Fuzzy panel data analysis

Muhammet O. Yalçin*, Nevin Güler Dincer, Serdar Demir
Dept. of Statistics, Faculty of Science, University of Mugla Sıtkı Koçman Muğla, Turkey
Corresponding author: moguzhanyalcin@mu.edu.tr

Abstract

In statistical and econometric researches, three types of data are mostly used as cross-section, time series and panel data. Cross-section data are obtained by collecting the observations related to the same variables of many units at constant time. Time series data are data type consisted of observations measured at successive time points for single unit. Sometimes, the number of observations in cross-sectional or time series data is insufficient for carrying out the statistical or econometric analysis. In that cases, panel data obtained by combining cross-section and time series data are often used. Panel data analysis (PDA) has some advantages such as increasing the number of observations and freedom degree, decreasing of multicollinearity, and obtaining more efficient and consistent predictions results with more data information. However, PDA requires to satisfy some statistical assumptions such as "heteroscedasticity", "autocorrelation", "correlation between units", and "stationarity". It is too difficult to hold these assumptions in real-time applications. In this study, fuzzy panel data analysis (FPDA) is proposed in order to overcome these drawbacks of PDA. FPDA is based on predicting the parameters of panel data regression as triangular fuzzy number. In order to validate the performance of efficiency of FPDA, FPDA, and PDA are applied to panel data consisted of gross domestic production data from five country groups between the years of 2005-2013 and the prediction performances of them are compared by using three criteria such mean absolute percentage error, root mean square error, and variance accounted for. All analyses are performed in R 3.5.2. As a result of analysis, it is observed that FPDA is an efficient and practical method, especially in case required statistical assumptions are not satisfied.

Keywords: Fixed effects; fuzzy logic; fuzzy panel data analysis; panel data; random effects.

1. Introduction

In statistical and econometric studies, the reliability of hypothesis tests and predictions depends on the selection of data being suitable for the statistical method and on being collected from reliable sources. In general, it is possible to group the data types used in statistical and econometric researches into three main titles as *time series, cross section, and panel data*. The time series consist of observations measured with the successive time intervals, and aims to predict the future values of time series by examining its past behavior. In time series, the number of units is equal to one. For analyzing time series, many statistical and soft computing methods such as Box-Jenkins (1976), Artificial Neural have been used so far (Shadab *et al.*, 2019; Wahid & Kim, 2017). The cross-section consists of "n" number of observations relating to a variable at a constant time point and regression analysis is generally used for analyzing cross section data (Bashir *et al.*, 2017; BuHamra *et al.*, 2018). In some cases, the length of time series or the number of units in cross-section data is not adequate for statistical and econometric analyses. PDA can be utilized for this kind of cases. The panel data are formed by combining the time series and cross section data (Yerdelen Tatoğlu, 2012). In other words, panel data

the dataset obtained by observing various characteristics of the same units through time (Ahn & Moon, 2001). On this basis, it is possible to claim that analyzing the panel data means also analyzing the time and cross-section data simultaneously. Besides, PDA increases the number of the observations, thus it provides more consistent and reliable predictions, enables to analyze the cross-section data for which the number of observations is not adequate, analyzing short time series, and reducing the multicollinearity problem.

In general, there are three objectives in studies conducted with panel data; i) revealing the change of the data in the cross-section through time, ii) explaining the change of the units one by one or together based on the other variables and iii) predicting each unit depending on the relevant independent variable (Hsiao, 2003). PDA is performed in 6 steps for these objectives. The first step is deciding on the form of the regression model that will be used for prediction. Here, there are two regression models one of which is called "Fixed Effects (FE)" and the other one "Random Effects (RE)". In the FE model, the differences between the units are expressed with a fixed term. In the RE model, the changes that occur according to the units of time are reflected with the error term. The Hausman test, which was developed in 1978 and which bears the name of its developer, is used to decide which of these regression models will be used (Hausman, 1978). The second step in PDA is to investigate whether there are dependencies between the cross sections by using the Breusch Pagan test (Breusch & Pagan, 1979). According to the result of this test, the generation of unit root tests which will be used to test the stationarity is decided. As a result of the Breusch Pagan test, if it is decided that there is cross section dependency between the units, the second generation panel data unit root tests are used; and if it is decided that there is no dependency between the units, the first generation panel unit root tests are used (Table 1). The fourth and fifth step of

the PDA consists of heteroscedasticity and autocorrelation testing, which alternate respectively. If the panel data under consideration satisfies all statistical assumptions mentioned, the prediction step is performed in the last step of the PDA.

As it may be understood here, the consistency and reliability of the prediction results obtained from the PDA depends on the satisfying of the assumptions mentioned before. Otherwise, the predictions may have deviations. But, since real time applications are considered, satisfying these assumptions is extremely difficult.

In the present study, a new method, the FPDA is proposed to protect the advantages of the PDA and to eliminate its limitations. In actual fact, FPDA is the adaptation of the fuzzy regression to panel data. In order to perform a fuzzy regression analysis, fuzzy regression has been suggested as an alternative method in case the observations do not ensure that the assumptions that are needed for the analysis are not satisfied, in other words, statistical regression analysis methods cannot be performed.

