Int. J. Nonlinear Anal. Appl. 13 (2022) 2, 1367–1375
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.6254

Idea about using ordinary least square by centroid methods for fuzzy pure spatial autoregressive model

Jaufar Mousa Mohammed[a,*], Maysoon M. Aziz[b], Ammer Fadel Tawfeeq[a]

[a]*Department Statistics, College Administration and Economics, University of Kirkuk, Iraq*

[b]*Department Mathematics, College Computer Sciences and Mathematics, University of Mosul, Iraq*

(Communicated by Javad Vahidi)

Abstract

Introduce some ideas about applied of spatial regression models for independent and dependent fuzzy variables. while the parameters crisp values, which are estimated by the ordinary least squares method. This paper has been formulated fuzzy Pure Spatial Autoregressive Model (FPSAM) from a fuzzy general spatial model, and applied for trapezoidal fuzzy number in the domain traffic accidents for a number of cities in Iraq for the year 2018 and that after converting the Trapezoidal fuzzy number into crisp values by centroid method, calculations the results by Matlab language.

Keywords: fuzzy spatial regression models, fuzzy pure spatial autoregressive model, centroid method
2020 MSC: 62A86

1 Introduction

Spatial econometrics is one of the concepts of traditional econometrics, Because it deals with the spatial phenomenon of each variable on the basis of place, these phenomena are specific and known measurements and the errors resulting from them are random variables that can be controlled by studying their behavior [4], As for fuzzy statistics, it has recently emerged after the emergence of the theory of fuzzy aggregates to be concerned with phenomena whose variables cannot be measured in points, but rather measured in periods, or what is described as uncertain cases or cases with fuzzy data because of its characteristics that make them unclear such as variables that belong in certain proportions to their aggregates. It has no complete affiliation, as well as linguistic variables that cannot be measured numerically, and there are variables that are measured roughly, but in fact they are ambiguous. As a result, fuzzy logic has become applied in many fields. The Artificial Intelligent model, especially in the field of artificial intelligence, is a technique that has a mechanical ability to find solutions to various scientific and applied problems, This is one of the reasons that prompted us to study the fuzzy logic in the general linear spatial regression, we get fuzzy pure spatial autoregressive model for fuzzy trapezoid data by centroid method, and we use least squares to estimated parameters [6].

*Corresponding author
Email addresses: jaufar@uokirkuk.edu.iq (Jaufar Mousa Mohammed), aziz_maysoon@uomosul.edu.iq (Maysoon M. Aziz), asas1964@uokirkuk.edu.iq (Ammer Fadel Tawfeeq)

Received: December 2021 *Accepted:* February 2022

2 Basic concepts in fuzzy logic

2.1 Fuzzy Set

It is set whose components have value of belonging, called the degree of membership, which are real numbers within the closed interval , and the degree membership is expressed as that represents the degree of belonging the element from the variable X to The fuzzy set A is written as:

$$A = (x, \mu_A(x)) : \quad \mu_A(x) : x \to [0, 1]$$

The memberships change from full or complete to non-membership, or partial membership.

2.2 Crisp Set

They are the elements that have a specific characteristic, which takes one of the two values, (1) when the element belongs to the set and (0) when the particular element does not belong to the set .it is called crisp set to distinguish it from the fuzzy set in concepts, let we have a set A known as a function and called the characteristic function as :

$$\chi_A(x) = \begin{cases} 0 & if \quad x \notin A \\ 1 & if \quad x \in A \end{cases}$$

2.3 Alpha Cat Set $\alpha-$ cat set

Let A fuzzy subset in universal set X , then we define an of A as:

$$A_{[\alpha]} = \{x \in X : \mu_A(x) \geq \alpha\}, \qquad \alpha \in [0, 1]$$

3 Strong $\alpha-$ cat set

Let A fuzzy subset in universal set X , then we define a strong of A as:

$$A_{[\alpha^+]} = \{x \in X : \mu_A(x) > \alpha\}, \qquad \alpha \in [0, 1]$$

3.1 Normalized Fuzzy Set (Core)

A fuzzy subset A in universal set X is called normalized (Core) if :

$$\sup_{x \in X} \mu_A(x) = 1$$

3.2 Convex Fuzzy Set

A fuzzy subset A in universal set X is called convex if :

$$\mu_A(t) \geq \min[\mu_A(r), \mu_A(s)] \; and \; t = \lambda r + (1 - \lambda)s \; where \; r, s \in R, \lambda \in [0, 1]$$

