E C O N O M I C S   B U L L E T I N

Bayesian Estimation of A Distance Functional Weight Matrix Model

Kazuhiko Kakamu
Department of Economics Finance, Institute for Advanced Studies

Abstract

This paper considers the distance functional weight matrix in spatial autoregressive and spatial error models from a Bayesian point of view. We considered the Markov chain Monte Carlo methods to estimate the parameters of the models. Our approach is illustrated with simulated data set.

We thank the Associate Editor and an anonymous referee for helpful comments and discussions. Of course, All remaining errors are ours.
Citation: Kakamu, Kazuhiko, (2005) "Bayesian Estimation of A Distance Functional Weight Matrix Model." *Economics Bulletin*, Vol. 3, No. 57 pp. 1−6
Submitted: October 18, 2005. **Accepted:** December 27, 2005.
URL: http://www.economicsbulletin.com/2005/volume3/EB−05C10018A.pdf

1 Introduction

Specification of the weight matrix is one of the problems in analyzing spatial data. There are two major approaches for specifying the weight matrix; contiguity dummy variables and distances among units. The latter is more flexible in that the spatial effects are different for different distances. However, the functional form of the weight matrix is determined in advance.

LeSage and Pace (2005) proposed a matrix exponential spatial specification (MESS) procedure. One of the advantages of this method, which they emphasize, is that it saves time because it does not require the calculation of a determinant. In addition, this method replaces the geometric pattern of decay in the spatial autoregressive (SAR) model with one of exponential decay. However this method can be only applied to the SAR model.

In empirical analysis, we are interested not only in the SAR model, but also in the spatial error model (SEM). Moreover, someone may be interested in both spatial interaction and spatial decay in intensities. In this paper, we propose a distance functional weight matrix model as an alternative method to capture both the intensity of spatial interaction and geometric pattern of decay and consider MCMC methods to estimate the parameters of the model. Our Bayesian approach is illustrated with simulated data set and we show the advantages of our approach.

The rest of this paper is organized as follows. In Section 2, we summarize the distance functional weight matrix model, which we propose in this paper and obtain a joint posterior distribution. Section 3 discusses the computational strategy of the MCMC method. In Section 4, our approach is illustrated using simulated data sets. Finally, brief conclusions are given in Section 5.

2 Distance functional weight matrix model

Let y_i denote a dependent variable for the ith unit, let x_i denote independent variables, where x_i is a $1 \times k$ vector of the ith unit and let w_{ij} denote the spatial weight on the jth unit with respect to the ith unit[1]. Then, the SAR model conditioned on parameters β, σ^2 and ρ is written as follows;

$$y_i = x_i\beta + \sum_{j=1}^{N} \rho \frac{w_{ij}}{\sum_{n=1}^{N} w_{in}} y_j + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2).$$

The parameter ρ measures the average influence of neighboring or contiguous obsevations on the observations. In this equation, we can capture the intensity of spatial interaction, but the functional form of the weight matrix is predetermined. On the other hand, LeSage and Pace (2005) propose the following MESS model;

$$y_i = x_i\beta + \sum_{j=1}^{N} \exp\left(\lambda \frac{w_{ij}}{\sum_{n=1}^{N} w_{in}}\right) y_j + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2).$$

[1]We will continue our discussion of the spatial autoregressive model. However, the techniques introduced here can be applied to SEM easily. The procedure to estimate SEM is introduced in Appendix A.

In this equation, λ captures the geometric pattern of decay in the SAR model, but the functional form of weight matrix is pre-determined as exponential and we cannot distinguish whether λ captures the intensity of spatial interaction or the geometric pattern of decay. In empirics, we are sometimes interested in the intensities both of spatial interaction and of decay simultaneously, and want to distinguish them. Therefore, we will propose the following distance functional weight matrix model with spatial lag conditioned on parameters β, σ^2, ρ and ϕ given by[2];

$$y_i = x_i\beta + \sum_{j=1}^{N} \rho \frac{w_{ij}^{\phi}}{\sum_{n=1}^{N} w_{in}^{\phi}} y_j + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \tag{1}$$

where ϕ and ρ express the intensities of spatial decay and spatial interaction, respectively. The corresponding likelihood function of the model (1) is as follows;

$$p(Y|\beta, \sigma^2, \rho, \phi, X, W) = (2\pi\sigma^2)^{-N/2} |I_N - \rho f(W|\phi)| \exp\left(-\frac{e'e}{2\sigma^2} \right), \tag{2}$$

where I_N is an $N \times N$ unit matrix, W denotes weight matrix (see e.g., Anselin, 1988) and $e = Y - \rho f(W|\phi)Y - X\beta$, where the ijth element of $f(W|\phi)$ is $f_{ij}(W|\phi) = \frac{w_{ij}^{\phi}}{\sum_{n=1}^{N} w_{in}^{\phi}}$.

