Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: http://www.tandfonline.com/loi/ubes20

Bayesian Analysis of Spatial Panel Autoregressive Models With Time-Varying Endogenous Spatial Weight Matrices, Common Factors, and Random Coefficients

Xiaoyi Han & Lung-Fei Lee

Bayesian Analysis of Spatial Panel Autoregressive Models With Time-Varying Endogenous Spatial Weight Matrices, Common Factors, and Random Coefficients

Xiaoyi HAN

Wang Yanan Institute for Studies in Economics (WISE), Department of Public Economics, MOE Key Laboratory of Econometrics, and Fujian Key Laboratory of Statistical Science, Xiamen University, Xiamen 361005, China (*han.293@buckeyemail.osu.edu*)

Lung-Fei LEE

Department of Economics, Ohio State University, Columbus, OH 43210 (*lee.1777@osu.edu*)

This article examines spatial panel autoregressive (SAR) models with dynamic, time-varying endogenous spatial weights matrices, common factors, and random coefficients. An empirical application is on the spillover effects of state Medicaid spending. Endogeneity of spatial weights matrices comes from the correlation of "economic distance" and the disturbances in the SAR equation. Common factors control for common shocks to all states and random coefficients may capture heterogeneity in responses. The Bayesian Markov chain Monte Carlo (MCMC) estimation is developed. Identification of factors and factor loadings, and model selection issues based upon the deviance information criterion (DIC) are explored. We find that a state's Medicaid related spending is positively and significantly affected by those of its neighbors. Both welfare motivated move and yardstick competition are possible sources of strategic interactions among state governments. Welfare motivated move turns out to be more a driving force for the interdependence and states do exhibit heterogenous responses.

KEY WORDS: Bayesian estimation; Common factors; Deviance Information Criterion; Time-varying endogenous spatial weight matrix; Random coefficients; Spatial dynamic panel model.

1. INTRODUCTION

Spatial econometric models have been receiving attention in various areas of economics, for example, Case, Hines, and Rosen (1993) and Baicker (2005) in public economics, Lin (2010) in the context of social interaction, Fingleton (2001) and Ertur and Koch (2007) for the study of growth and technological interdependence. The most widely used model is the spatial autoregressive (SAR) model. The SAR model is applied to study the setting where the outcome of a spatial unit is influenced by those of its neighbors. A spatial weights matrix denotes direct neighbors of all spatial units, which characterizes the structure of relative cross-sectional interactions and dependence across spatial units. The corresponding spatial parameter captures the strength of interaction. In the study of social interaction, a student's behavior can be directly influenced by his or her friends' behaviors. The friendship relations can be represented by the spatial (network) weights matrix, and peer effects are captured by spatial parameters. In the context of strategic interaction among governments, a local government's spending on some programs might be affected by those of its neighbors and the SAR model can also be interpreted as a reaction function. The spatial weights matrix specifies neighboring relationships for the governments, based upon measures of physical and/or economics distances, and spatial parameters capture the spillover effect.

For estimation of the SAR model in the cross-sectional setting, there are Ord (1975) and Lee (2004) for the maximum likelihood estimation (MLE) method, Anselin (1980) and Kelejian and Prucha (1998, 1999) for the IV methods, Lee (2007) for the generalized method of moments (GMM) method, and LeSage and Pace (2009) for the Bayesian Markov chain Monte Carlo (MCMC) method. With panel data, the quasi-maximum likelihood estimate (QMLE) method for dynamic or static SAR models with fixed effects can be found, for example, in Yu, de Jong, and Lee (2008, 2012) and Lee and Yu (2010). Baltagi et al. (2003, 2007), Baltagi, Egger, and Pfaffermayr (2013), and Kapoor, Kelejian, and Prucha (2007) investigated random effect models. The Bayesian MCMC method for a dynamic SAR model with random effects is used in Parent and LeSage (2012). A review of spatial panels can be found in Lee and Yu (2015). The estimation of the SAR model is carried out predominantly with the assumption that the spatial weights matrix is exogenous. This exogeneity assumption might not be reasonable in some empirical applications. In the study of social interaction, some latent variables might affect both the friendship formation decision and behavioral outcome. In the context of state

Journal of Business & Economic Statistics
October 2016, Vol. 34, No. 4
DOI: 10.1080/07350015.2016.1167058

spending, some economics variables such as income or gross domestic product (GDP) are used to construct spatial weights. These would give rise to (possibly) endogenous network or spatial weighting matrices. To tackle this issue, Qu and Lee (2015) considered the QMLE, IV, and GMM method for the estimation of an SAR model with an endogenous spatial weights matrix, in a cross-sectional setting. In the social interaction setting, Hsieh and Lee (2016) proposed an SAR model with endogenous friendship formation and adopt the Bayesian MCMC method for estimation. For the estimation of a spatial panel model with an endogenous spatial weights matrix, Kelejian and Piras (2014) suggested the IV method.

In addition to the spatial econometrics on local dependence, another strand of literature deals with cross-sectional strong dependence with common factor in panels. Unobserved common factors affect all the cross-sectional units (hence, strong dependence), but with different intensities, captured by the so-called factor loadings. It is more flexible than the conventional fixed time effect specification. There are various estimation methods of common factor models in both classical and Bayesian approaches for regression panel models. For the classical approach, Pesaran (2006) proposed a common correlated estimator (CCE) that uses cross-sectional averages of dependent variables and explanatory variables to control for the effect of common factors. Bai (2009) and Moon and Weidner (2009) investigated the nonlinear least-square method. In the Bayesian literature, Geweke and Zhou (1996) advocated a "lower-triangular" identification scheme for factors and factor loadings, and apply Gibbs sampler to analyze an arbitrage pricing theory model. Aguilar and West (2000) considered the Bayesian inference of a dynamic factor model for multivariate financial time series. Lopes and West (2004) investigated MCMC algorithms for static factor models and developed a reversible jump MCMC algorithm to select number of factors. Bai and Wang (2015) studied minimal identification conditions and Bayesian estimation of dynamic factor models.

