Estimation of the general spatial regression model (SAC) by the maximum likelihood method

Wadhah S. Ibrahim* a, Nawras Shanshool Mousa a

a College of Management and Economics, Dept. Of Statistic, Mustansiriya University, Baghdad, Iraq.

(Communicated by Madjid Eshaghi Gordji)

Abstract

That there are indicators or statistical transactions that have appeared in a large way in recent times to describe, summarize and analyse spatial data, when a study is done of many phenomena or a disease is studied, whether it is on humans or animals, we need to analyze the spatial data resulting from those phenomena, as it includes observations of the spatial units. For example, countries or provinces ... etc., all of these are linked to certain points or locations. The study uses the maximum likelihood method to estimate the parameters of the General Spatial Model by employing the model to study cancer which shows the relationship between the dependent variable Y represented by the number of patients and the explanatory variables represented (average age, tumor size, treatment, hormone, immunity) in light of the effect of spatial juxtaposition and using Rook neighboring criteria. One of the most important conclusions reached is the emergence of significant effects of some explanatory variables on the dependent variable Y, and the estimated values of the dependent variable Y are close to the real values of the same variable.

Keywords: The general spatial regression model, Spatial contiguity matrix, Rook neighboring criteria, Maximum Likelihood Method, Cancer.

1. Introduction

Spatial regression is methods for capturing the spatial dependence, and the spatial dependence of the regression model can be entered as relationships between the independent variables and the dependent variables [9]. As spatial regression depends largely on the data obtained by the researcher, any data that the researcher collects for any phenomenon he wants to study is not independent in itself, but rather depends on the place from which the data was taken. The spatial data is
distinguished from the time series data by the spatial arrangement of the observations, and spatial econometrics deals with spatial dependency and spatial heterogeneity. These characteristics may make traditional econometric techniques become unsuitable, and spatial economic measurement is concerned with following up on spatial effects such as the spatial dependence of observations in points different from the place.

The problem here is that the general linear regression model does not estimate spatial contiguity in its calculations, and this may lead to the loss of important data about the studied phenomenon, which ultimately affects the statistical results. Cancer in particular and its spread among neighboring areas varies from one area to another. This disease is considered one of the diseases that pose a danger to humans, as in many cases the health status of the neighboring areas is analyzed without following a correct scientific approach that takes into account the appropriate spatial or spatial effects to consider consideration.

2. Objective of the research

The aim of this research is to estimate the General Spatial Model, which suffers from the problem of spatial dependence, using the Maximum Likelihood Method, which describes the relationship between the dependent and represented variable (number of infected) and explanatory variables (average age, tumor size, treatment, hormone, immunity) for cancer under the Rook spatial contiguity criterion for the regular and modified spatial neighborhoods matrices ($W_{ij}$, $W_{ij}^{Adj}$).

3. Matrices of criterion and spatial response

The Rook contiguity criterion, one of the criteria for spatial contiguity, was used in the formation of the spatial contiguity matrix, as a method for this contiguity where a single value is taken if the two adjacent regions are finite and have a relationship between two regions on any side, meaning that ($W_{R=1}$), otherwise, the value will be zero ($W_{R=0}$) since the main diagonal elements of the matrix are zero because the point does not adjoin itself and that this contiguity has more than one point in one row of the matrix $W_{R}$, and the use of this matrix is more than others.[2].

Ten regions will be developed to determine the spatial contiguity matrix, as follows: A: represents the Kadhimiya area, B: represents the Adhamiya area, C: represents the city center, D: represents Palestine, E: represents the center of Rusafa, F: represents the new Baghdad, G: represents Eastern Karada, H: represents the safe, I: represent Al-Mansour, and J: represents the Karkh center.

Show the common boundaries between cell (A) and cells (B,C) and also between cell (B) and cells (A,D) through the matrix is as in the following formula:

$$W_{R} = \begin{bmatrix}
A & B & C & D & E & F & G & H & I & J \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{bmatrix}$$
It is noticeable that the Rook adjacency matrix shown in formula (3.1) can be formed as follows:

The point (A) is not adjacent to itself, so it takes the value ($W_{R11}= 0$), If the point (A) is adjacent to the point (B) it takes the value ($W_{R12} = 1$), If point (A) is adjacent to point (C) it takes the value ($W_{R13} = 1$), If point (A) is not adjacent to point (D) it takes the value ($W_{R14} = 0$), etc...

