Spatial agglomeration and spill-over analysis for Japanese prefectures during 1991-2000*

Kazuhiko Kakamu†
Department of Economics & Finance
Institute for Advanced Studies
and
Graduate School of Economics,
Osaka University

Wolfgang Polasek
Department of Economics & Finance
Institute for Advanced Studies

Hajime Wago
Graduate School of Economics
Nagoya University

April 21, 2007

Abstract

This paper considers the seemingly unrelated regression (SUR) model with spatial dependency from a Bayesian point of view. We consider Markov chain Monte Carlo methods to estimate the parameters of the model. We analyze the economics of agglomeration in Japanese prefecture during the period 1991 to 2000. From our empirical results, we found that the spatial autoregressive SUR model is the best model and that the economics of agglomeration decreased in this decade, but spatial interaction played an important role.

Key Words: Economics of agglomeration; Marginal likelihoods; Markov chain Monte Carlo (MCMC); Panel data; Seemingly unrelated regression (SUR); Spatial models.
JEL Classification: C11, C15, C23, R30.

*We would like to acknowledge comments of the participants at Valencia/ISBA 8th World Meeting on Bayesian Statistics in Benidorm (Alicante). The research for this work was supported by Grant-in-aid for Scientific Research (A)(1) 15200022.

†Corresponding author. *Mail*: Stumpergasse 56, 1060, Vienna, Austria. *Email*: kakamu@ihs.ac.at

1 Introduction

The seemingly unrelated regression (SUR) model was first proposed by Zellner (1962) and has been widely used for estimating system equations. Zellner (1971) and Box and Tiao (1973) studied the model from a Bayesian point of view and Percy (1992), Chib and Greenberg (1995) and Koop (2003) used the Markov chain Monte Carlo (MCMC) methods to estimate SUR models.

Since the seminal work by Anselin (1988), spatial interactions have become an important tool in econometrics. Spatial applications have become popular in applied sciences,like in economics and also social sciences; A special topic is the distribution of crime incidents, see e.g. Anselin (1988) or Kakamu *et al.* (2005) and in economics, Case (1991) studied the spatial patterns in household demand. However, all theses estimations are mostly done by univariate spatial models. In this paper, we consider the spatial SUR model, which takes the multivariate case into account, from a Bayesian point of view and construct the estimation methods of the model using MCMC methods. Also we introduce the marginal likelihood to compare the models with and without spatial interaction. Moreover, we illustrate our approach by estimating a regional production function for 47 Japanese regions and we study spatial spill-over effects.

For the regional production function, we study the economics of agglomeration with and without interregional spill-over effects using panel data during the period 1991 to 2000, which is called in Japan the "lost decade". Our estimation results show that the spatial autoregressive SUR model is the preferred model and that (1) average total factor productivity is the driving force of economic growth in manufacturing industries in Japan; (2) manufacturing industries in Japan became more labor intensive; (3) the economics of agglomeration became smaller over time; and (4) the spill-over effects play a small but important role.

The rest of this paper is organized as follows. In Section 2, we introduce the spatial SUR model. Section 3 describes the joint posterior distribution and discusses the computational approach by the MCMC method. Section 4 explains the MCMC estimation of the spatial autoregressive (SAR-SUR) and the spatial error model (SUR-SEM). We also derive how to calculate marginal likelihoods for the spatial model selection. In Section 5, we introduce the regional production function model and examine the economics of agglomeration in Japan during the period 1991 to 2000. Finally, brief conclusions are given in Section 5.

2 Spatial SUR Models

For the estimation of the regional production function we consider a multivariate regression model (a panel data set) where a cross sectional production function is estimated for each time point t across n regions. Let y_{it} denote the dependent variable on i-th unit (region) and the t-th equation, let x_{it}, which also includes the constant term, denote independent variables, where x_{it} is a $1 \times k$ vector of all the i-th units in the t-th equation, and let w_{ij} denote the ij-th element of the weight matrix W. Also suppose that

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}, \quad y_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{Nt} \end{pmatrix}, \quad X = \begin{pmatrix} X_1 & & \\ & \ddots & \\ & & X_T \end{pmatrix}, \quad X_t = \begin{pmatrix} x_{1t} \\ \vdots \\ x_{Nt} \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_T \end{pmatrix}, \quad D_\rho = diag(\rho_1, \cdots, \rho_T), \quad D_\rho \otimes W = \begin{pmatrix} \rho_1 W & & \\ & \ddots & \\ & & \rho_T W \end{pmatrix},$$

where the matrices are $y : NT \times 1$, $y_t : N \times 1$, $X : NT \times kT$, $X_t : N \times k$ and $\beta : kT \times 1$, respectively.

