Iraqi Statisticians Journal / Vol. 2 (2025): Special issue for ICSA2025 : 158-163

IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Using the Spatial Autoregressive Model to Study the Effect of Some Variables on the Mean Hemoglobin

Jinan Abdullah Anber [1], Ghiath Hameed Majeed[2], Mohammed Qadoury Abed [3], and Wadhah S. Ibrahim[4]

[1]Baghdad Technical College of Management, Dep. of Financial and Banking Sciences Techniques, Middle Technical University, Baghdad, Iraq.
[2]College of Basic Education, Dep. of Mathematics, Mustansiriyah University, Baghdad, Iraq
[3]A l-Mansour University College, Dep. of Accounting & Banking Science, Baghdad, Iraq.
[4]College of Management and Economics, Dep. Of Statistic, Mustansiriyah University, Baghdad, Iraq.

ARTICLE INFO

Article history:
Received 27/11/2024
Revised 28/11/2024
Accepted 14/1/2025
Available online 15/5/2025

Keywords:
Spatial Autoregressive Model
Maximum likelihood method
Rook Contiguity Criterion
Hemoglobin

ABSTRACT

Spatial estimation methods have shown their effectiveness in finding appropriate solutions that researchers encounter with regression models. In this research, the effect of some variables on hemoglobin will be studied at the level of Iraqi regions and. The maximum likelihood method will be used to estimate the spatial autoregressive model under the Rook contiguity spatial adjacency matrix. It has shown that the average hemoglobin data suffers from spatial dependence and that the variables (white blood cells, platelets, blood viscosity) have an effect on hemoglobin diseases.

1. Introduction

A lot of studies and researchers have mentioned the importance of the effectiveness of the spatial variable included in the spatial economic measurement model [7]. The importance of the spatial variable lies in studying its effect through the use of spatial adjacency matrices, which achieve a study of the variables that can influence the study's explanatory variables and thus influence the studied response variable. Spatial economic measurement models are used for economic, social, and even health studies. These models have proven their effectiveness in predicting future values with few errors from the reality of the phenomenon under study.

The health research dealt with predicting future disease cases and the way to prevent some of the effects of variables on these diseases, including, but not limited to, cancer, COVID-19, and others. So the research looks for the problem of studying the hemoglobin disease and the extent of the effect of some variables on this disease in the presence of the effect of the variable, and from this came the goal of the investigation, which includes studying this effect of the spatial variable with some variables, which are white blood cells, platelets, and blood viscosity, on the. Hemoglobin disease by using a spatial autoregressive model using the modified Rook adjacency matrix ($\mathcal{W}_R^{adj}$)[8].

* Corresponding author. E-**mail** address: dr_wadhah_stat@uomustansiriyah.edu.iq
https://doi.org/10.62933/ns1nxx54

2. Spatial Autoregressive Model

This model represents a special case of the general spatial autoregressive model which was proposed by Anselin [5]. The spatial autoregressive model can be expressed mathematically through the following formula:

$$\underline{\mathcal{Y}} = \lambda \mathcal{W} \underline{\mathcal{Y}} + \mathcal{X}\underline{\beta} + \underline{\mathcal{U}} \qquad (1)$$

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

Formula (1) can be written as follows:

$$\underline{\mathcal{Y}} = \left(\mathcal{X}\underline{\beta} + \underline{\mathcal{U}}\right)(1 - \lambda\mathcal{W})^{-1} \qquad (2)$$

So: $\underline{\mathcal{Y}}$ is a vector with dimensions (nx1) and represents the dependent variable; λ is the parameter of spatial dependence. $\mathcal{X}$: a matrix with dimensions (nxk) for explanatory variables. $\underline{\beta}$: a dimensional vector (kx1) of parameters related to the explanatory variables $\mathcal{X}$. $\underline{\mathcal{U}}$: A vector of errors with dimensions (nx1), which has a normal distribution with mean zero and variance $\sigma^2 I_n$. $\mathcal{W}$: A matrix of spatial adjacencies; its dimensions are (nxn), and it is fixed and predetermined. $\underline{\mathcal{U}}$: vector of errors (residuals) with dimension (nx1)[3,4].

