International Journal on Advanced Science Engineering Information Technology

Vol.15 (2025) No. 1
ISSN: 2088-5334

Comparing Some Adjacency Matrices to Estimate a Spatial Negative Binomial Regression Model

Basma Mohammad Lafta [a], Aseel Abdulrazzak Rasheed [a,*]

[a] *Collage of administration and Economics, Mustansiriyah University, Iraq*
Corresponding author: [*]aseelstat@uomustansiriyah.edu.iq

***Abstract*—The spatial regression model is used to illustrate the extent of the influence of independent variables on the dependent variable with the presence of spatial effects of adjacent locations. In such models, the dependent variable usually follows the normal distribution. In this research, a different case was studied, which is when the dependent variable is distributed in a negative binomial distribution, which is considered one of the important discrete distributions, and the basis of statistical models for count data. This distribution is suitable for data with overdispersion characteristics. In this research, a spatial negative binomial regression model is estimated using the maximum likelihood method of estimation, and based on the Queen adjacency criteria and the proposed longitudes to form the modified weight matrix, and simulation study is conducted to choose the best matrix from among the two used matrices. The results showed that the modified proposed longitude matrix is the best, as it was used to estimate the parameters of the negative binomial regression model using traffic accident data for 14 Iraqi governorates for the year 2022 as a response variable and based on the explanatory variables (temperature, rainfall, and amount of falling dust). The results showed that there is an effect of temperature and rainfall on the number of traffic accidents with spatial dependence.**

Keywords—Binary adjacency matrix; proposed longitude criterion; spatial negative binomial regression model.

Manuscript received 15 Sep. 2024; revised 29 Nov. 2024; accepted 5 Jan. 2025. Date of publication 28 Feb. 2025.

I. INTRODUCTION

Over the course of the last thirty to forty years, spatial models have been utilized in a broad variety of econometric applications. The modeling of spatial heterogeneity is the application that is utilized the most frequently in relation to the examination of census facts. In the first attempt to define spatial econometrics and its methodology, authors in [1] released a tiny volume with the title Spatial Econometrics. This volume was recognized as the first attempt. A number of publications on spatial analysis and spatial time series were published in the same year after they were first published. Inquiries and studies by [2], [3] include information that is associated with locations, spatial data is considered to be spatially oriented data. This data is distinct from other data in that it has two essential components, namely information denoting the place (spatial) and information that describes the attributes of the location. The geographic coordinates, such as latitude and longitude, are the subject of the information pertaining to location [4].

In contrast, descriptive information, often known as non-spatial information, includes facts like the density of the material and the sort of metal it is. There are distinctions in the circumstances of each of the places [5]. There is, however, a tight association between the state of each site and the condition of surrounding sites, and the correlation between the sites that were analyzed is referred to as spatial heterogeneity [6]. Among the several types of regression, one that is dependent on spatial variables is called spatial regression. When it comes to regression models. The link between the variables that are being explained and those that are being explained can be referred to as spatial dependency. Every piece of information that the researcher analyzes for any phenomena that he is interested in studying is not autonomous in and of itself; rather, it is reliant on the source of the information [7]. Studies in [8] and [9] aims to estimate a spatial negative binomial regression model (SAR), Using the maximum likelihood method, relying on the Quinn adjacency criterion, and a new adjacency criterion, relying on longitudes, and comparing them using the simulation method to reach the most appropriate matrix in obtaining an efficient estimate of the model, and then using it in the applied aspect in estimating the number of traffic accidents in the Iraqi governorates .

A. The spatial negative binomial regression model

An investigation of a spatial regression model is carried out with the purpose of elucidating the amount to which the independent variables have an impact on the dependent variable, taking into account the geographical impacts of neighboring sites. Researchers in the field of measurement have produced a number of spatial models that are concerned with spatial analysis processes, applying them to a wide variety of applications, and dealing with spatial dependency [10].

$$\mu^{SAR} = exp((I - \lambda\, w)^{-1} x\beta) \tag{1}$$

w is the spatial adjacency matrix with dimension (n×n).
λ is the spatial correlation parameter.
x is the matrix of variables.
β is a vector of parameters.

