Spatial Agglomeration and Spill-Over Analysis for Japanese Prefectures during 1991-2000

[1]Kazuhiko Kakamu, [2]Wolfgang Polasek and [3]Hajime Wago

[1]Faculty of Law and Economics,Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba, 263-8522, Japan,
[2]Department of Economics & Finance, Institute for Advanced Studies, Stumpergasse 56, 1060, Vienna, Austria.
[3]Department of Economics, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto, 603-8555, Japan.
wago@ism.ac.jp

Keywords: *MCMC; Panel data; Seemingly unrelated regression (SUR); Spatial models.*

EXTENDED ABSTRACT

This paper considers the seemingly unrelated regression (SUR) model with spatial dependencies from a Bayesian point of view. We consider Markov chain Monte Carlo (MCMC) methods to estimate the parameters of the models. We analyze the economics of agglomeration in Japanese prefecture during the period 1991 to 2000. From our empirical results, we found that the spatial error SUR model was the best model and that the economics of agglomeration and spill-over effects decreased in this decade.

Panel spatial data has been widely used in geographical statistics, regional science and so on. Although the analysis of panel spatial data is popular in several research areas, panel spatial models have been rarely examined in econometrics. One of the reasons may be the difficulties of evaluating the likelihoods of the models. However, because of the progress of the Markov chain Monte Carlo (MCMC) methods (see, *e.g.*, Chib, 2001 and Gamerman and Lopes, 2006 for recent advances of MCMC methods), it becomes easier to estimate the parameters of such models. For example, Kakamu *et al*. (2007) examined the spatial interaction of crime incidents in Japan using non-hierarchical panel spatial autoregressive model with heteroscedasticity and Kakamu and Wago (2007) showed the advantage of Bayesian panel spatial autoregressive model with hierarchical priors.

According to Anselin (1988), it is stated that the seemingly unrelated regression (SUR) model proposed by Zellner (1962) is applicable to panel data model. The advantage of the SUR model in panel models is in time varying parameters. If researchers are interested in the dynamics of the parameters, the SUR model is useful. Therefore, Anselin (1988) proposed the maximum likelihood method to estimate the model. However, as is pointed out by Kakamu and Wago (2007), the maximum likelihood methods involve the restriction problem of spatial parameters and Kakamu and Wago (2007) showed that the Bayesian method can avoid the problem.

From a Bayesian point of view, Zellner (1971) and Box and Tiao (1973) studied the model and Percy (1992) and Koop (2003) examined the Markov chain Monte Carlo (MCMC) methods to estimate the model. Chib and Greenberg (1995) extended the model with vector autoregressive and vector moving average errors of the first order. Thus, we propose a Bayesian approach for the estimation of the models (spatial autoregressive and spatial error models) using MCMC methods. We also introduce the marginal likelihood to compare the models with and without spatial interaction. Moreover, our approach is illustrated by real data set. In the real data example, we examined the economics of agglomeration with and without interregional spill-overs by panel data during the periods 1991 to 2000, which is called “lost decade”.

From the results, we found that the SEM-SUR model is the best model and serial correlation played an important role. In addition, following tendencies are confirmed: (1) Average TFP is the source of economic growth in Japanese manufacturing industries in this decade. (2) Manufacturing industries in Japan became more labor intensive. (3) The economics of agglomeration and spill-over effects became smaller over time and spill-over effect vanished in 1993.

Finally, we will mention about the remaining issues. As is pointed out above, the result that the Japanese manufacturing industries became labor intensive seems to be different from the situation in this decade in Japan. We cannot conclude what happened to the Japanese manufacturing industries in this decade only from our empirical results. However, it beyonds our analysis and it requires other moderate model or theory, like human capital theory. Moreover, we mentioned that the serial correlation plays an important role in this model. If we extend the model by Chib and Greenberg (1995) to spatial model, the serial correlation may become small. However, as our main purpose is to examine the SUR model with spatial dependencies, it also beyond our analysis, but our findings from Japanese cases represent an interesting firs step.

1 INTRODUCTION

Panel spatial data has been widely used in geographical statistics, regional science and so on. Although the analysis of panel spatial data is popular in several research areas, panel spatial models have been rarely examined in econometrics. One of the reasons may be the difficulties of evaluating the likelihoods of the models. However, because of the progress of the Markov chain Monte Carlo (MCMC) methods (see, *e.g.*, Chib, 2001 and Gamerman and Lopes, 2006 for recent advances of MCMC methods), it becomes easier to estimate the parameters of such models. For example, Kakamu *et al*. (2007) examined the spatial interaction of crime incidents in Japan using non-hierarchical panel spatial autoregressive model with heteroscedasticity and Kakamu and Wago (2007) showed the advantage of Bayesian panel spatial autoregressive model with hierarchical priors.

