空間計量経済学における空間重み行列の自動選択：空間ヘドニック分析への適用

瀬谷　創[1]・堤　盛人[2]・山形　与志樹[3]

[1]学生会員　国立環境研究所　地球環境研究センター（〒305-8506 つくば市小野川16-2）
E-mail:seya.hajime@nies.go.jp

[2]正会員　筑波大学准教授　システム情報系（〒305-8573 つくば市天王台1-1-1）
E-mail: tsutsumi@sk.tsukuba.ac.jp

[3]非会員　国立環境研究所　地球環境研究センター（〒305-8506 つくば市小野川16-2）
E-mail: yamagata@nies.go.jp

便益計測の代表的手法の一つである資産価値法では，資産としての不動産の価格をヘドニック・アプローチによって推定する方法が採られることが多い．言うまでもなく不動産データは，位置座標を持った空間データであり，位置に起因する空間的自己相関が存在する．近年，空間計量経済学と呼ばれる分野において，このような空間データの特性を考慮した『空間ヘドニック・アプローチ』を用いて分析を行う研究事例が蓄積されてきた．空間計量経済学のモデルは，データ間の自己相関関係を空間重み行列によって特定化する点に特徴があるが，その選択は多くの場合主観的であり，結果がその選択に依存するという点に本質的な課題がある．本研究では，リバーシブルジャンプMCMC法とsimulated annealing（焼きなまし法）を用いて，空間計量経済モデルにおいて重み行列と説明変数の自動選択を行う方法を考案し，ヘドニックモデルへの適用を行った．

Key Words : *Spatial econometrics, spatial weight matrix, model selection, reversible jump MCMC, simulated annealing*

1.　Introduction

The application of a hedonic approach to real estate data plays an important role in real estate market analysis. The recent progress of geographic information systems (GIS) has provided access to detailed attributes of real estate properties, including traditional transportation accessibility, land use control information, and considerable environmental information; hence, numerous potential variables are now available to a study. However, considering numerous explanatory variables in a regression model may lead to the problem of multicollinearity; therefore, in practice, model selection procedures using the test statistics (e.g., *t*-value), information criteria (e.g., Akaike's Information Criterion (AIC)), or Bayesian posterior model probabilities are routinely employed (e.g., Kitagawa, 1997; Leeb and Pötscher, 2005; Claeskens and Hjort, 2008; Magnus et al., 2010).

Another important aspect of real estate is spatial autocorrelation. Because real estate values and attributes are geographically distributed, such *spatial* characteristics of data must explicitly be modeled to obtain reliable estimates (e.g., Pace et al., 1998). One of the major research fields that considers the spatial autocorrelation is spatial econometrics (Anselin, 1988; LeSage and Pace, 2009). Applying spatial econometric models to the hedonic approach is often termed the "spatial hedonic approach" and has become popular in regional science and GIScience. Can (1990) is the first researcher to have introduced the spatial econometric technique to the hedonic approach; this work has been followed by numerous works on spatial hedonic approaches (Can, 1992; Brasington and Hite, 2005; Tsutsumi and Seya, 2008, 2009; Anselin and Lozano-Gracia, 2009; Koschinsky et al., 2011). The representative spatial econometric models are the spatial lag model (SLM) and the spatial error model (SEM) (e.g., Anselin and Bera, 1998). It is noteworthy that the former is occasionally referred to as the spatial autoregressive model (e.g., LeSage and Pace, 2009) and the latter is known as the simultaneous autoregressive model in the spatial statistics literature (see Arbia, 2006 for a discussion of the model name).

It might be true that spatial econometrics has achieved remarkable success (Anselin, 2010); practical difficulties in applying spatial econometric models include the specification of the spatial weight matrix (SWM), which affects the final analysis results. Some simulation studies have suggested that information criteria such as AIC are useful

when selecting weight matrices (e.g., Stakhovych and Bijmolt, 2008), but if many model candidates exist (e.g., when the selections of the explanatory variables are performed simultaneously), then the computational burden of calculating such criteria for each model is large. For example, suppose that the numbers of possible explanatory variables and SWMs are denoted as K and L, respectively; thus, the number of model candidates S is $S=2^K \times L$. Clearly when K assumes a large value, it is impossible to compute the AIC for each model within a reasonable computational time (e.g., $K=L=20$ yields $T=20,971,520$).

The present study develops an automatic model selection algorithm using the technique of reversible jump Markov Chain Monte Carlo (RJMCMC) combined with simulated annealing, which is the natural extension of the trans-dimensional simulated annealing (TDSA) algorithm by Brooks et al. (2003) to the spatial econometric model. The performance of our algorithm is examined with the well-known Boston housing dataset.

This paper comprises the following sections. Section 2 reviews the spatial econometric techniques, specifically focusing of the SWM specification, and Section 3 introduces the TDSA of Brooks et al. (2003). Section 4 applies the TDSA to the SLM, and Section 5 examines the performance of our algorithm. Finally, Section 6 concludes this paper.

2. Spatial econometric techniques

(1) spatial lag model

The present paper focuses on the SLM, which is expressed as

$$y = \rho W y + X\beta + \varepsilon \, , \ \varepsilon \sim N(0, \sigma^2 I) \, , \qquad (1)$$

where y is the $N \times 1$ vector of dependent variables, X is the $N \times K$ matrix of exogenous explanatory variables, β is the $K \times 1$ corresponding parameter vector, ε is the disturbance term, 0 denotes the $N \times 1$ vector whose element is given by 0, σ^2 is the variance of error, I is the $N \times N$ identity matrix, W is the $N \times N$ SWM, and ρ is the corresponding scalar parameter. The parameters of this model are usually estimated by the maximum likelihood method, the generalized method of moments, or the Bayesian MCMC; for more details, see Lee (2007), LeSage and Pace (2009), and Elhorst (2010).

(2) spatial weight matrix specification

For logical and identification purposes, the following assumptions are typically made for SWMs. First, W should be exogenous and invariant over time for the purpose of identification (Manski, 1993; Rincke, 2010) or parameter interpretation (LeSage and Pace, 2011). Kostov (2010) discussed that one reason for the popularity of the SWMs based on geographical distances is that their exogeneity is automatically ensured. Second, the diagonal elements of W are usually set to zero in order to prevent from predicting itself. Third, W is usually normalized so that the rows sum to unity (e.g., Fingleton, 2009).

Although the specification of W is a crucial step in spatial econometric models (e.g., Smith, 2009; Stakhovych and Bijmolt, 2009; LeSage and Pace, 2010), Anselin (2002) suggested that there is little formal guidance when choosing the correct spatial weights for the given application. Stakhovych and Bijmolt (2009) divided the literature on SWMs into three categories: (i) treating weights matrices as completely exogenous constructs, (ii) letting the data determine them, and (iii) estimating them.

The first approach is using geographical relations of observations or spatial units, such as k nearest neighbors or inverse distance. Here, distance is not limited to Euclidean distance, and many general distance measures can be used. LeSage and Polasek (2008) employed a transportation network to provide distance. The second approach is to use data to define SWMs. Getis and Aldstadt (2004) employed G statistics (Getis and Ord, 1992; Ord and Getis, 1995) to construct SWMs. Aldstadt and Getis (2006) employed the AMOEBA procedure, and Mur and Paelinck (2011) employed a p-median approach. Relatively few studies took the third approach. Bhattacharjee and Jensen-Butler (2005) estimated the SWM non-parametrically, while Fernández-Vázquez et al. (2009) used the maximum entropy approach. The other interesting approaches are summarized in Harris et al. (2011). It is also important to note that some alternative methods that do not use W have evolved, such as spatial filtering (Griffith, 2003; Tiefelsdorf and Griffith, 2007), the semi-parametric approach (Robinson, 2008), the spatial-HAC estimator (Kelejian and Prucha, 2007), and structural equation modeling (Folmer and Oud, 2008).

