

Estimation of COVID-19 infections in Iraqi governorates using generalized moments method in spatial autoregressive model

Suhad Ali Shaheed Al-Temimi, Rawaa Salh Al-Saffar

Statistics Department, Mustansiriyah University

ABSTRACT

At the end of 2019, a new type of virus that infects the human respiratory system was discovered in China, and it was briefly called COVID-19. In March 2020, the world Health Organization (WHO) declared Corona Virus a global pandemic. The Corona Virus is transmitted through air or through contact. The possibility of infection increases in the area or areas neighboring to the area that witnessed a community spread of the virus or when individuals return from that affected area to their areas of residence. Given the limited studies on the impact of affected neighboring areas or countries, this study focused on using the spatial autoregression model, one of the econometric models. Model parameters have been estimated using the Generalized Moment Method (GMM) which has the ability to correct the Endogeneity that occurs by the spatial regression variable as well as due to the endogenous variables. The results showed that the number of infections (Y_n) of Corona epidemic increases as there are infections in the surrounding areas and vice versa. This confirms the impact of spatial neighborhood on the spread of infections among neighboring governorates.

Keywords: Covide-19 , spatial Autoregressive model, GMM method, endogeneity

Corresponding Author:

Suhad Ali Shaheed Al-Temimi
statistics Department, Mustansiriyah University
Baghdad, Iraq
E-mail: dr.suhadali@uomustansiriyah.edu.iq

1. Introduction

In December of 2019, a new disease was discovered in Wu han , in central China, which was called COVID – 19. On March 11, 2020, the World Health Organization (WHO) classified COVID – 19 as a global pandemic. Given the lack of studies on the mechanism of spread of this epidemic between countries, this study uses spatial Autoregressive Model as it takes into account the spatial dimensions of the phenomenon data under study.

This model allows us to study the spatial effect or the spatial contiguity between the governorates and the increase of infections in those governorates. Anselin (over the period 1988-2001) discussed in detail the possibility of spatial interaction, represented by spatial Autocorrelation, and spatial structure, represented by spatial Heterogeneity.

The use of the Generalized Moment Method (GMM) estimator for estimating the spatial Autoregressive Model, developed by (Arellano and Bover 1995) and (Blundell and Bond 1998), is due to its ability to correct the Endogeneity that occurs by the spatial regression variable as well as due to the endogenous variables, treating some econometrics problems such as measurement errors and weakness of the Instrumental Variable (IV), control of the problem of Heteroskedasticity and Autocorrelation in the term of random error, and flexibility in the implemetation of mathematical operations [1].

2. Research objective

A study of the effect of spatial contiguity between governorates on the increase in the number of new cases of corona virus. Which The spread of the Corona pandemic in the Iraqi governorates linking with the degree of spatial convergence of each other

3. Research problem

There are few studies on Corona virus and the effect of spatial contiguity on increasing the number of infections during a specific period of time. The GMM technique using to estimate parameter in the spatial autoregression model because The consistency characteristic is one of the most important characteristics of the GMM estimator.

4. Literature review

In 1981, Cliff and Ord presented a method for dealing with spatial autocorrelation in linear regression models [2]. In 1988, Anselin and Griffith explained that when spatial dependence and, or spatial heterogeneity is neglected, the results of data analysis may become incorrect and as a result, special method must be used instead of those that follow the basic assumptions of econometrics [3].

In 2008, after studying Durbin Spatial Model, Elhorst, Piras and Arbia explained that the results are based on some methodological issues such as choosing the time period and including the fixed effects model. They used the Generalized Momentum Method (GMM) of estimation for the purpose of testing endogeneity that not only appears from endogenous variables at different levels, but also appears due to the prominent economic growth rates in neighboring economics [4].

In 2009, Jacobs , Ligthari and Vrijbrug studied dynamic models of panel data in the presence of both endogeneity and spatially correlated errors. This study was carried out through an expansion of the three-stage (GMM) method [5].

This method corrected the spatially correlated errors of the fixed effects of panel data models by taking the spatial regression and temporal regression for the dependent variable and for the explanatory variables added to the model.

The researchers have shown that the differences in the bias and the RMSE criterion between the spatial (GMM) estimates and the (GMM) estimates adopted in the research are considered minor differences, neglecting the spatially correlated errors.

In the same year, Agha & Vedeine studied the estimation of the spatial dynamic model for panel data using the (GMM) method to study the convergence of issues of European union regions [6]. The researchers have proposed two strategies for estimating the spatial dynamic model of panel data using (GMM). The first was the expansion of moment constraints for the (Arellano & Bond) estimator of the spatial dynamic autoregression model for panel data. The second strategy is based on the inclusion of spatial dependence in the error term to calculate the optimal spatial weights matrix. The research reached the conclusion that the (GMM) estimator controls both endogeneity and other problems of econometric model.