According to Anselin (1988), it is stated that the seemingly unrelated regression (SUR) model proposed by Zellner (1962) is applicable to panel data model. The advantage of the SUR model in panel models is in time varying parameters. If researchers are interested in the dynamics of the parameters, the SUR model is useful. Therefore, Anselin (1988) proposed the maximum likelihood method to estimate the model. However, as is pointed out by Kakamu and Wago (2007), the maximum likelihood methods involve the restriction problem of spatial parameters and Kakamu and Wago (2007) showed that the Bayesian method can avoid the problem.

From a Bayesian point of view, Zellner (1971) and Box and Tiao (1973) studied the model and Percy (1992) and Koop (2003) examined the Markov chain Monte Carlo (MCMC) methods to estimate the model. Chib and Greenberg (1995) extended the model with vector autoregressive and vector moving average errors of the first order. Thus, we propose a Bayesian approach for the estimation of the models (spatial autoregressive and spatial error models) using MCMC methods. We also introduce the marginal likelihood to compare the models with and without spatial interaction. Moreover, our approach is illustrated by real data set.

In the real data example, we examined the economics of agglomeration with and without interregional spillovers by panel data during the periods 1991 to 2000, which is called “lost decade”. [1] Empirical results show that the spatial error SUR model is the best model and that (1) Average TFP is the source of economic growth in Japanese manufacturing industries in this decade. (2) Manufacturing industries in Japan became more labor intensive. (3) The economics of agglomeration and spill-over effects became smaller over time and spill-over effect vanished in 1993.

The rest of this paper is organized as follows. In Section 2, we summarize the spatial autoregressive SUR model. Section 3 obtains a joint posterior distribution and discusses computational strategy of the MCMC method. In Section 4, we introduce the economic model and examine the economics of agglomeration in Japan during the period 1991 to 2000. Finally, brief conclusions and remaining issues are given in Section 5.

[1] In 1991, the collapse of bubbles occurred and Japan experienced the big recession, which Japan has never experienced. Therefore, it is the concern of Japanese economy to examine what happened in this decade.

2 SPATIAL AUTOREGRESSIVE SEEMINGLY UNRELATED REGRESSION MODEL

In this section, we will introduce the spatial autoregressive seemingly unrelated regression (SAR-SUR) model. [2] Let y_{it} and $\mathbf{x}_{it}$ be a dependent variable and a $1 \times k$ vector of covariates on ith unit ($i = 1, \cdots, N$) and tth period ($t = 1, \cdots, T$), respectively.

In the spatial model, the weight matrix $\mathbf{W}$ plays an important role. Therefore, we briefly explain the weight matrix, which we use in this paper. We use the contiguity weight matrix and the definition of contiguity obviously assumes the existence of a map, from which the boundaries can be discerned. For example, Figure 1 shows the 5 spatial units' example. In this figure, unit 1 is connected to unit 2, 3 and 4. Unit 2 is connected to 1, 3, 4 and 5 and so on.

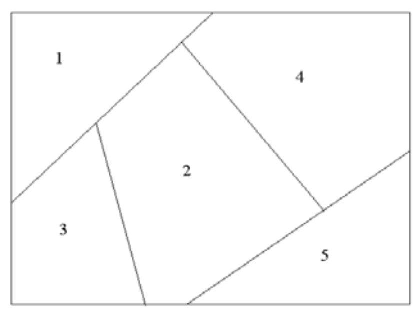

Figure 1. An example of contiguity weight matrix

[2] Spatial autoregressive model is widely used in examining spatial interaction (see *e.g.*, Kakamu *et al*., 2007). Therefore, we will introduce the model in this section. However, there are several spatial models (see *e.g.*, Anselin, 1988) and we will also examine the spatial error model (SEM) in this paper. It is possible to construct a spatial error SUR (SEM-SUR) model with some modifications of SAR-SUR model.