In fuzzy regression analysis, the relationship between the variables is not certain on the contrary to statistical regression analysis. For this reason, it gives more accurate results in case datasets include observations that have uncertainties and if the mathematical relationship between the variables is not defined clearly. The basic difference between the fuzzy regression analysis and statistical regression analysis is as follows; it is assumed in statistical regression analysis that the error terms from the incorrect measurement of observations, from the missing definition of the relationship between the variables, and from the independent variables that are not included in the model. On the other hand, in fuzzy regression analysis, it is assumed that the error terms from the uncertainty of the model coefficients, in other words, from the fuzziness. For this reason, while in statistical regression analysis, the difference between the real value and the predicted value is minimized, in fuzzy regression analysis, the

total distribution of the model coefficients is minimized (İşbilen Yücel, 2005). A lot of fuzzy regression models are proposed in the literature. In the present study, the fuzzy linear regression model, which was proposed by Tanaka *et al.*, (1982), was adapted to panel data. For the purpose of evaluating the efficiency of the proposed method, and comparing the prediction results with the PDA, the econometric relationship between 12 different socio-economic variables and Gross National Product Per Capita (GNP) was predicted. For this purpose, 5 panel data were used each of which consisted of 7 countries and 9 years between 2005-2013. Sixty panel data were analyzed separately for 12 variables and 5 Groups. As a result of the analyses, it was determined that the performance of the proposed method was better when compared with the PDA.

The organization of the study is as follows. The PDA and its steps are mentioned in the second part. In the third part, the proposed method is explained in detail. The fourth part includes the empirical results obtained in the study. In the last part, the study is concluded.

2. Panel data analysis

As mentioned before, PDA consists of 6 steps. In the first step, the model that will be used for prediction is decided. For this purpose, two regression models are used which are "FE" and "RE". In FE Model, the effect of the units is added to the model as fixed term (Baltagi, 1995). The FE model is defined as follows:

$$y_i = \sum_{j=1}^{N} D_{ij}\alpha_j + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i \quad (1)$$
$$i=1,2,\ldots,N*T$$

Here, N denotes the number of the units, T shows the number of the time points, y_i refers to the i. observation of the panel data, α_j shows the fixed parameters of j. unit, D_{ij} denotes the dummy variable that consists of 1s and 0s with $(N*T)$ xN-dimension, p shows the number of the independent variables, β_k shows that coefficient of the k. independent variable, x_{ik}, shows the i. observation of the

k. independent variable, ε_i shows the error term distributed normally with 0 mean and constant variance.

The D dummy variable matrix form in Equation (1) is as follows:

$$D = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & 0 & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \cdots & 1 \end{bmatrix}_{(N*T)xN} \quad (2)$$

To predict the parameters of the FE model, generally the Least Squares Method with Dummy Variable is used. This method is as follows:

$$J(X, \alpha, \beta) = \sum_{i=1}^{N*T} (y_i - (\sum_{j=1}^{N} D_{ij}\alpha_j + \sum_{k=1}^{p} \beta_k x_{ik}))^2 \quad (3)$$

Here, the α and β coefficients are predicted to make the objective function given in Equation 3 minimum.

In RE model, changes in the units or time are added to the model as a component of the error term. The RE model is defined as follows:

$$y_i = \bar{\bar{\alpha}} + \sum_{k=1}^{p} \beta_k x_{kit} + (\varepsilon_i + \mu_j) \quad (4)$$
$$i=1,2,\ldots,N*T$$

As seen in Equation 4, the RE model includes two types of error terms. While one of these two variables ε_i shows all the errors, μ_j shows the errors between the units according to the fixed time. For the purpose of predicting the parameters of the RE regression model, generally the Least Squares and Likelihood Methods are used.

For the purpose of prediction, Hausman test (Hausman, 1978) test is used to determine to use FE model or RE model. FE model is generally based on the hypothesis that there is a correlation between the units and the independent variables in the model. In here, null hypothesis is defined as *"there*

<div align="center">

Table 1. Panel Unit root tests

</div>

First Generation Unit Root Test	Second Generation Unit Root Test
1. Levin and Lin (1992)	1. Moon and Peron (2003)
2. Harris and Tzavalis (1999)	2. Choi (2002)
3. Levin, Lin and Chu (2002)	3. Phillips and Sul (2003)
4. Hadri (2000)	4. Pesaran (2007)
5. Breitung (2000)	5. Bai and Ng (2004)
6. Im, Pesaran and Shin (1997)	6. Breitung and Das (2003)
7. Mandala and Wu (1999)	7. Chang (2004)
8. Choi (2001)	
9. Im, Pesaran and Shin (2003)	

is no relation between the units and the independent variables" and the failure of rejecting this hypothesis shows that the RE model may be used, and the success in rejection shows that the FE model may be used.

The second step of the PDA is investigation of whether there is dependency between the cross sections. For this purpose, the Breusch Pagan Lagrange Test (Breusch & Pagan, 1979) and Pesaran (2007) CD test are used. Breusch Pagan Lagrange (1979) test may be used when the time dimension of the panel data is bigger than the cross section dimension, and the Pesaran CD test (Pesaran, 2007) may be used in both situations.

