Application of Spatial Error Model For Ordinary And Fuzzy Data With Comparison

Dalia Badee Omar
Department of Statistics Sciences, College of Administration & Economic, Kirkuk University, Iraq
dalia.badee@uokirkuk.edu.iq

Abstract

In this paper, an application of spatial data (ordinary & fuzzy) was studied by using a spatial regression model, which is the spatial error model (SEM), data used are 38 location for several Iraqi cities. In this study estimated parameter's model by ordinary least squares OLS and maximum likelihood MLE method for (ordinary and fuzzy data) then applied Simulation method to ensure the performance of the methods for estimating the model that used in our study, accordingly a program was designed in Matlab language , included the generation of data with different sample sizes (n =50, 100) default values for the parameters model, the spatial dependence parameter and a spatial weights matrix was used according the Rook Contiguity. After repeating the experiment 1000 times and comparison through statistical criteria (RMSE) the simulation results showed that the spatial error model for fuzzy data is better than the spatial error model for ordinary data.

Keywords: Spatial Error Model, Spatial data, Trapezoidal Fuzzy Number, Fuzzy spatial error model.

1 INTRODUCTION

The use of spatial econometrics was in the late 1971 by the researchers paelinck J.H. and Klaassen L. H. and it represents a scientific method for analyzing economic spatial series [1]. The analysis of spatial data has attracted a great interest in researches since the adoption of spatial in standard models, where the researchers put econometric a number of spatial models that are concerned with the processes of spatial analysis of data and its applications in a large number of uses and these spatial models are also concerned with the study of spatial effects between observations of the studied phenomena such as Economic, agricultural, service, health, etc. If the spatial dependence is not taken into account, this contradicts the assumptions of classic econometrics, which leads to biased and inconsistent estimations [2]. Therefore, spatial effects have an impact on the econometric model, which is called spatial econometric [3]. Spatial analysis, commonly referred to as geospatial data science is a geographical solution that combines data science with geographic solutions like geographic information systems (GIS) used in this paper to appear the Proximity relationship between regions [4]. This location is areas neighbor by limited or two areas connected a point. To test spatial autocorrelation in this study a researcher selected moran's test for spatial error autocorrelation model [5]. In statistics this tool is a measure of spatial autocorrelation depends for regression model. The aim of this research is to estimate the spatial error model, which suffers from the problem of correlation of errors for ordinary data and fuzzy data, then compare them using statistical comparison criteria to chosen the best model.

Manuscript received on: 15.09.2023
Accepted on: 22.10.2023
Published on: 30.11.2023
Issue DOI: doi.org/10.52688/21

2 SPATIAL DATA

Spatial data have characteristics, they contain attribute and locational information, spatial relationships are modeled with spatial weight matrix (W). The spatially data is characterized by spatial dependence between the observations of the data at different points, the observations are close to each other, it has a greater degree of spatial dependence than those far from the center, that is the strength of the spatial dependence between the observations decreases with the distance between them, and ignoring the spatial dependence may lead to the weakness of statistical methods for analyzing spatial data [6].

2.1 Spatial autoregressive error model (SEM)

The spatial regression model (SEM) is a special case of regression as the disturbances show spatial dependence, which is one of the problems in this model in the sense that the error limit is not independent, and it differs from the traditional model that assumes the independence of errors [7].

The spatial dependence of errors contradicts one of the basic assumptions of estimations in traditional econometrics which is the independence of error. In the case of neglecting the spatial dependence of errors, then the estimates and tests will be inefficient, meaning that the statistical analysis of a model is incorrect, so the spatial error model is appropriate when using spatial data to process The effects of spatial autocorrelation of errors, an SEM model is specified as:

$$Y = X\beta + u \quad (1)$$

$$u = \rho W u + e$$

where:

$\underline{e} \sim N(0, \sigma^2 I_n)$

The formula can be written as follows:

$$Y = X\beta + (I_n - \rho W)^{-1} e \quad (2)$$

Y: is a vector $(1\times n)$ dependent variable.

X: matrix $(n\times k)$ which is the independent variable.

ρ: parameter of spatial Regression.

W: is the spatial weights matrix$(n\times n)$.

β: is a vector(n × 1) Parameters to be estimated.

u: vector $(1\times n)$ spatial error correlation.

e: is the vector (1 × n) for error term.