3.3 Fuzzy number

A fuzzy subset A in universal set X is called fuzzy number if satisfy following condition :

1. convex fuzzy set.
2. normalized fuzzy set (maximum membership value is 1).
3. it's membership function is piecewise continuous.
4. It is defined in the real number [8].

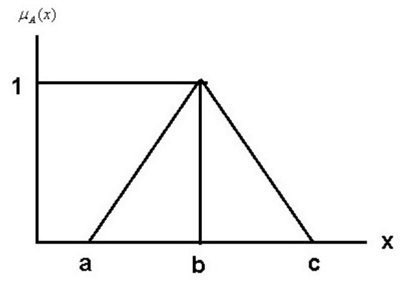

Figure 1: Triangular fuzzy number

3.4 Triangular fuzzy number

A fuzzy subset A in universal set X is called Triangular fuzzy number that expressed as $A = (a, b, c)$ where $a < b < c$ if has membership functions as:

$$\mu_A(x) = \begin{cases} \frac{x-a}{b-a} & if\ a \le x \le b \\ \frac{c-x}{c-b} & if\ b \le x \le c \\ 0 & Otherwise \end{cases} \tag{3.1}$$

3.5 Trapezoidal fuzzy number

A fuzzy subset A in universal set X is called Trapezoidal fuzzy number that expressed as $A + (a, b, c, d)$ where $a < b < c < d$ if has membership functions as [9]:

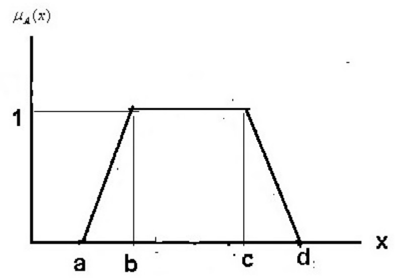

Figure 2: Trapezoidal fuzzy number

$$\mu_A(x) = \begin{cases} \frac{x-a}{b-a} & if\ a \le x \le b \\ 1 & if\ b \le x \le c \\ \frac{d-x}{d-c} & if\ c \le x \le d \\ 0 & Otherwise \end{cases} \tag{3.2}$$

3.6 Convert fuzzy number to crisp number (Defuzzification)

Let A fuzzy number we can transform A to crisp by centroid method this process is called defuzzification ,the centroid method has the following formula

$$A_c(x) = \frac{\int x\mu_{\tilde{A}}(x)dx}{\int \mu_{\tilde{A}}(x)dx = \frac{1}{3}(x_L + x_M + x_R)} \qquad if\ \ A\ \ Triangular \tag{3.3}$$

$$A_c(x) = \frac{\int x\mu_{\tilde{A}}(x)dx}{\int \mu_{\tilde{A}}(x)dx = \frac{1}{4}(x_L + x_{M_1} + x_{M_2} + x_R)} \qquad if\ \ A\ \ Trapezoidal \tag{3.4}$$

This method purposed by Sugeno in 1985 is the most commonly used technique and it is very accurate [2, 7, 12, 13].

3.7 Convert crisp number to fuzzy number (Fuzzyfication)

The convert process crisp number to fuzzy is called Fuzzyfication , and use membership function in convert which requires have range from zero and one, as shown in the following figure:

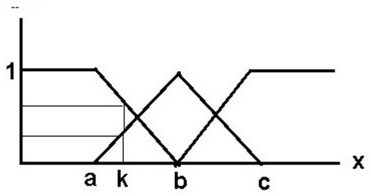

Figure 3: fuzzyfication

4 General Spatial Linear Regression Models

The form of general spatial model that contain both spatially lagged and error structure spatial correlation , As shown in the following formula

$$\begin{aligned} &y = \lambda W y + X\beta_1 + WX\beta_2 + u \qquad |\lambda| < 1 \\ &u = \rho W u + e \qquad |\rho| < 1 \end{aligned} \tag{4.1}$$