The third step in the PDA is the testing of the stationarity of the panel data. There are many panel data unit root tests for this purpose. These tests are generally separated into two classes as first generation and second generation unit root tests and the generation of the unit root test is determined according to the result of the cross section dependency test. As a result of the Breusch Pagan Lagrange(1979) or Pesaran CD test (2007), if it is concluded that there is cross section dependency between the units, the second generation unit root test is used, and if it is concluded that there no such dependency, the first generation unit root test is used. The panel unit root tests used to test the stationary are given in Table 1.

Heteroscedasticity is another hypothesis that needs to be tested. In this situation, the tests vary according to the model to be used for prediction. If FE model is preferred, in this case, the Wald Test (Engle, 1984) is used (Greene, 1993); and if RE model is preferred, the Brown and Forsythe test (Brown *et al.*, 1974) are used. Similarly, the tests that are used to investigate the autocorrelation varies according to the model. If the model is FE, the Baltagi test (Baltagi, 2005) is used to test the autocorrelation; and if it is RE model, the Lagrange (Baltagi *et al*, 2012) test is used.

The last step of the` PDA is performing the prediction according to the selected panel data regression model. As mentioned before, the reliability of the predictions in this step depends on satisfying of all the hypotheses tested in the previous steps. In the present study, which was conducted to eliminate this disadvantage of the PDA, the FPDA method, which required no constraints and hypotheses on the panel data, was proposed.

3. Fuzzy panel data analysis

FPDA proposed in this study is the adaptation of the fuzzy linear regression model proposed by Tanaka *et al.* (1982) to panel data, and is defined as follows.

$$\tilde{y}_i = \sum_{j=1}^{N} D_{ij}\,\tilde{\alpha}_j + \sum_{k=1}^{p} \tilde{\beta}_k x_{ik} \tag{5}$$

Here, $y_{\cdot i}$ shows the fuzzy value of the i. unit, x_{ik} refers to the i. observation of the k. independent variable, α_j refers to the fuzzy value of the fixed effects that shows the differences between the units, $\beta_{\cdot k}$ shows the fuzzy prediction value of the slope coefficients. It is possible to write the above equation in the form of matrix notation as follows:

$$\tilde{Y} = D\tilde{\alpha} + \tilde{\beta}X \qquad (6)$$

In a more explicit way, fuzzy panel data regression model that is proposed is as follows:

$$
\begin{bmatrix}
\tilde{y}_1 \\
\tilde{y}_2 \\
\vdots \\
\tilde{y}_T \\
\tilde{y}_{T+1} \\
\tilde{y}_{T+2} \\
\vdots \\
\tilde{y}_{2\bullet T} \\
\vdots \\
\tilde{y}_{(N-1)\bullet T+1} \\
\tilde{y}_{(N-1)\bullet T+2} \\
\vdots \\
\tilde{y}_{N\bullet T}
\end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
1 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 1 & 0 & \cdots & 0 \\
\vdots & & & \cdots & \\
0 & 0 & 0 & \cdots & 1 \\
0 & 0 & 0 & \cdots & 1 \\
\vdots & & & \vdots & \\
0 & 0 & 0 & \vdots & 1
\end{bmatrix}
\begin{bmatrix}
\tilde{\alpha}_1 \\
\tilde{\alpha}_2 \\
\vdots \\
\tilde{\alpha}_N
\end{bmatrix}
+
$$

$$
\begin{bmatrix}
\tilde{\beta}_1 \\
\tilde{\beta}_2 \\
\vdots \\
\tilde{\beta}_k
\end{bmatrix}
\cdot
\begin{bmatrix}
x_{11} & x_{12} & & x_{1p} \\
x_{21} & x_{22} & \cdots & x_{2p} \\
\vdots & \vdots & \cdots & \vdots \\
x_{T1} & x_{T2} & \vdots & x_{Tp} \\
\vdots & \vdots & \vdots & \vdots \\
x_{((N-1)\bullet T+1)1} & x_{((N-1)\bullet T+1)2} & \cdots & x_{((N-1)\bullet T+1)p} \\
x_{((N-1)\bullet T+2)1} & x_{((N-1)\bullet T+2)2} & \vdots & x_{((N-1)\bullet T+2)p} \\
\vdots & \vdots & & \vdots \\
x_{(N\bullet T)1} & x_{(N\bullet T)2} & \cdots & x_{(N\bullet T)p}
\end{bmatrix}
$$

$$(7)$$

Here, $\tilde{\alpha}_j = (\alpha_j, c_j)$ refers to the fuzzy constant parameters, $\tilde{\beta}_k = (b_k, d_k)$ refers to fuzzy slope coefficients, $\tilde{y}_i = (y_i, e_i)$ refers to the fuzzy prediction value. In the light of these data, the proposed fuzzy panel regression model may be reorganized as follows for i. observation:

$$(y_i, e_i) = (\alpha_j, c_j) + (b_1, d_1)x_{i1} + \qquad (8)$$
$$\cdots + (b_k, d_k)x_{ik} \quad i=1,2,\ldots,N*T$$

