IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Estimating the Fuzzy Regression Panel Data Model Based on Approximate Bayesian Computation with Application

Ahmed mutlag abdulateef[1], Emad Hazim Aboudi[2]

[1,] Department of Statistics, College of Administration and Economics, University of Baghdad Iraq, Baghdad
[2] Department of Statistics, College of Administration and Economics, University of Baghdad Iraq, Baghdad
[1]E-mail : ahmed.abd2101p@coadec.uobaghdad.edu.iq
[2]E-mail : emadhazim@coadec.uobaghdad.edu.iq

ARTICLE INFO

Article history:
Received 1 February 2025
Revised 2 February 2025
Accepted 7 March 2025
Available online 17 April 2025

Keywords:
Fuzzy Fixed Effect
Tanaka Fuzzy Fixed Effect
Quadratic Fuzzy Fixed Effect
Fuzzy Fixed Effect Least Absolute
approximate Bayesian computation

ABSTRACT

This study aims to estimate the parameters of fuzzy regression for panel data using a fixed regression model. To achieve greater accuracy in estimation, an approach is proposed that combines two estimation methods, leveraging the advantages of both probabilistic and traditional methods. The fixed regression model provides an integrated framework for analyzing cross-sectional and temporal data, contributing to a comprehensive analysis of panel data. This approach was applied to water pollution data of the Euphrates River using the fuzzy fixed regression model. The root mean square error (RMSE) criterion was used to compare different estimation methods. The results showed that the proposed methods generally outperformed the estimation methods, and gave better performance than probabilistic and traditional estimation methods. This study presents a new application for panel data using fuzzy regression, highlighting the benefit of combining traditional and probabilistic methods to achieve better estimations.

1. Introduction

Panel data combines cross-sectional and time series data. It can be defined as a set of t time series and cross-sectional data sets. In this context, observations at the cross-sectional level i represent cross-sectional data, while observations over a specified period j represent time series data. Alternatively, panel data can be defined as data obtained from m repeated observations of a phenomenon across N cross-sections during a specified time series. This allows the phenomenon under study to vary at two levels: horizontally across time and vertically across different cross-sections. It should be noted that most researchers use the terms "longitudinal data" and "panel data" interchangeably to refer to the same concept, without any difference in definition, meaning, or content. However, a few researchers distinguish between two types of panel data: "panel data," which refers to data with a large number of cross-sections and a short time period, and "cross-section-time series data," which includes a relatively smaller number of cross-sections and a reasonable length of time period.

There are many phenomena whose data are ambiguous, meaning that their value cannot be determined by a single value, and this data is represented by estimating a model that describes them and estimating their parameters, especially when the inputs are not ambiguous and the outputs and parameters are fuzzy, including panel data whose data contain ambiguity and fuzziness, and therefore the idea of this research resulted in estimating panel

Corresponding author E-mail address: ahmed.abd2101p@coadec.uobaghdad.edu.iq
https://doi.org/10.62933/jt7pmy71

data when the data of the dependent variable and parameters are fuzzy and the independent variable is accurate. Recently, fuzzy regression has not been sufficiently studied in the context of longitudinal data, although both topics are of great importance in different research fields. Fuzzy regression is an effective tool for handling uncertain or imprecise data, contributing to the development of more flexible models for data processing. panel data is a type of data that collects observations more different time periods for the same bodies, which provides deep insights into the development of variables over time. This study will review in-depth literature reviews on both fuzzy regression and panel data separately, with the aim of determine the theoretical foundations of each, and stressing potential overlaps. This contains reviewing the different applications of fuzzy regression and how it is used to address uncertainty in data, in addition to reviewing the approaches used with panel data and the importance of their integrated with more inclusiveanalytical models such as fuzzy regression.

[1] shows a fuzzy linear regression model with least absolute value based on two estimators with least absolute deviation (LAD). The model is determined to have crisp inputs, fuzzy outputs, and fuzzy parameters, processing the main challenges in fuzzy regression analysis. A new distance measure for triangular fuzzy numbers is proposed, which enhances the evaluation of differences between fuzzy variables. In addition, the study uses a similarity measure for triangular fuzzy numbers to evaluate the fit between observed and estimated values. Through three examples, the proposed model shows superior performance compared to existing fuzzy regression models based on the least squares (LS) method. The robustness of the model is evaluated, highlighting its ability to handle outliers effectively, and its application to datasets with missing values illustrates its reliability and adaptability in real world.

[2] in his study introduced a new approach to solve the fuzzy linear regression problem using crisp input and fuzzy output data, focusing on overcome some drawbacks of probabilistic methods and classic least squares methods. Probabilistic methods focus on the embedding feature, while least squares methods focusing on the central tendency. Therefore, Wang suggests a new fuzzy linear regression method based on approximate Bayesian computation (ABC), which is an alternative to classical methods in optimization the fuzzy regression model. The method uses the likelihood-free inference algorithm ABC to generate samples of unknown model parameters from the Bayesian posterior distribution, which can solution to overcome the harder of determining the likelihood function in the fuzzy environment.

In a study conducted by [3], the pollution of the Tigris River water in Baghdad was studied due to the presence of organic and inorganic materials that spoil the water quality, which negatively affects human health. The data represent the number of people infected with amoeba disease in both sides of Karkh and Rusafa as a dependent variable (Y) over the year 2018 at a monthly rate (12 months). The study relied on seven concentrations of pollutants as explanatory variables (Xi), which represent cross-sectional data for ten stations located on the banks of the Tigris River. Due to the use of temporal and cross-sectional data, longitudinal data models (Panel Data) were applied using parametric and non-parametric longitudinal models, with non-parametric estimators including the weighted and unweighted Nadaria-Watson estimator. The goal was to determine which of the estimators is the most efficient and gives the best model for predicting the number of recorded infections.

In [4] introduced dissertation, he discussed address the issue of selecting the most appropriate panel data model for studying and analyzing the value of industrial production and some of the influencing factors. The selection of the model involved determining the nature of the panel data used, relying on the Lagrange Multiplier Test, the Hausman Test, and comparisons between the estimated nonparametric models using the Nadaraya-Watson method, the Profile Least Squares method, and the Speckman method.

In the study of [5], fuzzy panel data analysis (FPDA) was show as a method to overcome some limitations of classical panel data analysis (PDA). Classical panel data requires statistical assumptions such as homogeneity, autocorrelation, and stationarity, which are often difficult to satisfy in practical applications. FPDA suggests to estimate the regression parameters of panel data using triangular fuzzy numbers, which helps to address these restrictions. To evaluation the efficiency of FPDA, it was apply with PDA to GDP data of five-country groups for the period 2005-2013. The best of the two models was compared using the criteria of mean absolute error (MAPE), root mean square error (RMSE), and (VAF). The results shown that FPDA is an effective and practical method especially in cases where the required statistical assumptions are not met.

In the study of [6], the problem of multicollinearity in fuzzy models that makes the fuzzy least squares estimator (FLSE) unsuitable for estimating a fuzzy regression model is addressed. The study relied on the fuzzy bridge regression (FBRE) method using triangular fuzzy numbers to overcome this problem, with the use of the (VIF) to detection multicollinearity when the crisp inputs and outputs and parameters are fuzzy. The results shown the better of the fuzzy bridge regression model in reduce the (MSE) through simulation experiments, indicate the efficiency of this model in provide accurate estimates under multicollinearity.