3 Posterior analysis

3.1 Joint posterior distribution

Given a prior density $p(\beta, \sigma^2, \rho, \phi)$ and the likelihood function given in (2), the joint posterior distribution can be expressed as

$$\begin{aligned} & p(\beta, \sigma^2, \rho, \phi | Y, X, W) \\ & \quad = p(\beta, \sigma^2, \rho, \phi) p(Y|\beta, \sigma^2, \rho, \phi, X, W) \\ & \quad = p(\beta, \sigma^2) p(\rho) p(\phi) p(Y|\beta, \sigma^2, \rho, \phi, X, W) \\ & \quad = p(\beta|\sigma^2) p(\sigma^2) p(\rho) p(\phi) p(Y|\beta, \sigma^2, \rho, \phi, X, W). \end{aligned} \tag{3}$$

And the prior distributions are as follows;

$$\begin{aligned} p(\beta|\sigma^2) \sim N(\beta_*, \sigma^2 A_*^{-1}), \quad p(\sigma^2) \sim G^{-1}(\nu_*/2, \lambda_*/2), \\ p(\rho) \sim U(-1, 1), \quad p(\phi) \propto const, \end{aligned}$$

where subscript $*$ denotes hyperparameters of the prior distributions, $G^{-1}(a, b)$ denotes an inverse gamma distribution with shape and scale parameters a and b. In addition, we suppose that the prior space of ρ is distributed *uniformly* over the interval $(-1, 1)$. Therefore, we restrict the prior space as $\rho \in (-1, 1)$.

3.2 Posterior simulation

The Markov chain sampling scheme can be constructed from the full conditional distributions of β, σ^2, ρ, ϕ.

[2] If we suppose a_i and b_i be the ith coordinates, the ijth element of the weight matrix, w_{ij}, is calculated as $w_{ij} = ((a_i - a_j)^2 + (b_i - b_j)^2)^{\frac{1}{2}}$.

3.2.1 Sampling ρ

From (3), the full conditional distribution of ρ is written as

$$\rho|\beta, \sigma^2, \phi, Y, X, W \propto |I_N - \rho f(W|\phi)| \exp\left(- \frac{e'e}{2\sigma^2} \right),$$

which cannot be sampled by standard methods (see e.g., LeSage, 2000). Therefore, we adopt the Metropolis algorithm (see e.g., Tierny, 1994).

The following Metropolis step is used: Sample ρ from

$$\rho = \rho^* + c\psi, \ \ \psi \sim N(0, 1),$$

where c is called a tuning parameter. Next, we evaluate the acceptance probability

$$\alpha(\rho^*, \rho) = min\left(\frac{p(\rho)}{p(\rho^*)}, 1 \right),$$

and finally set $\rho = \rho$ with probability $\alpha(\rho^*, \rho)$, otherwise $\rho = \rho^*$. It should be mentioned that the proposal value of ρ is not truncated to the interval $(-1, 1)$ since the constraint is part of the target density. Thus, if the proposed value of ρ is not within the interval, the conditional posterior is zero, and the proposal value is rejected with probability one. In the numerical example given below, we choose the tuning parameter value such that the acceptance rate is between 0.4 and 0.6 (see LeSage, 2000)[3].

3.2.2 Sampling ϕ

From (3), the full conditional distribution of ϕ can be written as

$$\phi|\beta, \sigma^2, \rho, Y, X, W \propto |I_N - \rho f(W|\phi)| \exp\left(- \frac{e'e}{2\sigma^2} \right),$$

which cannot be sampled by standard methods. Therefore, we again adopt the Metropolis algorithm.

The following Metropolis step is used: Sample ρ from

$$\phi = \phi^* + d\psi, \ \ \psi \sim N(0, 1),$$

where d is a tuning parameter. Next, we evaluate the acceptance probability

$$\alpha(\phi^*, \phi) = min\left(\frac{p(\phi)}{p(\phi^*)}, 1 \right),$$

and finally set $\phi = \phi$ with probability $\alpha(\phi^*, \phi)$, otherwise $\phi = \phi^*$. In the numerical example given below, we also choose the tuning parameter value such that the acceptance rate becomes between 0.4 and 0.6.

[3]If we choose the smaller values of c, it leads to slower exploration of the parameter space with a higher acceptance rate. On the other hand, if we choose the larger values of c, it leads to faster exploration of the parameter space with a lower acceptance rate. Taking this tardeoff into account, we choose the acceptance rate, which is between 0.4 and 0.6.