The spatial autocorrelation depends on the idea of bilateral contiguousness between the observations, so if there are two spatial units that have the same common boundaries of non-zero length, then they are considered contiguous and have a value (1), if it has not a common border then they are non-contiguous and have a value (0), according to the Rook Contiguity criterion. and thus it is possible to find the spatial adjacency matrix, which is known as the spatial weights matrix (W) that most commonly used based on geographic arrangement of the observations or contiguity. The spatial weights matrix can be also based on distance [8][9]. In this paper, we rely on Row-Standardized Weight Matrix , where the sum of each row in this adjust matrix is equal to one, as follow:

$$w_{ij}^{std} = \begin{bmatrix} \frac{w_{ij}}{\sum_{i=1}^{n} w_{ij}} & \text{if i neighbor j} & 0 < w_{ij}^{std} \leq 1 \\ 0 & if & otherwise \end{bmatrix} \quad (3)$$

2.2 Estimation methods

In this study, the parameters of the spatial error model for ordinary data were estimated by:

2.2.1 Maximum likelihood method (SEM) Model

The greatest possibility method was used, which gives the best estimate among several estimates of the model parameters .In the spatial error model, the concern is parameter $\boldsymbol{\rho}$, which is a coefficient that shows the correlation between the residuals, and the concentrated likelihood function for this model is as follows [10].

$$\text{Ln } L(\beta,\rho,\sigma^2) = -\frac{n}{2}\text{Ln}(2\pi) - \frac{n}{2}\text{Ln } \sigma^2 + Ln|\, I - \rho W| - \frac{1}{2\sigma^2}[(Y - X\beta)'\,(I - \rho W)'\,(I - \rho W)(Y - X\beta)] \quad (4)$$

We find the derivative of equation (4) with respect to β and σ^2 equating it to zero we get:

$$b_{ML} = [X'(I - \rho W)'(I - \rho W)X]^{-1}X'(I - \rho W)'(I - \rho W)\,Y]$$

$$e = [((I - \rho W)\,Y) - ((I - \rho W)\,X)\,b_{ML}] (5)$$

$$\sigma^2_{ML} = \frac{e'e}{n} \quad (6)$$

Ord suggested in 1975 that the determinant can be calculated as in below:

$$|\, I - \rho W \,| = \prod_{i=1}^{n}(1 - \rho w_i)$$

$$L_n\,(1 - \rho w_i) = \sum_{i=1}^{n} L_n\,(1 - \rho w_i)$$

After estimating the parameters of the model by eigen values for weights matrix and substituting the estimators into the concentrated likelihood function which contains one parameter (ρ), as follows:

$$L_c = \frac{-n}{2}\,[\frac{e'e}{n}] + \sum_{i=1}^{n} L_n\,(1 - \rho w_i) \quad (7)$$

2.2.2 Fuzzy sets

It is the group whose elements have a degree of affiliation or a membership degree and are real numbers within the closed interval [0,1]. The degree of membership is expresses by the affiliation function or the membership function, which represents the degree of belonging of the variable x to the fuzzy group and is written in the following form:

$$\boldsymbol{A} = \{(\boldsymbol{x}, \boldsymbol{\mu_A(x)}) : \boldsymbol{x} \in \boldsymbol{X} \quad : \quad \boldsymbol{\mu_A(x)} : \boldsymbol{X} \to [\boldsymbol{0}, \boldsymbol{1}]\}$$

If it is in an intermittently state, it will be as follows:

$$\boldsymbol{A} = \sum_{i=1}^{n} \frac{\mu_A(x_i)}{x_i} = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \ldots\ldots + \frac{\mu_A(x_n)}{x_n} \qquad (8)$$

If it a continuous state, it is written as follows:

$$\boldsymbol{A} = \left\{ \int \frac{\mu_A(x)}{x} \right\} \qquad (9)$$

2.2.3 Trapezoidal Fuzzy Number

The fuzzy subset A in the universal set X is called the trapezoidal fuzzy number which is expressed in the form A=(a,b,c,d) where a< b < c <d if it has an affiliation function as follows,(Wang et 2006) (Alavala,2008):

$$\mu_A(x) = \begin{bmatrix} \frac{x-a}{b-a} & if \quad a \le x \le b \\ 1 & if \quad b \le x \le c \\ \frac{d-x}{d-c} & if \quad c \le x \le d \\ 0 & other\ wise \end{bmatrix} \qquad (10)$$

3 FUZZYFICATION OF DATA

In this study, fuzzy trapezoidal data was used, a spatial error model was formulated for this data and the parameters of this model were estimated by two method [11].