Most of the existing literatures treat spatial econometric models and common factor models separately. However, in some empirical applications, both types of cross-sectional dependence may present. In Holley, Pesaran, and Yamagata (2010), housing prices at different states not only exhibit spatial correlation but may also face economic-wide shock such as oil prices change or technology change. In welfare competition, States' welfare spending might be affected by both the spending of their neighbors and common policy shocks from the federal government. These empirical examples call for the joint modeling of spatial dependence and common factors. Pesaran and Tosetti (2011) proposed the CCE of a linear regression panel data model with both spatial dependence and common factors in error terms. Holley, Pesaran, and Yamagata (2010) applied the CCE method in Pesaran (2006) and Pesaran and Tosetti (2011) to analyze changes in real house prices using State level data. Bai and Li (2014) developed the QMLE method for an SAR model with common factors, which influences both dependent and explanatory variables as in Pesaran (2006). All of them assume that the spatial weights matrix is exogenously given and time invariant. So far, little research has been done to study SAR panel models with endogenous spatial weight matrices, which may also be time-varying and with common factors.

In this article, we examine the specification and estimation of a SAR model that combines endogenous time-varying spatial weights matrices, common factors, and possibly random coefficients for responses, with a Bayesian estimation framework, applied to the study of strategic interaction among state governments on Medicaid related spending. We focus on two neighborliness: one based on geographical distance and the other on "economic" distance. These two distances represent two well-known sources of spillovers of states' welfare spending: welfare motivated move and yardstick competition. In the context of welfare competition, because one state's welfare spending or benefit level might be influenced by those of its neighbors to which its welfare recipients might move, the spatial weight matrix should be constructed based on geographical distance or migration flows. On the other hand, a state's welfare spending could also be affected by those of its demographically or economically similar neighbors because the state's residents might look at welfare spending in nearby similar states to decide whether their own government is wasting revenue and deserved to be voted out of office. This is called "yardstick competition," Besley and Case (1995). Thus, a government's neighbors can also refer to those who share similar demographic or economic characteristics. In this case, the spatial weights matrix should be specified based upon demographic or economic variables for distance. Endogeneity of spatial weights matrices comes from the correlation of economic distance and the disturbance in the SAR equation. Common factors control for common shocks to all states and factor loadings may capture heterogeneity in states' responses. Heterogenous responses are also allowed for some observed endogenous and exogenous interactions. For estimation, the Bayesian MCMC method is developed. Its computational tractability and convergence are supported by simulation results. Identification of time factor and factor loadings, and various model selection issues using deviance information criterion (DIC) are explored. We find that a state's Medicaid related spending is positively and significantly affected by the Medicaid related spending of its neighbors. As state governments respond to their geographically close and economically similar neighbors, so in the context of Medicaid spending, welfare motivated move and yardstick competition are both sources of strategic interactions among state governments.

Our model integrates recent literatures on spatial models and common factor models. Literature on spatial models with endogenous spatial weights matrices (Hsieh and Lee 2016; Qu and Lee 2015; Kelejian and Piras 2014) does not incorporate common factors while literature on common factor models (Bai 2009; Moon and Weidner 2009; Bai and Wang 2015) does not explore spatial dependence. Bai and Li (2014) and Pesaran and Tosetti (2011) considered the joint modeling of the two but they assume spatial weights matrices are exogenous. We manage to combine the time-varying endogenous spatial weights matrices and common factors to form a general complex SAR panel model, which is empirically motivated. One might question that our new model does not offer additional insights to the empirical example, given that model selection results end up with a restricted model. We argue that our modeling strategy follows the idea of general to specific approach in Campos, Ericsson, and Hendry (2005) and Doornik and Hendry (2014): propose an initial formulated general model that includes all relevant variables (new features), and search for the appropriate (possibly restricted) model by eliminating irrelevant variables (features) and retaining variables (features) that matter. This approach is very useful in the analysis of big data, as put forward by Doornik and Hendry (2014).

The article is organized as follows: Section 2 presents a general complex SAR model, which is used as an initial general model formulation in the strategy from general to specific models in Campos, Ericsson, and Hendry (2005). Such a strategy has been emphasized lately for big data analytic in Doornik and Hendry (2014). Section 3 specifies prior distributions and discusses the Bayesian MCMC estimation. Section 4 studies the identification of factor and factor loadings, and model selection issues. In particular, we aim to select the true model with a correct number of common factors. Section 5 summarizes simulation results on sampling properties of our Bayesian estimation method and model selection procedures. Section 6 provides results of our empirical study. Conclusions are drawn in Section 7.

2. THE MODELS

2.1 A General Complex SAR Panel Model

We start with an initial general formulated spatial model, which is a dynamic panel SAR model with endogenous time-varying spatial weights matrices and common factors. Empirical studies of spillover effect of state spending specify their spatial weights matrices based upon economic or demographic characteristics. In a panel data setting, as economic characteristics may change over time, those spatial weights matrices would be time-varying. Denote $w^b_{ij,t}$ the ijth element of W_{nt} before row-normalization. The ijth element of W_{nt} is specified as $w_{ij,t} = \frac{w^b_{ij,t}}{\sum_{k=1}^n w^b_{ik,t}},\ \ i \neq j;\ \ w_{ii,t} = 0,\ \ t = 1, 2, \ldots, T$. To specify $w^b_{ij,t}$, let d_{ij} be the geographical distance between i and j and $z_{it} = (z_{i1,t}, z_{i2,t}, \ldots, z_{ip,t})$ be a $1 \times p$ vector of i's demographic or economic characteristics at time t. Denote $Z_{nt} = (z'_{1t}, z'_{2t}, \ldots, z'_{nt})'$ be the $n \times p$ matrix of z_{it}'s. The demographic or economic distance between i and j is defined as $E_{ijr,t} = |z_{ir,t} - z_{jr,t}|,\ r = 1, 2, \ldots, p,\ t = 1, 2, \ldots, T$. $w^b_{ij,t}$ is constructed as