3.2.3 Sampling β and σ^2

If ρ and ϕ are given, then $Y^* = AY$ becomes a constant, where $A = I_N - \rho f(W|\phi)$, and the model is reduced to a linear regression model. Therefore, for β and σ^2, it can be easily verified that

$$\begin{aligned} \beta|\rho, \phi, \sigma^2, Y, X, W &\sim N(\tilde{\beta}, \sigma^2 \tilde{\Sigma}), \\ \sigma^2|\rho, \phi, Y, X, W &\sim G^{-1}\left(\frac{\hat{\nu}}{2}, \frac{\hat{\lambda}}{2}\right), \end{aligned}$$

where $\tilde{\beta} = (X'X + A_*)^{-1}(X'X\hat{\beta}^* + A_*\beta_*),\ \hat{\beta}^* = (X'X)^{-1}X'Y^*,\ \tilde{\Sigma} = (X'X + A_*)^{-1},\ \hat{\nu} = \nu_* + \nu,\ \hat{\lambda} = \lambda_* + \nu s^2 + (\beta_* - \hat{\beta}^*)'\tilde{\Omega}^{-1}(\beta_* - \hat{\beta}^*),\ \nu = N - k,\ s^2 = \nu^{-1}(Y^* - X\hat{\beta}^*)'(Y^* - X\hat{\beta}^*),\ \tilde{\Omega}^{-1} = (X'X)^{-1} + A^{-1}$. These parameters are easily sampled from a Gibbs sampler (see e.g., Gelfand and Smith, 1990).

4 Numerical example

To illustrate the Bayesian approach discussed in the previous section, y_i was generated from the normal distribution

$$y_i = 1.0 + 1.0x_{1i} + 1.0x_{2i} + \sum_{j=1}^{N} 0.6 \frac{w_{ij}^{-8}}{\sum_{n=1}^{N} w_{in}^{-8}} y_j + u_{it}, \quad u_{it} \sim N(0, 2)$$

where x_{1it} and x_{2it} were standard normal variates and where coordinates of W were log normal variates with mean 1 and variance 1000^2. For the prior distributions, the hyper-parameters are set as follows;

$$\beta_* = 0,\ A_* = 100^{-1} \cdot I_k,\ \nu_* = 2,\ \lambda_* = 0.01,$$

Since it is interesting to see the effect of misspecification of the weight matrix, we estimated the model with the restriction $\phi = 1$ and $\phi = -1$ as well as the model without the restriction. With the simulated data, we ran the MCMC algorithm, using 20000 iterations and discarding the first 5000 iterations. The chain was considered to have practically converged after 5000 iterations based on a diagnostic proposed by Geweke (1992). All results reported here were generated using Ox version 3.4 (see Doornik, 2001).

Table 1 shows the posterior estimates of the parameters[4]. From the table, we see that the estimated ϕ is far from the true value. However, from Figure 1, we see that the approximated posterior mode is around the true value[5]. Moreover, the other parameters are centered around the true value, which is not the case when $\phi = -1$ and $\phi = 1$. In the case of $\phi = -1$, not only is ρ overestimated, but β_0 is underestimated. In the case of $\phi = 1$, not only is ρ estimated with a different sign, but β_0 is overestimated. These results imply that the misspecification of the weight matrix leads to serious misspecification biases in ρ and β_0.

[4] The p-values of Geweke's diagnostic are $\beta_0 = 0.901$, $\beta_1 = 0.107$, $\beta_2 = 0.348$, $\rho = 0.567$, $\phi = 0.405$ and $\sigma^2 = 0.179$, respectively. It implies that all the chains converge.

[5] We also try a much smaller tuning parameter d to make the convergence fast. However, if we use it, the chain does not converge. Therefore the posterior distribution for the parameter ϕ so diffuse.

Table 1: Simulation results

	True value	without restriction		$\phi = -1$		$\phi = 1$	
β_0	1.00	0.909	(0.284)	0.351	(0.593)	4.316	(0.732)
β_1	1.00	1.009	(0.237)	1.050	(0.383)	1.111	(0.418)
β_2	1.00	0.742	(0.245)	1.151	(0.387)	1.139	(0.418)
ρ	0.60	0.605	(0.069)	0.756	(0.155)	−0.699	(0.254)
ϕ	−8.00	−26.322	(23.880)				
σ^2	2.00	2.350	(0.542)	6.097	(1.329)	7.079	(1.528)

Posterior means and standard deviations (in parentheses) are shown.

5 Conclusion

This paper has examined the distance functional weight matrix model from a Bayesian point of view. We expressed the joint posterior distribution and proposed MCMC methods to estimate the parameters of the model. We have illustrated our approach using simulated data.

From the results, we found serious misspecification biases in ρ and β_0. If we are interested in ρ and β_0, it may lead to misinterpretation. Therefore we can conclude that our approach can avoid the misspecification problem of the weight matrix if we use the distance data as a weight matrix. In addition, our approach is also superior to LeSage and Pace's (2005) approach in that our approach can be also applied to SEM. However, our approach is more time consuming because it requires the calculation of a determinant. In addition, the posterior distribution for the parameter ϕ is diffuse. These problems are remaining issues.

