Simulation study for comparison of spatial autoregressive probit estimation methods

F Novkaniza¹, A Djuraidah¹*, A Fitrianto¹ and I M Sumertajaya¹

¹Department of Statistics, IPB University, Bogor 16680, Indonesia
²Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia

*E-mail: anikdjuraidah@gmail.com

Abstract. One of probit model variant with spatial dependent is spatial autoregressive (SAR) probit model. In SAR probit model, the spatial dependence structure adds complexity to the estimation of parameters. There are four methods for estimating the parameter of SAR probit model; maximum likelihood, Bayes, linearized GMM, and conditional approximate likelihood. The purpose of this article is to choose the best estimation method from four methods describes above using some extensive simulation which can handle sample sizes with large observations and various value of spatial lag coefficient, provided the spatial weight matrix is in an inconvenient sparse form, as is for large data sets, where each observation neighbors only a few other observations. The best estimation method is chosen based on the shortest confidence interval for the mean of SAR probit estimation, lowest bias, and Root Mean Square Errors (RMSE) of prediction. It was found that conditional approximate likelihood method was the best among the four methods concerning confidence interval and bias, yet regarding estimating RMSE, maximum likelihood estimation performed better. Maximum likelihood, Bayes, and conditional approximate likelihood method were better than linearized GMM in SAR probit parameter estimation for large dataset.

1. Introduction

Spatial dependence is the measure of the degree to which one object correlated to other nearby objects. It plays an essential role in a wide range of empirical economic studies (Yu, 2017), such as regional and urban economics, environmental and public health (Jerret et al., 2010), social networks (Wong et al., 2006) and agricultural (Holloway et al., 2002). Ignoring the spatial dependence in the data may lead to inefficient, and or biased and inconsistent estimation (Anselin and Florax, 2013). Spatial regression is the formation of a regression model using location data. In spatial regression, observations at a site depend on observations located in other adjacent locations, so that there are spatial autocorrelation and spatial heterogeneity. In the presence of spatial autocorrelation in a model, a least square method cannot be employed to estimate the model parameters due to biased and inconsistent estimators. In many applications, we have to model dependent variables that reflect binary outcomes generated by spatially dependent processes. Spatial dependence in binary outcomes result in a situation were observed at one location are similar to others at nearby locations. Holloway, Shankara, and Rahman (2002) show that binary outcomes regarding the adoption of an agricultural program by Bangladeshi rice producers exhibited spatial dependence. One of probit variant models which focused on spatial autoregressive is Spatial Autoregressive (SAR) probit model. As well as spatial autocorrelation and if data varies spatially, it is reasonable to think that variances may vary.
also. Models typically imply heteroscedasticity and autocorrelation, but some specifications include autocorrelation without heteroscedasticity. Heteroscedasticity can be a potential problem in a spatial model with discrete dependent variables and makes the standard probit estimator inconsistent.

Probit models with spatial dependencies were first studied by McMillen (1992), and he proposed an EM algorithm to produce consistent maximum likelihood estimates for the model. In SAR probit model, the spatial dependence structure adds complexity to the estimation of parameters. The main assumption of the model is that the distribution of errors is known and is often assumed to be normal. Parameter estimation using a full maximum likelihood method is problematic because the likelihood function involves n integrals, where n is the sample size. To avoid the direct calculation of n-dimensional integration, several estimators have been proposed that can produce consistent estimates when data are spatially autocorrelated and heteroscedastic (e.g., Beron and Vijverberg 2000; Bhat and Guo 2004; Bolduc et al. 1997; Case 1992; Fleming 2002; Klier and McMillen 2008; LeSage 2000; McMillen 1992; Pinkse and Slade 1998). Case (1992), McMillen (1992) and Pinkse and Slade (1998) proposed estimators for the parameter of SAR probit model who becomes infeasible for large samples because they require the inversion of \( n \times n \) matrices. LeSage (2000) used Bayesian estimates through the Markov Chain Monte Carlo and Gibbs sampling, which sampled sequentially from the complete conditional distribution for all parameters. Klier and McMillen (2008) have proposed a linearized version of the GMM estimator that avoids the infeasible problem of inverting \( n \times n \) matrices when employing large samples and show that standard GMM reduces to a nonlinear two-stage least squares problem.