B. Binary contiguity matrix

Providing data allows the construction of a spatial weight matrix based on proximity. We assume that n refers to the number of spatial units. The spatial weight matrix with dimension n × n will be denoted by the symbol W, and it will be a positive, square, symmetric and non-random matrix. Each element within the matrix is denoted by W_{ij} at location i,j. A value is assigned to each pair at adjacent or non-adjacent locations by some predefined rules that define the spatial relationship between the locations. The general formula for the spatial weight matrix is as follows [11], [12].

$$W_{ij} = \{1 \quad if \quad i\ neighbour\ j \tag{2}$$

Through formula (2), the values of each element of the spatial weights' matrix are determined. Where if the sites i and j are adjacent, then the value of W_{ij}=1, and if they are not adjacent, then the value of W_{ij}=0. The elements of the main diagonal of the weight matrix are equal to zero because the regions are not adjacent to themselves [13], [14], [15].

C. Adjusted Weight Matrix

This matrix is based on the binary adjacency weights matrix W_ij after modifications have been made to it, in which the sum of the row equals one, as in the following formula [16], [17].

$$W_{ij\,adj} = \{\frac{\frac{W_{ij}}{\sum W_{ij}} \quad i\ neighbor\ j}{0 \quad other\,wise}\ 1 \tag{3}$$

There are several types of adjacencies to build the weight matrix, we will take the following:

1) Modified Queen Criterion (M-Q-C): This juxtaposition occurs by merging the Rook and Bishop matrix, i.e. the subscription becomes boundaries and points of the region. If the region shares a border or point, then the value of W_Q = 1, and if the region does not have a border or point with the neighboring region, then the value of W_Q = 0.

2) Contiguity criterion for proposed modified longitudes (C-C-P-M-L): This Contiguity is between two or more neighboring regions through which a common longitude passes, where the value of the region through which the longitude passes has a value of W_{ij}=1, and the common region through which the longitude does not pass has a value of W_{ij}=0.

D. The Maximum Likelihood Method for Estimating a Spatial Negative Binomial Regression Model

Maximum likelihood estimators are stable, highly efficient, and consistent. That is, the estimation process is done by making the estimates of the maximum likelihood function for the random variables as large as possible [18], [19].

$$f(y_i\,|\mu_i,\theta) = \frac{\Gamma(y_i+\theta^{-1})}{\Gamma(y_i+1)\Gamma(\theta^{-1})}\left(\frac{\theta^{-1}}{\mu_i^{SAR}+\theta^{-1}}\right)^{\alpha^{-1}}\left(\frac{\mu_i^{SAR}}{\mu_i^{SAR}+\theta^{-1}}\right)^{y_i} \tag{4}$$

Where $\theta \geq 0\,, i = 1,2,\ldots,n$

$$ln\ L(\theta\,,\beta) = \sum_{i=1}^{n}\left\{\left(\sum_{j=0}^{y-1} ln\,(j+\theta^{-1})\right) - ln\ (y_i\,!) - (y_i+\theta^{-1})ln\,(1+\theta\,\mu_i^{SAR}) + y_i\ ln\ \theta + y_i\ ln\ \mu_i^{SAR}\right\} \tag{5}$$

Since

$$\mu_i^{SAR} = exp\,[(I-\lambda\,w)^{-1}x\beta]$$

To obtain the maximum likelihood estimates for a spatial negative binomial regression model. Numerical estimation methods can be adopted, such as tRaphson method. Therefore, we must find the first and second derivatives of the model parameters [20], [21].

$$\frac{\partial\, Ln\, L}{\partial\lambda} = \sum_{i=1}^{n}\left\{-(y_i+\theta^{-1})\frac{\beta xw\ exp((I-\lambda w)^{-1}x\beta)}{\theta(I-\lambda w)^2\left(\frac{exp((I-\lambda w)^{-1}x\beta)}{\theta}+1\right)} + \frac{\beta xwy}{(I-\lambda w)^2}\right\} = 0 \tag{6}$$

Derivation with the ratio (β,θ) produces the following equations: [18].

$$\frac{\partial\, Ln\, L}{\partial\beta} = \sum_{i=1}^{n} x_i\frac{(y_i-\mu_i^{SAR})}{1+\theta\mu_i^{SAR}} = 0 \tag{7}$$

$$\frac{\partial\, Ln\, L}{\partial\theta} = \sum_{i=1}^{n}\left[\frac{1}{\alpha^2}\left(Ln(1+\theta\,\mu_i^{SAR}) + \frac{\theta(y_i-\mu_i^{SAR})}{1+\theta\,\mu_i^{SAR}}\right) + \psi\left(y_i+\frac{1}{\theta}\right) - \psi\left(\frac{1}{\theta}\right)\right] = 0 \tag{8}$$