The focus of this paper is the selection of the correct SWM from the many candidates. This selection step is required for categories (i) and (ii). Indeed, Bayesian techniques are useful for both variables and SWMs selection (Hepple, 1995a, b; LeSage and Fischer, 2008; Crespo, Cuaresma, and Feldkircher, 2010; Cotteleer et al., 2011), but they require the specification of the prior distributions (such as Zellner's g-prior) of the model parameters, and the model selection results are not always robust to the prior specification (see Seya and Tsutsumi, 2012). Stakhovych and Bijmolt (2008) suggested that applying a SWM selection procedure based on information criteria increases the probability of identifying its true specification. However, as stated above, if many model candidates exist, the computational burden of calculating such criteria for each model may be large.

3. Trans-Dimensional simulated annealing

(1) Fixed dimensional simulated annealing

Following Brooks et al. (2003), this paper first introduces the tradi-

tional fixed dimensional SA (FDSA), which does not allow for movement across the competing models. Given the objective function $f(\theta)$ that we wish to minimize over the p-dimensional vector of θ, the corresponding Boltzmann distribution admits a density $b_T(\theta) \propto \exp\{-f(\theta)/T\}$. The FDSA algorithm is summarized as follows (Andrieu et al., 2003; Brooks et al., 2003):

[1] Assume the current state T and θ.

[2] Propose $\tilde{\theta} = g(\theta, v)$, where v is drawn from a proposal $q(v)$ and g denotes any function such that $g\{g(\theta, v), \tilde{v}\} = \theta$.

[3] Accept the move to $\tilde{\theta}$ with probability

$$\alpha(\theta, \tilde{\theta}) = \min\left\{1, \frac{b_T(\tilde{\theta})q(\tilde{v})}{b_T(\theta)q(v)}\right\}.$$

[4] Repeat steps 2 and 3 for a specified number of iterations R.

[5] Lower the temperature T according to some predetermined schedule to "freeze" the chain at a point of minimum mass, and repeat steps 2–4 until some stopping criterion is met.

The final parameter θ will approximate the minimum of $f(\theta)$. SA can avoid the local minima by allowing the move not only to the higher probability state but also to the lower probability state with probability $\alpha(\theta, \tilde{\theta})$. FDSA is one of the most-used algorithms to minimize an objective function, which is adopted by the general optimization function `optim` of the R statistical language as an option. For more details regarding the FDSA, see Brooks and Morgan (1995) and Pham and Karaboga (2000).

(2) Trans-dimensional simulated annealing

The TDSA algorithm extends the FDSA algorithm to simultaneously consider the problem of model choice and parameter estimation. When we move between the competing models to identify the model that minimizes the objective function, we need to match the dimensions of the models for a meaningful comparison. Green's (1995) RJMCMC does so with the Jacobian. RJMCMC is a generalization of the Metropolis-Hastings (MH) algorithm (Hastings, 1970), to which a move across parameter spaces of different dimensionalities was introduced, while retaining a detailed balance. RJMCMC is a general algorithm and includes the MCMC model composition (MC3) approach, which is used in the Bayesian model averaging literature (Raftery et al., 1997; Hoeting et al., 1999), as is its special cases (Clyde, 1999). Johnson and Hoeting (2011) applied RJMCMC to the selection of a geostatistical model (see Cressie and Wikle, 2011). LeSage and Parent (2007) employed the MC3 approach for

averaging the SLM and SEM estimates. LeSage and Fischer (2008) and Cotteleer et al. (2011) extended the LeSage and Parent's (2007) algorithm by appending the SWMs selection step. Brooks et al. (2003) combined RJMCMC with SA, which they termed TDSA. This subsection introduces their proposed TDSA algorithm.

Let the observed data y be generated by a model $M_1, M_2, \ldots$ indexed by $s \in \mathcal{M}$. Corresponding to model M_s, there is the parameter vector $\theta_s \in \mathfrak{R}^{p_s}$ with corresponding likelihood function $L_s(y|\theta_s)$. Then, suppose that we wish to minimize of some general function $f(\theta_s, M_s)$ that is defined over both the parameter and model spaces. If we take

$$f(\theta_s, M_s) = -2\ln\{L_s(y \mid \theta_s)\} + 2p_s, \qquad (2)$$

then we obtain AIC. The associated Boltzmann distribution can be defined as

$$b_T(\theta_s, M_s) \propto \exp\{-f(\theta_s, M_s)/T\}. \qquad (3)$$

Brooks et al. (2003) noted that at temperature $T=1$, the Boltzmann distribution corresponds to a Bayesian posterior with a flat prior over the parameter space but to a prior for model M_s proportional to $\exp(-p_s/T)$.

Next, suppose that we wish to move (θ_s, M_s) to $(\theta_{\tilde{s}}, M_{\tilde{s}})$, where typically $p_s \neq p_{\tilde{s}}$. The TDSA algorithm can be summarized as follows:

[1] Assume the current state T and (θ_s, M_s).

[2] Propose $(\theta_{\tilde{s}}, M_{\tilde{s}}) = g_{s\tilde{s}}(\theta_s, v)$, where v is drawn from a proposal $q_{s\tilde{s}}(v)$ and where g denotes a function that satisfies $g_{\tilde{s}s}\{g_{s\tilde{s}}(\theta_s, v), \tilde{v}\} = \theta_s$. In other words, $p_s + \dim(v) = p_{\tilde{s}} + \dim(\tilde{v})$.

[3] Accept the move to $(\theta_{\tilde{s}}, M_{\tilde{s}})$ with probability

$$\alpha((\theta_s, M_s), (\theta_{\tilde{s}}, M_{\tilde{s}})) =$$
$$\min\left\{1, \frac{b_T(\theta_{\tilde{s}}, M_{\tilde{s}})j_{\tilde{s}s}q_{\tilde{s}s}(\tilde{v})}{b_T(\theta_s, M_s)j_{s\tilde{s}}q_{s\tilde{s}}(v)}\left|\frac{\partial\{g_{s\tilde{s}}(\theta_s, v), \tilde{v}\}}{\partial(\theta_s, v)}\right|\right\},$$

where $j_{s\tilde{s}}$ denotes the probability of proposing the new model $\tilde{s}$.

[4] Repeat steps 2 and 3 for a specified number of iterations R.

[5] Lower the temperature T according to some predetermined

schedule, and repeat steps 2–4 until some stopping criterion is met. Godsill (2001) proposed another acceptance probability, defined as

$$\alpha((\boldsymbol{\theta}_s, M_s), (\boldsymbol{\theta}_{\tilde{s}}, M_{\tilde{s}})) =$$

$$\min\left\{1, \frac{b_T(\boldsymbol{\theta}_{\tilde{s}}, M_{\tilde{s}}) j_{\tilde{s}s} q_{\tilde{s}s}(\boldsymbol{\theta}_s, M_s)}{b_T(\boldsymbol{\theta}_s, M_s) j_{s\tilde{s}} q_{s\tilde{s}}(\boldsymbol{\theta}_{\tilde{s}}, M_{\tilde{s}})}\right\}. \tag{4}$$

This acceptance probability explicitly includes the probability of proposing the move from s to $\tilde{s}$ and that the proposal is made directly in the new parameter space $\boldsymbol{\theta}_{\tilde{s}}$ rather than via the dimension-matching random variables $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$. In other words, this transformation corresponds to $\boldsymbol{\theta}_{\tilde{s}} = \boldsymbol{v}$ and $\boldsymbol{\theta}_s = \tilde{\boldsymbol{v}}$, which has a unity Jacobian. Thus, Godsill's (2001) formulation can avoid the need for a Jacobian term.

