{aligned}
\hat{m}_y &= -374.534 + 121.7348m_{x_1} + 1.2011m_{x_2} \\
&\quad + 0.8838m_{x_3} - 245.7830m_{x_4} \\
\hat{s}_y^{ML} &= -26.7245 + 101.4058s_{x_1}^{ML} + 1.5239s_{x_2}^{ML} - 222.9970s_{x_4}^{ML} \\
\hat{s}_y^{MR} &= -54.499 + 94.3404s_{x_1}^{MR} + 2.5159s_{x_2}^{MR} + 478.5705s_{x_4}^{MR} \\
\hat{s}_y^{VL} &= -44.1956 + 106.1367s_{x_1}^{VL} + 2.5432s_{x_2}^{VL} - 56.3952s_{x_4}^{VL} \\
\hat{s}_y^{VR} &= -51.2377 + 113.7085s_{x_1}^{VR} + 2.0319s_{x_2}^{VR} + 64.4922s_{x_4}^{VR}
\end{aligned} \tag{37}$$

In the above model, the parameters of the third explanatory variable, age (X_3), for estimating spreads $s_{x_3}^{ML}$, $s_{x_3}^{MR}$, $s_{x_3}^{VL}$ and $s_{x_3}^{VR}$ are zero, since ages are numeric in the dataset. In addition, the resulting optimal parameters for the right and left spreads of the membership/non-membership function for the other three explanatory variables are not the same, since TIFN observations in this example are not symmetrical. The performance analyses of this application are presented in the following section.

6 Discussion

From the above demonstrations and applications, this section provides more explanations and introduces some further investigations. In Example 1, Al-Qudaimi's (2021) approach is adopted to construct an IFRM with a negative TIFN explanatory variable. This approach derives the formulation processes of IFRMs, considering that TIFN explanatory variables may be negative, or their membership and/or non-membership functions are across zero in the domain variable. However, some unreasonable conditions are defined in the formulation processes, since they violate the basic definitions of an IFN so that the formulated IFRMs are problematic. The formulations are carried out by non-linear programming problems to determine the parameters of the IFRM, by which the resolution efficiency is influenced. In addition, this approach has the problem that wider spreads could be produced. Using the dataset from Al-Qudaimi(2021), the proposed approach is also applied to formulate an IFRM by five ordinary regression models, even for negative TIFN explanatory variables. Based on the performance comparisons, the proposed approach is much better than Al-Qudaimi's (2021) approach.

In Example 2, the parameter sign of $\hat{X}_1^{\hat{q}}$ in the IFRM from Arefi and Taheri's (2015) approach is negative; however, their formulation processes are derived based on the assumption that all parameters are positive TIFNs. Chen and Nien's (2020) approach and the proposed approach are not delimited by such a restriction. Furthermore, in this example, the spreads of the TIFN response and explanatory variables of the observations are all symmetrical TIFNs. They are defined as a proportion of the corresponding mean value so that mathematically the spread can be considered as a function of the corresponding mean value. Consequently,

the parameters of the estimated spreads in Eq. (36) are also proportional to the corresponding ones of the estimated mean value. For example, for $\hat{s}_y^{ML}$, the following relations can be obtained: $\hat{b}_0^m = \frac{1}{0.05}\hat{b}_0^{ML} = 20\hat{b}_0^{ML}$, $\hat{b}_1^m = \frac{0.1}{0.05}\hat{b}_1^{ML} = 2\hat{b}_1^{ML}$, and $\hat{b}_2^m = \frac{0.1}{0.05}\hat{b}_2^{ML} = 2\hat{b}_2^{ML}$. As another case, for $\hat{s}_y^{VL}$, the relations of $\hat{b}_0^m = \frac{1}{0.1}\hat{b}_0^{VL} = 10\hat{b}_0^{VL}$, $\hat{b}_1^m = \frac{0.15}{0.1}\hat{b}_1^{VL} = 1.5\hat{b}_1^{VL}$, and $\hat{b}_2^m = \frac{0.15}{0.1}\hat{b}_2^{VL} = 1.5\hat{b}_2^{VL}$ are determined. This means that only the estimated mean equation, $\hat{m}_y = \hat{b}_0^m + \hat{b}_1^m m_{x_1} + L + \hat{b}_p^m m_{x_p}$, is needed in the formulation processes, if the spreads are set as above. However, if the spreads of the observed TIFN response and explanatory variables are not set as above, as is the general case in real-world applications, the proposed approach can still be applied (the above proportional relations will not arise). In contrast, if the observations are not symmetrical TIFNs, the method proposed by Arefi and Taheri (2015) cannot be employed.