$$w^b_{ij,t} = \gamma_{ij,t} \times \tilde{w}^b_{ij,t},\ \ \tilde{w}^b_{ii,t} = 0;$$
$$\tilde{w}^b_{ij,t} = d_{ij}^{-\phi_0} E_{ij1,t}^{-\phi_1} E_{ij2,t}^{-\phi_2} \ldots E_{ijp,t}^{-\phi_p},\ \ i \neq j \tag{2.1}$$

for $i, j = 1, \ldots, n$ and $t = 1, \ldots, T$, where $\gamma_{ij,t}$ is a prespecified binary $0 - 1$ indicator, determining whether i and j are neighbors or not, for instance, whether i and j are bordering states or not; $\tilde{w}^b_{ij,t}$ captures the relative strength of the interaction, with ϕ_i's being some prespecified nonnegative parameters. If the effect of welfare motivated move (geographical distance) is the only interest, we may set $\phi_0 > 0$ and $\phi_i = 0$ for $i = 1, 2, \ldots, p$, as in Corrado and Fingleton (2012). Here, for a general formulation, $\phi_i's$ may not be restricted to zero. Equation (2.1) assumes the magnitude of interaction for each (i, j) pair depends on their distance measures. If i and j are relatively close in d_{ij} or $E_{ijr,t}$'s, the spillovers of j on i can be large, as long as $\phi_r \neq 0$.

A general formulated panel SAR model with W_{nt}'s can be

$$Y_{nt} = \lambda W_{nt} Y_{nt} + \psi_1 Y_{n,t-1} + \rho W_{n,t-1} Y_{n,t-1} + X_{n1t}\beta_1 + \Lambda f_t + V_{nt};$$
$$Z_{nt} = Y_{n,t-1}\psi_2 + X_{n2t}\beta_2 + \Omega \overline{f}_t + U_{nt},\ \ t = 1, 2, \ldots, T. \tag{2.2}$$

This model contains two sets of equations: an SAR outcome equation Y_{nt} and equations Z_{nt} for entries in W_{nt}. $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of dependent variables. X_{n1t} and X_{n2t} are $n \times k_1$ and $n \times k_2$ matrices of time-varying exogenous regressors. ψ_1 is a scalar parameter while $\psi_2 = (\psi_{21}, \psi_{22}, \ldots, \psi_{2p})$ is a $1 \times p$ row-vector of parameters. $Y_{n,t-1}$ affects both Y_{nt} in the SAR equation and Z_{nt} in the entry equations. The dynamic term $\psi_1 Y_{n,t-1}$ controls for the persistence in state's welfare spending while $\rho W_{n,t-1} Y_{n,t-1}$ would capture the diffusion (dynamic spillovers). The $V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ is an $n \times 1$ vector and $U_{nt} = (u'_{1t}, u'_{2t}, \ldots, u'_{nt})'$ is an $n \times p$ matrix of error terms. Assume $(v_{it}, u_{it}) \sim$ iid $\mathcal{N}_{p+1}(0, \left(\begin{smallmatrix} \sigma_v^2 & \sigma'_{vu} \\ \sigma_{vu} & \Sigma_u \end{smallmatrix}\right))$ across all i's and t's. The demographic or economic characteristics $z_{ir,t}$'s can be endogenous and so are the time-varying spatial weight matrices, as v and u's can be correlated with $\sigma_{vu} \neq 0$. f_t is the $q \times 1$ unobserved common factors at time t. Following Lopes and West (2004), assume that (i) f_t's are independently $\mathcal{N}_q(0, I_q)$ distributed for all t, (ii) f_t and (v_{it}, u_{it}) are independent for all t and i. The independent factors assumptions can be generalized to allow correlations of f_t's over time when a dynamic factor model is considered. Λ is the corresponding $n \times q$ factor loading matrix for the SAR outcome equation. $\Omega = (\Omega_1, \ldots, \Omega_p)$ where Ω_l denotes the $n \times q$ factor loading matrix for the lth set of entry equation, with $l = 1, 2, \ldots, p$, and their time factors are $\overline{f}_t = I_p \otimes f_t$. The f_t's simultaneously affect all the cross-sectional units in the outcome equation and entry equations, with different degrees captured by factor loading matrices Λ and Ω. The number of factors q is in general not known. In Section 4, we consider the model selection issue raised by possible different number of factors. In the context of Medicaid spending, Baicker (2005) mentioned that federally mandated Medicaid eligibility expansion in the 1980s and 1990s are common shocks to all states. Federal legislation requires state governments to expand their Medicaid program to cover more pregnant women and children. The impacts of this shock are heterogenous since some states have already covered the required groups while some have not. Another common shock mentioned by Baicker (2005) is the increase of overall medical price. States with larger Medicaid spending program would face a larger shock as medical price increases. When the data of federally mandated Medicaid expansion or medical price are not available, these effects can be captured by common factors.

It is known that the common factor specification with factor loadings may include the additive time and state fixed effects as a special case. Let q, the number of factors, be equal to $p + 2$, where p is the number of columns of Z_{nt}. And $f_t = (1, \alpha_{1t}, \alpha'_{2t})'$ be the corresponding $(p + 2) \times 1$ vector of factors. For any cross-sectional unit i, $\Lambda'_i f_t = c_{i1} + \alpha_{1t}$ with $\Lambda_i = (c_{i1}, 1, 0, \ldots, 0)'$. Similarly, denote $\Omega'_{l,i}$ as the ith row of Ω_l and $\alpha_{2t,l}$ as the lth element of α_{2t}. Then $\Omega'_{l,i} f_t = c_{i2,l} + \alpha_{2t,l}$ when $\Omega'_{l,i} = (c_{i2,l}, 0, \ldots, 1, \ldots, 0)$, where the "1" is located to the $(l + 2)$th position. Hence, the model in (2.2) may nest the spatial dynamic panel data (SDPD) models with additive individual and time effects as a special case.