Brooks et al. (2003) suggested that the only difference between FDSA and TDSA is that in the latter, we minimize over both parameters and models so that, at any temperature T, the distribution $b_T(\boldsymbol{\theta}_s, M_s)$ is explored via a combination of MCMC updates for within-model moves and RJMCMC updates for moves across models.

4.   Trans-dimensional simulated annealing for the spatial econometric model

(1)   Trans-dimensional simulated annealing applied to the linear regression model

This subsection applies the TDSA algorithm to the linear regression model, which we term the basic model (BM). The BM under the model M_s is given by

$$\boldsymbol{y} = X_s \boldsymbol{\beta}_s + \boldsymbol{\varepsilon}_s, \ \boldsymbol{\varepsilon}_s \sim N(\boldsymbol{0}, \sigma_s^2 \boldsymbol{I}). \tag{5}$$

Godsill (2001) proposed partial analytic RJMCMC (PARJ) in which some of the parameters are shared among the models. We assume that error variance is common to all the models, that is, $\sigma_s^2 = \sigma^2$, whereas $\boldsymbol{\beta}_s$ differs for each model. Given the data $\boldsymbol{y}$, the likelihood of the BM under M_s is given by

$$L_s(\boldsymbol{y} \,|\, \sigma^2, \boldsymbol{\beta}_s) = (2\pi\sigma^2)^{-N/2} \exp(-\boldsymbol{\varepsilon}_s' \boldsymbol{\varepsilon}_s / 2\sigma^2), \tag{6}$$

where $\boldsymbol{\varepsilon}_s' \boldsymbol{\varepsilon}_s = (\boldsymbol{y} - X_s \boldsymbol{\beta}_s)'(\boldsymbol{y} - X_s \boldsymbol{\beta}_s)$. The associated Boltzmann distribution can be defined as

$$b_T(\boldsymbol{\theta}_s, M_s) = L_s(\boldsymbol{y} \,|\, \sigma^2, \boldsymbol{\beta}_s)^{1/T} \exp(-p_s / T)$$

$$\propto (\sigma^2)^{-N/2T} \exp(-\boldsymbol{\varepsilon}_s' \boldsymbol{\varepsilon}_s / 2T\sigma^2) \exp(-p_s / T)$$

$$\propto (\sigma^2)^{-(N-K)/2T} \exp\{-(N-K)s_s^2 / 2T\sigma^2\}$$

$$\cdot (\sigma^2)^{-K/2T} \exp[-\{(\boldsymbol{\beta}_s - \hat{\boldsymbol{\beta}}_s)' X_s' X_s (\boldsymbol{\beta}_s - \hat{\boldsymbol{\beta}}_s)\}/2T\sigma^2]$$

$$\cdot \exp(-p_s / T), \tag{7}$$

where $\hat{\boldsymbol{\beta}}_s = (X_s' X_s)^{-1} X_s \boldsymbol{y}$ and $s_s^2 = (N-K)^{-1}(\boldsymbol{y} - X_s \hat{\boldsymbol{\beta}}_s)(\boldsymbol{y} - X_s \hat{\boldsymbol{\beta}}_s)$.

a)   Within model move

From eq. (7), we can easily obtain the conditional distribution of the error variance given the regression coefficients given by

$$b_T(\sigma^2 \,|\, \boldsymbol{\beta}_s, M_s)$$

$$\propto (\sigma^2)^{-(N-K)/2T} \exp\{-(N-K)s_s^2 / 2T\sigma^2\}$$

$$\propto InverseGamma\left(\frac{N-K}{2T} + 1, \frac{(N-K)s_s^2}{2T}\right). \tag{8}$$

Similarly, the conditional distribution for the regression coefficients given the error variance is given by

$$b_T(\boldsymbol{\beta}_s \,|\, \sigma^2, M_s)$$

$$\propto (\sigma^2)^{-K/2T} \exp[-\{(\boldsymbol{\beta}_s - \hat{\boldsymbol{\beta}}_s)' X_s' X_s (\boldsymbol{\beta}_s - \hat{\boldsymbol{\beta}}_s)\}/2T\sigma^2]$$

$$\propto N\left(\frac{\Sigma X_s' \boldsymbol{y}}{\sigma^2 T}, \Sigma\right), \tag{9}$$

where $\Sigma = \sigma^2 T (X_s' X_s)^{-1}$. The within move proceeds by generating $v_1 \sim b_T(\sigma^2 \,|\, \boldsymbol{\beta}_s, M_s)$ and $v_2 \sim b_T(\boldsymbol{\beta}_s \,|\, \sigma^2, M_s)$ and by setting $(\tilde{\sigma}^2, \tilde{\boldsymbol{\beta}}_s) = g(\sigma^2, \boldsymbol{\beta}_s, v_1, v_2) = (v_1, v_2)$ in step 2 of the annealing algorithm. We can skip step 3 because we are using Gibbs updates, and in this case the corresponding acceptance ratio in step 2 is 1 (see Brooks et al., 2003).

b)   Across (between) model move

We apply the previously explained Godsill's (2001) acceptance ratio to avoid the calculation of Jacobian term for dimension matching. Because we have assumed that the error variance is shared among the models, we can take the $b_T(\boldsymbol{\beta}_s \,|\, \sigma^2, M_s)$ for the proposal q. Thus, we can avoid the trial-and-error tuning of the proposal distribution. We also use a discrete random walk for $j_{\tilde{s}s}$ (Johnson and Hoeting, 2011). First, one of the explanatory variables is chosen

with probability $1/K$; if it is already in M_s, it is removed from $M_{\tilde{s}}$, and if it is not in M_s, it is added. Thus, $j_{\tilde{s}s}/j_{s\tilde{s}}=1$.

(2) Trans-dimensional simulated annealing applied to the spatial lag model

This subsection applies the TDSA algorithm to the SLM. The SLM under the model M_s is given by

$$y = \rho_s W_s y + X_s \beta_s + \varepsilon_s , \ \varepsilon_s \sim N(0, \sigma_s^2 I) . \quad (10)$$

We assume that the error variance and the spatial parameter are common to all the models, that is, $\rho_s = \rho$ and $\sigma_s^2 = \sigma^2$, whereas SWM and β_s differ for each model. This assumption is similar to the setting of Johnson and Hoeting (2011), who assumed that geostatistical parameters (nugget, range, partial-sill, and anisotropic parameters) are shared between the models but that regression coefficients differ. Given the data y, the likelihood of the SLM under M_s is given by

$$L_s(y \mid \rho, \sigma^2, \beta_s) =$$

$$(2\pi\sigma^2)^{-N/2} \mid I - \rho W_s \mid \exp(-\varepsilon_s' \varepsilon_s / 2\sigma^2) , \quad (11)$$

where $\varepsilon_s' \varepsilon_s = (y - \rho W_s y - X_s \beta_s)'(y - \rho W_s y - X_s \beta_s)$
$= (A_s y - X_s \beta_s)'(A_s y - X_s \beta_s)$, and $A_s = I - \rho W_s$. The associated Boltzmann distribution can be defined as

$$b_T(\theta_s, M_s) \propto L_s(y \mid, \sigma^2, \beta_s)^{1/T} \exp(-p_s/T)$$

$$\propto (\sigma^2)^{-N/2T} \mid I - \rho W_s \mid^{1/T} \exp(-\varepsilon_s' \varepsilon_s / 2T\sigma^2) \exp(-p_s/T)$$

$$\propto (\sigma^2)^{-(N-K)/2T} \exp\{-(N-K)s_s^2/2T\sigma^2\}$$

$$\cdot (\sigma^2)^{-K/2T} \exp[-\{(\beta_s - \hat{\beta}_s)' X_s' X_s(\beta_s - \hat{\beta}_s)\}/2T\sigma^2]$$

$$\cdot \exp(-p_s/T) \cdot \mid I - \rho W_s \mid^{1/T} , \quad (12)$$

where $\hat{\beta}_s = (X_s' X_s)^{-1} X_s A_s y$, $s_s^2 = (N-K)^{-1}(A_s y - X_s \hat{\beta}_s)(A_s y - X_s \hat{\beta}_s)$.