2. Methodology

2.1 Panel Data Models:

When Panel Data (panel data) combines the spatial or cross-sectional dimension and the temporal dimension, it combines the positive and negative aspects of these two dimensions. In addition to the problems that panel data can suffer from, such as the problem of heterogeneity of variance, the problem of error correlation, data instability, and characterization problems, there are resulting difficulties. The combination of these two dimensions consists of describing and clarifying the coefficients of the cross-sectional dimension, that is, whether the components of these coefficients have fixed or random effects. On this basis, these models were divided into three types to allow the analysis of this type of data into three sections.[7] [8]

2.2 Pooled Regression Model (PRM)

It is considered one of the simplest types of panel data models in which the regression coefficients are constant for all cross-sectional units over time, the aggregate panel data model differs from the multiple regression model because it neglects the effect of time, and is written according to the following formula:

$$Y_{it} = B_o + \sum_{k=1}^{K} B_k X_{kit} + U_{it} \quad \ldots 1$$

It represents:

Y_{it} : the value of the depended variable in the unit of section (i) at the time interval (t).

X_{kit} : the explanatory variables of the section (i) at the time interval (t).

B_o : represent The intercept parameter.

B_k : The vector represents the unknown regression coefficients.

U_{it}: represents the random error vector in the unit of section (i) at the time interval (t) with a mean of zero and variance σ_{ij}^2 Uit∼N(0,σ2).

2.3 Fixed Effects Regression Model (FEM)

It is also known as the Covariance Analysis Model and assumes that the effects model (B_{oi}) vary among fixed cross-sectional units. The intercept parameter remains constant across time periods, meaning that the value of the segment parameter does not change over time while the slope coefficients (B_{oi}) constant. This model assumes homogeneity in error variance for all cross-sections (observations) and assumes no autocorrelation within a specific time period between cross-sectional units. It is expressed by the following formula:[7]

$$Y_{it} = B_{oi} + \sum_{k=1}^{K} B_k X_{kit} + U_{it} \quad \ldots 2$$

B0i is a fixed effect specific to each cross-sectional unit i, and it is estimated using dummy variables or within-group transformations.

2.4 Random Effected Model (REM)

In the random effects model, the intercept parameter (B_o) changes randomly, and it is assumed either non-homogeneity of the error variance between the cross-sectional observations or the existence of an autocorrelation over time between the cross-sectional observations in a specific period of time. It is called the error components model or the composite error model because it contains two error components (Error) $\varphi_{ij} = u_i + v_{i_j}$, assumed that each a cross-sectional unit within the temporal effect is a condition that differs from the rest of the cross-sectional units, so the error components are combined, that is, between the difference within each cross-sectional unit across time periods in addition to the difference between the cross-sectional units , which can be formulated as follows: [9]

$Y_{it} = B_o + \sum_{k=1}^{K} B_k X_{kit} + \varphi_{ij}$ …3

Where $\varphi_{ij} = u_i + v_{i_j}$ represents the compound error.

2.5 Estimate Fixed Effects Regression Model (FERM)

$$Y_{it} = B_{oi} + \sum_{k=1}^{K} B_k X_{kit} + U_{it} \quad \ldots 4$$

(i=1,2 … , N) (t=1,2 … , T) (k=1,2 … , K). ($B_{oi} = \overline{B}_o + u_i$) represents the constant term for the cross section (i), ($\overline{B}_o$) the average constant term, and (u_i) represents the effect resulting from deleting the time-invariant cross-section variables. In other words, (u_i) represents a component that is constant over time and variable from one cross section to another (spatial effect).

Therefore, the model (4) It can be expressed as follows:

$$Y_{it} = \overline{B}_o + u_i + \sum_{k=1}^{K} B_{ik} X_{kit} + U_{it} \quad \ldots 5$$

under the assumption that (B_{oi}) are fixed parameters to be estimated with regression parameters (B_k) and ($Eu^2{}_{it} = \sigma_u^2$) and that (U_{it}) are independent random variables distributed by $Eu_{it} = 0$ and $\sigma_u^2 \mathrm{I}_{NT}$. This model is known as It is a deaf variable model and can be written as follows: [10]

$$Y_{it} = \sum_{i=1}^{N} B_{oi} D_{jt} + \sum_{k=1}^{K} B_{ik} X_{kit} + U_{it} \quad \ldots 6$$

Where is (D_{jt}) represents the dummy variables and takes values equal to zero or one, noting that:

$$D_{jt} = \begin{cases} 1 & \text{if i = j} \\ 0 & \text{if i} \neq \text{j} \end{cases}$$

The general form of the model is written as follows: [11] [10]

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} = \begin{bmatrix} j_T & 0 & \cdots & 0 & X_{s1} \\ 0 & j_T & \cdots & 0 & X_{s2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & j_T & X_{sN} \end{bmatrix} \begin{bmatrix} B_{01} \\ B_{02} \\ \\ B_{0N} \\ B_s \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}$$

More briefly, the above model can be rewritten after using the direct multiplication method As follows (Kronecker product): [10]

$$Y = [\mathrm{I}_N \otimes j_T X_s] \begin{bmatrix} B_{0\mathrm{i}} \\ B_s \end{bmatrix} + U \quad \ldots 7$$

Where $D = I_N \otimes j_T$.

Since:

$I_N \otimes j_T$: The matrix of dummy variables of the order (NT*N).

I_N: represents the unit matrix of the order (N*N).

($\mathrm{I}_N \otimes j_T X_s$) represents the matrix of independent variables excluding the constant term, and it is of the order NT*(N+K) and it is called the information matrix.

Under these assumptions, the ordinary least squares estimators with denominator variables (fixed effects model) will take the following form**:** [12]

$$b_{s(FE)} = [X'_s(\mathrm{I}_N \otimes D_T) X_s]^{-1} X'_s(\mathrm{I}_N \otimes D_T) Y = \left[\sum_{i=1}^{N} X'_{si} D_T X_{si}\right]^{-1} \sum_{i=1}^{N} X'_{si} D_T Y_i \quad \ldots 8$$

As $b_{s(FE)}$refers to the estimator of the regression coefficients for the deaf variable model, or it is directed to the estimates of the parameters of the fixed effects regression model, while the estimates of the parameters of the fixed terms (bar) are calculated from the formula:

$$b_{0i} = \bar{Y}_{i.} - \bar{X}_{i.} b_s \quad \ldots 9$$

i= 1,2,…,N

The above estimates have the property of the best linear unbiased estimate(BLUE) with a variance and covariance matrix:

$$Var - Cov(b_{(FE)}) = \sigma_e^2 \begin{bmatrix} \mathrm{TI}_N & (\mathrm{I}_N \otimes j_T)\, X_s \\ X'_s(\mathrm{I}_N \otimes j_T) & X'_s X_s \end{bmatrix}^{-1} \quad \ldots 10$$

Hence, the unbiased estimator of the variance σ_e^2 is calculated as follows:

$$S_e^2 = \frac{\varepsilon'\varepsilon}{NT - (N + K)}$$

Where:

$$\varepsilon = Y - [\mathrm{I}_N \otimes j_T\, X_s] \begin{bmatrix} b_{0\mathrm{i}} \\ b_s \end{bmatrix} = [\mathrm{I}_N \otimes D_T]Y - [\mathrm{I}_N \otimes D_T]X_s b_s \quad \ldots 11$$

2.6 Fuzzy Numbers

The fuzzy number $\widetilde{A}$ is characterized as a fuzzy set on the real number line R , subject to the following conditions:: [13]

(i) $\widetilde{A}$ is a normal and convex fuzzy set,

(ii) Its membership function $\widetilde{A}$ is upper semicontinuous,

(iii) The α-level set $\widetilde{A}$ is bounded for each α ∈[0; 1].