From Fig. 4, with smaller spreads, the coverage of the estimated TIFN responses by the proposed approach better matches that of the observed ones, and thus the distances between them are smaller than those from the other two approaches. As reflected by $\hat{Y}_{LSM}^{\hat{q}}$, $\hat{Y}_{LAD}^{\hat{q}}$ and $\hat{Y}_{CN}^{\hat{q}}$ in Fig 4, they have almost equivalent most likely values but greatly different spreads, leading to $\hat{Y}_{CN}^{\hat{q}}$ having lower estimation error in terms of distance measures, as shown in Table 4. Evidently, this indicates that, individually considering fuzziness and randomness of IFNs, the proposed approach can significantly improve the estimation performance, since the wider spreads will not be produced.

In the application case, the TIFN response and explanatory variables are not symmetrical, so Arefi and Taheri's approach (2015) cannot be employed. By applying the proposed approach, the estimated TIFN response variable can be determined to be asymmetrical. The total absolute distance measure ($D_{p=1}$) between the observed and estimated weights is 4027.08 in this case, which is quite close to the value of 4012.83 obtained using the traditional regression method with the mean value of the dataset. Particularly, the explanatory variable, age, is numeric in the dataset so that only randomness exists for this variable. By applying the proposed approach, fuzziness characterized by spreads of this variable in the four ordinary regression models are ignored, as shown in (37), so that this variable does not add the fuzziness to the estimated TIFN response, i.e., the spread is not increased from this variable. This shows that the proposed approach can effectively be applied to estimation purposes for TIFN or mixed-type observations in uncertain environments.

In sum, some important features of the proposed approach can be briefly described as follows. With the consideration of fuzziness and randomness of TIFN observations, the

proposed approach formulates IFRMs by five ordinary regression models to characterize the five components of the estimated TIFN response. A mathematical programming problem is formulated to determine the optimal parameters of the five ordinary regression models. The resulting parameters are numerical, and the five components of the estimated TIFN response are directly determined using those parameters. Such formulation processes have some advantages in that they improve on the weaknesses of existing approaches as follows. Firstly, the wide spreads are not produced in the estimated TIFN response, since intuitionistic fuzzy arithmetic is not used and the spreads are individually determined. In contrast to existing approaches, the parameter signs of IFRMs are not a concern in the formulation processes. The proposed approach can deal with various data types of observations, such as asymmetrical IFNs, negative IFNs, or even numerical numbers. In addition to the performance in estimation purposes, the proposed approach is general and flexible in real-world cases with uncertain information.

7 Conclusion

With the consideration of fuzziness and randomness of TIFN observations, this study applied mathematical programming approaches based on the criterion of the least absolute deviations to build IFRMs in terms of five ordinary regression models representing the components of the estimated TIFN response variable, respectively. Each component of the estimated TIFN response variable is determined by the corresponding ordinary regression model. There are many advantages with the proposed approach. Unlike an existing approach, the observations used in this study are not limited to symmetrical TIFNs. The randomness and fuzziness of TIFN response and explanatory variables are included in the formulation processes so that each component of the estimated TIFN response variable can be individually estimated, subject to the definition of a TIFN, thus avoiding wider spreads that may include unnecessary information. This formulation process also solves the problem of determining the parameter signs in IFRM approaches beforehand, since the proposed approach applies mathematical programming to estimate each component of the estimated TIFN response variable. In particular, besides the TIFN explanatory variables, observations that include numeric variables are also allowed in the proposed approach, increasing its applicability. Demonstrations and comparisons showed that the proposed approach outperforms existing approaches in terms of similarity and distance measures. In future research, different kinds of regression models, not limited to linear models, representing the components of the estimated TIFN response variable could be investigated for possibly enhancing the model performance.

Compliance with Ethical Standards:

Funding This study was funded by the Ministry of Science and Technology, Taiwan, Republic of China, under contracts MOST 104-2410-H-006-054-MY3 and MOST 107-2410-H-006-041-MY2.

Conflict of Interest L.-H. Chen declares that he has no conflict of interest. S.-H. Nien declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Al-Qudaimi A (2021) Ishita Approach to Construct an Intuitionistic Fuzzy Linear Regression Model. *Fuzzy Optimization and Modeling Journal* 2 (2):1-11
- Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, Dongarra J, Du Croz J, Greenbaum A, Hammarling S, McKenney A, Sorensen D (2017) LINGO The Modeling Language and Optimizer. LINDO Systems, Chicago, Illinois
- Arefi M, Taheri SM (2015) Least-Squares Regression Based on Atanassov's Intuitionistic Fuzzy Inputs-Outputs and Atanassov's Intuitionistic Fuzzy Parameters. *IEEE Transactions on Fuzzy Systems* 23 (4):1142-1154. doi:10.1109/Tfuzz.2014.2346246
- Atanassov KT (1986) Intuitionistic Fuzzy-Sets. *Fuzzy Sets and Systems* 20 (1):87-96. doi:10.1016/S0165-0114(86)80034-3
- Atanassov KT (1999) Intuitionistic fuzzy sets: Theory and applications. Physica-Verlag, New York. doi:10.1007/978-3-7908-1870-3
- Boukezzoula R, Jaulin L, Desrochers B, Foulloy L (2021) Thick gradual sets and their computations: Application for determining the uncertain zone explored by an underwater robot. *Engineering Applications of Artificial Intelligence* 102:104287
- Chakraborty D, Jana DK, Roy TK (2014) Arithmetic operations on generalized intuitionistic fuzzy number and its applications to transportation problem. *Opsearch* 52 (3):431-471. doi:10.1007/s12597-014-0194-1
- Chan KY, Engelke U (2017) Varying Spread Fuzzy Regression for Affective Quality Estimation. *IEEE Transactions on Fuzzy Systems* 25 (3):594-613. doi:10.1109/Tfuzz.2016.2566812
- Chen LH, Hsueh CC (2007) A mathematical programming method for formulating a fuzzy regression model based on distance criterion. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics* 37

