

Simulation Study for Comparison of Maximum Likelihood and Bayesian Method in Spatial Autoregressive Models with Heteroskedasticity

Fitri Ramadhini¹, Anik Djuraidah², Aji Hamim Wigena³
{ramadhini_fitri@apps.ipb.ac.id¹, anikdjuraidah@apps.ipb.ac.id², aji-hw@apps.ipb.ac.id³}

Department of Statistics, IPB University, Bogor, 16680, Indonesia^{1,2,3}

Abstract. Generally spatial regression considers only one of the spatial effects, namely spatial dependence or heteroskedasticity between areas. Spatial autoregressive (SAR) models take only into account the dependence on the response variable. Most of SAR estimators are valid if there is no violation in the error assumption. Estimation of SAR parameters with heteroskedasticity using maximum likelihood (ML) method gives bias and inconsistent estimators. An alternative method that can be used is Bayesian method. Bayesian method solves heteroskedasticity by modeling the structure of variance-covariance matrix. Simulation data is used to evaluate the Bayesian method in estimating parameters of SAR model with heteroskedasticity. The results indicate that Bayesian method provides bias parameter estimates relatively small and consistent compared to the ML method.

Keywords: Bayesian, heteroskedasticity, maximum likelihood, spatial dependence, spatial regression

1 Introduction

Area data is the observations in certain geographical areas. This data is obtained from the aggregation of several observation values that correspond to the specified area. Area data has spatial effects, namely spatial dependence and heterogeneity [1]. Spatial dependence can be described by linear regression, which is autoregressive on the response, error, or combination of the two [2] [3]. Models with dependence on response are called spatial autoregressive models (SAR), models with dependence on error are called spatial error models (SEM), and the models with combination of the two are called the generalized spatial models (GSM). Spatial heterogeneity can be modeled by geographically weighted regression (GWR) [4].

The estimation of SAR parameters using maximum likelihood (ML) method [5] depends on the assumption that errors are normal and homogeneous. Other estimation of SAR, those are two-stage least squares (S2SLS) method [3] with the estimators are consistent and normal asymptotic and generalized method of moments (GMM) [6] that produces efficient estimators with easier computing than ML method. However, these estimators are inconsistent if there is heteroskedasticity [7] [8].

Generally, spatial regression only considers one of the spatial effects. In its application, the spatial heterogeneity or heteroskedasticity problem often occurs in cross section data. Heteroskedasticity can occur in aggregation data because of a process of averaging data with many different observations at the time of aggregation [9] [10]. This causes the ML estimator to be invalid. This problem can be solved by modifying the GMM method [11] and Bayesian approach [12] [1]. Bayesian method solves heteroskedasticity by modeling the structure of variance-covariance matrix that allows flexible models for each distribution of spatial data with high accuracy [13].

This study aims to evaluate the bias of the parameters estimator using SAR model with Bayesian method in the condition of heteroskedasticity. The evaluation is based on simulation that compared to ML method. The smaller the bias values, the estimation of parameters will be better and consistent.

2 Material and Method

2.1 SAR Model with Bayesian Approach

The SAR model with heteroskedasticity can be expressed as follows:

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}; \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \quad (1)$$

with \mathbf{y} is an $n \times 1$ vector of response variables, \mathbf{X} is an $n \times p$ matrix of explanatory variable, $\boldsymbol{\beta}$ is a $p \times 1$ vector of coefficient regression parameter, \mathbf{W} is an $n \times n$ spatial weight matrix, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of no autocorrelation error, and ρ is a spatial lag coefficient or spatial autocorrelation.

Estimation using Bayesian method assumes that $\boldsymbol{\Sigma} = \sigma_0^2 \mathbf{V}$, with $\mathbf{V} = \text{diag}(v_1, v_2, \dots, v_n)$, and $v_i = \frac{\sigma_i^2}{\sigma_0^2}$ for $i=1, \dots, n$. This assumption indicates that the heteroskedasticity specification has two components: (i) the constant component σ_0^2 , and (ii) a component v_i that varies between observations [8]. SAR model in equation (1) can be written as:

$$\mathbf{y} = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \quad (2)$$

with $\boldsymbol{\varepsilon}^* = (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon}$, and $\boldsymbol{\varepsilon}^* \sim \text{MVN}(\mathbf{0}, \sigma_0^2 \mathbf{V} ((\mathbf{I} - \rho \mathbf{W}^T)^{-1} (\mathbf{I} - \rho \mathbf{W})^{-1}))$. Let $\mathbf{v} = (v_1, \dots, v_n)'$ is an $n \times 1$ vector and the likelihood function of this model can be written as:

$$L(\mathbf{y} | \boldsymbol{\theta}, \sigma_0^2, \mathbf{v}) = (2\pi)^{-\frac{n}{2}} (\sigma_0^2)^{-\frac{n}{2}} \prod_{i=1}^n v_i^{-\frac{1}{2}} |\mathbf{S}(\rho)| \times \exp \left[-\frac{1}{2} (\mathbf{S}(\rho) \mathbf{y} - \mathbf{X} \boldsymbol{\beta})' (\sigma_0^2 \mathbf{V})^{-1} (\mathbf{S}(\rho) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right] \quad (3)$$

with $\mathbf{S}(\rho) = (\mathbf{I} - \rho \mathbf{W})$ and $\boldsymbol{\theta}$ is hyperparameter for ρ and $\boldsymbol{\beta}$. The posterior function for the Bayesian method with prior distribution assumptions can be stated as:

$$p(\boldsymbol{\theta}, \sigma_0^2, \mathbf{v}, r | \mathbf{y}) \propto L(\mathbf{y} | \boldsymbol{\theta}, \sigma_0^2, \mathbf{v}) \pi(\boldsymbol{\beta}) \pi(\sigma_0^2) \pi(\mathbf{v}) \pi(r) \pi(\rho)$$

$$\begin{aligned}
& \propto (2\pi)^{-\frac{n}{2}} (\sigma_0^2)^{-\frac{n}{2}} \prod_{i=1}^n v_i^{\frac{1}{2}} |\mathbf{S}(\rho)| \\
& \quad \times \exp \left[-\frac{1}{2} (\mathbf{S}(\rho) \mathbf{y} - \mathbf{X} \boldsymbol{\beta})' (\sigma_0^2 \mathbf{V})^{-1} (\mathbf{S}(\rho) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right] \\
& \quad \times |\mathbf{T}|^{\frac{1}{2}} \exp \left(-\frac{1}{2} \boldsymbol{\beta}' \mathbf{T}^{-1} \boldsymbol{\beta} \right) \\
& \quad \times (\sigma_0^2)^{-(a+1)} \exp \left(\frac{-b}{\sigma_0^2} \right) \\
& \quad \times \left(\frac{r}{2} \right)^{\frac{nr}{2}} \left[\Gamma \left(\frac{r}{2} \right) \right]^{-n} \prod_{i=1}^n v_i^{-\frac{r+2}{2}} \exp \left\{ -\frac{r}{2v_i} \right\} \\
& \quad \times r^{m-1} \exp \left\{ -\frac{r}{k} \right\} \times p(\rho)
\end{aligned} \tag{4}$$

2.2 Data

This study uses simulation data to evaluate Bayesian method. Simulation data was generated by the following processes:

1. The model is SAR without intercept, $\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with the number of observations $n=30, 90, 900$ and \mathbf{y} is $n \times 1$ vector of response variable.
2. Explanatory variables $\mathbf{X}=(\mathbf{x}_1, \mathbf{x}_2)$ and the coefficients $\boldsymbol{\beta}=(\beta_1, \beta_2)'$ with $\boldsymbol{\beta}=(1, 2)'$. Explanatory variables \mathbf{x}_1 dan \mathbf{x}_2 are $n \times 1$ vector and generated from standard normal distribution, $N(0, 1)$. This step is repeated 1000 times.
3. The autoregressive coefficients on the SAR models are $\rho = (-0.8, -0.5, -0.2, 0.0, 0.2, 0.5, 0.8)$ to describe spatial autocorrelation.
4. Weight matrices \mathbf{W} were generated by the following steps:
 - a. The first matrix based on circular world matrix [14] with the following steps:
 - 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 $i=2, \dots, n/3$. While for the first row all elements are zero except in position $(1, 2)$ and $(1, n)$. The number of neighbors in these rows are 2.
 - 2) The second $n/3$ rows of each element are zero except in positions $(j, j \pm r)$, $j = \frac{n}{3} + 1, \dots, \frac{2n}{3}$; $r=1, 2, \dots, 5$. The number of neighbors in these rows are 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 $j=2n/3+1, \dots, n-1$. 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 are 2.
 - 4) This matrix then is row standardized.
 - b. The second matrix is the row standardized spatial weight matrix based on the small group interaction [15]. This matrix is a diagonal block matrix which each block describes group interaction. Block size comes from uniform distribution, $U(3, 20)$. Suppose $\{g_1, \dots, g_G\}$ as a group with G is the number of group. The group size is m_i , with $i=1, \dots, G$. The block from group i is $\mathbf{B}_i = \frac{1}{m_i-1} (\mathbf{l}_{m_i} \mathbf{l}_{m_i}' - \mathbf{I}_{m_i})$, with \mathbf{l}_{m_i} is $m_i \times 1$ vector of 1. The weight matrix is $\mathbf{W}_n = \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_G)$.
5. The heteroskedastic pattern is generated using the number of neighboring units [14].

Suppose h_i is the number of neighbors from the i unit, then the error $\boldsymbol{\varepsilon}$ with element ε_i is generated by the following technique:

$$\varepsilon_i = \sigma_i \varphi_i \quad (5)$$

$$\sigma_i = 0.5 \frac{h_i}{\sum_{j=1}^n h_j/n} \text{ with } \varphi_i \sim N(0,1) \quad (6)$$

2.3 Steps of Analysis

The steps of analysis in this study are as follows:

1. Generating simulation data
2. Estimating parameters with Bayesian method, which is defining the distribution of priors for each hyperparameter then finding the posterior distribution using Markov chain Monte Carlo (MCMC). The assumptions of the priors in the SAR model are [16]:
 - β_i from normal distribution $\beta_i \sim N(c, T)$, for $i=1, \dots, n$, with T is a large value of variance.
 - $\sigma_0^2 \sim IG(a, b)$ where $IG(a, b)$ is the inverse gamma distribution with shape $a > 0$ and scale $b > 0$.
 - $\frac{r}{v_i} \sim \chi^2(r)$, for $i=1, \dots, n$, with degree of freedom r is from gamma distribution, $r \sim G(m, k)$.
 - $\rho \sim U(\frac{1}{\lambda_{\min}}, \frac{1}{\lambda_{\max}})$ with $U(\cdot)$ is uniform distribution where λ_{\min} and λ_{\max} are minimum and maximum eigen values of \mathbf{W} .
3. Evaluating the bias ($\hat{\beta}$) of the Bayesian method compared to the ML method.