2.7 LR-type fuzzy number and cross-sectional notation

One of the most commonly used fuzzy numbers in the literature is the LR-type fuzzy numbers proposed by Dubois and Prade (1980). In fuzzy numbers of type LR, L(x) and R(x) are characteristic functions that show the left and right parts of the fuzzy number, respectively ([14]). The membership function for a fuzzy number of LR type denoted by A,

$$\widetilde{A}_{(x_i)} = \begin{cases} L\left(\frac{a-x}{b}\right) & \mathrm{x} \le \mathrm{b} \\ R\left(\frac{x-a}{c}\right) & x \ge b \end{cases} \quad \ldots 12$$

$\widetilde{A}_{(x_i)}$ the membership function of the fuzzy number A, which represents the degree of membership. The fuzzy set is flat between the outputs of the function, where L (left) and R (right) represent two forms of a function, either Triangular or Trapezoidal, as described by ([15]).

2.8 Triangular fuzzy number

Triangular fuzzy numbers are usually expressed as A= (a, b, c). An example representation of triangular fuzzy numbers, which is one of the most commonly used fuzzy numbers.[16]

The Triangular function can be represented in the following formula:

$$\mu_{\widetilde{A}(x_i)} = \begin{cases} \dfrac{(\mathrm{x} - \mathrm{a})}{(\mathrm{b} - \mathrm{a})} & a \le \mathrm{x} \le \mathrm{b} \\ 1 & x = b \\ \dfrac{(\mathrm{c} - \mathrm{x})}{(\mathrm{c} - \mathrm{b})} & \mathrm{b} \le \mathrm{x} \le \mathrm{c} \end{cases}$$

Here, b is defined as the peak (center) of the triangular fuzzy number, a and c, respectively, as the lower and upper boundary values. In order for the triangular fuzzy number to be symmetrical, the left and right spreads must be of equal magnitude, that is, the values a and c must be equidistant from the center value.

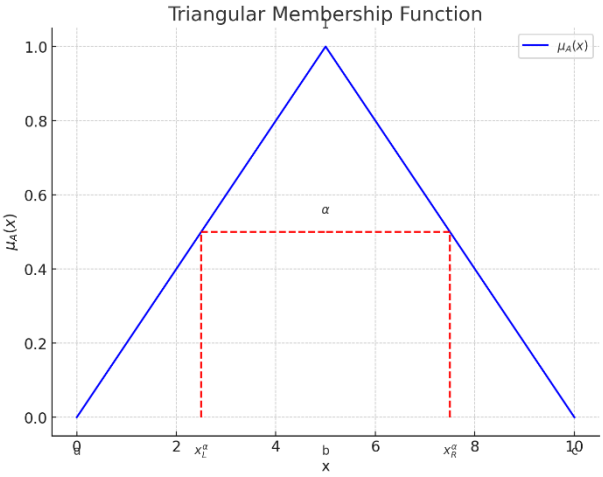

Figure 1. Triangular membership function

2.9 Fuzzy Linear Regression (FLR)

The inception of the concept of Fuzzy sets is attributed to L.A. Zadeh in 1965 to manipulate the probabilistic and uncertainty of data and information

The linear fuzzy regression model estimates the significant relationship between the response variable and the independent variables in a fuzzy with a linear function. [17].

Uncertainty in fuzzy regression, if the relationship between the independent variables and the dependent variable is fuzzy or if the data themselves are fuzzy, leads to the following types of fuzzy regression. [18] [19]

1- Crisp input and fuzzy output with fuzzy coefficients.(CIFO)

$\widetilde{\boldsymbol{Y}} = \mathbf{f}(\boldsymbol{x}, \widetilde{\boldsymbol{\beta}}) = \widetilde{\boldsymbol{\beta}}_{\mathbf{0}} + \widetilde{\boldsymbol{\beta}}_{\mathbf{1}} \boldsymbol{x}_{\boldsymbol{i1}} + \widetilde{\boldsymbol{\beta}}_{\mathbf{2}} \boldsymbol{x}_{\boldsymbol{i2}} \ldots + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{p}} \mathbf{x}_{\boldsymbol{ip}} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}} \quad , \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \cdots \boldsymbol{p}$

٢- Fuzzy input and fuzzy output with crisp coefficients. (FIFO)

$\widetilde{\boldsymbol{Y}} = \mathbf{f}(\tilde{\mathbf{x}}, \widetilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}_{\mathbf{0}} + \boldsymbol{\beta}_{\mathbf{1}} \tilde{\mathbf{x}}_{\boldsymbol{i1}} + \boldsymbol{\beta}_{\mathbf{2}} \tilde{\mathbf{x}}_{\boldsymbol{i2}} \ldots + \boldsymbol{\beta}_{\boldsymbol{p}} \tilde{\mathbf{x}}_{\boldsymbol{ip}} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}} \quad , \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \cdots \boldsymbol{p}$

3- Fuzzy input and fuzzy output with fuzzy coefficients. (FIFO)

$\widetilde{\boldsymbol{Y}} = \mathbf{f}(\tilde{\mathbf{x}}, \widetilde{\boldsymbol{\beta}}) = \widetilde{\boldsymbol{\beta}}_{\mathbf{0}} + \widetilde{\boldsymbol{\beta}}_{\mathbf{1}} \tilde{\mathbf{x}}_{\boldsymbol{i1}} + \widetilde{\boldsymbol{\beta}}_{\mathbf{2}} \tilde{\mathbf{x}}_{\boldsymbol{i2}} \ldots + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{p}} \tilde{\mathbf{x}}_{\boldsymbol{ip}} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}} \quad , \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \cdots \boldsymbol{p}$

This research will adopt Fuzzy Model 1, where the inputs are crisp, and the outputs and parameters are fuzzy (CIFO). This model was chosen due to the nature of water pollution data, where outputs are often vague language expressions describing water quality or pollution levels, while inputs such as pH, and other physical and chemical factors are precise, measurable values.

$\tilde{Y} = \mathrm{f}(x, \tilde{\beta}) = \tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} \ldots + \tilde{\beta}_p \mathrm{x}_{ip} + \tilde{\varepsilon}_i \quad , i = 1,2, \cdots p \quad \ldots 13$

$\tilde{\beta}$**:** Fuzzy parameters model.

$\tilde{Y}$: Fuzzy dependent variable, $x_{i1}, x_{i2} \ldots x_{ip}$: Crisp independent variables.