- (3):705-712. doi:10.1109/tsmcb.2006.889609
- Chen LH, Hsueh CC (2009) Fuzzy Regression Models Using the Least-Squares Method Based on the Concept of Distance. *IEEE Transactions on Fuzzy Systems* 17 (6):1259-1272. doi:10.1109/Tfuzz.2009.2026891
- Chen LH, Nien SH (2020) Mathematical programming approach to formulate intuitionistic fuzzy regression model based on least absolute deviations. *Fuzzy Optimization and Decision Making* 19:191-210
- Choi SH, Buckley JJ (2008) Fuzzy regression using least absolute deviation estimators. *Soft Computing* 12 (3):257-263. doi:10.1007/s00500-007-0198-3
- Fong D, Hong IH (2010) The Cournot Production Game with Entropy Spreads. Paper presented at the The 40th International Conference on Computers Industrial Engineering, Awaji, Japan,
- Freund RJ, Wilson WJ (2003) *Statistical methods*. 2 edn. Academic Press, New York
- Grzegorzewski P Distances and orderings in a family of intuitionistic fuzzy numbers. In: 3rd Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT '03), Zittau, Germany, 2003. IRAFM, pp 223-227
- Guha D, Chakraborty D (2010) A Theoretical Development of Distance Measure for Intuitionistic Fuzzy Numbers. *International Journal of Mathematics and Mathematical Sciences* 2010:1-25. doi:10.1155/2010/949143
- Hesamian G, Akbari MG, Roozbeh M (2020) Intuitionistic Fuzzy Partial Logistic Regression Model Using Ridge Methodology. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 28 (4):527-543
- Ho CH, Chang PT, Hung KC, Lin KP (2019) Developing intuitionistic fuzzy seasonality regression with particle swarm optimization for air pollution forecasting. *Industrial Management & Data Systems* 119 (3):561-577. doi:10.1108/imds-02-2018-0063
- Kao C, Chyu CL (2002) A fuzzy linear regression model with better explanatory power. *Fuzzy Sets and Systems* 126 (3):401-409. doi:10.1016/S0165-0114(01)00069-0
- Kelkinnama M, Taheri SM (2012) Fuzzy least-absolutes regression using shape preserving operations. *Information Sciences* 214:105-120. doi:10.1016/j.ins.2012.04.017
- Liu HW, Shi KQ (2000) Intuitionistic fuzzy numbers and intuitionistic distribution numbers. *Journal of Fuzzy Mathematics* 8 (4):909-918
- Mahapatra GS, Roy TK (2009) Reliability evaluation using triangular intuitionistic fuzzy numbers arithmetic operations. *World Academy of Science, Engineering and Technology* 3 (2):350-357
- Parvathi R, Malathi C, Akram M, Atanassov KT (2013) Intuitionistic fuzzy linear regression analysis. *Fuzzy Optimization and Decision Making* 12 (2):215-229. doi:10.1007/s10700-012-9150-9
- Patriota AG (2013) A classical measure of evidence for general null hypotheses. *Fuzzy Sets and Systems* 233:74-88. doi:10.1016/j.fss.2013.03.007
- Spiliotis M, Garrote L (2021) Unit hydrograph identification based on fuzzy regression analysis. *Evolving Systems*:1-22. doi:10.1007/s12530-021-09380-7
- Tanaka H, Uejima S, Asai K (1982) Linear-Regression Analysis with Fuzzy Model. *Ieee T Syst Man Cyb* 12 (6):903-907
- Watada J, Wang SM, Pedrycz W (2009) Building Confidence-Interval-Based Fuzzy Random Regression Models. *IEEE Transactions on Fuzzy Systems* 17 (6):1273-1283. doi:10.1109/Tfuzz.2009.2028331
- Xu ZS, Yager RR (2006) Some geometric aggregation operators based on intuitionistic fuzzy sets. *International Journal of General Systems* 35 (4):417-433. doi:10.1080/03081070600574353
- Yolcu OC, Bas E, Egrioglu E, Yolcu U (2020) A new intuitionistic fuzzy functions approach based on hesitation margin for time-series prediction. *Soft Computing* 24 (11):8211-8222
- Zadeh LA (1965) Fuzzy Sets. *Information and Control* 8 (3):338-353. doi:10.1016/S0019-9958(65)90241-X