$$\text{Bias}(\hat{\beta}_i) = E(\hat{\beta}_i) - \beta_i \quad (7)$$

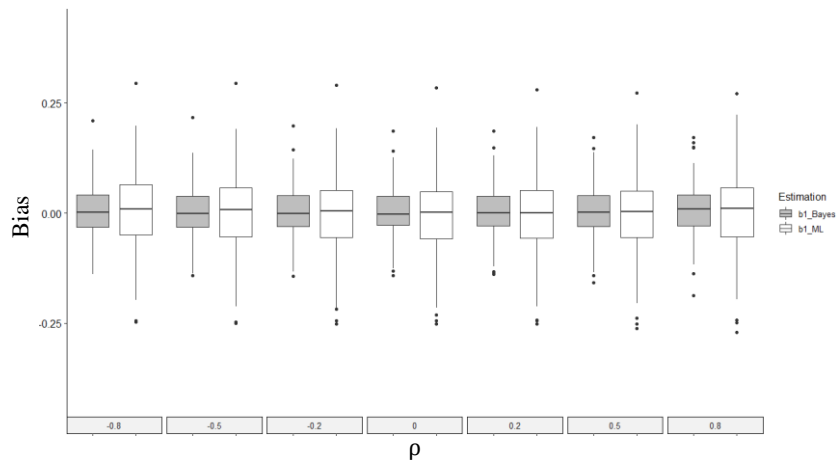
with β_i is the parameter coefficient β of i^{th} simulation, $\hat{\beta}_i$ is the value of the estimated coefficient β of i^{th} simulation.

3 Result and Discussion

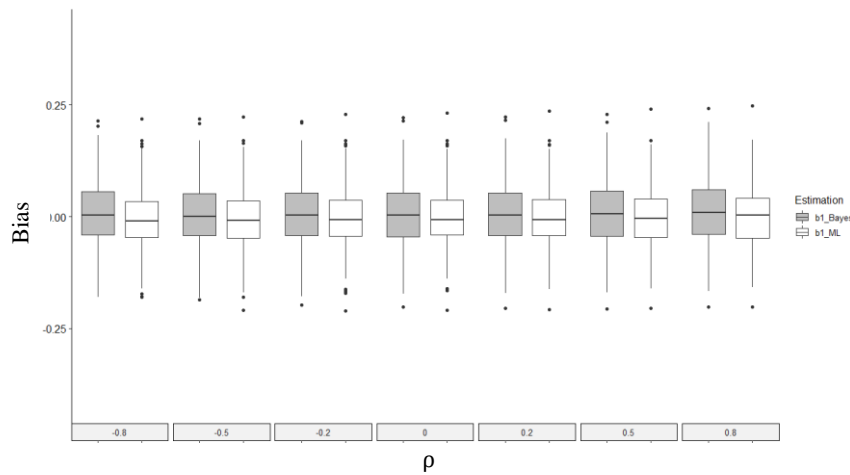
The Bayesian and maximum likelihood method are evaluated based on bias values of the parameter estimator. Each bias values for the Bayesian method and maximum likelihood (ML) are presented in the form of boxplots in **Figure 1** to **Figure 9** for the circular world / CW (a) matrix and small group interaction / SGI (b). The bias values of the Bayesian method are compared with the ML method. Boxplot of the bias values for Bayesian method is presented next to the ML method bias values for easy comparison. The gray boxplot is a representation of the bias values for the Bayesian method. While the white boxplot is a representation of the bias values for the ML method. A small bias value indicates a good estimate of the parameter, while a small bias variation indicates consistent parameter estimate.

3.1 The result for β_1 estimator

Figure 1 shows the simulation results at $n = 30$ for the β_1 estimator. Bayesian method in the results of the simulation with the CW matrix at $n = 30$ gives the bias values and bias variations of the estimator β_1 smaller than the ML method. The bias values in the first matrix do not show quite different results for each spatial autocorrelation. These results are different for simulations with the SGI matrix. The Bayesian method in the SGI matrix gives greater bias values than the ML method but have mean of around 0, while the ML method tends to be under estimate.



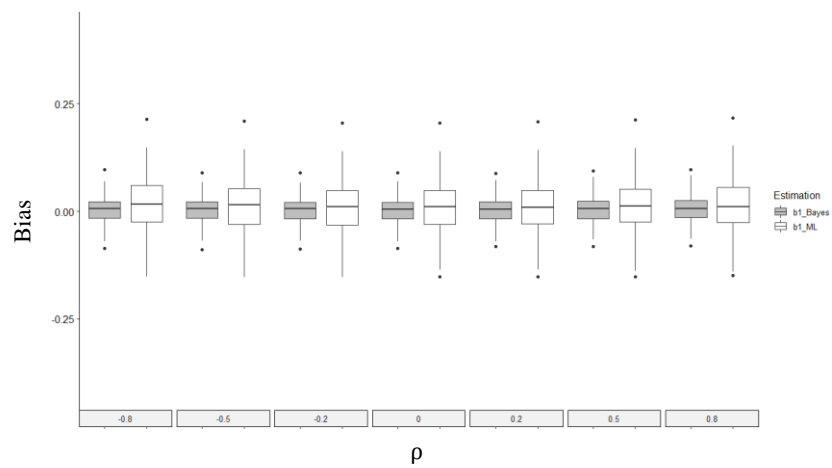
(a)