$\tilde{\varepsilon}_i$: Fuzzy random erorr.

The most widely known and used methods for estimate parameter fuzzy regression are:

1. **Possibilistic regression.**
2. **Fuzzy Least Squares Method.**

2.10 Fuzzy Fixed Effect Panel Data Model

we introduce and develop a statistical regression model We will develop Fuzzy Fixed Effect Panel Data (FFEPD) Model Based on a above section Fixed Effect Linear Regression (FELR), according to the following:

$$\tilde{Y}_{it} = \tilde{B}_{oi} \oplus \sum_{k=1}^{K} \tilde{B}_{ik} X_{kit} \oplus \tilde{U}_{it} \quad \ldots 14$$

Where $\oplus$ as the fuzzy addition operator, Then (FFELR) model can be rewritten as:

$$\hat{\tilde{Y}}_i = \sum_{j=1}^{N} D_{ij} \tilde{a}_j \oplus \sum_{k=1}^{p} \tilde{\beta}_k x_{ik} \quad \ldots 15$$

$\hat{\tilde{Y}}_i$:A vector with rank (T*1) of observations of the fuzzy dependent variable for the cross section (i).

D_{ij} **:**matrix of dummy variables with order (N*T)*N representing fixed effects.

$\tilde{a}_j$ **:**represents fuzzy intercepts specific to each cross-sectional unit. the constant term parameter of the fuzzy regression model for cross section (i).

x_{ik}: ordered matrix (T*k) of observations of explanatory variables for cross-section (i).

$\tilde{\beta}_k$**:** a vector rank (k*1) of fuzzy regression parameters for the cross section (i).

It is possible to write the above equation in using matrices as follows:

$$\hat{\tilde{Y}} = D\tilde{a} \oplus \tilde{\beta} X \quad \ldots 16$$

Given that:

$\tilde{a}_j = (a_j, c_j)$:fuzzy constant parameter.

$\tilde{\beta}_k = (b_k, \mathrm{d}_k)$: fuzzy slope coefficients.

$\tilde{y}_i = (y_i, e_i)$: fuzzy prediction value.

2.11 Tanaka Fuzzy Fixed Effect Panel Data Model

The Fuzzy Fixed Effect Linear Regression (FFELR) model introduced in this research constitutes an adaptation of the fuzzy linear regression model originally proposed by Tanaka et al. (1982,1987,1989) to Fuzzy Fixed Effect Linear Regression (FFELR) model. In this case (FFELR) objective function and the constraints desired are as follows:[5]

$$Min\, J = N.\mathrm{T} \left(\sum_{j=1}^{N} c_j D_j \right) + d_1 \sum_{i=1}^{N*T} |x_{i1}| + d_2 \sum_{i=1}^{N*T} |x_{i2}| + \cdots + d_k \sum_{i=1}^{N*T} |x_{ik}| \quad \ldots 17$$

Here, the J denotes the total uncertainty, representing the total fuzziness encompassed within the model. The constraints are as follows:

Constraint 1:$\sum_{j=1}^{N} \alpha_j \boldsymbol{D_j} + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} + (1-h)\left(\left(\sum_{j=1}^{N} c_j \boldsymbol{D_j}\right) + d_1|x_{i1}| + d_2|x_{i2}| + \cdots + d_k|x_{ik}|\right) \geq y_i + (1-h)e_i \qquad i = 1,2, \ldots N*T$

Constraint ٢:$\sum_{j=1}^{N} \alpha_j \boldsymbol{D_j} + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} + (1-h)\left(\left(\sum_{j=1}^{N} c_j \boldsymbol{D_j}\right) + d_1|x_{i1}| + d_2|x_{i2}| + \cdots + d_k|x_{ik}|\right) \leq y_i - (1-h)e_i \qquad i = 1,2, \ldots N*T$

Constraint ٣:

$$c_j > 0 \quad j = 1,2, \dots, N$$
$$d_i > 0 \quad i = 1,2, \dots, K$$

Here, if Constraint 3 equals or exceeds Spread 0, representing a distance measure, it implies that Constraint 1 and Constraint 2 must encompass all y_{ij} values within the lower and upper limits of the fuzzy predictions. As elucidated, the aforementioned minimization problem entails several constraints. Following the algorithm's determination of fuzzy parameters, the lower limit, upper limit, and midpoint of the fuzzy value are computed as follows:

$$y_{i\,lower} = \hat{y}_i - e_i = \sum_{j=1}^{N} \alpha_j \boldsymbol{D_j} + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} - \left(\left(\sum_{j=1}^{N} c_j \boldsymbol{D_j} \right) + d_1|x_{i1}| + d_2|x_{i2}| + \cdots + d_k|x_{ik}| \right) \quad \ldots 18$$

The mid-point of the fuzzy prediction:

$y_{i\,midpoint} = \hat{y}_i = \sum_{j=1}^{N} \alpha_j \boldsymbol{D_j} + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} \quad \ldots 19$

The upper limit of the fuzzy prediction:

$$y_{i\,upper} = \hat{y}_i + e_i = \sum_{j=1}^{N} \alpha_j \boldsymbol{D_j} + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} + \left(\left(\sum_{j=1}^{N} c_j \boldsymbol{D_j} \right) + d_1|x_{i1}| + d_2|x_{i2}| + \cdots + d_k|x_{ik}| \right) \quad \ldots 20$$

The h in the constraints is referred to as the fuzziness level, with a value ranging between 0 and 1. This value is designated by the user at the outset of the algorithm and signifies the degree of reliance on the dataset.

2.12 Quadratic Programming Fuzzy Fixed Effect Panel Data Model

In this method, we develop Fuzzy Fixed Effect Panel Data (FFEPD) Model using the Quadratic Programming (QP) approach. In the fundamental formulation of QP, we employ the sum of squared spreads of the estimated outputs as the objective function. Additionally, we consider minimizing the sum of squared spreads of the estimated outputs, denoted as:

$$Min\, J = N.T\left(\boldsymbol{c^t} D_j \boldsymbol{c}\right) + \boldsymbol{d^t}\left(\sum_{i=1}^{N*T} |x_i|\,|x_i|^t\right)\boldsymbol{d} + \xi\, \boldsymbol{\alpha^t}\, \boldsymbol{\alpha} \quad \ldots 21$$

Where:

$\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n)^t, \boldsymbol{c} = (c_0, c_1, c_2, \dots, c_n)^t, \boldsymbol{d} = (d_0, d_1, d_2, \dots, d_n)^t$

under the constraint (subject to)؛

Constraint 1:

$$\sum_{j=1}^{N} \alpha_j \boldsymbol{D_j} + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} + (1-h)\left(\left(\sum_{j=1}^{N} c_j \boldsymbol{D_j} \right) + d_1|x_{i1}| + d_2|x_{i2}| + \cdots + d_k|x_{ik}| \right) \geq y_i + (1-h)e_i \quad , i = 1,2, \dots N * T$$