Here, all of $\tilde{\alpha}_j, \tilde{\beta}_k$ and $\hat{\tilde{y}}_i$ s have triangle membership functions as shown below:

As seen in Figure 1, α_j, b_k and y_i refers to the central points of the triangular fuzzy numbers.

c_j, d_k and e_i shows the semi-spread of the constant, slope, and prediction values, respectively. Here, it is possible to see that y_i's central value is equal to $\alpha_j + b_1 x_{i1} + \ldots + b_k x_{ik}$; and the total spread, in other words, the total error is equal to $c_j + d_1 x_{i1} + \ldots + d_k x_{ik}$. In this case, the membership value for the parameter or prediction values is computed as follows:

$$
\mu_{(\alpha_s)} =
\begin{cases}
1 - \dfrac{\alpha_j - \alpha_s}{c_j} & \alpha_j - c_j \le \alpha_s \le \alpha_j \\
1 - \dfrac{\alpha_s - \alpha_j}{c_j} & \alpha_j \le \alpha_s \le \alpha_j + c_j \quad \text{for any } s \quad (9) \\
0 & \text{In other cases}
\end{cases}
$$

$$
\mu_{(b_r)} =
\begin{cases}
1 - \dfrac{b_k - b_r}{d_k} & b_k - d_k \le b_r \le b_k \\
1 - \dfrac{b_r - b_k}{d_k} & b_k \le b_r \le b_k + d_k \quad (10) \\
0 & \text{In other cases}
\end{cases}
$$

$$
\mu_{(y_h)} =
\begin{cases}
1 - \dfrac{y_i - y_h}{e_i} & y_i - e_i \le y_h \le y_i \\
1 - \dfrac{y_h - y_i}{e_i} & y_i \le y_h \le y_i + e_i \quad (11) \\
0 & \text{In other cases}
\end{cases}
$$

As seen in Equations (9), (10), and (11), there is no additional error term in the fuzzy panel data regression model. In this model, the error is distributed to the model coefficients based on the spread, for this reason, the total error of the model is obtained by summing all the spreads. In the fuzzy panel data regression model, the purpose is to predict the fuzzy parameters that will make the total spread minimum. In this case, FPDA objective function and the constraints desired are as follows:
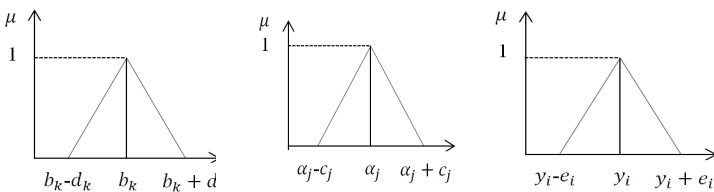
$$Min\ J = N\left(\sum_{j=1}^{N} c_j\,D_J\right) + d_1\sum_{i=1}^{N*T}|x_{i1}|$$
$$+ d_2\sum_{i=1}^{N*T}|x_{i2}| + \cdots$$
$$+ d_k\sum_{i=1}^{N*T}|x_{ik}| \qquad (12)$$

Here, J refers to the total uncertainty, in other words, the total fuzziness. The limitations, on the other hand, are as follows:

Constraint 1:
$c_1, c_2, \ldots, c_N \geq 0,\ d_1, d_2, \ldots, d_k \geq 0$

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

Constraint 3: $\sum_{i=1}^{N} \alpha_j\,D_J + b_1 x_{i1} +$
$b_2 x_{i2} + \cdots + b_k x_{ik} - (1-h)\left(\left(\sum_{j=1}^{N} c_j\,D_J\right) + d_1|x_{i1}| + d_2|x_{i2}| + \cdots + d_k|x_{ik}|\right) \leq y_i - (1-h)e_i \qquad i=1,\ldots,N*T$

Here, if the constraint 1 is equal to or bigger than spreads 0, which corresponds to a distance measure, this means that the constraint 2 and constraint 3 must cover all the y_{ij} of the lower and upper limits of the fuzzy predictions. As it may be understood here, the above minimization problem has   number of constraints. According to the fuzzy parameters obtained as a result of the algorithm, the lower, upper limit and midpoint of the fuzzy value computed as follows. The lower limit of the fuzzy prediction:

$$y_{i_{lower}} = \hat{y}_i - e_i = \sum_{j=1}^{N} \alpha_j\,D_J + b_1 x_{i1} +$$
$$b_2 x_{i2} + \cdots + b_k x_{ik} - \left(\left(\sum_{j=1}^{N} c_j\,D_J\right) + d_1|x_{i1}| + d_2|x_{i2}| + \cdots + d_k|x_{ik}|\right)$$
$$(13)$$

The mid-point of the fuzzy prediction:

$$y_{midpoint} = \hat{y}_i = \sum_{j=1}^{N} \alpha_j\,D_J + \qquad (14)$$
$$b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik}$$

The upper limit of the fuzzy prediction:

$$y_{i_{upper}} = \hat{y}_i + e_i = \sum_{j=1}^{N} \alpha_j\,D_J + \qquad (15)$$
$$b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} \qquad +$$
$$\left(\left(\sum_{j=1}^{N} c_j\,D_J\right) + d_1|x_{i1}| +$$
$$d_2|x_{i2}| + \cdots + d_k|x_{ik}|\right)$$

The h in the constraints is called as the fuzziness level, and has a value between [0, 1] (Ergün, 2011). This value is defined by the user in the starting point of the algorithm, and reflects how much the user relies on the dataset (İşbilen, 2005).