3.1 Fuzzy spatial regression error model (FSEM)

The parameter of fuzzy spatial error model $\rho \neq 0$ we get model as follow:

$$y_c = z_c\mathrm{B} + u_c$$

$$|\rho| < 1 \qquad (11)$$

$$u_c = \rho W u_c + e_c$$

Where: $e_c \sim N(0, \sigma^2_{\epsilon n} I_n)$

then:

$$u_c = (\mathrm{I} - \rho \mathrm{W})^{-1} e_c$$

As follow:

$$y_c = z_c\mathrm{B} + (\mathrm{I} - \rho \mathrm{W})^{-1} e_c \qquad (12)$$

Where:

y_c: is (1× p) vector which is the Centroid of the fuzzy number trapezoidal of the dependent variable.

z_c: matrix (k× n) which is the Centroid of the fuzzy number trapezoidal of the independent variable.

I: Identity matrix (n× n).

ρ: parameter of spatial regression

W: is the spatial weights matrix(n× n).

β: is a vector(1 × n) Parameter to be estimated.

e_c: is the vector (1 × p) for error term.

We notice in this model that when $\rho = 0$ there is no spatial correlation between the errors of the adjacent observations and this means that we get the normal fuzzy regression model, and if ρ is not equal to zero, we get a model with a fuzzy spatial correlation between the errors of the adjacent observations[13]. Estimation methods for fuzzy (FSEM) Model: This model estimated by two method [12].

3.2 Ordinary Least Square (OLS) for (FSEM)

When we use OLS method to estimate parameters (defined by the formula 11) we will assume that the parameter is known and estimated in advance [13].

$$y_c = z_c\mathrm{B} + u_c$$

$$u_c = (\mathrm{I} - \rho \mathrm{W})^{-1} e_c = B^{-1} e_c$$

$$y_c = z_c\mathrm{B} + B^{-1} e_c$$

$$e_c = y_c\mathrm{B} - \mathrm{B}z_c\mathrm{B} \qquad (13)$$

$$e_c' e_c = (\mathrm{B}' y_c' - \beta' z_c B')(y_c\mathrm{B} - \mathrm{B}z_c\beta) = y_c B B y_c' - 2y_c' B' B z_c \beta + \beta' z_c' B' B z_c \beta$$

By making the derivation with respect to (β) and equal to zero, we get:

$$\hat{\beta}= (z_c'(B'B)z_c)^{-1}\ y_c'(B'B)\ z_c \qquad (14)$$

Where $(B'B)$= V

$\hat{\beta}=\ (z_c'Vz_c)^{-1}\ y_c'V\ z_c$

The variance as follows:

$$\sigma^2=\frac{(y_c\text{B} - \text{B}z_c\hat{\beta})'\ ((y_c\text{B} - \text{B}z_c\hat{\beta})}{n-k} \qquad (15)$$

3.3 Maximum likelihood for (FSEM)

In the spatial error model, the focus is on the parameter $\boldsymbol{\rho}$ as it shows the amount of correlation between fuzzy errors and the model is estimated by the MLE method as follows:

$y_c = z_c\text{B} + u_c$

$u_c = \rho W u_c +\ e_c$

$y_c = z_c\text{B} + \rho W u_c +\ e_c$

And compensate for the error value in the likelihood function:

$$\text{L}(\beta, \rho,\ \sigma^2\ /\ y_c\ , z_c) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\ |\text{I} - \rho W| \exp[-\frac{1}{2\sigma^2}\ \text{e}'(\text{I} - \rho W)'\ (\text{I} - \rho W)e'] \qquad (16)$$

$\text{Ln L}= -\frac{n}{2}\ \text{Ln}(2\pi) - \frac{n}{2}\text{Ln}\ \sigma^2 + Ln|\ \text{I} - \rho W| - \frac{1}{2\sigma^2}\ [(y_c - z_c\text{B})'\ (\text{I} - \rho W)'\ (\text{I} - \rho W)\ (y_c - z_c\text{B})]$