A Spatial error model (SEM)

The distance functional weight matrix model with spatial error conditioned on parameters β, σ^2 ρ and ϕ is written as follows;

$$y_i = x_i\beta + \sum_{j=1}^{N} \rho \frac{w_{ij}^{\phi}}{\sum_{n=1}^{N} w_{in}^{\phi}} (y_j - x_j\beta) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2). \tag{4}$$

Then, we will introduce the likelihood function of the model (4) as follows;

$$p(Y|\beta, \sigma^2, \rho, \phi, X, W) = (2\pi\sigma^2)^{-N/2} |I_N - \rho f(W|\phi)| \exp\left(-\frac{e'e}{2\sigma^2} \right),$$

where $e = Y - X\beta - \rho f(W|\phi)(Y - X\beta)$.

Suppose $X^* = AX$ and if we replace the e above and change the sampling scheme of β and σ^2 as follows;

$$\begin{aligned} \beta|\rho, \phi, \sigma^2, Y, X, W &\sim N(\tilde{\beta}, \sigma^2\tilde{\Sigma}), \\ \sigma^2|\rho, \phi, Y, X, W &\sim G^{-1}\left(\frac{\hat{\nu}}{2}, \frac{\hat{\lambda}}{2}\right), \end{aligned}$$

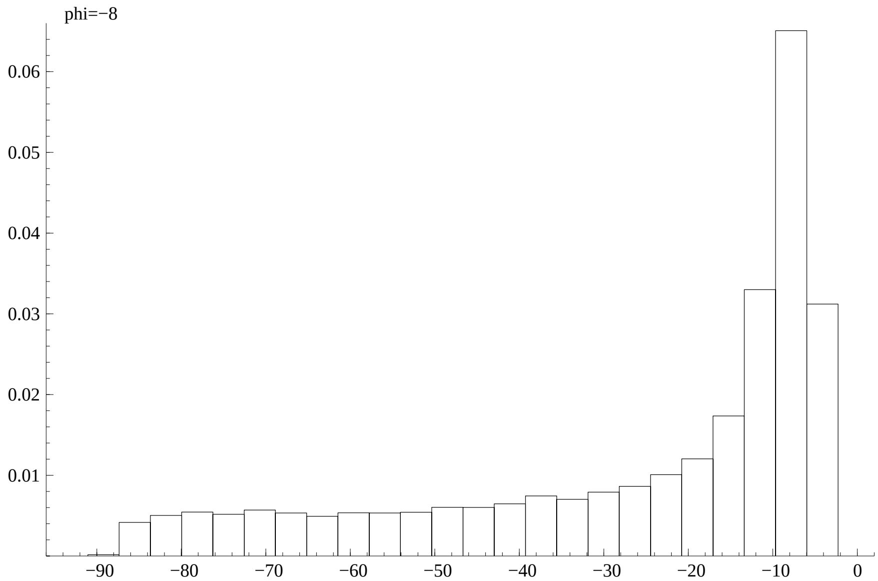

Figure 1: The approximate posterior distribution of ϕ

where $\tilde{\beta} = (X^{*'}X^* + A_*)^{-1}(X^{*'}X^*)\hat{\beta}^* + A_*\beta_*)$, $\hat{\beta}^* = (X^{*'}X^*)^{-1}X^{*'}Y^*$, $\tilde{\Sigma} = (X^{*'}X^* + A_*)^{-1}$, $\hat{\nu} = \nu_* + \nu$, $\hat{\lambda} = \lambda_* + \nu s^2 + (\beta_* - \hat{\beta}^*)'\tilde{\Omega}^{-1}(\beta_* - \hat{\beta}^*)$, $\nu = N - k$, $s^2 = \nu^{-1}(Y^* - X^*\hat{\beta}^*)'(Y^* - X^*\hat{\beta}^*)$, $\tilde{\Omega}^{-1} = (X^{*'}X^*)^{-1} + A^{-1}$, then we can apply our method to the spatial error model.

References

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer.

Doornik, J. A. (2001) *Ox: Object Oriented Matrix Programming, 3.0*, Timberlake Consultants Press, London.

Gelfand, A. E. and A. F. M. Smith (1990) "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, **85**, 398–409.

Geweke, J. (1992) "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Means," in *Bayesian Statistics 4*, (eds J. M. Bernard, J. O. Berger, A. P. Dawid and A. F. M. Smith), Oxford University Press, Oxford.

LeSage, J. P. (2000) "Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models," *Geographical Analysis*, **32**, 19–35.

LeSage, J. P. and R. K. Pace (2005) "Matrix Exponential Spatial Specification," forthcoming in *Journal of Econometrics.*

Tierney, L. (1994) " Markov Chains for Exploring Posterior Distributions (with discussion)," *Annals of Statistics*, **22**, 1701–1762.