4. Empirical results

In this part, the econometric relationship between the GNP per capita and 12 different socio-economic variables was predicted for the purpose of evaluating the efficiency of the proposed model and comparing the prediction results with the PDA. For this purpose, 5 panel data consisting of 9 years between 2005-2013 for 7 countries were used. Sixty panel data were analyzed separately for 12 variables and for 5 groups in total. The variables used in the study and the countries in teach group are given in Table 2. All data sets are downloaded from the web site of https://data.worldbank.org/. All analyzes are implemented in R program and R code is written for FPDA (Appendix 1).

In the rest of the paper, the result of PDA, FPDA and the comparison results are given.

4.1. Result of PDA

In this study, first of all, Hausman test was performed. The hypothesis relating to Hausman test is as follows:

H_0:There is no correlation between independent variables and units

H_1: *There is correlation betwe independent variables and units*

The p-values of Hausman test and the panel data regression model selected for each group and each variable are given in Table 3.

Table 3. Hausman Test Results

	1.Group		2.Group		3.Group		4.Group		5.Group	
	Model	p	Model	p	Model	p	Model	p	Model	p
X_1	RE	0.12	RE	0.89	RE	0.97	RE	0.45	RE	0.69
X_2	FE	0.00	RE	0.10	FE	0.00	FE	0.02	FE	0.00
X_3	RE	0.95	FE	0.00	RE	0.74	FE	0.00	FE	0.00
X_4	FE	0.00	RE	0.22	RE	0.41	RE	0.29	FE	0.00
X_5	FE	0.00	RE	0.07	RE	0.72	FE	0.00	FE	0.00
X_6	FE	0.00	RE	0.47	RE	0.99	FE	0.00	FE	0.00
X_7	RE	0.43	FE	0.02	FE	0.01	FE	0.01	FE	0.00
X_8	FE	0.01	FE	0.01	FE	0.00	RE	0.50	RE	0.92
X_9	RE	0.06	RE	0.20	RE	0.08	RE	0.57	RE	0.71
X_{10}	FE	0.00	RE	0.58	RE	0.58	FE	0.00	RE	0.16
X_{11}	FE	0.00	FE	0.03	FE	0.00	RE	0.45	FE	0.00
X_{12}	FE	0.00	RE	0.09	FE	0.01	FE	0.04	RE	0.25

Table 4. PC Unit Root Test Results

	Group 1		Group 2		Group 3		Group 4		Group 5	
	PC	S/NS	PC	S/NS	PC	S/NS	PC	S/NS	PC	S/NS
X_1	0.02	S	0.00	S	0.77	NS	0.95	NS	0.55	NS
X_2	0.01	S	0.69	NS	0.04	S	0.01	S	0.01	S
X_3	0.02	S	0.80	NS	0.032	S	0.40	NS	0.25	NS
X_4	0.00	S	0.78	NS	0.00	S	0.88	NS	0.02	S
X_5	0.00	S	0.17	NS	0.01	S	0.01	S	0.02	S
X_6	0.01	S	0.04	S	0.01	S	0.01	S	0.05	S
X_7	0.65	NS	0.05	S	0.74	NS	0.05	S	0.04	S
X_8	1.00	NS	0.47	NS	0.61	NS	0.75	NS	0.05	S
X_9	0.10	NS	0.77	NS	0.00	S	0.05	S	0.05	S
X_{10}	0.70	NS	0.50	NS	0.31	NS	0.66	NS	0.7	NS
X_{11}	1.00	NS	0.49	NS	0.78	NS	0.95	NS	1.00	NS
X_{12}	0.70	NS	0.12	NS	0.42	NS	0.01	S	0.22	NS

In Table 3, p-values are used to decide the result of test. The p-values being smaller than 0.05 means that the null hypothesis cannot be rejected, and if they are equal or bigger than 0.05, it means that the null hypothesis can be rejected. In this respect, based on the results given in Table 3, it is possible to claim that the FE model must be preferred for 31 of the 60 panel data, and the RE model must be preferred for 29.

After the selection of the model, the Breusch Pagan Test was applied to each of the 60 panel data. As a result of the Breusch Pagan Test, it was observed that all the significance values were smaller than 0.05, and for this reason, there is cross section dependency in all of the 60 panel data. As a result of this, the PC test, which is one of the second generation unit root tests, was used to test the stationary of the panel data. Table 4 shows the results of the PC test.