We find the derivative of equation () with respect to β and σ^2 equating it to zero we get:

$y_c{}^* = (\text{I} - \rho W)\ y_c\ \ ,\ z_c{}^* = (\text{I} - \rho W)\ Z_c$

$$\hat{\beta}_{\text{MLE}}= (\text{z}'_c{}^*\ \text{z}_c{}^*)^{-1}\text{z}'_c{}^*\ \text{y}_c{}^* \qquad (17)$$

When putting equation () in equation () we get:

$\hat{\beta}_{MLE}\ =\ [z'_c(\text{I} - \rho W)'(\text{I} - \rho W)z_c]^{-1}z'_c(\text{I} - \rho W)'(\text{I} - \rho W)\ y_c]$

$\mathbf{e_c}= y_c - z_c\hat{\beta}_{MLE}$

$$\sigma^2{}_{MLE}=\frac{\text{e}'_c\text{e}_c}{n} \qquad (18)$$

After substituting the value of $\hat{\beta}_{MLE}$ and $\sigma^2{}_{MLE}$ into the likelihood function, we get

$$L_c= -\frac{n}{2}\ \text{Ln}[\frac{1}{2}\ \mathbf{e'_c}(\text{I} - \rho W)'(\text{I} - \rho W)\mathbf{e_c}] + Ln|\ \text{I} - \rho W| \qquad (19)$$

By using the iterative method for concentrated likelihood function ,we can obtained (ρ) value.

4 MODEL FITTING

The spatial autocorrelation patterns of (SEM) model for ordinary data and fuzzy data were compared using statistics coefficient of root mean square error. Root Mean Squares Error (RMSE): It measures the average difference between values predicted by a predictive model and the actual values. It has widely been used for evaluating the accuracy of a prediction. The smaller value of criterion is the best model.

The formula for calculating this criterion is:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\ y_i - \hat{y}_i)^2}{n-k-1}} \qquad (20)$$

5 EXPERIMENTAL PART

Simulation can be considered a way to check the performance of the theoretical method before using it with the ordinary data. In this paper the implementation of simulation experiments includes for spatial regression error model, as they are executed as follows:

The experiment was repeated (R= 1000) times after determining the different sample sizes (n=50,100) and assuming values for the spatial dependence parameter (ρ = 0.8). Generating the variables represented by the random error limit and the independent variable (X, e) which are distributed as follows:

X~$\mathbf{N}(\mathbf{0}, \boldsymbol{\sigma}\ \mathbf{2}\ )$, e ~$\mathbf{N}(\mathbf{0}, \mathbf{1})$, where $\boldsymbol{\sigma}$ (6.5)

Generating initial value of parameters ($\boldsymbol{\beta}\mathbf{0}$=6.370, $\boldsymbol{\beta}\mathbf{1}$=0.113, $\boldsymbol{\beta}\mathbf{2}$= 0.632, $\boldsymbol{\beta}\mathbf{3}$ =0.202).

Generating Row- Standard Weight Matrix, according to the Rook Contiguity. After generating the independent variable and the random error variable and Row-standardized weight matrix and determining the values of the spatial parameter (ρ), they are compensated in spatial regression model in order to obtain the dependent variable (Y). Comparison the SEM, FSEM spatial regression model for ordinary and fuzzy data, according to comparison criteria (RMSE).

Figure 1: show the neighbor areas of the region.

The tables (1,2) show the results from SEM model that estimate by MLE method for different samples and the spatial dependence parameter (ρ)= 0.8.

Table 1: estimation method for (SEM) Model.

	n=50				RMSE
β	6.718	0.611	0.472	0.625	0.4673

Table 2: estimation method for (SEM) Model

	n=100				RMSE
β	6.033	0.431	0.505	0.187	0.3532

These two tables show the results for (SEM) model estimation by maximum likelihood method ,in table (2) the RMSE value is less than RMSE value in table (1) this mean that the sample size has an effect, when the sample size larger the model be best. As for the fuzzy data ,Simulation method was used to generate random samples of spatial error regression model with different sizes(n=50, 100) and the number of independent variables is three. Then these samples(data) were fuzzed by trapezoidal fuzzy number type, according the Chen method with the existence of a function of belonging to each of the independent and dependent variables, the Maximum Likelihood method and the Ordinary Least Square method were used to estimate the parameters of the studied model, a comparison was made between the estimation methods at different sample sizes and the root mean square error criterion was adopted (RMSE) .