For the model in Equation (2.2), loadings and factors cannot be separately identified, even after normalization restrictions on the covariance matrix of f_t's are imposed. For any orthogonal $q \times q$ matrix Q, we have the observationally equivalent relations $\Lambda f_t = \Lambda Q Q' f_t$ and $\Omega \overline{f}_t = \Omega(I_p \otimes QQ' f_t)$. To identify factors and loadings up to sign identification, we need to pin down the matrix Q. As it contains q^2 elements, q^2 restrictions are needed to determine f_t and the loadings. Even though the factor loadings in the SAR outcome and entry equations are different, still only q^2 restrictions are needed for this model, because once Λ and f_t's in the SAR equation are identified, Ω_l in the lth entry equation will also be identified for all $l = 1, 2, \ldots, p$. Assuming the covariance matrix of f_t equaling to I_q only provides us with $\frac{q(q+1)}{2}$ restrictions. To achieve identification, further $\frac{q(q-1)}{2}$ restrictions are needed. In a model, some researchers may possess the view that f_t's are only introduced to control for latent common shocks. In such a view, we are only interested in the products Λf_t's and $\Omega_l f_t$'s. Without imposing the remaining $\frac{q(q-1)}{2}$ restrictions, f_t and Λ or Ω_l's cannot be separately identified but their products may still be identified. Therefore, the estimation of λ, β's and Σ would not be affected even if identification of Λ, Ω_l's, and f_t's are not achieved. More discussions on identification of the factors and the loadings in the model will be in Section 4.

Other than the factors and factor loadings, we follow the idea of identification in the Bayesian theory (Kadane 1974; Hsiao 1983; Poirier and Tobias 2003) to demonstrate the identification of the remaining parameters, λ, Σ, and all β's. Specifically, we check whether data (likelihood) brings information to update the posterior distributions of those parameters. A simulation study is conducted to show that the posterior distributions of those parameters do collapse to their true values for a large sample in Section 5 and Figure 1(a).

Assume the initial period, $Y_{n,0}$ is exogenously given. Let $S_{nt}(\lambda) = I_n - \lambda W_{nt}$ and $A_{nt}(\lambda, \psi_1, \rho) = S_{nt}^{-1}(\lambda)(\psi_1 I_n + \rho W_{n,t-1})$. The reduced form of the SAR equation in Equation (2.2) is $Y_{nt} = A_{nt}(\lambda, \psi_1, \rho) Y_{n,t-1} + S_{nt}^{-1}(\lambda)(X_{n1t}\beta_1 + \Lambda f_t + V_{nt})$. According to Lee and Yu (2012), with time-varying spatial weight matrices, a sufficient condition for the stability is $|\lambda| < 1$ and $|\lambda| + |\psi_1| + |\rho| < 1$, for row-normalized W_{nt}'s. This restriction would be imposed in the priors and the sampling steps for λ, ψ_1, and ρ. Alternatively, if one is not interested in the dynamic features of the model and set $\rho = \psi_1 = 0$ and $\psi_2 = 0$, we have a restricted static model, where the stability condition would be reduced to $|\lambda| < 1$.

2.2 Extension to Allow Random Effects in Coefficients

The model in Equation (2.2) assumes that, in the SAR equation, all states react the same way to the average spending of its neighbors (λ), to the average lagged spending of its neighbors (ρ), to its own characteristics (β_1), and to common factors (Λ). While individual responses to common time factors can be heterogenous, reactions of states might also be heterogenous to other observed factors. Similarly, in the entry equation, a state's GDP or income per capita might be influenced by its lagged spending (ψ_2), demographic characteristics (β_2), and common factors (Ω) differently. This motivates random effects in coefficients on the parameters λ, ψ_1, ρ, β_1, Λ, ψ_2, β_2, and Ω.

Let λ_i be the response of state i's spending to its neighbors' spending, ψ_{1i} be the response of i's spending to its own lagged spending, ρ_i be the response of i's spending to the lagged spending of its neighbors, and β_{1i} be the response of i's spending to its own characteristics. Denote $\mathcal{L} = \mathrm{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, $\mathcal{R} = \mathrm{Diag}(\rho_1, \rho_2 \ldots, \rho_n)$, $\Psi_1 = \mathrm{Diag}(\psi_{11}, \psi_{12}, \ldots, \psi_{1n})$, $\overline{X}_{n1t} = \mathrm{Diag}(x_{11,t}, x_{21,t}, \ldots, x_{n1,t})$, and $\overline{\beta}_1 = (\beta_{11}', \beta_{12}', \ldots, \beta_{1n}')'$. Furthermore, for the entry equation, recall that $\psi_2 = (\psi_{21}, \ldots, \psi_{2p})$ is a $1 \times p$ row vector while $\beta_2 = (\beta_{21}, \ldots, \beta_{2p})$ is a $k_2 \times p$ matrix of coefficients. Let $\overline{\beta}_{2i,l} = (\psi_{2i,l}', \beta_{2i,l}')'$ be the lth $(k_2+1) \times 1$ vector for individual i, and $\overline{\beta}_{2l} = (\overline{\beta}_{21,l}', \ldots, \overline{\beta}_{2n,l}')'$ be the collection of $\overline{\beta}_{2i,l}$'s for all i's, for $l = 1, 2, \ldots, p$. Denote $\overline{\beta}_2 = (\overline{\beta}_{21}, \overline{\beta}_{22}, \ldots, \overline{\beta}_{2p})$. Also denote $\overline{x}_{i2t} = (y_{i,t-1}, x_{i2t})$, and $\overline{X}_{n2t} = \mathrm{Diag}(\overline{x}_{12t}, \overline{x}_{22t}, \ldots, \overline{x}_{n2t})$. The extended dynamic SAR model with random coefficients is

$$
\begin{aligned}
Y_{nt} &= \mathcal{L} W_{nt} Y_{nt} + \Psi_1 Y_{n,t-1} + \mathcal{R} W_{n,t-1} Y_{n,t-1} \\
&\quad + \overline{X}_{n1t}\overline{\beta}_1 + \Lambda f_t + V_{nt}, \\
Z_{nt} &= \overline{X}_{n2t}\overline{\beta}_2 + \Omega f_t + U_{nt}, \ t = 1, 2, \ldots, T. \qquad (2.3)
\end{aligned}
$$