E. Moran's coefficient test

It is a tool for measuring spatial dependence in the studied data, it is symbolized by the symbol I and is analogous to the Durbin Watson test in time series data. Its value ranges between (1+, 1-), as the closer the value of the Moran's coefficient is to (1+), the spread pattern of the data is close. However, if the value of the Moran's coefficient approaches (1-), the spread pattern of the data is divergent, but if the value approaches (0).) The spread of data is random. The formula for Moran's coefficient is [22], [23], [24].

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(y_i - \underline{y}\right)\left(y_j - \underline{y}\right)}{\sum_{i=1}^{n} \left(y_i - \underline{y}\right)^2} \quad (9)$$

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}$$

whereas:
S The sum of the elements in the matrix W.
n sample size.
W is a matrix of adjacency weights with dimensions n×n.

F. Mean Squares Error

It is calculated for the model for all matrices used. This criterion is one of the important comparison criteria in regression models. There is an inverse relationship between the MSE and the significance of the model, where the lower the MSE, the greater the significance of the model and vice versa. The formula for this standard [25], [26] is as follows:

$$MSE = \frac{1}{p} \sum_{i=1}^{p} \left(\widehat{B_i} - B\right)^2 \quad (10)$$

whereas:
p: represents the number of times the experiment was repeated
(B_i) represents the estimated value of parameter i
B represents the real parameter value

II. Materials and Method

A. Stages of the simulation experiment

The R 4.3.0 programming language was used to write the simulation program, which is attached in Appendix A. The written program includes four basic stages for estimating a negative binomial regression mod. At this stage, the default values for the parameters are chosen considering the presence of three explanatory variables [27][28][29]. Different default values were chosen for the parameters and these values are shown in Table 1.

TABLE I
DEFAULT VALUES FOR PARAMETERS

Model	β_0	β_1	β_2	β_3	θ	λ
1	0.1	0.1	0.1	0.1	1/2	0.2
2	2	0.07	-0.02	0.2	1/2	0.2
3	0.5	0.3	0.2	0.4	1/2	0.2
4	0.1	0.1	0.1	0.1	1/8	0.2
5	2	0.07	-0.02	0.2	1/8	0.2
6	0.5	0.3	0.2	0.4	1/8	0.2
7	0.1	0.1	0.1	0.1	1/2	0.8
8	2	0.07	-0.02	0.2	1/2	0.8
9	0.5	0.3	0.2	0.4	1/2	0.8
10	0.1	0.1	0.1	0.1	1/8	0.8
11	2	0.07	-0.02	0.2	1/8	0.8
12	0.5	0.3	0.2	0.4	1/8	0.8

B. Generating data

At this stage, the explanatory variables are generated as being generated from a uniform distribution [30]. The error is generated from a negative binomial distribution, and then the dependent variable is collected according to the approved model. Each experiment was repeated 5000 times. Three different sample sizes were chosen (56, 112, 168).

III. Results and Discussion

A. Contiguity criterion for proposed modified longitudes (C-C-P-M-L) Modified Queen Criterion (M-Q-C)

The MSE criteria was utilized for the aim of evaluating several weight matrices for parameters and locating the optimal weight matrix. This is because the matrix that has the lowest MSE value is deemed to be superior. As shown in Tables 2,3 ,4 ,5,6,7,8,9,10,11,12 and 13.

TABLE II
ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE FIRST CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	**C-C-P-M-L**	0.09669	0.09961	0.09954	0.10011	0.46839	0.20474	0.14247
	M-Q-C	0.06654	0.10047	0.10040	0.09980	0.46854	0.20644	0.14342
112	**C-C-P-M-L**	0.10745	0.09951	0.09976	0.09969	0.48586	0.20004	0.06598
	M-Q-C	0.06932	0.10060	0.09989	0.10052	0.48249	0.20002	0.06793
168	**C-C-P-M-L**	0.13152	0.09841	0.10012	0.10023	0.48943	0.19992	0.03846
	M-Q-C	0.08893	0.10008	0.10026	0.09992	0.48983	0.19987	0.04290