Therefore, the following weight matrix is constructed.

	unit 1	*unit* 2	*unit* 3	*unit* 4	*unit* 5
unit 1	0	1	1	1	0
unit 2	1	0	1	1	1
unit 3	1	1	0	0	0
unit 4	1	1	0	0	1
unit 5	0	1	0	1	0

In this weight matrix, the diagonal elements take 0 and the connections are expressed by 1. Thus, let w_{ij} denote the spatial weight on jth unit with respect to ith unit, that is, the ijth element of $\mathbf{W}$. In the SAR-SUR model, the coefficients are constant across space, but vary for each time period. The error terms are temporally correlated i.e., there is a constant covariance between errors for different time periods for the same spatial units. Then, the SAR-SUR model conditioned on parameters ρ_t, $\boldsymbol{\beta}_t$ and ω_{ts} for $t, s = 1, \cdots, T$ is written as follows:

$$y_{it} = \rho_t \sum_{j=1}^{N} w_{ij} y_{jt} + \mathbf{x}_{it}\boldsymbol{\beta}_t + \epsilon_{it}, \text{ with } E[\epsilon_{it}\epsilon_{is}] = \omega_{ts} \tag{1}$$

In matrix form, the equation for each time period t becomes:

$$\mathbf{y}_t = \rho_t \mathbf{W}\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \text{ with } E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_s'] = \omega_{ts}\mathbf{I}_N \tag{2}$$

where $\mathbf{y}_t$ and $\boldsymbol{\epsilon}_t$ are $N \times 1$ vectors, $\mathbf{W}$ is a $N \times N$ weight matrix, $\mathbf{X}_t$ is a $N \times k$ matrix of independent variables and $\mathbf{I}_N$ is a $N \times N$ unit matrix. The equations for time periods 1 to T are combined as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{bmatrix} = \begin{bmatrix} \rho_1\mathbf{W} & 0 & \cdots & 0 \\ 0 & \rho_2\mathbf{W} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_T\mathbf{W} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{bmatrix} + \begin{bmatrix} \mathbf{X}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{X}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{X}_T \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_T \end{bmatrix} \tag{3}$$

or, grouped

$$\mathbf{y} = (\mathbf{D} \otimes \mathbf{W})\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \Omega \otimes \mathbf{I}_N) \tag{4}$$

where $\mathbf{y}$ is a $NT \times 1$ vector of dependent variables, $\mathbf{D}$ is a $T \times T$ diagonal matrix with $\rho_1, \cdots, \rho_T$ as its elements, $\mathbf{X}$ is a $NT \times kT$ block diagonal matrix, $\boldsymbol{\epsilon}$ is a $NT \times 1$ error vector, and Ω is a $T \times T$ matrix with ω_{ts} as its elements.

Then, the likelihood function of the model (4) is given as:

$$L(\mathbf{y}|\mathbf{D}, \boldsymbol{\beta}, \Omega^{-1}, \mathbf{X}, \mathbf{W}) \propto \left\{ \prod_{t=1}^{T} |\mathbf{I}_N - \rho_t \mathbf{W}| \right\} |\Omega^{-1}|^{\frac{N}{2}} \exp\left\{ -\frac{\mathbf{e}'(\Omega^{-1} \otimes \mathbf{I}_N)\mathbf{e}}{2} \right\} \tag{5}$$

where $\mathbf{e} = \mathbf{y} - (\mathbf{D} \otimes \mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$.

3 POSTERIOR ANALYSIS

3.1 Joint posterior distribution

Since we adopt a Bayesian approach, we complete the model by specifying the prior distribution over the parameters. Therefore, we apply the following prior;

$$\pi(\mathbf{D}, \boldsymbol{\beta}, \Omega^{-1}) = \left\{ \prod_{t=1}^{T} \pi(\rho_t) \right\} \pi(\boldsymbol{\beta})\pi(\Omega^{-1})$$

Given a prior density $\pi(\mathbf{D}, \boldsymbol{\beta}, \Omega^{-1})$ and the likelihood function given in (5), the joint posterior distribution can be expressed as

$$\begin{aligned} &\pi(\mathbf{D}, \boldsymbol{\beta}, \Omega^{-1}|\mathbf{y}, \mathbf{X}, \mathbf{W}) \\ &\quad \propto \pi(\mathbf{D}, \boldsymbol{\beta}, \Omega^{-1}) L(\mathbf{y}|\mathbf{D}, \boldsymbol{\beta}, \Omega^{-1}, \mathbf{X}, \mathbf{W}), \\ &\quad \propto \left\{ \prod_{t=1}^{T} \frac{1}{\lambda_{\max}^{-1} - \lambda_{\min}^{-1}} \mathbf{1}_{[\lambda_{\min}^{-1}, \lambda_{\max}^{-1}]}(\rho_t) \right\} \\ &\qquad \times \exp\left\{ -\frac{(\boldsymbol{\beta} - \boldsymbol{\beta}_*)'\Sigma_*^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_*)}{2} \right\} \\ &\qquad \times |\Omega^{-1}|^{\frac{\nu_* - T - 1}{2}} \exp\left\{ -\frac{\mathrm{tr}(\Omega^{-1}\Omega_*^{-1})}{2} \right\} \\ &\qquad \times \left\{ \prod_{t=1}^{T} |\mathbf{I}_N - \rho_t \mathbf{W}| \right\} |\Omega^{-1}|^{\frac{N}{2}} \\ &\qquad \times \exp\left\{ -\frac{\mathbf{e}'(\Omega^{-1} \otimes \mathbf{I}_N)\mathbf{e}}{2} \right\} \end{aligned} \tag{6}$$

where $\mathbf{1}_{[a,b]}(x)$ is an indicator function, which takes 1 when x is in the interval between a and b.