Where

X: a matrix of non stochastic regression

W: weight matrix

$e/X = i.i.d.N(0, \sigma^2_{en} I_n)$

β_1, β_2, vectors (n*1) of parameter require estimate him

λ, ρ: parameters spatial regression

u: residues spatially associated

e: vector (n*1) random error

the model (4.1) can be written as

$$\begin{aligned} &y = \lambda W y + X\beta_1 + Z\beta_2 + u \qquad |\lambda| < 1 \\ &u = \rho W u + e \qquad |\rho| < 1 \end{aligned} \tag{4.2}$$

Where $Z = [X, WX]$ and $\beta = [\beta_1, \beta_2]$, this model is called spatial Autoregressive by Autoregressive error structure and Includes many spatial econometric models they are as follows :

1. $\beta = 0$ and $\lambda = 0$ or $\rho = 0$ is called pure spatial autoregressive model.
2. $\lambda = 0$ and $\rho = 0$ is called lagged independent variable model.
3. $\lambda \neq 0$ and $\rho = 0$ is called spatial lag model.
4. $\lambda = 0$ and $\rho \neq 0$ is called spatial error model [1, 3].

4.1 Pure Spatial Autoregressive Model (PSAM)

In model (4.2), let $\beta = 0$ and $\lambda = 0$ or $\rho = 0$ and we assuming again $e/X = i.i.d.N(0, \sigma^2_{en} I_n)$ and matrix W non stochastic , so the model given simple spatial autoregression , thus when $\rho = 0$ we have model :

$$y = \lambda W y + e \qquad |\lambda| < 1 \tag{4.3}$$

So by Ordinary Least Square (OLS) we estimate the parameter $\hat{\lambda}$

$$\hat{\lambda} = (y'W'Wy)^{-1} y'Wy \tag{4.4}$$

And when $\lambda = 0$, we get

$$y = \rho W y + e \qquad |\rho| < 1 \tag{4.5}$$

Again by Ordinary Least Square (OLS) we estimate the parameter $\hat{\rho}$ is equal [3, 5]

$$\hat{\rho} = (y' W' W y)^{-1} y' W y \tag{4.6}$$

4.2 Ordinary Least Square For Fuzzy Pure Spatial Autoregressive Model by centroid

The Fuzzy Pure Spatial Autoregressive Model when $\rho = 0$ is written as

$$\tilde{y} = \lambda W \tilde{y} + \tilde{e} \qquad |\lambda| < 1 \tag{4.7}$$

Where

$\tilde{y}$: vector (n*1) is triangular or trapezoidal fuzzy number are depend variable.

λ: parameter spatial regression is crisp number.

W : spatial weights matrix is crisp number.

$\tilde{e}$: vector (n*1) is triangular or trapezoidal fuzzy number are random error [14].

Idea this paper convert triangular or trapezoidal fuzzy number to crisp number as (3.3) or (3.4). Then in this case we can written (4.7) as

$$y_c = \lambda W y_c + e_c \qquad |\lambda| < 1 \tag{4.8}$$

When $\rho = 0$

$$\begin{aligned} &\Longrightarrow y_c - \lambda W y_c = e_c \\ &\Longrightarrow (I - \lambda W) y_c = e_c \\ &\Longrightarrow y_c = (I - \lambda W)^{-1} e_c \end{aligned}$$

And when $\lambda = 0$, we get

$$\begin{aligned} y_c &= \rho W y_c + e_c \qquad |\rho| < 1 \\ &\Longrightarrow y_c - \rho W y_c = e_c \\ &\Longrightarrow (I - \rho W) y_c = e_c \\ &\Longrightarrow y_c = (I - \rho W)^{-1} e_c \end{aligned}$$

where

y_c: vector (n*1) is centroid of triangular or trapezoidal fuzzy number are depend variable.

e_c: vector (n*1) is centroid of triangular or trapezoidal fuzzy number are random error