The most remarkable feature of the spatial autoregressive model is the use of the spatial adjacency matrix, which is based on the spatial variable, and through it the effect of the spatial variable is studied with the rest of the explanatory variables on the response variable.
There are a lot of studies in which researchers dealt with different types of spatial adjacency matrices [1,6]. The formula for the modified Rook matrix can be as follows:

$$\sum_{j=1}^{n} \mathcal{W}_{ij}^{adj} = 1 \qquad j = 1,2,3,\dots,n \qquad (3)$$

$$\mathcal{W}_R^{adj} = \mathcal{W}_{ij}^{adj} = \begin{cases} \dfrac{\mathcal{W}_{ij}}{\sum \mathcal{W}_{ij}} & \text{if} \quad \mathcal{W}_{ij} = 1 \\ 0 & \text{if} \quad \mathcal{W}_{ij} = 0 \end{cases} \qquad (4)$$

3. The MLE to estimate the parameters of SAR:

It is one of the important estimation methods because it gives the best estimate of the model parameters among several methods. In order to estimate the model parameters, we rewrite the model in formula (1) as follows [2,8]**:**

$$\underline{\mathcal{U}} = (I - \lambda\mathcal{W})\underline{\mathcal{Y}} - \mathcal{X}\underline{\beta} \qquad (5)$$

The formula of the potential function for the spatial autoregressive model is as follows:

$$L\left(\underline{\beta}, \sigma^2, \lambda\right) = (2\pi\,\sigma^2)^{\frac{-n}{2}} + \,|\,I - \lambda\mathcal{W}|\exp\left((1/2\sigma^2)\underline{\mathcal{U}}'\underline{\mathcal{U}}\right) \qquad (6)$$

So:

$$\underline{\mathcal{U}}'\underline{\mathcal{U}} = \underline{\mathcal{Y}}'(I - \lambda\mathcal{W})'(I - \lambda\mathcal{W})\underline{\mathcal{Y}} - 2\underline{\beta}'\mathcal{X}'(I - \lambda\mathcal{W})\underline{\mathcal{Y}} + \underline{\beta}'\mathcal{X}'\mathcal{X}\underline{\beta} \qquad (7)$$

To estimate the model parameters, we derive the potential function as in the following formula:

$$\frac{\partial LnL(\underline{\beta}, \sigma^2, \lambda)}{\partial \underline{\beta}} = -\left(\frac{1}{2\sigma^2}\right) 2\mathcal{X}'(I - \lambda\mathcal{W})\underline{\mathcal{Y}} + 2\mathcal{X}'\mathcal{X}\,\underline{\hat{\beta}}_{mle} \qquad (8)$$

By making formula (8) equal to zero and simplifying it, we obtain:

$$\underline{\hat{\beta}}_{mle} = (\mathcal{X}'\mathcal{X})^{-1}\,\mathcal{X}'(I - \lambda\mathcal{W})\underline{\mathcal{Y}} \qquad (9)$$

$$\hat{\sigma}^2{}_{mle} = \frac{(\underline{\mathcal{Y}} - \lambda\mathcal{W}\underline{\mathcal{Y}} - \mathcal{X}\,\underline{\hat{\beta}}_{mle})'(\underline{\mathcal{Y}} - \lambda\mathcal{W}\underline{\mathcal{Y}} - \mathcal{X}\,\underline{\hat{\beta}}_{mle})}{n} \qquad (10)$$

It is noted in the formulas referred to (9) and (10) that it is not possible to find both model parameters ($\underline{\hat{\beta}}_{mle}$ و $\hat{\sigma}^2{}_{mle}$) if the spatial dependence parameter λ is unknown. This is why Ord proposed in 1975 AD, a formula was developed to calculate the determinant $|I - \lambda\mathcal{W}|$ as follows[7]:

$$Ln\,|I - \lambda\mathcal{W}| = \sum_{i=1}^{n} Ln\,(1 - \lambda\omega_i) \qquad (11)$$

So: ω_i Eigen values for the spatial weights matrix $\mathcal{W}$.