Denote $\overline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)'$, $\overline{\rho} = (\rho_1, \rho_2, \ldots, \rho_n)'$, and $\overline{\psi}_1 = (\psi_{11}, \psi_{12}, \ldots, \psi_{1n})'$ as a collection of spatial random coefficients. Let $S_{nt}(\overline{\lambda}) = I_n - \mathcal{L} W_{nt}$ and $A_{nt}(\overline{\lambda}, \overline{\psi}_1, \overline{\rho}) = S_{nt}^{-1}(\overline{\lambda})(\Psi_1 + \mathcal{R} W_{n,t-1})$. The reduced form of the SAR equation in (2.3) is $Y_{nt} = A_{nt}(\overline{\lambda}, \overline{\psi}_1, \overline{\rho}) Y_{n,t-1} + S_{nt}^{-1}(\overline{\lambda})(\overline{X}_{n1t}\overline{\beta}_1 + \Lambda f_t + V_{nt})$. The corresponding stability condition is $||A_{nt}(\overline{\lambda}, \overline{\psi}_1, \overline{\rho})||_\infty < 1$. Notice that $||A_{nt}(\overline{\lambda}, \overline{\psi}_1, \overline{\rho})||_\infty < ||S_{nt}^{-1}(\overline{\lambda})||_\infty \times (||\Psi_1||_\infty + ||\mathcal{R}||_\infty ||W_{n,t-1}||_\infty)$. According to Horn and Johnson (1985), with $||\mathcal{L} W_{nt}||_\infty < 1$, $S_{nt}(\overline{\lambda}) = (I_n - \mathcal{L} W_{nt})$ is invertible. With a row-normalized W_{nt}, $||S_{nt}^{-1}(\overline{\lambda})||_\infty \leq \frac{1}{1 - ||\mathcal{L}||_\infty}$. Therefore, $||A_{nt}(\overline{\lambda}, \overline{\psi}_1, \overline{\rho})||_\infty < \frac{1}{1 - ||\mathcal{L}||_\infty} \times (||\Psi_1||_\infty + ||\mathcal{R}||_\infty)$. To ensure stability, we assume $||\mathcal{L}||_\infty + ||\Psi_1||_\infty + ||\mathcal{R}||_\infty < 1$. This stability condition would be imposed on the sampling step of $\overline{\lambda}$, $\overline{\psi}_1$, and $\overline{\rho}$.

3. BAYESIAN MCMC ESTIMATION

3.1 The Likelihood Function

Let $\{Y_{nt}\}$ and $\{Z_{nt}\}$ denote the collections of all Y_{nt}'s and Z_{nt}'s for $t = 1, 2, \ldots, T$. Denote $\Psi = (\lambda, \psi_1, \rho)$ and $\tilde{\beta}_2 = (\psi_2, vec(\beta_2)')'$. Let $\beta = (\beta_1', \tilde{\beta}_2')'$, which is of dimension $k_1 + p(1 + k_2)$. Let $\Sigma = \begin{pmatrix} \sigma_v^2 & \sigma_{vu}' \\ \sigma_{vu} & \Sigma_u \end{pmatrix}$ be the $(p+1) \times (p+1)$ covariance matrix of v_{it}'s and u_{it}'s. Define $\Sigma_t = I_n \otimes \Sigma$ and an $n(p+1) \times 1$ column vector $M_{t|f} = \begin{pmatrix} S_{nt}(\lambda)Y_{nt} - \overline{A}_{nt}(\psi_1, \rho)Y_{n,t-1} - X_{n1t}\beta_1 - \Lambda f_t \\ vec(Z_{nt} - Y_{n,t-1}\psi_2 - X_{n2t}\beta_2 - \Omega \overline{f}_t) \end{pmatrix}$, where $S_{nt}(\lambda) = I_n - \lambda W_{nt}$, $\overline{A}_{nt}(\psi_1, \rho) = \psi_1 I_n + \rho W_{n,t-1}$, and Σ_t is the corresponding $n(p+1) \times n(p+1)$ covariance matrix. Note that $|\Sigma_t| = |\Sigma|^n$. The likelihood function from (2.2) at time t is

$$
\begin{aligned}
& f(Y_{nt}, Z_{nt} | \Lambda, \Omega, f_t, \Psi, \beta, \Sigma) \\
& \quad \propto |I_n - \lambda W_{nt}| \times |\Sigma|^{-\frac{n}{2}} \times \exp\left(-\frac{1}{2} M_{t|f}' \Sigma_t^{-1} M_{t|f}\right).
\end{aligned} \qquad (3.1)
$$