In 2010, Lee & Yu examined (GMM) estimator for the spatial dynamic panel data model with fixed effects when n is large but T is small relative to n [7]. The two researchers demonstrated that by excluding spatial fixed effects, the convergence properties of estimators are achieved.

Also they conclude that there are suitable quadratic moment conditions to handle spatial effects and not just linear moment conditions, after comparing the spatial dynamic model for panel data with the dynamic model for panel data [7-8].

In 2011, the researcher Atra Sami Ghani , in his PhD thesis, titled Biz methods in analyzing the spatial economic measurement model for the purpose of identifying suitable alternative methods to traditional economic measurement methods that can be used in dealing with spatial data [9].

In 2013, Baltagi, B.H, Fingleton , B and Pirrotte , A focused on the estimation and predictive performance of several estimators for the dynamic and autoregressive spatial lag panel data model with spatially correlated disturbances. The main idea of their research was to mix non-spatial and spatial instruments to obtain consistent estimates of the Parameters [10].

They used Monte Carlo Simulations to compare the short – run and long – run effects and evaluate the predictive efficiencies of optimal and various suboptimal predictors using the Root Mean Square Error (RMSE) criterion. Finally, they proposed a spatial GMM estimator under the assumptions that the model includes temporal and spatial lags on the endogenous variable together with SAR-RE disturbances.

Also, in 2013, Osman and Suleyman demonstrated that maximum likelihood (ML) estimator for spatial autoregressive models is generally inconsistent when heteroskedasticity is not taken into account in the estimation. Thus, they used the robust generalized method of moment (GMM) estimation for the spatial model allowing for a spatial lag not only in the dependent variable but also in the disturbance term [11].

As a result, they showed the consistency of the robust GMM estimator and determined its asymptotic distribution.

In 2016, the researcher Al-Tamimi, Suhad, in her PhD thesis, addressed some parameterized methods to estimate the Spatial Dynamic Panel Data (SDPD) model [12].

This thesis discussed two spatial models. First, when there are fixed spatial effects. Second, when there are fixed spatial temporal effects in addition to fixed spatial effects.

In 2017, Jay M, Erin E, Mevin B, Ephram M, and Marie – Josee stated that ecological data often show pattern, which can be modeled as autocorrelation [13].

They designed conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models to model spatially autocorrelated data. The study reached, by using maximum likelihood and Bayesian methods, to determine estimators of statistically significant characteristics.

5. Estimate spatial autoregressive model (SARM)

One of the most widely used model of spatial mutual influences is one that proposed by Cliff and Ord (1973).

This model was originally based on the model considered by Whittle (1954).

The model can be expressed according to the following formula:

$$Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n \quad \dots(1)$$

Where as:

Y_n : represents the $(n \times 1)$ vector of observations of the dependent variable on every spatial point $(n=1,2,\dots,N)$.

X_n : represents the $(n \times k)$ matrix of exogenous variables.

β : a vector of $(k \times 1)$ represents interaction parameters of exogenous variables.

W_n : A matrix of dimensions $(n \times n)$ represents the matrix of spatial weights and describes the spatial arrangement of units within the sample and all diagonal elements are zero.

λ : represents the parameter of interaction with respect to the spatially lagging dependent variable.

$W_n Y_n$: represents the spatially lagging dependent variable.

ϵ_n : represents the $(n \times 1)$ vector of regression disturbances.

The direct estimation of model (1) results in a biased and inconsistent estimate of most of the model.

Parameters according to the Maximum Likelihood method (direct approach as named by Yu and Lee in 2010).

Yu and Lee explained that there are several proposals for estimating the model in equation (1) for the purpose of obtaining estimators with convergent properties and consistent estimates of the estimated parameters, noting that all estimation methods are applied after proving that the parameters are fully diagnosed.

Also, the random errors (ϵ_{nt}) in equation (1) when they are spatially correlated, they become as in the following equation:

$$\epsilon_n = \gamma \epsilon_n + \lambda_2 W_n \epsilon_n + \epsilon_n \quad \dots(2)$$

Where as:

ϵ_n : A vector of dimensions $(n \times 1)$ reflects the specifications of the random error term of the model, and in the general model, it is series and spatially correlated.

γ : A sequential autocorrelation coefficient with respect to the random error term.

λ_2 : A spatial autocorrelation coefficient with respect to the random error term.

At that point, the model in equation (1) is called the spatial error model, but if the errors are independent and have a symmetrical distribution, then the general model is called the Spatial Lag Model (SLM).