Constraint 2:

$$\sum_{j=1}^{N} \alpha_j \boldsymbol{D_j} + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik}$$

$$- (1-h)\left(\left(\sum_{j=1}^{N} c_j \boldsymbol{D_j}\right) + d_1|x_{i1}| + d_2|x_{i2}| + \cdots + d_k|x_{ik}|\right)$$

$$\leq y_i - (1-h)e_i \qquad i = 1,2,\ldots N * T$$

Constraint 3:

$$c_j > 0 \quad j = 1,2,\ldots,N$$
$$d_i > 0 \quad i = 1,2,\ldots,K$$

2.13 Fuzzy Fixed Effect Least Absolute Estimation Method (FFELAE)

we apply the least absolute deviation estimators to construct the fuzzy least absolute linear regression model with crisp inputs, fuzzy output and fuzzy parameters, introduce a distance between triangular fuzzy numbers,[1]. we interduce developed Fuzzy Fixed Effect Least Absolute Linear Regression Model, and use the similarity measure of triangular fuzzy numbers to evaluate the fitting of the observed and estimated values. Through to minimize the least absolute distance between observed fuzzy outputs and estimated ones, we have the objective function as follows:

$$S = Min \sum_{i=1}^{n} \left| \tilde{y}_i - \sum_{j=1}^{N} D_{ij}\, \tilde{a}_j - \sum_{j=1}^{p} \tilde{\beta}_j x_{ij} \right| \quad \ldots 22$$

by minimizing the Absolute distances between observed and predicted fuzzy output data as follows:

$$S = \arg\min \sum_{i=1}^{n} d\ \left| \tilde{y}_i - \hat{\tilde{Y}}_i \right| \qquad 23$$

The resulting estimators Fuzzy Fixed Effect Least Absolute estimate denoted by:

$$,\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_{0i} \\ \tilde{\beta}_{ki} \end{pmatrix}$$

$$d\left|\tilde{y}_i, \hat{\tilde{Y}}_i\right| = \left|{y_i}^C - \hat{\tilde{Y}}_i^{\,C}\right| + \left|\left[({y_i}^C - {y_i}^L) - \left(\tilde{\hat{y}}_i^{\,C} - \hat{\tilde{Y}}_i^{\,L}\right)\right]\right| + \left|\left[({y_i}^C - {y_i}^U) - \left(\tilde{\hat{y}}_i^{\,C} - \hat{\tilde{Y}}_i^{\,U}\right)\right]\right| \quad \ldots 24$$

$$\tilde{y}_i = ({y_i}^C, {y_i}^L, {y_i}^U) \qquad \hat{\tilde{Y}}_i = \left(\hat{\tilde{Y}}_i^{\,C}, \hat{\tilde{Y}}_i^{\,L}, \hat{\tilde{Y}}_i^{\,U}\right)$$

$\hat{\tilde{Y}}_i$: The estimated value of the fuzzy response variabl

2.14 Fuzzy Linear Regression for Panel Data Using Approximate Bayesian Computation

In this study, we interduce a novel fuzzy linear regression approach tailored for panel data analysis, to advantage the Approximate Bayesian Computation (ABC) framework. The method integrates the central tendency characteristics of least squares estimation with the inclusion properties inherent in possibilistic approaches, creating a hybrid model suitable for analyzing fuzzy data. A Bayesian model is formulated to address the complexities of fuzziness in panel data regression. Given the inherent challenges in defining a precise likelihood function under conditions of fuzziness, we adopt a likelihood-free algorithm based on the rejection-ABC methodology. Finally, the convergence and performance of the proposed algorithm are evaluated using a panel dataset, providing evidence of its effectiveness and robustness in handling fuzzy regression for temporal and individual-specific variations.

In the first stage, the fuzzy outputs are transformed into crisp values by applying any deblurring method and based on these crisp outputs, the centers of the fuzzy coefficients α are estimated using classical least squares estimation.

In the second stage, the spreed of the fuzzy coefficients e are estimated by applying elements of Bayesian statistics. We consider e to be fuzzy random variables and assume that

the joint prior distribution of $e = (c_j, d_k)$ is denoted by π(e) . Based on Bayes' equation, the posterior distribution $\pi(e \mid \tilde{y})$ given the observed data $\tilde{y}$ = [$\tilde{y}$1, $\tilde{y}$2, . . . , $\tilde{y}$n] can be calculated by:

$$\pi(e \mid \tilde{y}) \propto L(e, \tilde{y})\pi(e) \quad \ldots 25$$

To our knowledge, there is no generally accepted definition of the probability function $L(e, \tilde{y})$ for a model with fuzzy coefficients, Therefore, classical Bayesian MCMC sampling algorithms, e.g., Metropolis and Gibbs, are not applicable to solving Eq. (25). The probability-free rejection sampling algorithm and its version called the approximate Bayesian rejection (ABC) algorithm provide us with a solution to overcome the problem of not having a well-defined probability function. Let e be a sample from its prior distribution π(e) and α0 be the least-squares estimate of the centers of the fuzzy coefficients, then the symmetric triangular fuzzy coefficients can be denoted by $(\alpha 0, e)^T$.

According to the concept of sampling rejection, samples from the prior distribution can be transformed into samples from the posterior distribution by retaining each sampled value with a probability proportional to the probability. Therefore, with probability L(e, ỹ) equal to probability Pr(Ỹe = ỹ), samples from the posterior distribution π(e|ỹ) in Eq. (25) can be obtained sequentially by first sampling the prior π(e) and then retaining the samples obtained if the computed output Ỹ e equals the observed data ỹ.

The algorithm Rejection Sampling used to generate independent samples from the posterior distribution is discussed, but it is noted that its use may be impractical due to the very low acceptance rate and high computational cost. This means that the observational errors in the observed data make it difficult to generate fuzzy results that exactly match the observed results using random samples from the $\pi(e)$ distribution. In other words, the probability $Pr(Ye = y)$ is approximately zero. Therefore, obtaining valid samples at an acceptance rate close to zero requires huge computational resources. To make the process easier, the acceptance condition $Ye = y$ can be relaxed to a more flexible condition of $d(Ye, y) \leq \epsilon$ where $\epsilon > 0$ is a threshold, and $d(Ye, y) \geq 0$is a distance function that measures the difference or divergence between the two data sets. Different measures of distance or similarity can be used to construct this function.

This leads to the development of a new algorithm called Algorithm 1 that relies on rejecting non-matching samples using the Approximate Bayesian Computation - ABC method. This algorithm does not produce samples from the exact posterior distribution $(\pi(e|\tilde{y})$, but rather produces samples from the approximate distribution:

$$\pi_{d,\epsilon}(e|y) \propto Pr(d(Y_e, y) \leq \epsilon)\pi(e) \quad \ldots 26$$

Where this distribution can be thought of as an approximate posterior distribution conditional on the event $d(Y_e, y) \leq \epsilon$.

