Comparison of Maximum Likelihood and Generalized Method of Moments in Spatial Autoregressive Model with Heteroskedasticity

Rohimatul Anwar[1], Anik Djuraidah[2], Aji Hamim Wigena[3]
{rohimatul_anwar@apps.ipb.ac.id[1], anikdjuraidah@apps.ipb.ac.id[2], aji_hw@apps.ipb.ac.id[3]}

Student of DepartementStatistics, IPB University, Bogor, 16680, Indonesia[1]
Lecturer of DepartementStatistics, IPB University, Bogor, 16680, Indonesia[2,3]

Abstract. Spatial dependence and spatial heteroskedasticity are problems in spatial regression. Spatial autoregressive regression (SAR) concerns only to the dependence on lag. The estimation of SAR parameters containingheteroskedasticityusing the maximum likelihood estimation (MLE) method provides biased and inconsistent. The alternative method is the generalized method of moments (GMM). GMM uses a combination of linear and quadratic moment functions simultaneously so that the computation is easier than MLE. The bias is used to evaluate the GMM in estimating parameters of SAR model with heteroskedasticity disturbances in simulation data. The results show that GMM provides the bias of parameter estimates relatively consistent and smaller compared to the MLE method.

Keywords: Heteroskedasticity, spatial autoregressive, maximum likelihood, generalized moment method.

1 Introduction

Spatial dependence and spatial heteroskedasticity are problems inspatial data [1]. Lesage [2] stated that spatial dependence can be described in regression models, such as autoregressive response, error, or both. Models with dependencies in response are called spatial autoregressive models (SAR).Fotheringham[3] stated that spatial heteroskedasticity can be described using geographically weighted regression (GWR).

Ord [4] considered the maximum likelihood (ML) for the estimation of the regression model. Kelejian and Prucha [5]extended that the MLE estimator is inconsistent in heteroskedasticity disturbances. Anselin[1] introduced the two-stage least squares (S2SLS) method. Kelejian and Prucha [6] introduced the generalized method of moments (GMM). GMM does not require a distribution assumption of the disturbance andcomputationally easier than the ML methods [7].

The results of Kelejian and Prucha's research [5] showed that the estimation method is valid if the assumption of errors is stochastic and identical normal. However, heteroscedasticitycan occur in aggregation data. In this case,heteroskedasticity originates from a data averaging process with many different observations at the time of aggregation [6]. Kelejian and Prucha [7]developed the GMM method into a robust form that has been proven to be consistent if there is heteroskedasticity. Combination of linear and quadratic in moment functionsare simultaneously assumed by GMM.

ICSA 2019, August 02-03, Bogor, Indonesia

DOI 10.4108/eai.2-8-2019.2290489

This study evaluated the MLE and GMM methods for the solution in heteroscedasticity disturbances in the SAR model. In the simulation data, the bias value is used for the evaluation.

2Materials and Methods

2.1 SAR Model with GMM Approach

SAR model specification is considered $\mathbf{y} = \rho\mathbf{W}\,\mathbf{y}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$where$\boldsymbol{\varepsilon}$ ~N($\mathbf{0}$, $\sigma^2\mathbf{I}$), $\mathbf{y}$is the n×1 vector of a dependent variable, $\mathbf{X}$is the n×k matrix predictor, $\boldsymbol{\beta}$is the k×1 vector of regression coefficient parameter, $\mathbf{W}$is the n×n spatial weight matrix, $\boldsymbol{\varepsilon}$is the n×1 vector of disturbances (or innovations), and ρ is the spatial autoregressive parameter[1]. Kelejian and Prucha [7] motivated to control spatial autocorrelations in the model and then reduced the form of the modelas follows:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$$
$$(\mathbf{I} - \rho\mathbf{W})\,\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$$
$$\mathbf{S}\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$$
$$\mathbf{y} = \mathbf{S}^{-1}\mathbf{X}\boldsymbol{\beta}+\mathbf{S}^{-1}\boldsymbol{\varepsilon}$$

$\mathbf{S}=(\mathbf{I} - \rho\mathbf{W})$is the n × n matrix and $\mathbf{X}$ is the nxk matrix a non-stochastic of the independent variable, $\mathbf{W}$is the nxn matrix a weighting dimension with the main diagonal of zero, and $\boldsymbol{\varepsilon}$ is vector $\boldsymbol{\varepsilon}$ with dimension nx1. Let's $\mathbf{W^*}$is the n × n matrix a weighting such that:

$$\mathbf{W^*y} = \mathbf{W^*S}^{-1}\mathbf{X}\boldsymbol{\beta}+\mathbf{W^*S}^{-1}\boldsymbol{\varepsilon} = \mathbf{G}\mathbf{X}\boldsymbol{\beta}+\mathbf{G}\boldsymbol{\varepsilon} \quad (1)$$

Liu *et al.* [8] states $\mathbf{W^*y}$is called the spatial distance of the dependent variable and is a non-stochastic function ($\mathbf{GX}\boldsymbol{\beta}$) and stochastic ($\mathbf{G}\boldsymbol{\varepsilon}$), with $\mathbf{W^*y}$ is correlated with $\boldsymbol{\varepsilon}$ or can be stated as follows: $E[(\mathbf{G}\boldsymbol{\varepsilon})'\boldsymbol{\varepsilon}]= E[(\mathbf{W^*}(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon})'\boldsymbol{\varepsilon}]=\sigma^2\mathrm{tr}\left(\mathbf{W^*}(\mathbf{I} - \rho\mathbf{W})^{-1}\right)\neq 0$, so that this condition fulfils heteroscedastic conditions. GMM estimates parameters in this heteroscedastic condition by weighting the combination of linear and quadratic moment functions. Non-stochastic and stochastic functions are used in shaping the moment function for the GMM estimator, with $\boldsymbol{\theta}=(\rho, \boldsymbol{\beta})'$ with the true parameter value is $\boldsymbol{\theta}_0=(\rho_0,\boldsymbol{\beta}_0)'$. Liu and Lee [9] define for $\mathbf{Q}$ is a matrix constructed from functions $\mathbf{W^*}$ and $\mathbf{X}$. Based on equation (1), $\mathbf{G}= \mathbf{W^*S}^{-1}$then $\mathbf{Q}=(\mathbf{GX}\boldsymbol{\beta}, \mathbf{X})$is the non-stochastic part that forms the moment function of the population$\mathbf{Q}'\boldsymbol{\varepsilon}$. Let P be the size matrix n×n with $\mathrm{tr}(\mathbf{P})=\mathrm{tr}\left(\mathbf{G}\text{-}\mathrm{Diag}(\mathbf{G})\right)=0$, so we get the population moment function in the form of $\boldsymbol{\varepsilon}'\,\mathbf{P}_j\boldsymbol{\varepsilon}$is obtained from the orthogonal form of the moment function as follows:

$$E(\mathbf{Q}'\boldsymbol{\varepsilon}) = \mathbf{Q}'E(\boldsymbol{\varepsilon})=\mathbf{0}_{(k+1)\times 1} \quad (2)$$
$$E(\boldsymbol{\varepsilon}'\,\mathbf{P}_j\boldsymbol{\varepsilon}) = E\left(\mathrm{tr}(\boldsymbol{\varepsilon}'\mathbf{P}_j\boldsymbol{\varepsilon})\right)=E\left(\mathrm{tr}(\mathbf{P}_j\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}')\right)$$
$$=\mathrm{tr}\left(\mathbf{P}_j E(\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}')\right)=\mathrm{tr}\,(\mathbf{P}_j\boldsymbol{\Sigma})=0 \quad (3)$$

where $\boldsymbol{\Sigma}=\mathrm{diag}\,(\sigma_1^2,\ldots,\sigma_n^2)$is a variation of the matrixn×n. Because of$E(\mathbf{Q}'\boldsymbol{\varepsilon}) =0$ and $E(\boldsymbol{\varepsilon}'\,\mathbf{P}_j\boldsymbol{\varepsilon}) =0$, meaning that there is no longer a correlation between $\mathbf{P}_j$ and $\boldsymbol{\varepsilon}$.The moment function based on $\mathbf{Q}$is linear and based on $\mathbf{P}_j$ is quadratic in ε with j = 1,2, ..., n. The choice of the $\mathbf{P}_j$quadratic moment matrix is asymptotically efficient as a GMM estimator [11]. The moment parameter function of the spatial model together with GMM is a combination of the linear and quadratic moment functions as follows:

$$\mathbf{g_n}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Q}'\boldsymbol{\varepsilon}(\boldsymbol{\theta}) \\ \boldsymbol{\varepsilon}(\boldsymbol{\theta})'\mathbf{P}_j\boldsymbol{\varepsilon}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\varepsilon}(\boldsymbol{\theta})'\left(\mathbf{G}\text{-Diag}(\mathbf{G})\right)\boldsymbol{\varepsilon}(\boldsymbol{\theta}) \\ (\mathbf{GX\beta},\mathbf{X})'\boldsymbol{\varepsilon}(\boldsymbol{\theta}) \end{pmatrix} \tag{4}$$

Where $\boldsymbol{\varepsilon}(\boldsymbol{\theta})=\mathbf{S\,y}-\mathbf{X\beta}$.Suppose that $\boldsymbol{\Omega}$ is a matrix of moment functions. $\boldsymbol{\Omega}$ consist of variance and covariance that are linear and quadratic[1] in $\boldsymbol{\varepsilon}$.

$$\boldsymbol{\Omega}=\mathrm{E}[\mathbf{g_n}(\boldsymbol{\theta_0})\mathbf{g_n}'(\boldsymbol{\theta_0})]=\begin{pmatrix} \mathrm{tr}\left(\boldsymbol{\Sigma}\mathbf{P}^*\left(\mathbf{P}^{*'}\boldsymbol{\Sigma}+\boldsymbol{\Sigma}\mathbf{P}^*\right)\right) & \mathbf{0}_{1\times(k+1)} \\ \mathbf{0}_{1\times(k+1)} & \mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q} \end{pmatrix} \tag{5}$$

[2] $\boldsymbol{\Sigma}=$ diag $(\sigma_1^2,\ldots,\sigma_n^2)$ and $\mathbf{P}^* = (\mathbf{G}\text{-Diag}(\mathbf{G}))$, the parameters of the spatial model are simultaneously suspected by a combination of linear and quadratic in moment functions by GMM approach. GMM robust estimator specification is considered:

$$\hat{\boldsymbol{\theta}}=\begin{pmatrix}\hat{\rho} \\ \hat{\boldsymbol{\beta}}\end{pmatrix}=\mathrm{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\mathbf{g_n}'(\boldsymbol{\theta})\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{g_n}(\boldsymbol{\theta}) \tag{6}$$

where $\widehat{\boldsymbol{\Omega}}$ is a consistent estimator for $\boldsymbol{\Omega}$.

2. 2 Data

Simulation data was used in this study. Simulation is conducted with 1000 replications with the number of observations n=30, n=90 and n=900. Simulation data will be considered as the following scenario:

1. The SAR model is specification considered $\mathbf{y}_i = \rho\mathbf{Wy}+\mathbf{X}_i\boldsymbol{\beta}+\boldsymbol{\varphi}_i$.
2. There are two independent variables without an intercept, $\mathbf{X}_i=(\mathbf{X}_{1i}, \mathbf{X}_{2i})$ where $i = 1,2,\ldots,$ n and $\boldsymbol{\beta}=(\beta_1, \beta_2)'$ and coefficient $\boldsymbol{\beta}=(1, 2)'$.
3. $\mathbf{X}_1$ dan $\mathbf{X}_2$ are the $n \times 1$vector and generated from normal distribution N(0,1).
4. ρ is the autoregressive coefficient in the SAR model. The coefficientρthat use for simulation are (-0.8, -0.5, -0.2, 0,0.2, 0.5, 0.8).
5. There are two matrices of spatial weights to be used.
   a) Circular World (CW) Matrix
      1. The first $n/3$ rows (except the first row) all elements are zero except in positions $(i, i+1)$ and $(i, i-1)$, for $i = 2, \ldots, n/3$. While for the first row all elements are zero except in position (1,2) and $(1, n)$. The number of neighbors in these rows is 2.
      2. The second $n/3$ rows of each element are zero except in positions $(j, j \pm r)$,$j = \frac{n}{3}+1, \ldots, \frac{2n}{3}$; $r = 1,2,\ldots,5$. The number of neighbors in these rows is 10.
      3. The third $n/3$ rows (except the last row) all elements are zero except in positions $(j, j+1)$ and $(j, j-1)$, for $j = 2n/3+1, \ldots, n-1$. While for the last row all elements are zero except in position (n, 1) and (n, n-1). The number of neighbors in these rows is 2.
      4. This matrix then is row standardized.

