Regional Science and Urban Economics 43 (2013) 816–837

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

Bayesian estimation and model selection for spatial Durbin error model with finite distributed lags ☆

Xiaoyi Han *, Lung-fei Lee

Department of Economics, Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 3 December 2012
Received in revised form 29 March 2013
Accepted 17 April 2013
Available online 23 April 2013

JEL classification:
C11
C21

Keywords:
Spatial Durbin error model
Spatial autoregressive model
Matrix exponential spatial specification
Smoothness prior
Marginal likelihood
Bayesian estimation

ABSTRACT

In this paper we investigate a spatial Durbin error model with finite distributed lags and consider the Bayesian MCMC estimation of the model with a smoothness prior. We study also the corresponding Bayesian model selection procedure for the spatial Durbin error model, the spatial autoregressive model and the matrix exponential spatial specification model. We derive expressions of the marginal likelihood of the three models, which greatly simplify the model selection procedure. Simulation results suggest that the Bayesian estimates of high order spatial distributed lag coefficients are more precise than the maximum likelihood estimates. When the data is generated with a general declining pattern or a unimodal pattern for lag coefficients, the spatial Durbin error model can better capture the pattern than the SAR and the MESS models in most cases. We apply the procedure to study the effect of right to work (RTW) laws on manufacturing employment.

1. Introduction

Spatial econometric models applied in regional science and geography have been receiving more attention in various areas of economics. The most popular spatial econometric model is the spatial autoregressive (SAR) model. The SAR model implies a geometrical decay pattern of spatial spillover effects or externalities from levels of neighbors' exogenous characteristics in its reduced form. A geometrical decay pattern of spatial distributed lags is not the only choice of modeling spatial externalities. There are other models which display a different pattern of spatial externalities. Recently, LeSage and Pace (2007) introduce the matrix exponential spatial specification (MESS) model, which exhibits an exponential declining pattern of spatial externalities. These two spatial lag patterns do not exhaust other possible patterns. Furthermore, both the SAR model and the MESS model incorporate global spatial externalities in the sense that they relate all the neighbors in the system to each other. If one wants to capture local spatial externalities,[1] these two models might not be appropriate. In practice, if one wants to incorporate local externalities in the model and there were no formal theoretical guidance on which pattern of spatial externalities to choose, a possible solution is to propose a spatial Durbin error model (SDEM)[2] with finite distributed lags of exogenous regressors, which does not impose strong restrictions on the pattern of spillover effects or externalities. Then we are facing a non-nested model selection problem among a SDEM model, the SAR model and the MESS model. Hence, in addition to the estimation of a SDEM model, it is of interest to construct a model discrimination procedure for them.

In this paper we propose a finite lag SDEM model with a smoothness prior in order to accommodate more flexible patterns of local spatial externalities. We consider the Bayesian MCMC estimation of the SDEM model and the corresponding Bayesian model selection

☆ We are grateful to the editor and two anonymous referees for their helpful comments.
* Corresponding author. Tel.: +1 614 607 1414; fax: +1 614 292 3906.
E-mail addresses: han.293@osu.edu (X. Han), lee.1777@osu.edu (L. Lee).

[1] As mentioned by Anselin (2003), local spatial externalities would be appropriate when the proper spatial range of the explanatory variables is the location itself and its immediate neighbors. For more discussion regarding global and local spatial externalities, see Anselin (2003).

[2] We thank two referees for pointing out that the original name "SLX" for the model is somehow confusing. We have changed the name of the model to SDEM throughout the paper. Related literature about the SDEM model are mentioned in Section 2.

0166-0462/$ – see front matter
http://dx.doi.org/10.1016/j.regsciurbeco.2013.04.006

procedure for the SDEM model, the SAR model and the MESS model. We focus on Bayesian estimation because a direct maximum likelihood (ML) estimation for high order lag coefficients of the SDEM model might be imprecise due to multicollinearity among lagged regressors.[3] As motivated by Shiller's smoothness prior for distributed lag model, we may impose a smoothness prior on the lag coefficients in order to obtain better estimates.

For model selection among non-nested models, both classical approach and Bayesian approach are available in the literature. However, the classical non-nested tests might be unreliable due to imprecise estimates for high order lag coefficients of a SDEM model. For Bayesian approach, Zellner (1971) has set forth the basic theory and the model selection procedure, which involves calculating and comparing the posterior probabilities of competitive models and is feasible for competitive non-nested models. With posterior probabilities for competitive models one can see whether those models are close competitors or one model just dominates all others. Hepple (1995a,b) extends the Bayesian model selection procedure into non-nested spatial models.[4] In particular, he has derived expressions of marginal likelihoods for a number of spatial models including the SAR model and the spatial error model, which greatly simplify the calculation of model posterior probabilities. LeSage and Pace (2007) derive an expression of the marginal likelihood of the MESS model, which could be used to produce Bayesian model comparison procedure for the MESS model and other models. LeSage and Parent (2007) also extend the Bayesian model selection procedure for linear regression models into the SAR model and the spatial error model. Their focus is on comparing models with different matrices of explanatory variables.[5] We can also derive an expression of the marginal likelihood of the SDEM model, which simplifies the model selection procedure.

The paper is organized as follows: Section 2 introduces the SDEM model with a smoothness prior and specifies the model selection problem. Section 3 considers Bayesian MCMC estimation of the SDEM model. Section 4 discusses Bayesian model selection procedure. Section 5 summarizes simulation results to investigate sampling properties of our Bayesian estimation method and model selection procedure. Section 5 includes an empirical study of the effect of right to work (RTW) on manufacturing employment. Conclusions are drawn in Section 6. Technical details and tables are given in Appendix A.

2. The models

Consider a spatial autoregressive (SAR) model

$$Y_n = \lambda W_n Y_n + c l_n + X_n \beta + V_n \tag{2.1}$$

where l_n is a $n \times 1$ column vector of ones, and c is the coefficient of the intercept term. X_n is a $n \times k$ dimensional matrix of nonstochastic exogenous variables. W_n is a spatial weight matrix with known constants with a zero diagonal. The error terms in $V_n = (v_{n1}, v_{n2}, \ldots, v_{nn})'$ are assumed to be i.i.d normally distributed with mean 0 and variance σ^2. The reduced form of the SAR model reviews its implication in spatial externalities in terms of a geometric declining pattern of spatial distributed lags on exogenous regressors,

$$Y_n = (I_n - \lambda W_n)^{-1} c l_n + X_n \beta + \sum_{q=1}^{\infty} W_n^q X_n \lambda^q \beta + (I_n - \lambda W_n)^{-1} V_n. \tag{2.2}$$

Here the nonzero elements of rows of W_n with $q \geq 1$ represent the qth order neighbors. Then the specification in Eq. (2.2) has spillover effects or externalities generated by the regressors xs from one's different level of neighbors being geometrically declining.

As an alternative to the SAR specification, LeSage and Pace (2007) introduce the MESS model with the specification $S_n^{ex}(\mu) Y_n = c l_n + X_n \beta + V_n$, of which the reduced form is

$$Y_n = S_n^{ex}(\mu)^{-1} c l_n + S_n^{ex}(\mu)^{-1} X_n \beta + S_n^{ex}(\mu)^{-1} V_n, \tag{2.3}$$

where $S_n^{ex}(\mu) = e^{\mu W_n} = I_n + \sum_{q=1}^{\infty} \frac{1}{q!} (\mu W_n)^q$. The model introduces an exponential decay pattern of spatial externalities via its spatial distributed lags on regressors.

However, according to Eqs. (2.2) and (2.3), both the SAR model and the MESS model may be rather "restrictive" because they impose strong restrictions on the pattern of spatial externalities. Moreover, Eqs. (2.2) and (2.3) are infinite summations for lagged regressors $W_n X_n$s, implying that both models only incorporate global spatial externalities because they relate all the neighbors in the system to each other. To allow for a more flexible pattern of spatial externalities and incorporate local spatial externalities, a possible specification is the following finite spatial distributed lag model:

$$\begin{aligned} Y_n = {} & c l_n + X_n \beta + W_n X_n \beta_1 + W_n^2 X_n \beta_2 + \ldots + W_n^m X_n \beta_m \\ & + (I_n - \rho W_n)^{-1} V_n. \end{aligned} \tag{2.4}$$

Here we combine local externalities for X_n with global externalities for V_n in the model.[6] We do not impose strong restrictions on the lag coefficients $\beta_1, \ldots, \beta_m$. As a result, the model should be able to accommodate more flexible patterns of spatial externalities. LeSage and Pace (2009) label the model in Eq. (2.4) as a spatial Durbin error model (SDEM). Some literatures have discussed other versions of the SDEM model, for example, Lacombe et al. (2012) and LeSage and Christina (2012). However, neither of them includes higher order spatial distributed lag terms of X_n in the model.[7] In empirical research, we usually do not have an economic theory that tells us which pattern of spatial externalities to choose. We are facing a non-nested model selection problem among the SDEM model, the SAR model and the MESS model.

However, one concern about the SDEM model in Eq. (2.4) is that, when we use the ML method to directly estimate the model, estimates of high order coefficients βs might be poor due to possible multicollinearity among the lagged regressors $X_n, W_n, X_n, \ldots, W_n^m X_n$.[8] In practice, one might expect that those coefficients would change smoothly, so we borrow the idea of smoothness prior from Shiller (1973) to impose some random restrictions on the lag coefficients and use the Bayesian MCMC method to estimate the SDEM model. For model selection, we follow the procedure advocated in Zellner (1971) and compare the posterior probabilities for the three models.

[3] There might be multicollinearity among the lagged exogenous regressors in the SDEM model. This is similar to the source of multicollinearity in the spatial Durbin model (Anselin, 1988). See Section 2 for more discussions.

[4] Hepple has considered a bunch of non-nested model selection problems for spatial models. For example, comparing SAR models with different spatial weight matrices or comparing spatial error models with different spatial weight matrices. For more details, see Hepple (1995a,b).

[5] LeSage and Parent deal with the cases where the number of possible models that consist of different combinations of candidate explanatory variables is too large. Calculation of posterior probabilities for all models is difficult. They rely on a Markov Chain Monte Carlo model composition methodology proposed by Madigan and York (1995). For more details, see LeSage and Parent (2007).

[6] Here local externalities are incorporated by the lagged exogenous regressors $W_n X_n$s while global externalities are captured by $(I_n - \rho W_n)^{-1} V_n$. The model in Eq. (2.4) is also a generalized version of Eq. (30) in Anselin (2003), with higher order lagged exogenous regressors $W_n X_n$s. See Anselin (2003) for more discussions.

[7] Lacombe et al. (2012) only include the first order spatial distributed lag in their SDEM model while LeSage and Christina (2012) include two different W_ns in the model.

[8] The estimates of the lag coefficients might have a large variance.

3. The SDEM model with smoothness prior

3.1. The smoothness prior

Shiller (1973) proposes a smoothness prior for the lag coefficients of the linear distributed lag models:

$$y_t = \sum_{i=0}^{q-1} x_{t-i}\beta_i + v_t$$

where q is a known constant representing the lag length and β_is are the unknown lag coefficients. OLS estimate of the β_is might provide us with an erratic or jagged shape because of multicollinearity among x_ts. In many circumstances, it is unreasonable to believe that a stable relationship should take such a form. Instead we would expect the estimates of lag coefficients to form a “smooth” or “simple” curve. Thus Shiller suggests to impose random restrictions on β_is. That is, he assumes a normal prior for differences of β_is and considers a Bayesian approach to estimate the model.

Here we consider a similar smoothness prior for the lag coefficients in the SDEM model. Let $Z_n = (l_n, X_n, W_n, X_n, \ldots, W_n^m X_n)$ and $\gamma = (c, \beta', \beta'_1, \ldots, \beta'_m)' = (c, \delta)$. Also let $\theta = (\rho, \gamma', \sigma^2)'$ be the parameter vector. The SDEM model can be rewritten as:

$$Y_n = Z_n\gamma + (I_n - \rho W_n)^{-1} V_n. \tag{3.1}$$

Let $R_n(\rho) = I_n - \rho W_n$. The Cochrane–Orcutt transformation of the model is

$$R_n(\rho) Y_n = R_n(\rho) Z_n \gamma + V_n.$$

Denote $Y_n(\rho) = R_n(\rho) Y_n$ and $Z_n(\rho) = R_n(\rho) Z_n$. Then the likelihood function of the model is:

$$f(Y_n|\theta) \propto \left(\sigma^2\right)^{-\frac{n}{2}} |R_n(\rho)| \exp\left(-\frac{1}{2\sigma^2}(Y_n(\rho) - Z_n(\rho)\gamma)'(Y_n(\rho) - Z_n(\rho)\gamma)\right). \tag{3.2}$$

We want to define a prior for the lag coefficient vector δ. Following Shiller (1973), we first introduce the $d+1$ difference matrix and define the $d+1$ difference of δ. Let R be a $(m-d) \times (m+1)$ matrix of $d+1$ differences, with rank $p = m - d$. For instance, if $m = 4$ and $d = 1$, the second difference matrix is:

$$R = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{pmatrix}.$$

With the $d+1$ difference matrix, the $d+1$ difference of δ is defined as $u = R_d\delta$ where $R_d = R \otimes I_k$. For example, with $m = 4$, $k = 1$ and $d = 1$, the second difference u is

$$u = \begin{pmatrix} \beta - 2\beta_1 + \beta_2 \\ \beta_1 - 2\beta_2 + \beta_3 \\ \beta_2 - 2\beta_3 + \beta_4 \end{pmatrix}.$$

Note that u is just a linear transformation of δ. The idea of the smoothness prior is to assume u to being spherically normally distributed with mean 0 and variance $L^2\sigma^2$, with L representing one's prior belief regarding how small the $d+1$ difference should be. Then the $d+1$ difference for δ would be relatively smooth due to the smooth nature of the normal distribution. The prior of δ can be fully determined by the prior of u. Specifically, if we assume a uninformative prior for $\beta, \ldots, \beta_d$,

$$\pi(\beta_h) \propto \mathcal{U}(-C_h, C_h), \quad h = 1, 2, \ldots, k$$
$$\pi\left(\beta_{ij}\right) \propto \mathcal{U}\left(-C_{ij}, C_{ij}\right), \quad i = 1, 2, \ldots, d, \quad j = 1, 2, \ldots, k,$$

where C_hs and C_{ij}s are known constants, β_h is the hth component of the k-dimensional parameter vector β_h, β_{ij} is the jth component of the k-dimensional parameter vector β_i, and $\mathcal{U}(-c, c)$ is the uniform density on the interval $(-c, c)$. With $(d+1) \times k$ priors for $\beta, \ldots, \beta_d$ and u being $(m-d) \times k$, we can have the one–one mapping from these priors to the $(m+1) \times k$ dimensional δ. Explicitly, the prior distribution of δ is

$$\pi\left(\delta|\sigma^2\right) \propto \pi\left(u|\sigma^2\right) \propto \left(\frac{1}{L^2\sigma^2}\right)^{\frac{(m-d)k}{2}} \exp\left(-\frac{u'u}{2L^2\sigma^2}\right). \tag{3.3}$$

Here we assume the prior for δ is conditional on σ^2 since this will greatly simplify our calculation for the posterior probability of the model. Note that it is a common theme in the literature.[9] For example, LeSage and Parent (2007) assume the g-prior in Zellner (1986) for the coefficients of the explanatory variables, conditional on σ^2 in the SAR model and the spatial error model. Without this assumption, the resulting marginal likelihood will not have a simple form and some other techniques for calculating the marginal likelihood would have to be used, e.g. Chib (1995) or Chib and Jeliazkov (2001).

Alternatively, we can consider a more informative smoothness prior for δ, in which we assume a jointly normal prior for $\beta, \ldots, \beta_d$ and u, where $u = R_{di}\delta$ with R_{di} being a $(m+1)k \times (m+1)k$ nonsingular matrix. Specifically, $R_{di} = \tilde{R} \otimes I_k$ and $\tilde{R}$ is augmented with $(d+1)$ rows of unit row vectors to the $d+1$ order difference matrix. For instance, if $m = 4$ and $d = 1$,

$$R_{di} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{pmatrix}.$$

The informative smoothness prior of δ is

$$\pi\left(\delta|\sigma^2\right) \propto \pi\left(u|\sigma^2\right) \propto \left(L^2\sigma^2\right)^{-\frac{(m+1)k}{2}} \exp\left(-\frac{u'u}{2L^2\sigma^2}\right). \tag{3.4}$$

3.2. Bayesian MCMC estimation of the SDEM model with a smoothness prior

We illustrate the Bayesian MCMC estimation of the SDEM model with an informative smoothness prior. The estimation is carried out using Gibbs sampling where one step Metropolis–Hastings sampling is involved at each draw. Recall that $\theta = (\rho, c, \delta', \sigma^2)'$. We applied a set of conjugate prior for σ^2, c, the informative smoothness prior for δ (from the prior of u) and a uniform prior for ρ. So the priors for θ are:

$$\begin{aligned} &\pi(\rho) \sim \mathcal{U}\left(-\frac{1}{\tau_n}, \frac{1}{\tau_n}\right) \\ &\pi\left(c|\sigma^2\right) \sim \mathcal{N}\left(\bar{c}, p^2\sigma^2\right) \\ &\pi\left(\mu|\sigma^2\right) \sim \mathcal{N}\left(0, L^2\sigma^2 I_{(m+1)k}\right) \\ &\pi\left(\sigma^2\right) \sim \mathcal{IG}\left(\frac{a}{2}, \frac{b}{2}\right), \end{aligned} \tag{3.5}$$

[9] We thank a referee for pointing out this.

where $\mathcal{N}\left(\overline{c}, p^2\sigma^2\right)$ is the normal density with mean $\overline{c}$ and variance $p^2\sigma^2$, $\mathcal{IG}(\frac{a}{2}, \frac{b}{2})$ is the inverse gamma density with shape parameter $\frac{a}{2}$ and scale parameter $\frac{b}{2}$, and $\mathcal{U}\left(-\frac{1}{\tau_n}, \frac{1}{\tau_n}\right)$ is the uniform density on the interval $\left(-\frac{1}{\tau_n}, \frac{1}{\tau_n}\right)$.[10] Here the value of ρ is restricted in the interval $\left(-\frac{1}{\tau_n}, \frac{1}{\tau_n}\right)$, where $\tau_n = \min\{\max_{1 \le i \le n}\sum_{j=1}^{n} w_{ij}, \max_{1 \le j \le n}\sum_{i=1}^{n} w_{ij}\}$.[11] p, L, a and b are all prior parameters. Combine the likelihood function with those priors, the posterior distribution of θ is

$$\pi(\theta|Y_n) \propto \left(\sigma^2\right)^{-\left(\frac{n+a+(m+1)k+1}{2}+1\right)}|R_n(\rho)| \times \exp\left\{-\frac{1}{2\sigma^2}\left[(Y_n(\rho)-Z_n(\rho)\gamma)'(Y_n(\rho)-Z_n(\rho)\gamma)+b + \frac{(c-\overline{c})^2}{p^2}+\delta'\frac{R'_d}{L}\frac{R_d}{L}\delta\right]\right\}. \tag{3.6}$$