6. Spatial autocorrelation and spatial heterogeneity

Spatial autocorrelation and spatial heterogeneity are two characteristics associated with spatial data that must be taken into consideration when developing a model of spatial dynamic phenomena.

The spatial autocorrelation is defined as the existence of a functional relationship between what happens at a certain point in space, which is expressed statistically according to the following formula ($Cov[y_i y_j] \neq 0$ for $i \neq j$), spatial autocorrelation is considered to be of great importance because most of the results are based on it in studies dealing with spatial econometrics.

The term spatial heterogeneity relates to variation in relationships over space, or, intrinsic characteristics unevenly distributed over space. Also, spatial heterogeneity refers to the instability of error variances, which can be expressed according to the following formula:

$$(\text{var} \text{Var}(\epsilon_i) = \sigma_r^2 \text{ when } i \in L_r)$$

Where:

L : represents a group of spatial units.

r : represents the location of those units.

The spatial autocorrelation between the two units (j,i) is relatively dependent on their location and therefore positive spatial autocorrelation appears when the values of the random variable (low / high) are concentrated in neighboring spatial units.

7. Moran's test (1- Test)

This test is considered one of the most important tests to verify the presence or absence of a spatial autocorrelation between the cross section.

For the SDPD model, the test equation for the vector (Y_{nt}) is:

$$I = \left(\frac{R}{S_0} \right) Y'_{nt} W_n Y_{nt} (Y'_{nt} Y_{nt})^{-1} \quad \dots(3)$$

Whereas:

(S_0) represents $(\sum_{i=1}^n \sum_{j=1}^n W_{ij})$.

- (W_{ij}) represents (ij^{th}) for the matrix of spatial weights. W_n

- (R) represents the number of regions (cross sections) taken in the study.

The test hypothesis is based on the null hypothesis, that is, there is no spatial autocorrelation ($H_0: \lambda = 0$). Accepting or rejecting null hypothesis is through a comparison with (Z), which is calculated according to the following formula:

$$Z = \frac{I - E(I)}{\sqrt{\text{var}(I)}} \quad \dots(4)$$

Where ($E(I)$ and $\text{var}(I)$) represent respectively the expectation and variance of the Moran test, and it is calculated according to the following formula:

$$E(I) = \frac{-1}{R-1}$$

$$\text{Var}(I) = \frac{RS_4 - S_3 S_1 (1-2R)}{(R-1)(R-2)(R-3)(\sum_i \sum_j w_{ij})^2} \quad \dots(5)$$

$$S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2$$

$$S_2 = \sum_i (\sum_i w_{ij} + \sum_j w_{ji})^2$$

$$S_3 = \frac{R^{-1} (Y'_{nt} Y_{nt})^2}{(R^{-1} Y'_{nt} Y_{nt})^2}$$

$$S_4 = (R^2 - 3R + 3)S_1 - RS_2 + 3 (\sum_i \sum_j w_{ij})^2$$

The value of the Moran test (1-test) is calculated under the assumption of a normal distribution and that the value of (Z) can be used as the value of (P-value) and be compared with the value of (1-test). It should be noted here that the Moran test (1-test) is calculated for each year separately.

8. Generalized method of moments (GMM)

The estimation by using Generalized Method of Moments (GMM) is based on moment functions that are constructed from the same model. In 2007, (Lee) developed a specific methodology for constructing moment

functions in which he used the linear and quadratic moment functions of the (SAR) model based on converting the model in equation (1) to the following reduced form:

$$Y_n = (I_n - \lambda W_n)^{-1} X_n \beta_0 + (I_n - \lambda W_n)^{-1} \epsilon_n \quad \dots (6)$$

Then , above equation is multiplied by the matrix of spatial weights (W_n) to become:

$$W_n Y_n = G_n X_n \beta_0 + G_n \epsilon_n \quad \dots(7)$$

Where ($G_n = W_n(I_n - \lambda W_n)^{-1}$). The endogenous variable ($W_n Y_n$) on the left – hand side of equation (7) represents a function with respect to the non- random part represented by ($G_n X_n \beta_0$) and a function with respect to the random part by ($G_n \epsilon_n$).

Therefore, Lee and Yu (2007-2010) formulated moments in terms of the random part and non- random part .

The non- random part was measured by:

$$Q_n = [Z_n] \quad \dots(8)$$

On represents the vector of relevance and efficient instrumental variables to obtain consistent estimators with respect to both ($W_n Y_n$) and have fixed column dimensions greater or equal to ($k_x + 3$) . The matrix of instrumental variables that are included in the linear moments function is:

$$Q_{1n} = [Q'_{11}, Q'_{12}, \dots, Q'_{1n}]' \quad \dots (9)$$

In equation (9) above , the matrix is computed by taking the expectation for the instrumental variable (Z_{nt}) , which represents the matrix of dimensions ($k_x + 2$) from the following variables:

$$Z_n = (W_n Y_n, X_n)$$

Equation (9) can be calculated with the initial values of the parameters (λ_0, β'_0) .