[1]Appendix 1 can be used to derive Ω matrices ithis section
[2]Appendix 2

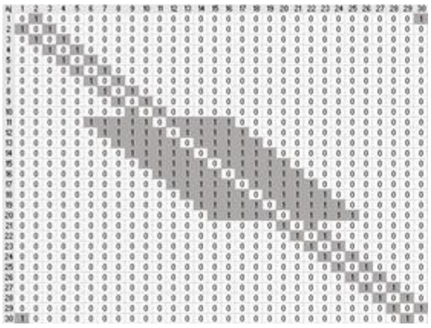


Fig. 1. Circular world matrix

b) Small-Group Interaction (SGI) Matrix
This matrix isintroduced by Lin dan Lee (2010). For each sample size n, we generate random groups where the size of each group is drawn from Uniform (3,20) distribution. For each group, if the group size is greater than 10, thenthe variance equal to the group size; otherwise, thevariance is square of the inverse of the group size.

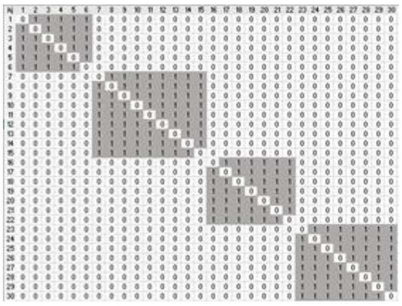


Fig. 2. Small group interaction matrix

6. Heteroskedasticity is generated by h_i in the number of neighbours in the units i and$\varphi_i$is the size of the sampleφ_i~N (0,1). ε_i generated by $\varepsilon_i=\sigma_i\varphi_i$, where$\sigma_i=c\frac{h_i}{\sum_{j=1}^{n} h_j/n}$.
7. The coefficient c used is the signal-noise ratio of the model and c=0.5 is the optimal coefficient.

2.3 Data Analysis Procedure

Data Analysis used *software R Studio 3.5.2* for simulation data. Following Kelejian and Prucha[6] the estimators will be GMM estimators corresponding to the following:

1. Generate simulation data.
2. Estimate parameter by GMM approach.
3. Evaluate estimation parameters based on bias values (Bias-$\hat{\beta}$)

$$\text{Bias-}\hat{\beta}=\frac{1}{s}\sum_{i=1}^{s}\left[\beta_i-\hat{\beta}_i\right] \quad (7)$$

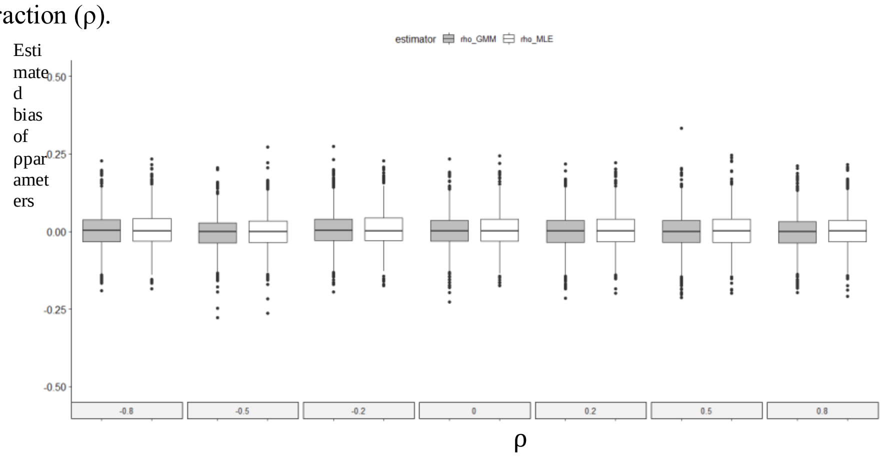
3Result and Discussion

Figures 3-11 has presented bias of estimated parameters forρ,β_1, and $\beta_2$from the MLE and GMM methods. The matrix is used a circular world (CW) matrix that illustrates heteroscedastic conditions in spatial units globally, and small group interactions (SGI) matrix illustrate heteroscedastic conditions in small group interactions..

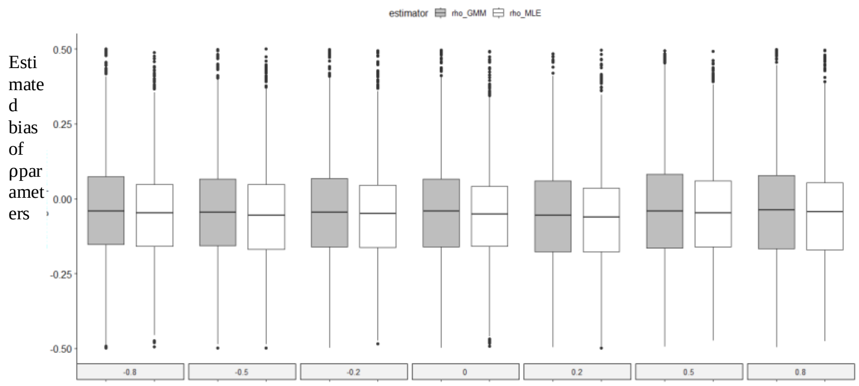
Parameter of *ρ*

Figure 3.is presented**c**omparison of the estimated bias values of the parameters ρ GMM and MLE with n = 30 in the matrix [a]CW matrix and [b]SGI matrix. The bias of the CW matrix is smaller than SGI matrix. Based on **Figure 3. [a]**, the simulation results almost give the same pattern. The bias value of the parameter ρ is around 0.25 (positive or negative), that both GMM and MLE provide the estimated bias value of the maximum parameter of 0.25. GMM gives the estimated bias value ρ smaller than the MLE. Based on **Figure 3. [b]**, the estimated bias value of the ρ parameter is 0.5 (positive or negative), that both GMM and MLE provide the estimated parameter bias value of 0.5. MLE gives the estimated bias value ρ smaller than the GMM method.

The difference in spatial interactions (ρ) does not significantly influence the bias value, that the type of spatial interaction (positive or negative) does not significantly influence the estimated parameter bias value. In addition, it can be seen that the pattern of value distribution to types of spatial interactions has asymmetric shape and low diversity for each type of spatial interaction (ρ).