The modified spatial adjacency matrix can be found which is denoted by the symbol $W_{ij}^{Adj}$.

It is calculated by the following formula.

$$W_{ij}^{Adj} \begin{cases} \frac{W_{ij}}{\sum W_{ij}} & i \text{ neighbor } j \\ 0 & \text{other wise} \end{cases}$$

That is, each value of any row of the ordinary spatial adjacency matrix $W_{ij}$ divide by the total row as shown in the matrix below:

$$W_{ij}^{Adj} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

4. Study model

The general spatial regression model consists of two parts: the spatial lag and a spatially linked error structure. (SAC) represents an appropriate approach to modeling this type of dependence in error and it is explained as in the following:

$$Y = \rho W_1 Y + X \beta + u \quad , \quad u = \lambda W_2 u + \varepsilon \quad , \quad \varepsilon \sim N(0, \sigma^2 \varepsilon) \tag{4.1}$$

Since: $Y$: it represents a vector ($n*1$) of the dependent variable. $X$: represents an array ($n*k$) of explanatory variables. $W_1$, $W_2$: they represent spatial weight matrices with dimensions ($n*n$), where($W_1$) represents the spatial adjacency matrix between the views and the contiguous regions and ($W_2$) it represents the contiguity between the views and the city center, usually the proximity relationship or the distance function, which are fixed and predetermined. They can be equal $W_1=W_2$ what do you represent $W_1$ [4]. $\rho$: represents a parameter of spatial dependence. $\beta$: it represents a vector of parameters ($k*1$) that are associated with the matrix of explanatory variables $X$. $\lambda$: the spatial autoregressive parameter of errors is the spatial lag coefficient of the error and $u$: spatially related errors.[4].

5. The maximum likelihood method

Estimation of spatial regression models is usually performed by estimating the greatest possible, where the probability of the joint distribution (possibility) of all observations is maximized with respect to the number of relevant parameters. It is strong in small deviations from the assumption of normality, and it is also considered one of the most important methods because it gives the best estimate of the parameter among several possible estimates.[7].
To extract the estimation equations, it is explained as follows:

\[ \varepsilon = u - \lambda Wu \]  \hfill (5.1)

\[ u = (I - \lambda W)^{-1} \varepsilon \]  \hfill (5.2)

\[ Y(1 - \rho W) = (I - \lambda W)^{-1} \varepsilon \]  \hfill (5.3)

\[ \varepsilon^T \varepsilon = [(I - \lambda W) Y(I - \rho W) - (I - \lambda W) X\beta]^T [(I - \lambda W) Y(I - \rho W) - (I - \lambda W) X\beta] \]  \hfill (5.4)

\[ L(\beta, \rho, \lambda, \sigma^2/Y, X) = -n^2 \ln(2\pi) - n^2 \ln \sigma^2 + \ln |I - \rho W| + \ln |I - \lambda W| - \frac{1}{2\sigma^2} \varepsilon^T \varepsilon \]  \hfill (5.5)

\[ \frac{\partial (\beta, \rho, \lambda, \sigma^2/Y, X)}{\partial \beta} = -\frac{1}{2\sigma^2} \left[ -2X'(I - \lambda W) Y(I - \rho W) + X'(I - \lambda W)'(I - \lambda W) X\hat{\beta}_{mle} \right] \]  \hfill (5.6)

\[ \hat{\beta}_{mle} = \left[X'A'AX\right]^{-1}[X'\hat{\gamma}A,YB], \text{ and } A = (I - \lambda W) \]  \hfill (5.7)

To estimate the value of the correlation parameter (\( \rho \)) they are found using iterative methods for the probability function, as follows:

\[ |I - \rho W| = \prod_{i=1}^{n} (1 - \rho w_i) \]  \hfill (5.8)

\[ \ln |I - \rho W| = \sum_{i=1}^{n} \ln (1 - \rho w_i) \]  \hfill (5.9)