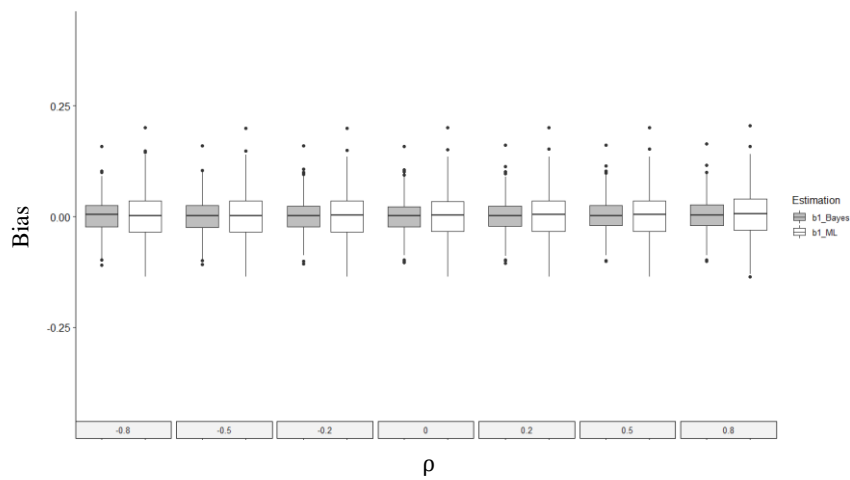
(b)

Fig. 1. Comparison of the bias values of the Bayesian and ML method at $n = 30$ for the β_1 estimator on the (a) CW and (b) SGI matrices.

The simulation results at $n = 90$ for the estimator β_1 are shown in **Figure 2**. The Bayesian method at $n = 90$ gives smaller bias values than the ML method for each matrix. The bias values at $n = 90$ are also smaller than $n = 30$, this shows that the Bayesian method is more consistent. The consistency of the Bayesian method can also be seen from the smaller variances compared to the ML method. The bias values for the CW and SGI matrices tend to be the same for each spatial autocorrelation.



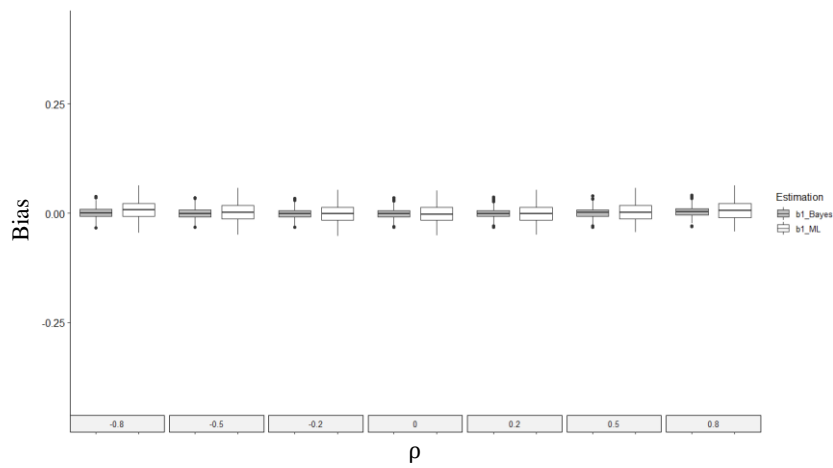
(a)



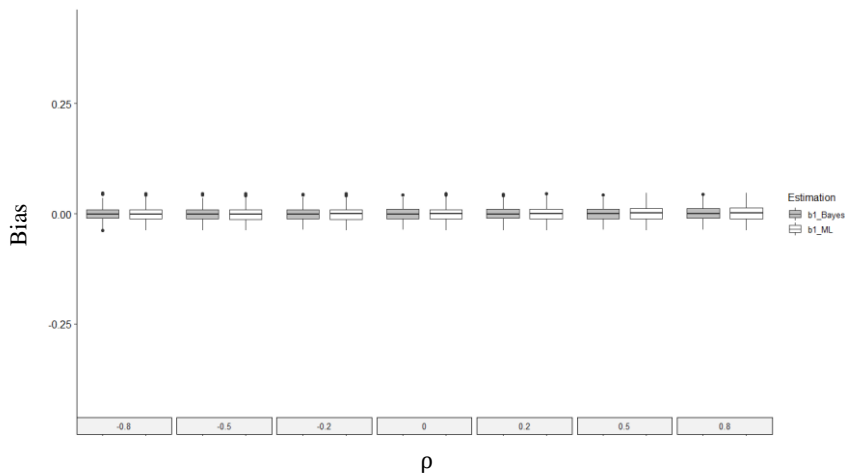
(b)

Fig. 2. Comparison of the bias values of the Bayesian and ML method at $n = 90$ for the β_1 estimator on the (a) CW and (b) SGI matrices.

Figure 3 shows the simulation results at $n = 900$ for the β_1 estimator. The simulation results at $n = 900$ show that the greater the number of observations, the smaller the bias values for each method. The Bayesian method gives smaller bias values than the ML method on the CW matrix. The SGI matrix does not give quite different results, but the Bayesian method has symmetric distributions with the mean of around 0 and the variances are quite small while the ML method has distribution that extend to the right. That is, the ML method on the SGI matrix tends to provide estimates that over estimate.



(a)

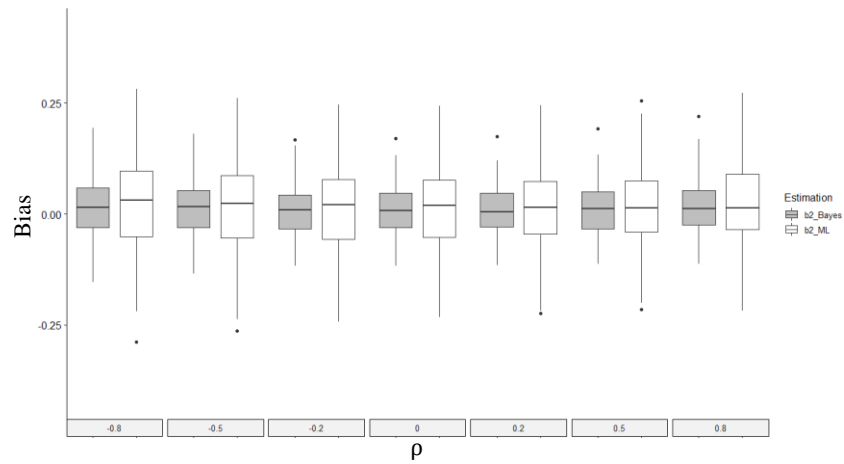


(b)