(a)

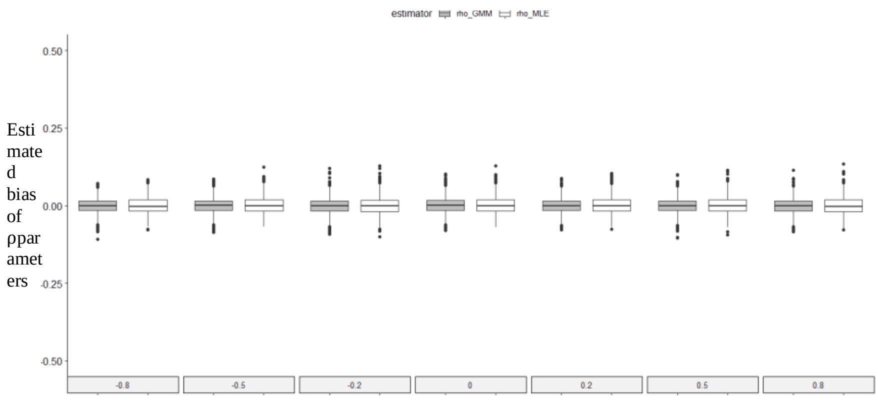


ρ
(b)

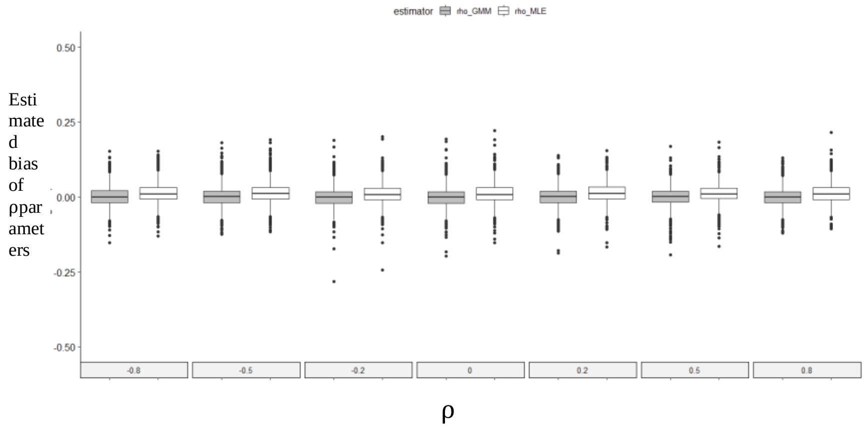
Fig. 3. Comparison of the bias values of ρ for GMMand ML method at $n = 30$in matrix [a] CW and [b] SGI

Figure 4.is presented a comparison of estimated bias values of GMM and MLE parameters at the number of observations n = 90 in the [a]CW matrix and [b]SGI matrix. The bias value at the number of observations n = 90 is smaller than at the number of observations n = 30.The bias of the CW matrix is smaller than SGI matrix.

Based on **Figure 4.**the GMM gives an estimated parameter bias value smaller than MLE. The CW matrix gives a smaller bias value than the SGI matrix. The consistency of GMM can be seen from the smaller variance compared to the MLE method. **Figure 4. [a]** the difference in spatial interactions (ρ) does not significantly affect the bias value, that the type of spatial interaction (positive or negative) does not significantly affect the estimated value of the parameter estimate. whereas**Figure 4. [b]** the difference in spatial interactions (ρ) significantly influences the alleged bias value when ρ = 0.2 (positive and negative).



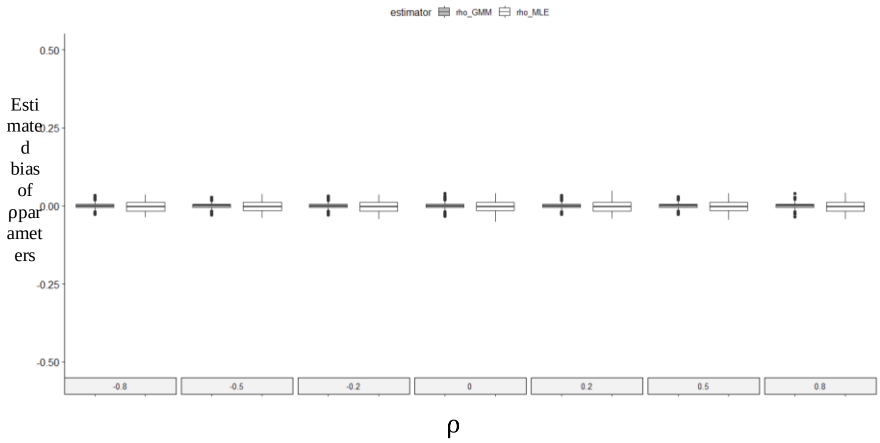
ρ
(a)



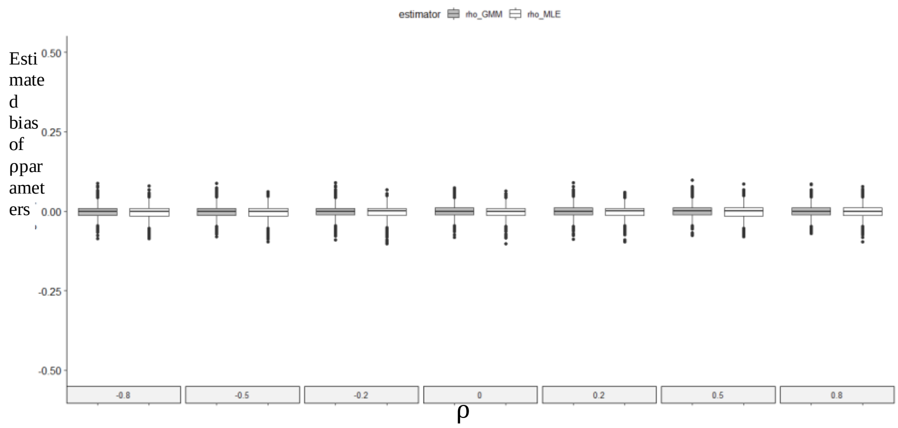
(b)

Fig. 4. Comparison of the bias values of ρ for GMM and ML method at $n = 90$in the matrix [a] CW and [b] SGI

Figure 5.is presented the simulation results at the number of observationsn = 900 for the estimator ρ. The bias value at the number of observations n = 900 is smaller than at the number of observations n = 30 and n = 90.The CW matrix gives a smaller bias value than the SGI matrix.On the CW matrix, the GMM provides smaller bias values than MLE. Whereas on the SGI matrix, the bias value of the GMM is smaller than the MLE method. The difference in spatial interactions (ρ) does not significantly influence the estimated bias value. The shape of the bias value is symmetric.