When performing the orthogonality parameter transformation, the random error term matrix becomes:

$$\epsilon_{1n} = [\epsilon'_{11}, \epsilon'_{12}, \dots, \epsilon'_{1n}]'$$

Where:

$$\epsilon_n = (I_n - \lambda_0 W_n) Y_n - X_n \beta_0 \quad \dots(10)$$

Then the formula for the linear moments function is ($Q'_{1n} \epsilon_{1n}$) .

As for the random part represented by ($G_n \epsilon_n$) , it is measured by (P_{Ln}) , where ($L=1,2,\dots, m$) represents the number of moment functions , and (P_{Ln}) is a matrix with dimensions ($L \times n$) and is specified regularly (UB) in the sum of row and column are in absolute terms , and (P_{Ln}) are calculated according to the following formula :

$$P_{nL} = \left(G_n - \left(\frac{tr G_n}{n} I_n \right) \right) \quad \dots(11)$$

Thus , the quadratic moment functions will take the form ($\epsilon'_n(\theta) P_{nL} \epsilon_n(\theta)$) and that the orthogonality condition is fulfilled in the case that the random error term is (i.i.d) and thus :

$$E(\epsilon'_n(\theta) P_{nL} \epsilon_n(\theta)) = tr(P_{nL} (\epsilon'_n(\theta) \epsilon_n(\theta))) = 0 \quad \dots(12)$$

Thus , for both stochastic and non- stochastic parts , the instrumental variables (IV) arise in such a way that they are related to ($W_n Y_{nt}$) but not associated with (ϵ_{nt}) and therefore linear and quadratic moment functions can be written as a function in terms of the parameters ($\theta = (\lambda_0, \beta'_0)$) and as follows:

$$g_n(\theta) = \begin{pmatrix} \epsilon_n'(\theta) \mathbf{P}_{n,1} \epsilon_n(\theta) \\ \vdots \\ \epsilon_n'(\theta) \mathbf{P}_{n,m} \epsilon_n(\theta) \\ \mathbf{Q}'_n \epsilon_n(\theta) \end{pmatrix} \quad \dots(13)$$

The moment functions in equation (13) include both linear and quadratic moments and are called systematic or finite moments. Liu and others (2010), proved that given the set of moment function ($g_{nT}(\theta)$) any other moment functions that can be added to this set is redundant. Also, they showed that the ML estimator is characterized by the set of moment functions $g_n(\theta)$, therefore, the GMM estimator based on these moment functions is asymptotically equivalent to the ML estimator. They suggested, when the errors are (i.i.d), another best set of quadratic moment functions so that the optimal GMM estimator is asymptotically more efficient than the quasi ML estimator.

9. Consistency and asymptotically distribution for GMME under moment conditions

The consistency characteristic is one of the most important characteristics of the GMM estimator under the presence of the linear and quadratic moment conditions in equation (13). In order to prove this characteristic, the identification conditions for the model must be studied in equation (1) with the following assumption:

$$\delta = (\beta')'$$

$$S_n = (I_n - \lambda W_n)$$

$$S_n(\lambda) = S_n + (\lambda_0 - \lambda)W_n$$

Then, the formation of $(\epsilon_{nt}^*(\theta))$, can be expanded as follows:

$$\epsilon_n(\theta) = f_n(\theta) + S_n(\lambda)S_n^{-1}\epsilon_n \quad \dots(14)$$

Whereas:

$$\left. \begin{aligned} \epsilon_n &\equiv \epsilon_n(\theta_0) \\ f_n(\theta) &\equiv S_n(\lambda)S_n^{-1}Z_t \delta_0 - Z_n \delta \\ &= Z_n(\delta_0 - \delta) + (\lambda_0 - \lambda)G_n Z_n \delta_0 \end{aligned} \right\} \quad \dots(15)$$

For the moment functions in equation (13) to be identified, the sufficient condition is to prove that:

$$P \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{Q}'_n [\mathbf{Z}_n, \mathbf{G}_n \mathbf{Z}_n \delta_0] \quad \dots(16)$$

Which has full rank ($K_2 + 1$), whereas

$$\mathbf{Z}_n = (Z'_n, \dots, Z'_n)' \quad : \quad \mathbf{G}_n = I_n \otimes G_n$$

As a result of (\mathbf{Z}_n) involving spatial regressions, this condition will generally be fulfilled as long as $(\delta \neq 0)$.