As well as to estimate the value of the parameter (\( \lambda \)) they are found using iterative methods as follows:

\[ |I - \lambda W| = \prod_{i=1}^{n} (1 - \lambda w_i) \]  \hfill (5.10)

\[ \ln |I - \lambda W| = \sum_{i=1}^{n} \ln (1 - \lambda w_i) \]  \hfill (5.10)

\[ s^2 = \frac{\varepsilon^T \varepsilon}{n} \]  \hfill (5.11)

6. Moran’s Test

This test is a measure to show whether there is a spatial dependency in the data or not, and it is a general measure and depends on a model (GLM) \[ Y = X \beta + \varepsilon \]. Or how one of the observations is similar to the other observations surrounding it in each region and neighboring regions through the matrix of spatial weight [\( S \)], the idea in Moran’s test depends on that the close things have more relationship than the distant things of any phenomenon related to each other, if the value of Moran’s coefficient close to (3.1) means that there is a spatial autocorrelation [9]. Moran’s formula is as follows:

\[ I_m = n \left( \varepsilon' w \varepsilon \right) / S_0 \left( \varepsilon' \varepsilon \right) \]  \hfill (6.1)

Since: \( S_0 \): the sum of each element in the W matrix. W: square weight matrix (n*n). n : sample volume. \( \varepsilon \): the error vector (residuals) has dimensions (n*1).
When a standard row is used and the row sum is equal to (3.1) in this case \( n = S_0 \) it simplifies the above formula as follows:

\[
I_m = \varepsilon'w\varepsilon / \varepsilon'\varepsilon
\]

(6.2)

To find out if the value of Moran’s coefficient is \( I_m \) statistically significant at a certain degree of confidence, the Moran \((Z)\) test is used:

\[
Z = (I - E(I)) / \sqrt{\text{var}(I)}
\]

(6.3)

\[
E(I) = \text{tr}(MW) / (n - k)
\]

(6.4)

\[
\text{var}(I) = \text{tr}(MWMW' + MWMW) + (\text{tr}(MW))^2 / (n-k) (n-k+2) - (E(I))^2
\]

(6.5)

Since: \( M = I - X (X'X) X' \): A deaf matrix is square and symmetric. \( \text{tr}(MWMW') \): the sum of the diagonal elements of the matrix. \( k \): The number of explanatory variables.

7. Lagrange Multiplier Test

The multiplexed Lagrange test is more used than Moran’s test because Moran is used only to test the spatial dependency whether it exists or not, and it is not possible to test what is the alternative model of the GLM model by Moran’s test, while the Lagrange test gives what is the alternative model \((\rho \) or \( \lambda \)). This approach indicates that if \( \rho \) is important or \( \lambda \) or both, [9].

7.1. Lagrange Multiplier for \( (\rho) \)

\( H_0 : \rho = 0 \) Spatial dependence exists

\( H_1 : \rho \neq 0 \) At least one of \( \rho \) is not equal to 0 i.e. there is no spatial dependence

The test format is as follows:

\[
\text{LM}_\rho = \frac{(\varepsilon'WXb)^2}{D}
\]

(7.1)

\[
D = \frac{(WXb)'MA(WXb)}{S^2} + \text{tr} \left( W'W + WW' \right)
\]

(7.2)

\( S^2 \) : is the error variance of the general linear regression model.

We compare the calculated value with the tabular value of \( \chi^2 \) \((1, \alpha)\) and then hypotheses are determined.