(a)

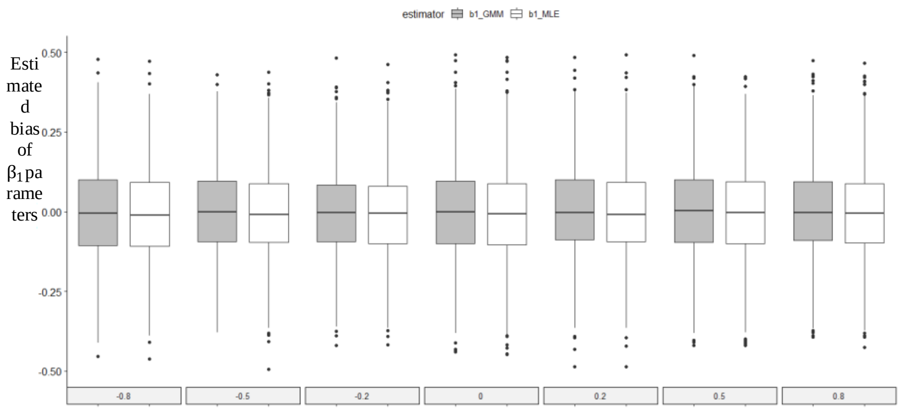


(b)

Fig. 5. Comparison of the bias values of ρ for GMMand MLE method at $n = 900$in the matrix [a] CW and [b] SGI

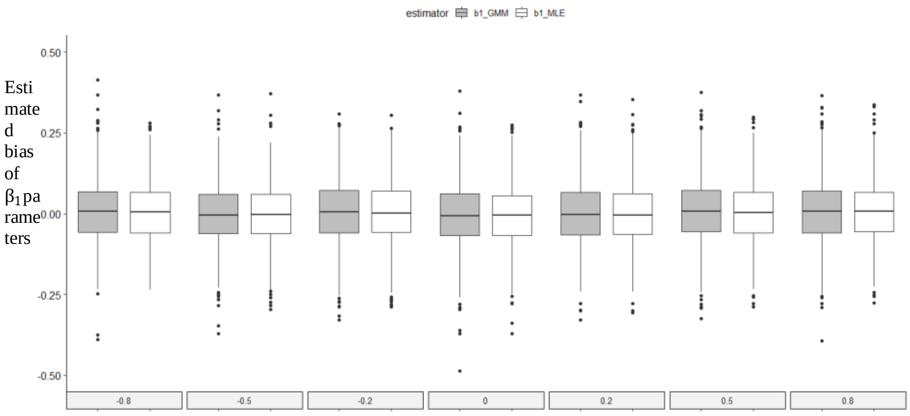
Parameter β_1

Figure 6.is presented the bias values for the estimated β_1 parameter at the number of observations n=30. The SGI matrix gives a smaller bias value than the CW matrix.On the CW matrix, the different types of spatial autocorrelations do not affect the estimated bias value of the β_1 parameter. However, on the SGI matrix, the highest bias value is the GMM and MLE method when there is no spatial autocorrelation ($\rho = 0$). The bias value in the CW matrix is not different results for each spatial autocorrelation. These results are different for simulations with the SGI matrix. The GMM and MLE have a mean of around 0 and symmetric in shape.



ρ

(a)

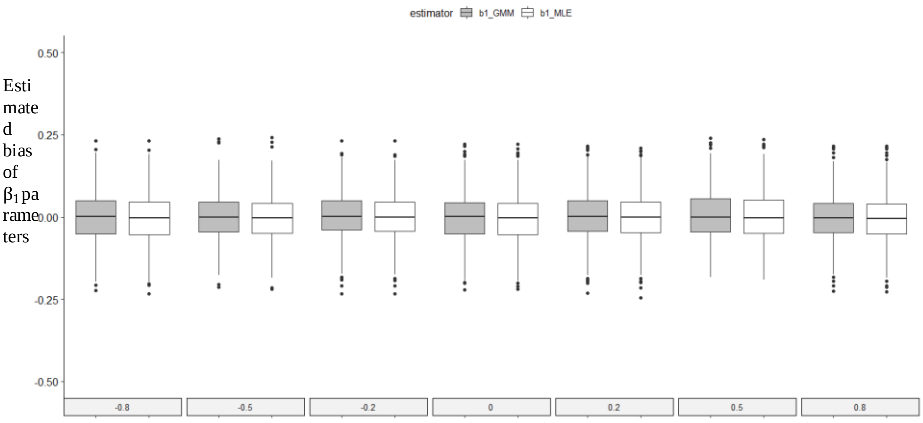


ρ

(b)

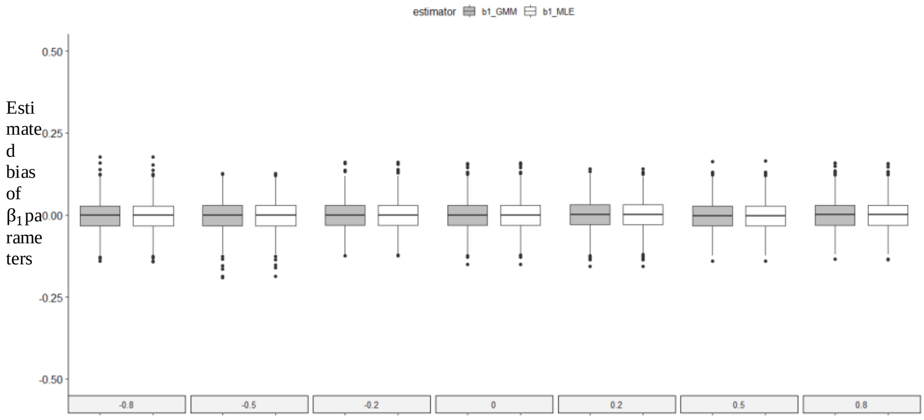
Fig. 6. Comparison of the bias values of β_1 for GMMand MLE method at $n = 30$in the matrix [a] CW and [b] SGI

Figure 7.is presented the β_1 bias at the number of observations $n = 90$.The bias value at the number of observations n = 90 is smaller than at the number of observations n = 30.The bias of the SGI matrix is smaller than CW matrix. The GMM provides smaller bias values than the MLE. Consistency can be seen from the smaller variance compared to the MLE method. The bias values is the same for each spatial autocorrelation. The different types of spatial autocorrelations do not affect the bias value β_1.