Second method to estimate Fuzzy coefficients centers and spreads We use the Bayesian method to estimate the parameter We assume that α and e are two random variables, each with a prior probability distribution, $\pi 1(\alpha)$,π2 (e), We will have the sales distribution of the observations We find Bayes' equation, the posterior distribution $\pi(\alpha, e \mid \tilde{y})$ given the observed data $\tilde{y}$ = [$\tilde{y}$1, $\tilde{y}$2, . . . , $\tilde{y}_{NT}$] can be calculated by:

$$\pi(\alpha, e \mid \tilde{y}) \propto L(\alpha, e, \tilde{y})\pi_1(\alpha)\pi_2(e) \quad \ldots 27$$

Algorithm 1: Rejection approximate Bayesian computation

1. for $q = 1$ to NT
2. repeat
3. Generate **e** from the prior distribution $\pi(\mathbf{e})$.
4. Compute: $\tilde{\boldsymbol{Y}}_{\boldsymbol{e}} = [\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_n]^T$

Where:

$$\tilde{\boldsymbol{Y}}_{\boldsymbol{e}} = \left(\alpha_j D_j, c_j D_j\right) + (b_k xi, d_k |xi|)$$

$$j = 1,2,\dots,N \ , i = 1,2,\dots,nt$$

5. until $d(\tilde{Y}_e, \tilde{y}) = 1 - \bar{S}(\tilde{Y}_e, \tilde{y})$ or

I = MaxIt, terminate the loop.

6. . $\boldsymbol{e}^{(q)} \leftarrow \boldsymbol{e}$
7. end

2.15Algorithm ABC: Fuzzy Regression Parameter Estimation Using approximate Bayesian computation

To estimate the parameters of a fuzzy regression using approximate Bayesian arithmetic, with uncertainty in both the data and the fuzzy parameters, we present two estimation algorithms as follows:

Algorithm 2: Fuzzy regression panel data model based on approximate Bayesian computation (ABC1)

1. for $q = 1$ to NT
2. repeat
3. Estimate $a_p = (\alpha_j, b_k)$ by **OLS**.
4. Estimate $e_p = (c_j, d_k)$ by **TL** Tanaka method or **QP** Quadratic method.
5. Generate **e** from the prior distribution $\pi(\boldsymbol{e})$: $exp(length(e_p), rate = \frac{1}{k_1 \cdot e_p})$
6. Compute: $\tilde{\boldsymbol{Y}}_{\boldsymbol{e}} = [\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_n]^T$

Where:

$$\tilde{\boldsymbol{Y}}_{\boldsymbol{e}} = \left(\alpha_j D_j, c_j D_j\right) + (b_k xi, d_k |xi|)$$

$$j = 1,2,\dots,N \ , i = 1,2,\dots,nt$$

7. until $d(\tilde{Y}_e, \tilde{y}) = 1 - \bar{S}(\tilde{Y}_e, \tilde{y})$

or I = MaxIt, terminate the loop.

8. . $\boldsymbol{e}^{(q)} \leftarrow \boldsymbol{e}$
9. end

Algorithm 3: Fuzzy regression panel data model based on approximate Bayesian computation (ABC2)

1. for $q = 1$ to NT
2. repeat
3. Estimate $a_p = (\alpha_j, b_k)$ by **OLS**.
4. Estimate $e_p = (c_j, d_k)$ by **TL** Tanaka method or **QP** Quadratic method.
5. Generate $\mathbf{e}$ from the prior distribution

$$\pi(\mathbf{e}) : exp(length(e_p), rate = \frac{1}{k_1 \cdot e_p})$$

6. Generate a from the prior distribution $\pi(a) : N(a_p, \Sigma)$ Where:

$$\Sigma = \text{diag}(k_1 a_{p_0}, k_2 a_{p_1}, \dots, k_m a_{p_m})$$

7. Compute: $\tilde{\boldsymbol{Y}}_{\boldsymbol{e}} = [\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_n]^T$

Where:

$$\tilde{\boldsymbol{Y}}_{\boldsymbol{e}} = (\alpha_j D_j, c_j D_j) + (b_k xi, d_k |xi|), j = 1,2,\dots,N \ , i = 1,2,\dots,nt$$

8. until $d(\tilde{Y}_e, \tilde{y}) = 1 - \bar{S}(\tilde{Y}_e, \tilde{y})$ or I = MaxIt, terminate the loop.
9. . $\boldsymbol{a}^{(q)} \leftarrow \boldsymbol{a}$,$\boldsymbol{e}^{(q)} \leftarrow \boldsymbol{e}$

2.16 Description of Study Area and Sampling Collection

The case study is situated in the water quality of the Euphrates River, all sites along the Euphrates River from north to south were selected, bringing their number to 11 sites. Samples were collected monthly for 12 months during 2023, and these samples were obtained from the Iraqi Ministry of Environment.

The research methodology is grounded in the standard specification for the River and Public Water Pollution Control System No. 417 of 2009. This framework guides the examination and interpretation of various water quality parameters, including pH, NO3, PO4, TDS, and SO4. The study provides insights into the current state of water quality in Iraq and highlights the challenges and potential solutions for maintaining and improving these vital resources.

Table 1: Iraqi Standard Specifications No. 417 for the Year 2009 - Second Update

Parameters	pH	TDS	NO3	PO4	DO
Acceptable Limits	6.5 – 8.5	< 1000 mg/l	< 50 mg/l	Not specified	Not specified

To provide a comprehensive and summarized evaluation of the tests, a standardized score ranging from 0 to 100 is utilized through the Water Quality Index (WQI). The resulting value of this index reflects the level of water quality; a lower score indicates poor water quality, while a higher score suggests good water quality.

Table 2: Classification of Water Quality Based on Water Quality Index (WQI) Values

دليل جودة المياه(Water Quality Index)				
No.	River Water Condition حالة مياه النهر	What dose each score mean توضيح درجة التلوث	Color اللون	Range المدى
1	(Excellent) ممتاز	Mostly healthy and thriving صحي بشكل ممتاز		95-100
2	(Good) جيد	Impact by pollution but still resilient متأثر بالتلوث ويبقى جيداً		80-94
3	(Fair) مقبول	Significantly impacted by pollution متأثر بالتلوث بشكل ملحوظ ولكن يبقى مقبولاً		65-79
4	(Bad) رديء	Severely impacted by pollution متأثر بالتلوث بشكل كبير ويمكن معالجته		45-64
5	(Very Bad) مرفوض	threat to human health and native species يشكل تهديد لصحة الانسان وبقية الاحياء		0-44

Table 3: RMSE criterion results of real data **OLS + TL**

Method	RMSE
Fuzzy Tanaka	22.0451
Fuzzy Quadratic	21.5526
PFLAS	6.2105
ABC1TL	6.0016
ABC2TL	7.6193