Another method for estimating SAR probit parameter is conditional approximate likelihood estimation by Martinetti and Geniaux (2017). They proposed an estimator that relies on the approximation of the likelihood function, that follows a multivariate normal distribution which parameters depend on the spatial structure of the observations. This approximation is inspired by the univariate conditioning procedure proposed by Mendell and Elston (1974), with some modifications to improve the prediction accuracy and speed of computation. The conditional estimator relies on the approximation of the likelihood function, that follows a multivariate normal distribution which parameters depend on the spatial structure of the observations. Martinetti and Geniaux (2017) mentioned that this likelihood approximation method is the best estimation method that can handle large sample sizes in which many neighborhoods are much less than large sample sizes.

The purpose of the study is to determine the best method among the four methods described above: maximum likelihood, Bayes, linearized GMM, and conditional approximate likelihood method for estimating the parameter of SAR probit model. It is important to choose estimation of SAR probit model which leads a better consistency and efficiency by considering some important aspects such as confidence intervals of the estimates, bias, and RMSE of prediction to choose the best method. In this article, we did a simulation study to compare four different methods of SAR probit estimation in the case of large sample sizes, in which the number of neighboring observations is much less than the sample size. In the next section, we first introduce SAR probit models and explain four different methods for the estimation procedure of the SAR probit model in detail. Data processing were conducted using the R package. Package spatial probit (Wilhelm and de Matos, 2013) was used to fit Bayesian estimation of a SAR probit model using the MCMC method and Gibbs sampling, McSpatial (McMillen, 2013) was implemented for the Maximum Likelihood Estimation (MLE) and Linearized Generalized Method of Moments (GMM) estimation, and the result was compared to the conditional approximate likelihood estimation, fitted with ProbitSpatial (Martinetti and Geniaux, 2017). Simulations were varied based on the number of observations and the number of nearest neighbors.
2. Methods of Spatial Autoregressive Probit Estimation

Spatial Autoregressive probit model defined by

\[ y = \rho W y + X \beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n) \]  

\[ y = (I_n - \rho W)^{-1}(X \beta + \epsilon) \]  

\[ y_i = \begin{cases} 1, & \text{if } z_i \geq 0 \\ 0, & \text{if } z_i < 0 \end{cases} \]

where \( W \) is an \( n \times n \) spatial weights matrix, \( \rho \) is the spatial autoregressive parameter in \([-1,1]\), \( y \) is \( n \times 1 \) vector observed value of binary limited dependent variable that reflects the presence or absence of a feature in each region or observation, \( z \) is the unobserved latent dependent variable, and \( X \) is an \( n \times k \) matrix of explanatory variables. If \( \rho = 0 \), the spatial probit model collapses to the standard probit model. Spatial effects are present if \( \rho \) is not equal to zero; values of \( y_i \) are influenced by nearby values of \( y_i \), where “nearby” is defined implicitly by the prespecified entries of the weight matrix. SAR probit model relaxes the strict interdependence assumption used in standard probit models by allowing changes in one explanatory variable for one observation to impact the values of other observations within a neighboring distance as defined by the spatial weights matrix \( W \). The inverting \( n \times n \) matrix \( W \) contains the information on the spatial relationship between observations. Spatial weight matrix \( W \) is usually constructed as a function of the distance between observations or other contiguity measures, such as using nearest neighbors. We denote \( w_{ij} = 1 \), if observations \( i \) and \( j \) are contiguous, while \( w_{ij} = 0 \) otherwise has been constructed using nearest neighbors. The spatial weight matrix \( W \) of the SAR probit model is symmetric, row-standardized, and by convention, the diagonal entry is set to zero.

In the SAR probit model, the structure of spatial dependence adds complexity in parameter estimation. For SAR probit model, most used estimation techniques are full-information maximum likelihood (McMillen, 1992). The principle of maximum likelihood estimation is in which the value of \( \beta \) maximizes the probability of observing the given sample. Maximization of the log-likelihood function obtained parameters. Some iterations were necessary to find the maximum of the log-likelihood function and as a classical method. It requires hours to estimate just for a small sample problem. One iteration in the maximum likelihood estimation of SAR probit model involves calculating the expected value of the latent dependent variable and estimating new coefficients by maximizing the log-likelihood function. Estimates can be very slow because each iteration requires \( n \times n \) matrix inverse, so the maximum likelihood estimation procedure is limited to small sample size. Also, a further loss of this estimator is not provided consistent covariance matrix estimator. The models estimated in McMillen (1992) required about 40 main iterations with as many as 15 sub-iterations to maximize the log-likelihood function. An \( n \times n \) matrix must be inverted in each main iteration, which limits the estimation procedure's application to small problems. Beron and Vijverberg (2000) report maximum likelihood estimation quite difficult because of estimation times for a SAR probit model requiring many hours for a 49 observation problem. Another disadvantage of these estimators is that consistent covariance matrix estimates are not available.