In this cross-sectional SUR model, the equations can be correlated and therefore we define by Ω the covariance matrix across equations. Then the spatial autoregressive SUR (or SAR-SUR) model contains the parameters $(\beta, \Omega^{-1}, D_\rho)$ is written as follows;

$$y = X\beta + (D_\rho \otimes W)y + \epsilon, \quad \epsilon \sim \mathcal{N}\left(0, \Omega \otimes I_N\right), \tag{1}$$

where I_N is $N \times N$ unit matrix.

Lemma: The likelihood function L of the model (**??**) is given by;

$$L(y|\beta, \Omega^{-1}, D_\rho, X, W) = \frac{1}{\sqrt{2\pi}^{NT}} |\Omega^{-1}|^{N/2} |I_{NT} - (D_\rho \otimes W)| \exp\left[-\frac{e'(\Omega^{-1} \otimes I_N)e}{2}\right], \quad (2)$$

where $e = y - X\beta - (D_\rho \otimes W)y$.

Proof: The model is $y = X\beta + (D_\rho \otimes W)y + \epsilon$, $\epsilon \sim \mathcal{N}(0, \Omega \otimes I_N)$. We rewrite it as $y = F^{-1}X\beta + F^{-1}\epsilon$, where $F = (I_{NT} - D_\rho \otimes W)$. Then, the variance of $F^{-1}\epsilon$ is:

$$\begin{aligned} Var(F^{-1}\epsilon) &= J(F^{'-1})(\Omega \otimes I_N)J(F^{-1}) \\ &= \frac{1}{|F'^{-1}|^{\frac{1}{2}}|F^{-1}|^{\frac{1}{2}}}(\Omega \otimes I_N) \\ &= \frac{1}{|F^{-1}|}(\Omega \otimes I_N) \\ &= |F|(\Omega \otimes I_N) \\ &= |I_{NT} - D_\rho \otimes W|(\Omega \otimes I_N). \end{aligned} \quad (3)$$

Therefore, the determinant appears in (**??**), as the Jacobian of the transformation is $|\frac{\partial \epsilon}{\partial y}| = |F|$. Q.E.D.

3 Posterior analysis of the spatial SUR models

3.1 Joint posterior distribution of SAR-SUR model

Since we adopt a Bayesian approach, we complete the model by specifying the prior distribution for all the parameters. Therefore, we assume the following priors;

$$p(\beta, \Omega^{-1}, D_\rho) = p(\beta)p(\Omega^{-1}) \prod_{t=1}^{T} p(\rho_t),$$

and

$$\beta \sim \mathcal{N}(\beta_*, \Sigma_*), \quad \Omega^{-1} \sim \mathcal{W}(\Omega_*, \nu_*),$$
$$\rho_t \sim \mathcal{U}(\lambda_{min}^{-1}, \lambda_{max}^{-1}), \quad t = 1, \cdots, T,$$

where $\mathcal{W}(A, b)$ denotes a Wishart distribution with scale parameter A and degrees of freedom b. λ_{min} and λ_{max} denote the minimum and maximum eigenvalues of 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 $\rho_t \in (\lambda_{min}^{-1}, \lambda_{max}^{-1})$.

Using the prior density $p(\beta, \Omega^{-1}, D_\rho)$ and the likelihood function as in (**??**), the joint posterior distribution can be expressed as

$$
\begin{aligned}
&p(\beta, \Omega^{-1}, D_\rho | y, X, W) \\
&\quad\quad \propto \; p(\beta, \Omega^{-1}, D_\rho) L(y|\beta, \Omega^{-1}, D_\rho, X, W) \\
&\quad\quad \propto \; |\Sigma_*|^{-1} \exp\left[-\frac{(\beta - \beta_*)\Sigma_*^{-1}(\beta - \beta_*)}{2} \right] \\
&\quad\quad\quad \times |\Omega^{-1}|^{(\nu_* - T - 1)/2} \exp\left[-\frac{tr(\Omega^{-1}\Omega_*^{-1})}{2} \right] \\
&\quad\quad\quad \times |\Omega^{-1}|^{N/2} |I_{NT} - (D_\rho \otimes W)| \exp\left[-\frac{e'(\Omega^{-1} \otimes I_N)e}{2} \right]. \qquad (4)
\end{aligned}
$$