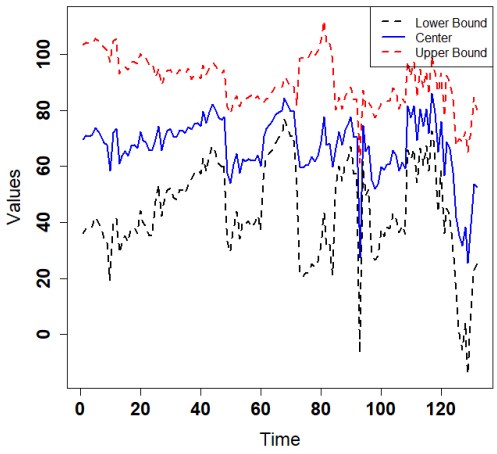

Figure 2 : Tanaka Fuzzy Fixed Effect Panel Data Model

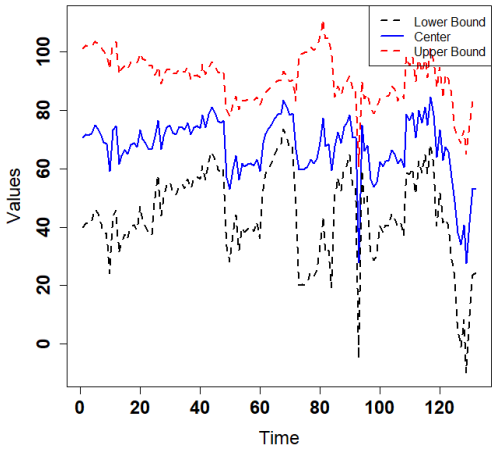

Figure 3 : Quadratic Fuzzy Fixed Effect Panel Data Model

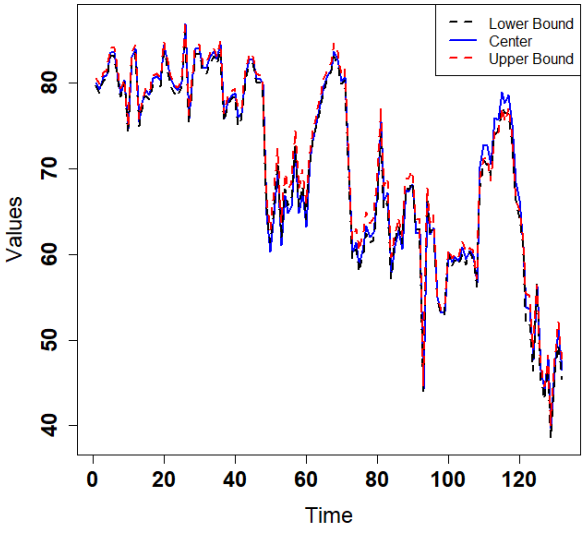

Figure 4 : Fuzzy Fixed Effect Least Absolute Estimation Method

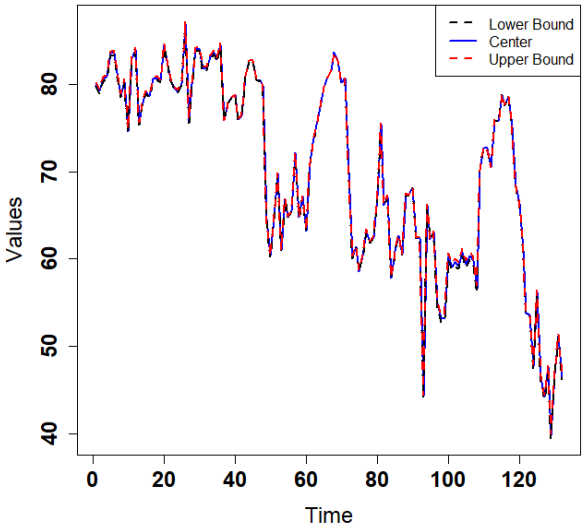

Figure 5 : ABC1TL

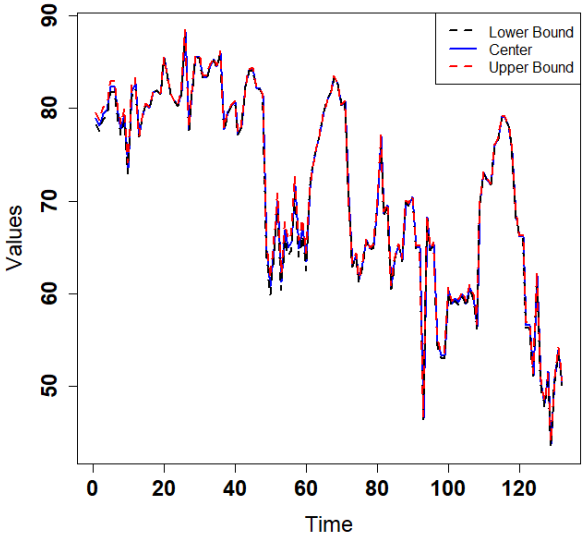

Figure 6: ABC2TL

Table 4: RMSE criterion results of real data

OLS + QP

Method	RMSE
Fuzzy Tanaka	22.0451
Fuzzy Quadratic	21.5526
PFLAS	6.2105
ABC1QP	5.9986
ABC2QP	6.0603