TABLE III
ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE SECOND CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	**C-C-P-M-L**	1.92344	0.07168	-0.01903	0.20049	0.46678	0.20280	0.13347
	M-Q-C	1.98511	0.06968	-0.01958	0.19985	0.46631	0.19877	0.14436
112	**C-C-P-M-L**	1.96347	0.07134	-0.02041	0.19970	0.48607	0.19989	0.06361
	M-Q-C	2.02933	0.06862	-0.02025	0.20004	0.48564	0.19922	0.06393
168	**C-C-P-M-L**	2.02786	0.06871	-0.01988	0.19993	0.49027	0.20052	0.04235
	M-Q-C	1.99045	0.07005	-0.01985	0.20007	0.48930	0.20010	0.04031

TABLE IV

ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE THIRD CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	C-C-P-M-L	0.48286	0.30021	0.19963	0.39964	0.47002	0.20211	0.13128
	M-Q-C	0.51903	0.29909	0.20016	0.39878	0.46518	0.20301	0.14935
112	C-C-P-M-L	0.51721	0.29854	0.20058	0.40021	0.48490	0.20007	0.06304
	M-Q-C	0.47991	0.30036	0.20020	0.39972	0.48316	0.20015	0.06576
168	C-C-P-M-L	0.50526	0.29944	0.19986	0.40032	0.48766	0.20012	0.03999
	M-Q-C	0.50360	0.29983	0.19983	0.39946	0.49129	0.20086	0.04237

TABLE V

ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE FOURTH CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	C-C-P-M-L	0.09517	0.10008	0.10007	0.09972	0.11493	0.20102	0.03508
	M-Q-C	0.11805	0.09908	0.09979	0.10012	0.11441	0.19945	0.03841
112	C-C-P-M-L	0.10842	0.09953	0.09982	0.10038	0.12059	0.19901	0.01724
	M-Q-C	0.09406	0.10015	0.10000	0.09994	0.12104	0.19827	0.01983
168	C-C-P-M-L	0.09752	0.10003	0.10024	0.09989	0.12086	0.20053	0.01106
	M-Q-C	0.09623	0.10010	0.10008	0.10002	0.12222	0.20023	0.01235

TABLE VI

ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE FIFTH CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	C-C-P-M-L	2.00596	0.06955	-0.02007	0.20009	0.11589	0.19783	0.03411
	M-Q-C	1.99248	0.07015	-0.02007	0.19991	0.11789	0.19750	0.03539
112	C-C-P-M-L	1.98632	0.07043	-0.02021	0.20016	0.12041	0.19962	0.01657
	M-Q-C	2.01169	0.06963	-0.02032	0.19986	0.12139	0.20025	0.01810
168	C-C-P-M-L	1.99459	0.07024	-0.01991	0.19978	0.12216	0.19982	0.01062
	M-Q-C	1.98829	0.07036	-0.02005	0.20005	0.12199	0.20149	0.01179

TABLE VII

ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE SIXTH CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	C-C-P-M-L	0.48414	0.30038	0.19983	0.40027	0.11638	0.20080	0.03047
	M-Q-C	0.48135	0.30041	0.20004	0.40034	0.11522	0.20351	0.03342
112	C-C-P-M-L	0.50928	0.29939	0.20020	0.40007	0.12116	0.20014	0.01463
	M-Q-C	0.48501	0.30048	0.19989	0.40016	0.12079	0.20064	0.01505
168	C-C-P-M-L	0.48725	0.30040	0.20007	0.40009	0.12122	0.20017	0.01038
	M-Q-C	0.50116	0.29992	0.20000	0.39989	0.12191	0.19993	0.01061

TABLE VIII

ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE SEVENTH CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	C-C-P-M-L	0.11476	0.09850	0.10014	0.10036	0.46630	0.80077	0.13060
	M-Q-C	0.05070	0.10082	0.10079	0.09999	0.46726	0.80164	0.15005
112	C-C-P-M-L	0.10801	0.09895	0.10039	0.10058	0.48237	0.80047	0.06430
	M-Q-C	0.06484	0.10094	0.10014	0.10019	0.48334	0.79963	0.06839
168	C-C-P-M-L	0.08800	0.10036	0.09973	0.10021	0.49019	0.79962	0.04180
	M-Q-C	0.10254	0.09966	0.09968	0.10030	0.48880	0.80114	0.04413