And by Ordinary Least Square (OLS) we estimate the parameter $\hat{\lambda}$:

$$\begin{aligned} y_c &= \lambda W y_c + e_c \qquad |\lambda| < 1 \\ e_c &= y_c - \lambda W y_c \\ e_c' &= (y_c - \lambda W y_c)' = y_c' - \lambda y_c' W' \\ e_c' e_c &= (y_c' - \lambda y_c' W')(y_c - \lambda W y_c) = y_c' y_c - 2\lambda y_c' W y_c + \lambda^2 y_c' W' W y_c \end{aligned}$$

We derive this quantity with respect to λ and equal to zero we get:

$$\hat{\lambda} = (y_c' W' W y_c)^{-1} y_c' W y_c \tag{4.9}$$

Similarly when $\lambda = 0$ we get:

$$y_c = \rho W y_c + e_c \qquad |\rho| < 1 \tag{4.10}$$

And by (OLS) we get :

$$\hat{\rho} = (y_c' W' W y_c)^{-1} y_c' W y_c \tag{4.11}$$

5 Spatial weights matrix (Rook contiguity)

It is a square matrix which it elements positive value, and it is not necessary to be symmetric , and it is based on geographic arrangement of the observations, or contiguity , i.e. the relations among location with other location in one row of the matrix and the diagonal elements in the matrix are equal to zero ,the Spatial weights matrix by Rook contiguity is equal to 1 if the two areas (locations) neighbor by limited and have relation between the two areas (locations) in any side, and it is equal to 0 otherwise. This matrix used in applications more than the others [10].

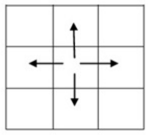

Figure 4: Shows the Rook weight matrix

6 Moran Test for Spatial Regression

It is a general measure depends on the general linear model GLM and uses for autocorrelation coefficient (called the Moran coefficient because Moran is the name of the Scientist that find the test), the Moran formula is:

$$I = \frac{n(e'we)}{S_0(e'e)} \tag{6.1}$$

where :

$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$ sum of all elements of matrix W.

W is Weights Matrix (neighborhood) it is square $n \times n$.

n: number observation (size sample).

e: vector residual dimensions $(n \times 1)$.

where we using row – standardized where sum of row equal to 1 in this case $(n = S_0)$ that is work to simplify the Moran's formula as follows :

$$I = \frac{(e'we)}{(e'e)} \tag{6.2}$$

To known if the value of the Moran coefficient it is Indicator Statistics in certain degree of confidence we must use moran (Z) test with Hypotheses:

$H_0 : \lambda = 0, \ \ \theta = 0$ there is no spatial dependence

H_1 : at least one of $\lambda \neq 0 \ \ or \ \theta \neq 0$ there is spatial dependence

$$\begin{aligned} Z &= \frac{I - E(I)}{\sqrt{V(I)}} \\ E(I) &= \frac{n(tr(MW))}{S_0(n-k)} \\ V(I) &= \frac{tr(MWMW') + tr(MW)^2 + (tr(MW))^2}{(n-k)(n-k+1)}(\frac{n}{S_0})^2 - (E(I))^2 \end{aligned} \tag{6.3}$$

Where

$M = I - X(X'X)^{-1}X'$ Idempotent Matrix that is $(n \times n)$ and symmetric

tr: Sum diagonal element .

k: Number of explanatory variables.

The calculated value Z is compared with value of$Z$ table for $\alpha/2$, and where Moran test is significant that is mean relation between geographic location that refers to use spatial regression model and not enough general linear model (GLM) and we have spatial dependency [4, 11].