3.2 Posterior simulation of the SAR-SUR model

Based on the joint posterior distribution given in (**??**), we can now derive conditional distributions for the MCMC estimation. The Markov chain sampling scheme is now constructed from the following full conditional distributions of β, Ω^{-1} and ρ_t for $t = 1, \cdots, T$.

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

We know $|I_{NT} - (D_\rho \otimes W)| = \prod_{t=1}^{T} |I_N - \rho_t W|$ by the properties of a block diagonal matrix. Using (**??**), the full conditional distribution of ρ_t is proportional to

$$
p(\rho_t | \beta, \Omega^{-1}, D_{\rho_{-t}}, y, X, W) \propto |I_N - \rho_t W| \exp\left(-\frac{tr(E'E\Omega^{-1})}{2} \right), \qquad (5)
$$

where E is the residual matrix with $vec(E) = e$. Since this expression cannot be sampled by standard methods, we adopt a random-walk Metropolis step at this point of the sampler (see e.g., Tierney, 1994). [1]

[1] We also tried to apply the independence chain. However, we recommend the random-walk chain, because it is faster. Some more discussions are found in Appendix B.

We sample ρ_t^{new} from

$$\rho_t^{new} = \rho_t^{old} + c\phi, \quad \phi \sim \mathcal{N}(0,1). \tag{6}$$

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

$$\alpha(\rho_t^{old}, \rho_t^{new}) = min\left(\frac{p(\rho_t^{new})}{p(\rho_t^{old})}, 1\right), \tag{7}$$

where p is the full conditional distribution in (**??**) and, of course, ρ_t in E also changes to ρ_t^{new} in $p(\rho_t^{new})$. 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). 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).

In the sampling of ρ_t for $t = 1, \cdots, T$, the following density is used;

$$\begin{gathered} p(\rho_1|\beta, \Omega^{-1}, \rho_2^{old}, \cdots, \rho_T^{old}), \\ p(\rho_2|\beta, \Omega^{-1}, \rho_1, \rho_3^{old}, \cdots, \rho_T^{old}), \\ \vdots \\ p(\rho_T|\beta, \Omega^{-1}, \rho_1, \cdots, \rho_{T-1}). \end{gathered}$$

3.2.2 Sampling the other parameters

The full conditional distribution for β is given by

$$\begin{aligned} & p(\beta|\Omega^{-1}, D_\rho, y, X, W) \\ & \qquad \propto \ \exp\left[-\frac{(\beta-\beta_*)\Sigma_*^{-1}(\beta-\beta_*)}{2}\right] \exp\left[-\frac{e'(\Omega^{-1}\otimes I_N)e}{2}\right] \\ & \qquad \propto \ \mathcal{N}(\beta_{**}, \Sigma_{**}), \end{aligned} \tag{8}$$

where $\beta_{**} = \Sigma_{**}(X'(\Omega^{-1} \otimes I_N)(y - (D_\rho \otimes W)y) + \Sigma_*^{-1}\beta_*)$, $\Sigma_{**} = (X'(\Omega^{-1} \otimes I_N)X + \Sigma_*^{-1})^{-1}$.

The full conditional distribution for Ω^{-1} is given by

$$
\begin{aligned}
& p(\Omega^{-1}|\beta, D_\rho, y, X, W) \\
& \quad \propto \quad |\Omega^{-1}|^{(\nu_*-T-1)/2} \exp\left[-\frac{tr(\Omega^{-1}\Omega_*^{-1})}{2}\right] |\Omega^{-1}|^{N/2} \exp\left[-\frac{tr(\Omega^{-1}E'E)}{2}\right] \\
& \quad \propto \quad \mathcal{W}(\Omega_{**}, \nu_{**}), \qquad (9)
\end{aligned}
$$

where $\Omega_{**} = (E'E + \Omega_*^{-1})^{-1}$ and $\nu_{**} = N + \nu_*$.