In the panel data stationary test, while

the null hypothesis is in the form of *"the panel data is not stationary"*, the alternative hypothesis is *"the panel data is stationary"*. If the PC significance values are equal to or bigger than 0.05, the panel data are not stationary, if not, the panel data are stationary. In Table 4, S means that the panel data are stationary, and NS means the panel data are not stationary. In this context, 28 of the 60 panel data in

Table 4 are stationary, and 32 are not stationary. The statistical tests here upon will be performed over stationary panel data. The reason for this is to perform the comparisons of the performances of the FPDA and PDA for the situations in which PDA is usable and reliable.

The significance values for the stationary panel data and for each group are given in Table 5.

Table 5. Heteroscedasticity Test Results

Group 1		Group 2		Group 3		Group 4		Group 5	
Var.	p	Var.	p	Var.	p	Var.	p	Var.	p
X_1	0.53	X_1	0.00	X_2	0.00	X_2	0.00	X_2	0.00
X_2	0.00	X_6	0.00	X_3	0.99	X_5	0.00	X_4	0.00
X_3	0.99	X_7	0.00	X_4	0.00	X_6	0.00	X_5	0.00
X_4	0.00			X_5	0.00	X_7	0.00	X_6	0.00
X_5	0.00			X_6	0.00	X_9	0.00	X_7	0.00
X_6	0.00			X_9	0.92	X_{12}	0.00	X_8	0.00
								X_9	0.00

Table 6. Autocorrelation Test

Group 1		Group 3	
Var.	Sig.	Var.	Sig.
X_1	0.01	X_3	0.20
X_3	0.08	X_9	0.14

In the heteroscedasticity testing, the significance values being equal to and bigger than 0.05 means that the variance between the groups is equal. Based on this, it is possible to claim based on Table 5 that the variance is equal between the groups in panel data related with X_1 and X_3 from the 1st group and with X_3 and X_9 variables from the 3rd group. The last test in the PDA is autocorrelation testing. When the analysis is continued with the panel data that do not have any heteroscedasticity problem the results given in Table 6 are obtained for the autocorrelation testing.

While in the autocorrelation testing, the null hypothesis is in the form of *"there is no autocorrelation"*, the alternative hypothesis is *"there is autocorrelation"*. In this case, the

significance values being equal to or bigger than 0.05 means that there is no autocorrelation problem in the panel data in question. It is possible to claim based on Table 6 that there are no autocorrelation problems in the panel data related with 1st group X_3 variable, and in the panel data related with the X_3 and X_9 variables in the 3rd group.

4.2. Comparison results

In this section, the comparison results of the PDA and FPDA methods are given. For the purpose of increasing the comparisons, all the stationary data were included in the analysis. The Root Mean Square Error (RMSE), Mean Absolute Percentage Error

(MAPE) and Variance Accounted For (VAF) criteria, which are computed below, were used as the comparison criteria.

$$MSE = \sqrt{\frac{\sum_{i=1}^{N*T}(y_i - \hat{y}_i)^2}{N*T}} \quad (16)$$

$$MAPE = \frac{\sum_{i=1}^{N*T}|y_i - \hat{y}_i|/y_i}{N*T} * 100 \quad (17)$$

$$VAF = \left(1 - \frac{var(y_i - \hat{y}_i)}{var(y_i)}\right) x 100 \quad (18)$$

Table 7. Comparison Criteria for Group 1

Variable	RMSE		MAPE		VAF	
	PDA	FPDA	PDA	FPDA	PDA	FPDA
X_1	5472.41	2157.60	12.28	4.72	70.38	83.49
X_2	2026.12	2418.78	3.99	4.88	85.27	79.22
X_3	4933.45	1586.74	9.83	3.27	14.72	91.01
X_4	2238.72	2709.25	4.55	5.73	82.02	77.60
X_5	1858.08	2113.99	3.73	4.17	87.61	84.23
X_6	1525.19	1630.58	3.32	3.68	91.65	90.57

Table 8. Comparison Criteria for Group 2

Variable	RMSE		MAPE		VAF	
	PDA	FPDA	PDA	FPDA	PDA	FPDA
X_1	11888.98	4216.26	37.13	11.77	26.57	90.22
X_6	13925.33	3226.15	50.35	8.99	10.78	94.78
X_7	1996.43	5016.83	6.11	9.21	97.77	86.12

Table 9. Comparison Criteria for Group 3

Variable	RMSE		MAPE		VAF	
	PDA	FPDA	PDA	FPDA	PDA	FPDA
X_2	3197.51	3423.18	5.41	6.14	90.28	88.18
X_3	9936.87	3070.34	19.63	5.52	3.70	90.40
X_4	9633.85	8243.26	18.76	12.75	0.55	38.62
X_5	9698.17	4274.97	18.85	7.33	1.00	81.62
X_6	9548.10	2963.90	19.75	7.15	6.88	91.10
X_9	13291.90	3376.35	21.32	7.28	82.42	88.40