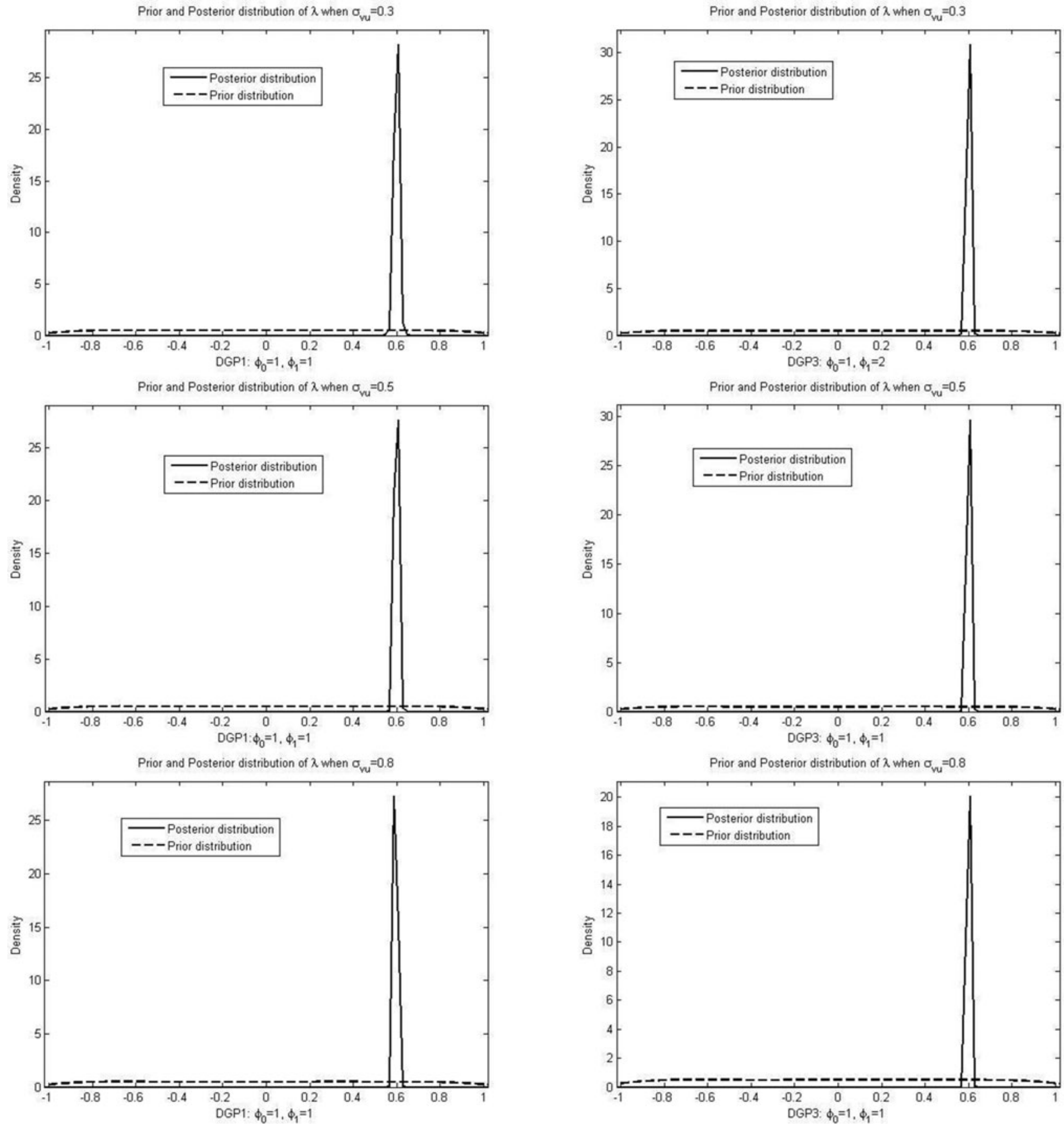

Figure 1. Graphs of simulation and empirical results (a) Prior and Posterior distribution of λ in simulation: DGP1 and DGP3 (*continued on following page*).

The likelihood function conditional on Z_{nt} is

$$f(Y_{nt}|Z_{nt},\Lambda,\Omega,f_t,\Psi,\beta,\Sigma) \propto \left(\sigma_\xi^2\right)^{-\frac{n}{2}} \times |I_n - \lambda W_{nt}| \times \exp\left(-\frac{H'_{nt|f}H_{nt|f}}{2\sigma_\xi^2}\right), \quad (3.2)$$

with $H_{nt|f} = S_{nt}(\lambda)Y_{nt} - \psi_1 Y_{n,t-1} - \rho W_{n,t-1}Y_{n,t-1} - X_{n1t}\beta_1 - \Lambda f_t - (Z_{nt} - Y_{n,t-1}\psi_2 - X_{n2t}\beta_2 - \Omega\overline{f}_t)\eta$, where $\eta = \Sigma_u^{-1}\sigma_{vu}$ and $\sigma_\xi^2 = \sigma_v^2 - \sigma'_{vu}\Sigma_u^{-1}\sigma_{vu}$. Equation (3.2) would be useful when we sample Ψ.

3.2 The MCMC Algorithm

Denote $\theta = (\Psi, \beta, \Sigma)$. Let $\Lambda = (\Lambda_1, \Lambda_2, \ldots, \Lambda_n)'$, where $\Lambda_i = (\Lambda_{i1}, \ldots, \Lambda_{iq})'$ is a $q \times 1$ column vector of Λ_{ij}'s. Also let $\Omega_l = (\Omega_{l,1}, \Omega_{l,2}, \ldots, \Omega_{l,n})'$, where $\Omega_{l,i} = (\Omega_{l,i1}, \ldots, \Omega_{l,iq})'$ is a $q \times 1$ column vector for $l = 1, 2, \ldots, p$. The Λ and $\Omega = (\Omega_1, \ldots, \Omega_p)$ are the loading matrices for f_t, where $f_t \sim \mathcal{N}_q(0, I_q)$ for $t = 1, 2, \ldots, T$. Given the stability condition, the prior of $\Psi = (\lambda, \psi_1, \rho)$ is $\pi(\Psi) = \pi(\lambda) \times \pi(\rho|\lambda) \times \pi(\psi_1|\lambda, \rho)$, with $\pi(\lambda) \sim \mathcal{U}(-1, 1)$, $\pi(\rho|\lambda) \sim \mathcal{U}(-1 + |\lambda|, 1 - |\lambda|)$, and $\pi(\psi_1|\lambda, \rho) \sim \mathcal{U}(-1 +$