7.2. Lagrange Multiplier for \((\lambda)\)

\( H_0 : \lambda = 0 \) The spatial dependence is in the wrong

\( H_1 : \lambda \neq 0 \) At least one of \( \lambda \) is not equal to zero, the spatial dependence does not exist by mistake

Where rejecting the null hypothesis and accepting the alternative hypothesis means that the spatial dependence exists and the alternative model is \((\lambda)\)

\[
\text{LM}_\lambda = \frac{(\varepsilon'wz)^2}{T}
\]

(7.3)

\[
T = \text{tr} \left( [w + w'] w \right)
\]

(7.4)

To compare \((\text{LM}_\rho, \text{LM}_\lambda)\) with a tabular value of \( \chi^2 \) \((1, \alpha)\), where the Lagrange test \((\rho \) or \( \lambda \)) for spatial dependence in each of them needs a strong test and a strong role for the \((\rho \) and \( \lambda \)) model.
8. The practical side

Despite the performance of the health side and the hospitals related to it and its support in increasing the medical requirements of doctors, nurses and other cadres, but there is an important matter, which is how to cut and dissipate diseases between areas that have a direct impact on human life, as well as the impact of neighboring areas on the spread of diseases, in this research was Focusing on the spatial aspect, i.e. spatial juxtapositions to know its impact on the spread of diseases between regions, knowing the distribution of disease incidence and building a spatial model for predicting cancer, as this aspect included estimating the general spatial regression model using the regular and modified spatial juxtapositions $W_{ij}$, $W_{ij}^{Adj}$. In light of the spatial juxtaposition criterion Rook, and using the spatial data collected by taking a random sample and according to the geographical areas of the two sides of Karkh and Rusafa in Baghdad governorate from hospitals affiliated with the Baghdad Health Department / Ministry of Health, as they were collected from records and drums belonging to each patient, as well as using the Department of The Central Statistical Organization of the Ministry of Planning in designing the map that includes the Karkh and Rusafa areas according to their administrative division. The data collected was characterized by the following variables:

- Cancer disease was used, which shows the relationship between the dependent variable $Y$, represented by the number of patients, and (5.1) of the explanatory variables were used with their levels after agreement on them with the specialized doctors, as follows:
  - $Y$: represents the number of injured;
  - $X_1$: represents the average age number;
  - $X_2$: represents the Tumor size;
  - $X_3$: represents the hormone;
  - $X_4$: represents the immunity;
  - $X_5$: represents the treatment.

The data in the above was obtained through a questionnaire for people with cancer, which included all the ten regions of the Karkh and Rusafa sides of Baghdad Governorate, according to the administrative division and knowledge, And Figure 1 represents a map of the Karkh and Rusafa sides of the city of Baghdad, divided by the ten regions according to the administrative division:
After the application of the statistical program Matlab, the study model test was conducted in formula (4.1) using Moran’s Z test in formula (6.3) to detect the spatial dependence From the table (3.1) below, which shows the some Statistical indicate, it was found that the value of Moran’s Z-test statistic is 9.606904 when using the ordinary spatial adjacency matrix  $W_{ij}$ and that its P-Value is (1.96) which is less than the significance level 0.05 and this indicates the significance of the test, meaning that there is a spatial dependence between the Karkh and Rusafa areas of Baghdad governorate.

Table 1: shows the some Statistical indicate

<table>
<thead>
<tr>
<th>F</th>
<th>R²</th>
<th>R²adj</th>
<th>Z</th>
<th>LM₀ SACML</th>
<th>LM₁ SACML</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.93335</td>
<td>0.3663</td>
<td>0.3176</td>
<td>9.6069</td>
<td>3.857293</td>
<td>244.8722</td>
<td>0.53095699</td>
</tr>
</tbody>
</table>

The data was analyzed according to the usual spatial adjacency matrix  $W_{ij}$ with dimension (85×85) by using the greatest possibility method in estimating the parameters of the general spatial regression model (SAC) in formula (4.1) and also under the Rook criterion, as these parameters were tested using the F test and extracting the significant value of the test and comparing it With the significance level of 0.05, so we note that the F-test value equals 14.93335 and its P-Value is ( 0.217185) which is less than the significance level 0.05. This indicates that the differences are significant, that is, there is at least one of the explanatory variables (age rate, tumor size Treatment, hormone, immunity) have a significant effect on the Y-dependent variable (the number of infected), It is also noted that the value of the coefficient of the value of the coefficient of determination $R^2$ is 0. 366343, which indicates that 36% of the differences in the number of injured people under spatial
influences are caused by the explanatory variables.

\[ R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (8.1) \]

\[ R_{adj}^2 = 1 - \frac{(1 - R^2)(n - 1)}{(n - k - 1)} \quad (8.2) \]

Table 2: shows the real and estimated values of the dependent variable Y of using the usual spatial adjacency matrix \( W_{ij} \) under the Rook.