By substituting the value of each of the following formulas (9), (10), (11) into the maximum likelihood function formula (7), we obtain the concentrated likelihood function, which is a non-linear function, and through the use of iterative methods for the concentrated likelihood function in the formula (12) The spatial dependence parameter λ for the spatial autoregressive (SAR) model will be obtained[3]:

$$CLf = \frac{-n}{2} Ln\left[\frac{(e_{ols} - \lambda e_L)'((e_{ols} - \lambda e_L)}{n}\right] + \sum_{i=1}^{n} Ln\,(1 - \lambda\omega_i) \quad (12)$$

Whereas

$$\underline{e_{ols}} = \underline{\mathcal{Y}} - \mathcal{X}\,\underline{\hat{\beta}}_{ols},$$

$$\underline{\mathrm{e}_{\mathrm{L}}} = \mathcal{W}\underline{\mathcal{Y}} - \mathcal{X}\,\underline{\hat{\beta}}_{L}$$

$$\underline{\hat{\beta}}_{ols} = (\mathcal{X}'\mathcal{X})^{-1}\,\mathcal{X}'\underline{\mathcal{Y}},$$

$$\underline{\hat{\beta}}_{L} = (\mathcal{X}'\mathcal{X})^{-1}\,\mathcal{X}'\mathcal{W}\underline{\mathcal{Y}}$$

4. The tests of Spatial dependence:

Two types of tests are adopted to confirm the effect of the dependence of the spatial variable on the variables of the studied phenomenon, they are[1,:

4.1. Moran Coefficient Test:

The test is a statistical tool to measure spatial dependence in the data of the phenomenon to be studied, and we can symbolize it with the symbol (I_{MC}), and it corresponds and is similar to the Durbin Watson test in time series data, (Akkar, year) and the value of the test ranges between (1+,1), and the test formula is as follows[2,7]:

$$I_{MC} = \frac{n(\underline{\mathcal{U}}'\mathcal{W}\,\underline{\mathcal{U}})}{S_\theta(\underline{\mathcal{U}}'\underline{\mathcal{U}})} \;, S_\theta = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathcal{W}_{\mathrm{ij}} \quad (13)$$

The asymptotic distribution for Moran's statistic test has been reached, as it conforms to the standard normal distribution and its formula is as follows:

$$\mathcal{Z}_{\mathrm{I}_{MC}} = \frac{\mathrm{I}_{MC} - \mathrm{E}(\mathrm{I}_{MC})}{\sqrt{\mathcal{V}(\mathrm{I}_{MC})}} \;,$$

$$\mathrm{E}(\mathrm{I}_{MC}) = \frac{tr(\mathcal{M}\mathcal{W})}{n-k} \quad (14)$$

$$\mathcal{V}(I_{MC}) = \frac{tr(\mathcal{M}\mathcal{W}\mathcal{M}\mathcal{W}') + tr(\mathcal{M}\mathcal{W})^2 + (tr(\mathcal{M}\mathcal{W}))^2}{(n-k)(n-k+1)} - \left(E(I_{MC})\right)^2$$

$$\mathcal{M} = (I - (\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}')$$

So: tr: the sum of the elements of the central diameter. k represents the number of explanatory variables. $\mathcal{M}$: represents a solid matrix, which is square and symmetric.

It can be said that if the calculated value of ($\mathcal{Z}_{\mathrm{I}_{MC}}$) is greater than the tabulated value of ($\mathcal{Z}_{Ta}$) at a certain level of significance, we accept the alternative hypothesis, and this means that there is a spatial dependence between the components of the studied phenomenon. However, if it is the opposite, this leads to accepting the null hypothesis, meaning that there is no spatial dependence between the components of the studied phenomenon.

4.2. Lakrange multiplied test

It is a test to reveal the presence of spatial dependence and the suitability of the model to the data for the phenomenon studied, and the formula for the Lakrange multiplied test for the parameter (λ) is as follows[8]**:**

$$L\mathcal{M}_\lambda = \frac{\left(\frac{\underline{\underline{\mathcal{U}}}'\mathcal{W}\,\underline{\underline{\mathcal{U}}}}{S^2}\right)^2}{tr\,[(\mathcal{W} + \mathcal{W}')\mathcal{W}\,]} \quad (15)$$

S^2: Error variance of the general linear regression model. To compare ($L\mathcal{M}_\lambda$) with a tabular value of $\chi^2_{(1,\alpha)}$.

5. The practical aspect

The main motivation for studying the average hemoglobin volume in people and its effect on the percentage of blood volume in the human body in light of the effect of spatial adjacencies is to know the extent of the average hemoglobin volume in all regions, especially adjacent regions, and the extent of the effect of adjacencies. The data was obtained from the Ministry of Health and is represented by the following variables:

$\mathcal{Y}$: represents the average volume of hemoglobin in the blood. $\mathcal{X}1$: represents the average white blood cell count. $\mathcal{X}2$: represents blood viscosity. $\mathcal{X}3$: average platelet count.