ρ

(a)

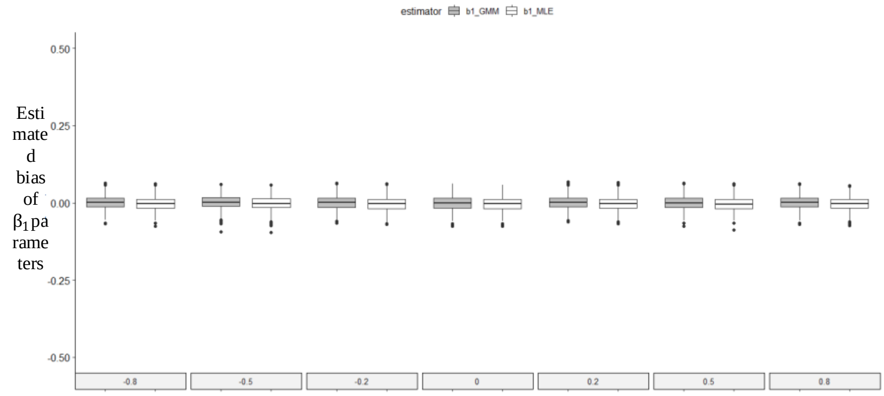


ρ

(b)

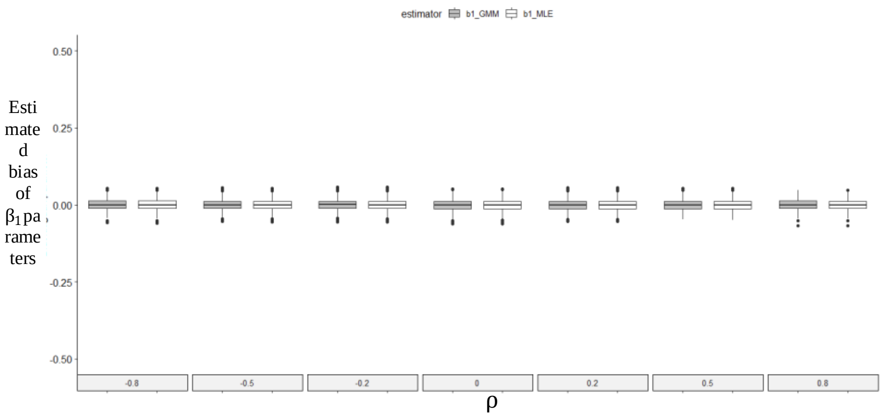
Fig. 7. Comparison of the bias values of β_1 for GMM and MLE method at n = 90 in matrix (a) CW and (b) SGI

Figure 8. is presented a comparison of the estimated bias value of β_1 at the number of observations n = 900 in matrix [a] CW and [b] SGI. The bias value at the number of observations n = 90 is smallest than at the number of observations n = 30 and n=90. The bias of the SGI matrix is smaller than CW matrix.The GMM provides smaller bias values than the MLE. This shows that GMM is more consistent.The difference in spatial interactions (ρ) does not significantly affect the bias value, that the type of spatial interaction (positive or negative) does not significantly affect the alleged bias value of the β_1 parameter. The two matrices do not give very different results, havesymmetrical shapes with an average of around 0 and small variations.



ρ

(a)

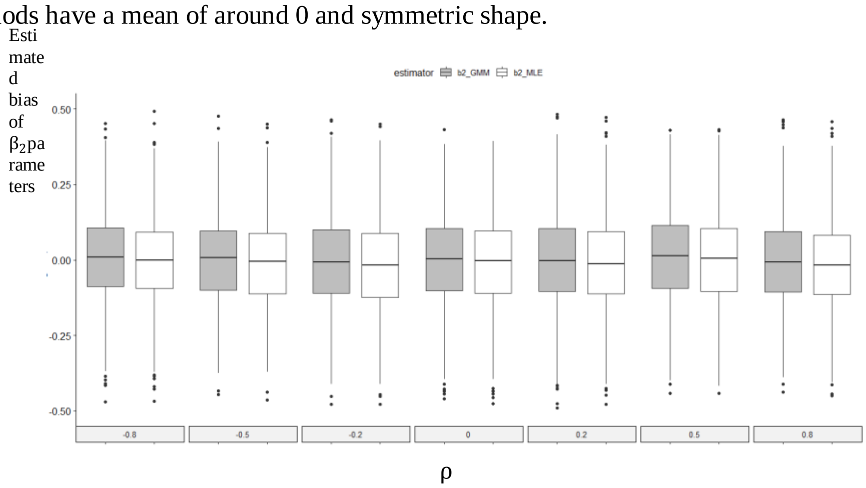


(b)

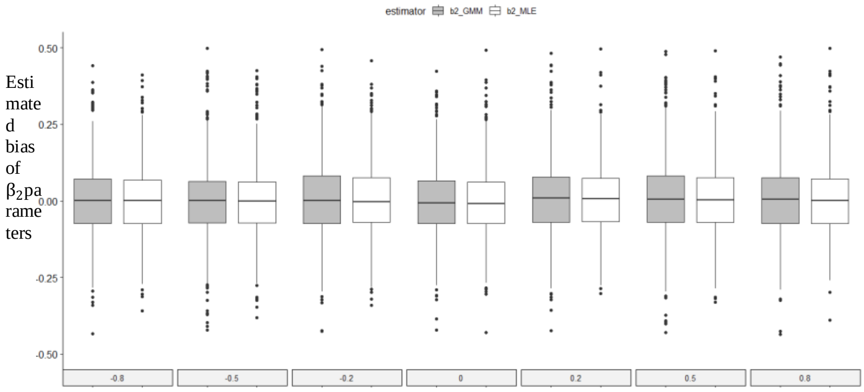
Fig. 8. Comparison of the bias values of β_1 for GMM and MLE method at $n = 900$ in the matrix [a] CW and [b] SGI

Parameter β_2

Figure 9is presented the bias values for the estimated β_2 parameter at the number of observations n=30. The SGI matrix gives a smaller bias value than the CW matrix.On CW mtatrix, the bias of GMM and MLE is quite high, meaning that the variation of the bias is quite large.Wheres on SGI matrix, the bias of the GMM provides a smlles than the MLE. Both methods have a mean of around 0 and symmetric shape.