Finally, we assume the following prior distributions;

$$\begin{aligned} \rho_t &\sim \mathcal{U}(\lambda_{\min}^{-1}, \lambda_{\max}^{-1}), \\ \boldsymbol{\beta} &\sim \mathcal{N}(\boldsymbol{\beta}_*, \Sigma_*), \\ \Omega^{-1} &\sim \mathcal{W}(\Omega_*, \nu_*) \end{aligned}$$

where $\mathcal{W}(\mathbf{M}, a)$ denotes an wishart distribution with scale matrix $\mathbf{M}$ and degree of freedom a. $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalues of $\mathbf{W}$. As is shown in Sun *et al*. (1999), it is well known that $\lambda_{\min}^{-1} < 0$ and $\lambda_{\max}^{-1} > 0$ and ρ_t must lie in the interval. Therefore, we restrict the prior space as $\rho_t \in (\lambda_{\min}^{-1}, \lambda_{\max}^{-1})$.

3.2 Posterior simulation

Since the joint posterior distribution is given by (6), we can now use MCMC methods. The Markov chain sampling scheme can be constructed from the full conditional distributions of ρ_t for $t = 1, \cdots, T$, $\boldsymbol{\beta}$ and Ω^{-1}.

3.2.1 Sampling ρ_t for $t = 1, \cdots, T$

From (6), the full conditional distribution of ρ_t is written as:

$$\pi(\rho_t|\mathbf{D}_{-t}, \boldsymbol{\beta}, \Omega^{-1}, \mathbf{y}, \mathbf{X}, \mathbf{W}) \propto |\mathbf{I}_N - \rho_t \mathbf{W}| \exp\left\{-\frac{\mathrm{tr}(\Omega^{-1}\mathbf{E}'\mathbf{E})}{2}\right\} \quad (7)$$

where $\mathbf{D}_{-t}$ is $\mathbf{D}$ without ρ_t and $\mathbf{E}$ is $N \times T$ matrix with vec$(\mathbf{E}) = \mathbf{e}$, which cannot be sampled by standard methods. Therefore, we adopt a following random-walk Metropolis algorithm (see *e.g.*, Tierney, 1994).

Sample ρ_t^{new} from

$$\rho_t^{new} = \rho_t^{old} + c\phi, \quad \phi \sim \mathcal{N}(0,1). \quad (8)$$

The scaler c is called tuning parameter and ρ_t^{old} is the parameter of the previous sampling. Next, we evaluate the acceptance probability

$$\alpha(\rho_t^{old}, \rho_t^{new}) = \min\left\{\frac{\pi(\rho_t^{new}|\mathbf{D}_{-t}, \boldsymbol{\beta}, \Omega^{-1}, \mathbf{y}, \mathbf{X}, \mathbf{W})}{\pi(\rho_t^{old}|\mathbf{D}_{-t}, \boldsymbol{\beta}, \Omega^{-1}, \mathbf{y}, \mathbf{X}, \mathbf{W})}, 1\right\} \quad (9)$$

Finally we set $\rho_t = \rho_t^{new}$ with probability $\alpha(\rho_t^{old}, \rho_t^{new})$, otherwise $\rho_t = \rho_t^{old}$. The scalar c is tuned to produce an acceptance rate between 10% and 30% as is suggested in Holloway *et al.* (2002). [3] It should be mentioned that the proposal density of ρ_t is not truncated to the interval $(\lambda_{min}^{-1}, \lambda_{max}^{-1})$ since the constraint is part of the target density. Thus, if the proposal value of ρ_t is not within the interval, the conditional posterior is zero, and the proposal value is rejected with probability one (see Chib and Greenberg, 1998).