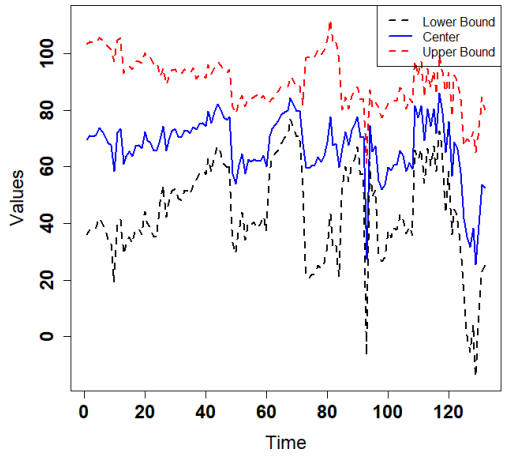

Figure 7 : Tanaka Fuzzy Fixed Effect Panel Data Model

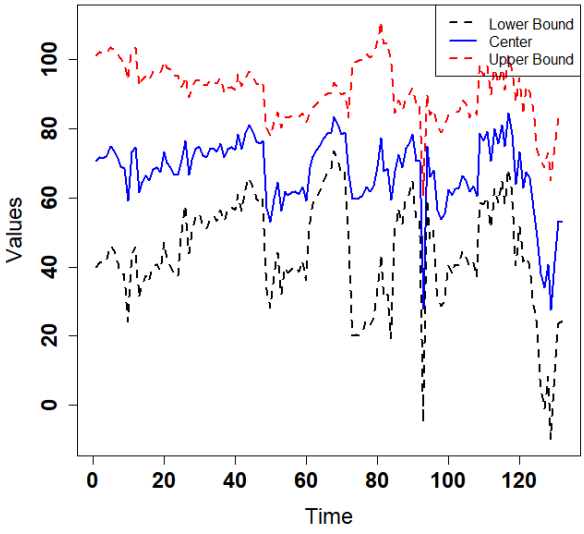

Figure 8 : Quadratic Fuzzy Fixed Effect Panel Data Model

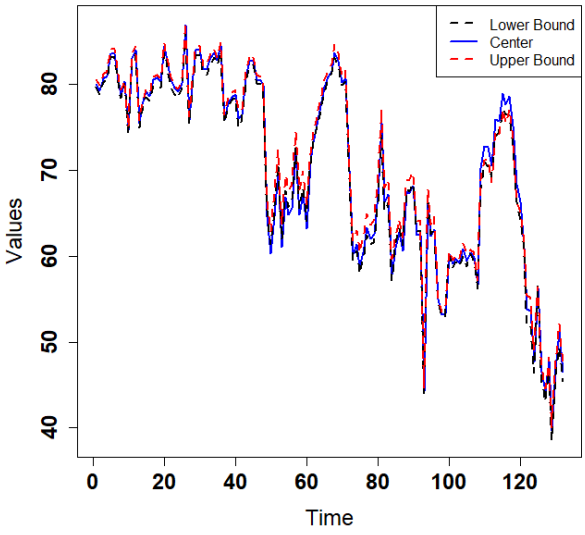

Figure 9 : Fuzzy Fixed Effect Least Absolute Estimation Method

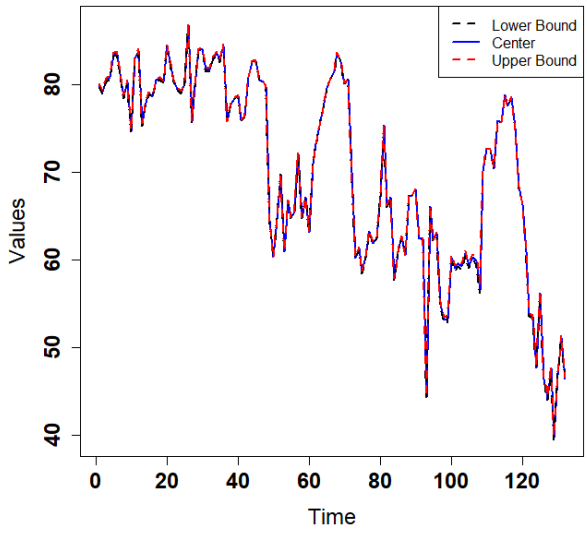

Figure 10 : ABC1QP

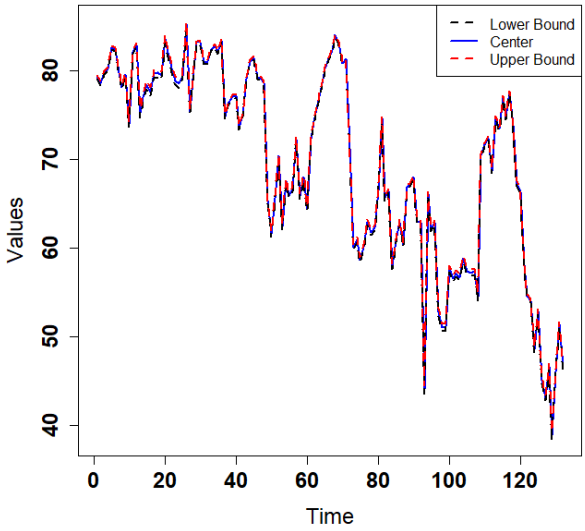

Figure 11 : ABC1QP

3. Conclusions

1. Classical methods (Fuzzy Tanaka and Fuzzy Quadratic) performed poorly with very high RMSE values, indicating that they are less accurate in prediction.
2. The improved methods (PFLAS and ABC) fared significantly better in accuracy, recording very low RMSE values.
3. Among the improved methods, ABC1 (either TL or QP) was the best performer, with a slight additional improvement when using Quadratic Programming (QP).
4. Using Quadratic Programming showed a slight improvement over Tanaka Linear in some ways, enhancing the effectiveness of this model with advanced methods.
5. Traditional methods such as Fuzzy Tanaka and Fuzzy Quadratic give less accurate

results (high RMSE), reflecting the difficulty of these methods in dealing with real data effectively.

6. More advanced methods such as PFLAS and ABC (either TL or QP) achieve much better performance (lower RMSE).
7. Between ABC1 and ABC2, we observe that ABC1 with either TL or QP gives more accurate results.
8. When using from TL to QP slightly improves performance, especially for ABC1, suggesting that using Quadratic Programming (QP) gives an additional advantage.

References

[1] J. Li, W. Zeng, J. Xie, and Q. Yin, "A new fuzzy regression model based on least absolute deviation," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 54–64, 2016.

[2] N. Wang, M. Reformat, W. Yao, Y. Zhao, and X. Chen, "Fuzzy Linear regression based on approximate Bayesian computation," *Appl. Soft Comput.*, vol. 97, p. 106763, 2020.

[3] A. M. Hassan and A. T. Rahem, "Comparison Of The Estimators Of The General Least Squares Method Of The Parametric Model With The Estimators Of The Nadaria-Watson Weighted And Unweighted Method Of The Non-Parametric Model Of Longitudinal Data.," *Iraqi J. Econ. Sci.*, vol. 19, no. 70, 2021.

[4] R. T. K. Al-Adly and E. H. Aboodi, "Using some methods of estimating longitudinal data models with a practical application," *Master's Thesis-University Baghdad, Coll. Adm. Econ.*, 2021.

[5] M. O. Yalçın, N. Güler Dinçer, and S. Demir, "Fuzzy panel data analysis," *Kuwait J. Sci.*, vol. 48, no. (3), p. pp(1-13), 2021, doi: doi.org/10.48129/kjs.v48i3.8810.

[6] R. E. Kareem and M. J. Mohammed, "Fuzzy Bridge Regression Model Estimating via Simulation," *J. Econ. Adm. Sci.*, vol. 29, no. 136, pp. 60–69, 2023.

[7] C. Hsiao, *Analysis of Panel Data*, Second Edi. 2003.

[8] Z. Y. Algamal, "Selecting Model in Fixed and Random Panel Data Models," *IRAOI J. Stat. Sci.*, vol. 12, no. 1, 2012.

[9] B. H. Baltagi, "Econometric Analysis of Panel Data, John Wiley&Sons Ltd," *West Sussex, Engl.*, 2005.

[10] P. Das, "Econometrics in theory and practice," *Springer*, vol. 10, pp. 978–981, 2019.

[11] C. Hsiao, *Analysis of Panel Data*, Third Edit. New York, NY 10013-2473, USA: Cambridge University Press., 2014.

[12] N. A. Abd and N. H. Fadhil, "The comparison between longitudinal data regression models in estimating and analyzing investment functions for productive economic sectors in Iraq for the period (1995-1996)," *J. Adm. Econ.*, no. 122, 2019.

[13] H.-C. Wu, "Fuzzy estimates of regression parameters in linear regression models for imprecise input and output data," *Comput. Stat. Data Anal.*, vol. 42, no. 1–2, pp. 203–217, 2003.

[14] D. Dubois and H. Prade, "Systems of linear fuzzy constraints," *Fuzzy sets Syst.*, vol. 3, no. 1, pp. 37–48, 1980.

[15] A. H. Ali, M. Aljanabi, and M. A. Ahmed, "Fuzzy generalized Hebbian algorithm for large-scale intrusion detection system," *Int. J. Integr. Eng.*, vol. 12, no. 1, pp. 81–90, 2020.

[16] M. C. J. Anand and J. Bharatraj, "Theory of triangular fuzzy number," *Proc. NCATM*, vol. 80, 2017.

[17] S. Yeylaghi, M. Otadi, and N. Imankhan, "A new fuzzy regression model based on interval-valued fuzzy neural network and its applications to management," *Beni-Suef Univ. J. basic Appl. Sci.*, vol. 6, no. 2, pp. 106–111, 2017.

[18] A. R. Arabpour and M. Tata, "Estimating the parameters of a fuzzy linear regression model," *Iran. J. Fuzzy Syst.*, vol. 5, no. 2, pp. 1–19, 2008.

[19] M. Haggag, "A new fuzzy regression model by mixing fuzzy and crisp inputs," *Am Rev Math Stat*, vol. 6, pp. 9–25, 2018.