7 Practical Part

7.1 Estimate initial value

In this part of the paper, we used, real data these data represent the number of deaths $\tilde{y}_i$resulting from traffic accidents in Iraq for six governorates (Baghdad, Anbar, Diyala, Salah al-Din, Kirkuk, Nineveh) by 38 observations. And this traffic accidents as Crash accidents $\tilde{X}_1$ and Overturn accidents $\tilde{X}_2$ and Run over accidents $\tilde{X}_3$,this data represent trapezoidal fuzzy number and has membership function, in this paper explains given idea about dealing with such data, in this peper transform fuzzy data into Crisp data, by the formula

$$x_c = \frac{1}{4}(x_L + x_{M_1} + x_{M_2} + x_R) \qquad y_c = \frac{1}{4}(y_L + y_{M_1} + y_{M_2} + y_R)$$

So the fuzzy multiple regression formula for this data is

$$y_{ci} = \beta_0 + \beta_1 X_{c1i} + \beta_2 X_{c2i} + \beta_3 X_{c3i} + e_{ci}$$

Where:

y_{ci}: the number deaths in traffic accidents (D.A)

X_{c1i}: Crash accidents (C.A)

X_{c2i}: Overturn accidents (O.A)

X_{c3i}: Run over accidents (R.A). We get β_i by ordinary least square

Table 1: Estimated initial value of beta

Model	Constant	C.A	O.A	R.A
Beta	7.3078	0.0544	0.7951	0.2653

7.2 Moran Test

The initial values that estimated are used in the Moran test, which we obtained the following result:

Table 2: result moran test

I	Z(I)
-0.0372	0.7878

Since the value of Z(I) is less than the tabular value of Z in terms of $(\alpha/2)$, then we accept the null hypothesis, which states that the data is spatially dispersed and randomly.

7.3 Estimate parameter FPSAM By Using (OLS)

To estimate the parameters of the fuzzy autoregressive model, we calculated spatial matrix weights between locations with 38×38 by (Rook contiguity), and we are have two cases, the first case when $\rho = 0$ then the form of the model is as follows:

$$y_c = \lambda W y_c + e_c$$

Second case when $\lambda = 0$

$$y_c = \rho W y_c + e_c$$

And after estimating the parameter λ and ρ by the ordinary least squares method as in formula (4.9) and (4.11).

Table 3: Estimate coefficient

Estimate coefficient **FPSAM**	
$\hat{\rho} = 0$	$\hat{\lambda} = 0.0385$
$\hat{\lambda} = 0$	$\hat{\rho} = 0.0385$

Since $\hat{\rho} = \hat{\lambda}$ so the fuzzy pure spatial autoregressive model (FPSAM) is :

$$y_{ci} = 0.0385 \sum_{j=1}^{38} W_{ij} y_{ci}, \qquad i = 1, \dots, 38$$

This model is also called spatial autoregressive model of first order and as it does not include any independent variable, that meaning it is based on the idea of explaining the phenomenon for itself by considering the independent variable as the same dependent variable y, but only with a spatial lag .

Table 4: centroid data and results error

	Location Cities	y_c	X_{1c}	X_{2c}	X_{3c}	$\hat{y}$	e
1	Mosul	52	6	1	68	8.4622	43.5378
2	AL-Hamdaniya	16	26	5	20	8.4622	7.5378
3	Telkaif	16	6	1	4	8.4622	7.5378
4	Sinjar	4	6	1	4	8.4622	- 4.4622
5	Tel afar	16	6	5	4	8.4622	7.5378
6	AL-Hatra	16	6	1	4	10.1546	5.8454
7	Maqmoor	16	6	1	4	8.4622	7.5378
8	Kirkuk	40	46	13	52	7.0775	32.9225
9	Daquq	28	26	21	20	7.0775	20.9225
10	Debes	28	6	13	4	7.0775	20.9225
11	Tikrit	16	66	1	52	28.9253	-12.9253
12	Tuz kurmato	4	6	5	4	28.9253	-24.9253
13	Samara	16	6	5	4	28.9253	-12.9253
14	Baled	4	6	1	4	28.9253	-24.9253
15	AL- Dor	4	6	1	4	28.9253	-24.9253
16	AL- shargat	4	6	5	4	28.9253	-24.9253
17	Baquba	88	186	13	132	11.2316	76.7684
18	AL- meqdadia	16	46	9	4	11.2316	4.7684
19	AL-Kalus	76	66	17	20	11.2316	64.7684
20	Kanaqeen	4	46	13	4	11.2316	-7.2316
21	Baladrouz	16	6	9	4	11.2316	4.7684
22	Al-Rasafa	64	106	17	132	12.4625	51.5375
23	AL-aadamia	16	46	13	52	12.4625	3.5375
24	AL-sader2	16	26	9	36	12.4625	3.5375
25	AL-sader1	4	26	5	52	12.4625	-8.4625
26	AL-Karek	40	126	17	116	12.4625	27.5375
27	AL-Kademia	4	66	9	68	12.4625	-8.4625
28	AL-Mahmoodia	40	26	5	20	12.4625	27.5375
29	Abu-griab	16	6	1	4	12.4625	3.5375
30	AL-Taremia	4	6	1	4	12.4625	-8.4625
31	AL-Madayn	40	6	1	20	12.4625	27.5375
32	AL-Rumadi	16	146	5	20	16.4628	-0.4628
33	Heet	4	6	1	4	16.4628	-12.4628
34	AL-Faloga	4	26	9	4	16.4628	-12.4628
35	Anah	4	6	1	4	16.4628	-12.4628
36	Haditha	4	6	5	4	16.4628	-12.4628
27	AL- Rutba	16	6	5	4	16.4628	-0.4628
38	AL-qaem	28	6	5	4	16.4628	11.5372