As for the identification of the matrix of instrumental variables (IVs) in equation (9), (Q_n) must be strictly identified. As mentioned above, the dimensions of the columns of (Q_n) are fixed for all values of (n) and with full rank, and thus (Q_n) is fully identified. The identification condition requires that the following be fulfilled:

$$P \lim_{n \rightarrow \infty} \frac{1}{n(T-1)} \mathbf{Q}'_n \mathbf{Q}_n \quad \dots (17)$$

After proving the identification , the optimal estimate for the GMM estimator at (θ_0) values is the following:

$$\hat{\theta}_{o,n} = \underset{\theta \in \Theta}{\operatorname{argmin}} g'_n(\theta) \hat{\Sigma}_n^{-1} g_n(\theta) \quad \dots (18)$$

Where $(\operatorname{argmin}_{\theta \in \Theta})$ means the estimator that achieves a lower estimate than the rest of the estimators within the parameter space (θ) . $\hat{\Sigma}_n$ represents the information matrix of moment functions $(g_n(\theta))$ and when talking the inverse , we get the variance and covariance matrix while the number of its rows is greater than or equal to $(k_z + 1)$. Under the assumption that the random errors are naturally distributed (i.i.d) , then (Σ_n) is calculated according to the following formula:

$$\Sigma_n = \sigma_0^4 \begin{pmatrix} \frac{1}{n} \omega'_{nm} \omega_{nm}^s & 0_{m \times (k_z+1)} \\ 0_{(k_z+1) \times m} & \frac{1}{\sigma_0^2} \frac{1}{n} \mathbf{Q}'_n \mathbf{Q}_n \end{pmatrix} \quad \dots (19)$$

Whereas

$$\omega_{nm} = [\operatorname{vec}_D(\mathbf{P}_{n,1}), \dots, \operatorname{vec}_D(\mathbf{P}_{n,m})]$$

$\operatorname{vec}_D(\mathbf{P}_{n,L})$, denotes the vertical vector, which is formulated by the main diagonal elements of the matrix $(\mathbf{P}_{n,j})$, and the same applies to:

$$\omega_{nm}^s = \operatorname{vec}_D(\mathbf{P}_{n,1}^s), \dots, \operatorname{vec}_D(\mathbf{P}_{n,m}^s)$$

$\operatorname{vec}_D(\mathbf{P}_{n,L}^s)$, denotes the vertical vector , wich is formulated by the main diagonal elements of the matrix $(\mathbf{P}_{n,L}^s)$ and is calculated through

$$\mathbf{P}_{n,L}^s = \mathbf{P}_{n,L} + \mathbf{P}'_{n,L}$$

As for the asymptotically normal distribution of the estimator of the generalized moment method in the presence of the linear and quadratic moment functions $(\hat{\theta}_{o,n})$, it is possible to have a specific distribution such as the normal distribution at the initial values (θ_0) where the amount $(\frac{g_n(\theta_0)}{\sqrt{n(T-1)}})$ has a distribution asymptotical to the normal distribution . on the basis of the central limit theory (CLT) , when (n) is large , the variance of the estimator $(\hat{\theta}_n)$ is as in equation (19) after taking its inverse .

10. The simulation

For the purpose of implementing a simulation to generate explanatory variables that follow a normal distribution , the (Box-Muller) method was used to generate explanatory variables that follow a regular normal distribution $(U(0,1))$ and then convert those variables into independent random variables (Z_r) that follow a standard normal distribution according to the following formula:

$$\left. \begin{aligned} Z_r &= (-2LnU_r)^{\frac{1}{2}} \cdot \cos(2\pi U_r) \\ Z_{r+1} &= (-2LnU_r)^{\frac{1}{2}} \cdot \sin(2\pi U_r) \end{aligned} \right\} \dots (20)$$

To transform the variables from the standard normal distribution to the normal distribution with mean (μ) and variance (σ^2) , the following formula is used :

$$X_i = \mu + \sigma^2 Z \quad , \quad i = 1,2, \dots n \quad \dots (21)$$

As for the random error generation (ϵ_{nt}), it is generated according to the independent normal distribution ($\epsilon_{nt} \sim N(0, \sigma_\epsilon^2)$) and by using the formula below :

$$e_i = 0 + \sigma_\epsilon^2 Z, i = 1, 2, \dots, n \quad \dots (22)$$

On the application side of this research, it has been assumed that the value of the variance is equal to (1).