a) Within model move

Eq. (12) shows that we can easily obtain the conditional distribution of the error variance given the regression coefficients and the spatial parameter, given by

$$b_T(\sigma^2 \mid \beta_s, \rho, M_s)$$

$$\propto (\sigma^2)^{-(N-K)/2T} \exp\{-(N-K)s_s^2/2T\sigma^2\}$$

$$\propto InverseGamma\left(\frac{N-K}{2T}+1, \frac{(N-K)s_s^2}{2T}\right) . \quad (13)$$

Similarly, the full conditional distribution for the regression coefficients given the error variance and the spatial parameter is given by

$$b_T(\beta_s \mid \sigma^2, \rho, M_s)$$

$$\propto (\sigma^2)^{-K/2T} \exp[-\{(\beta_s - \hat{\beta}_s)' X_s' X_s(\beta_s - \hat{\beta}_s)\}/2T\sigma^2]$$

$$\propto N\left(\frac{\Sigma X_s' A_s y}{\sigma^2 T}, \Sigma\right), \quad (14)$$

where $\Sigma = \sigma^2 T(X_s' X_s)^{-1}$. The spatial parameter is sampled using a random walk MH algorithm (Kakamu, 2009; Seya et al., 2012).

$$\text{Sample } \tilde{\rho} = \rho + cz , \ z \sim N(0,1) ,$$

where c is called the tuning parameter. The tuning parameter c was incremented or decremented when the acceptance rate moved below 0.30 or above 0.50, leading to an acceptance rate of approximately 0.40 after a burn-in period. The acceptance probability is calculated in the following manner:

$$\alpha(\rho, \tilde{\rho}) = \min\left\{1, \frac{\mid I - \tilde{\rho} W_s \mid^{1/T} \exp(\varepsilon_{\tilde{s}}' \varepsilon_{\tilde{s}} / 2\sigma^2)}{\mid I - \rho W_s \mid^{1/T} \exp(\varepsilon_s' \varepsilon_s / 2\sigma^2)}\right\} . \quad (15)$$

The within move proceeds by generating $v_1 \sim b_T(\sigma^2 \mid \beta_s, \rho, M_s)$, $v_2 \sim b_T(\beta_s \mid \sigma^2, \rho, M_s)$, and $v_3 \sim b_T(\rho \mid \beta_s, \sigma^2, M_s)$ and then setting $(\tilde{\sigma}^2, \tilde{\beta}_s, \tilde{\rho}) = g(\sigma^2, \beta_s, \rho, v_1, v_2, v_3) = (v_1, v_2, v_3)$ in step 2 of the annealing algorithm.

b) Across (between) model move

Because we have assumed that the error variance and the spatial parameter are shared among the models to avoid a long RJMCMC run, our across-model move is composed of the move across the explanatory variables and SWMs. The former is the same as in the BM case. That is, we randomly select one of the explanatory variables with probability $1/K$; if it is already in M_s, it is removed in $M_{\tilde{s}}$, and if it is not in M_s, it is added. We do not consider the drop and

Table 1. Variable description (Kostov, 2010)

Variable	Description
MEDV	Median values of owner-occupier housing in thousands of US dollars
LON	Tract point longitude in decimal degrees
LAT	Tract point latitude in decimal degrees
CRIM	Per capita crime
ZN	Proportion of residential land zoned for lots over $25\,000\,\text{ft}^2$ per town
INDUS	Proportion of nonretail business acres per town
CHAS	An indicator: 1 if tract borders Charles River; 0 otherwise
NOX	Nitric oxides concentration (parts per 10 million) per town
RM	Average number of rooms per dwelling
AGE	Proportion of owner-occupied units built prior to 1940
DIS	Weighted distance to five Boston employment centres
RAD	Index of accessibility to radial highways per town
TAX	Property-tax rate per US $10\,000$ per town
PTRATIO	Pupil – teacher ratio per town
B	Calculated as $1000(N_{\text{Black}}-0.63)^2$ where N_{Black} is the proportion of Blacks
LSTAT	Percentage of lower status population

add move in this study. We assume the $b_T(\boldsymbol{\beta}_s \mid \sigma^2, \rho, M_s)$ for the proposal q.

For the latter move, the SWM candidates must be logically predetermined. We adopt a popular k-nearest-neighbors criterion. That is, we prepared 30 weighting matrices (from $k=1$ to 30) and randomly choose one SWM from the candidate. Then the acceptance probability is calculated in the following manner:

$$\alpha(\boldsymbol{W}_s, \boldsymbol{W}_{\tilde{s}}) = \min\left\{1, \frac{\mid \boldsymbol{I} - \rho \boldsymbol{W}_{\tilde{s}} \mid^{1/T} \exp(\boldsymbol{\varepsilon}_{\tilde{s}}' \boldsymbol{\varepsilon}_{\tilde{s}} / 2\sigma^2)}{\mid \boldsymbol{I} - \rho \boldsymbol{W}_s \mid^{1/T} \exp(\boldsymbol{\varepsilon}_s' \boldsymbol{\varepsilon}_s / 2\sigma^2)}\right\}.$$

5. Illustration of the proposed algorithm

(1) Data

To examine the performance of our proposed algorithm, we use the well-known Boston housing dataset. This dataset was originally provided by Harrison and Rubinfield (1978), and Gilley and Pace (1996) augmented the dataset with longitude-latitude of the observations. Pace and Gilley (1997) suggested that these data exhibit various problems common to many hedonic pricing or mass appraisal models. For example, not all variables exhibit the proper sign, that is, the AGE variable (see Table 1) is insignificant and positive, and a high positive spatial autocorrelation exists among the observations. Pace and Gilley (1997) constructed two hedonic pricing models based on BM and SEM with this dataset and found that SEM would successfully yield the significantly negative estimate of the AGE variable. This curious result may be caused by the impacts of spatially

autocorrelated omitted variables on the included variables that were successfully incorporated into the model by using the SEM. Kostov (2010) indicated that this dataset is one of the most popular datasets, and it has stimulated a whole industry that has used this and other datasets to examine and compare alternative statistical methods. Indeed, many studies have employed this dataset (e.g., Pace and Gilley, 1997; Kostov, 2010; LeSage and Pace, 2004, 2010).

We adopted the same variables to Kostov (2010), as indicated in Table 1. Following Kostov (2010), the natural logarithms of MEDV, DIS, RAD, and LSTAT are taken, while the squares of NOX and RM are taken to capture some of the underlying nonlinearities, resulting in ln(MEDV), ln(DIS), ln(RAD), ln(LSTAT), NOX^2, and RM^2. With regard to the descriptive statistics of the data, see Kostov (2010). The numbers of observations and the explanatory variables (including the intercept) are 506 and 14, respectively. Because we consider the 30 different SWMs, the number of possible model candidates is $2^{13} \times 30 = 245760$, where the intercept is always assumed to be in the model.