8 Recommendations

1. Using maximum likelihood method in estimating parameters and comparing them with the method of least squares for determine the best method for estimating parameters by Root Mean Squares Error or Mean Absolute Percentage Error .
2. Fuzzyfication the data into triangular fuzzy number and comparing it with the trapezoidal fuzzy number for showing a difference between them in the estimate.
3. There are other methods than the centroid method in dealing with fuzzy data such as alpha cat ($\alpha - cat$).
4. Study and application real data on other fuzzy spatial models.

References

[1] A.S. Ahmar and M.K. Aidid, *Crime Modeling using Spatial Regression Approach*, J. Phys.: Conf. Ser. **954** (2018), no. 1, 012013.

[2] Q.N.N. Al-Kazaz and H.J.K. Al-Saadi, *Comparing bayes estimation with maximum likelihood estimation of generalized inverted exponential distribution in case of fuzzy data*, J. Econ. Admin. Sci. **23** (2017), no. 101.

[3] G. Arbia, *A primer for spatial econometrics with applications in R*, Springer, 2014.

[4] B.O. Dalia and F.T. Amer, *Estimating parameters of some spatial regression models with experimental and applied study*, Department of Statistics, College of Administration and Economics, University of Kirkuk, 2020.

[5] M.M. Fischer and J. Wang, *Spatial data analysis: models, methods and techniques*, Springer Science & Business Media, 2011.

[6] M.E. Hassan and A.T. Heba, *Fuzzy regression analysis*, Iraqi J. Statist. Sci. **10** (2006), 61–84.

[7] C. Kao and C.L. Chyu, *A fuzzy linear regression model with better explanatory power*, Fuzzy Sets Syst. **126** (2002), no. 3, 401–409

[8] K.H. Lee, *First course on fuzzy theory and applications*, Springer Science & Business Media, 2004.

[9] T.J. Ross, *Fuzzy logic with engineering applications*, John Wiley & Sons, 2005.

[10] S.Q. Hadi and O.A.M. Ali, *Estimating spatial regression models for poverty rates in the districts of Iraq for the year 2012*, Master's Thesis in Statistics, College of Administration and Economics, University of Baghdad, 2014

[11] E. Tony Smith, *Spatial data analysis*, School of Engineering and Applied Science, University of Pennsylvania, 2015.

[12] Y.J. Wang and H.S. Lee, *The revised method of ranking fuzzy numbers with an area between the centroid and original points*, Comput. Math. Appl. **55** (2008), no. 9, 2033-2042

[13] Y.M. Wang, J.B. Yang, D.L. Xu and K.S. Chin, *On the centroids of fuzzy numbers*, Fuzzy Sets Syst. **157** (2006), no. 7, 919–926.

[14] A. Zhang, *Statistical analysis of fuzzy linear regression model based on centroid method*, Appl. Math. **7** (2016), no. 7, 579.