Recall that $\gamma = (c,\delta')'$. Let $\overline{\gamma} = \left(\overline{c}, O'_{(m+1)k\times 1}\right)'$ and $M = \begin{pmatrix} \frac{1}{p} & O \\ O & \frac{R_d}{L}\end{pmatrix}$. Then we have

$$\frac{(c-\overline{c})^2}{p^2}+\frac{\delta' R'_d R_d \delta}{L^2} = (\gamma-\overline{\gamma})'M'M(\gamma-\overline{\gamma}).$$

So the posterior distribution for θ can be rewritten as

$$\pi(\theta|Y_n) \propto \left(\sigma^2\right)^{-\left(\frac{n+a+(m+1)k+1}{2}+1\right)}|R_n(\rho)| \times \exp\left\{-\frac{1}{2\sigma^2}\left[(Y_n(\rho)-Z_n(\rho)\gamma)'(Y_n(\rho)-Z_n(\rho)\gamma) + b + (\gamma-\overline{\gamma})'M'M(\gamma-\overline{\gamma})\right]\right\}. \tag{3.7}$$

With conjugate priors, γ and σ^2 can be directly sampled by Gibbs sampling steps since their posterior conditional distributions have closed forms.[12] But we must use the Metropolis–Hastings (M–H) algorithm for sampling the spatial parameter ρ.[13] Note that the full conditional distribution for ρ is nonstandard due to the presence of W_n. Therefore, a "standard" Gibbs sampling step fails but a Metropolis step works.[14] The whole sampling procedure is:

1. Initial values σ_0^2, ρ_0
2. Given σ_{t-1}^2 and ρ_{t-1} from last iteration:
   (a) Sample $\gamma_t|\rho_{t-1},\sigma_{t-1}^2$
   (b) Sample $\sigma_t^2|\rho_{t-1},\gamma_t$
   (c) Sample $\rho_t|\gamma_t,\sigma_t^2$
3. Repeat step 2 for a larger number of draws and burn in the first B draws.[15]

3.2.1. Sampling γ and σ^2

Denote $\tilde{Y}_n(\rho) = \left(Y'_n(\rho), (M\overline{\gamma})'\right)'$ and $\tilde{Z}_n(\rho) = \left(Z'_n(\rho), M'\right)'$. Also let $A_1(\rho) = \tilde{Z}'_n(\rho)\tilde{Z}_n(\rho)$ and $T_\gamma(\rho) = A_1(\rho)^{-1}\tilde{Z}'_n(\rho)\tilde{Y}_n(\rho)$. The conditional posterior distribution of γ, given ρ and σ^2 is

$$\gamma\big|\rho,\sigma^2 \sim \mathcal{N}\left(T_\gamma(\rho), \sigma^2 A_1^{-1}(\rho)\right). \tag{3.8}$$

The conditional posterior distribution for σ^2 given other parameters is still an inverse gamma distribution,

$$\sigma^2\big|\rho,\gamma \sim IG\left(\frac{a_p}{2},\frac{b_p}{2}\right) \tag{3.9}$$

where

$$a_p = n+a+(m+1)k+1$$
$$b_p = b+(\gamma-\overline{\gamma})'M'M(\gamma-\overline{\gamma})+(Y_n(\rho)-Z_n(\rho)\gamma)'(Y_n(\rho)-Z_n(\rho)\gamma).$$

3.2.2. Sampling for ρ

The expression for the conditional posterior distribution of ρ is:

$$\pi\left(\rho\big|\gamma,\sigma^2\right) \propto |R_n(\rho)|\exp\left(-\frac{1}{2\sigma^2}(Y_n(\rho)-Z_n(\rho)\gamma)'(Y_n(\rho)-Z_n(\rho)\gamma)\right). \tag{3.10}$$

We adopt a Metropolis–Hastings sampling procedure for ρ. We use a normal distribution along with a tuned random walk procedure suggested by Holloway et al. (2002) to produce candidate value for ρ. Denote ρ^* as the candidate value of ρ. Also we label the current value as ρ^c. The relationship between the candidate value and the present value is $\rho^* = \rho^c + h_\rho \times \mathcal{N}(0,1)$, where h_ρ is the tuning parameter for ρ and $\mathcal{N}(0,1)$ is the standard normal distribution. As suggested by LeSage and Pace (2009), the goal of tuning the proposal from the normal distribution is to ensure that the M–H sampling procedure moves over the entire conditional distribution. We also adjust the tuning parameters during the MCMC sampling process. If the acceptance rate falls below 0.4, we adjust $h'_\rho = h_\rho/1.1$. But if the acceptance rate rises above 0.6, we set $h'_\rho = 1.1 \times h_\rho$. The idea is that when the acceptance rate is too low, we decrease the variance of the normal distribution to make a new proposal closer to the current value. But if the acceptance rate is too high, we increase the variance so the new proposal can range more widely over the domain of the parameters.

However, one concern about this procedure is that one might sample from the wrong marginal distribution when the "automatic" step-size adjustment are made during sampling.[16] One solution is to fix the tuning parameter after the burn-in period. We also consider a modified version of the tuned random walk procedure.[17] That is, we collect values of h_ρ during burn-in and use the mean of those

[10] Here we use a uniform prior for ρ. Apparently the prior density function has the value $\frac{1}{b-a}$ on (a,b), which changes as a and b change. An alternative to the uniform prior would be a four parameter beta prior $\mathcal{B}(d,d,\rho_{\min},\rho_{\max})$, introduced by LeSage and Parent (2007), with the density function $\pi(\rho) \sim \frac{1}{Beta(d,d)}\frac{(\rho-\rho_{\min})^{d-1}(\rho_{\max}-\rho)^{d-1}}{(\rho_{\max}-\rho_{\min})^{2d-1}}$ on the support $(\rho_{\min}, \rho_{\max})$. So researchers can modify the parameter space of ρ by setting different values of $\rho_{\min}$ and $\rho_{\max}$. For example, LeSage and Parent (2007) use this beta prior and set $\rho_{\min} = -1$ and $\rho_{\max} = 1$.

[11] This interval is suggested by Kelejian and Prucha (2010) in which $I_n - \rho W_n$ is nonsingular for all values of ρ. τ_n can be viewed as a single scaling normalization factor for W_n. As pointed out by a referee, the interval $\left(\frac{1}{\lambda_{\min}}, 1\right)$ may also be used for ρ, where λ_n is the minimum eigenvalue of W_n. Compared with the eigenvalues of W_n, it is relatively easy to compute for a large sample size (Kelejian and Prucha, 2010). Furthermore, theoretically, it has the advantage that if the original matrix is symmetric, the rescaled weights matrix will remain symmetric. In our Monte Carlo study and empirical application, τ_n is equal to 1. So that it is not different from many practice researchers who adopt the more convenient parameter space $(-1,1)$.

[12] The Gibbs sampler allows one to obtain draws from some complicated posterior distribution by sampling sequentially from conditional distributions. Gelfand and Smith (1990) demonstrate its use in many statistical applications. Casella and George (1992) provide a simple introduction to the Gibbs sampler. More discussions of the implementation of the Gibbs sampler for spatial models can be found in LeSage and Pace (2009).

[13] The M–H algorithm is first proposed by Metropolis et al. (1953) and generalized by Hastings (1970). One can rely on the M–H algorithm to sample from some unconditional or conditional distribution when the distributional form is nonstandard. Chib and Greenberg (1995) provide a detailed introductory exposition of the algorithm. Conditions regarding the convergence of Markov chain for the M–H algorithm can be found in Tierney (1994) and Chib and Greenberg (1996). Implementations of the M–H algorithm for various spatial models can be found in LeSage and Pace (2009).

[14] We thank a referee for suggesting to emphasize on this need.

[15] In this paper we burn in the first 20% samples.

[16] We thank a referee for pointing out this.

[17] This suggestion comes from a referee. We thank for that.

values as the tuning parameter during "real" sampling. The modified Bayesian estimation results of the SDEM models are summarized in Tables 3, 4, 6 and 8.

4. Model comparison

One advantage of Bayesian method is that it can provide a formal approach for model comparison even when the competing models are non-nested. Our focus is on comparing spatial models with different patterns of spatial externalities. Let M_i and M_j denote the ith and jth competing models. Also let Y_n represent the data and θ_i represent the vector of parameters of M_i. If we specify the prior probabilities $P(M_i)$ and $P(M_j)$ for the two models, as well as the prior density $\pi(\theta)$ for the parameters, the posterior odds, which is the ratio of the products of the prior and the marginal likelihood of the two models, is:

$$\underbrace{\frac{P(M_i|Y_n)}{P(M_j|Y_n)}}_{\text{Posterior odds}} = \underbrace{\frac{P(M_i)}{P(M_j)}}_{\text{Prior odds}} \times \underbrace{\frac{f_i(Y_n|M_i)}{f_j(Y_n|M_j)}}_{\text{Bayes odds}}$$

where $f_i(Y_n|M_i) = \int f_i(Y_n|\theta_i, M_i)\pi(\theta_i, M_i)d\theta_i$.

We usually assume the same prior for all models. Thus, we only need to pay attention to the Bayes factor, which is just the ratio of the two model's marginal likelihoods. The model with a larger

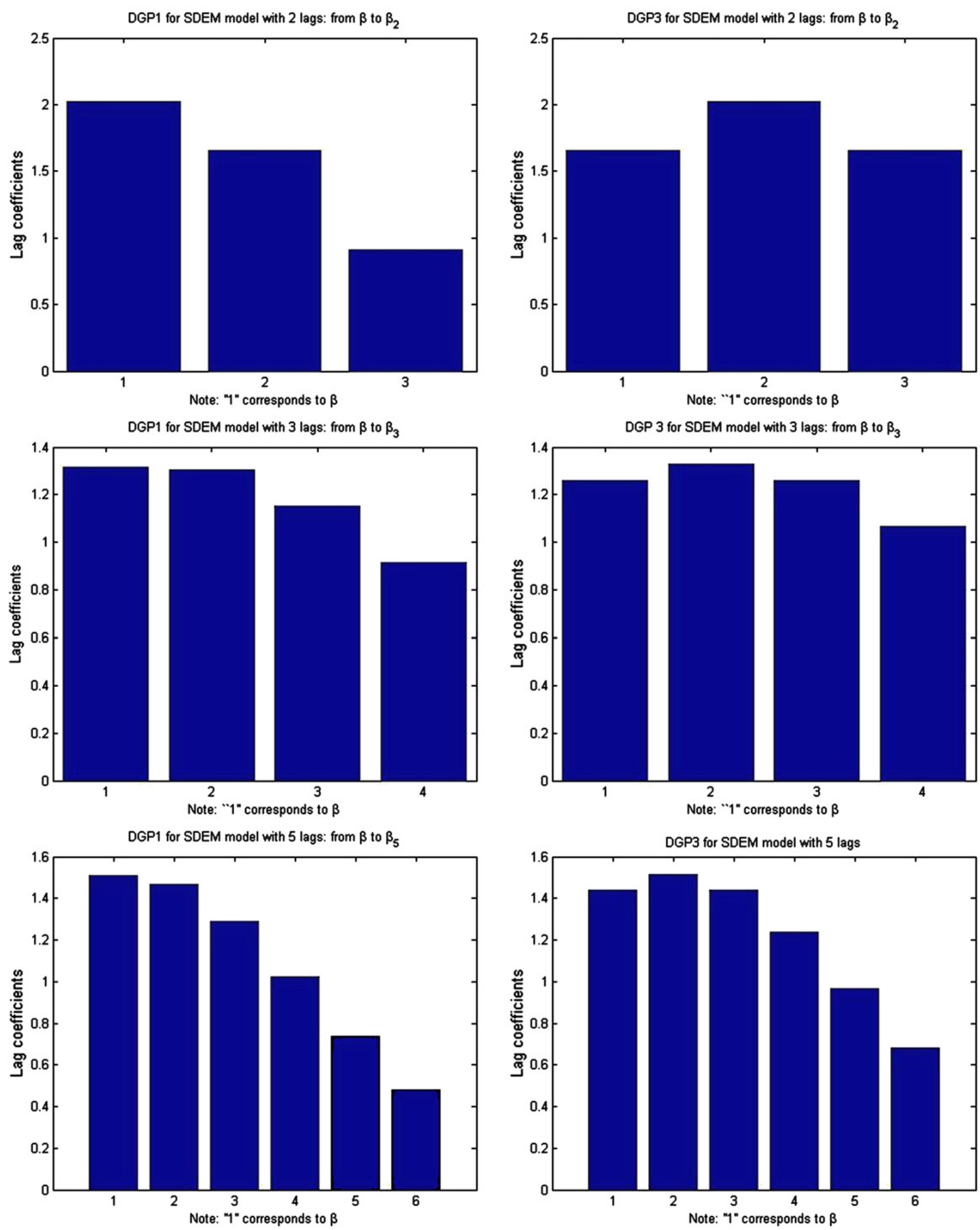

Fig. 1. DGPs for the SDEM model.

marginal likelihood has larger posterior probability and hence is more likely to be the model that generates the data. If we have K models involved, the posterior probability of model k is:

$$P(M_k|Y_n) = \frac{P(M_k)f_k(Y_n|M_k)}{\sum_{j=1}^K P\left(M_j\right)f_j\left(Y_n|M_j\right)} = \frac{P(M_k)\int f_k(Y_n|\theta_k, M_k)\pi(\theta_k, M_k)d\theta_k}{\sum_{j=1}^K P\left(M_j\right)\int f_j\left(Y_n\middle|\theta_j, M_j\right)\pi\left(\theta_j, M_j\right)d\theta_j}. \tag{4.1}$$

Assume the same prior probabilities for all K models, the posterior probability can be rewritten as

$$P(M_k|Y_n) = \frac{f_k(Y_n|M_k)}{\sum_{j=1}^K f_j\left(Y_n|M_j\right)}. \tag{4.2}$$

Therefore, to obtain the posterior probability of model k, we only need to calculate the marginal likelihood for each model and then the corresponding posterior odds, i.e., $f_k(Y_n|M_j)/f_k(Y_n|M_k)$'s. The following three subsections discuss the simplified expressions of the marginal likelihood for the SDEM model with an informative smoothness prior, and those of the SAR the MESS models.

4.1. Marginal likelihood of the SDEM model with an informative smoothness prior

Consider the marginal likelihood of the SDEM model with an informative smoothness prior. Given its priors in Eqs. (3.5) and (3.7), the marginal likelihood of the SDEM model is

$$\begin{aligned} f(Y_n|M_{SDEM}) &= \int \pi\left(\gamma|\sigma^2\right)\pi\left(\sigma^2\right)\pi(\rho)f\left(Y_n|\gamma,\rho,\sigma^2\right)d\gamma d\sigma^2 d\rho \\ &= C^0_{SDEM}\int |R_n(\rho)|\left(\sigma^2\right)^{-\left(\frac{n+(m+1)k+1+a}{2}+1\right)} \\ &\times \exp\left\{-\frac{1}{2\sigma^2}\left[(Y_n(\rho)-Z_n(\rho)\gamma)'(Y_n(\rho)-Z_n(\rho)\gamma)\right]\right\} \\ &+(\gamma-\overline{\gamma})'M_2(\gamma-\overline{\gamma})+b\}d\gamma d\sigma^2 dp \end{aligned} \tag{4.3}$$

with $C^0_{SDEM} = (2\pi)^{-\frac{n+(m+1)k+1}{2}} \times L^{-(m+1)k} \times |R_d| \times p^{-1} \times (2\tau_n)^{-1} \times \frac{\frac{b}{2}^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)}$ and $M_2 = M'M$. Using the properties of the multivariate normal probability density function (pdf) and the inverse gamma pdf, we analytically integrate Eq. (4.3) with respect to γ and σ^2.[18] Let $A_1(\rho) = Z_n'(\rho)Z_n(\rho) + M_2$ and $\tilde{\gamma}(\rho) = A_1(\rho)^{-1'}\left(Z_n'(\rho)Z_n(\rho)\hat{\gamma}(\rho) + M_2\overline{\gamma}\right)$ with $\hat{\gamma}(\rho) = (Z_n'(\rho)Z_n(\rho))^{-1}Z_n'(\rho)Y_n(\rho)$. We can derive the following simplified equation for the marginal likelihood of the SDEM model, which is just an integration with respect to ρ:

$$f(Y_n|M_{SDEM}) = C_{SDEM}\int_\rho |R_n(\rho)||A_1(\rho)|^{-\frac{1}{2}}[Q_1(\rho)+Q_2(\rho)+Q_3(\rho)+b]^{-\frac{n+a}{2}}d\rho \tag{4.4}$$

where

$$\begin{aligned} C_{SDEM} &= (2\pi)^{-\frac{n}{2}} \times L^{-(m+1)k} \times |R_d| \times p^{-1} \times (2\tau_n)^{-1} \times \frac{\frac{b}{2}^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)} \times 2^{\frac{n+a}{2}} \\ &\times \Gamma\left(\frac{n+a}{2}\right) \\ Q_1(\rho) &= (Y_n(\rho)-Z_n(\rho)\hat{\gamma}(\rho))'(Y_n(\rho)-Z_n(\rho)\hat{\gamma}(\rho)), \\ Q_2(\rho) &= \hat{\gamma}(\rho)Z'_n(\rho)Z_n(\rho)\hat{\gamma}(\rho) - \tilde{\gamma}'(\rho)Z'_n(\rho)Z_n(\rho)\tilde{\gamma}(\rho), \\ Q_3(\rho) &= \overline{\gamma}'M_2\overline{\gamma} - \tilde{\gamma}'(\rho)M_2\tilde{\gamma}(\rho). \end{aligned} \tag{4.5}$$

Therefore, we can rely on univariate numerical integration to calculate Eq. (4.4) and the marginal likelihood of the SDEM model can be evaluated by a univariate numerical quadrature.[19]

4.2. Marginal likelihood of the SAR model

For the case of the SAR model, $Y_n = \lambda W_n Y_n + c l_n + X_n\beta + V_n$, let $X_{1n} = (l_n, X_n)$ and $\beta_1 = (c, \beta')'$. Denote $S_n(\lambda) = I_n - \lambda W_n$ and $Y_n(\lambda) = S_n(\lambda)Y_n$. The SAR model can be rewritten as

$$Y_n(\lambda) = X_{1n}\beta_1 + V_n. \tag{4.6}$$

Let $\theta^{sar} = \left(\lambda, \beta_1', \sigma^2\right)'$ represent the parameter of the SAR model. The likelihood function is

$$f(Y_n|\theta^{sar}) = (2\pi)^{-\frac{n}{2}}(\sigma)^{-n}|S_n(\lambda)|\exp\left\{-\frac{(Y_n(\lambda)-X_{1n}\beta_1)'(Y_n(\lambda)-X_{1n}\beta_1)}{2\sigma^2}\right\}. \tag{4.7}$$

We assume the following priors for θ^{sar}[20]:

$$\begin{aligned} &\lambda \sim \mathcal{U}(-\tau_n, \tau_n) \\ &\beta_1|\sigma^2 \sim \mathcal{N}\left(\overline{B}_1, p^2\sigma^2 I_{k+1}\right) \\ &\sigma^2 \sim \mathcal{IG}\left(\frac{a_1}{2}, \frac{b_1}{2}\right). \end{aligned} \tag{4.8}$$

Therefore, the marginal likelihood of the SAR model is

$$\begin{aligned} f(Y_n|M_{SAR}) &= \int \pi\left(\beta_1|\sigma^2\right)\pi\left(\sigma^2\right)\pi(\lambda)f\left(Y_n|\beta_1,\lambda,\sigma^2\right)d\beta_1 d\sigma^2 d\lambda \\ &= C^0_{sar}\int |S_n(\lambda)|\sigma^{-(n+k+1+a_1+2)} \\ &\times \exp\left\{-\frac{1}{2\sigma^2}\left[(Y_n(\lambda)-X_{1n}\beta_1)'(Y_n(\lambda)-X_{1n}\beta_1)\right.\right. \\ &\left.\left.+\frac{1}{p^2}(\beta_1-\overline{\beta}_1)'(\beta_1-\overline{\beta}_1)+b_1\right]\right\}d\beta_1 d\sigma^2 d\lambda \end{aligned} \tag{4.9}$$

where $C^0_{sar} = (2\pi)^{-\frac{n+k+1}{2}} \times p^{-(k+1)} \times (2\tau_n)^{-1} \times \frac{\frac{b_1}{2}^{\frac{a_1}{2}}}{\Gamma\left(\frac{a_1}{2}\right)}$. Using the properties of the multivariate normal pdf and the inverse gamma pdf, we analytically integrate Eq. (4.9) with respect to β_1 and σ^2. Let $\hat{\beta}_1(\lambda) = \left(X_{1n}'X_{1n}\right)^{-1}X_{1n}'Y_n(\lambda)$, $A_2 = X_{1n}'X_{1n} + I_{k+1}/p^2$ and $\tilde{\beta}_1(\lambda) = A_2^{-1}\left(X_{1n}'X_{1n}\hat{\beta}_1(\lambda) + \frac{\overline{\beta}_1}{p^2}\right)$. The marginal likelihood of the SAR model is

$$f(Y_n|M_{SAR}) = C_{sar}\int_\lambda |S_n(\lambda)|[Q_1(\lambda)+Q_2(\lambda)+Q_3(\lambda)+b_1]^{-\frac{n+a_1}{2}}d\lambda \tag{4.10}$$

[18] See Appendix A for more details.