(2) Simulation results

Thus far, various cooling schedules for the annealing algorithms have been proposed. Brooks et al. (2003) suggested that when we set $T = 10\lambda^t$, acceptable performances seem to be achieved for $\lambda \geq 0.95$ and for $R \geq 500$. Hence, following Brooks et al. (2003), we adopt this cooling schedule and set $\lambda = 0.95$. We take the initial temperature of $T_0 = 10$, which is reduced every 1000 iterations (i.e., $R = 1000$). We stopped the annealing procedure when sufficient cooling was attained ($T \approx 0.02$), which corresponds to the $t = 122$ temperature reduction

Table 2. Parameter estimates of the BM and SLM

	Coef.	t	Coef.	z
INTERCEPT	4.558	29.5	2.186	12.1
CRIM	–0.01186	–9.53	–0.007841	–7.77
ZN	0.00008016	0.159	0.0004855	1.20
INDUS	0.0002395	0.101	0.0009184	0.484
CHAS	0.0914	2.75	0.009495	0.353
NOX	–0.638	–5.64	–0.2844	–3.07
RM	0.006328	4.82	0.007062	6.68
AGE	0.00009074	0.172	–0.00006746	–0.160
DIS	–0.1913	–5.73	–0.1605	–5.97
RAD	0.09571	5.00	0.07873	5.12
TAX	–0.0004203	–3.43	–0.0003609	–3.66
PTRATIO	–0.03112	–6.21	–0.009155	–2.17
B	0.0003637	3.53	0.0002742	3.28
LSTAT	–0.3712	–14.8	–0.2494	–11.9
Spatial parameter ρ			0.5074	LR(187.5)
Adjusted R^2	0.800			
Log likelihood			243.7	
AIC	–269.91(ML)		–455.5	
Variance of error	0.0333		0.0214	

AIC for the BM is calculated using the ML estimates.

Table 3. Parameter estimates of the model specified by the TDSA algorithm

(In the column labeled as "Model," 1 denotes included and 0 denotes not included in the model).

	Model	Coef.
INTERCEPT	1	2.277
CRIM	1	–0.007979
ZN	0	
INDUS	0	
CHAS	0	
NOX	1	–0.2871
RM	1	0.006779
AGE	0	
DIS	1	–0.1583
RAD	1	0.07898
TAX	1	–0.0003315
PTRATIO	1	–0.01128
B	1	0.0002737
LSTAT	1	–0.2588
Spatial parameter	0.5006	
Log likelihood	242.8	
AIC	–461.7	
Residual variance	0.0226	

(Fig. 1). Hence, the number of whole iterations is 122000. The program is coded with the R statistical language.

Table 2 provides the OLS estimates of the BM and ML estimates of the SLM in case that all of the explanatory variables are introduced. Here, the number of nearest neighbors k for constructing W is set to six because it minimized the log-likelihood (or maximized the AIC in

the same sense because the numbers of adopted variables are fixed) (fig. 2). Note that the estimates of the AGE variable are positive for the BM but negative for the SLM (although they are not statistically significant). Also, the result of the LR test suggests that a high positive spatial lag dependence among the observations may exist.

The TDSA outputs are shown in figs. 3 through 6. The acceptance rate for spatial parameter ρ is 0.456, and the computation time is 13.08 hours. Because we directly coded the method with an interpreted language (i.e., R) and not a compiled language (e.g., C++), a relatively large computation time was required.

Fig. 3 (a) provides a trace plot of the number of nearest neighbors k for W each time that the temperature is reduced (every 1000th iteration). It is found that the chain converges rapidly to the best model (i.e., k=6) as the temperature is decreased. Figs. 4 and 5 provide the corresponding plots of error variance and spatial parameter.

Figs. 3 (b) and 6 provide the trace plot of the number of nearest neighbors k and the regression coefficients of CRIM and ZN variables in the first and the last 1000 iterations. Note that these variables frequently move from one to zero when T is high (i.e., T=10), while they are frozen at one or zero when T is low (i.e., $T \approx 0.02$). Similar movement was obtained for the other variables. Table 3 presents the estimates of the model specified by the TDSA. We find that the variables whose coefficients are not statistically significant at a 5% level in table 2 are not selected, and therefore the AIC value is improved as compared to full model specification.
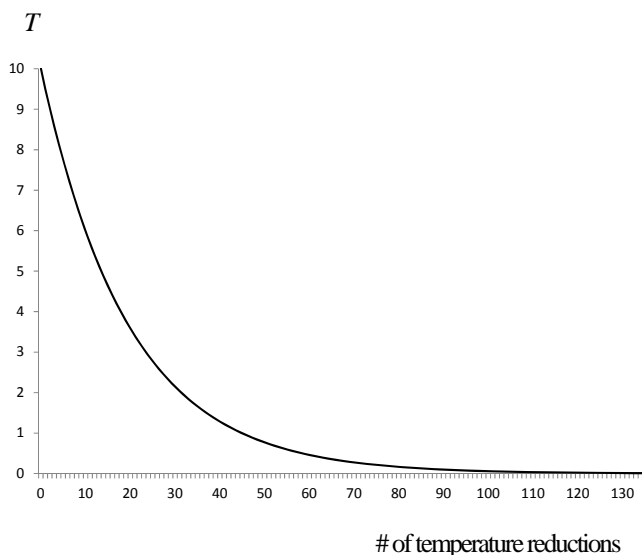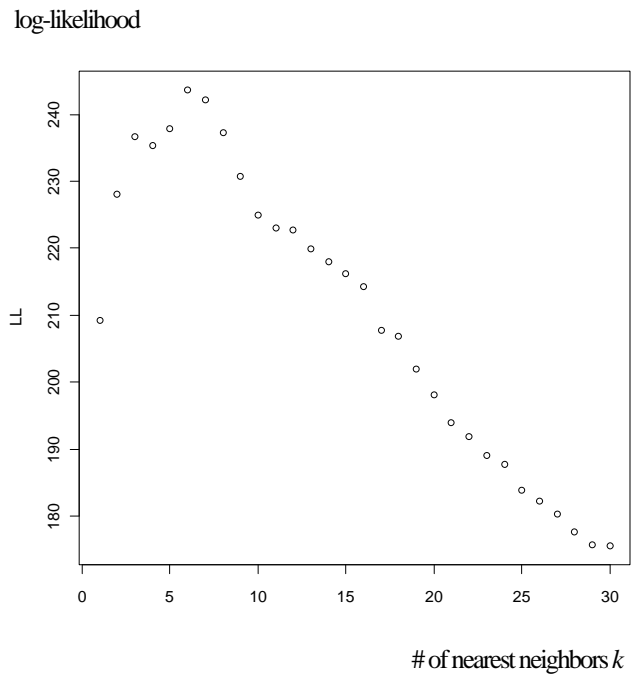
log-likelihood