Table 10. Comparison Criteria for Group 4

Variable	RMSE		MAPE		VAF	
	PDA	FPDA	PDA	FPDA	PDA	FPDA
X_2	2549.04	2643.60	8.37	8.66	88.04	87.15
X_5	2731.36	3048.40	9.07	9.99	86.28	84.17
X_6	1906.69	2090.65	5.02	5.81	93.31	92.15
X_7	1801.80	2394.76	6.12	7.18	94.03	90.14
X_9	7103.82	3033.43	22.92	8.27	29.30	83.08
X_{12}	2891.30	2970.66	9.55	9.99	84.62	83.79

Table 11. Comparison Criteria for Group 5

Variable	RMSE		MAPE		VAF	
	PDA	FPDA	PDA	FPDA	PDA	FPDA
X_2	3584.21	4205.42	10.26	10.38	97.40	96.55
X_4	3909.59	4169.37	9.77	8.66	96.90	96.50
X_5	3638.29	4302.22	10.56	10.01	97.33	96.46
X_6	3465.33	4452.63	6.42	11.53	97.58	96.34
X_7	2975.09	3769.98	6.66	8.30	97.58	96.33
X_8	22334.51	5511.43	54.43	15.30	0.30	93.91
X_9	21033.88	4398.93	61.87	10.43	26.20	96.43

In Table 7, Table 8, Table 9, Table 10 and Table 11, the situations in which the performance of the FPDA is better are given as underlined. When these situations are evaluated, it was observed that the FPDA provides the best results especially for the predicted panel data with the RE model especially in the PDA. In other situations, the model gives results that is close to the PDA. The better performance of the FPDA especially in RE models may be explained as follows. The FPDA may be considered as the fuzzy version of the fixed effects model. If the panel data covers especially the stationarity hypothesis, the FE model will give mostly better results since it is based on minimization of the difference between the real value and the predicted value. However, obtaining close result from FPDA may be considered as an important advantage since it does not require any hypothesis or limitation on panel data.

5. Conclusion

The most important objective of the PDA is to determine the relationship between two or more variables in the cross-section data that have time dimension. For this objective, panel data regression models are used. Panel data regression models are divided into two groups as FE and RE regression models according to the form of the effects that are not observed through unit or time in the model. For performing the PDA, firstly it must be decided to which model the dataset fits more. However, even if the model that

fits the dataset is selected, the reliability of the regression models obtained as a result of the PDA depends on satisfying of some statistical hypotheses such as heteroscedasticity, autocorrelation, correlation between the units, and stationarity, which is extremely difficult in real-time applications.

In order to overcome these disadvantages of PDA, FPDA, which requires no statistical assumptions, is proposed in this study. FPDA is the fuzzy version of the FE regression model. Here, the reason why the FE panel data regression model was preferred is that in RE model, the unit effects are added to the model as error term; however, in fuzzy regression model, there are no other error terms, and the error is considered as the dispersion of the model parameters. The most important superiority of the proposed method is that it merges the PDA, which has the ability of performing the time series analysis simultaneously for multiple units, and the fuzzy regression, which has the ability of making decisions under uncertainty without any constraints and statistical assumptions.

In order to evaluate the efficiency of the proposed model, and to compare the prediction results with the PDA, the econometric relationship between 12 different socio-economic variables and GDP was predicted. For this purpose, 5 panel data were used consisting of 9 time points between 2005-2013 each for 7 countries. A total of 60 panel data were analyzed separately for 5 Groups and 12 variables.

The stationary panel data were used for

the purpose of comparing the performances of the PDA and FPDA, and the non-stationary panel data is excluded from the analysis. The RMSE, MAPE and VAF goodness of fit criteria is used for the comparisons. As a result of the analyzes, the following conclusions are obtained;

1. The PDA is better in 14 of 28 stationary panel data at, and the FPDA method prediction performance is better in the other 14.
2. There are no serious differences between the prediction performances of both methods in case the PDA is better; however, there are serious differences between the prediction performances in case the FPDA is better.
3. The PDA gives better results in panel data in FE model, the FPDA gives better results in panel data in RE model.
4. It is more suitable to use the PDA for prediction in case the panel data in question does not meet the required statistical hypotheses, and in the contrary situation, the FPDA is better.

The most important disadvantage of the FPDA proposed in this study is that it only consider the relationship between dependent and independent variables in the fuzzy panel regression model. We are planning to develop Dynamic Fuzzy Panel Data Analysis model that contains at least one lagged variable of dependent variable in our future work.

References

Ahn, S.C. & Moon, H.R. (2001) Large-N and Large-T Properties of Panel Data Estimators and the Hausman Test, USC Cleo Research Paper, 01-20.

Bai, J., and Ng, S. (2004) A Panic Attack on Unit Roots and Cointegration, Econometrica, **72**(4): 1127-1177.

Baltagi, B.H. (1995) Econometric Analysis of Panel Data, John Wiley & Sons, New York, 458s.

Baltagi, B.H. (2005) Econometric Analysis of Panel Data, 3.Baskı, John Wiley & Sons, New York, 378s.

Baltagi, B., Feng, Q. & Kao, C. (2012) A Lagrange Multiplier Test for Cross-Sectional Dependence in a Fixed Effects Panel Data Model, Center for Policy Research. 193.