TABLE IX

ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE FIFTH CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	C-C-P-M-L	2.03958	0.06750	-0.02037	0.20072	0.47238	0.80153	0.13667
	M-Q-C	2.02322	0.06855	-0.02068	0.20000	0.46885	0.79856	0.14178
112	C-C-P-M-L	2.03827	0.06822	-0.01992	0.19960	0.48476	0.79857	0.06272
	M-Q-C	1.96796	0.07029	-0.01944	0.20073	0.48648	0.79932	0.07059
168	C-C-P-M-L	1.95004	0.07169	-0.01990	0.19978	0.49172	0.79976	0.04114
	M-Q-C	2.01001	0.06961	-0.02003	0.19950	0.49104	0.80075	0.04478

TABLE X

ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE NINTH CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	C-C-P-M-L	0.43081	0.30189	0.19995	0.40043	0.47015	0.79623	0.13330
	M-Q-C	0.50463	0.29903	0.20067	0.39960	0.46771	0.79498	0.13771
112	C-C-P-M-L	0.49181	0.30027	0.19959	0.39963	0.48266	0.79872	0.06278
	M-Q-C	0.47587	0.30063	0.19951	0.40015	0.48305	0.80057	0.07026
168	C-C-P-M-L	0.50086	0.29975	0.19995	0.39994	0.48919	0.80098	0.04018
	M-Q-C	0.49312	0.30009	0.20023	0.39977	0.48725	0.79963	0.04487

TABLE XI

ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE TENTH CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	**C-C-P-M-L**	0.09174	0.09990	0.10049	0.10015	0.11521	0.79954	0.03522
	M-Q-C	0.10529	0.09985	0.09943	0.10022	0.11646	0.79984	0.03535
112	**C-C-P-M-L**	0.09534	0.10001	0.10009	0.10014	0.12016	0.79980	0.01647
	M-Q-C	0.10867	0.09961	0.09991	0.10011	0.11950	0.79886	0.01890
168	**C-C-P-M-L**	0.09777	0.10022	0.10009	0.10044	0.12182	0.79912	0.01159
	M-Q-C	0.09322	0.10012	0.10005	0.10013	0.12198	0.80013	0.01177

TABLE XII

ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE ELEVENTH CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	**C-C-P-M-L**	2.00991	0.06959	-0.02019	0.19990	0.11573	0.79825	0.03380
	M-Q-C	1.97368	0.07064	-0.01981	0.20035	0.11656	0.79931	0.03612
112	**C-C-P-M-L**	1.99964	0.07007	-0.02020	0.19979	0.12143	0.80027	0.01607
	M-Q-C	1.99641	0.07001	-0.01989	0.20007	0.12039	0.79807	0.01768
168	**C-C-P-M-L**	1.99669	0.06999	-0.01992	0.20018	0.12232	0.80039	0.01105
	M-Q-C	1.99708	0.07003	-0.01987	0.19984	0.12288	0.79999	0.01154

TABLE XIII

ESTIMATION VALUES OF PARAMETERS AND MSE OF MATRICES FOR THE FIFTH CASE

n	Matrix	β0	β1	β2	β3	θ	λ	MSE
56	**C-C-P-M-L**	0.50365	0.29962	0.20000	0.39990	0.11610	0.79879	0.03005
	M-Q-C	0.52298	0.29888	0.19968	0.40036	0.11531	0.79842	0.03459
112	**C-C-P-M-L**	0.51507	0.29928	0.19994	0.40026	0.12111	0.80022	0.01476
	M-Q-C	0.50282	0.29993	0.20002	0.39988	0.12176	0.79940	0.01548
168	**C-C-P-M-L**	0.49362	0.30027	0.20004	0.39972	0.12188	0.79924	0.01006
	M-Q-C	0.49073	0.30036	0.19988	0.40001	0.12211	0.79981	0.01029

From the Tables, we conclude that the best matrix for estimating the spatial negative binomial regression model is the proposed modified longitude matrix, and therefore we will rely on it in the applied aspect.