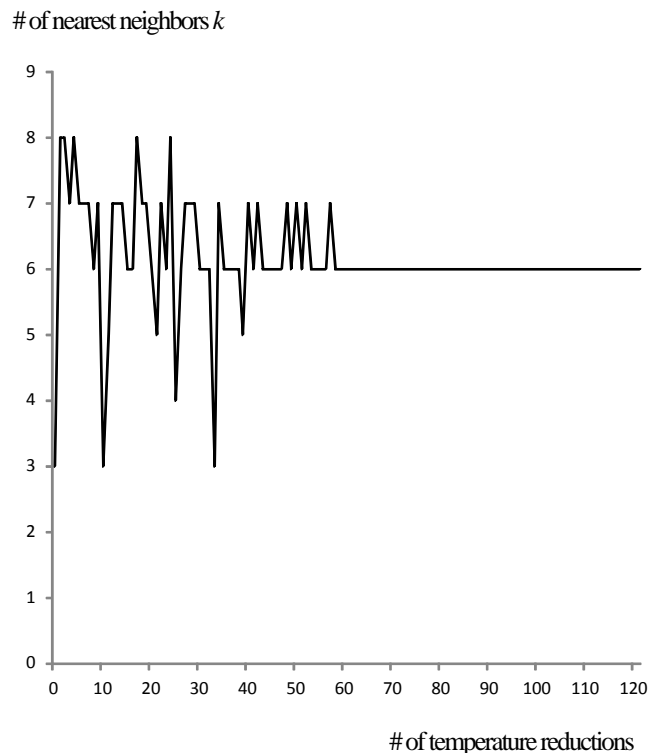**Figure 2.** Log-likelihood (LL) value versus the number of nearest neighbors

# of nearest neighbors k

# of temperature reductions

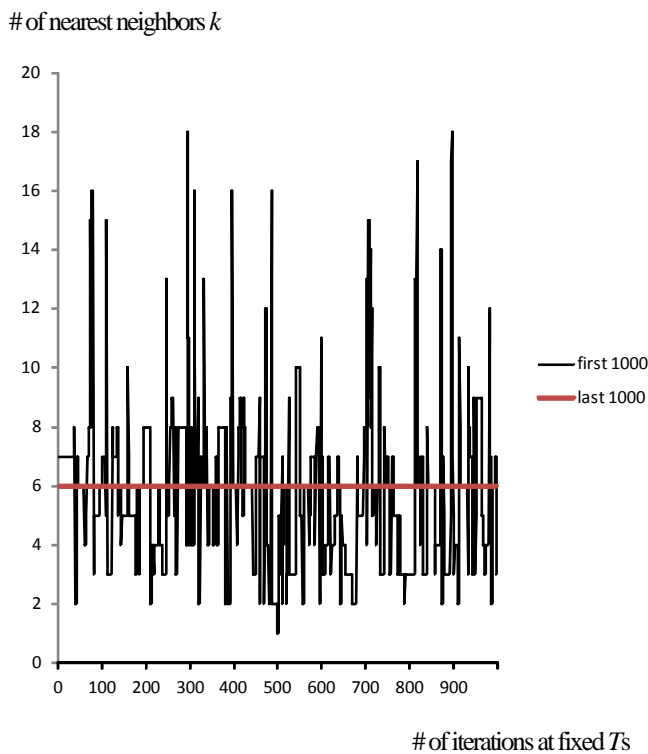**Figure 3 (a)**. Trace plot of the number of the nearest neighbors versus number of temperature reductions for TDSA

T

# of temperature reductions

Figure 1. Cooling schedule

# of nearest neighbors k

# of iterations at fixed Ts

Figure 3 (b). Trace plot of number of the nearest neighbors
in the first and last 1000 iterations

Spatial parameter ρ

# of temperature reductions
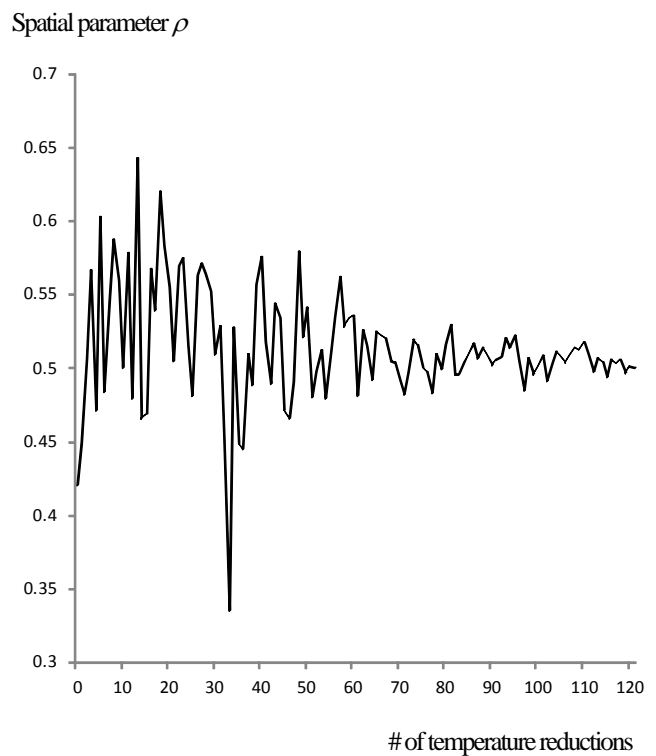
Figure 5. Trace plot of the number of the spatial parameter versus
number of temperature reductions for TDSA

Error variance σ^2
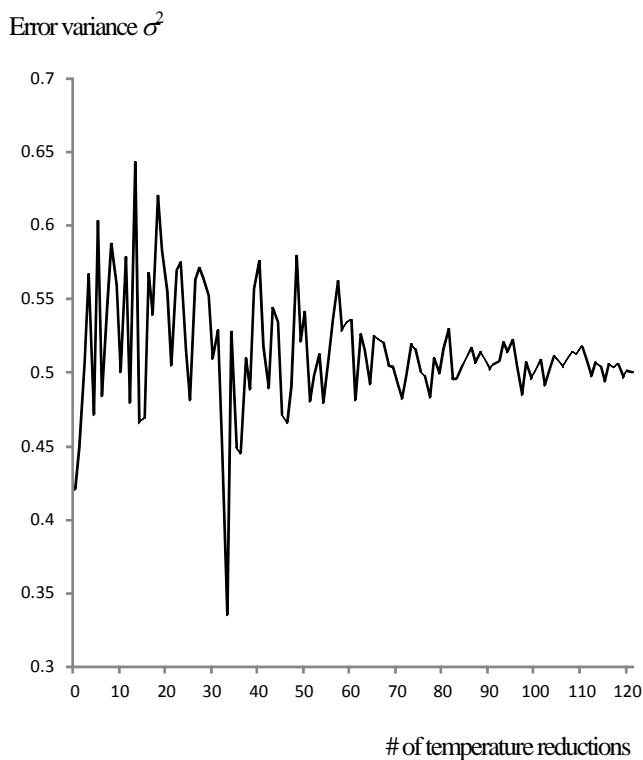

# of temperature reductions

Figure 4. Trace plot of the number of the error variance versus
number of temperature reductions for TDSA

6. Concluding remarks

Although the specification of W is crucial step in spatial econometric models, Anselin (2002) suggested that there is little formal guidance when choosing the correct spatial weights for the given application. Some simulation studies have suggested that the information criteria such as AIC are useful for the weight matrices selection, but if many model candidates exist, then the computational burden of calculating such criteria for each model would be large. The present study developed an automatic model selection algorithm for the SLM using the TDSA algorithm. The performance of our algorithm is examined using the well-known Boston housing dataset.