[19] The codes are available http://xiaoyihan.weebly.com/research.html.

[20] Here we assume a normal prior for β_1 rather than the g-prior used in LeSage and Parent (2007) because the focus of the model selection is comparing different patterns of spatial externalities. We might want to assume "similar" priors for β_1 and $\gamma = (c,\delta')'$. The smoothness prior for δ is actually a normal prior. So we also impose a normal prior for β_1 conditional on σ^2.

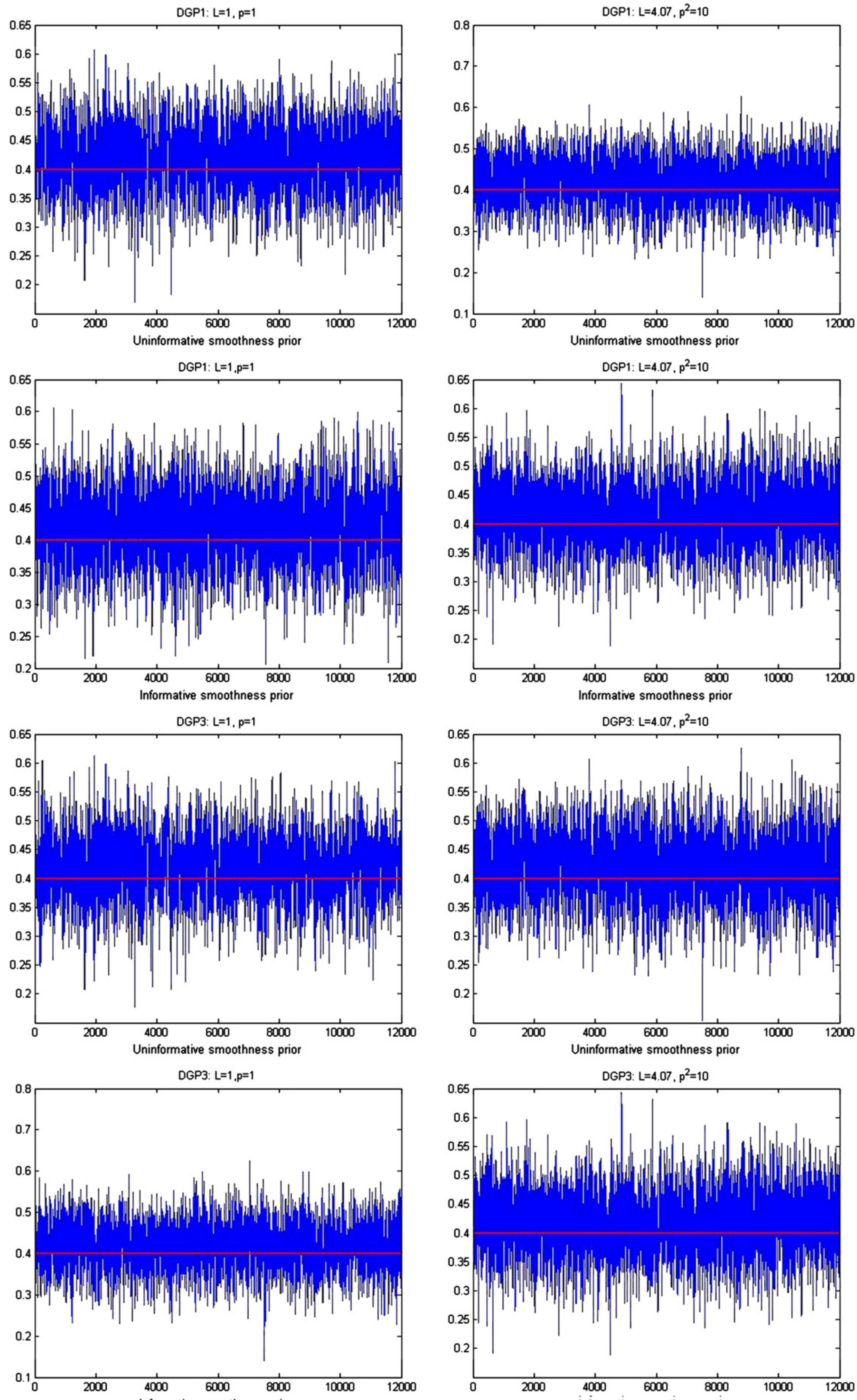

Fig. 2. Trace plots of ρ for SDEM model with 2 lags: DGP1 and DGP3.

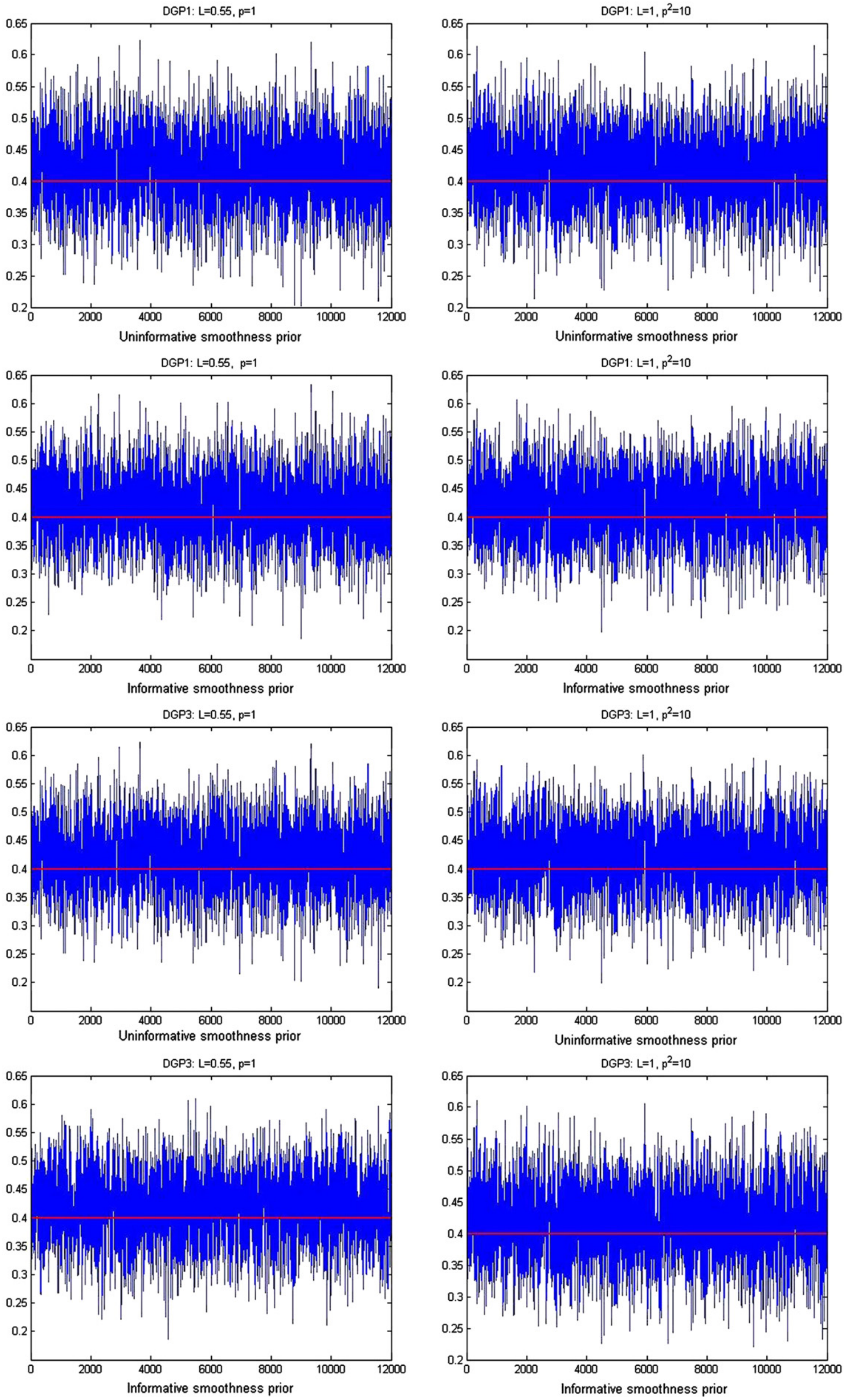

Fig. 3. Trace plots of ρ for SDEM model with 3 lags: DGP1 and DGP3.

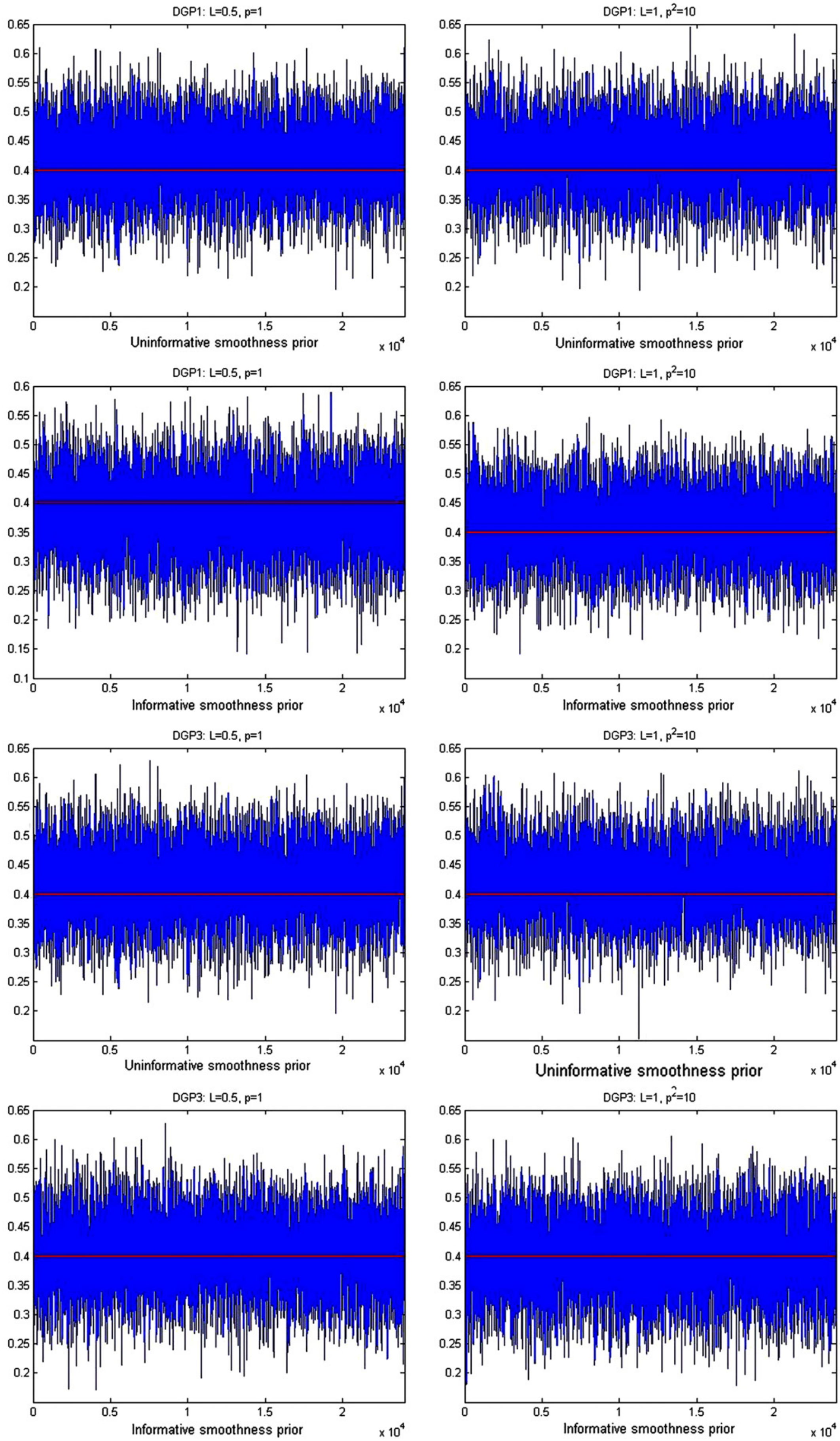

Fig. 4. Trace plots of ρ for SDEM model with 5 lags: DGP1 and DGP3.

Table 1
Estimation of the SDEM model with 2 lags.

		ML		Uninformative				Informative			
				$L=1, p=1$		$L=4.07, p=10$		$L=1, p=1$		$L=4.07, p=10$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
DGP1	ρ	0.3840	0.0609	0.3916	0.0592	0.3886	0.0617	0.3927	0.0557	0.3863	0.0631
	c	1.9999	0.0817	1.9877	0.0802	1.9907	0.0753	1.9787	0.0802	1.9948	0.0827
	β	2.0177	0.0612	2.0130	0.0586	2.0191	0.0635	2.0070	0.0618	2.0201	0.0593
	β_1	1.6291	0.1479	1.6100	0.1286	1.6451	0.1497	1.6023	0.1295	1.6522	0.1442
	β_2	0.9410	0.2572	0.9609	0.2304	0.9087	0.2803	0.9656	0.2449	0.9073	0.2604
	σ^2	0.9882	0.0652	1.0050	0.0645	0.9988	0.0679	1.0157	0.0635	0.9893	0.0623
DGP2	ρ	0.3893	0.0596	0.3922	0.0605	0.3988	0.0586	0.3956	0.0596	0.3977	0.0614
	c	1.9981	0.1213	1.9943	0.1202	2.0024	0.1160	1.9862	0.1211	1.9947	0.1195
	β	1.2636	0.0870	1.2637	0.0744	1.2731	0.0868	1.2503	0.0817	1.2553	0.0904
	β_1	1.0424	0.2102	1.0421	0.1734	1.0496	0.2025	1.0022	0.1810	1.0205	0.1899
	β_2	0.5682	0.3707	0.5913	0.3488	0.5280	0.3725	0.6165	0.3294	0.5885	0.4053
	σ^2	1.9857	0.1319	1.9960	0.1323	1.9824	0.1338	1.9909	0.1323	1.9719	0.1220
DGP3	ρ	0.3847	0.0624	0.3891	0.0613	0.3937	0.0633	0.3942	0.0541	0.3921	0.0625
	c	1.9994	0.0851	1.9856	0.0835	2.0003	0.0816	1.9850	0.0793	2.0005	0.0852
	β	1.6483	0.0644	1.6399	0.0603	1.6515	0.0636	1.6279	0.0548	1.6499	0.0600
	β_1	2.0081	0.1520	1.9764	0.1334	2.0146	0.1586	1.9256	0.1265	2.0071	0.1529
	β_2	1.6859	0.2782	1.7470	0.2454	1.6768	0.2705	1.7396	0.2322	1.6779	0.2693
	σ^2	0.9893	0.0665	1.0043	0.0654	0.9927	0.0688	1.0140	0.0662	0.9937	0.0674

DGP1: $(\rho_0, c_0, \sigma_0^2, \beta_0, \beta_{10}, \beta_{20}) = (0.4, 2, 1, 2.0185, 1.6526, 0.9070)$;
DGP2: $(\rho_0, c_0, \sigma_0^2, \beta_0, \beta_{10}, \beta_{20}) = (0.4, 2, 2, 1.2616, 1.0329, 0.5669)$;
DGP3: $(\rho_0, c_0, \sigma_0^2, \beta_0, \beta_{10}, \beta_{20}) = (0.4, 2, 1, 1.6526, 2.0185, 1.6526)$.

where

$$
\begin{aligned}
C_{sar} &= (2\pi)^{-\frac{n}{2}} \times (2\tau_n)^{-1} \times p^{-(k+1)} \times \frac{\frac{b_1}{2}^{\frac{a_1}{2}}}{\Gamma\left(\frac{a_1}{2}\right)} \times |A|^{-\frac{1}{2}} \times 2^{\frac{n+a_1}{2}} \\
&\quad \times \Gamma\left(\frac{n+a_1}{2}\right), \\
Q_1(\lambda) &= \left(Y_n(\lambda) - X_{1n}\hat{\beta}_1(\lambda)\right)' \left(Y_n(\lambda) - X_{1n}\hat{\beta}_1(\lambda)\right), \\
Q_2(\lambda) &= \hat{\beta}_1(\lambda)' X_{1n}' X_{1n} \hat{\beta}_1(\lambda) - \tilde{\beta}_1(\lambda)' X_{1n}' X_{1n} \tilde{\beta}_1(\lambda), \\
Q_3(\lambda) &= \overline{\beta}'_1 \frac{I_{k+1}}{p^2} \overline{\beta}_1 - \tilde{\beta}_1(\lambda)' \frac{I_{k+1}}{p^2} \tilde{\beta}_1(\lambda).
\end{aligned}
\tag{4.11}
$$

This marginal likelihood of the SAR model can also be evaluated by a univariate numerical integration.