These fcd's are easily sampled through the Gibbs sampler (see e.g., Gelfand and Smith, 1990).

4 The SUR spatial error model (SUR-SEM)

The spatial error SUR (SEM-SUR) model conditioned on parameters $(\beta, \Omega^{-1}, D_\rho)$ is written as follows;

$$
y = X\beta + (D_\rho \otimes W)u + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \Omega \otimes I_N), \qquad (10)
$$

where $u = y - X\beta$.

Then, we will introduce the likelihood function L of the model (**??**) as follows;

$$
\begin{aligned}
& L(y|\beta, \Omega^{-1}, D_\rho, X, W) \\
& \quad = \quad \frac{1}{\sqrt{2\pi}^{NT}} |\Omega^{-1}|^{N/2} |I_{NT} - (D_\rho \otimes W)| \exp\left[-\frac{e'(\Omega^{-1} \otimes I_N)e}{2}\right] \qquad (11)
\end{aligned}
$$

where $e = y - X\beta - (D_\rho \otimes W)(y - X\beta)$.

The full conditional distributions of the model are as follows:

$$
\begin{aligned}
p(\rho_t|\beta, \Omega^{-1}, D_{\rho_{-t}}, y, X, W) \quad & \propto \quad |I_N - \rho_t W| \exp\left(-\frac{tr(E'E\Omega^{-1})}{2}\right), \qquad (12) \\
p(\beta|\Omega^{-1}, D_\rho, y, X, W) \quad & \propto \quad \mathcal{N}(\beta_{**}, \Sigma_{**}), \qquad (13) \\
p(\Omega^{-1}|\beta, D_\rho, y, X, W) \quad & \propto \quad \mathcal{W}(\Omega_{**}, \nu_{**}), \qquad (14)
\end{aligned}
$$

where $\beta_{**} = \Sigma_{**}((X - (D_\rho \otimes W)X)'(\Omega^{-1} \otimes I_N)(y - (D_\rho \otimes W)y) + \Sigma_*^{-1}\beta_*)$, $\Sigma_{**} = ((X - (D_\rho \otimes W)X)'(\Omega^{-1} \otimes I_N)(X - (D_\rho \otimes W)X) + \Sigma_*^{-1})^{-1}$, $\Omega_{**} = (E'E + \Omega_*^{-1})^{-1}$ and $\nu_{**} = N + \nu_*$. Eq. (**??**) cannot be sampled by standard methods. Therefore, we also adopt the random-walk Metropolis algorithm as is stated in Section 3.2.1.

4.1 Model selection by marginal likelihoods

For model M_k, let $L(y|\theta_k, M_k)$ and $p(\theta_k|M_k)$ be the likelihood and prior for the model, respectively. Then, the marginal likelihood of the model is defined as

$$m(y) = \int L(Y|\theta_k, M_k)p(\theta_k|M_k)d\theta_k.$$

Since the marginal likelihood can be written as

$$m(y) = \frac{L(y|\theta_k, M_k)p(\theta_k|M_k)}{p(\theta_k|y, M_k)},$$

Chib (1995) suggests to estimate the log marginal likelihood from the expression

$$\log m(y) = \log L(y|\theta_k^*, M_k) + \log p(\theta_k^*|M_k) - \log p(\theta_k^*|y, M_k),$$

where θ_k^* is a particular high density point (typically the posterior mean or the maximum likelihood (ML) estimate). He also provides a computationally efficient method to estimate the posterior ordinate $p(\theta_k^*|y, M_k)$ in the context of Gibbs sampling and Chib and Jeliazkov (2001) provides the method in the context of Metropolis-Hasting sampling. In the SUR-SAR model, for example, we set $\theta_k = (\beta, \Omega, D_\rho)$ and estimate the posterior ordinate $p(\theta_k^*|y, M_k)$ via the decomposition

$$\begin{aligned} & p(\theta_k^*|y, M_k) \\ &= p(\beta^*|D_\rho^*, \Omega^*, y, X, W)p(\Omega^*|\beta^*, D_\rho^*, y, X, W) \\ & \quad \prod_{t=1}^{T} p(\rho_t|\beta^*, \Omega^*, D_{\rho_{-t}}^*, y, X, W). \end{aligned}$$