3.2.2 Sampling $\boldsymbol{\beta}$ and Ω^{-1}

The full conditional distributions for $\boldsymbol{\beta}$ and Ω^{-1} are as follows:

$$\begin{aligned} \pi(\boldsymbol{\beta}|\mathbf{D}, \Omega^{-1}, \mathbf{y}, \mathbf{X}, \mathbf{W}) &\propto \mathcal{N}(\boldsymbol{\beta}_{**}, \Sigma_{**}), \\ \pi(\Omega^{-1}|\mathbf{D}, \boldsymbol{\beta}, \mathbf{y}, \mathbf{X}, \mathbf{W}) &\propto \mathcal{W}(\Omega_{**}, \nu_{**}), \end{aligned}$$

with $\bar{\mathbf{y}} = \mathbf{y} - (\mathbf{D} \otimes \mathbf{W})\mathbf{y}$, $\boldsymbol{\beta}_{**} = \Sigma_{**}\{\mathbf{X}'(\Omega^{-1} \otimes \mathbf{I}_N)\bar{\mathbf{y}} + \Sigma_*^{-1}\boldsymbol{\beta}_*\}$, $\Sigma_{**} = \{\mathbf{X}'(\Omega^{-1} \otimes \mathbf{I}_N)\mathbf{X} + \Sigma^{-1}\}^{-1}$, $\Omega_{**} = (\mathbf{E}'\mathbf{E} + \Omega_*^{-1})^{-1}$ and $\nu_{**} = N + \nu_*$.

These parameters are easily sampled from Gibbs sampler (see *e.g.*, Gelfand and Smith, 1990).

[3] In Holloway *et al.* (2002), the scalar c is selected to make the acceptance rate between 40% and 60%. However, we considered the inefficiency factor and select the scalar c to make the inefficiency factor smallest.

4 EMPIRICAL EXAMPLE

4.1 A regional economic model for Japan during 1991-2000

We extend the regional production function model of Kanemoto *et al.* (1996) to the spatial agglomeration model for 47 Japanese prefectures during the decade from 1991 to 2000. We estimate aggregate production functions for prefectures to derive the dynamics of agglomeration economies and interregional spill-overs [4]. An aggregate production function in a prefecture is written as $Y = F(L, K, S)$, where L, K, S, and Y are respectively the employment, the private capital, the spill-overs, and the total production (or value added) in a prefecture. We assume that in the absence of agglomeration economies the production function exhibits constant returns to scale with respect to labor and capital inputs. The degree of agglomeration economies can then be measured by the degree of increasing returns to scale of the estimated production function.

This approach is justified if we assume that technological externalities exist between firms in a prefecture. For example, suppose a firm in a prefecture receives external benefits from urban agglomeration, measured by the total employment L, and from spill-overs S. Assuming that the firm uses labor n and (private) capital k as inputs, we can write its production function as $f(n, k, L, S)$. For expositional simplicity, we assume that all firms are identical. The total production in a prefecture is then $Y = mf(L/m, K/m, L, S)$, where m is the number of firms in a prefecture. Free entry of firms guarantees that the size of an individual firm is determined such that the production function of an individual firm, $f(n, k, L, S)$, exhibits constant returns to scale with respect to n and k. This condition determines the number of firms m as a function of other variables, $m = m^*(L, K, S)$. The aggregate production function is then

$$F(L,K,S) = m^*(L,K,S) \, f\left(\frac{L}{m^*(L,K,S)}, \frac{K}{m^*(L,K,S)}, L, S\right)$$

This aggregate production function satisfies

$$\begin{aligned} F_L(L,K,S) &= m\left[\frac{1}{m}f_n + f_L\right] + m_L^*[f - nf_n - kf_k] \\ &= f_n(n,k,L,S) + mf_L(n,k,L,S), \end{aligned}$$

where subscripts denote partial derivatives and the second square bracket equals zero because of

[4] We have to mention that our main concern is in the transition of agglomeration economies, although Kanemoto *et al.* (1996) is interested not only in the magnitude of agglomeration economies but also in testing Tokyo is too large or not. Therefore, we only focus on the estimation of agglomeration economies in this paper.

constant-returns-to-scale condition. The last term mf_L measures the marginal benefits of agglomeration economies.

Although a variety of functional forms are possible for the urban production function, we consider a simple Cobb-Douglas type:

$$Y_{it} = A_{it} K_{it}^{\alpha_t} L_{it}^{\gamma_t}. \tag{10}$$

The magnitude of agglomeration economies can be measured by the degree of scale economy, $\alpha_t + \gamma_t - 1$.