4.3. Marginal likelihood of the MESS model

The MESS model can be written as $Y_n(\mu) = X_{1n}\beta_1 + V_n$ with $Y_n(\mu) = S_n^{ex}(\mu)\, Y_n$, $X_{1n} = (l_n, X_n)$ and $\beta_1 = (c, \beta')'$. Let $\theta^{ex} = \left(\mu, \beta'_1, \sigma^2\right)'$. The likelihood function for the MESS model is

$$
f\left(Y_n | \theta^{ex}\right) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left\{ -\frac{(Y_n(\mu) - X_{1n}\beta_1)'(Y_n(\mu) - X_{1n}\beta_1)}{2\sigma^2} \right\}. \tag{4.12}
$$

Table 2
Estimation of the SDEM model with 3 lags.

		ML		Uninformative				Informative			
				$L=0.55, p=1$		$L=1, p=10$		$L=0.55, p=1$		$L=1, p=10$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
DGP1	ρ	0.3840	0.0692	0.3928	0.0608	0.3896	0.0624	0.3893	0.0610	0.3914	0.0636
	c	1.9934	0.0756	1.9796	0.0741	2.0027	0.0832	1.9881	0.0790	2.0032	0.0793
	β	1.3138	0.0659	1.3110	0.0583	1.3158	0.0554	1.2879	0.0538	1.3072	0.0580
	β_1	1.2792	0.2451	1.2521	0.1441	1.2796	0.1928	1.1407	0.1259	1.2295	0.1769
	β_2	1.1655	0.4143	1.1374	0.1056	1.1334	0.1741	1.1297	0.1085	1.1232	0.1987
	β_3	0.9246	0.6819	1.0017	0.2738	0.9658	0.3933	1.1874	0.2414	1.0487	0.3545
	σ^2	0.9892	0.0645	1.0063	0.0655	0.9951	0.0684	1.0157	0.0617	0.9974	0.0637
DGP2	ρ	0.3837	0.0580	0.3903	0.0580	0.3951	0.0591	0.3914	0.0575	0.3919	0.0542
	c	1.9881	0.1167	1.9802	0.1116	1.9922	0.1066	1.9748	0.1151	1.9961	0.1074
	β	1.0452	0.0830	1.0513	0.0828	1.0665	0.0836	1.0250	0.0719	1.0426	0.0793
	β_1	0.9914	0.3376	1.0335	0.2022	1.0350	0.2631	0.8790	0.1813	0.9702	0.2333
	β_2	0.9434	0.5821	0.9275	0.1652	0.8952	0.2502	0.9041	0.1464	0.8884	0.2461
	β_3	0.7869	0.9495	0.7781	0.3850	0.7132	0.5243	0.9836	0.3583	0.8905	0.4943
	σ^2	1.9827	0.1297	1.9841	0.1371	1.9788	0.1356	2.0032	0.1297	1.9843	0.1405
DGP3	ρ	0.3851	0.0601	0.3923	0.0591	0.3906	0.0573	0.3943	0.0611	0.3981	0.0610
	c	1.9948	0.0811	1.9824	0.0800	1.9976	0.0786	1.9912	0.0799	1.9991	0.0834
	β	1.2574	0.0631	1.2567	0.0558	1.2550	0.0587	1.2281	0.0558	1.2443	0.0577
	β_1	1.3468	0.2310	1.3094	0.1396	1.2989	0.1925	1.1585	0.1315	1.2416	0.1820
	β_2	1.2855	0.4103	1.2409	0.1117	1.2397	0.1701	1.2226	0.0973	1.2344	0.1806
	β_3	1.0232	0.6476	1.1287	0.2600	1.1458	0.3721	1.3441	0.2375	1.2060	0.3785
	σ^2	0.9934	0.0630	1.0094	0.0635	0.9950	0.0652	1.0202	0.0650	0.9959	0.0672

DGP1: $(\rho_0, c_0, \sigma_0^2, \beta_0, \beta_{10}, \beta_{20}, \beta_{30}) = (0.4, 2, 1, 1.3180, 1.3035, 1.1535, 0.9135)$;
DGP2: $(\rho_0, c_0, \sigma_0^2, \beta_0, \beta_{10}, \beta_{20}, \beta_{30}) = (0.4, 2, 2, 1.0544, 1.0428, 0.9228, 0.7308)$;
DGP3: $(\rho_0, c_0, \sigma_0^2, \beta_0, \beta_{10}, \beta_{20}, \beta_{30}) = (0.4, 2, 1, 1.2579, 1.3298, 1.2579, 1.0648)$.

Table 3
Estimation of the SDEM model with 2 lags: fix step size.

		ML		Uninformative				Informative			
				$L=1, p=1$		$L=4.07, p=10$		$L=1, p=1$		$L=4.07, p=10$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
DGP1	ρ	0.3840	0.0609	0.3947	0.0605	0.3946	0.0625	0.3882	0.0580	0.3902	0.0632
	c	1.9999	0.0817	1.9800	0.0811	1.9971	0.0809	1.9874	0.0802	2.0018	0.0781
	β	2.0177	0.0612	2.0169	0.0600	2.0188	0.0643	2.0078	0.0600	2.0178	0.0650
	β_1	1.6291	0.1479	1.6419	0.1331	1.6531	0.1441	1.6000	0.1312	1.6455	0.1485
	β_2	0.9410	0.2572	0.9282	0.2391	0.9066	0.2478	0.9691	0.2642	0.9033	0.2685
	σ^2	0.9882	0.0652	0.9973	0.0683	0.9901	0.0688	1.0097	0.0675	0.9887	0.0662
DGP2	ρ	0.3893	0.0596	0.3920	0.0631	0.3951	0.0638	0.3911	0.0628	0.3953	0.0621
	c	1.9981	0.1213	1.9817	0.1169	1.9974	0.1080	1.9855	0.1175	2.0042	0.1168
	β	1.2636	0.0870	1.2701	0.0831	1.2689	0.0828	1.2544	0.0815	1.2620	0.0865
	β_1	1.0424	0.2102	1.0443	0.1898	1.0452	0.1923	1.0097	0.1723	1.0362	0.2079
	β_2	0.5682	0.3707	0.5947	0.3318	0.5549	0.3660	0.6012	0.3404	0.5836	0.3844
	σ^2	1.9857	0.1319	1.9855	0.1280	1.9794	0.1273	1.9874	0.1259	1.9769	0.1290
DGP3	ρ	0.3847	0.0624	0.3975	0.0648	0.3863	0.0604	0.3876	0.0596	0.3924	0.0597
	c	1.9994	0.0851	1.9876	0.0727	1.9928	0.0808	1.9836	0.0749	1.9915	0.0801
	β	1.6483	0.0644	1.6426	0.0592	1.6525	0.0624	1.6327	0.0620	1.6576	0.0693
	β_1	2.0081	0.1520	1.9842	0.1203	2.0042	0.1521	1.9398	0.1339	2.0184	0.1490
	β_2	1.6859	0.2782	1.7275	0.2359	1.6649	0.2559	1.7611	0.2438	1.6318	0.2997
	σ^2	0.9893	0.0665	1.0006	0.0607	0.9946	0.0692	1.0067	0.0647	0.9811	0.0649

DGP1: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20}) = (0.4,2,1,2.0185,1.6526,0.9070)$;
DGP2: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20}) = (0.4,2,2,1.2616,1.0329,0.5669)$;
DGP3: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20}) = (0.4,2,1,1.6526,2.0185,1.6526)$.

We impose the following priors for θ^{ex}:

$$\begin{aligned}
&\mu \sim \mathcal{N}\left(\overline{\mu}, \xi^2\right)\\
&\beta_1|\sigma^2 \sim \mathcal{N}\left(\overline{\beta}_1, p^2\sigma^2 I_{k+1}\right)\\
&\sigma^2 \sim \mathcal{IG}\left(\frac{a_2}{2}, \frac{b_2}{2}\right).
\end{aligned} \tag{4.13}$$

Hence, the marginal likelihood of the MESS model is

$$\begin{aligned}
f(Y_n|M_{MESS}) &= \int \pi\left(\beta_1|\sigma^2\right)\pi\left(\sigma^2\right)\pi(\mu) f\left(Y_n|\beta_1,\mu,\sigma^2\right) d\beta_1 d\sigma^2 d\mu\\
&= C^0_{mess}\int(\sigma)^{-(n+k+1+a_2+2)}\exp\left(-\frac{(\mu-\overline{\mu})^2}{2\xi^2}\right)\\
&\quad\times\exp\left\{-\frac{1}{2\sigma^2}\left[(Y_n(\mu)-X_{1n}\beta_1)'(Y_n(\mu)-X_{1n}\beta_1)\right.\right.\\
&\quad\left.\left.+\frac{1}{p^2}(\beta_1-\overline{\beta}_1)'(\beta_1-\overline{\beta}_1)\right\}\right] d\beta_1 d\sigma^2 d\mu
\end{aligned} \tag{4.14}$$

Table 4
Estimation of the SDEM model with 3 lags: fix step size.

		ML		Uninformative				Informative			
				$L=0.55, p=1$		$L=1, p=10$		$L=0.55, p=1$		$L=1, p=10$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
DGP1	ρ	0.3840	0.0692	0.3905	0.0622	0.3950	0.0600	0.3889	0.0627	0.3926	0.0636
	c	1.9934	0.0756	1.9826	0.0796	1.9996	0.0812	1.9839	0.0721	2.0016	0.0787
	β	1.3138	0.0659	1.3172	0.0566	1.3193	0.0566	1.2912	0.0535	1.3097	0.0548
	β_1	1.2792	0.2451	1.2683	0.1338	1.2990	0.1834	1.1460	0.1330	1.2295	0.1717
	β_2	1.1655	0.4143	1.1381	0.1039	1.1258	0.1737	1.1178	0.1037	1.1174	0.1812
	β_3	0.9246	0.6819	0.9802	0.2686	0.9229	0.3453	1.1680	0.2622	1.0587	0.3491
	σ^2	0.9892	0.0645	1.0014	0.0686	0.9947	0.0701	1.0269	0.0634	0.9955	0.0602
DGP2	ρ	0.3837	0.0580	0.3944	0.0618	0.3945	0.0660	0.3980	0.0612	0.3945	0.0592
	c	1.9881	0.1167	1.9746	0.1091	2.0132	0.1189	1.9719	0.1140	1.9905	0.1150
	β	1.0452	0.0830	1.0498	0.0704	1.0525	0.0794	1.0300	0.0766	1.0530	0.0809
	β_1	0.9914	0.3376	0.9961	0.1951	1.0469	0.2692	0.9027	0.1940	1.0088	0.2462
	β_2	0.9434	0.5821	0.9118	0.1493	0.9314	0.2440	0.9034	0.1579	0.9008	0.2339
	β_3	0.7869	0.9495	0.8209	0.3695	0.7548	0.5106	0.9368	0.3645	0.8230	0.5124
	σ^2	1.9827	0.1297	1.9908	0.1270	1.9749	0.1296	1.9860	0.1331	1.9667	0.1269
DGP3	ρ	0.3851	0.0601	0.3937	0.0634	0.3927	0.0615	0.3894	0.0591	0.3907	0.0680
	c	1.9948	0.0811	1.9843	0.0804	2.0004	0.0798	1.9864	0.0796	1.9971	0.0815
	β	1.2574	0.0631	1.2596	0.0542	1.2577	0.0566	1.2228	0.0503	1.2452	0.0544
	β_1	1.3468	0.2310	1.3082	0.1421	1.3022	0.1825	1.1557	0.1274	1.2467	0.1786
	β_2	1.2855	0.4103	1.2368	0.1078	1.2202	0.1816	1.2153	0.0995	1.2295	0.1827
	β_3	1.0232	0.6476	1.1201	0.2437	1.1409	0.3651	1.3475	0.2345	1.2218	0.3703
	σ^2	0.9934	0.0630	1.0084	0.0673	0.9967	0.0682	1.0220	0.0666	1.0019	0.0687

DGP1: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30}) = (0.4,2,1,1.3180,1.3035,1.1535,0.9135)$;
DGP2: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30}) = (0.4,2,2,1.0544,1.0428,0.9228,0.7308)$;
DGP3: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30}) = (0.4,2,1,1.2579,1.3298,1.2579,1.0648)$.

Table 5
Estimation of the SDEM model with 5 lags.

		ML		Uninformative				Informative			
				$L=0.5, p=1$		$L=1, p=10$		$L=0.5, p=1$		$L=1, p=10$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
DGP1	ρ	0.3834	0.0624	0.3924	0.0627	0.3903	0.0608	0.3887	0.0584	0.3894	0.0581
	c	1.9971	0.0793	1.9845	0.0787	1.9990	0.0855	1.9806	0.0762	2.0006	0.0779
	β	1.5068	0.0733	1.5098	0.0512	1.5082	0.0602	1.4733	0.0505	1.4991	0.0583
	β_1	1.4435	0.3179	1.4428	0.1388	1.4443	0.1714	1.2892	0.1318	1.4167	0.1554
	β_2	1.2876	1.1222	1.2743	0.1769	1.2752	0.2974	1.2097	0.1707	1.2719	0.2919
	β_3	1.0996	1.7266	1.0423	0.1089	1.0464	0.2275	1.1002	0.1106	1.0839	0.2171
	β_4	0.7239	4.0335	0.7641	0.1473	0.7646	0.1648	0.8838	0.1389	0.7911	0.1685
	β_5	0.4161	3.9624	0.4563	0.0480	0.4679	0.0309	0.6170	0.0259	0.4411	0.0335
	σ^2	0.9814	0.0633	1.0054	0.0653	0.9928	0.0612	1.0393	0.0672	1.0029	0.0671
DGP2	ρ	0.3837	0.0639	0.3937	0.0633	0.3941	0.0639	0.3930	0.0604	0.3901	0.0568
	c	2.0086	0.1200	1.9941	0.1183	1.9972	0.1183	1.9901	0.1095	2.0066	0.1165
	β	1.2480	0.0992	1.2519	0.0736	1.2531	0.0784	1.2247	0.0730	1.2433	0.0797
	β_1	1.2055	0.4553	1.2048	0.2020	1.1968	0.2422	1.0630	0.1898	1.1625	0.2181
	β_2	1.1609	1.3744	1.0763	0.2437	1.0871	0.3853	0.9908	0.2507	1.0740	0.3890
	β_3	0.9579	2.3516	0.8809	0.1509	0.8999	0.3142	0.9090	0.1589	0.9330	0.3018
	β_4	0.3819	5.1613	0.6384	0.2070	0.6388	0.2330	0.7452	0.1950	0.6643	0.2304
	β_5	0.4585	5.1290	0.3932	0.0543	0.3635	0.0813	0.5472	0.0319	0.3318	0.0645
	σ^2	1.9735	0.1330	1.9944	0.1329	1.9758	0.1252	2.0024	0.1279	1.9708	0.1287
DGP3	ρ	0.3825	0.0623	0.3915	0.0633	0.3913	0.0586	0.3911	0.0594	0.3964	0.0589
	c	2.0030	0.0858	1.9915	0.0799	1.9987	0.0883	1.9843	0.0823	2.0097	0.0825
	β	1.4392	0.0737	1.4397	0.0553	1.4380	0.0576	1.4012	0.0497	1.4214	0.0587
	β_1	1.5064	0.3240	1.4884	0.1403	1.5098	0.1595	1.3408	0.1284	1.4450	0.1683
	β_2	1.4572	1.0590	1.4030	0.1665	1.4512	0.2521	1.3544	0.1718	1.4474	0.2734
	β_3	1.2732	1.7736	1.2261	0.1032	1.2590	0.2160	1.2948	0.1132	1.3223	0.2326
	β_4	0.8565	3.9158	0.9836	0.1495	0.9731	0.1571	1.1003	0.1495	1.0179	0.1668
	β_5	0.7145	3.9557	0.6686	0.0804	0.6265	0.0605	0.8733	0.0208	0.6513	0.0349
	σ^2	0.9841	0.0638	1.0051	0.0626	0.9942	0.0669	1.0330	0.0674	0.9961	0.0653

DGP1: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30},\beta_{40},\beta_{50}) = (0.4,2,1,1.5109,1.4662,1.2875,1.0229,0.7354,0.4784)$;
DGP2: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30},\beta_{40},\beta_{50}) = (0.4,2,2,1.2590,1.2218,1.0729,0.8524,0.6128,0.3987)$;
DGP3: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30},\beta_{40},\beta_{50}) = (0.4,2,1,1.4400,1.5139,1.4400,1.2395,0.9653,0.6802)$.

Table 6
Estimation of the SDEM model with 5 lags: fix step size.

		ML		Uninformative				Informative			
				$L=0.5, p=1$		$L=1, p=10$		$L=0.5, p=1$		$L=1, p=10$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
DGP1	ρ	0.3834	0.0624	0.3903	0.0586	0.3913	0.0613	0.3930	0.0567	0.3911	0.0603
	c	1.9971	0.0793	1.9872	0.0799	2.0000	0.0829	1.9863	0.0831	1.9939	0.0856
	β	1.5068	0.0733	1.5039	0.0550	1.5036	0.0559	1.4732	0.0514	1.4963	0.0541
	β_1	1.4435	0.3179	1.4450	0.1406	1.4587	0.1745	1.2839	0.1234	1.4060	0.1488
	β_2	1.2876	1.1222	1.2878	0.1784	1.3219	0.2776	1.1959	0.1661	1.2864	0.2689
	β_3	1.0996	1.7266	1.0494	0.1095	1.0532	0.2134	1.0852	0.1072	1.1053	0.2072
	β_4	0.7239	4.0335	0.7573	0.1457	0.7361	0.1624	0.8772	0.1311	0.7900	0.1589
	β_5	0.4161	3.9624	0.4682	0.0394	0.4078	0.0514	0.6593	0.0341	0.4235	0.0326
	σ^2	0.9814	0.0633	0.9994	0.0646	0.9888	0.0644	1.0318	0.0643	0.9934	0.0638
DGP2	ρ	0.3837	0.0639	0.3924	0.0601	0.3901	0.0596	0.3912	0.0637	0.4031	0.0566
	c	2.0086	0.1200	1.9874	0.1159	2.0015	0.1248	1.9885	0.1137	1.9988	0.1161
	β	1.2480	0.0992	1.2486	0.0780	1.2479	0.0875	1.2243	0.0740	1.2432	0.0782
	β_1	1.2055	0.4553	1.1963	0.1919	1.1902	0.2166	1.0779	0.1758	1.1930	0.2103
	β_2	1.1609	1.3744	1.0669	0.2223	1.0869	0.3499	1.0163	0.2318	1.1405	0.3794
	β_3	0.9579	2.3516	0.8758	0.1455	0.8919	0.3059	0.9143	0.1420	0.9604	0.2911
	β_4	0.3819	5.1613	0.6423	0.2083	0.6573	0.2294	0.7172	0.2024	0.6181	0.2179
	β_5	0.4585	5.1290	0.3887	0.0304	0.4372	0.0558	0.4490	0.0424	0.2562	0.0557
	σ^2	1.9735	0.1330	1.9885	0.1367	1.9663	0.1298	2.0050	0.1288	1.9754	0.1275
DGP3	ρ	0.3825	0.0623	0.3906	0.0613	0.3995	0.0531	0.3837	0.0599	0.3912	0.0624
	c	2.0030	0.0858	1.9820	0.0816	1.9973	0.0837	1.9856	0.0788	2.0061	0.0754
	β	1.4392	0.0737	1.4385	0.0550	1.4397	0.0580	1.4062	0.0532	1.4282	0.0568
	β_1	1.5064	0.3240	1.4890	0.1410	1.4961	0.1669	1.3336	0.1302	1.4376	0.1491
	β_2	1.4572	1.0590	1.4072	0.1672	1.4035	0.2714	1.3375	0.1708	1.4045	0.2939
	β_3	1.2732	1.7736	1.2320	0.1092	1.2239	0.2181	1.2838	0.1075	1.3056	0.2215
	β_4	0.8565	3.9158	0.9944	0.1491	0.9899	0.1619	1.1098	0.1412	1.0269	0.1729
	β_5	0.7145	3.9557	0.7481	0.0428	0.7630	0.0716	0.8773	0.0308	0.6610	0.0776
	σ^2	0.9841	0.0638	1.0037	0.0647	0.9947	0.0646	1.0331	0.0636	0.9944	0.0603

DGP1: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30},\beta_{40},\beta_{50}) = (0.4,2,1,1.5109,1.4662,1.2875,1.0229,0.7354,0.4784)$;
DGP2: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30},\beta_{40},\beta_{50}) = (0.4,2,2,1.2590,1.2218,1.0729,0.8524,0.6128,0.3987)$;
DGP3: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30},\beta_{40},\beta_{50}) = (0.4,2,1,1.4400,1.5139,1.4400,1.2395,0.9653,0.6802)$.