The terms $p(\beta^*|D_\rho^*, \Omega^*, y, X, W)$ and $p(\Omega^*|\beta^*, D_\rho^*, y, X, W)$ are calculated from the Gibbs output (see Chib, 1995) and $p(\rho_t|\beta^*, \Omega^*, D_{\rho_{-t}}^*, y, X, W)$ is calculated from the Metropolis-Hasting output (see Chib and Jeliazkov, 2001). Using marginal likelihoods, we can now compare the following models: SUR, SAR-SUR, and SEM-SAR. [2]

[2] Note that if we drop the last term in **??** $\prod_{t=1}^{T} p(\rho_t|\beta^*, \Omega^*, D_{\rho_{-t}}^*, y, X, W)$, then we get the posterior ordinate of the ordinary SUR model.

5 Estimating the regional production function

5.1 A regional economic model for Japan 1991-2000

We extend the regional production function model of Kanemoto *et al.* (1996) to include a spatial agglomeration parameter for the 47 Japanese prefectures during the decade from 1991 to 2000. We estimate the Cobb-Douglas production function using all 47 prefectures and we are interested in the magnitude of urban agglomeration and interregional spill-over effects. The regional production function in a prefecture is specified 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 spatial extension of the production function is justified if there exist technological externalities between firms across prefectures. For example, suppose a firm in a prefecture receives external benefits from an urban agglomeration, measured by the total employment L, and the spill-overs are measured by S. Assuming that the firm uses labor n and (private) capital k as inputs, we can write the production function as $f(n, k, L, S)$. For 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

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

This aggregate production function satisfies

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

where subscripts denote partial derivatives and the second expression in square brackets equals zero because of constant-returns-to-scale condition. The last term mf_L measures the marginal benefits of the urban agglomeration economies.

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

$$Y_{it} = A_{it}K_{it}^{\alpha_t}L_{it}^{\gamma_t}. \tag{15}$$

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

Next, we will introduce the interregional spill-overs. We specify an interregional spill-over parameter using the spatial econometrics approach. Suppose that the interregional spill-overs could change the total factor productivity A_{it} and is related to the regional weights w_{ij} and the ratio $\frac{Y_{jt}}{L_{jt}}$. Then we yield the following expression for the total factor productivity:

$$A_{it} = A_{0t} \times \prod_j \left(\frac{Y_{jt}}{L_{jt}}\right)^{\rho_t w_{ij}}, \tag{16}$$

where ρ is interpreted as the intensity of the spatial interactions. Substitute (**??**) for (**??**) and rearrange terms to obtain the equation;

$$\ln\left(\frac{Y_{it}}{L_{it}}\right) = \ln A_{0t} + \rho_t \sum_j w_{ij} \ln\left(\frac{Y_{jt}}{L_{jt}}\right) + a_{1t}\ln\left(\frac{K_{it}}{L_{it}}\right) + a_{2t}\ln(L_{it}), \tag{17}$$

where $\alpha_t = a_{1t}$ and $\gamma_t = a_{1t} + a_{2t}$. Then, the economics of agglomeration coefficient is measured by a_2 (see also Kanemoto *et al.*, 1996) and thus can be estimate by a SAR model. However, as we use a time panel data set, we have to consider the correlation among periods. Therefore, we will use our SAR-SUR model[3].

[3] If we replace the spill-over in (**??**) 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. Furthermore, if we assume that all A_{it}'s are equal across regions, then the model reduces to simple (i.e. non-spatial) SUR model.

Note that an alterntive estimation equation can be obtained by

$$\ln\left(\frac{Y_{it}}{K_{it}}\right) = \ln A_{0t} + \rho_t \sum_j w_{ij} \ln\left(\frac{Y_{jt}}{K_{jt}}\right) + a_{1t} \ln\left(\frac{L_{it}}{K_{it}}\right) + a_{2t} \ln(K_{it}). \quad (18)$$

5.2 Estimation results

Before we describe the empirical results, we explain the data set we use 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 the 'added value of the manufacturing industries', the total capital is the 'amount at hand of permanent assets' and the total employment is the number of full time employees, which exclude part time labor. As a weight matrix W, we use the contiguity matrix of Japanese prefectures as in Kakamu *et al.* (2005), which represents the spatial connection or the 'geography' of economic activities [4] and the average number of dummy variables are 4. For the prior distributions we have specified the following hyper-parameters;

$$\beta_* = 0,\ \Sigma_* = 100 I_{kT},\ \Omega_* = 100 I_T,\ \nu_* = T + 1,$$

Finally, we ran the MCMC algorithm using 10'000 iterations and discarded the first 5000 iterations. All programming results reported here were generated by the Ox programming language, version 4.02 (see Doornik, 2001).