Our algorithm does not require any somewhat subjective prior settings like the inclusion/exclusion probabilities of the stepwise regression. Moreover, this algorithm may still be feasible in a similar computational time even when the number of explanatory variables or SWMs is much larger. LeSage and Pace (2010) noted that specification of W is "the biggest myth" in spatial econometrics; hence, continued efforts should be devoted to this topic for the further evaluation of spatial econometrics.
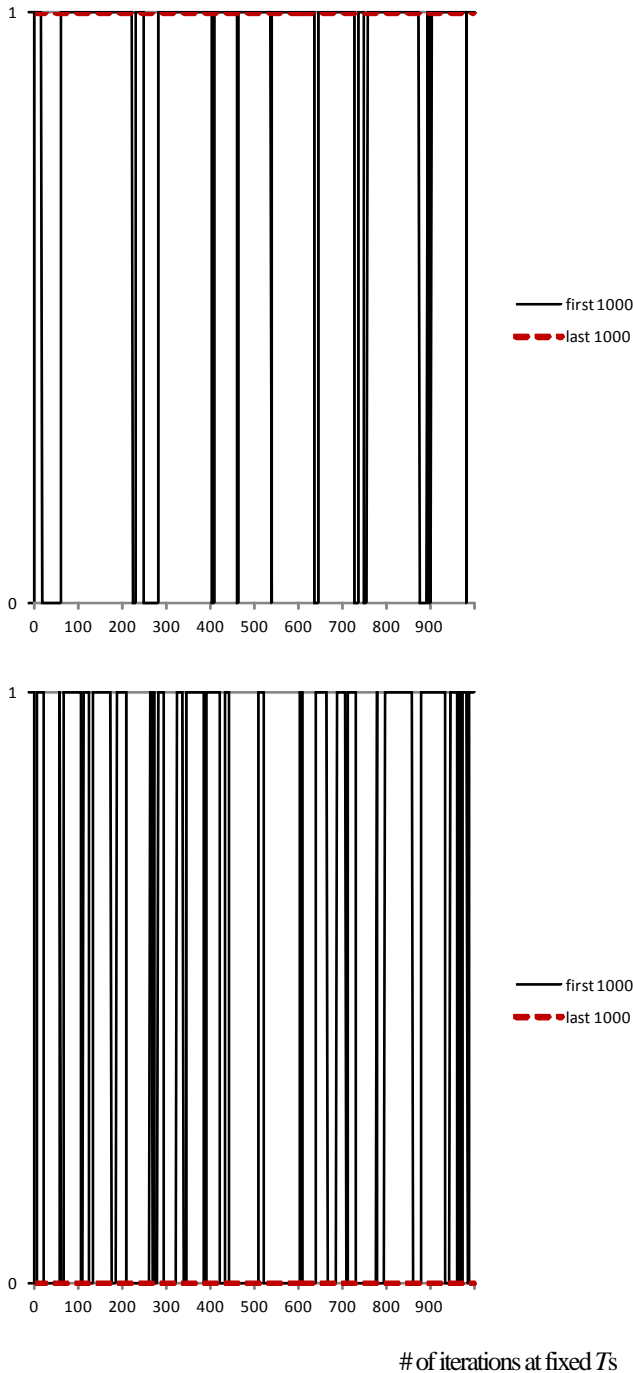
Figure 6. Trace plot of the CRIM (TOP) and ZN (bottom) variables in the first and last 1000 iterations

(1 denotes included and 0 denotes not included in the model)

References

1) Aldstadt, J., and Getis, A. (2006). Using AMOEBA to create a spatial weights matrix and identify spatial clusters, *Geographical Analysis*, 38 (4), 327–343.

2) Andrieu, C., de Freitas, N., Doucet, A., and Jordan, M.I. (2003). An introduction to MCMC for machine learning, *Machine Learning*, 50 (1-2), 5–43.

3) Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

4) Anselin, L. (2002). Under the food: Issues in the specification and interpretation of spatial regression models, *Agricultural Economics*, 27 (3), 247–267.

5) Anselin, L. (2010). Thirty years of spatial econometrics, *Papers in Regional Science*, 89 (1), 3–25.

6) Anselin, L., and Bera, A.K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics, *Handbook of Applied Economic Statistics* (eds. Ullah, A., and Giles, D.E.), 237–289, Marcel Dekker, New York.

7) Anselin, L., and Lozano-Gracia, N. (2009). Spatial hedonic models, *Palgrave Handbook of Econometrics 2* (eds. Patterson, K., and Mills, T.C.), 1213–1250, Palgrave Macmillan, Basingstoke.

8) Arbia, G. (2006). *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*, Springer-Verlag, Berlin.

9) Bhattacharjee, A., and Jensen-Butler, C. (2005). Estimation of spatial weights matrix in a spatial error model, with an application to diffusion in housing demand, CRIEFF Discussion Paper No. 0519, University of St. Andrews, UK.

10) Brasington, D.M., and Hite, D. (2005). Demand for environmental quality: A spatial hedonic analysis, *Regional Science and Urban Economics*, 35 (1), 57–82.

11) Brooks, S.P., Friel, N., and King, R. (2003). Classical model selection via simulated annealing, *Journal of the Royal Statistical Society: Series B*, 65 (2), 503–520.

12) Brooks, S.P., and Morgan, B.J.T. (1995). Optimization using simulated annealing, *The Statistician*, 44 (2), 241-257

13) Can, A. (1990). The Measurement of neighborhood dynamics in urban house prices, *Economic Geography*, 66 (3), 254–272.

14) Can, A. (1992). Specification and estimation of hedonic housing price models, *Regional Science and Urban Economics*, 22 (3), 453–474.

15) Claeskens, G., and Hjort, N.L. (2008). *Model Selection and Model Averaging (Cambridge Series in Statistical and Probabilistic Mathematics)*, Cambridge University Press, Cambridge.

16) Clyde, M. (1999). Comment on 'Bayesian model averaging: a tutorial, *Statistical Science*, 14 (4), 401–404.

17) Cotteleer, G., Stobbe, T., and van Kooten, G.C. (2011). Bayesian model averaging in the context of spatial hedonic pricing: An application to farmland values, *Journal of Regional Science*, available online.

18) Crespo Cuaresma, J., and Feldkircher, M. (2010). *Spatial filtering, model uncertainty and the speed of income convergence in Europe*, working papers 160, Oesterreichische Nationalbank.

19) Cressie, N., and Wikle, C. (2011). *Statistics for Spatio-Temporal Data*, John Wiley & Sons, New York.

20) Elhorst, J.P. (2010). Applied spatial econometrics: Raising the bar, *Spatial Economic Analysis*, 5 (1), 9–28.

21) Fernández-Vázquez, E., Mayor-Fernández, M., and Rodríguez-Vález, J. (2009). Estimating spatial autoregressive models by GME-GCE techniques, *International Regional Science Review*, 32 (2), 148–172.

22) Fingleton, B. (2009). Spatial autoregression, *Geographical Analysis*, 41 (4), 385–391.

23) Folmer, H., and Oud, J. (2008). How to get rid of W: a latent variables approach to modelling spatially lagged variables, *Environment and Planning A*, 40 (10), 2526–2538.

24) Getis, A., and Aldstadt, J. (2004). Constructing the spatial weights matrix using a local statistic, *Geographical Analysis*, 36 (2), 90–104.

25) Getis, A., and Ord, J.K. (1992). The analysis of spatial association by use of distance statistics, *Geographical Analysis*, 24 (3), 189–206.

26) Gilley, O.W., and Pace, R.K. (1996). On the Harrison and Rubinfeld data, *Journal of Environmental Economics and Management*, 31 (3), 403–405.

27) Godsill, S. (2001). On the relationship between Markov Chain Monte Carlo methods for model uncertainty, *Journal of Computational and Graphical Statistics*, 10 (2), 230–248.

28) Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, 82 (4), 711–732.