Table 1: Shows the results of the two tests to detect the presence of spatial dependence and the suitability of the data to the SAR model.

Moran I	$\mathrm{E}(\mathrm{I}_{MC})$	$\mathcal{V}(I_{MC})$	$\mathcal{Z}_{\mathrm{I}_{MC}}$	$\mathcal{LM}_{\lambda}$	**P-Value**
10.62071	-0.09128	0.00315	4.74515	0.5342e+06	0.0002

From table (1), we can observe the following:

1- The results of the Moran test I are at a significance level (0.05) under the null and alternative hypotheses and when using the modified spatial weight matrix under the Rook contiguity criterion ($\mathcal{W}_R^{adj}$), which is shown in the table as significant, and this indicates that the data suffers from a spatial dependence.

2- As for the Lakrange multiplied test ($\mathcal{LM}_{\lambda}$) and when using the modified spatial adjacency matrix ($\mathcal{W}_R^{adj}$), the results shown in Table (1) have significant differences after comparing with the value of $\chi 2\,(1)$ which is equal to 3.841 at a significance level of (0.05) under the two null hypotheses. The alternative is: $\mathrm{H}_0\!:\gamma = 0$ v.s $\mathrm{H}_1\!:\gamma \neq 0$. This indicates the suitability of the parametric spatial regression (SAR) model to the data under the Rook criterion adjacency to the modified spatial adjacency matrix.

Table 2: shows the real and estimated values of the dependent variable $\mathcal{Y}$ (average volume of hemoglobin in the blood) for the (SAR) model under the modified Rook adjacency criterion

Seq.	$\mathcal{Y}$	$\hat{\mathcal{Y}}$	Seq.	$\mathcal{Y}$	$\hat{\mathcal{Y}}$	Seq.	$\mathcal{Y}$	$\hat{\mathcal{Y}}$
1	15.6	16.2586	35	28.9	20.75686	69	22.7	28.04331
2	18	10.4697	36	27.1	31.67304	70	14.2	22.51365
3	33.4	30.4299	37	14.5	18.03501	71	24.6	20.17282
4	18.1	10.0008	38	32.5	31.64742	72	27	21.07596
5	28.5	21.6027	39	14.8	14.92748	73	26.4	23.36136
6	22.6	19.82652	40	29.1	28.71492	74	18.7	20.49127
7	29	30.83314	41	32.9	23.92726	75	10.8	13.96714
8	37	27.94752	42	16.9	17.78384	76	26	30.16027
9	10.9	19.9562	43	15.3	19.86147	77	19.2	17.48229
10	28.4	26.66548	44	17.1	20.80452	78	37.3	27.01755
11	16.4	20.10693	45	19.7	21.29526	79	15.9	18.53791
12	28	25.93379	46	20.7	18.97886	80	25	28.79479
13	28.6	30.65026	47	25.3	18.3014	81	28.5	33.42864
14	18.9	16.41474	48	16.8	20.29043	82	27.9	22.70528
15	15.5	20.61307	49	27.6	21.01607	83	20.3	23.60111
16	32.2	28.41085	50	17.3	20.0622	84	30.8	31.90561
17	18.7	16.869	51	16.8	16.78182	85	23.9	25.66404
18	16.3	19.09818	52	17.1	20.31742	86	25.2	19.70668
19	17.1	22.33729	53	24.7	28.79486	87	18.1	19.03025
20	28.2	23.69451	54	28.3	20.75678	88	30.3	31.0269
21	17.1	27.71868	55	31.1	29.97849	89	31.1	28.70193
22	9	12.81922	56	17.3	20.1467	90	10	11.07385

23	14.9	22.02083	57	11.3	15.03441	91	26.9	19.69752
24	10.9	11.60777	58	24.4	21.8718	92	27.1	22.70361
25	18.2	21.97422	59	17.8	25.32512	93	29.7	19.14904
26	30.3	30.51361	60	34.5	31.13784	94	33.5	33.31069
27	27	20.2951	61	29.2	28.48956	95	31.6	29.57477
28	17.5	16.45262	62	23.6	25.39244	96	27.5	31.04049
29	16.1	20.28659	63	29.9	28.80119	97	28.8	26.61364
30	14.6	20.73807	64	18.6	17.08889	98	21.5	21.82183
31	15.1	17.35372	65	27.9	29.45145	99	21.3	24.33911
32	28.6	22.53195	66	28	23.68999	100	31.9	35.08109
33	15.3	18.19675	67	22	16.1954	101		
34	16.5	19.30246	68	19	16.6825	102		