Other methods for SAR probit estimation are weighted least squares (McMillen, 1992), generalized method of moment estimators (Pinkse and Slade, 1998) and linearized generalized method of moments estimators (Klier and McMillen, 2008). They show that standard GMM reduces to a nonlinear two-stage least squares problem. LeSage (2000) use Bayesian estimation via Markov Chain Monte Carlo methods that sample sequentially from conditional distributions for all parameters based on Gibbs sampling to simulate the probabilities. In SAR probit model, the dependence leads to a truncated multivariate normal distribution (TMVN) for the latent \( z \) parameters from which we need to sample these parameters. But in the MCMC sampling scheme for estimating the SAR probit model, samples...
were drawn sequentially from the conditional posterior distributions for the model parameters $\beta, \rho, z$. The conditional distribution of the parameter vector $z$ takes the form of a truncated multivariate distribution. Parameter $\rho$ can be sampled from $p(\rho | \beta, z)$ using the Metropolis-Hastings method. Joint conditional distribution for the model parameters $p(\beta, \rho | z) = p(\beta | \rho, z, y)$ and individual conditional posterior distributions for the parameters $p(\beta | \rho, z), p(\rho | \beta, z)$ that are the same as in case a continuous variable.

The latest method for SAR probit estimation was proposed by Martinetti and Geniaux (2017) who use approximate likelihood estimation which based on the full maximization of the likelihood of an approximate multivariate normal distribution function. It allows to fit SAR probit models for large datasets. The function is based on the maximization of the approximate likelihood function. The approximation is inspired by the Mendell and Elston algorithm (1978) for computing multivariate normal probabilities and take advantage of the sparsity of the spatial weight matrix, known as a conditional method and performs relatively well regarding the accuracy of the estimated parameters. The SAR probit model reformulated from Equation (2) and matrix variance-covariance of the error terms is:

$$\Sigma = \sigma^2((I_n - \rho W)^{-1}((I_n - \rho W)^{-1})^T)$$

The likelihood function is:

$$L(\beta, \rho) = \Phi_n(\mathbf{x} \in \mathbf{A} | \Sigma) = \frac{1}{(2\pi)^{n/2} \| \Sigma \|^{1/2}} \int_{\mathbf{A}_1} \cdots \int_{\mathbf{A}_n} e^{-\frac{1}{2} x^T \Sigma^{-1} x}$$

where $\mathbf{A} = \{A_i\}_{i\in\{1,...,n\}} = (a_i, b_i)_{i\in\{1,...,n\}}$ and

$$a_i = \begin{cases} ((I_n - \rho W)^{-1} X \beta), & \text{if } y_i = 0 \\ -\infty, & \text{if } y_i = 1 \end{cases}$$

$$b_i = \begin{cases} ((I_n - \rho W)^{-1} X \beta), & \text{if } y_i = 1 \\ \infty, & \text{if } y_i = 0 \end{cases}$$

By maximizing the corresponding likelihood function in the form of an n-dimensional integral, estimates of the $\beta$ and $\rho$ are obtained. Multivariate normal probabilities cannot be computed exactly because there exists no closed formula for solving integral in equation (4). Hence, Martinetti and Geniaux (2017) use the modified version of the Mendell and Elston approximation method (1974) by rewriting the multivariate normal probabilities as the product of univariate conditional probabilities. The algorithm is based on the Cholesky decomposition of the variance-covariance matrix. The estimator of $\beta$ and $\rho$ were shown to be consistent and efficient properties.

In this paper, data were generated based on the SAR probit model in equation (2). We did 100 simulations, varied by the number of nearest neighbor’s $k$ and true value coefficient spatial lag $\rho$. Table 1 presents the variations in the generated data for simulation.

| Parameter | Sample Size | $k$ | Number of Simulation (Nsim) | $\rho$ |
|-----------|-------------|-----|-----------------------------|-------|
| $\beta = (0,1,-2)$ | 500 | 6 | 100 | 0.90 |
| $\beta = (0,1,-2)$ | 500 | 6 | 100 | 0.55 |
| $\beta = (0,1,-2)$ | 500 | 3 | 100 | 0.9 |
| $\beta = (0,1,-2)$ | 500 | 3 | 100 | 0.55 |