The dependent variable was generated directly from the models used in the simulation experiments and by using the explanatory variables that were generated with the corresponding parameters (β) in addition to the errors in equation (19). Also, it has been assumed that the values of the parameters (β) are equal to (1) and in the form of the vertical vector whose dimensions are (1 × 18) and its elements represent the unity, and thus :

$$Y_i = X_{r,i} \beta_r + e_i \quad \dots (23)$$

With regard to the matrices of spatial weights, the Queen contiguity matrix has been chosen which has the advantage of being Row –Normalized as shown in Appendix (1). As for the proposed spatial weight matrix, it was adopted on the basis of distances between adjacent governorate centers (W_n) and as shown in Appendix (2).

Table 1. The results of the simulation experiment to obtain the initial values

	λ_0	β_0
$\sigma = 0.1$	SD=0.071	SD=0.041
	RMSE=0.071	RMSE=0.041
$\sigma = 0.5$	SD=0.092	SD=0.043
	RMSE=0.092	RMSE=0.043
$\sigma = 0.9$	SD=0.095	SD=0.049
	RMSE=0.095	RMSE=0.049

It is noted from the above table that the best initial value is when ($\sigma = 0.1$) for obtaining the lowest (SD) and lowest (RMSE) which will be relied upon in estimating the model according to the GMM.

11. The application

Data published by the Ministry of Health in Iraq for the period from February 2, 2020 to October 1, 2020 has been used for eighteen governorates (including the Kurdistan region) that represent cross-sections (n=18). The data related to the research subject included the following :

- Number of infection with corona virus (representing the dependent variable Y_{nt}).
- Number of health institutions across Iraq (explanatory variable X_n).

For the purpose of detecting the existence of spatial autocorrelation with respect to the spatial regression variable ($W_n Y_n$), the Moran's test was calculated according to the aforementioned equation (3). The null hypothesis states that there is no spatial autocorrelation, ($H_0: \lambda = 0$), and that the hypothesis is accepted or rejected by comparison with the standard value of (Z) as in equation (4).

According to Table 2, the results obtained indicate that the value of the calculated test (1- test) is higher than the standard value of (Z) and therefore we reject the null hypothesis and accept the existence of a spatial autocorrelation of the variable (Y_n) for the eighteen governorates, that is, the number of infections (Y_n) interacts spatially with cross-section representing the governorates of Iraq during the period under study.

Table 2. The results of the spatial autocorrelation test

Null Hypothesis	Z-statistics*	Moran's I test*	Year
Reject H_0	0.75265	2.3368	/2)Feb (2020/ /1)Oct(2020/

- This results were obtained by the researcher using Eviews- 9

Table 3. The estimation of the model parameters for spatial autocorrelation (SAR) by GMM method.

Parameters	Estimator	
	Estimate	T-test
λ	2.0252	0.02258
β	-0.5852	1.0254

Sig. level: 5% ($|t - stat.| > 1.64$)

According to Table 3, the value of the autocorrelation parameter (λ) is positive and this indicates the increase in the number of Covid-19 infections at the level of the eighteen governorates due to the presence of spatial autocorrelation resulting from the absence of natural or administrative obstacles between the governorates.

As for the explanatory variable (X) that represents the number of health institutions, it adversely affects the variable of the number of infections with coronavirus, as an increase of one unit in the number of health institutions reduces the number of infections by (0.5852).

12. Conclusions

- 1- The GMM method has a number of advantages when it is used to estimate model parameters. The most important of these advantages is achieving the least variance, approaching zero, and least bias, approaching the true value of the parameter.
- 2- The explanatory variable (number of health institutions) affects positively and effectively in medically controlling Covid-19 disease. The results showed that increasing one unit in health institutions reduces the number of cases of the disease.
- 3- The autoregression variable (Y_n) causes a significant increase in the number of infections at the level of the eighteen governorates.