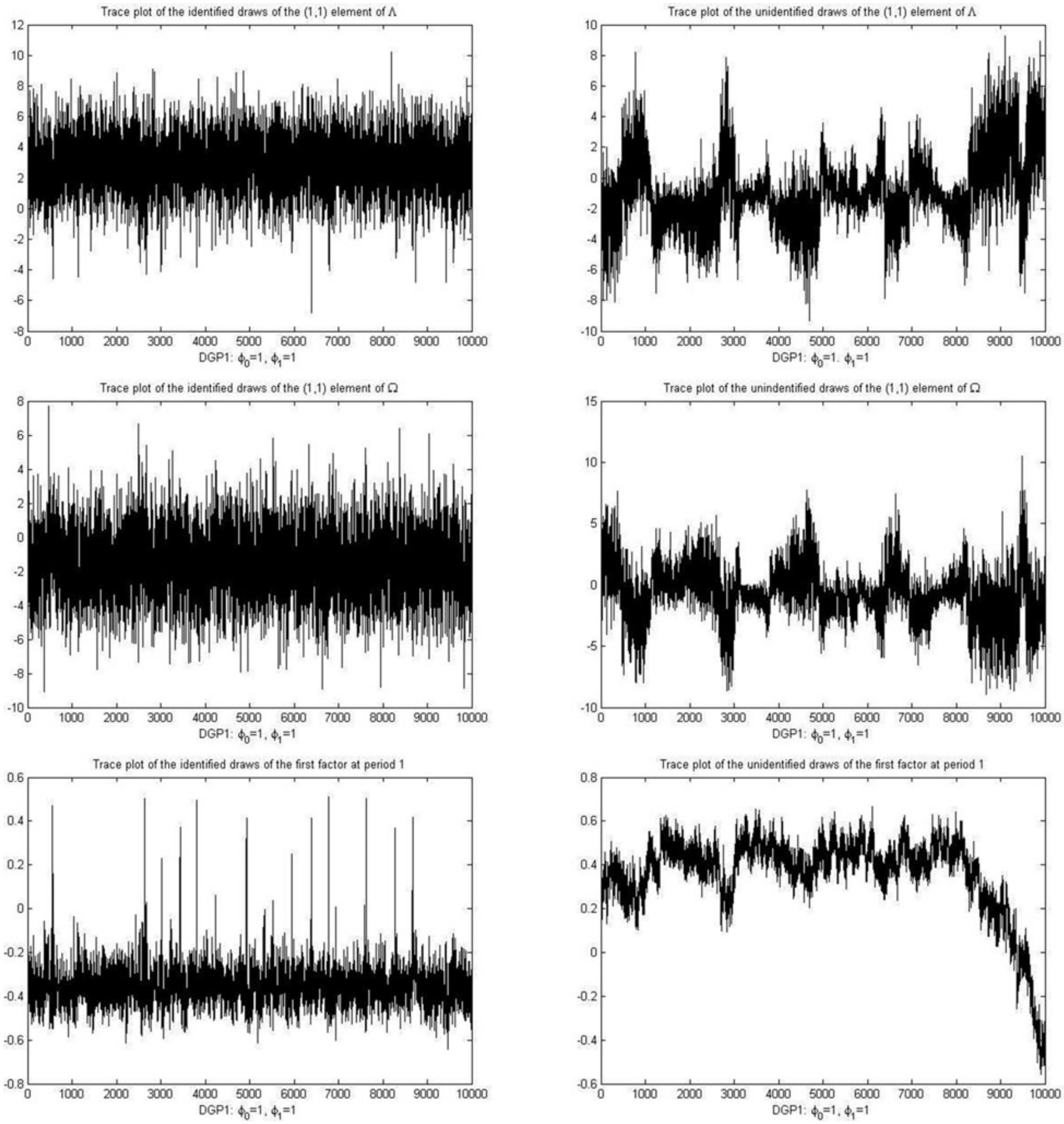

Figure 1. Graphs of simulation and empirical results (b) Trace plots of elements of identified and unidentified f_t, Λ and Ω in simulation: DGP1 *(continued from previous page)*.

$|\lambda| + |\rho|, 1 - |\lambda| - |\rho|)$, where $\mathcal{U}(a, b)$ is the uniform random variable over the intervals (a, b). The priors of β, Σ, Λ_i's, and $\Omega_{l,i}$'s are

$$\beta = (\beta_1', \tilde{\beta}_2')' \sim \mathcal{N}_{k_1+p(1+k_2)}(\beta_O, B_O),$$

$$\Sigma = \begin{pmatrix} \sigma^2 & \sigma_{vu}' \\ \sigma_{vu} & \Sigma_u \end{pmatrix} \sim \mathcal{IW}_{p+1}(\nu, \nu R),$$

$$\Lambda_i \sim \mathcal{N}_q(\Lambda_O, \tau_1 I_q),\ i = 1, 2, \ldots, n,$$

$$\Omega_{l,i} \sim \mathcal{N}_q(\Omega_O, \tau_2 I_q),\ i = 1, 2, \ldots, n,\ l = 1, 2, \ldots p, \quad (3.3)$$

where $\mathcal{N}_m(b_O, B_O)$ is the multivariate normal m-dimensional vector with mean b_O, and covariance matrix B_O for any m, b_O, and B_O; $\mathcal{IW}_{p+1}(\nu, \nu R)$ has an inverse-Wishart distribution with ν degrees of freedom and a symmetric positive definite scale matrix νR.