B. Good Matching of Data

The relationship between climate and transportation is ancient and close since the beginning of the process of trade exchange on a large scale between human societies. The climate was a contributing factor to this trade exchange, especially in the period when humans used sailing ships in human transportation and trade. We will study land transport, as it is also affected by climatic events, and to a lesser extent. Weather events may uproot railways, as can heavy rains, and disruption of traffic and traffic signals as a result of fog or as a result of freezing, although they have a significant impact on traffic safety.

In order to ensure that the data of the dependent variable (the number of traffic accidents) is subject to a negative binomial distribution, goodness-of-fit tests provided by the ready-made statistical program Easy Fit 5.6 were used. These tests are the Kolmogorov-Smirnov (KS) Kolmogorov-Smirnov test and the Anderson-Darling (AD) test. These tests test the following hypothesis:

H_0: y~Negative Binomial

H_1: y≁Negative Binomial

It is clear from Table 6 that the probability value of the Kolmokrov-Smirnov (KS) test is (0.17823) greater than the significance level (0.05), which indicates acceptance of the null hypothesis that the data follows a negative binomial distribution, and this is supported by the Anderson-Darlink (AD) test. The probability value of this test is equal to (2.5018), which is greater than the significance level (0.05). The results of the two tests were as in Table 14.

TABLE XIV

GOODNESS-OF-FIT TESTS FOR THE VARIABLE NUMBER OF TRAFFIC ACCIDENTS

	KS	AD
P- Value	0.17823	2.5018

C. Moran's spatial dependence test

The results of Moran's test for the spatial dependence of the proposed length matrix were obtained, and the results are as shown in Table 15:

TABLE XV

MORAN TEST RESULTS

Moran I	Expected	Variance	Z-Statistic	P-Value
0.18	-0.018	0.0005	9.02	<0.0001

The results indicate that there is spatial dependence based on the P-Value, which is less than the significance level of 0.05.

D. Model estimation

The maximum likelihood method was used to estimate the parameters of the spatial binomial regression model based on the proposed longitudinal weight matrix because it gave very satisfactory results according to the simulation results as shown in Table 16

TABLE XVI

PARAMETER ESTIMATES

Parameters	β_0	β_1	β_2	β_3	θ	Λ
Estimates	-1.89131	0.09610	0.01544	-0.00148	2.60530	0.87472

Based on the results on the Figure 1, the observations of the dependent variable were estimated, and the following Figure shows a graph of the estimated values with the actual values of the number of traffic accidents in Iraq for the year 2022.

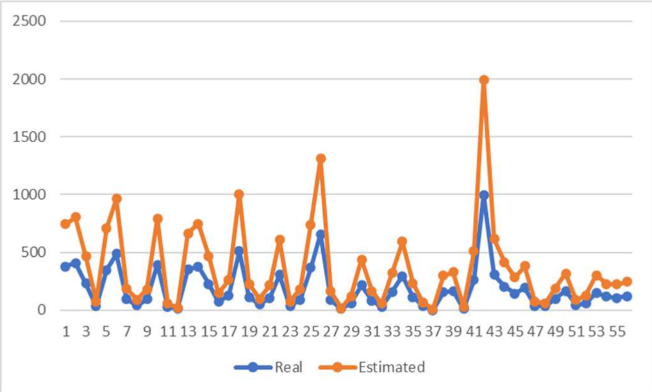

Fig. 1 Currently, estimated values of the dependent variable

IV. Conclusion

The spatial negative binomial regression model proved its efficiency in light of the two adjacent matrices adopted in the research with large sample sizes. This is evident from the decrease in the mean square error values, at a sample size of 168, where the weight matrix for the modified longitudes showed its efficiency when estimating the spatial negative binomial regression model. This is due to the low values of the mean square error (MSE).In the applied side and based on the results of Moran's coefficient, it was concluded that the count data represented by the number of traffic accidents in the Iraqi governorates suffers from spatial dependency, and notice the correlation of the dependent variable. The number of traffic accidents has a positive, significant relationship with the two explanatory variables: the general average temperature and the average rainfall. That is, as temperatures increase, traffic accidents increase with them. The greater the amount of rain falling, the greater the number of traffic accidents depending on the type of road, the third explanatory variable, the amount of falling dust, is inversely related to the number of traffic accidents. We recommend Using a spatial negative binomial regression model in the economic and health fields

Acknowledgment

No grant did not fund this research.