Table 7
Misspecified estimation of the SDEM model with 3 lags.

		Uninformative				Informative			
		$L=1$, $p=1$		$L=4.07$, $p=10$		$L=1$, $p=1$		$L=4.07$, $p=10$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
DGP1	ρ	0.3934	0.0591	0.3943	0.0626	0.3953	0.0588	0.4020	0.0602
	c	1.9864	0.0834	2.0070	0.0790	1.9880	0.0825	1.9970	0.0838
	β	1.3192	0.0580	1.3189	0.0656	1.3098	0.0589	1.3229	0.0643
	β_1	1.5452	0.1309	1.5481	0.1608	1.5152	0.1323	1.5522	0.1536
	β_2	1.5974	0.2414	1.6120	0.2752	1.6564	0.2359	1.5852	0.2768
	β_3	NA	NA	NA	NA	NA	NA	NA	NA
	σ^2	1.0060	0.0632	0.9979	0.0977	1.0050	0.0669	0.9950	0.0958
DGP2	ρ	0.3920	0.0626	0.3904	0.0626	0.4004	0.0604	0.3901	0.0629
	c	1.9844	0.1138	1.9962	0.1222	1.9835	0.1144	2.0101	0.1129
	β	1.0559	0.0808	1.0617	0.0893	1.0445	0.0784	1.0465	0.0850
	β_1	1.2264	0.1901	1.2564	0.2098	1.2114	0.1812	1.2316	0.2017
	β_2	1.2753	0.3370	1.2674	0.3782	1.2873	0.3260	1.2705	0.3890
	β_3	NA	NA	NA	NA	NA	NA	NA	NA
	σ^2	1.9738	0.1300	1.9874	0.1313	1.9907	0.1297	1.9842	0.1281
DGP3	ρ	0.3946	0.0618	0.3952	0.0582	0.3924	0.0600	0.4013	0.0613
	c	1.9846	0.0810	1.9962	0.0806	1.9874	0.0844	2.0004	0.0841
	β	1.2558	0.0609	1.2623	0.0616	1.2435	0.0626	1.2565	0.0632
	β_1	1.6167	0.1340	1.6127	0.1627	1.5679	0.1286	1.6165	0.1582
	β_2	1.8023	0.2502	1.7566	0.2982	1.8279	0.2485	1.7832	0.2864
	β_3	NA	NA	NA	NA	NA	NA	NA	NA
	σ^2	1.0101	0.0739	0.9986	0.0651	1.0135	0.0617	0.9985	0.0687

DGP1: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30}) = (0.4,2,1,1.3180,1.3035,1.1535,0.9135)$;
DGP2: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30}) = (0.4,2,2,1.0544,1.0428,0.9228,0.7308)$;
DGP3: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30}) = (0.4,2,1,1.2579,1.3298,1.2579,1.0648)$.

with $C^0_{mess} = (2\pi)^{-\frac{n+k+1}{2}} \times \left(2\pi\xi^2\right)^{-\frac{1}{2}} \times p^{-(k+1)} \times \frac{\frac{b_2}{2}^{\frac{a_2}{2}}}{\Gamma(\frac{a_2}{2})}$. Using the properties of the multivariate normal pdf and the inverse gamma pdf, we analytically integrate Eq. (4.14) with respect to β_1 and σ^2. Let $\hat{\beta}_1(\mu) = \left(X'_{1n}X_{1n}\right)^{-1}X'_{1n}Y_n(\mu)$, $A_2 = X'_{1n}X_{1n} + I_{k+1} \times p_2^{-2}$ and $\tilde{\beta}_1(\mu) = A_2^{-1}\left(X'_{1n}X_{1n}\hat{\beta}_1(\mu) + \frac{\overline{\beta}_1}{p^2}\right)$. The marginal likelihood for the MESS model is

$$f(Y_n|M_{MESS}) = C_{mess}\int_\mu \exp\left(-\frac{(\mu-\overline{\mu})^2}{2\xi^2}\right)[Q_1(\mu)+Q_2(\mu)+Q_3(\mu)+b_2]^{-\frac{n+a_2}{2}}d\mu \tag{4.15}$$

where

$$\begin{aligned} C_{mess} &= (2\pi)^{-\frac{n}{2}} \times \left(2\pi\xi^2\right)^{-\frac{1}{2}} \times p^{-(k+1)} \times \frac{\frac{b_2}{2}^{\frac{a_2}{2}}}{\Gamma(\frac{a_2}{2})} \times |A_2|^{-\frac{1}{2}} \times 2^{\frac{n+a_2}{2}} \\ &\quad \times\Gamma\left(\frac{n+a_2}{2}\right), \\ Q_1(\mu) &= \left(Y_n(\mu)-X_{1n}\hat{\beta}_1(\mu)\right)'\left(Y_n(\mu)-X_{1n}\hat{\beta}_1(\mu)\right), \\ Q_2(\mu) &= \hat{\beta}_1(\mu)'X'_{1n}X_{1n}\hat{\beta}_1(\mu)-\tilde{\beta}_1(\mu)'X'_{1n}X_{1n}\tilde{\beta}_1(\mu), \\ Q_3(\mu) &= \overline{\beta}'_1\frac{I_{k+1}}{p^2}\overline{\beta}_1-\tilde{\beta}_1(\mu)'\frac{I_{k+1}}{p^2}\tilde{\beta}_1(\mu). \end{aligned} \tag{4.16}$$

Following LeSage and Pace (2007), we evaluate the marginal likelihood in Eq. (4.15) for μ over a range of support $[\mu_{\min}, \mu_{\max}]$, where μ has most of its prior support.[21]

5. Simulation study

5.1. Monte Carlo simulation setup

In this section, we apply the Bayesian estimation algorithm and model selection procedure outlined above to simulated data sets. The study consists of two parts. In the first part, we focus on the performance of the Bayesian estimation of the finite lag SDEM model with smoothness priors. We would like to know how the smoothness prior would affect estimation. In the second part, we utilize the model selection procedure on three data generating processes (DGP): the SDEM model, the SAR model and the MESS model. We want to see whether the model selection procedure could select the right pattern of spatial externalities by assigning the true model with the highest posterior probability. In the experiment, the finite lag SDEM models we consider are the following three cases:

$$\begin{aligned} &2\text{ lags}: Y_n = cl_n + X_n\beta + W_nX_n\beta_1 + W_n^2X_n\beta_2 + (I_n-\rho W_n)^{-1}V_n; \\ &3\text{ lags}: Y_n = cl_n + X_n\beta + W_nX_n\beta_1 + W_n^2X_n\beta_2 + W_n^3X_n\beta_3 + (I_n-\rho W_n)^{-1}V_n \\ &5\text{ lags}: Y_n = cl_n + X_n\beta + W_nX_n\beta_1 + W_n^2X_n\beta_2 + W_n^3X_n\beta_3 + W_n^4X_n\beta_4 \\ &\qquad + W_n^5X_n\beta_5 + (I_n-\rho W_n)^{-1}V_n. \end{aligned}$$

For estimation, we set the number of experiment repetitions to be 300.[22] For each repetition, the total sample size is 450. We consider 3

Table 8
Misspecified estimation of the SDEM model with 3 lags: fix step size.

		Uninformative				Informative			
		$L=1$, $p=1$		$L=4.07$, $p=10$		$L=1$, $p=1$		$L=4.07$, $p=10$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
DGP1	ρ	0.3926	0.0610	0.3940	0.0607	0.3913	0.0572	0.3968	0.0632
	c	1.9849	0.0781	1.9900	0.0813	1.9837	0.0801	2.0001	0.0771
	β	1.3204	0.0591	1.3200	0.0653	1.2887	0.0506	1.325	0.0601
	β_1	1.5463	0.1301	1.5518	0.1439	1.435	0.1057	1.5497	0.1354
	β_2	1.6092	0.2505	1.5847	0.2665	1.6973	0.2010	1.5776	0.2627
	β_3	NA	NA	NA	NA	NA	NA	NA	NA
	σ^2	1.0061	0.0656	1.0023	0.0654	1.0321	0.0675	0.9935	0.0658
DGP2	ρ	0.3900	0.0639	0.3920	0.0627	0.3923	0.0591	0.3952	0.0607
	c	1.9954	0.1135	1.9961	0.1196	1.9906	0.1175	1.9972	0.1264
	β	1.0507	0.0793	1.0536	0.0856	1.0437	0.0818	1.0571	0.0858
	β_1	1.2234	0.1794	1.2431	0.1984	1.2121	0.1913	1.2439	0.2184
	β_2	1.3171	0.3368	1.3033	0.3820	1.3467	0.3481	1.2488	0.3854
	β_3	NA	NA	NA	NA	NA	NA	NA	NA
	σ^2	1.9914	0.1374	1.9793	0.1314	1.9954	0.1213	1.9673	0.1253
DGP3	ρ	0.3969	0.0588	0.4084	0.0571	0.3923	0.0612	0.4017	0.0620
	c	1.9871	0.0826	2.0000	0.0801	1.9836	0.0803	1.9986	0.0837
	β	1.2651	0.0596	1.2574	0.0642	1.2492	0.0582	1.2659	0.0624
	β_1	1.6126	0.1319	1.6072	0.1375	1.5746	0.1357	1.6133	0.1523
	β_2	1.7655	0.2462	1.7818	0.2798	1.8147	0.2428	1.7518	0.2668
	β_3	NA	NA	NA	NA	NA	NA	NA	NA
	σ^2	1.0022	0.0677	0.9972	0.0664	1.0126	0.0660	0.9918	0.0631

DGP1: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30}) = (0.4,2,1,1.3180,1.3035,1.1535,0.9135)$;
DGP2: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30}) = (0.4,2,2,1.0544,1.0428,0.9228,0.7308)$;
DGP3: $(\rho_0,c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30}) = (0.4,2,1,1.2579,1.3298,1.2579,1.0648)$.

[21] As suggested by LeSage and Pace (2007), with a row-normalized spatial weight matrix W_n, $\mu_{\max}$ and $\mu_{\min}$ could be determined by the correspondence $\lambda = 1 - \exp(\mu)$. In our experiment, we set $\lambda_{\max} = 0.9999$ and $\lambda_{\min} = -0.9999$. Then we have $\mu_{\min} = -9.2103$ and $\mu_{\max} = 0.6931$.

[22] LeSage and Pace (2007) and LeSage and Parent (2007) do not include Monte Carlo studies. They only apply their Bayesian estimation method or model selection procedure on some empirical data sets. Other Bayesian econometrics paper, e.g., George et al. (2008), does include some Monte Carlo studies with the number of experiment repetitions being 100.

Table 9
Model frequencies for the SDEM model with 2 lags, the SAR model and the MESS model: small sample.

DGP	Model/prior	$L=1$						$L=4.07$					
		$p=1,\xi=1$			$p=\sqrt{10},\xi=\sqrt{3}$			$p=1,\xi=1$			$p=\sqrt{10},\xi=\sqrt{3}$		
		Low	Mod	High	Low	Mod	High	Low	Mod	High	Low	Mod	High
SDEM1	SDEM with 2 lags	0.57	0.16	0.41	0.75	0.32	0.59	0.46	0.1	0.31	0.63	0.72	0.75
	SAR	0.21	0.41	0.48	0.16	0.39	0.33	0.26	0.45	0.57	0.25	0.16	0.18
	MESS	0.22	0.43	0.11	0.09	0.29	0.08	0.28	0.45	0.12	0.12	0.12	0.07
SDEM2	SDEM with 2 lags	0.62	0.21	0.49	0.84	0.05	0.66	0.12	0.04	0.05	0.28	0.22	0.24
	SAR	0.34	0.55	0.36	0.16	0.3	0.29	0.8	0.69	0.78	0.67	0.72	0.73
	MESS	0.04	0.24	0.15	0	0.05	0.05	0.08	0.27	0.17	0.05	0.06	0.03
SDEM3	SDEM with 2 lags	0.94	0.7	0.72	0.99	0.89	0.92	0.91	0.63	0.08	0.99	0.83	0.27
	SAR	0.04	0.12	0.08	0	0.04	0.02	0.05	0.21	0.48	0.01	0.1	0.37
	MESS	0.02	0.18	0.2	0.01	0.07	0.06	0.04	0.16	0.44	0	0.07	0.36
SAR	SDEM with 2 lags	0.01	0.04	0	0.15	0.19	0	0	0	0	0.01	0.01	0
	SAR	0.86	0.84	0.92	0.84	0.77	0.82	0.85	0.79	0.94	0.97	0.94	0.81
	MESS	0.13	0.12	0.08	0.01	0.04	0.18	0.15	0.21	0.06	0.02	0.05	0.19
MESS	SDEM with 2 lags	0.03	0	0	0.15	0.18	0	0	0	0	0	0.03	0
	SAR	0.91	0.68	0.1	0.83	0.71	0.04	0.88	0.54	0.11	0.98	0.85	0.03
	MESS	0.06	0.32	0.9	0.02	0.11	0.96	0.12	0.46	0.89	0.02	0.12	0.97

Low: $\rho_0=0.2$, $\lambda_0=0.2$ and $\mu_0=-0.0969$;
Mod: $\rho_0=0.4$, $\lambda_0=0.4$ and $\mu_0=-0.5108$;
High: $\rho_0=0.8$, $\lambda_0=0.8$ and $\mu_0=-1.6094$;
SDEM1: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20})=(2,1,2.0185,1.6526,0.9070)$;
SDEM2: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20})=(2,2,1.2616,1.0329,0.5669)$;
SDEM3: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20})=(2,1,1.6526,2.0185,1.6526)$;
SAR: $(c_0,\sigma_0^2,\beta_0)=(2,1,1)$;
MESS: $(c_0,\sigma_0^2,\beta_0)=(2,1,1)$.

DGPs for each model. The DGP1 and DGP2 of the SDEM model with 2 lags are:

$$\begin{aligned}&\rho=0.4;c=2;\sigma^2=1;\\&\tilde{\beta}_i=8\times\frac{\exp\left(-(i-1)^2/5\right)}{\sqrt{5\pi}},i=1,2,3;\\&\beta=\tilde{\beta}_1=2.0185,\beta_1=\tilde{\beta}_2=1.6526,\beta_2=\tilde{\beta}_3=0.9070.\end{aligned}\tag{5.1}$$

$$\begin{aligned}&\rho=0.4;c=2;\sigma^2=2;\\&\tilde{\beta}_i=5\times\frac{\exp\left(-(i-1)^2/5\right)}{\sqrt{5\pi}},i=1,2,3;\\&\beta=\tilde{\beta}_1=1.2616,\beta_1=\tilde{\beta}_2=1.0329,\beta_2=\tilde{\beta}_3=0.5669.\end{aligned}\tag{5.2}$$

And the DGP3 with 2 lags is:

$$\begin{aligned}&\rho=0.4;c=2;\sigma^2=1;\\&\tilde{\beta}_i=8\times\frac{\exp\left(-(i-2)^2/5\right)}{\sqrt{5\pi}},i=1,2,3;\\&\beta=\tilde{\beta}_1=1.6526,\beta_1=\tilde{\beta}_2=2.0185,\beta_2=\tilde{\beta}_3=1.6526.\end{aligned}\tag{5.3}$$

That is, we consider a moderate spatial interaction effect of $\rho=0.4$ for the model. The lag parameter β_is are generated as the ordinates of the normal density. The differences for the three DGPs are on the feature of those β_is and σ^2. For DGP1, the β_is are declining in neither a geometrical pattern nor an exponential pattern. For DGP2, the declining

Table 10
Model frequencies for the SDEM model with 2 lags, the SAR model and the MESS model: large sample.

DGP	Model/prior	$L=1$						$L=4.07$					
		$p=1,\xi=1$			$p=\sqrt{10},\xi=\sqrt{3}$			$p=1,\xi=1$			$p=\sqrt{10},\xi=\sqrt{3}$		
		Low	Mod	High	Low	Mod	High	Low	Mod	High	Low	Mod	High
SDEM1	SDEM with 2 lags	1	1	1	1	1	0.99	1	1	0.99	1	1	1
	SAR	0	0	0	0	0	0.01	0	0	0.01	0	0	0
	MESS	0	0	0	0	0	0	0	0	0	0	0	0
SDEM2	SDEM with 2 lags	1	0.85	0.84	1	0.94	0.93	1	0.52	0.5	1	0.76	0.65
	SAR	0	0.11	0.16	0	0.04	0.07	0	0.3	0.5	0	0.18	0.35
	MESS	0	0.04	0	0	0.02	0	0	0.18	0	0	0.06	0
SDEM3	SDEM with 2 lags	1	1	1	1	1	1	1	1	1	1	1	0.99
	SAR	0	0	0	0	0	0	0	0	0	0	0	0
	MESS	0	0	0	0	0	0	0	0	0	0	0	0.01
SAR	SDEM with 2 lags	0	0.01	0	0.02	0.05	0.02	0	0	0	0	0.03	0
	SAR	1	0.99	1	0.98	0.95	0.98	1	1	1	0.99	0.97	1
	MESS	0	0	0	0	0	0	0	0	0	0.01	0	0
MESS	SDEM with 2 lags	0.03	0	0	0.14	0.05	0	0	0	0	0	0	0
	SAR	0.46	0.97	0	0.48	0.93	0	0.42	0.97	0	0.64	0.98	0
	MESS	0.51	0.03	1	0.38	0.02	1	0.58	0.03	1	0.36	0.02	1

Low: $\rho_0=0.2$, $\lambda_0=0.2$ and $\mu_0=-0.0969$;
Mod: $\rho_0=0.4$, $\lambda_0=0.4$ and $\mu_0=-0.5108$;
High: $\rho_0=0.8$, $\lambda_0=0.8$ and $\mu_0=-1.6094$;
SDEM1: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20})=(2,1,2.0185,1.6526,0.9070)$;
SDEM2: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20})=(2,2,1.2616,1.0329,0.5669)$;
SDEM3: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20})=(2,1,1.6526,2.0185,1.6526)$;
SAR: $(c_0,\sigma_0^2,\beta_0)=(2,1,1)$;
MESS: $(c_0,\sigma_0^2,\beta_0)=(2,1,1)$.