The purpose of the four methods comparison is to find the most efficient method, regarding the accuracy of parameter estimation. Using a cumulative mean plot of parameter estimation for each method, we investigate the convergence of parameter estimation to look at the consistency of the results among the repetition of simulations. We also use the confidence interval for the mean of
parameter estimation, bias, and Root Mean Square Error (RMSE) of prediction, which will be useful for choosing the best method. The bias of an estimator is the difference between this estimator's expected value and the true value of the parameter being estimated. RMSE is used to evaluate models by summarizing the differences between the actual (observed) and predicted values. RMSE gives the standard deviation of the model prediction error. A smaller value indicates a better model performance. For comparing the estimation methods, the following step will be conducted:

**Step 1.** Generate data with \( n = 500 \)
We generate matrix observations of two explanatory variables from a normal distribution and error from a standard normal distribution. As in equation (3), we create a spatial weight matrix \( W \) based on contiguity of \( k \) nearest neighbor, \( k = 6 \), and \( k = 3 \). The vector of SAR probit model parameter is \( \beta = (0, 1, -2) \), Coefficient of spatial lag \( \rho = 0.9 \), and \( 0.55 \) for \( k = 6 \) and \( \rho = 0.9 \), and \( 0.55 \) for \( k = 3 \). We compute binary response variable \( y \) based on latent variable \( z \) as in equation (2)

**Step 2.** Using R package McSpatial, we estimate SAR probit model parameter by Maximum Likelihood Estimation (MLE) method with syntax `spprobitml` and Linearized GMM with syntax `spprobit`.

**Step 3.** Using R package spatialprobit, we estimate SAR probit model parameter by Bayes estimation method with syntax `sarprobit`. As in Wilhem (2013), the idea in Bayes estimation is to sample from a posterior distribution of the model parameters \( p(z|\beta, \rho, y) \) given the data \( y \) and prior distributions \( p(z), p(\beta), \) and \( p(\rho) \). Sampling for the posterior distribution \( p(z, \beta, \rho|y) \) can be realized by a Markov Chain Monte Carlo and Gibbs sampling scheme, where sampling from the following three conditional densities \( p(z|\beta, \rho, y), p(\beta|z, \rho, y) \) and \( p(\rho|z, \beta, y) \).

**Step 4.** Using R package ProbitSpatial, we use conditional approximate likelihood that was proposed by Martinetti and Geniaux (2017) with syntax `SpatialProbitFit` for estimating SAR probit parameters. It relies on a standard probit estimation which applies to the model estimated conditional on \( q \)

**Step 5.** Step 1 to 4 are repeated 100 times for each method and compute the 95% confidence intervals of SAR probit parameter estimation means

**Step 6.** Compute the bias from each parameter estimation method and computed prediction value of binary response for assessing RMSE of prediction.

3. Results and Discussion

3.1. Data scenario 1
We did 100 simulations with data size \( n = 500 \) and \( k = 6 \). We computed SAR probit parameter estimation for each method; MLE, linearized GMM, Bayes, and conditional approximate likelihood. We constructed a cumulative mean plot for every SAR probit parameter estimation with a variable value of spatial lag as the following:
1. Cumulative Mean Plot of $\hat{\beta}_0$

![Cumulative Mean Plot of $\hat{\beta}_0$ for $\rho = 0.9$ and $\rho = 0.55$](image)

(a) $\rho = 0.90$  
(b) $\rho = 0.55$

**Figure 1.** Cumulative Mean Plot of $\hat{\beta}_0$ for $\rho = 0.9$ and $\rho = 0.55$

2. Cumulative Mean Plot of $\hat{\beta}_1$

![Cumulative Mean Plot of $\hat{\beta}_1$ for $\rho = 0.9$ and $\rho = 0.55$](image)

(a) $\rho = 0.90$  
(b) $\rho = 0.55$

**Figure 2.** Cumulative Mean Plot of $\hat{\beta}_1$ for $\rho = 0.9$ and $\rho = 0.55$

3. Cumulative Mean Plot of $\hat{\beta}_2$

![Cumulative Mean Plot of $\hat{\beta}_2$ for $\rho = 0.9$ and $\rho = 0.55$](image)

(a) $\rho = 0.90$  
(b) $\rho = 0.55$

**Figure 3.** Cumulative Mean Plot of $\hat{\beta}_2$ for $\rho = 0.9$ and $\rho = 0.55
From Figure 1-3, we can see that parameter estimation from linearized GMM is the worst method because the estimation is so far away from the real value of SAR probit parameter (beta). We also can check this by a cumulative mean plot of \((\hat{\beta}_i - \beta_i), i = 0, 1, 2\) in figure 4-6.