29) Griffith, D.A. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*, Springer-Verlag, Berlin.

30) Harris, R., Moffat, J., and Kravtsova, V. (2011). In search of 'W', *Spatial Economic Analysis*, 6 (3), 249–270.

31) Harrison, D.Jr., and Rubinfeld, D.L. (1978). Hedonic housing prices and the demand for clean air, *Journal of Environmental Economics and Management*, 5 (1), 81–102.

32) Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57 (1), 97–109.

33) Hepple L.W. (1995a). Bayesian techniques in spatial and network econometrics: 1. Model comparison and posterior odds, *Environment and Planning A*, 27 (3), 447–469.

34) Hepple, L.W. (1995b). Bayesian techniques in spatial and network econometrics: 2. Computational methods and algorithms, *Environment and Planning A*, 27 (4), 615–644.

35) Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. (1999). Bayesian model averaging: A tutorial (with Discussion), *Statistical Science*, 14 (4), 382–417, [Corrected version.], Correction: 15, 193–195.

36) Johnson, D.S. and Hoeting, J.A. (2011). Bayesian multimodel inference for spatial regression models, *PLoS ONE*, 6 (11): e25677. doi:10.1371/journal.pone.0025677.

37) Kakamu, K. (2009). Small sample properties and model choice in spatial models: A Bayesian approach, *Far East Journal of Applied Mathematics*, 34, 31–56.

38) Kakamu, K. and Wago, H. (2008). Small-sample properties of panel spatial autoregressive models: Comparison of the Bayesian and maximum likelihood methods, *Spatial Economic Analysis*, 3 (3), 305–319.

39) Kelejian, H.H., and Prucha, I.R. (2007). HAC estimation in a spatial framework, *Journal of Econometrics*, 140 (1), 131–154.

40) Kitagawa, G. (1997). Information criteria for the predictive evaluation of Bayesian models, *Communications in Statistics, Theory and Methods*, 26 (9), 2223–2246.

41) Koschinsky, J., Lozano-Gracia, N., and Piras, G. (2011). The welfare benefit of a home's location: An empirical comparison of spatial and non-spatial model estimates, *Journal of Geographical Systems*, available online.

42) Kostov, P. (2010). Model boosting for spatial weighing matrix selection in spatial lag models, *Environment and Planning B*, 37 (3), 533–549.

43) Lee, L.-F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models, *Journal of Econometrics*, 137 (2), 489–514.

44) Leeb, H., and Pötscher, B.M. (2005). Model selection and inference: Facts and fiction, *Econometric Theory*, 21 (1), 21–59.

45) LeSage, J.P., and Fischer, M. (2008). Spatial growth regressions: Model specification, estimation and interpretation, *Spatial Economic Analysis*, 3 (3), 275–304.

46) LeSage, J.P., and Pace, R.K. (2004). Models for spatially dependent missing data, *Journal of Real Estate Finance and Economics*, 29 (2), 233–254.

47) LeSage, J.P. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*, Taylor & Francis, Boca Raton.

48) LeSage, J.P., and Pace, R.K. (2010). The biggest myth in spatial econometrics, available at SSRN: *http://ssrn.com/abstract=1725503*

49) LeSage, J.P., and Pace, R.K. (2011). Pitfalls in higher order model extensions of basic spatial regression methodology, paper presented at the North American Meetings of the Regional Science Association International, Miami, FL, November, 2011.

50) LeSage, J.P., and Parent, O. (2007). Bayesian model averaging for spatial econometric models, *Geographical Analysis*, 39 (3), 241–267.

51) LeSage, J.P., and Polasek, W. (2008). Incorporating transportation network structure in spatial econometric models of commodity flows, *Spatial Economic Analysis*, 3 (2), 225–245.

52) Magnus, J.R., Powell, O., and Prüfer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics, *Journal of Econometrics*, 154 (2), 139–153.

53) Manski, C.F. (1993). Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies*, 60 (3), 531–542.

54) Mur, J., and Paelinck, J.H.P. (2011). Deriving the W–matrix via p–

median complete correlation analysis of residuals, *the Annals of Regional Science*, available online.

55) Ord, J.K., and Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application, *Geographical Analysis*, 27 (4), 286–306.

56) Pace, R.K., Barry, R., and Sirmans, C.F. (1998). Spatial statistics and real estate, *Journal of Real Estate Finance and Economics*, 17 (1), 5–13.

57) Pace, R.K., and Gilley, O.W. (1997). Using the spatial configuration of the data to improve estimation, *Journal of Real Estate Finance and Economics*, 14 (3), 333–340.

58) Pham, D.T., and Karaboga, D. (2000). *Intelligent Optimisation Techniques: Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks*, Springer-Verlag, London.

59) Raftery, A.E., Madigan, D., and Hoeting, J.A. (1997). Bayesian model averaging for linear regression models, *Journal of the American Statistical Association*, 92 (437), 179–191.

60) Rincke, J. (2010). A commuting-based refinement of the contiguity matrix for spatial models, and an application to local police expenditures, *Regional Science and Urban Economics*, 40 (5), 324–330.

61) Robinson P. M. (2008). Developments in the analysis of spatial data, *Journal of the Japan Statistical Society*, 38 (1), 87–96.

62) Seya, H., and Tsutsumi, M. (2012). Application of model averaging techniques to spatial hedonic land price models, *Econometrics: New Research* (eds. Mendez, S.A. and Vega, A.M.), *Nova Science Publishers*, New York.

63) Seya, H., Tsutsumi, M. and Yamagata, Y. (2012). Income convergence in Japan: A Bayesian spatial Durbin model approach, *Economic Modelling*, 29 (1), 60–71.

64) Smith, T.E. (2009). Estimation bias in spatial models with strongly connected weight matrices, *Geographical Analysis*, 41 (3), 307–332.

65) Stakhovych, S., and Bijmolt, T.H.M. (2009). Specification of spatial models: A simulation study on weights matrices, *Papers in Regional Science*, 88 (2), 389–408.

66) Tiefelsdorf, M., and Griffith, D. (2007). Semiparametric filtering of spatial auto-correlation: The eigenvector approach, *Environment and Planning A*, 39 (5), 1193–1221.

67) Tsutsumi, M., and Seya, H. (2008). Measuring the impact of large-scale transportation project on land price using spatial statistical models, *Paper in Regional Science*, 87 (3), 385–401.

68) Tsutsumi, M., and Seya, H. (2009). Hedonic approaches based on spatial econometrics and spatial statistics: Application to evaluation of project benefits, *Journal of Geographical Systems*, 11 (4), 357–380.

AUTOMATIC SELECTION OF A SPATIAL WEIGHT MATRIX IN SPATIAL ECONOMETRICS: APPLICATION TO A SPATIAL HEDONIC APPROACH

Hajime SEYA, Morito TSUTSUMI, and Yoshiki YAMAGATA

The application of a hedonic approach to real estate data plays an important role in property market analysis and urban planning. The recent progress of spatial econometrics offers a new technique called the "spatial hedonic approach," which considers the element of spatial autocorrelation among property values and attributes that are geographically distributed. The practical difficulties in applying spatial econometric models include the specification of the spatial weight matrix (SWM), which affects the final analysis results. Some simulation studies suggest that information criteria such as AIC are useful for the spatial weight matrices selection, but if many model candidates exist (e.g., when the selections of explanatory variables are performed simultaneously), then the computational burden of calculating such criteria for each model is large. The present study developed an automatic model selection algorithm using the technique of reversible jump MCMC combined with the simulated annealing. The performance of this algorithm is verified using the well-known Boston housing dataset.