Next, we will introduce the interregional spill-overs. We examine interregional spill-overs using the spatial econometrics technique. Suppose that the interregional spill-overs change the total factor productivity A_{it} and is related to the regional weight w_{ij} and Y_{jt}/L_{jt}. This yields the following total factor productivity:

$$A_{it} = A_{0t} \times \prod_j \left(\frac{Y_{jt}}{L_{jt}} \right)^{\rho_t w_{ij}} \tag{11}$$

ρ is the intensity of spatial interaction. Substitute (11) for (10) and rearrange the equation, to obtain;

$$\ln\left(\frac{Y_{it}}{L_{it}}\right) = \ln A_{0t} + \rho_t \sum_j w_{ij} \ln\left(\frac{Y_{jt}}{L_{jt}}\right) + a_{1t} \ln\left(\frac{K_{it}}{L_{it}}\right) + a_{2t} \ln(L_{it}) \tag{12}$$

where $\alpha_t = a_{1t}$ and $\gamma_t = a_{1t} + a_{2t}$. Then, the economics of agglomeration is evaluated by a_2 (see Kanemoto *et al*., 1996) and we can estimate by spatial autoregressive model. However, as we use the panel data, we have to consider the correlation among periods. Therefore, we will use our SAR-SUR model.[5]

4.2 Empirical results

Before examining empirics, we explain the data used in this paper. Our data set stems from the Census of Manufactures prepared by the Ministry of International Trade and Industry (MITI) of Japan. For 47 prefectures, the total production is added values of manufacturing industries, the total capital is the amount on hand of permanent assets and the total employment is the common labors, which exclude part time labor. As a weight matrix $\mathbf{W}$, we use the contiguity dummy variables proposed by Kakamu *et al*. (2007), which considers the connection of economic activities[6] and average number of dummy variables are 4. Using the given following hyper-parameters;

$$\boldsymbol{\beta}_* = \mathbf{0}, \quad \Sigma_* = 100 \times \mathbf{I}_{kT},$$
$$\Omega_* = 100 \times \mathbf{I}_T, \quad \nu_* = T + 1,$$

we ran the MCMC algorithm using 100000 iterations and discarding the first 50000 iterations. All results reported here were generated using Ox version 4.10 (see Doornik, 2006). [7]

Table 1. Log marginal likelihood

SAR-SUR	SEM-SUR	SUR
335.323	357.766	344.764

First of all, we have to consider which model is the best model, that is, whether there exists interregional spill-over or not and which kind of spill-over effect exists if the spill-over effext exists. For the model choice, we use the marginal likelihoods proposed by Chib (1995).

Table 1 shows the log marginal likelihoods of all models: SAR-SUR, SEM-SUR and SUR models. From the results, we can find that the SEM-SUR model is the best model and we will show the result of SEM-SUR model hereafter. However, it is reasonable to compare the SEM-SUR model with the simple SUR model, because it implies that we compare the economics of agglomeration with and without spill-over effects, we also introduce the results of the SUR model.

Table 2 shows the coefficient estimates ($\ln A_0, a_1, a_2, \rho$) of the SEM-SUR model (above) and SUR model (below) in the selected years, respectively. Figure 2 also shows the (marginal) posterior distributions of the coefficients by box plots. First of all, we see that all the coefficients except for ρ in 1993-2000 do not include zero in the 95% credible interval. When we compare the results with and without spill-overs, in Table 2, can find that all the posterior means in SEM-SUR are larger than those in SUR model. It implies that if we ignore the spill-over effects, we might misinterpret the source

[5] If we replace the spill-over in (12) to $A_{it} = A_{0t} \times \prod_j (Y_{jt}/A_{0t} L_{jt}^{\alpha} K_{jt}^{\gamma})^{\rho_t w_{ij}}$, then we can construct SEM-SUR model. Of course, if we assume A_{it} is equal in all regions, it is reduced to simple SUR model.

[6] All except one (Okinawa) Japanese prefectures are situated on the four major islands, Hokkaido, Honshu, Shikoku and Kyushu. But these four islands are connected by train and roads, despite the fact that islands are separate geographical entities. But for example, the most northern island Hokkaido is connected by the Seikan railway tunnel to Honshu. And Honshu is connected by the Awaji and Seto Bridge to Shikoku, and the southern island of Kyushu is also connected by the Kanmon Tunnel and Bridge to Honshu. Therefore, Okinawa is the only prefecture which is independent of all other prefectures. In addition, the weight matrix is row standardized, that is, $\sum_{j=1}^{N} w_{ij} = 1$, for identification.

[7] The code as well as the derivation of the algorithm were checked using the joint distribution test described in Geweke (2004) and the convergence of MCMC is checked by convergence diagnostic proposed by Geweke (1992).