1. Cumulative Mean Plot of \((\hat{\beta}_0 - \beta_0)\)

![Figure 4](image1.png)

- (a) \(\rho = 0.90\)
- (b) \(\rho = 0.55\)

**Figure 4.** Cumulative Mean Plot of \((\hat{\beta}_0 - \beta_0)\) for \(\rho = 0.9\) and \(\rho = 0.55\)

2. Cumulative Mean Plot of \((\hat{\beta}_1 - \beta_1)\)

![Figure 5](image2.png)

- (a) \(\rho = 0.90\)
- (b) \(\rho = 0.55\)

**Figure 5.** Cumulative Mean Plot of \((\hat{\beta}_1 - \beta_1)\) for \(\rho = 0.9\) and \(\rho = 0.55\)

3. Cumulative Mean Plot of \((\hat{\beta}_2 - \beta_2)\)

![Figure 6](image3.png)

- (a) \(\rho = 0.90\)
- (b) \(\rho = 0.55\)

**Figure 6.** Cumulative Mean Plot of \((\hat{\beta}_2 - \beta_2)\) for \(\rho = 0.9\) and \(\rho = 0.5\)
Table 2. Confidence interval for mean of parameter estimation

| Parameter | Method      | 95% CI for mean $\bar{\beta}_0$                      |
|-----------|-------------|-------------------------------------------------------|
| $\beta_0 = 0$ | Bayes      | (-0.0079, 0.0146)                                      |
|           | MLE        | (-0.0113, 0.0157)                                      |
|           | Lin GMM    | (-0.1000, 0.0352)                                      |
|           | App-Lik-Cond | (-0.0124, 0.0262)                                     |
| $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ |
| $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.55$ |
| $\rho = 0.55$ | $\rho = 0.55$ | $\rho = 0.55$ |
| $\beta_1 = 1$ | Bayes      | (0.6613, 0.7030)                                       |
|           | MLE        | (1.0476, 1.1626)                                       |
|           | Lin GMM    | (0.3539, 0.3788)                                       |
|           | App-Lik-Cond | (0.8314, 0.9131)                                     |
| $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ |
| $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.55$ |
| $\rho = 0.55$ | $\rho = 0.55$ | $\rho = 0.55$ |
| $\beta_2 = -2$ | Bayes      | (-1.3887, -1.3273)                                     |
|           | MLE        | (-2.3600, -2.1935)                                     |
|           | Lin GMM    | (-0.7533, -0.7175)                                     |
|           | App-Lik-Cond | (-1.8601, -1.7360)                                    |
| $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ |
| $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.55$ |
| $\rho = 0.55$ | $\rho = 0.55$ | $\rho = 0.55$ |

Table 3. Bias of SAR probit estimation

| Parameter | Method     | Bias of $\bar{\beta}_0$                      |
|-----------|------------|-----------------------------------------------|
| $\beta_0 = 0$ | Bayes | 0.0033                                        |
|           | MLE       | 0.0022                                        |
|           | Lin GMM   | -0.0324                                       |
|           | App-Lik-Cond | 0.0069                                       |
| $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ |
| $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.55$ |
| $\rho = 0.55$ | $\rho = 0.55$ | $\rho = 0.55$ |
| $\beta_1 = 1$ | Bayes | -0.3178                                       |
|           | MLE       | 0.1051                                        |
|           | Lin GMM   | -0.6341                                       |
|           | App-Lik-Cond | -0.1277                                      |
| $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ |
| $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.55$ |
| $\rho = 0.55$ | $\rho = 0.55$ | $\rho = 0.55$ |
| $\beta_2 = -2$ | Bayes | 0.6420                                        |
|           | MLE       | -0.2798                                       |
|           | Lin GMM   | 1.2646                                        |
|           | App-Lik-Cond | 0.2019                                       |
| $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ |
| $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.55$ |
| $\rho = 0.55$ | $\rho = 0.55$ | $\rho = 0.55$ |
3.2. Data scenario 2
We constructed a cumulative mean plot for every SAR probit parameter estimation with vary value of spatial lag with $n = 500, k = 3, Nsim = 100$ as the following in Figure 7-9:

1. Cumulative Mean Plot of $\hat{\beta}_0$

![Cumulative Mean Plot of $\hat{\beta}_0$ for $\rho = 0.90$ and $\rho = 0.55$](image)

Figure 7. Cumulative Mean Plot of $\hat{\beta}_0$ for $\rho = 0.90$ and $\rho = 0.55$

2. Cumulative Mean Plot of $\hat{\beta}_1$

![Cumulative Mean Plot of $\hat{\beta}_1$ for $\rho = 0.90$ and $\rho = 0.55$](image)

Figure 8. Cumulative Mean Plot of $\hat{\beta}_1$ for $\rho = 0.90$ and $\rho = 0.55$

2. Cumulative Mean Plot of $\hat{\beta}_2$

![Cumulative Mean Plot of $\hat{\beta}_2$ for $\rho = 0.90$ and $\rho = 0.55$](image)

Figure 9. Cumulative Mean Plot of $\hat{\beta}_2$ for $\rho = 0.90$ and $\rho = 0.55
We also check the cumulative mean plot of \((\beta_i - \beta_i), i = 0,1,2\) in figure 10-12.

1. Cumulative Mean Plot of \((\beta_0 - \beta_0)\)

![Cumulative Mean Plot of \((\beta_0 - \beta_0)\)](image)

(a) \(\rho = 0.90\)  
(b) \(\rho = 0.55\)

**Figure 10.** Cumulative Mean Plot of \((\beta_0 - \beta_0)\) for \(\rho = 0.9\) and \(\rho = 0.55\)

2. Cumulative Mean Plot of \((\beta_1 - \beta_1)\)

![Cumulative Mean Plot of \((\beta_1 - \beta_1)\)](image)

(a) \(\rho = 0.90\)  
(b) \(\rho = 0.55\)

**Figure 11.** Cumulative Mean Plot of \((\beta_1 - \beta_1)\) for \(\rho = 0.9\) and \(\rho = 0.55\)

3. Cumulative Mean Plot of \((\beta_2 - \beta_2)\)

![Cumulative Mean Plot of \((\beta_2 - \beta_2)\)](image)

(a) \(\rho = 0.90\)  
(b) \(\rho = 0.55\)

**Figure 12.** Cumulative Mean Plot of \((\beta_2 - \beta_2)\) for \(\rho = 0.9\) and \(\rho = 0.40\)
Table 4. Confidence interval for mean of parameter estimation.

| Parameter | Method    | 95% CI for mean $\hat{\beta}_0$ | $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.40$ |
|-----------|-----------|----------------------------------|---------------|---------------|---------------|---------------|
| $\beta_0 = 0$ | Bayes    | (-0.0180, 0.0004)               | (-0.0122, 0.0070) | (-0.0122, 0.0126) | (-0.0109, 0.0148) |
|           | MLE      | (-0.0206, 0.0055)               | (-0.0171, 0.0054) | (-0.0137, 0.0119) | (-0.0123, 0.0138) |
|           | Lin GMM  | (0.0009, 0.0621)                | (-0.0154, 0.0049) | (-0.0098, 0.0080) | (-0.0099, 0.0101) |
|           | App-Lik-Cond | (-0.0335, -0.0014)           | (-0.0168, 0.0082) | (-0.0131, 0.0126) | (-0.0122, 0.0141) |

| Parameter | Method    | 95% CI for mean $\hat{\beta}_1$ | $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.40$ |
|-----------|-----------|----------------------------------|---------------|---------------|---------------|---------------|
| $\beta_1 = 1$ | Bayes    | (0.4825, 0.5156)                | (0.8307, 0.8679) | (0.9898, 1.0375) | (1.0028, 1.0488) |
|           | MLE      | (1.0660, 1.1587)                | (1.0156, 1.0693) | (1.0016, 1.0502) | (0.9985, 1.0452) |
|           | Lin GMM  | (0.2792, 0.3016)                | (0.5056, 0.5302) | (0.7207, 0.7555) | (0.8480, 0.8841) |
|           | App-Lik-Cond | (0.7582, 0.8216)           | (0.9718, 1.0238) | (0.9985, 1.0471) | (0.9979, 1.0445) |

| Parameter | Method    | 95% CI for mean $\hat{\beta}_2$ | $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.40$ |
|-----------|-----------|----------------------------------|---------------|---------------|---------------|---------------|
| $\beta_2 = -2$ | Bayes    | (-1.0035, -0.9605)               | (-1.7102, -1.6480) | (-2.0876, -2.0600) | (-2.1145, -2.0301) |
|           | MLE      | (-2.2781, -2.1158)               | (-2.1139, -2.0292) | (-2.1089, -2.0217) | (-2.1050, -2.0215) |
|           | Lin GMM  | (-0.5917, -0.5660)               | (-1.0341, -0.9954) | (-1.5160, -1.4620) | (-1.7746, -1.7187) |
|           | App-Lik-Cond | (-1.6136, -1.5219)           | (-2.0236, -1.9412) | (-2.1029, -2.0158) | (-2.1030, -2.0197) |