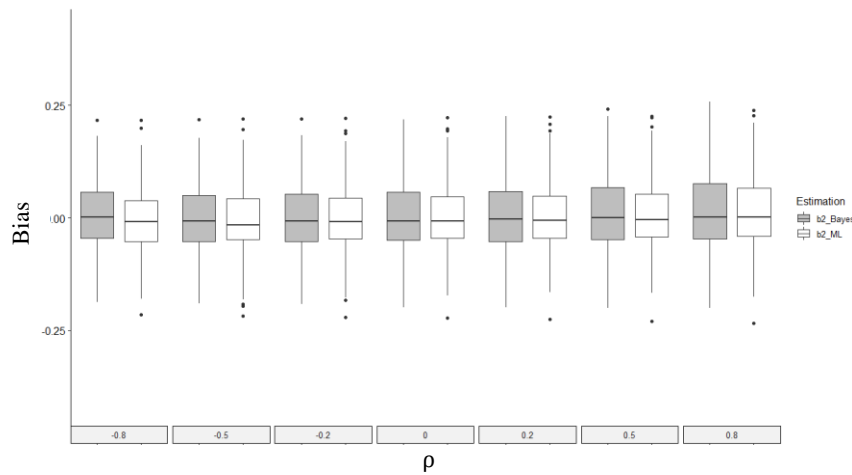
Fig. 3. Comparison of the bias values of the Bayesian and ML method at $n = 900$ for the β_1 estimator on the (a) CW and (b) SGI matrices.

3.2 The result for β_2 estimator

The simulation results at $n = 30$ for the β_2 estimator are presented in **Figure 4**. Bayesian method in the simulation results with the CW matrix on the $n = 30$ provides smaller β_2 estimator bias values than the ML method. These results are different for simulations with the SGI matrix. The Bayesian method in the SGI matrix gives greater bias than the ML method but has mean of around 0 while the ML method tends to under estimate.



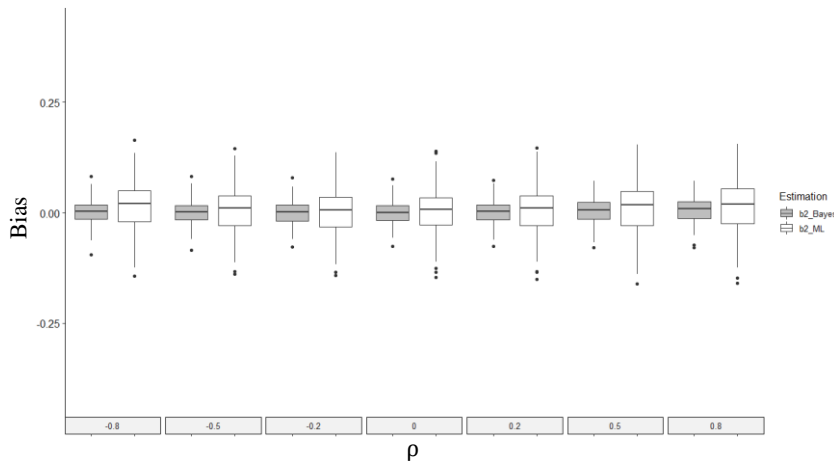
(a)



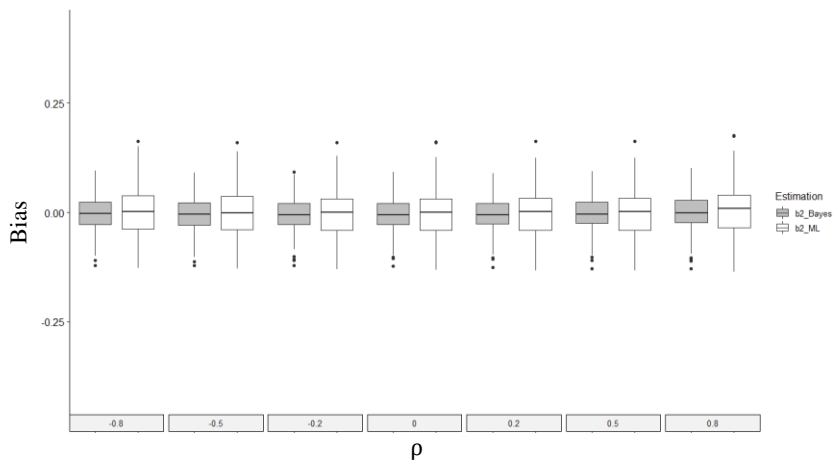
(b)

Fig. 4. Comparison of the bias values of the Bayesian and ML method at $n = 30$ for the β_2 estimator on the (a) CW and (b) SGI matrices.

Figure 5 shows the simulation results at $n = 90$ for the β_2 estimator. The β_2 estimator of the Bayesian method at $n = 90$ with the CW and SGI matrices gives smaller bias values than the ML method. These bias values are also smaller than at $n = 30$ with mean of around 0 and smaller variance. The distribution patterns for each spatial autocorrelation do not show quite different results.