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th>Y_hat</th>
<th></th>
<th>Y</th>
<th>Y_hat</th>
<th></th>
<th>Y</th>
<th>Y_hat</th>
<th></th>
<th>Y</th>
<th>Y_hat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>34.30839</td>
<td>23</td>
<td>25</td>
<td>59.95965</td>
<td>45</td>
<td>50</td>
<td>92.41552</td>
<td>67</td>
<td>8</td>
<td>32.7784</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>75.64811</td>
<td>24</td>
<td>24</td>
<td>48.22248</td>
<td>46</td>
<td>49</td>
<td>99.64901</td>
<td>68</td>
<td>3</td>
<td>14.33738</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>71.95312</td>
<td>25</td>
<td>42</td>
<td>78.66708</td>
<td>47</td>
<td>202</td>
<td>406.5243</td>
<td>69</td>
<td>14</td>
<td>35.01177</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>37.3608</td>
<td>26</td>
<td>40</td>
<td>85.61975</td>
<td>48</td>
<td>94</td>
<td>174.8316</td>
<td>70</td>
<td>42</td>
<td>60.6197</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>50.00752</td>
<td>27</td>
<td>246</td>
<td>487.8845</td>
<td>49</td>
<td>38</td>
<td>71.01928</td>
<td>71</td>
<td>39</td>
<td>78.02265</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>76.93921</td>
<td>28</td>
<td>76</td>
<td>129.0298</td>
<td>50</td>
<td>8</td>
<td>25.40555</td>
<td>72</td>
<td>18</td>
<td>33.43757</td>
</tr>
<tr>
<td>7</td>
<td>184</td>
<td>321.1754</td>
<td>29</td>
<td>19</td>
<td>42.81988</td>
<td>51</td>
<td>1</td>
<td>16.89708</td>
<td>73</td>
<td>42</td>
<td>80.21287</td>
</tr>
<tr>
<td>8</td>
<td>59</td>
<td>113.6461</td>
<td>30</td>
<td>26</td>
<td>43.52387</td>
<td>52</td>
<td>4</td>
<td>16.9209</td>
<td>74</td>
<td>25</td>
<td>56.68784</td>
</tr>
<tr>
<td>9</td>
<td>26</td>
<td>53.69172</td>
<td>31</td>
<td>94</td>
<td>127.6302</td>
<td>53</td>
<td>4</td>
<td>24.34964</td>
<td>75</td>
<td>124</td>
<td>248.8755</td>
</tr>
<tr>
<td>10</td>
<td>27</td>
<td>46.64975</td>
<td>32</td>
<td>59</td>
<td>112.7837</td>
<td>54</td>
<td>3</td>
<td>18.60641</td>
<td>76</td>
<td>43</td>
<td>96.49516</td>
</tr>
<tr>
<td>11</td>
<td>45</td>
<td>74.52909</td>
<td>33</td>
<td>29</td>
<td>56.14945</td>
<td>55</td>
<td>29</td>
<td>72.4466</td>
<td>77</td>
<td>21</td>
<td>45.2633</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>50.35123</td>
<td>34</td>
<td>19</td>
<td>44.46662</td>
<td>56</td>
<td>49</td>
<td>92.70032</td>
<td>78</td>
<td>11</td>
<td>36.16298</td>
</tr>
<tr>
<td>13</td>
<td>49</td>
<td>80.2287</td>
<td>35</td>
<td>16</td>
<td>35.0339</td>
<td>57</td>
<td>241</td>
<td>480.4666</td>
<td>79</td>
<td>18</td>
<td>29.27032</td>
</tr>
<tr>
<td>14</td>
<td>43</td>
<td>88.28071</td>
<td>36</td>
<td>20</td>
<td>47.73456</td>
<td>58</td>
<td>86</td>
<td>159.3303</td>
<td>80</td>
<td>19</td>
<td>32.06947</td>
</tr>
<tr>
<td>15</td>
<td>35</td>
<td>64.38263</td>
<td>37</td>
<td>14</td>
<td>43.63961</td>
<td>59</td>
<td>33</td>
<td>65.22252</td>
<td>81</td>
<td>17</td>
<td>43.16309</td>
</tr>
<tr>
<td>16</td>
<td>40</td>
<td>80.45722</td>
<td>38</td>
<td>80</td>
<td>183.1086</td>
<td>60</td>
<td>12</td>
<td>31.40908</td>
<td>82</td>
<td>10</td>
<td>36.86993</td>
</tr>
<tr>
<td>17</td>
<td>34</td>
<td>80.29568</td>
<td>39</td>
<td>25</td>
<td>54.40315</td>
<td>61</td>
<td>21</td>
<td>49.08453</td>
<td>83</td>
<td>9</td>
<td>27.45094</td>
</tr>
<tr>
<td>18</td>
<td>217</td>
<td>458.8325</td>
<td>40</td>
<td>7</td>
<td>24.60467</td>
<td>62</td>
<td>24</td>
<td>52.03069</td>
<td>84</td>
<td>36</td>
<td>79.9632</td>
</tr>
<tr>
<td>19</td>
<td>77</td>
<td>147.5653</td>
<td>41</td>
<td>28</td>
<td>56.85412</td>
<td>63</td>
<td>25</td>
<td>55.16188</td>
<td>85</td>
<td>28</td>
<td>46.29183</td>
</tr>
<tr>
<td>20</td>
<td>22</td>
<td>45.30557</td>
<td>42</td>
<td>42</td>
<td>86.01443</td>
<td>64</td>
<td>28</td>
<td>71.51414</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>10</td>
<td>25.06506</td>
<td>43</td>
<td>38</td>
<td>61.68985</td>
<td>65</td>
<td>118</td>
<td>247.3315</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>20</td>
<td>48.01115</td>
<td>44</td>
<td>47</td>
<td>87.89959</td>
<td>66</td>
<td>30</td>
<td>61.27334</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: shows the real and estimated values of the dependent variable (Y).