Table 5. Bias of SAR probit estimation.

| Parameter | Method    | Bias of $\hat{\beta}_0$ | $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.40$ |
|-----------|-----------|--------------------------|---------------|---------------|---------------|---------------|
| $\beta_0 = 0$ | Bayes    | -0.0088                   | -0.0026       | 0.0002        | 0.0020        |
|           | MLE      | -0.0076                   | -0.0058       | -0.0009       | 0.0008        |
|           | Lin GMM  | 0.0315                    | -0.0052       | -0.0009       | 0.0001        |
|           | App-Lik-Cond | -0.0174                | -0.0043       | -0.0003       | 0.0009        |

| Parameter | Method    | Bias of $\hat{\beta}_1$ | $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.40$ |
|-----------|-----------|--------------------------|---------------|---------------|---------------|---------------|
| $\beta_1 = 1$ | Bayes    | -0.5010                   | -0.1507       | 0.0137        | 0.0258        |
|           | MLE      | 0.1123                    | 0.0424        | 0.0259        | 0.0218        |
|           | Lin GMM  | -0.7096                   | -0.4831       | -0.2619       | -0.1340       |
|           | App-Lik-Cond | -0.2101               | -0.0021       | 0.0228        | 0.0212        |

| Parameter | Method    | Bias of $\hat{\beta}_2$ | $\rho = 0.90$ | $\rho = 0.75$ | $\rho = 0.55$ | $\rho = 0.40$ |
|-----------|-----------|--------------------------|---------------|---------------|---------------|---------------|
| $\beta_2 = -2$ | Bayes    | 1.0180                    | 0.3209        | -0.0468       | -0.0723       |
|           | MLE      | -0.1970                   | -0.0715       | -0.0653       | -0.0632       |
|           | Lin GMM  | 1.4211                    | 0.9852        | 0.5110        | 0.2533        |
|           | App-Lik-Cond | 0.4322                 | 0.0176        | -0.0593       | -0.0614       |
Based on the cumulative mean plot of parameter estimation, the confidence interval of mean and bias of estimator for each method, we found that linearized GMM estimation result was so far from a value of parameter setting. So we were assessing RMSE of prediction for three other methods; MLE, Bayes estimation, and conditional approximate likelihood estimation in Table 6.

Table 6. RMSE of Prediction.

| Method   | RMSE of Prediction (nobs=100, k=6, Ns im=100)  | RMSE of Prediction (nobs=100, k=3, Ns im=100) |
|----------|----------------------------------------------|-----------------------------------------------|
|          | $\rho = 0.90$                      | $\rho = 0.75$                      | $\rho = 0.55$                      | $\rho = 0.90$                      | $\rho = 0.75$                      | $\rho = 0.55$                      |
| Bayes    | 0.3707                               | 0.3629                               | 0.3609                               | 0.3889                               | 0.3720                               | 0.3617                               | 0.3592                               |
| MLE      | 0.3509                               | 0.3609                               | 0.3603                               | 0.3523                               | 0.3667                               | 0.3670                               | 0.3610                               | 0.3591                               |
| App-Lik-Cond | 0.3825                           | 0.3642                           | 0.3611                           | 0.4022                           | 0.3736                           | 0.3619                           | 0.3585                           |

We found that conditional approximate likelihood was the overall best method, but it is worst when it comes to estimating RMSE of prediction in a large value of coefficient spatial lag.

References

[1] L. Anselin, R. Florax, and S.J. Rey. 2015. Advances in Spatial Econometrics: Methodology, Tools and Applications", US: Springer; 2015.

[2] L. Anselin, A. K. Bera, R. Florax and M.J. Yoon. 1996. Simple Diagnostic Tests for Spatial Dependence, Regional Science and Urban Economics. 26:77-104. DOI: 10.1016/0166-0462(95)02111-6.