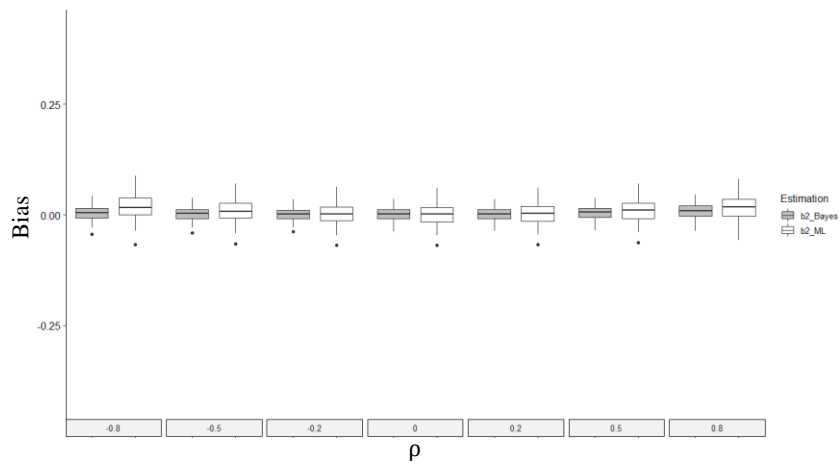
(a)



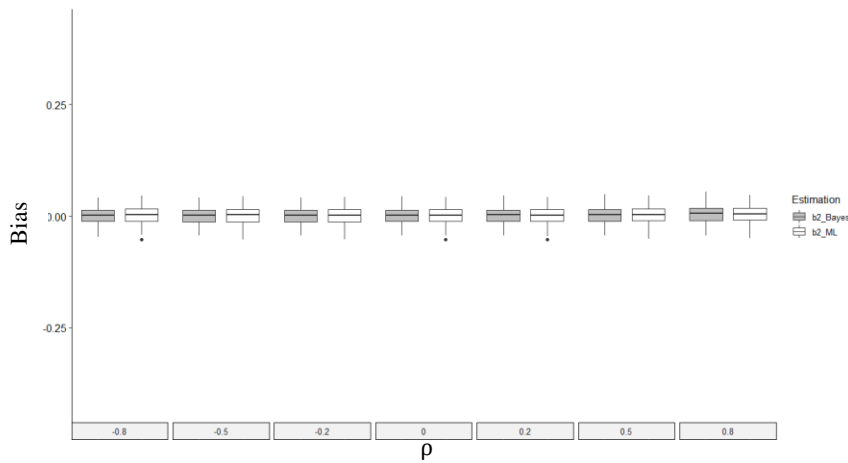
(b)

Fig. 5. Comparison of the bias values of the Bayesian and ML method at $n = 90$ for the β_2 estimator on the (a) CW and (b) SGI matrices.

Figure 6 shows the simulation results at $n = 900$ for the β_2 estimator that the greater the number of observations, the smaller the bias values for each method. The Bayesian method gives smaller bias values compared to the ML method on both the CW matrix and the SGI matrix. On the CW matrix, the Bayesian method has symmetric patterns that have mean of around 0 with small variances except that $\rho = 0.8$ tends to provide over-estimating estimators. The bias values in the SGI matrix do not show different results for each method and spatial autocorrelation.



(a)

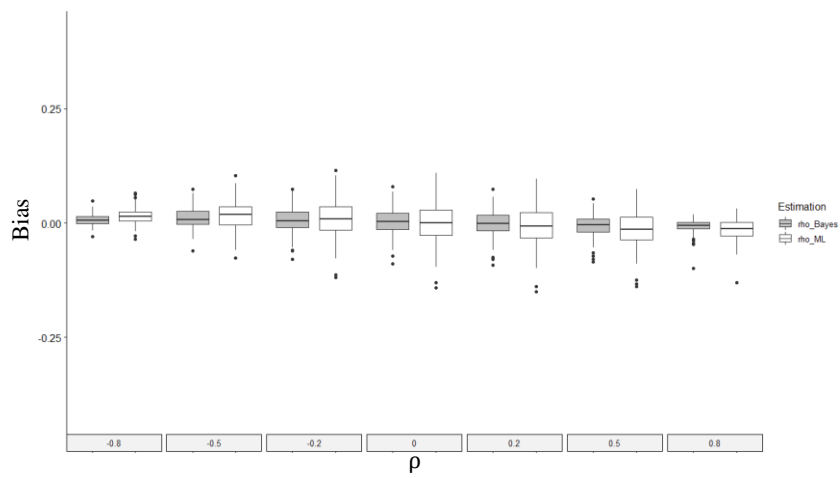


(b)

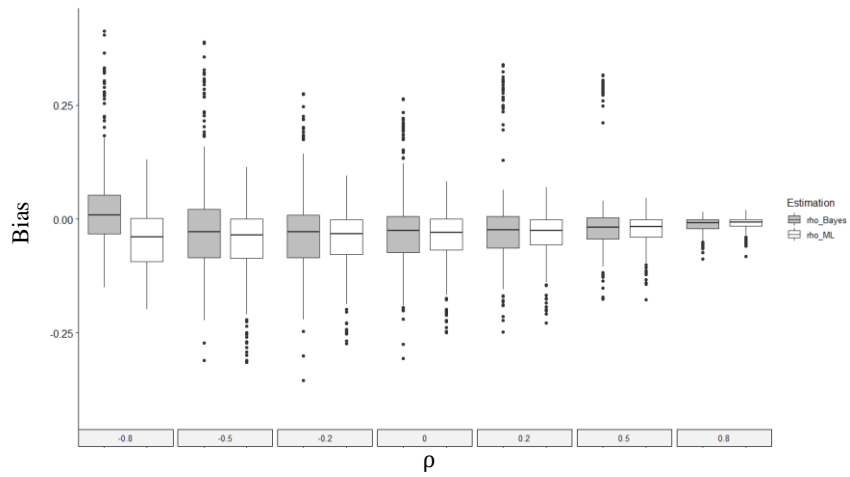
Fig. 6. Comparison of the bias values of the Bayesian and ML method at $n = 900$ for the β_2 estimator on the (a) CW and (b) SGI matrices.