Table 11
Model frequencies for the SDEM model with 3 lags, the SAR model and the MESS model: small sample.

DGP	Model/prior	$L=0.55$						$L=1$					
		$p=1, \xi=1$			$p=\sqrt{10}, \xi=\sqrt{3}$			$p=1, \xi=1$			$p=\sqrt{10}, \xi=\sqrt{3}$		
		Low	Mod	High	Low	Mod	High	Low	Mod	High	Low	Mod	High
SDEM1	SDEM with 3 lags	0.84	0.24	0.02	0.93	0.68	0.11	0.89	0.69	0.1	1	0.92	0.48
	SAR	0.14	0.57	0.68	0.04	0.25	0.59	0.05	0.21	0.67	0	0.05	0.3
	MESS	0.05	0.19	0.3	0.03	0.07	0.3	0.06	0.1	0.23	0	0.03	0.22
SDEM2	SDEM with 3 lags	0.86	0.61	0.30	0.97	0.9	0.77	0.88	0.48	0.11	0.99	0.83	0.77
	SAR	0.08	0.31	0.48	0.03	0.08	0.08	0.08	0.41	0.62	0.01	0.14	0.08
	MESS	0.06	0.08	0.22	0	0.02	0.15	0.04	0.11	0.27	0	0.03	0.15
SDEM3	SDEM with 3 lags	0.83	0.56	0.02	0.99	0.79	0.15	0.98	0.79	0.1	1	0.94	0.53
	SAR	0.11	0.28	0.54	0.01	0.16	0.53	0.01	0.13	0.56	0	0.01	0.21
	MESS	0.06	0.16	0.44	0	0.05	0.32	0.01	0.08	0.34	0	0.05	0.26
SAR	SDEM with 3 lags	0.01	0	0	0.09	0.13	0.01	0.01	0	0	0.11	0.03	0
	SAR	0.84	0.79	0.92	0.89	0.81	0.89	0.89	0.82	0.94	0.86	0.94	0.86
	MESS	0.15	0.21	0.08	0.02	0.06	0.1	0.1	0.18	0.06	0.03	0.03	0.14
MESS	SDEM with 3 lags	0.01	0	0	0.03	0.08	0	0.02	0.01	0	0.06	0.06	0
	SAR	0.88	0.59	0.06	0.94	0.79	0.03	0.86	0.61	0.12	0.92	0.83	0.04
	MESS	0.11	0.41	0.94	0.03	0.13	0.97	0.12	0.38	0.88	0.02	0.11	0.96

Low: $\rho_0=0.2$, $\lambda_0=0.2$ and $\mu_0=-0.0969$;
Mod: $\rho_0=0.4$, $\lambda_0=0.4$ and $\mu_0=-0.5108$;
High: $\rho_0=0.8$, $\lambda_0=0.8$ and $\mu_0=-1.6094$;
SDEM1: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,1,1.3180,1.3035,1.1535,0.9135)$;
SDEM2: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,2,1.0544,1.0428,0.9228,0.7308)$;
SDEM3: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,1,1.2579,1.3298,1.2579,1.0648)$;
SAR: $(c_0,\sigma_0^2,\beta_0)=(2,1,1)$;
MESS: $(c_0,\sigma_0^2,\beta_0)=(2,1,1)$.

pattern of β_is is similar to DGP1 except that the magnitudes of the β_is become smaller. Moreover, for DGP3, the β_is trace out a unimodal shape with the maximum value at β_i. Similarly, the DGP1 and DGP2 of the SDEM model with 3 lags are:

$$\begin{aligned}&\rho=0.4; c=2; \sigma^2=1;\\ &\tilde{\beta}_i=10\times\frac{\exp\left(-(i-1.4)^2/18\right)}{\sqrt{18\pi}}, i=1,2,3,4;\\ &\beta=\tilde{\beta}_1=1.3180, \beta_1=\tilde{\beta}_2=1.3035,\\ &\beta_2=\tilde{\beta}_3=1.1535, \beta_2=\tilde{\beta}_4=0.9135.\end{aligned} \tag{5.4}$$

$$\begin{aligned}&\rho=0.4; c=2; \sigma^2=2;\\ &\tilde{\beta}_i=8\times\frac{\exp\left(-(i-1.4)^2/18\right)}{\sqrt{18\pi}}, i=1,2,3,4; \beta=\tilde{\beta}_1=1.0544,\\ &\beta_1=\tilde{\beta}_2=1.0428, \beta_2=\tilde{\beta}_3=0.9228, \beta_3=\tilde{\beta}_4=0.7308\end{aligned} \tag{5.5}$$

and the DGP3 with 3 lags is

$$\begin{aligned}&\rho=0.4; c=2; \sigma^2=1;\\ &\tilde{\beta}_i=10\times\frac{\exp\left(-(i-2)^2/18\right)}{\sqrt{18\pi}}, i=1,2,3,4; \beta=\tilde{\beta}_1=1.2579,\\ &\beta_1=\tilde{\beta}_2=1.3298, \beta_2=\tilde{\beta}_3=1.2579, \beta_3=\tilde{\beta}_4=1.0648.\end{aligned} \tag{5.6}$$

Table 12
Model frequencies for the SDEM model with 3 lags, the SAR model and the MESS model: large sample.

DGP	Model/prior	$L=0.55$						$L=1$					
		$p=1, \xi=1$			$p=\sqrt{10}, \xi=\sqrt{3}$			$p=1, \xi=1$			$p=\sqrt{10}, \xi=\sqrt{3}$		
		Low	Mod	High	Low	Mod	High	Low	Mod	High	Low	Mod	High
SDEM1	SDEM with 3 lags	1	1	0.73	1	1	0.87	1	1	0.85	1	1	0.96
	SAR	0	0	0.26	0	0	0.13	0	0	0.14	0	0	0.04
	MESS	0	0	0.01	0	0	0	0	0	0.01	0	0	0
SDEM2	SDEM with 3 lags	1	1	0.58	1	1	0.86	1	1	0.51	1	1	0.81
	SAR	0	0	0.42	0	0	0.14	0	0	0.49	0	0	0.18
	MESS	0	0	0	0	0	0	0	0	0	0	0	0.01
SDEM3	SDEM with 3 lags	1	1	0.91	1	1	0.98	1	1	0.99	1	1	0.99
	SAR	0	0	0.07	0	0	0.01	0	0	0.01	0	0	0.01
	MESS	0	0	0.02	0	0	0.01	0	0	0	0	0	0
SAR	SDEM with 3 lags	0	0	0	0.01	0.03	0.06	0	0	0	0.01	0.09	0.08
	SAR	1	1	1	0.98	0.97	0.94	0.99	1	1	0.99	0.91	0.92
	MESS	0	0	0	0.01	0	0	0.01	0	0	0	0	0
MESS	SDEM with 3 lags	0.03	0	0	0.05	0.03	0	0.01	0	0	0.05	0.04	0
	SAR	0.47	0.89	0	0.57	0.91	0	0.51	0.9	0	0.73	0.85	0
	MESS	0.5	0.11	1	0.38	0.06	1	0.48	0.1	1	0.22	0.11	1

Low: $\rho_0=0.2$, $\lambda_0=0.2$ and $\mu_0=-0.0969$;
Mod: $\rho_0=0.4$, $\lambda_0=0.4$ and $\mu_0=-0.5108$;
High: $\rho_0=0.8$, $\lambda_0=0.8$ and $\mu_0=-1.6094$;
SDEM1: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,1,1.3180,1.3035,1.1535,0.9135)$;
SDEM2: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,2,1.0544,1.0428,0.9228,0.7308)$;
SDEM3: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,1,1.2579,1.3298,1.2579,1.0648)$;
SAR: $(c_0,\sigma_0^2,\beta_0)=(2,1,1)$;
MESS: $(c_0,\sigma_0^2,\beta_0)=(2,1,1)$.

Table 13
Misspecified model frequencies for the SDEM model with 3 lags, the SAR model and the MESS model: small sample.

DGP	Model/prior	$L=1$						$L=4.07$					
		$p=1,\xi=1$			$p=\sqrt{10},\xi=\sqrt{3}$			$p=1,\xi=1$			$p=\sqrt{10},\xi=\sqrt{3}$		
		Low	Mod	High	Low	Mod	High	Low	Mod	High	Low	Mod	High
SDEM1	SDEM with 2 lags	0.91	0.65	0.14	0.98	0.96	0.48	0.68	0.36	0.02	0.94	0.66	0.11
	SAR	0.08	0.21	0.48	0	0.03	0.27	0.24	0.42	0.75	0.05	0.25	0.59
	MESS	0.01	0.14	0.38	0.02	0.01	0.25	0.08	0.22	0.23	0.01	0.09	0.3
SDEM2	SDEM with 2 lags	0.89	0.46	0.37	0.98	0.95	0.88	0.48	0.14	0.03	0.62	0.33	0.12
	SAR	0.07	0.44	0.44	0.01	0.05	0.07	0.48	0.72	0.74	0.37	0.63	0.59
	MESS	0.04	0.1	0.19	0.01	0	0.05	0.04	0.14	0.23	0.01	0.04	0.29
SDEM3	SDEM with 2 lags	0.93	0.8	0.12	1	0.95	0.43	0.94	0.65	0.04	1	0.97	0.47
	SAR	0.05	0.14	0.48	0	0.03	0.35	0.04	0.23	0.62	0	0.03	0.33
	MESS	0.02	0.06	0.4	0	0.02	0.22	0.02	0.12	0.34	0	0	0.2

Low: $\rho_0=0.2$, $\lambda_0=0.2$ and $\mu_0=-0.0969$;
Mod: $\rho_0=0.4$, $\lambda_0=0.4$ and $\mu_0=-0.5108$;
High: $\rho_0=0.8$, $\lambda_0=0.8$ and $\mu_0=-1.6094$;
SDEM1: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,1,1.3180,1.3035,1.1535,0.9135)$;
SDEM2: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,2,1.0544,1.0428,0.9228,0.7308)$;
SDEM3: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,1,1.2579,1.3298,1.2579,1.0648)$;

Table 14
Misspecified model frequencies for the SDEM model with 3 lags, the SAR model and the MESS model: large sample.

DGP	Model/prior	$L=1$						$L=4.07$					
		$p=1,\xi=1$			$p=\sqrt{10},\xi=\sqrt{3}$			$p=1,\xi=1$			$p=\sqrt{10},\xi=\sqrt{3}$		
		Low	Mod	High	Low	Mod	High	Low	Mod	High	Low	Mod	High
SDEM1	SDEM with 2 lags	1	1	0.75	1	1	0.78	1	1	0.57	1	1	0.79
	SAR	0	0	0.24	0	0	0.21	0	0	0.41	0	0	0.21
	MESS	0	0	0.01	0	0	0.01	0	0	0.02	0	0	0
SDEM2	SDEM with 2 lags	1	1	0.44	1	1	0.62	1	1	0.18	1	1	0.35
	SAR	0	0	0.52	0	0	0.38	0	0	0.81	0	0	0.64
	MESS	0	0	0.04	0	0	0	0	0	0.01	0	0	0.01
SDEM3	SDEM with 2 lags	1	1	0.87	1	1	0.94	1	1	0.72	1	1	0.84
	SAR	0	0	0.12	0	0	0.05	0	0	0.26	0	0	0.14
	MESS	0	0	0.01	0	0	0.01	0	0	0.02	0	0	0.02

Low: $\rho_0=0.2$, $\lambda_0=0.2$ and $\mu_0=-0.0969$;
Mod: $\rho_0=0.4$, $\lambda_0=0.4$ and $\mu_0=-0.5108$;
High: $\rho_0=0.8$, $\lambda_0=0.8$ and $\mu_0=-1.6094$;
SDEM1: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,1,1.3180,1.3035,1.1535,0.9135)$;
SDEM2: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,2,1.0544,1.0428,0.9228,0.7308)$;
SDEM3: $(c_0,\sigma_0^2,\beta_0,\beta_{10},\beta_{20},\beta_{30})=(2,1,1.2579,1.3298,1.2579,1.0648)$;

Finally, the DGP1 and DGP2 of the SDEM model with 5 lags are

$$\begin{aligned}&\rho=0.4;c=2;\sigma^2=1;\\&\tilde{\beta}_i=12\times\frac{\exp\left(-(i-1.2)^2/20\right)}{\sqrt{20\pi}},i=1,2,3,4,5,6;\\&\beta=\tilde{\beta}_1=1.5109,\beta_1=\tilde{\beta}_2=1.4662,\beta_2=\tilde{\beta}_3=1.2875,\\&\beta_3=\tilde{\beta}_4=1.0229,\beta_4=\tilde{\beta}_5=0.7354,\beta_5=\tilde{\beta}_6=0.4784.\end{aligned}\tag{5.7}$$

$$\begin{aligned}&\rho=0.4;c=2;\sigma^2=2;\\&\tilde{\beta}_i=10\times\frac{\exp\left(-(i-1.2)^2/20\right)}{\sqrt{20\pi}},i=1,2,3,4,5,6;\\&\beta=\tilde{\beta}_1=1.2590,\beta_1=\tilde{\beta}_2=1.2218,\beta_2=\tilde{\beta}_3=1.0729,\\&\beta_3=\tilde{\beta}_4=0.8524,\beta_4=\tilde{\beta}_5=0.6128,\beta_5=\tilde{\beta}_6=0.3987.\end{aligned}\tag{5.8}$$

The DGP3 is

$$\begin{aligned}&\rho=0.4;c=2;\sigma^2=1;\\&\tilde{\beta}_i=12\times\frac{\exp\left(-(i-1.2)^2/20\right)}{\sqrt{20\pi}},i=1,2,3,4,5,6;\\&\beta=\tilde{\beta}_1=1.4400,\beta_1=\tilde{\beta}_2=1.5139,\beta_2=\tilde{\beta}_3=1.4400,\\&\beta_3=\tilde{\beta}_4=1.2395,\beta_4=\tilde{\beta}_5=0.9653,\beta_5=\tilde{\beta}_6=0.6802.\end{aligned}\tag{5.9}$$

The bar charts of the DGP1 and DGP3 for the β_is of the SDEM model with 2, 3 and 5 lags are depicted in Fig. 1. The error terms v_{ni}s of the SDEM model are generated as i.i.d sample from the normal distribution with different variances. For the DGP1 and DGP3, the variance of v_{ni}s is 1; while for DGP2, the variance is set to be 2. As a result, the variation in the error terms is relatively larger in DGP2 than in DGP1 and DGP3. The exogenous regressors X_n are generated from the standard normal distribution. The spatial weight matrix W_n is constructed by the function "makeneighborsw",[23] which generates a row-normalized spatial weight matrix based on m nearest neighbors. In the experiment, m is set to be 5.

We consider both the uninformative and informative smoothness prior for the SDEM model. The values of the prior parameters are:

$$\begin{aligned}&\text{SDEM with 2 lags}:\bar{c}=0,p=1\text{ or }10;L=1\text{ or }4.07;\\&\text{SDEM with 3 lags}:\bar{c}=0,p=1\text{ or }10;L=0.55\text{ or }1;\\&\text{SDEM with 5 lags}:\bar{c}=0,p=1\text{ or }10;L=0.5\text{ or }1;\\&a=6,b=4;d=1.\end{aligned}$$

With $d=1$, we only consider second order difference for the lag coefficient vector δ. Two settings of the prior parameters are used for c and δ. Specifically, for the model with 2 lags, $p=1$ or 10 is used, respectively reflecting a tight and moderate normal prior for

[23] This function is taken from LeSage's Matlab code for spatial econometrics, which can be found at http://www.spatial-econometrics.com/.

Table 15
Descriptive statistics for variables (N = 427).

Variable	Mean	Maximum	Minimum	Standard deviation
Manufacturing employment as a percentage of private wage and salary employment	18.44	46.18	0.717	11.47
Percentage of population aged 18–64	58.65	79	48.5	4.12
Percentage of population who are female	50.33	55.6	37.2	1.74
Percentage of population nonwhite	11.24	83.6	0.3	12.39
Percentage of population age 25 or above who are high school graduates or higher	77.81	95.3	46.1	8.92
Percentage of population aged 25 or above with a bachelor's degree or higher	15.95	60.2	5.4	7.38
Right to work dummy variable	0.52	1	0	0.5
Small business survival index	41.29	52.15	24.88	7.02

c; $L = 1$ or 4.07 is used, respectively reflecting a tight and moderate smoothness prior for δ. For the model with 3 and 5 lags, a tight and moderate normal prior for c with $p = 1$ or 10 is still used, but we change the tight and moderate smoothness prior to $L = 0.55$ or 1 and $L = 0.5$ or 1, respectively. The estimation procedures we considered are: the ML estimation and the Bayesian estimation with informative and uninformative smoothness priors. For Bayesian estimation, we run Markov chains of different lengths for different models. For models with 2 and 3 lags, the Markov chain length is 15,000 in each repetition except that for DGP2, the chain length increases to 30,000.[24] For model with 5 lags, the Markov chain length is 30,000 for all DGPs.[25] We burn in the first 20% draws[26] of each chain and collect every *10*th draws as final draws.[27] Some selected trace plots of the Bayesian estimates of ρ in the Monte Carlo study are depicted in Figs. 2–4, to demonstrate the convergence of the MCMC sampler.

Lastly, besides estimation of the above three models, we also check the performance of misspecified Bayesian estimator for the SDEM model with 3 lags. We try the Bayesian estimation for a model with 2 lags when the DGP is the SDEM model with 3 lags. We want to see how a misspecification can affect the Bayesian estimates.