13. Recommendations

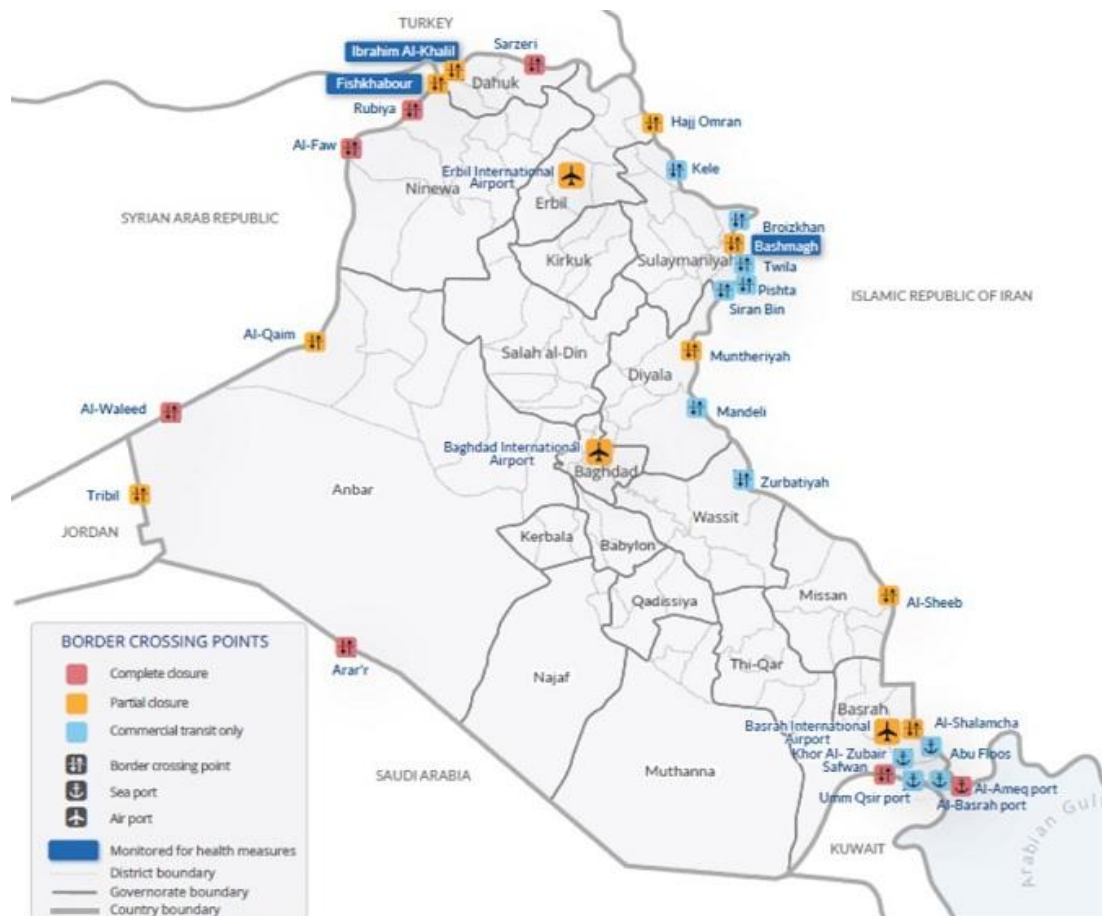
- 1- It is preferable to use GMM in estimation the spatial autoregression model because of its effectiveness in treating Endogeneity in the model and in providing consistent specifications for the estimated parameters.
- 2- It is necessary to test the use of other methods for parameters estimation in the spatial autoregression model such as Two stage least square method (2SLS) and Three stage least square method (3SLS)
- 3- It is important to choose other variables to detect their effects on increasing the infections of Covid-19 disease in Iraq and how to reduce them.

References

- [1] K. Ord, "Estimation Methods for Models of Spatial Interaction". Journal of the American Statistical Association, vol.70, no.349, pp. 120–126, 1975.
- [2] A.D. Cliff and J.K. Ord, "Spatial Processes, Models & Applications", Pion Limited, London, 1981.
- [3] L. Anselin and D. Griffith, "Do Spatial Effects Really Matter in Regression Analysis", Papers, Regional Science Association, vol. 65, pp.11-34, 1988.
- [4] J. P. Elhorst, G. Piras and G. Arbia, "Growth and Convergence in a Multiregional Model with Space-Time Dynamics". Faculty Economics and Business University of Groningen, the Netherland, Geographical Analysis, pp. 338-355, 2008.
- [5] J.P. Jacobs, J. Ligthart and H. Vrijburg. "Dynamic Panel data Models Featuring Endogenous Intraction and Spatially Correlated Errors". International Center for Public Policy, Working Paper Series 0915, Andrew Young School of Policy Studies, Georgia State University, 2009.
- [6] S. Bouayad-Agha, and L. Vadrine. "Estimation strategies for a spatial dynamic panel using GMM. A new approach to the convergence issue of European regions," Spatial Economic Analysis, vol.5, no. 2, pp. 205-227, 2010.
- [7] L. F. Lee and J. Yu, "Efficient GMM Estimation of Spatial Dynamic Panel Data Models with Fixed Effects", Ohio State University, 2010.
- [8] L. F. Lee and J. Yu, "Estimation of Spatial Panels". The essence of Knowledge (now), Boston-Delft, 2011.