References

[1] O. J. Adeyemi, R. Paul, and A. Arif, "An assessment of the rural-urban differences in the crash response time and county-level crash fatalities in the United States," *J. Rural Health*, vol. 38, no. 4, pp. 999–1010, Oct. 2021, doi: 10.1111/jrh.12627.

[2] T. E. Adeyeye et al., "Estimating policy-relevant health effects of ambient heat exposures using spatially contiguous reanalysis data," *Environ. Health*, vol. 18, no. 1, Apr. 2019, doi: 10.1186/s12940-019-0467-5.

[3] P. Ashmore et al., "Spatiotemporal and socioeconomic risk factors for dengue at the province level in Vietnam, 2013–2015: Clustering analysis and regression model," *Trop. Med. Infect. Dis.*, vol. 5, no. 2, p. 81, May 2020, doi: 10.3390/tropicalmed5020081.

[4] F. Cappelli et al., "Climate change and armed conflicts in Africa: Temporal persistence, non-linear climate impact and geographical spillovers," *Econ. Polit.*, vol. 40, no. 2, pp. 517–560, Jun. 2022, doi: 10.1007/s40888-022-00271-x.

[5] S. de F. Marques and C. S. Pitombo, "Local modeling as a solution to the lack of stop-level ridership data," *J. Transp. Geogr.*, vol. 112, p. 103682, Oct. 2023, doi: 10.1016/j.jtrangeo.2023.103682.

[6] J. M. Hilbe, Negative Binomial Regression, Mar. 2011, doi:10.1017/cbo9780511973420.

[7] J. L. Gittleman and M. Kot, "Adaptation: Statistics and a null model for estimating phylogenetic effects," *Syst. Zool.*, vol. 39, no. 3, p. 227, Sep. 1990, doi: 10.2307/2992183.

[8] L. Anselin and A. Bera, "Spatial dependence in linear regression models with an introduction to spatial econometrics," in *Handbook of Applied Economic Statistics*, A. Ullah and D. Giles, Eds. New York, NY, USA: Marcel Dekker, 1998, pp. 237–289, doi: 10.1201/9781482269901-36.

[9] L. Anselin, "Thirty years of spatial econometrics," *Pap. Reg. Sci.*, vol. 89, no. 1, pp. 3–26, Mar. 2010, doi: 10.1111/j.1435-5957.2010.00279.x.

[10] L. Anselin, *Spatial Econometrics: Methods and Models*. Springer Netherlands, 1988, doi: 10.1007/978-94-015-7799-1.

[11] M. Leiva, F. Vasquez-Lavín, and R. D. Ponce Oliva, "Do immigrants increase crime? Spatial analysis in a middle-income country," *World Dev.*, vol. 126, p. 104728, Feb. 2020, doi: 10.1016/j.worlddev.2019.104728.

[12] T. Lokhande and Y. Xie, "Spatial spillover effects of urban decline in Southeast Michigan," *Appl. Geogr.*, vol. 158, p. 103031, Sep. 2023, doi: 10.1016/j.apgeog.2023.103031.

[13] M. R. Abonazel and O. Saber, "A comparative study of robust estimators for Poisson regression model with outliers," *J. Stat. Appl. Probab.*, vol. 9, pp. 279–286, 2020.

[14] A. Maiti et al., "Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States," *Sustain. Cities Soc.*, vol. 68, p. 102784, May 2021, doi: 10.1016/j.scs.2021.102784.

[15] M. Morales-Otero and V. Núñez-Antón, "Comparing Bayesian spatial conditional overdispersion and the Besag–York–Mollié models: Application to infant mortality rates," *Mathematics*, vol. 9, no. 3, p. 282, Jan. 2021, doi: 10.3390/math9030282.

[16] M. Mukhsar et al., "Dynamical spatial model of heavy metals in Kendari Bay using Bayesian geographical weighted regression," *J. Phys.: Conf. Ser.*, vol. 1899, no. 1, p. 012107, May 2021, doi: 10.1088/1742-6596/1899/1/012107.

[17] A. Noda, T. Yamanouchi, K. Kobayashi, and J. Nishihiro, "Temporal continuity and adjacent land use exert different effects on richness of grassland specialists and alien plants in semi-natural grassland," *Appl. Veg. Sci.*, vol. 25, no. 3, Jul. 2022, doi: 10.1111/avsc.12682.