Furthermore, for model selection, the number of experiment repetition is reduced to 100.[28] We consider both a small sample size of 100 and a large sample size of 1200 in each repetition. We report the frequency in which the true model is assigned the highest posterior probability. The DGPs are the SDEM models with 2 or 3 lags, the SAR model and the MESS model. Specifically, the DGPs of the SDEM model are the ones in Eqs. (5.1)–(5.6), except that we consider three different values of ρ: $\rho = 0.2$, $\rho = 0.4$ and $\rho = 0.8$. The $\rho = 0.2$ corresponds to a weak spatial interaction effect, $\rho = 0.4$ corresponds to a moderate spatial interaction effect and $\rho = 0.8$ corresponds to a strong spatial interaction effect. The DGP for the SAR model is:

$$\lambda = 0.2 \ or \ 0.4 \ \text{or} \ 0.8, c = 2, \beta = 1, \sigma^2 = 1. \tag{5.10}$$

The DGP for the MESS model is:

$$\mu = -0.0969 \ or \ -0.5108 \ or \ -1.6094,^{29} c = 2, \ \beta = 1, \ \sigma^2 = 1. \tag{5.11}$$

The priors of the parameters of the SAR model are specified as Eq. (4.8), and the values of the prior parameters are

$$p = 1 \ or \ \sqrt{10}; \ a_1 = 6, \quad b_1 = 4; \quad \overline{\beta}_1 = O_{k+1}.^{30} \tag{5.12}$$

The priors of the parameters of the MESS model are the same as Eq. (4.13) and the values of prior parameters are

$$p = 1 \ or \ \sqrt{10}; \ a_2 = 6, \quad b_2 = 4; \quad \overline{\beta}_1 = O_{k+1}; \quad \overline{\mu} = 0, \quad \xi = 1 \quad or \quad \sqrt{3}. \tag{5.13}$$

Specifically, $L = 1$ or 4.07, or $L = 0.55$ or 1, is used for δ, representing either a tight or moderate informative smoothness prior. The $p = 1$ or $\sqrt{10}$ is used to reflect a tight and moderate normal priors with mean 0 for β_1.[31] $\xi = 1$ *or* $\sqrt{3}$ is applied to the normal prior for μ, reflecting a continuum from tight to moderate. Finally, we evaluate the marginal likelihood for each model based on the simplified expressions in Section 4. Then posterior probabilities are computed and compared. Lastly, we investigate a misspecified SDEM model for model comparison. We evaluate the posterior probability of the SDEM model with 2 lags when the DGP is the model with 3 lags. We want to see whether the SDEM model, even misspecified, can explain some features of the true DGP or not, compared with the SAR and the MESS models.

5.2. Results

Tables 1–2 summarize estimation results for the SDEM model with 2 and 3 lags. For model with 2 lags, ML estimates and Bayesian estimates of the spatial parameter ρ, the coefficients c and β are similar for all DGPs. Estimates of lag coefficients β_1 and β_2 are also similar for the ML approach and the Bayesian approach, except that the Bayesian estimates have a slightly smaller standard deviation in most cases. For model with 3 lags, ML estimates and the Bayesian estimates of ρ, c and β are still similar. However, estimates of the high order lag coefficients β_2 and β_3 differ for the two approaches. For all DGPs, the standard deviations of Bayesian estimates of β_2 and β_3 are smaller than those of ML estimates. Furthermore, for both informative and uninformative smoothness priors, the smaller L is, the smaller

[24] We apply the method in Raftery and Lewis (1992) to decide the adequate length of our MCMC sampler. Raftery and Lewis provide an answer to the adequate length of the chain of draws based on the accuracy of the posterior summaries desired by the user. Based on their method, the maximum number of draws needed for the SDEM models with 2 and 3 lags is around 4000 for all DGPs. Here we increase the length of our sampler to 15,000 or 30,000 to ensure the convergence of the chains.

[25] With more lags in the model, the autocorrelation of ρ in the sampling process becomes higher. Based on Raftery and Lewis' method, the maximum number of draws needed for the SDEM model with 5 lags is around 7000. So we increase the length of Markov chains to make sure the chains eventually converge.

[26] The estimation results change little when we burn in the first 50% draws.

[27] As suggested by a referee, given that the chains are well behaved as exhibited by the trace plots, thinning 10% might be too much and unnecessary. We also double check some estimation results and redraw some trace plots of the SDEM model without thinning the chains. The estimation results and the trace plots turn out to be similar to the previous ones.

[28] We have increased the repetition number to 400 but the model frequencies change little.

[29] LeSage and Pace (2007) provide a correspondence between μ and λ, which is $\lambda = 1 - \exp(\mu)$. So $\mu = -0.0969$ corresponds to $\lambda = 0.0924$, which refers to a weak interaction effect. $\mu = -0.5108$ corresponds to $\lambda = 0.4$, which refers to a moderate interaction effect. $\mu = -1.6094$ corresponds to $\lambda = 0.8$, which refers to a strong interaction effect.

[30] Here $k = 1$.

[31] We also set the prior parameter p of the intercept coefficient c in the SDEM model to be $\sqrt{10}$.

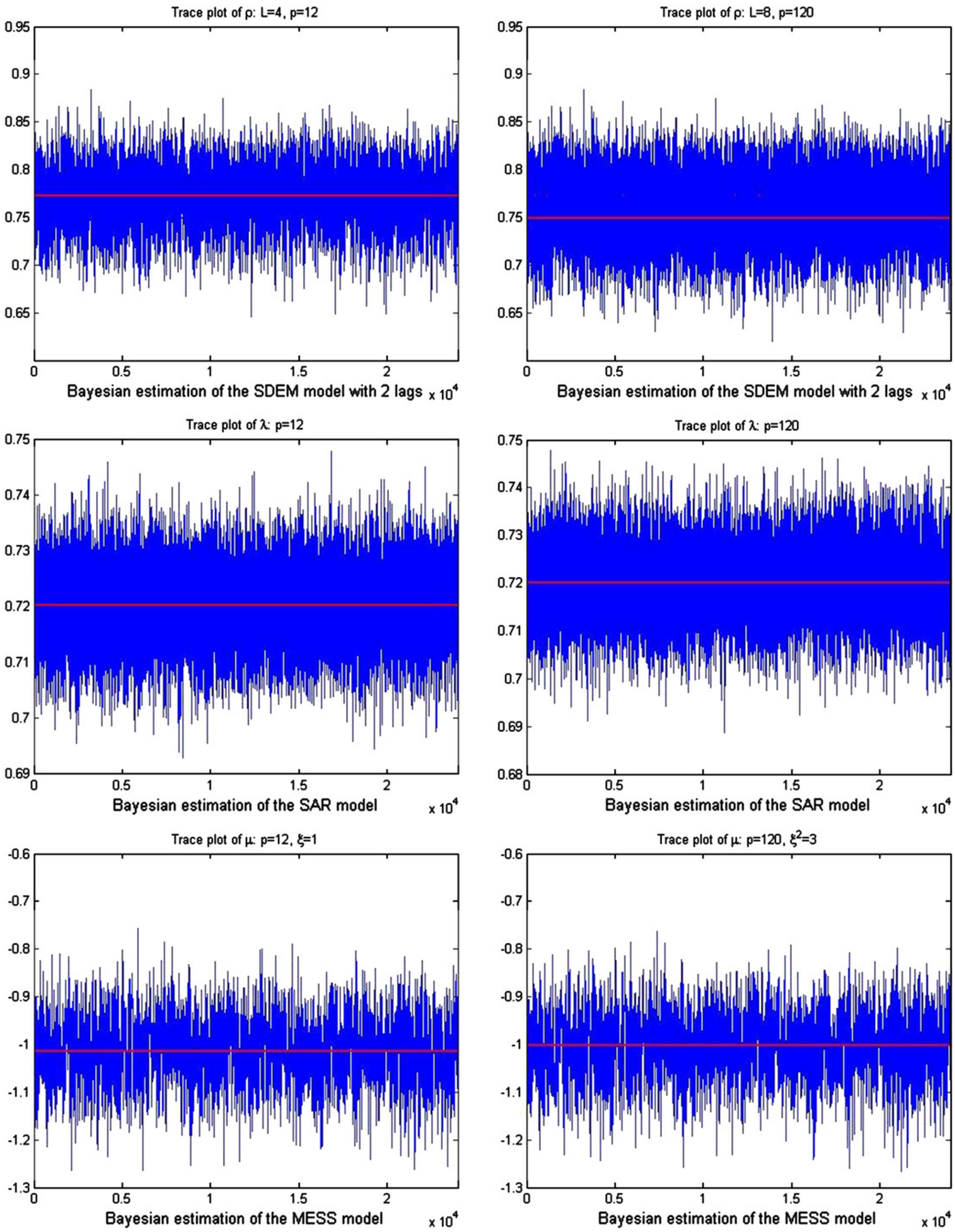

Fig. 5. Trace plots of ρ, λ and μ in empirical application.

the standard deviations of most Bayesian estimates of β_is are. This suggests that a tighter smoothness prior might provide us with Bayesian estimates with smaller standard deviations. Finally, Tables 3–4 provide estimation results for the SDEM model with 2 and 3 lags when the step size h_ρ is fixed in the M–H algorithm. The results are similar to that in Tables 1 and 2.

Table 5 gives the estimation results for the SDEM model with 5 lags. Estimates for ρ, c and β are still similar for the two approaches. But the standard deviations of MLEs for the β_is are larger than those of Bayesian estimates, especially for β_4 and β_5. In particular, for DGP2, most Bayesian estimates of β_4 are closer to the true value with smaller standard deviations, compared with the ML estimates. Table 6 provides estimation results for the SDEM model with 5 lags when h_ρ is fixed in the M–H algorithm. The results are similar to that in Table 5.

Table 7 provides the misspecified Bayesian estimation results of the SDEM model with 3 lags, in which we estimate the model with 2 lags but the DGP is the model with 3 lags. The Bayesian estimates for ρ, c and β are still fine. But there are larger biases for estimates of β_1 and β_2, compared with the Bayesian estimates for the model with 3 lags in Table 2. Table 8 summarizes the misspecified Bayesian estimation results of the SDEM model with 3 lags, when h_ρ is fixed in the M–H algorithm. The results are similar to that in Table 7.

Table 9 summarizes model frequencies for the SDEM model with 2 lags, the SAR model and the MESS model when the sample size is 100. When the DGPs are the SDEM model, the Bayesian routines are not so

Table 16
Selected ML estimation results for SDEM, SAR and MESS models.

Independent variable	SDEM	SAR	MESS
Constant	−327.3961 (−3.99)***	−41.4515 (−2.74)***	−53.4244 (−3.24)***
PCT of population aged 18–64	0.3934 (1.93)*	0.3041 (2.96)***	0.3968 (3.56)***
PCT of population female	0.6169 (1.92)*	0.6148 (3.09)***	0.7904 (3.66)***
PCT of population nonwhite	−0.050 (−1.04)	−0.0563 (−2.00)**	−0.0686 (−2.25)**
PCT of population age 25+, high school graduates	−0.3468 (−3.16)***	−0.0469 (−0.86)	−0.0530 (−0.89)
PCT of population aged 25+, with a bachelor's degree or higher	−0.1660 (−1.61)	−0.2805 (−3.95)***	−0.3411 (−4.43)***
RTW dummy variable	2.7405 (3.90)***	2.6552 (4.10)***	2.8999 (4.10)***
SBSI	0.1459 (2.47)**	0.1273 (2.66)***	0.1459 (2.79)***
$W_n \times$ PCT of population aged 25+ with a bachelor's degree or higher	−0.1795 (−0.70)	NA	NA
$W_n^2 \times$ PCT of population aged 25+ with a bachelor's degree or higher	−0.9593 (−2.13)**	NA	NA
$W_n \times$ RTW dummy	−3.3675 (−1.29)	NA	NA
$W_n^2 \times$ RTW dummy	2.6261 (0.43)	NA	NA
ρ	0.7260 (21.67)***		
λ		0.7200 (23.01)***	
μ			−0.9983 (−13.36)***

PCT: percentage.
SBSI: small business survival index.
t-Statistics are in parentheses.
* Significant at the 10% level.
** Significant at 5% level.
*** Significant at 1% level.

Table 17
Bayesian estimation of the SDEM model with 2 lags for empirical data.

Independent variable	$L=4, p=12$			$L=4, p=120$		
	Mean	S.D.	95% CI	Mean	S.D.	95% CI
Constant	−112.0046	59.5559	(−229.1331, 3.9758)	−310.4615	88.1479	(−479.4242, −133.9901)
PCT of population aged 18–64	0.4193	0.2015	(0.0235, 0.8108)	0.4004	0.2024	(0.0073, 0.8001)
PCT of population female	0.6583	0.3181	(0.0254, 1.2858)	0.6294	0.3179	(0.0124, 1.2542)
PCT of population nonwhite	−0.0392	0.0478	(−0.1328, 0.0545)	−0.0479	0.0479	(−0.1431, 0.0464)
PCT of population aged 25+, high school graduates	−0.3386	0.1090	(−0.5528, −0.1228)	−0.3467	0.1104	(−0.5650, −0.1308)
PCT of population aged 25+, with a bachelor's degree or higher	−0.1819	0.1028	(−0.3815, 0.0214)	−0.1704	0.1027	(−0.3728, 0.0318)
RTW dummy	2.8051	0.6917	(1.4370, 4.1578)	2.7643	0.6942	(1.4034, 4.1196)
SBSI	0.1531	0.0588	(0.0373, 0.2689)	0.1480	0.0590	(0.0323, 0.2634)
$W_n \times$ PCT of population aged 25+ with a bachelor's degree or higher	−0.1181	0.2551	(−0.6193, 0.3763)	−0.1847	0.2555	(−0.6787, 0.3191)
$W_n^2 \times$ PCT of population aged 25+ with a bachelor's degree or higher	−0.5632	0.4612	(−1.4669, 0.3443)	−0.9111	0.4604	(−1.8115, −0.0037)
$W_n \times$ RTW dummy	−3.2090	2.5930	(−8.3171, 1.8389)	−3.2259	2.6062	(−8.3054, 1.9086)
$W_n^2 \times$ RTW dummy	0.4815	0.3597	(−11.9032, 12.9748)	2.6135	6.3111	(−9.6535, 15.1801)
ρ	0.7720	0.0319	(0.7072, 0.0319)	0.7493	0.0334	(0.6822, 0.8138)

CI: credible interval; PCT: percentage; SBSI: small business survival index.
Number of iterations is 30,000. We burn in the first 20% draws.

good at pointing to the proper model in some cases. This is so, in particular for DGP2, in which we have a larger variation in the error terms. With moderate or high spatial dependence the SAR model can capture some features of the SDEM DGP. For DGP3, in which we have a unimodal shape for the lag coefficients, the SDEM model has the largest posterior probability in most cases. When the DGP is the SAR model, the frequency of the SAR model is the largest in all cases. Finally, if the DGP is the MESS model, the SAR model can capture some features of the MESS model in cases of low and moderate spatial dependence. But as interaction becomes stronger, the MESS model tends to have the largest frequency.[32]

Table 10 summarizes model frequencies for the SDEM model with 2 lags, the SAR model and the MESS model when the sample size increases to 1200. For the SDEM model, the Bayesian routines work much better than the small sample setting. The frequency of the SDEM model with 2 lags is the largest in most cases for all DGPs. The only exception is that, for DGP2, with moderate or high spatial dependence and a tight normal prior for β_1, the SAR model is able to capture some features of the SDEM DGP. When the DGP is the SAR model, the frequency of the SAR model is the largest in all cases. Moreover, the SAR model is still able to capture some features of the MESS DGP unless we have strong spatial interaction effect.

Table 11 provides model frequencies for the SDEM model with 3 lags, the SAR model and the MESS model when the sample size is 100. Note that the DGPs of the SAR model and the MESS model are the same as in Table 9. But the Ls used are different from those in Table 9. For the SDEM model with 3 lags, the Bayesian routines are better at pointing out the proper model, compared with the small sample size setting of the SDEM model with 2 lags. However, with high spatial dependence and a tight normal prior for β_1, the SAR model can capture some features of the SDEM model. When the DGP is the SAR model, the frequency of the SAR model is still largest in all cases. If the DGP is the MESS model, the SAR model is a good substitute of the MESS model unless we have strong spatial dependence.

Table 12 provides model frequencies for the SDEM model with 3 lags, the SAR model and the MESS model when the sample size increases to 1200. The SDEM model still has the highest frequencies for all DGPs from the SDEM model. When the DGP is the SAR model, the frequency of the SAR model is the largest in all cases. If the DGP is the MESS model, the SAR model is a good substitute of the MESS model unless we have strong spatial dependence.

Table 13 summarizes the model frequencies for the misspecified SDEM model, the SAR model and the MESS model when the sample

[32] This is consistent with Han and Lee (2013). In their Monte Carlo studies they find that with strong spatial dependence the J-test statistics can have good power.

Table 18
Bayesian estimation of the SDEM model with 2 lags for empirical data: fix step size.

	$L=4, p=12$			$L=4, p=120$		
Independent variable	Mean	S.D.	95% CI	Mean	S.D.	95% CI
Constant	−113.7841	59.2085	(−228.9005, 3.9486)	−310.5735	88.6935	(−481.8439, −134.3320)
PCT of population aged 18–64	0.4185	0.2014	(0.0216, 0.8112)	0.3992	0.2000	(0.0044, 0.7909)
PCT of population female	0.6556	0.3180	(0.0274, 1.2765)	0.6270	0.3161	(0.0112, 1.2455)
PCT of population nonwhite	−0.0401	0.0481	(−0.1343, 0.0548)	−0.0472	0.0482	(−0.1418, 0.0473)
PCT of population aged 25+, high school graduates	−0.3392	0.1085	(−0.5514, −0.1271)	−0.3455	0.1094	(−0.5598, −0.1287)
PCT of population aged 25+, with a bachelor's degree or higher	−0.1804	0.1031	(−0.3842, 0.0205)	−0.1706	0.1029	(−0.3733, 0.0299)
RTW dummy	2.7959	0.6861	(1.4406, 4.1434)	2.7604	0.6967	(1.3988, 4.1249)
SBSI	0.1522	0.0590	(0.0359, 0.2690)	0.1473	0.0587	(0.0320, 0.2611)
$W_n \times$ PCT of population aged 25+ with a bachelor's degree or higher	−0.1162	0.2557	(−0.6143, 0.3853)	−0.1861	0.2574	(−0.6885, 0.3185)
$W_n^2 \times$ PCT of population aged 25+ with a bachelor's degree or higher	−0.5639	0.4609	(−1.4611, 0.3440)	−0.9146	0.4668	(−1.8197, 0.0052)
$W_n \times$ RTW dummy	−3.2619	2.5900	(−8.3418, 1.8256)	−3.2312	2.6049	(−8.3433, 1.8943)
$W_n^2 \times$ RTW dummy	0.4262	6.3598	(−12.0111, 12.9401)	2.5585	6.3100	(−9.7479, 15.0344)
ρ	0.7712	0.0318	(0.7056, 0.8314)	0.7500	0.0344	(0.6793, 0.8149)

CI: credible interval; PCT: percentage; SBSI: small business survival index.
Number of iterations is 30,000. We burn in the first 20% draws.

Table 19
Bayesian estimation of the SAR model for empirical data.