3.2 The result for ρ estimator

The Bayesian method at $n = 30$ for ρ estimator is shown in **Figure 7**. The simulation results show that the CW matrix gives smaller bias values compared to the ML method. The bias values decrease with the amount of spatial autocorrelation (either negative or positive). In the SGI matrix, the bias values of the Bayesian method are greater than the ML method and tend to under estimate except at $\rho = -0.8$ which is over estimate. The bias values on the SGI matrix show better results when spatial autocorrelation is positive and get smaller with the magnitude of negative spatial autocorrelation.



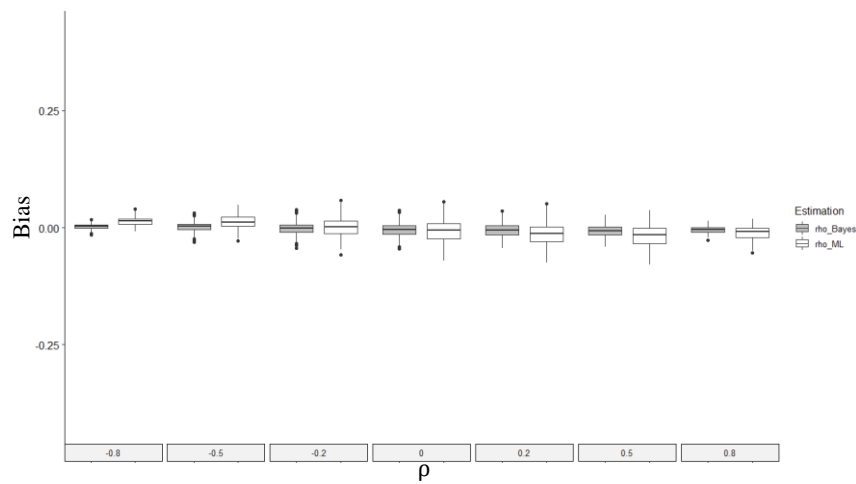
(a)



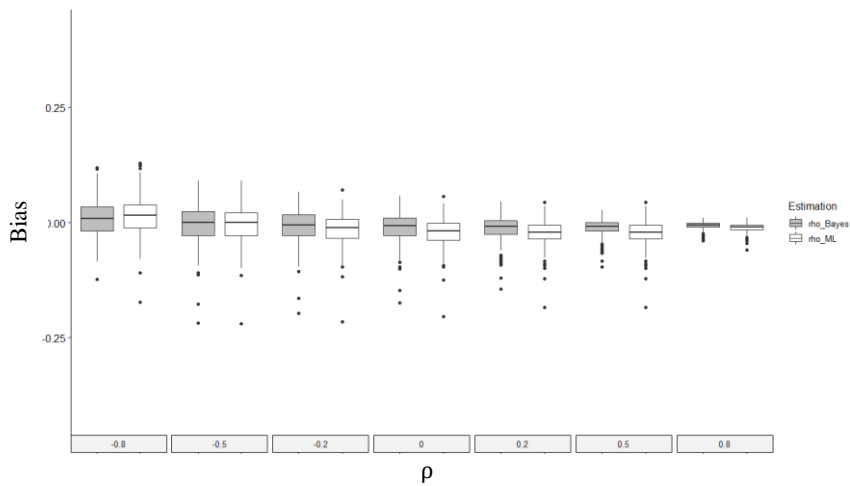
(b)

Fig. 7. Comparison of the bias values of the Bayesian and ML method at $n = 30$ for the ρ estimator on the (a) CW and (b) SGI matrices.

The simulation results at $n = 90$ for ρ estimator are shown in **Figure 8**. Based on **Figure 8** on the CW matrix shows the same pattern at $n = 30$ (**Figure 7**) but with smaller bias values. The Bayesian method gives smaller bias values compared to the ML method. The bias values decrease with the amount of spatial autocorrelation (either negative or positive). On the SGI matrix, the bias values of the Bayesian method tend to be the same as the ML method but will get smaller as the magnitude of the positive spatial autocorrelation.



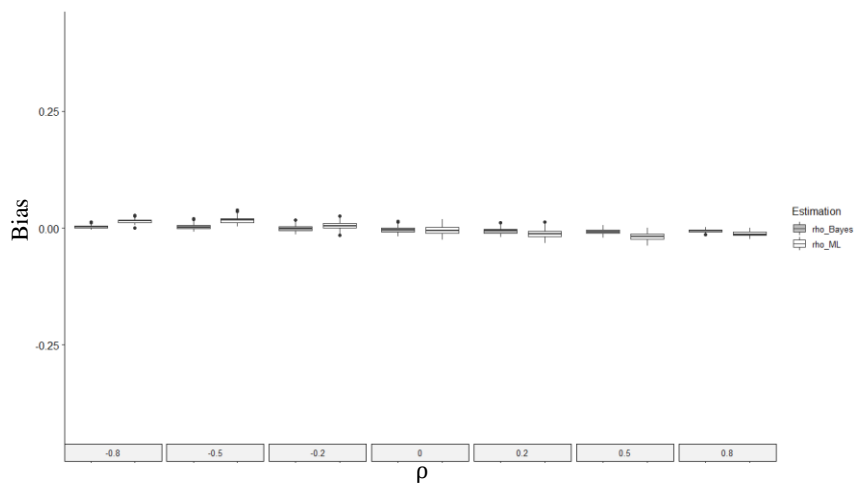
(a)