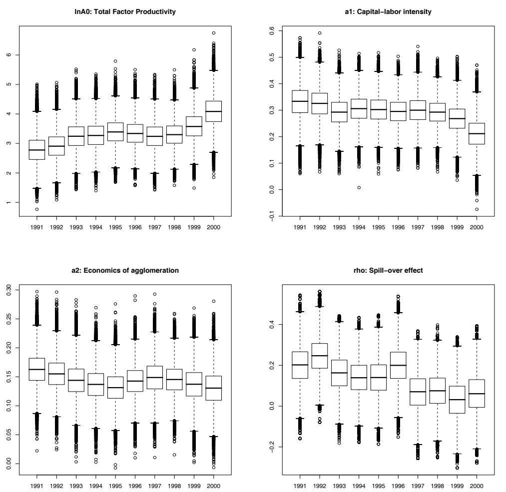

Figure 2. Box plots of $\ln A_0$, a_1, a_2 and ρ

Table 2. Empirical results: Posterior means and standard deviations (in parenthesis)

	Spatial error SUR model					
	1991		1993		1995	
$\ln A_0$	2.792	(0.490)	3.254	(0.476)	3.400	(0.456)
a_1	0.332	(0.062)	0.293	(0.055)	0.303	(0.053)
a_2	0.163	(0.029)	0.144	(0.029)	0.131	(0.028)
ρ	0.201	(0.098)	0.162	(0.092)	0.140	(0.091)
	1996		1998		2000	
$\ln A_0$	3.348	(0.452)	3.305	(0.440)	4.092	(0.522)
a_1	0.295	(0.052)	0.293	(0.050)	0.211	(0.059)
a_2	0.143	(0.027)	0.145	(0.027)	0.130	(0.031)
ρ	0.198	(0.097)	0.075	(0.091)	0.061	(0.098)
	SUR model					
	1991		1993		1995	
$\ln A_0$	3.009	(0.526)	3.493	(0.496)	3.607	(0.473)
a_1	0.295	(0.064)	0.262	(0.055)	0.278	(0.052)
a_2	0.165	(0.031)	0.141	(0.032)	0.128	(0.030)
	1996		1998		2000	
$\ln A_0$	3.665	(0.473)	3.478	(0.453)	4.292	(0.534)
a_1	0.252	(0.052)	0.275	(0.049)	0.193	(0.058)
a_2	0.140	(0.030)	0.141	(0.028)	0.125	(0.032)

of economic growth because $\ln A_0$, a_1 and a_2 are underestimated in case of SUR model. Hence, when we focus on the result of SEM-SUR model in Figure 2, there are following tendencies in each parameter. The posterior means of $\ln A_0$ became larger over time. It means that the average TFP was a driving force of economic growth in Japanese manufacturing industries in this decade. However, if we consider the fact that Japan experienced serious recession, we might be able to conclude that the average TFP was not so strong force for economic growth compared with the other sources of economic growth. On the other hand, the capital-labor intensity, a_1, became smaller over time, especially in 2000. It implies that manufacturing industries in Japan became more labor intensive[8]. The coefficient of the economics of agglomeration, a_2, also became smaller over time.

[8] In spite of the fact that the Japanese manufacturing industries are automated, which means capital intensive, after the oil shock drastically, the result is opposite to the fact. It is interesting result but it may be caused by the data set problem or it needs more detail analysis to examine what happened to the manufacturing industries in this decade.

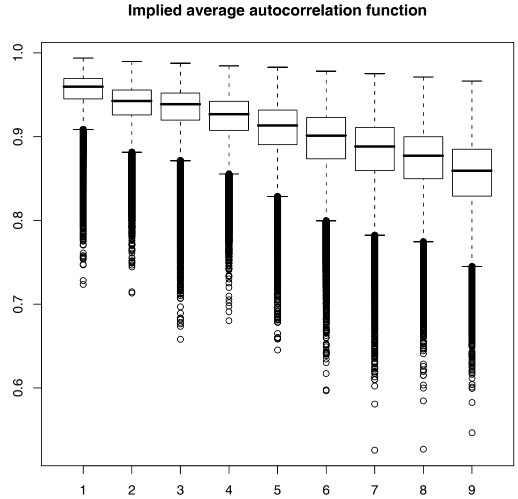

Figure 3. Box plot autocorrelation function

Moreover, concerning the spill-over effect, ρ, we note that the effect decline and vanished in 1993. From the evidences, we conclude that the power of economics of agglomeration and spill-over effects became smaller and smaller after the collapse of bubbles in 1991, that is, the big recession in the 1990s, which is also called "lost decade", might be associated with a declining importance of the economics of agglomeration and spill-over effects.

Figure 3 shows the box plots (of off-diagonal elements of correlation matrix) in SEM-SUR model, which can be interpreted as an implied autocorrelation function. It shows that the serial correlation plays an important role in analysing production function in Japanese manufacturing industries. The high serial correlation can be also seen on box plots in Figure 3.