Table 3: estimation method for (FSEM) model

N	OLS		MLE	
	ρ	RMSE	ρ	RMSE
50	-0.1483	0.3127	-0.2673	0.2198
100	-0.9745	0.2374	-0.6201	0.2377

In Table (3) show the (FSEM) model that contain three independent variable that the result of RMSE in MLE method less than the RMSE in ordinary lease square ,that's mean the best method for estimating is MLE in this model for fuzzy data in different sample sizes.

6 CONCLUSIONS

If we compare the results SEM model for the ordinary data and FSEM for the fuzzy data that estimated by MLE method , we notice that the RMSE for different sample size is less in (FSEM) model for fuzzy data than the RMSE for (SEM) model for ordinary data, that's mean the best model estimated by the MLE method is FSEM for fuzzy data .we recommend the ordinary data can be fuzzed into triangular fuzzy number compared with fuzzy data of the trapezoidal fuzzy number type . As for the spatial dependence parameter (ρ) in FSEM model , the table(3) shows the correlation strength is stronger when the sample size is 50 but when the sample size is larger the value of (ρ) decreases. The strength relationship between the observations decreases as we moved away from the center.

7 REFERENCE

[1] Alavala, C.R., 2008. "Fuzzy logic and neural networks: basic concepts & application". New Age International.

[2] Ali Mohammed Ali Chichan, Mustafa Ali Fakhri, Zana najim(2021). " Comparison Between Iterative Maximum Likelihood Estimators Method And Bayesian Method For Estimating Logistic Regression Model Parameters With Practical Application "Journal of Economics And Administrative Sciences, Published by Kirkuk University,2015 volume(10)Issue (2), pp(259-270).

[3] Chen, G., & Pham, T. T. (2000). "Introduction to fuzzy sets, fuzzy logic, and fuzzy control systems". CRC press

[4] Hesamian, G., & Akbari, M. G. (2020). "A kriging method for fuzzy spatial data". International Journal of Systems Science, 51(11), 1945- 1958

[5] Higazi, S. F., Abdel-Hady, D. H., & Al-Oulfi, S. A. (2013)."Application of spatial regression models to income poverty ratios in middle delta contiguous counties in Egypt". Pakistan Journal of Statistics and Operation Research,93-110.

[6] Jaufar Mousa Mohammed , Maysoon M. Aziz , Amer Fadhil Tawfeeq (2022). ” Fuzzy Estimations in Spatial Statistics”, Computer Science and Mathematics University of Mosul.

[7] Jaufar Mousa Mohammed (2014).” Estimation Multivariate data points in spatial statistics with application”, College of Administration and Economics at Kirkuk University, journal of Economics And Administrative Sciences, Published by Baghdad University , 2015 volume(21) Issue (85),pp(410-423).

[8] Karim, A., Faturohman, A., Suhartono, S., Prastyo, D. D., & Manfaat, B. (2017). "Regression Models for Spatial Data: An Example from Gross Domestic Regional Bruto in Province Central Java". Journal Economic Pembangunan: Kajian Masalah Ekonomi dan Pembangunan, 18(2), 213-224

[9] Lou, M., Zhang, H., Lei, X., Li, C., & Zang, H. (2016). "Spatial autoregressive models for stand top and stand mean height relationship in mixed Quercus mongolica broadleaved natural stands of northeast China". Forests, 7(2), 43.

[10] Lu, J., & Zhang, L. (2011). "Modeling and prediction of tree height–diameter relationships using spatial autoregressive models". Forest Science, 57(3), 252-264.

[11] Sham Azad Rahim,(2016)."Comparison Between Classical and Spatial Regression Techniques using Fuzzy Logic". University of Sulaimani College of Commerce

[12] Wati, A. D. A., & Khikmah, L. (2020). "Modeling Spatial Error Model (SEM) On Human Development Index (IPM) In Central Java 2018". J Int Comp & He Inf, 1(2).

[13] Zhang, A. (2016): "Statistical analysis of fuzzy linear regression model based on centroid method". Applied Mathematics, 7(07), 579.