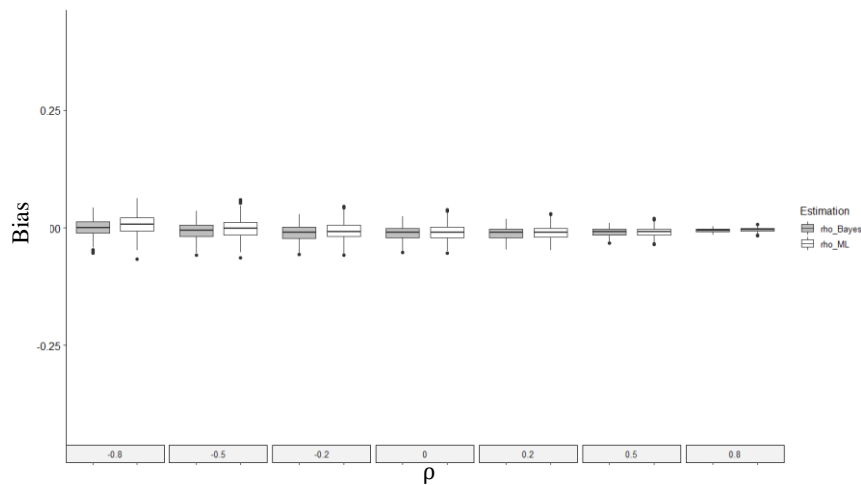
(b)

Fig. 8. Comparison of the bias values of the Bayesian and ML method at $n = 90$ for the ρ estimator on the (a) CW and (b) SGI matrices.

Figure 9 shows the simulation results at $n = 900$ for ρ estimator. The simulation results at $n = 900$ for ρ estimator indicate that the greater the number of observations, the smaller the bias values will be for each method. The Bayesian method give smaller bias values compared to the ML method on the CW matrix. The bias values decrease with the amount of spatial autocorrelation (either negative or positive). In the SGI matrix, the bias values of the Bayesian method are smaller than the ML method. These values decrease with the amount of positive spatial autocorrelation.



(a)



(b)

Fig. 9. Comparison of the bias values of the Bayesian and ML method at $n = 900$ for the ρ estimator on the (a) CW and (b) SGI matrices.

Based on the simulation results each number of observations for the two matrices shows that overall the Bayesian method gives the bias values of the estimated parameters for β_1 , β_2 , and ρ are smaller than the ML method with symmetric distribution patterns with mean around 0 and small variances. In the SGI matrix with $n = 30$ the Bayesian method has greater bias values than the ML method for each parameter estimator. However, the bias values of the Bayesian method are smaller than the ML method at 90 and 900 observations. This can occur because the heteroskedasticity pattern in the SGI matrix is not sufficiently visible for small observations. This shows that the Bayesian method is more consistent than the ML method. Theoretically the ML method is inconsistent with heteroskedastic problems [8].

4 Conclusion

The Bayesian method can be used for SAR modeling containing heteroskedasticity. Evaluations on the Bayesian method give smaller bias values and smaller bias variances compared to the maximum likelihood method, indicating the consistency of the Bayesian method. The bias values for the estimated regression coefficient parameter decrease as the number of observations increases, while the estimator of the spatial autocorrelation parameter is most consistent with Bayesian on both large and small numbers of observations.

Acknowledgments

This work is fully supported by Kemenristek DIKTI (Kementerian Riset Teknologi dan Pendidikan Tinggi) of Indonesia.

References

- [1] Anselin, L: *Spatial Econometrics: Methods and Models*. Dordrecht (NL): Kluwer Academic Publ (1988)
- [2] LeSage JP: *Spatial Econometrics*. Toledo (ES): Department of Economics University of Toledo (1998)
- [3] Anselin L: *Spatial Econometrics*. Dallas (GB): University of Texas (1999)
- [4] Fotheringham AS, Brunson C, Charlton M: *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester (GB): J Wiley (2002)
- [5] Ord K: Estimation methods for models of spatial interactions. *Journal of the American Statistical Association*. 70(349): 120-126 (1975)
- [6] Kelejian HH, Prucha IR: A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40(2):509–533 (1999)
- [7] Doğan O, Taşpınar S: GMM estimation of spatial autoregressive models with moving average disturbances. *Reg. Sci. Urban Econ*. 43(6):903–926. doi:10.1016/j.regsciurbeo.2013.09.002 (2013)
- [8] Doğan O, Taşpınar S: Spatial autoregressive models with unknown heteroskedasticity: A comparison of Bayesian and robust GMM approach. *Reg. Sci. Urban Econ*. 45(1):1–21 (2014)
- [9] Griffiths WE: *A Companion to Theoretical Econometrics*. New Jersey (US): Blackwell Publ (2007)
- [10] Lee LF, Liu X: Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory*. 26(1):187–230 (2010)
- [11] Kelejian HH, Prucha IR: Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*. 157(1): 53-67 (2010)

- [12] Anselin L: A note on small sample properties of estimators in a first-order spatial autoregressive model. *Environ Plan A*. 14(8):1023–1030 (1982)
- [13] Koop G: *Bayesian Econometrics*. Chichester (GB): J Wiley (2003)
- [14] Arraiz I, Drukker DM, Kelejian HH, Prucha IR: A spatial cliff-ord-type model with heteroskedastic innovations: small and large sample results. *J Reg Sci*. 50(2):592–614 (2010)
- [15] Lin X, Lee LF: GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*. 157(1): 34-52 (2010)
- [16] LeSage JP, Pace RK: *Introduction to Spatial Econometrics*. New York (US): Taylor and Francis Group (2009)