	$p=12$			$p=120$		
Independent variable	Mean	S.D.	95% CI	Mean	S.D.	95% CI
Constant	−27.9891	12.4024	(−52.6765, −3.6341)	−39.4747	14.7558	(−68.7888, −11.0081)
PCT of population aged 18–64	0.2371	0.0908	(0.0588, 0.4181)	0.2940	0.0997	(0.0989, 0.4925)
PCT of population female	0.4623	0.1705	(0.1245, 0.7996)	0.5927	0.1928	(0.2176, 0.9771)
PCT of population nonwhite	−0.0640	0.0271	(−0.1162, −0.0111)	−0.0574	0.0273	(−0.1111, −0.0029)
PCT of population aged 25+, high school graduates	−0.0717	0.0516	(−0.3800, −0.1177)	−0.0503	0.0534	(−0.4126, −0.1414)
PCT of population aged 25+, with a bachelor's degree or higher	−0.2487	0.0669	(−0.3800, −0.1177)	−0.2758	0.0690	(−0.4126, −0.1414)
RTW dummy	2.5362	0.6403	(1.2845, 3.8123)	2.6325	0.6461	(1.3528, 3.8929)
SBSI	0.1206	0.0472	(0.0274, 0.2138)	0.1260	0.0478	(0.0336, 0.2185)
λ	0.7202	0.0057	(0.7083, 0.7321)	0.7201	0.0055	(0.7073, 0.7327)

CI: credible interval; PCT: percentage; SBSI: small business survival index.
Number of iterations is 30,000. We burn in the first 20% draws.

size is 100. When the DGP is the SDEM model with 3 lags, the frequency of the misspecified SDEM model with 2 lags is the highest in many cases. However, for DGP1 and DGP2, with moderate or high spatial dependence and a moderate normal prior for β_1, the SAR model can capture some features of the DGP better than the SDEM and the MESS models.

Table 14 summarizes the model frequencies for the misspecified SDEM model, the SAR model and the MESS model when the sample size increases to 1200. When the DGP is the SDEM model with 3 lags, the frequency of the misspecified SDEM model with 2 lags is the highest in most cases. This implies that the SDEM model, even misspecified, can still capture some features of the DGP better than the SAR and the MESS models.

6. An empirical illustration

In this section, we apply the Bayesian estimation method and the model selection procedure to the data set in Kalenkoski and Lacombe (2006).[33] Kalenkoski and Lacombe (2006) reexamine the effect of right-to-work (RTW) laws on manufacturing employment[34] after controlling for spatial dependence. Their data set consists of 427 counties, which are selected if they locate on the border of a RTW state and a non-RTW state. The dependent variable is manufacturing's employment as a percentage of total wage and salary employment in 2000. The key explanatory variable is the RTW dummy. Besides, we include the small business survival index (SBSI), which is a measure of the business climate for firms in a state, the percentage of population aged 18–64, and other characteristics of the labor market as explanatory variables. Descriptive statistics of all explanatory variables are summarized in Table 15. The empirical specifications we consider are:

$$
\begin{aligned}
&Y_n = cl_n + X_n\beta + W_nX_n\beta_1 + W_n^2X_n\beta_2 + (I_n - \rho W_n)^{-1}V_n,\\
&Y_n = \lambda W_nY_n + cl_n + X_n\beta + V_n,\\
&S_n^{ex}(\mu)Y_n = cl_n + X_n\beta + V_n.
\end{aligned}
\tag{6.1}
$$

The estimation methods we adopt are the ML method, and the Bayesian estimation method for the SDEM model with smoothness prior, the SAR model and the MESS model. We consider the informative smoothness prior for the SDEM model. The priors for the SAR model and the MESS model are the same as those in Eqs. (4.8) and (4.13). The value of the prior parameters is:

$$
\begin{aligned}
&SDEM\ \ with\ \ 2\ \ lags: L = 4\ \ or\ \ 8;\ \ p = 12\ \ or\ \ 120;\ \ a = 6, b = 4;\\
&SAR: p = 12\ \ or\ \ 120; a_1 = 6, b_1 = 4;\\
&MESS: p = 12\ \ or\ \ 120; \xi = 1\ \ or\ \ \sqrt{3};\ \ a_2 = 6,\ \ b_2 = 4.
\end{aligned}
$$

Specifically, $L=4$, $p=12$, and $\xi=1$ refer to a "tight" prior while $L=8$, $p=120$, and $\xi=\sqrt{3}$ refer to a "moderate" prior. The trace plots of ρ, λ and μ in the empirical study are depicted in Fig. 5. Furthermore, we evaluate the marginal likelihood of each model based on the expressions given in Section 4 and compare their posterior probabilities.

Table 16 provides ML estimation results for the three models. The estimate of ρ in the SDEM model is 0.7260 and highly significant. The estimates of λ and μ in the SAR and MESS model are, respectively, 0.7200 and −0.9983. Both are significant at the 1% level. Therefore, all

[33] We thank Kalenkoski and Lacombe for providing their data set.
[34] There are a large literature regarding the effect of RTW laws on industrial employment. See Moore and Newman (1985) and Moore (1998) for more discussions.

Table 20
Bayesian estimation of the MESS model for empirical data.

Independent variable	$p=12, \xi=1$			$p=120, \xi=\sqrt{3}$		
	Mean	S.D.	95% CI	Mean	S.D.	95% CI
Constant	−35.7935	13.9610	(−62.7098, −8.5029)	−50.9230	16.5378	(−82.9973, −19.1507)
PCT of population aged 18–64	0.3075	0.1022	(0.1032, 0.5049)	0.3842	0.1125	(0.1645, 0.6079)
PCT of population female	0.5875	0.1934	(0.2113, 0.9643)	0.7616	0.2176	(0.3443, 1.1863)
PCT of population nonwhite	−0.0771	0.0301	(−0.1366, −0.0188)	−0.0701	0.0301	(−0.1289, −0.0106)
PCT of population aged 25+, high school graduates	−0.0840	0.0570	(−0.1954, 0.0264)	−0.0572	0.0587	(−0.1720, 0.0583)
PCT of population aged 25+, with a bachelor's degree or higher	−0.2978	0.0745	(−0.4428, −0.1533)	−0.3355	0.0774	(−0.4878, −0.1839)
RTW dummy	2.7542	0.7052	(1.3655, 4.1407)	2.8787	0.7048	(1.4880, 4.2813)
SBSI	0.1371	0.0524	(0.0350, 0.2388)	0.1447	0.0521	(0.0428, 0.2480)
μ	−1.0142	0.0683	(−0.8823, −1.1479)	−1.0020	0.0694	(−1.1375, −0.8710)

CI: credible interval; PCT: percentage; SBSI: small business survival index.
Number of iterations is 30,000. We burn in the first 20% draws.

Table 21
Posterior model probabilities for the empirical data.

Model/Prior	$L=4, p=12, \xi=1$	$L=8, p=120, \xi=\sqrt{3}$
SDEM with 2 lags	0.0000	0.0000
SAR	1.0000	1.0000
MESS	0.0000	0.0000

three models suggest a strong spatial interaction effect. The estimated coefficients of the RTW dummy are all positive and significant, implying that the RTW laws do have a positive effect on manufacturing employment, other things equal. Tables 17–18 summarize Bayesian estimation results for the SDEM model. We rely on the Bayesian 95% credible intervals to decide whether the estimated coefficients are significantly different from zero or not. The Bayesian estimates of ρ are above 0.74 and significantly different from zero for both tight and moderate priors, implying a strong spatial dependent pattern. The Bayesian estimates of the RTW dummy coefficient are positive and significantly different from zero, indicating that the RTW law tends to increase manufacturing employment, all else equal. However, the Bayesian estimates of high order coefficients of the RTW dummy are not significantly different from zero, suggesting that the average RTW status of a county's first or second order neighbor might not have a strong impact on the manufacturing employment of that county.[35] Moreover, with a moderate prior the Bayesian estimates are closer to the ML estimates. Tables 19 and 20 summarize the Bayesian estimation results of the SAR model and the MESS model. We still get strong and significant spatial interaction effects. The estimated coefficients of the RTW dummy are positive and significant for the two models. Again, with moderate priors, the Bayesian estimates are similar to the ML estimates.

Finally, Table 21 gives the posterior probabilities for the three models. With a tight or moderate prior, the posterior probabilities of the SAR model are much higher than those of the SDEM model and MESS model. Therefore, a geometrical decline pattern of spatial externalities is more compatible with the data set.

7. Conclusion

In this paper we investigate a finite SDEM model and consider the Bayesian MCMC estimation of the model with a smoothness prior. We study also the corresponding Bayesian model selection procedure for the SDEM model, the SAR model and the MESS model. We derive expressions of marginal likelihoods of the three models, which greatly simplify the model selection procedure. Simulation results suggest that the Bayesian estimates of high order lag coefficients are more precise than the ML estimates. When the data is generated with a general declining pattern or a unimodal pattern for lag coefficients, the SDEM model can better capture the pattern than the SAR and the MESS models in most cases.

Using the dataset of Kalenkoski and Lacombe (2006), the SDEM model with an informative smoothness prior, the SAR model and the MESS model and corresponding Bayesian model selection procedure are applied to study the relationship between RTW laws and manufacturing employment. The empirical results show a positive relationship between RTW laws and manufacturing employment in each of the three models. The model selection results suggest that, among the three models, the SAR model would be the one to capture the pattern of spatial externalities for the data set.

Appendix A. Analytical integration for the marginal likelihood of the SDEM model with informative smoothness prior

Recall that the marginal likelihood for the SDEM model is

$$\begin{aligned} f(Y_n|M_{SDEM}) &= \int \pi\left(\gamma|\sigma^2\right)\pi\left(\sigma^2\right)\pi(\rho) f\left(Y_n|\gamma,\rho,\sigma^2\right) d\gamma d\sigma^2 d\rho \\ &= C^0_{SDEM} \int |R_n(\rho)| \sigma^{-[n+(m+1)k+1+a+2]} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2}\left[(Y_n(\rho)-Z_n(\rho)\gamma)'(Y_n(\rho)-Z_n(\rho)\gamma)\right]\right. \\ &\quad \left. +(\gamma-\overline{\gamma})' M_2(\gamma-\overline{\gamma}) + b\right\} d\gamma d\sigma^2 d\rho. \end{aligned}$$

The first step is to integrate out γ. We simplify the term $(Y_n(\rho)-Z_n(\rho)\gamma)'(Y_n(\rho)-Z_n(\rho)\gamma)+(\gamma-\overline{\gamma})'M_2(\gamma-\overline{\gamma})$. Note that

$$\begin{aligned} (Y_n(\rho)-Z_n(\rho)\gamma)'(Y_n(\rho)-Z_n(\rho)\gamma) &= (Y_n(\rho)-Z_n(\rho)\hat{\gamma}(\rho))'(Y_n(\rho)-Z_n(\rho)\hat{\gamma}(\rho)) \\ &\quad + (\gamma-\hat{\gamma}(\rho))' Z'_n(\rho) Z_n(\rho)(\gamma-\hat{\gamma}(\rho)) \end{aligned} \tag{A.1}$$

where $\hat{\gamma}(\rho) = \left(Z'_n(\rho)Z_n(\rho)\right)^{-1} Z'_n(\rho)Y_n(\rho)$. Let $Q_1(\rho) = (Y_n(\rho)-Z_n(\rho)\hat{\gamma}(\rho))'(Y_n(\rho)-Z_n(\rho)\hat{\gamma}(\rho))$, then we have

$$\begin{aligned} &(Y_n(\rho)-Z_n(\rho)\gamma)'(Y_n(\rho)-Z_n(\rho)\gamma)+(\gamma-\overline{\gamma})'M_2(\gamma-\overline{\gamma}) \\ &\quad = Q_1(\rho)+(\gamma-\hat{\gamma}(\rho))'Z'_n(\rho)Z_n(\rho)(\gamma-\hat{\gamma}(\rho))+(\gamma-\overline{\gamma})'M_2(\gamma-\overline{\gamma}). \end{aligned} \tag{A.2}$$

Moreover, let $A_1(\rho) = Z'_n(\rho)Z_n(\rho)+M_2$ and $\tilde{\gamma}(\rho) = A_1(\rho)^{-1}\left(Z'_n(\rho)Z_n(\rho)\hat{\gamma}(\rho)+M_2\overline{\gamma}\right)$,

$$\begin{aligned} &(\gamma-\hat{\gamma}(\rho))'Z'_n(\rho)Z_n(\rho)(\gamma-\hat{\gamma}(\rho))+(\gamma-\overline{\gamma})'M_2(\gamma-\overline{\gamma}) \\ &\quad = (\gamma-\tilde{\gamma}(\rho))'A_1(\rho)(\gamma-\hat{\gamma}(\rho))+Q_2(\rho)+Q_3(\rho) \end{aligned} \tag{A.3}$$

---

[35] This is consistent with the ML estimates. In particular, the estimated coefficients of $W_n \times$ *RTW dummy* and $W_n^2 \times$ *RTW dummy* are not statistically significant in Table 16.

with $Q_2(\rho)=\hat{\gamma}(\rho)'Z'_n(\rho)Z_n(\rho)\hat{\gamma}(\rho)-\tilde{\gamma}'(\rho)Z'_n(\rho)Z_n(\rho)\tilde{\gamma}(\rho)$ and $Q_3(\rho)=\overline{\gamma}'M_2\overline{\gamma}-\tilde{\gamma}'(\rho)M_2\tilde{\gamma}(\rho)$. According to Eqs. (A.1)–(A.3), we have

$$\begin{aligned}&(Y_n(\rho)-Z_n(\rho)\gamma)'(Y_n(\rho)-Z_n(\rho)\gamma)+(\gamma-\overline{\gamma})'M_2(\gamma-\overline{\gamma})\\&=Q_1(\rho)+Q_2(\rho)+Q_3(\rho)+\left(\gamma-\tilde{\gamma}(\rho)\right)'A_1(\rho)\left(\gamma-\tilde{\gamma}(\rho)\right).\end{aligned}\tag{A.4}$$

Therefore, using the property of multivariate normal pdf to integrate out γ, we get

$$\begin{aligned}f(Y_n|M_{SDEM})=&\,C^1_{SDEM}\int_\rho|A_1(\rho)|^{-\frac{1}{2}}|R_n(\rho)|\\&\int_{\sigma^2}\sigma^{-(n+a+2)}\exp\left(-\frac{1}{2\sigma^2}(b+Q_1(\rho)+Q_2(\rho)+Q_3(\rho))\right)d\sigma^2d\rho,\end{aligned}\tag{A.5}$$

with $C^1_{SDEM}=(2\pi)^{-\frac{n}{2}}\times L^{-(m+1)k}\times|R_d|\times p^{-1}\times(2\tau_n)^{-1}\times\frac{\frac{b}{2}^{\frac{a}{2}}}{\Gamma(\frac{a}{2})}$.

The second step is to integrate out σ^2. Let $v=\frac{Q(\rho)}{2\sigma^2}=\frac{b+Q_1(\rho)+Q_2(\rho)+Q_3(\rho)}{2\sigma^2}$, then

$$\begin{aligned}&\sigma^2=Q(\rho)(2v)^{-1}\\&\frac{d\sigma^2}{dv}=-Q(\rho)\left(2v^2\right)^{-1}.\end{aligned}\tag{A.6}$$

Combine Eqs. (A.5) and (A.6),

$$f(Y_n|M_{SDEM})=C^1_{SDEM}\times\int_\rho|A_1(\rho)|^{-\frac{1}{2}}|R_n(\rho)|\int_v Q(\rho)^{-\frac{n+a}{2}}2^{-\frac{n+a}{2}}v^{-\frac{n+a-2}{2}}\exp(-v)dvd\rho.\tag{A.7}$$

Using the property of the inverse gamma pdf, we have

$$f(Y_n|M_{SDEM})=C^1_{SDEM}\int_\rho|R_n(\rho)||A_1(\rho)|^{-\frac{1}{2}}Q(\rho)^{-\frac{n+a}{2}}d\rho.\tag{A.8}$$

with $C_{SDEM}=C^1_{SDEM}\times2^{\frac{n+a}{2}}\times\Gamma\left(\frac{n+a}{2}\right)$.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Boston.
Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. International Regional Science Review 26 (2), 153–166.
Casella, G., George, E., 1992. Explaining the Gibbs sampler. The American Statistician 46, 167–174.
Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321.
Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. The American Statistician 4, 327–335.
Chib, S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. Econometric Theory 12, 409–431.
Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis–Hastings output. Journal of the American Statistical Association 96, 270–281.
Gelfand, A., Smith, A., 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85, 398–409.
George, E., Sun, D., Ni, S., 2008. Bayesian stochastic search for VAR model restrictions. Journal of Econometrics 142, 553–580.
Han, X., Lee, L.F., 2013. Model selection using *J*-test for the spatial autoregressive model vs. the matrix exponential spatial model. Regional Science and Urban Economics 43, 250–271.
Hastings, W., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.
Hepple, L.W., 1995a. Bayesian techniques in spatial and network econometrics: 1. model comparison and posterior odds. Environment and Planning A 27, 447–469.
Hepple, L.W., 1995b. Bayesian techniques in spatial and network econometrics: 2. computational methods and algorithms. Environment and Planning A 27, 615–644.
Holloway, G., Shankara, B., Rahman, S., 2002. Bayesian spatial probit estimation: a primer and an application to *HYV* rice adoption. Agricultural Economics 27 (3), 383–402.
Kalenkoski, C., Lacombe, D., 2006. Right-to-work laws and manufacturing employment: the importance of spatial dependence. Southern Economic Journal 73 (2), 402–418.
Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53–67.
Lacombe, D., Holloway, G., Shaughnessy, T., 2012. Bayesian estimation of the spatial Durbin error model with an application to voter turnout in the 2004 presidential election. International Regional Science Review (http://irx.sagepub.com/content/early/2012/07/13/0160017612452133).
LeSage, J., Christina, L., 2012. The impact of migration on social capital: do migrants take their bowling balls with them? Growth and Change 43, 1–26.
LeSage, J., Pace, R.K., 2007. A matrix exponential spatial specification. Journal of Econometrics 140, 190–214.
LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. CRC Press, Boca Raton.
LeSage, J., Parent, O., 2007. Bayesian model averaging for spatial econometric models. Geographical Analysis 39, 241–267.
Madigan, D., York, J., 1995. Bayesian graphical models for discrete data. International Statistical Review 63, 215–232.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M., Teller, A., Teller, E., 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1092.
Moore, W.J., 1998. The determinants and effects of right-to-work laws: a review of the recent literature. Journal of Labor Research 19, 445–469.
Moore, W.J., Newman, R.J., 1985. The effects of right-to-work laws: a revew of the literature. Industrial & Labor Relations Review 38, 571–585.
Raftery, A.E., Lewis, S., 1992. How many iterations in the Gibbs sampler? In: Bernardo, J.M., B.J.D.A., Smith, A. (Eds.), Bayesian Statistics. Oxford University Press.
Shiller, R., 1973. A distributed lag estimator derived from smoothness priors. Econometrica 41 (4), 775–788.
Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). The Annals of Statistics 22, 1701–1762.
Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. J. Wiley and Sons, Inc., New York.
Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti. North-Holland Elsevier, Amsterdam.