[3] C. Anne. 1992. Neighborhood Influence and Technological Change. Regional Science and Urban Economics. 22(3): 491–508. DOI: 10.1016/0166-0462(92)90041-X.

[4] B. H. Baltagi, S.H. Song and W. Koh. 2003. Testing Panel Data Regression Models with Spatial Error Correlation. Journal of Econometrics. 117(1): 123-150. DOI: 0.1016/S0304-4076(03)00120-9.

[5] K. J. Beron and W. P. M. Vijverberg. 2004. Probit in a spatial context: a Monte Carlo analysis, Advances in Spatial Econometrics. Springer Berlin Heidelberg;169–195.

[6] C. R. Bhat, J. A. Guo. 2004. Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modelling. Transportation Research Part B: Methodological. 38(2): 147-168. DOI: 10.1016/S0191-2615(03)00005-5.

[7] D. Bolduc, B. Fortin, and S. Gordon. 1997. Multinomial Probit Estimation of Spatially Interdependent Choices: An Empirical Comparison of Two New Techniques. International Regional Science Review. 20(1-2): 77-101. DOI: 10.1177/016001769702000105.

[8] M. Fleming. 2002. "Techniques for Estimating Spatially Dependent Discrete Choice Models." In: Anselin, L., Florax, R.J., Rey, S.J. (Eds.). Advances in Spatial Econometrics. Springer, Heidelberg.

[9] G. Holloway G, B. Shankar B, and S. Rahman. 2002. Bayesian spatial probit estimation: a primer and an application to HYV rice adoption. Agricultural Economics. 27(3): 383-402. DOI: 10.1016/S0169-5150(02)00070-1.

[10] M. Jerret, S. Gale and C. Kontgis. 2010. Spatial Modeling in Environmental and Public Health Research, International Journal of Environmental Research and Public Health. 7(4): 1302-1329. DOI: 10.3390/ijerph7041302

[11] T. Klier, D. P. McMillen. 2008. Clustering of Auto Supplier Plants in the United States: Generalized Method of Moments Spatial Logit for Large Samples, Journal of Business and Economic Statistics. 26:460-471.

[12] S. Y. H. Kao and A. K. Bera. 2016. Spatial Regression: The Curious Case of Negative Spatial
Dependence. Working Paper.

[13] J. P. LeSage. 2000. Bayesian estimation of limited dependent variable spatial autoregressive models. Geographical Analysis. 32(1): 19–35.

[14] J. P. LeSage and R. K. Pace. 2009. Introduction to Spatial Econometrics, Chapman and Hall/CRC.

[15] D. Martinetti and G. Geniaux. 2017. Approximate likelihood estimation of spatial probit models, Regional Science and Urban Economics. 64: 30-45. DOI: 10.1016/j.regsciurbeco.2017.02.002.

[16] D. Martinetti D and G. Geniaux. 2016. ProbitSpatial: Probit with Spatial Dependence, SAR and SEM Models, http://CRAN.Rproject.org/package=ProbitSpatial

[17] D. P. McMillen. 1992. Probit with Spatial Autocorrelation, Journal of Regional Science. 32(3): 335-348. DOI: 10.1111/j.1467-9787.1992.tb00190.x.

[18] McMillen DP. 2013. McSpatial: Nonparametric spatial data analysis, http://CRAN.Rproject.org/package=McSpatial.

[19] N. R. Mendell and R. C. Elston. 1974. Multifactorial qualitative traits: genetic analysis and prediction of recurrence risks, Biometrics. 30(1): 41–57. DOI: 10.2307/2529616.

[20] J. Pinkse and M. E. Slade. 1998. Contracting in space: an application of spatial statistics to discrete choice models. Journal of Econometrics. 85(1): 125–154. DOI: 10.1016/S0304-4076(97)00097-3.

[21] S. Wilhelm and M. G. D. Matos. 2013. spatialprobit: SpatialProbitModels. http://CRAN.Rproject.org/package=spatialprobit.

[22] S. Wilhelm S and M. G. D. Matos. 2013. Estimating Spatial Probit Models in R, The R Journal. 5(1).

[23] L. H. Wong, P. Pattison, and G. A. Robins. 2006. Spatial model for Social Network, Physica A: Statistical Mechanics and its Applications. 360(1): 99-120. DOI: 10.1016/j.physa.2005.04.029.

[24] Y. Yu. 2017. Empirical Spatial Econometric: Application to China’s Economy, Emerging Markets Finance & Trade. 53:1939-1942. DOI: 10.1080/1540496X.2017.1363489.