Table **??** shows the marginal likelihoods of all models: SAR-SUR, SEM-SUR, SUR and independent SAR models. Note, that the SAR-SUR model is the best model. Therefore, we will refer only to the parameters of the SAR-SUR model hereafter. We have also added the results of the SUR model, since we want to compare the SAR-SUR models with the simple SUR model.

[4] All except one of the Japanese prefectures (Okinawa) are situated on the four major islands of Japan: Hokkaido, Honshu, Shikoku and Kyushu. All these four islands are connected by train and roads, and eases the fact that islands are separate geographical entities but also connected by ships and ferries. For example, the most northern island Hokkaido is connected by the Seikan railway tunnel to the main island Honshu. And Honshu is connected by 2 large bridge systems, the Awaji and Seto bridge to Shikoku, and the southern island of Kyushu is also connected by the Kanmon tunnel and a bridge to Honshu. Therefore, the island of Okinawa is the only prefecture which is independent of all other prefectures.

Table **??** shows the coefficient estimates of SAR-SUR model. First of all, note that all the coefficients are 'significant' in the Bayesian sense of lying outside the 4σ interval around zero. [5]. The (marginal) posterior distributions of the coefficients, $\ln A_0$, a_1 a_2 and ρ, are shown in Figure **??** by box plots. Note that in general, the marginal distributions of $\ln A_0$ became larger over time. This means that the average TFP is a driving force for the economic growth in Japanese manufacturing industries. On the other hand, the capital-labor intensity a_1 became smaller over time. This could show that the manufacturing industries in Japan became slightly more labor intensive. The economics of agglomeration coefficient, a_2, also became smaller. From this evidence, we conclude that the power of economics of agglomeration is getting smaller and smaller after the collapse of economic bubble in 1991. Therefore the 1990 decade, which is sometimes called the "lost decade", is associated with a declining importance of the era of the economics of agglomeration. Finally, for the spill-over effect, ρ_t, we note that all the parameters are estimated significantly but small and they make a large jump in 2000.

Table **??** shows the variance-covariance and correlation matrices of the empirical example and Figure **??** shows the box plots (of off-diagonal elements), which can be interpreted as implied autocorrelation function. From the correlation matrix, we see that there exist serious time-correlation among equations. It shows that the serial correlation plays an important role in analysing production function in manufacturing industries. This high high serial correlations can also be seen from the box plots in Figure **??**.

Table The entries **??** and **??** show the posterior means, the variance-covariance and correlation matrices of the simple SUR model. Note that the SUR correlation matrix Ω for the simple SUR and the SAR-SUR model are quite similar. Furthermore, we find that many coefficients are of the same size, but the coefficients $\ln A_0$ and a_2 are overestimated if we ignore the possibility of spatial interactions. This shows that if we ignore spatial interaction we might misinterpret the source of economic growth because the $\ln A_0$ and a_2, the TFP and

[5] i.e. significant means that the 95% credible interval does not include zero.

the economics of agglomeration coefficients are overestimated in case assuming a simple SUR model. Finally we note that the spatial ρ_t coefficients, i.e. the spill-over effects play a small but important role on our approach, because they increase the model fit substantially.

Summarizing our estimation results, we are making the following implications: (1) Average TFP seems to be the major source of economic growth in the manufacturing industries of Japan. (2) Manufacturing industries have become more labor intensive. (3) The economics of agglomeration became smaller over time, but (4) The spill-over effects play a small but important role. From the SUR correlation matrix we see that the manufacturing production function in Japan is serially correlated.

6 Conclusions

This paper studied the SUR model with spatial dependency from a Bayesian point of view. We derived the joint posterior distribution, and proposed MCMC methods to estimate the parameters of the model. We have illustrated our approach using Japanese real data and analyzed the economics of agglomeration in Japan during the decade from 1991 to 2000.