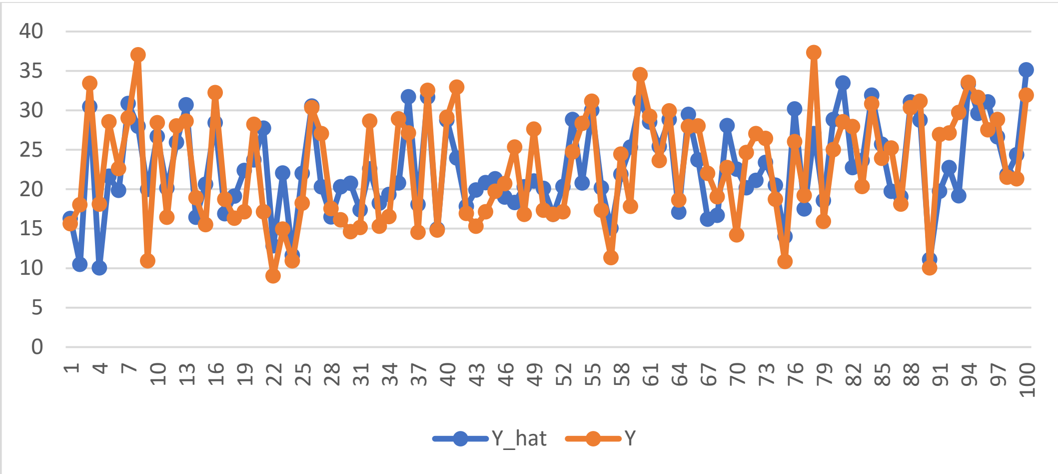

Figure 1: Shows the graph for the real and estimated values of the dependent variable.

6. Conclusions

1. Based on the results of the Moran coefficient test, it was concluded that the average hemoglobin data in the blood suffers from spatial dependence.
2. Through the Lakrange multiplied test, the suitability of the parametric spatial regression (SAR) model was revealed in light of the juxtaposition of the Rook criterion.
3. The variables (white blood cells, platelets, and blood viscosity) have an effect on haemoglobin disease.

4. Recommendations

1. We recommend comparing the spatial adjacency matrix with the distance matrix, which depends on the coordinates of people's locations and not the center of spatial units.
2. For future studied For future studies, a comparison will be made between the spatial adjacency matrix and the matrix that combines the adjacency length of the spatial units and the distance factor between observations, which is called Combined Distance-Boundary Weights.

References

[1] Akkar, A. A (1992). (2021). Estimation the spatial Durban regression model for anemia patients sample in some region of Al-Karkh/Baghdad. *Journal of Administration and Economics*, (128), 275-293.

[2] Akkar, A. A., Anber, J. A., & Ibrahim, W. S. (2023). The Study of Spatial Effect on Some Variables of Water Pollution in Baghdad Regions. *Al-Rafidain University College For Sciences*, (54).

[3] Anselin, L. (1992)"Spatial Data Analysis With GIS: An Introduction to Appalachian the Social

Sciences", National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA 93106.

[4] Anselin, L. (2010). Thirty years of spatial econometrics. Papers in regional science, 89(1), 3-26.

[5] Anselin, L. (2013). *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.

[6] Ibrahim, W. S., Majeedb, G. H., & Hussain, W. J. (2021). Comparison and estimation of a Spatial Autoregressive (SAR) model for cancer in Baghdad Regions. *Int. J. Agricult. Stat. Sci. Vol*, *17*(1), 1921-1927.

[7] Ibrahim, W. S., & Mousa, N. S. (2022). Estimation of the general spatial regression model (SAC) by the maximum likelihood method. *International Journal of Nonlinear Analysis and Applications*, *13*(1), 2947-2957.

[8] Ibrahim, W. S., & Mousa, N. S. (2022). Estimation of the parameters of the spatial general regression model (SAC) by the method of the greatest possibility (MLE) of the adjacency matrix using simulation. *journal of the college of basic education*, *28*(114/علمي).