From the table below, which shows the some Statistical indicate, it was found that the value of Moran’s Z-test statistic is 9.4636 when using the ordinary spatial adjacency matrix $W_{ij}^{Adj}$ and that its P-Value is (9.4636) which is less than the significance level 0.05 and this indicates the significance of the test, meaning that there is a spatial dependence between the Karkh and Rusafa areas of Baghdad governorate.

<table>
<thead>
<tr>
<th>F</th>
<th>$R^2$</th>
<th>$R^2_{adj}$</th>
<th>Z</th>
<th>$LM_p$ SAC</th>
<th>$LM_{SAC}$</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.74235</td>
<td>0.288</td>
<td>0.541</td>
<td>9.4636</td>
<td>4.019951</td>
<td>243.7283</td>
<td>0.529613874</td>
</tr>
</tbody>
</table>

The data was analyzed according to the usual spatial adjacency matrix $W_{ij}^{Adj}$ with dimension $(85\times85)$ by using the greatest possibility method in estimating the parameters of the general spatial regression model (SAC) in formula (4.1) and also under the Rook criterion, as these parameters were tested using the F test and extracting the significant value of the test and comparing it With the significance level of 0.05, so we note that the F-test value equals 14.74235 and its P-Value is (0.217185) which is less than the significance level 0.05. This indicates that the differences are significant, that is, there is at least one of the explanatory variables (age rate, tumor size Treatment, hormone, immunity) have a significant effect on the Y-dependent variable (the number of infected).

It is also noted that the value of the coefficient of the value of the coefficient of determination $R^2$ is 0.288089, which indicates that 28% of the differences in the number of injured people under spatial influences are caused by the explanatory variables.