(a)

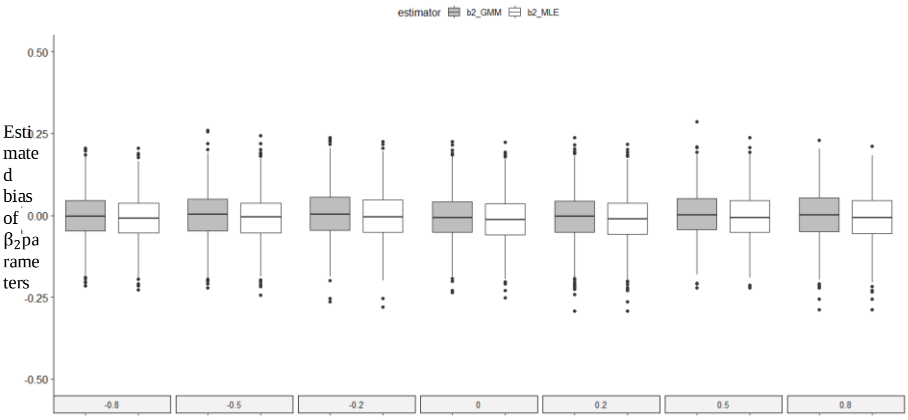


ρ

(b)

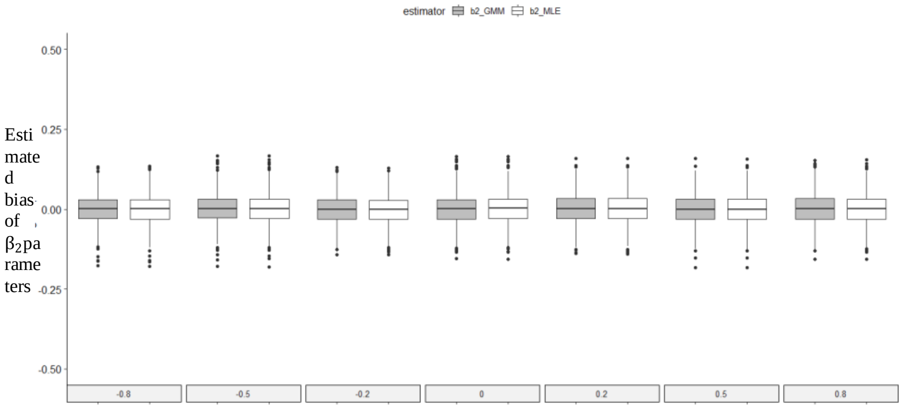
Fig. 9. Comparison of the bias values of β_2 for GMM and MLE method at $n = 30$ in matrixin matrix [a] CW and [b] SGI

Figure 10.is presented the β_2 bias at the number of observations $n = 90$.The bias value at the number of observations n = 90 is smaller than at the number of observations n = 30.The bias of the SGI matrix is smaller than CW matrix. The GMM provides smaller bias values than the MLE. Consistency can be seen from the smaller variance compared to the MLE method. The different types of spatial autocorrelations do not affect the bias value β_2.



ρ

(a)

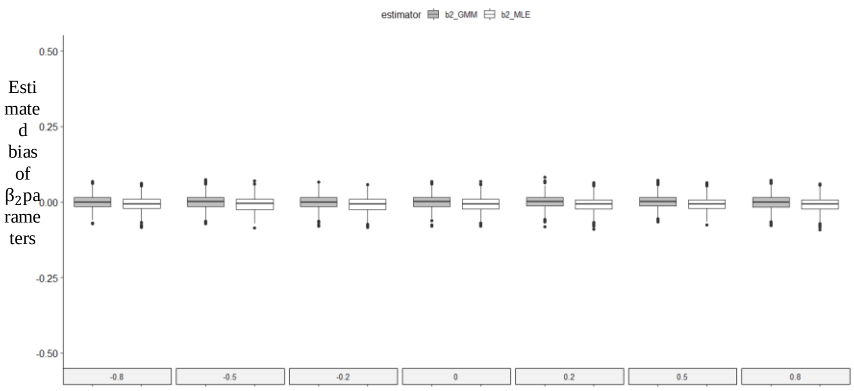


ρ

(b)

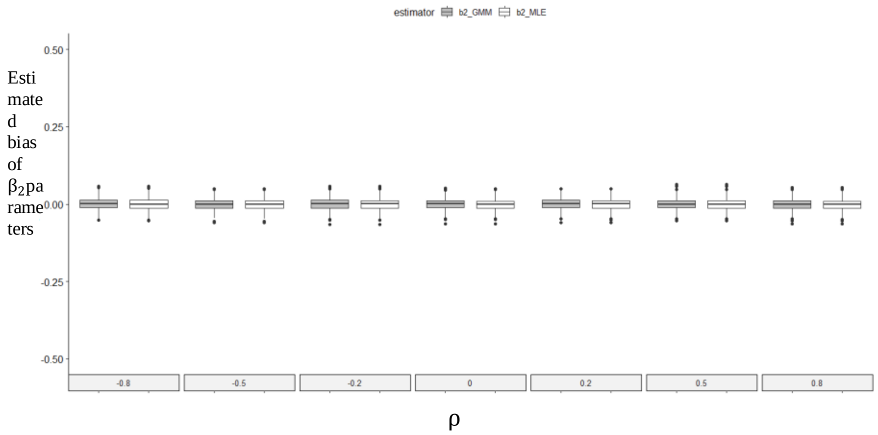
Fig. 10. Comparison of the bias values of β_2 for GMM and MLE method at $n = 90$ in the matrix(a) CW and (b) SGI

Figure11.is presented a comparison of the estimated bias value of β_2 at the number of observations n = 900 in matrix [a] CW and [b] SGI. The bias value at the number of observations n = 90 is smallest than at the number of observations n = 30 and n=90. The bias of the SGI matrix is smaller than CW matrix.The GMM provides smaller bias values than the MLE. This shows that GMM is more consistent.The difference in spatial interactions (ρ) does not significantly affect the bias value, that the type of spatial interaction (positive or negative) does not significantly affect the alleged bias value of the β_2 parameter. The two matrices do not give very different results, havesymmetrical shapes with an average of around 0 and small variations.



ρ

(a)



(b)

Fig. 11. Comparison of the bias values of β_2 for GMM and MLE method at $n = 900$ in the matrix(a) CW and (b) SGI

Based on the type of matrix, the CW matrix gives the result that the bias value of the β_1 and $\beta_2$is large, but the bias value of the ρ parameter is small. The SGI matrix gives a bias value of the β_1 and $\beta_2$is small, but the bias value of each parameter ρ is large.Based on these results, the CW matrix is better used to estimate the parameter ρ, but the SGI matrix is better used to estimate the β_1 and β_2.

The overall simulation results show that the greater the number of observations, the smaller the bias value for each method. GMM gives the bias value of each parameter which is estimated to be ρ, β_1 and β_2 smaller than MLE. Doğan and Taşpinar [7] state that theoretically, MLE is not consistent with heteroscedastic problems, and has been proven by Liu et al. [8] so that the GMM method is more appropriate to be used in estimating parameters when heteroscedastic occurs.