[18] M. Yang, Y. Huang, Y. Guo, W. Zhang, and B. Wang, "Ultra-short-term wind farm cluster power prediction based on FC-GCN and trend-aware switching mechanism," *Energy*, vol. 290, p. 130238, Mar. 2024, doi: 10.1016/j.energy.2024.130238.

[19] M. Kim and J. Song, "Seismic damage identification by graph convolutional autoencoder using adjacency matrix based on structural modes," *Earthq. Eng. Struct. Dyn.*, vol. 53, no. 2, pp. 815–837, Nov. 2023, doi: 10.1002/eqe.4047.

[20] Y. Yang, J. Guo, Y. Li, and J. Zhou, "Forecasting day-ahead electricity prices with spatial dependence," *Int. J. Forecast.*, vol. 40, no. 3, pp. 1255–1270, Jul. 2024, doi: 10.1016/j.ijforecast.2023.11.006.

[21] Z. Li, Y. Wei, G. Chen, K. Lu, and X. Zheng, "Learning dynamics of multi-level spatiotemporal graph data for traffic flow prediction," *Comput. Commun.*, vol. 223, pp. 26–35, Jul. 2024, doi: 10.1016/j.comcom.2024.05.007.

[22] F. T. Al-Maliky and J. S. Chiad, "Study and analysis the flexion moment in active and passive knee prosthesis using back propagation neural network predictive," *J. Braz. Soc. Mech. Sci. Eng.*, vol. 44, no. 11, Oct. 2022, doi: 10.1007/s40430-022-03850-y.

[23] J. Gan, Q. Su, L. Li, Y. Ju, and L. Li, "Urban traffic accident frequency modeling: An improved spatial matrix construction method," *J. Adv. Transp.*, vol. 2025, no. 1, Jan. 2025, doi: 10.1155/atr/1923889.

[24] M. M. Yahya, A. M. Al-Mushehdany, and H. J. M. Alalkawi, "Evaluation of buckling of 2024-T3 under high temperatures," *Int. J. Heat Technol.*, vol. 40, no. 4, pp. 947–952, Aug. 2022, doi: 10.18280/ijht.400411.

[25] G. S. Gill, W. Cheng, M. Singh, and Y. Li, "Comparative evaluation of alternative Bayesian semi-parametric spatial crash frequency models," *J. Traffic Transp. Eng. (Engl. Ed.)*, vol. 12, no. 1, pp. 151–166, Feb. 2025, doi: 10.1016/j.jtte.2022.01.005.

[26] A. Al-Mushehdany, M. M. Yahya, E. K. Ibrahim, and H. J. M. Alalkawi, "Nano reinforcement technique as a tool for enhancement the mechanical and fatigue properties," *Curved Layer. Struct.*, vol. 9, no. 1, pp. 345–351, Jan. 2022, doi: 10.1515/cls-2022-0026.

[27] I. O. B. Al-Fahad, A. D. Hassan, B. M. Faisal, and H. Kadhim Sharaf, "Identification of regularities in the behavior of a glass fiber-reinforced polyester composite of the impact test based on ASTM

D256 standard," *East.-Eur. J. Enterp. Technol.*, vol. 4, no. 7 (124), pp. 63–71, Aug. 2023, doi: 10.15587/1729-4061.2023.286541.

[28] I. O. B. Al-Fahad, H. Kadhim Sharaf, L. N. Bachache, and N. K. Bachache, "Identifying the mechanism of the fatigue behavior of the composite shaft subjected to variable load," *East.-Eur. J. Enterp. Technol.*, vol. 3, no. 7 (123), pp. 37–44, Jun. 2023, doi: 10.15587/1729-4061.2023.283078.

[29] K. H. Salman, A. H. Elsheikh, M. Ashham, M. K. A. Ali, M. Rashad, and Z. Haiou, "Effect of cutting parameters on surface residual stresses in dry turning of AISI 1035 alloy," *J. Braz. Soc. Mech. Sci. Eng.*, vol. 41, no. 8, Aug. 2019, doi: 10.1007/s40430-019-1846-0.

[30] B. M. Faisal et al., "Finite element analysis for aluminium alloy 7075-T6 subjected to the bending-torsional variable phase of fatigue loading," *J. Adv. Res. Appl. Mech.*, vol. 128, no. 1, pp. 62–71, Nov. 2024, doi: 10.37934/aram.128.1.6271.