After a statistical comparison of MAPE it was found that the modified Rook was the best.
Table 4: shows the real and estimated values of the dependent variable $Y$ of using the usual spatial adjacency matrix $W_{ij}^{Adj}$ under the Rook.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>34.17303</td>
<td>23</td>
<td>25</td>
<td>59.79909</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>75.51326</td>
<td>24</td>
<td>24</td>
<td>48.06132</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>71.8199</td>
<td>25</td>
<td>42</td>
<td>78.50752</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>37.22444</td>
<td>26</td>
<td>40</td>
<td>85.45901</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>49.87278</td>
<td>27</td>
<td>246</td>
<td>487.5282</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>76.80437</td>
<td>28</td>
<td>76</td>
<td>128.8662</td>
</tr>
<tr>
<td>7</td>
<td>184</td>
<td>321.0572</td>
<td>29</td>
<td>19</td>
<td>42.65768</td>
</tr>
<tr>
<td>8</td>
<td>59</td>
<td>113.5154</td>
<td>30</td>
<td>26</td>
<td>43.36621</td>
</tr>
<tr>
<td>9</td>
<td>26</td>
<td>53.5577</td>
<td>31</td>
<td>94</td>
<td>127.475</td>
</tr>
<tr>
<td>10</td>
<td>27</td>
<td>46.48807</td>
<td>32</td>
<td>59</td>
<td>112.6325</td>
</tr>
<tr>
<td>11</td>
<td>45</td>
<td>74.36824</td>
<td>33</td>
<td>29</td>
<td>55.99358</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>50.19739</td>
<td>34</td>
<td>19</td>
<td>44.26188</td>
</tr>
<tr>
<td>13</td>
<td>49</td>
<td>80.07474</td>
<td>35</td>
<td>14</td>
<td>43.43479</td>
</tr>
<tr>
<td>14</td>
<td>43</td>
<td>88.13064</td>
<td>36</td>
<td>20</td>
<td>47.52983</td>
</tr>
<tr>
<td>15</td>
<td>35</td>
<td>64.23056</td>
<td>37</td>
<td>28</td>
<td>56.74786</td>
</tr>
<tr>
<td>16</td>
<td>40</td>
<td>80.30701</td>
<td>38</td>
<td>80</td>
<td>182.9124</td>
</tr>
<tr>
<td>17</td>
<td>34</td>
<td>80.14413</td>
<td>39</td>
<td>25</td>
<td>54.19868</td>
</tr>
<tr>
<td>18</td>
<td>217</td>
<td>458.7071</td>
<td>40</td>
<td>7</td>
<td>24.39865</td>
</tr>
<tr>
<td>19</td>
<td>77</td>
<td>147.4183</td>
<td>41</td>
<td>28</td>
<td>56.74786</td>
</tr>
<tr>
<td>20</td>
<td>22</td>
<td>45.15236</td>
<td>42</td>
<td>42</td>
<td>85.91037</td>
</tr>
<tr>
<td>21</td>
<td>10</td>
<td>24.85871</td>
<td>43</td>
<td>38</td>
<td>61.58434</td>
</tr>
<tr>
<td>22</td>
<td>20</td>
<td>47.80571</td>
<td>44</td>
<td>47</td>
<td>87.79591</td>
</tr>
</tbody>
</table>

Figure 3: shows the real and estimated values of the dependent variable $(Y)$.
9. Conclusions

1. It was found that the value of the F-test of the general spatial regression model (SAC) when using the regular and modified spatial adjacency matrixes \( W_{ij}, W_{ij}^{Adj} \) has significant differences, meaning that there is at least one explanatory variable that has a significant effect on the dependent variable \( Y \).

2. It was shown through the graph that the estimated values of the dependent variable \( Y \) when using the regular and modified spatial adjacency matrixes \( W_{ij}, W_{ij}^{Adj} \) are close to the real values of the same variable.

3. Show by the value of the coefficient of determination \( R^2 \) when using the matrix \( M W_{ij} \) that there are 36% of the differences, and when using the \( W_{ij}^{Adj} \) matrix, the percentage of differences is 28% that occurred for the dependent variable \( Y \), and this is caused by the explanatory variables.

References