4Conclusion

GMM can be used in data that is known to be autoregressive in the response variable and also heteroscedastic. Based on the results of simulation data that GMM provides a smaller bias value than MLE. Heteroscedastic conditions in spatial units as well as in the form of interactions of groups of spatial units provide consistent predictor results. Different types of spatial interactions do not affect the estimated value of the bias of the parameters. This data is very appropriate to apply the estimation of the SAR model with GMM.

5 Appendix

1. Suppose that Ω is a matrix of moment functions consisting of variance and covariance that are linear and quadratic in $\boldsymbol{\varepsilon}$. For a $\mathbf{G}$ quadrilateral matrix, set $\mathrm{Diag}(\mathbf{G})=(g_{11},...,g_{nn})'$'is a vector of diagonal matrix elements $\mathbf{G}$.

$$\boldsymbol{\Omega} = \mathrm{var}(\mathbf{g}_n(\boldsymbol{\theta}_0))$$
$$= E[\mathbf{g}_n{}^{\mathbf{2}}(\boldsymbol{\theta}_0)] - (E[\mathbf{g}_n(\boldsymbol{\theta}_0)])^2$$
$$= E[\mathbf{g}_n(\boldsymbol{\theta}_0)\mathbf{g}_n{}'(\boldsymbol{\theta}_0)] - 0$$
$$= E[\mathbf{g}_n(\boldsymbol{\theta}_0)\mathbf{g}_n{}'(\boldsymbol{\theta}_0)]$$

2. Let $\mathbf{X}_n$, $\mathbf{Y}_n$ and $\mathbf{Z}_n$ be nxn matrices.$\mathbf{X}_n$ and $\mathbf{Y}_n$ have zero diagonal elements, and $\mathbf{Z}_n$ has uniformly bounded row and column sums in absolute value. Assume $\boldsymbol{\Sigma} = \mathrm{diag}\,(\sigma_1^2,....,\sigma_n^2)$.

   a. $E(\boldsymbol{\varepsilon}'_n\mathbf{X}'_n\boldsymbol{\varepsilon}'_n.\boldsymbol{\varepsilon}'_n\mathbf{Y}'_n\boldsymbol{\varepsilon}'_n) = \Sigma_{i=1}^n\Sigma_{j=1}^n\mathbf{x}_{n,ij}(\mathbf{y}_{n,ij}+\mathbf{x}_{n,ji})\boldsymbol{\sigma}_{ni}^2\boldsymbol{\sigma}_{nj}^2 = \mathrm{tr}\,(\boldsymbol{\Sigma}_n\mathbf{X}_n(\mathbf{Y}'{}_n\boldsymbol{\Sigma}_n + \mathbf{Y}_n\boldsymbol{\Sigma}_n))$

   b. $E(\varepsilon_n\boldsymbol{\Sigma}_n\boldsymbol{\varepsilon})^2 = \Sigma_{i=1}^n\mathbf{z}_{n,ii}^2[E(\boldsymbol{\varepsilon}_{ni}^2) - 3\,\boldsymbol{\sigma}_{ni}^4] + (\Sigma_{i=1}^n\mathbf{z}_{n,ii}^2\boldsymbol{\sigma}_{ni}^2) + \Sigma_{i=1}^n\Sigma_{j=1}^n\mathbf{c}_{n,ij}^2(\mathbf{z}_{n,ij}\mathbf{z}_{n,ji})\boldsymbol{\sigma}_{ni}^2\boldsymbol{\sigma}_{nj}^2$
   $= \Sigma_{i=1}^n\mathbf{z}_{n,ii}^2[E(\boldsymbol{\varepsilon}_{ni}^4) - 3\,\boldsymbol{\sigma}_{ni}^4] + \mathrm{tr}^2(\Sigma_n\mathbf{Z}_n) + \mathrm{tr}\,(\boldsymbol{\Sigma}_n Z_n\mathbf{Z}'_n\boldsymbol{\Sigma}_n + \boldsymbol{\Sigma}_n\mathbf{Z}_n\boldsymbol{\Sigma}_n\mathbf{Z}_n)$

   c. $\mathrm{Var}(\boldsymbol{\varepsilon}_n\mathbf{Z}_n\varepsilon) = \Sigma_{i=1}^n\mathbf{z}_{n,ii}^2[[E(\boldsymbol{\varepsilon}_{ni}^4) - 3\,\boldsymbol{\sigma}_{ni}^4] + \Sigma_{i=1}^n\Sigma_{j=1}^n\mathbf{z}_{n,ij}(\mathbf{z}_{n,ij}\mathbf{z}_{n,ji})\boldsymbol{\sigma}_{ni}^2\boldsymbol{\sigma}_{nj}^2$
   $= \Sigma_{i=1}^n\mathbf{z}_{n,ii}^2[E(\boldsymbol{\varepsilon}_{ni}^4) - 3\,\boldsymbol{\sigma}_{ni}^4] \quad + \mathrm{tr}\,(\boldsymbol{\Sigma}_n\mathbf{Z}_n\mathbf{Z}'_n\boldsymbol{\Sigma}_n + \boldsymbol{\Sigma}_n\mathbf{Z}_n\boldsymbol{\Sigma}_n\mathbf{Z}_n)$

6 References

[1]B. S. Yandell and L. Anselin. 2006. *Spatial Econometrics: Methods and Models.*, vol. 85, no. 411.
[2]J. LeSage. 1999. *Spatial Econometrics Toolbox.A companion to Theor. Econom.*
[3]Fotheringham AS, Brunsdon C, Charlton M. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* England (GB): John Wiley & Sons, LTD.
[4]K, Ord. 1975. *Estimation methods for models of spatial interaction.J. Am. Stat. Assoc.* vol. 70. no. 349. pp. 120–126.
[5]H. H. Kelejian and I. R. Prucha. 1999. *A generalized moments estimator for the autoregressive parameter in a spatial model.* Int. Econ. Rev. (Philadelphia), vol. 40. no. 2. pp. 509–533.
[6]H. H. Kelejian and I. R. Prucha. 2010. *Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances.* J. Econom. vol. 157. no. 1. pp. 53–67.
[7]O, Doğan and S, Taşpinar. 2014. *Spatial autoregressive models with unknown heteroskedasticity: A comparison of Bayesian and robust GMM approach.* Reg. Sci. Urban Econ. vol. 45. no. 1. pp.1–21.
[8]Liu X, Lee LF, Bollinger CR. 2010. An efficient GMM estimator of spatial autoregressive models. *J. Econom.* 159(2) 303–319. doi:10.1016/j.jeconom.2010.08.001
[9]Lee L. 2006. Efficient GMM Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances. Science (80).