- [9] A. S. G. Khudair, "Biz methods in analyzing the spatial economic measurement model with practical application" . PhD thesis in Statistics, College of Administration and Economics, University of Baghdad, 2011.
- [10] B.H. Baltagi, B. Fingleton and A. Pirotte, "Estimating and Forecasting a Dynamic Spatial Panel Data Model", The Department of Economics, University of Oxford, John Wiley & Sons Ltd, 2013.
- [11] O. Dogan and S. Taspınar, "GMM Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances", Ph.D. Program in Economics CUNY Graduate Center 365 Fifth Avenue New York, 2013.
- [12] S. A. Sh. Al-Tamimi, "A comparative study of some methods of estimating the spatial dynamic model of (Panel Data) with practical application", PhD thesis, Department of Statistics, College of Administration and Economics, Mustansiriyah University, 2016.
- [13] J. M. Ver Hoef, E. E. Peterson, M. B. Hooten, E. M. Hanks, and M.-J. Fortin. "Spatial autoregressive models for statistical inference from ecological data." *Ecological Monographs*, vol.88, no. 1, pp.36-59, 2018.
- [14] J. P. Elhorst," Spatial Econometrics From Cross-Sectional Data To Spatial Panels ". Springer Heidelberg, New York, 2014.
- [15] <https://www.who.int/gho/database/ar/>.

Appendix



Province	Cases*	Deaths*	Recovered*	health institutions**
Baghdad	94,187	1,650	29,115	480
Basra	29,055	318	4,463	246
Wasit	19,816	160	4,154	159
Karbala	19,068	187	4,194	121
Dhi Qar	17,565	460	6,005	294
Najaf	16,634	113	4,453	183
AL-Sulaimaniya	16,430	415	5,914	546
Erbil	15,814	97	2,699	390
Al Qadisiyah	13,246	169	2,817	110
Babil	13,149	33	2,308	239
Maysan	12,723	195	4,192	121
Kirkuk	12,348	204	2,434	221
Duhok	10,194	0	299	249
Muthanna	9,367	62	725	110
Diyala	8,752	131	2,983	121
Nineveh	7,262	37	549	317
Salah Al-Deen	6,667	83	1,497	208
Al Anbar	3,924	45	726	276
Total	326,201	4,741	87,434	4391

Source:

* The Iraqi Ministry of Health / daily update until 1/10/2020.

** Ministry of Planning / Central Statistical Organization / Annual Statistical Abstract 2017 pg. 433.

Demonstrates the matrix of spatial weights (W_n) on the basis of inter-governorate contexts

Sulaimaniya	Duhok	Erbil	Nineveh	Kirkuk	Basra	Maysan	Dhi Qar	Muthanna	Najaf	Karbala	Qadisiyah	Wasit	Al Anbar	Babil	Salah Al-Deen	Diyala	Baghdad	Province
0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0.25	0.25	0.25	0	Baghdad
0.2	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0.2	0.2	0	0.2	Diyala
0	0	0	0.2	0.2	0	0	0	0	0	0	0	0	0.2	0	0	0.2	0.2	Salah Al-Deen
0	0	0	0	0	0	0	0	0	0.1429	0.1429	0.1429	0.1429	0.1429	0	0	0.1429	0.1429	Babil
0	0	0	0.1667	0	0	0	0	0	0.1667	0.1667	0	0	0	0.1667	0.1667	0	0.1667	Al Anbar
0	0	0	0	0	0	0.2	0.2	0	0	0	0.2	0	0	0.2	0	0.2	0	Wasit
0	0	0	0	0	0	0	0.2	0.2	0.2	0	0	0.2	0	0.2	0	0	0	Qadisiyah
0	0	0	0	0	0	0	0	0	0.3333	0	0	0	0.3333	0.3333	0	0	0	Karbala
0	0	0	0	0	0	0	0	0.2	0	0.2	0.2	0	0.2	0.2	0	0	0	Najaf
0	0	0	0	0	0.25	0	0.25	0	0.25	0	0.25	0	0	0	0	0	0	Muthanna
0	0	0	0	0	0.2	0.2	0	0.2	0	0	0.2	0.2	0	0	0	0	0	Dhi Qar
0	0	0	0	0	0.3333	0	0.3333	0	0	0	0	0.3333	0	0	0	0	0	Maysan
0	0	0	0	0	0	0.3333	0.3333	0.3333	0	0	0	0	0	0	0	0	0	Basra
0.3333	0	0.3333	0	0	0	0	0	0	0	0	0	0	0	0	0.3333	0	0	Kirkuk
0	0.25	0.25	0	0	0	0	0	0	0	0	0	0	0.25	0	0.25	0	0	Nineveh
0.2	0.2	0	0.2	0.2	0	0	0	0	0	0	0	0	0	0	0.2	0	0	Erbil
0	0	0.5	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Duhok
0	0	0.25	0	0.25	0	0	0	0	0	0	0	0	0	0	0.25	0.25	0	Sulaimaniya