Bashir, H., Ahmad, S., S., Urooj, R. & Nawaz M. (2017) Modeling some plant species distribution against environmental gradients using multivariate regression models, Kuwait Journal of Science, **44**(4):119-128.

Box, G.E.P. & Jenkins G.M. (1976) Time series Analysis: forecasting and control, Revised Edition, San Francisco, Holden Days.

Breitung, J. (2000) The Local Power of Some Unit Root Tests for Panel Data, In: Nonstationary Panels, Panel Cointegration and Dynamic Panels, Elsevier, s161-177.

Breitung, J. & Das, S. (2003) Panel Unit Root Test Under Cross Sectional Dependence, Statistica Neerlandica, **59**(4), 414-433.

Breusch, T.S. & Pagan, A.R. (1979) A Simple Test for Heteroskedasticity and Random Coefficient Variation, Econometrica, **47**(5), 1287-1294

Brown, M. B., & Forsythe, A. B. (1974) Robust tests for the equality of variances. Journal of the American Statistical Association. **69**: 364–367.

BuHamra, S., Al-Kandari, N., & Al-Harbi M. (2018) Parametric and nonparametric bootstrap: an analysis of indoor air data from Kuwait, Kuwait Journal of Science, **45**(2):22-29

Chang, Y. (2004) Bootstrap Unit Root Test in Panel with Cross-Sectional Dependency, Journal of Econometrics, **120**(2), 263-293.

Choi, I. (2001) Unit Root Tests for Panel Data, Journal of International Money and Finance, Elsevier, **20**(2), 249-272.

Choi, I. (2002) Combination Unit Root Tests for Cross-Sectionally Correlated Panels, Mimeo, Hong Kong, Hong Kong University of Science and Technology.

Engle, R.F. (1984) Wald Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. In Hand-book of Econometrics, Vol. 11, edited by Z. Griliches and M.D. Intriligator, pp. 775-826. Amsterdam:North Holland.

Ergün, E. (2011) A case study on fuzzy regression and semiparametric addivite regression,, Master Thesis, University of Muğla Sıtkı Koçman, , Muğla, s108.

Greene, H.W. (1993) Econometrics Analysis, McMillan, New York, s791.

Hadri, K. (2000) Testing for Stationarity in Heterogeneous Panel Data, The Econometrics Journal, **3**(2), 148-161

Harris, R.D.F. & Tzavalis, E. (1999) Inference for Unit Roots in Dynamic Panels Where the Dimension is Fixed, Journal of Econometrics, **91**:201-226.

Hausman, J.A. (1978) Specification Test in Econometrics, Econometrica, **6**(6), 1251-1271.

Hsiao, C. (2003) Analysis of Panel Data, 2.Baskı, Cambridge University Press, Cambridge, 384s.

Im, K.S., Pesaran, M.H. & Shin, Y. (1997) Testing for Unit Roots in Heterogeneous Panels, Mimeo, 1-30.

Im, K.S., Pesaran, M.H. & Shin, Y. (2003) Testing for Unit Roots in Heterogeneous Panels, Journal of Economics, **115**:53-74.

İşbilen Yücel, L. (2005) Bulanık Regresyon: Türkiye'de 1980-2004 Döneminde Kayıt Dışı Ekonominin Bulanık Yöntemlerle Tahminine İlişkin Bir Uygulama, Master Thesis, İstanbul University, İstanbul, 129s.

Levin, A. & Lin, C.F. (1992) Unit Root Test in Panel Data: Asymptotic and Finite Sample Properties, University of California, **108**:1-24.

Levin, A., Lin, C., Chu, J. & Shang, C. (2002) Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties, Journal of Econometrics, **108**:1-24.

Maddala, G.S. & Wu, S. (1999) A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, Oxford Bulletin of Econometrics and Statistics, special issue, 631-652.

Moon, H.R., Perron, B. & Phillips P.C.B. (2003) Incidental Trends and The Power of Panel Unit Root Test, Journal of Econometrics, **141**(2): 416-459.

Pesaran, M.S. (2007) A Simple Panel Unit Root Test in The Presence of Cross Section Dependence, Journal of Applied Econometrics, **22**:265-312.

Phillips, P.C.B. & Sul, D. (2003) Dynamic Panel Estimation and Homogeneity Testing Under Cross-Section Dependence, Econometrics Journal, **6**:217-259.

Shadab, A., Said, S.& Ahmad, S. (2019) Box-Jenkins multiplicative ARIMA modeling for prediction of solar radiation:a case study, International Journal of Energy and Water Resources, **3**:305-318.

Tanaka, H., Uejima, S. & Asai, K. (1982) Linear Regression Analysis with Fuzzy Model, IEEE Systems, Trans. Systems Man Cybernet, 203-907.

Wahid, F., Kim, D., H., (2017) Short-term energy consumption prediction in Korean residental buildings using optimized multi-layer perceptron, Kuwait Journal of Science, **44**(2), 67-77.

Yerdelen Tatoğlu, F. (2012) Panel Veri Ekonometrisi (Stata Uygulamalı), 3.baskı, Beta Yayınları, 334s.