Summarizing, we draw the following conclusions: (1) Average TFP is a strong source for economic growth in manufacturing industries in Japan. (2) Manufacturing industries in Japan have become more labor intensive in the 1990s. (3) The economics of agglomeration became smaller over time, but (4) The spill-over effects play a small but important role. These effects do not show any specific trend except for an outlier in the year 2000 which could be an indicator of the existenc3e of a possible slight spatial interrelationship and a potential structural change. This finding needs to be explored in more detail in some future studies. Not surprisingly, we see from the SUR correlation matrix that the manufacturing production function in Japan is strongly serially correlated across the years. A box plot of all theses same' off-diagonal correlation elements of the estimated SUR covariance-matrix is given in Figure 2.

Appendix A

Alternatives for sampling ρ First, we will introduce the independence Metropolis-Hastings proposal step to sample ρ_t for the Gibbs sampler. The following step is used: Sample ρ_t^{new} from

$$\begin{aligned} \rho_t^{new} &\sim \mathcal{N}(\hat{\rho}_t, \hat{\sigma}_t^2), & (19) \\ \hat{\rho}_t &= \{(Wy_t)'(Wy_t)\}^{-1}(Wy_t)'(y_t - X_t\beta_t), & (20) \\ \hat{\sigma}_t^2 &= \omega_t\{(Wy_t)'(Wy_t)\}^{-1}, & (21) \end{aligned}$$

where ω_t is the tt-th element of Ω.

Next, we evaluate the acceptance probability

$$\alpha(\rho_t^{old}, \rho_t^{new}) = min\left(\frac{p(\rho_t^{new})/q(\rho_t^{new})}{p(\rho_t^{old})/q(\rho_t^{old})}, 1\right), \quad (22)$$

where p is the full conditional distribution in (**??**), q is the proposal density given in (**??**) and, of course, ρ_t in E also changes to ρ_t^{new} in $p(\rho_t^{new})$. Finally we set $\rho_t = \rho_t^{new}$ with probability $\alpha(\rho_t^{old}, \rho_t^{new})$, otherwise $\rho_t = \rho_t^{old}$. 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).

Figure **??** shows the autocorrelation function (ACF) of two Metropolis-Hastings proposals for ρ_t in 1991. The estimated parameters are very similar and the decay of the acf's are also quite similar. However, in sampling ρ_t by an independence step, we need to calculate $\hat{\rho}_t$ and $\hat{\sigma}_t^2$ in (**??**) and (**??**) for each t and each iteration. Because of the calculations, it takes a longer time for convergence of the sampler than using a random-walk proposal. Therefore, we used the random-walk proposals in all the MCMC programs.

References

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

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

[3] Case, A.C. (1991) "Spatial patterns in Household Demand," *Econometrica*, **59**, 953–965.

[4] Chib, S. and E. Greenberg (1995) "Hierarchical Analysis of SUR Models with Extensions to Correlated Serial Errors and Time-Varying Parameter Models," *Journal of Econometrics*, **68**, 339–360.

[5] Chib, S. and E. Greenberg (1998) "Analysis of Multivariate Probit Models," *Biometrika*, **88**, 347–361.

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

[7] 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.

[8] Kakamu, K., W. Polasek and H. Wago (2005) "Spatial Interaction of Crime Incidents in Japan," *Proceedings of International Congress on Modelling and Simulation*, 407–413.

[9] Kanemoto, Y., T. Ohkawara and T. Suzuki (1996) "Agglomeration Economies and a Test for Optimal City Size in Japan," *Journal of the Japanese and International Economies*, **10**, 379-398.

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

[11] Percy, D.F. (1992) "Prediction for Seemingly Unrelated Regression," *Journal of the Royal Statistical Society Series B*, **54**, 243–252

[12] Sun, D., R.K. Tsutakawa and P.L. Speckman (1999) "Posterior Distribution of Hierarchical Models using CAR(1) Distributions," *Biometrika*, **86**, 341–350.

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

[14] 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.

[15] Zellner, A. (1971) *An Introduction to Bayesian Inference in Econometrics*, John Wiley , New York.

Figure 1: Box plots of $\ln A_0$, a_1, a_2 and ρ

Figure 2: Box plot autocorrelation function

Figure 3: Autocorrelation function for random-walk and independence chains