5 CONCLUSIONS

This paper examined the SUR model with spatial dependencies from a Bayesian point of view. We expressed the joint posterior distribution of the model, and proposed MCMC methods to estimate the parameters of the model. We have illustrated our approach using Japanese manufacturing industries' data and examined the economics of agglomeration in Japan during the period 1991 to 2000.

From the results, we found that the SEM-SUR model is the best model and serial correlation played an important role. In addition, following tendencies are confirmed: (1) Average TFP is the source of economic growth in Japanese manufacturing industries in this decade. (2) Manufacturing industries in Japan became more labor intensive. (3) The economics of agglomeration and spill-over effects became smaller over time and spill-over effect vanished in 1993.

Finally, we will mention about the remaining issues. As is pointed out above, the result that the Japanese manufacturing industries became labor intensive seems to be different from the situation in this decade in Japan. We cannot conclude what happened to the Japanese manufacturing industries in this decade only from our empirical results. However, it beyonds our analysis and it requires other moderate model or theory, like human capital theory. Moreover, we mentioned that the serial correlation plays an important role in this model. If we extend the model by Chib and Greenberg (1995) to spatial model, the serial correlation may become small. However, as our main purpose is to examine the SUR model with spatial dependencies, it also beyond our analysis, but our findings from Japanese cases represent an interesting firs step.

6 ACKNOWLEDGEMENTS

We would like to acknowledge to John Geweke, Ivan Jeliazkov, Yasuhiro Omori and Ruey Tsay for their helpful comments. We are also grateful to seminar/conference participants at Valencia/ISBA 8th World Meeting on Bayesian Statistics in Benidorm (Alicante), 61st European Meeting of the Econometric Society in Vienna, Joint Statistical Meetings 2006 in Seattle, 2006 Japanese Economic Association Autumn Meeting at Osaka City University and International Workshop on Bayesian Statistics and Applied Econometrics at Tohoku University. The research for this work was supported by Grant-in-aid for Scientific Research (A)(1) 15200022. This work is done while Kazuhiko Kakamu was staying at Institute for Advanced Studies.

7 REFERENCES

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

Box G.E.P., Tiao G.C. (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA.

Chib S. (1995), Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, **90**, 1313–1321.

Chib A. (2001), Markov chain Monte Carlo methods: Computation and inference, *Handbook of Econometrics: volume 5*, Heckman J.J., Leamer E. (eds). Amsterdam: North Holland, 3569–3649.

Chib S, Greenberg E. (1995), Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models, *Journal of Econometrics*, **68**, 339–360.

Chib S, Greenberg E. (1998), Analysis of multivariate probit models, *Biometrika*, **88**, 347–361.

Chib S, Jeliazkov I. (2001), Marginal likelihood from the Metropolis-Hastings output, *Journal of the American Statistical Association*, **96**, 270–281.

Doornik JA. (2006), *Ox: An Object Oriented Matrix Programming Language*, Timberlake Consultants Press, London.

Gamerman D, Lopes H.F. (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd edition)*, Chapman & Hall/CRC.

Gelfand A.E., Smith A.F.M. (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*, Bernard J.M., Berger J.O., Dawid A.P., Smith A.F.M. (eds). Oxford University Press, Oxford.

Geweke J. (2004), Getting it right: Joint distribution tests of posterior simulators, *Journal of the American Statistical Association*, **99**, 799–804.

Kakamu K, Polasek W, Wago H. (2007), Spatial interaction of crime incidents in Japan, forthcoming in *Mathematics and Computers in Simulation*.

Kakamu K, Wago H. (2007), Small sample properties of panel spatial autoregressive models: Comparison of Bayesian and maximum likelihood methods, *mimeo*.

Kanemoto Y, Ohkawara T, Suzuki T. (1996), Agglomeration economies and a test for optimal city size in Japan, *Journal of the Japanese and International Economies*, **10**, 379-398.

Koop G. (2003), *Bayesian Econometrics*, John Wiley & Sons Ltd., West Sussex, England.

Percy DF. (1992), Prediction for seemingly unrelated regression, *Journal of the Royal Statistical Society Series B*, **54**, 243–252.

Sun D, Tsutakawa R.K, Speckman P.L. (1999), Posterior distribution of hierarchical models using CAR(1) distributions, *Biometrika*, **86**, 341–350.

Tierney L. (1994), Markov chains for exploring posterior distributions (with discussion), *Annals of Statistics*, **22**, 1701–1762.

Zellner A. (1962), An efficient method of estimating seemingly unrelated regressions and test for aggregation bias, *Journal of the American Statistical Association*, **57**, 348–368.

Zellner A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Weily, New York.