Denote $\{f_t\}$ as the collection of f_t, $t = 1, 2, \ldots, T$ and $\pi(\{f_t\})$ as the normal density for $\{f_t\}$. Applying Bayes' theorem, the joint posterior density function of θ, Λ, Ω, and $\{f_t\}$ is

$$\pi(\theta, \Lambda, \Omega, \{f_t\}|\{Y_{nt}\}, \{Z_{nt}\}) \propto \pi(\theta) \times \pi(\Lambda) \times \pi(\Omega)$$
$$\times \pi(\{f_t\}) \times f(\{Z_{nt}\}|\Omega, \{f_t\}, \theta)$$
$$\times f(\{Y_{nt}\}|\{Z_{nt}\}, \Lambda, \Omega, \{f_t\}, \theta),$$

where a $\pi(.)$ on the right-hand side represents a density of a prior distribution. For simplicity, exogenous $\{X_{n1t}\}$, $\{X_{n2t}\}$, and d_{ij}'s

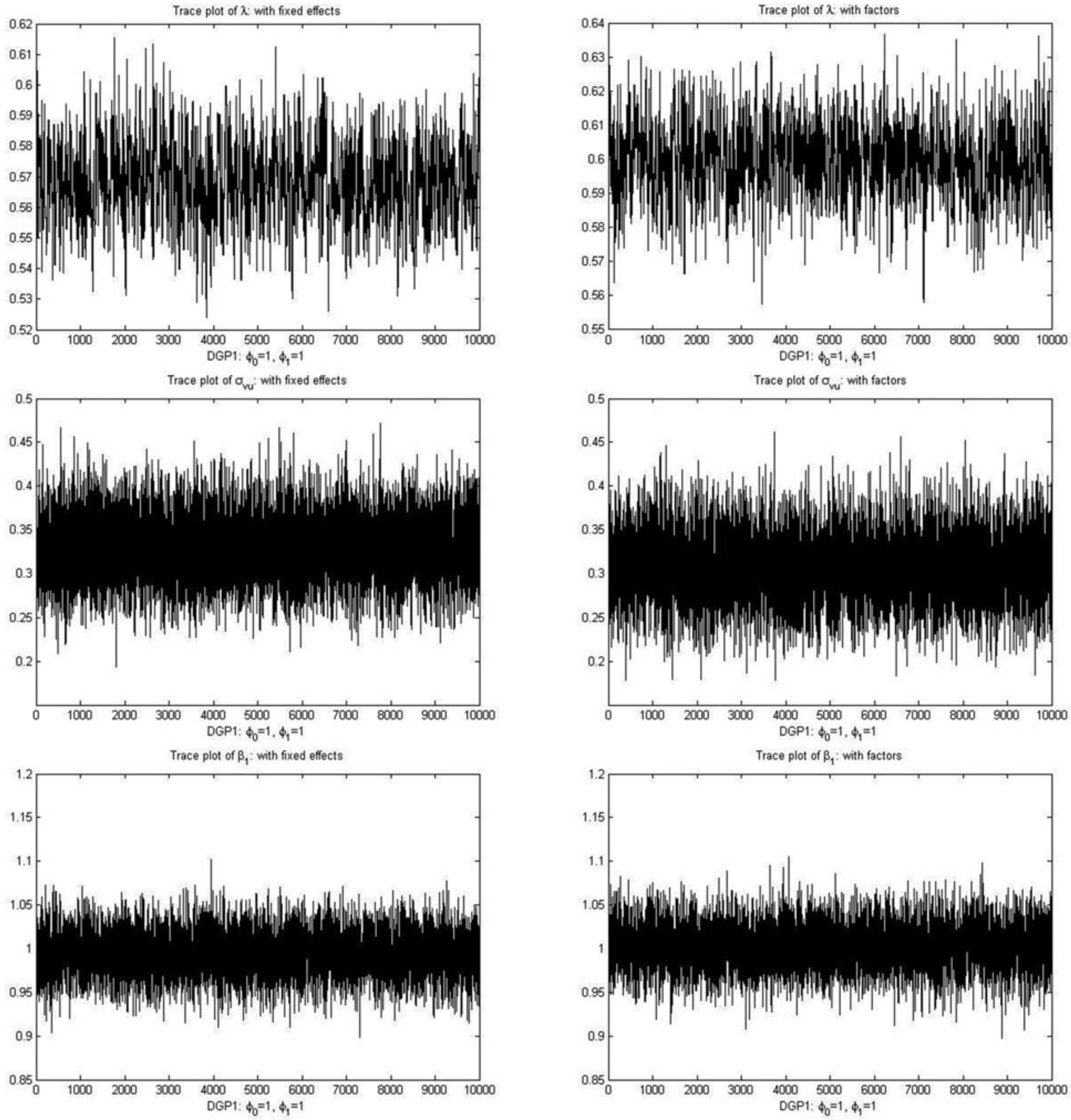

Figure 1. Graphs of simulation and empirical results (c) Trace plots of λ, σ_{vu} and β_1 in simulation: DGP1 *(continued from previous page)*.

are suppressed. With conjugate priors, β, Σ, Λ_i's, $\Omega_{l,i}$'s, and f_t's can be sampled via Gibbs sampling steps. But a Metropolis–Hastings (M-H) step is needed to sample $\Psi = (\lambda, \psi_1, \rho)$. With the presence of W_{nt}, the coefficients ρ and ψ_1 of predetermined variables do not have conjugate priors, and cannot be updated using a Gibbs sampling step. So we follow Parent and LeSage (2012) to use one M-H step to sample Ψ. We use the adaptive Metropolis (AM) algorithm in Haario, Saksman, and Tamminen (2001) and Roberts and Rosenthal (2009). Unlike the standard M-H algorithm, which uses the random walk proposal with covariance matrix equaling to an identity matrix, the AM algorithm uses historical MCMC draws to construct the covariance matrix of the proposal distribution. Let Ψ be the $k \times 1$ parameter vector to be updated. At the iteration g, the historical MCMC draws of Ψ is $(\Psi^{(0)}, \Psi^{(1)}, \ldots, \Psi^{(g-1)})$. The AM proposal by Roberts and Rosenthal (2009) is

$$f_g(\Psi|\Psi^{(0)}, \Psi^{(1)}, \ldots, \Psi^{(g-1)}) = \begin{cases} \mathcal{N}_k(\Psi^{(g-1)}, 0.1^2 I_k/k) & g \le 2k \\ (1-\delta)\mathcal{N}_k(\Psi^{(g-1)}, \mathrm{cov}(\Psi^{(0)}, \Psi^{(1)}, \ldots, \Psi^{(g-1)})2.38^2/k) & \\ \quad + \delta\mathcal{N}_k(\Psi^{(g-1)}, 0.1^2 I_k/k) & g > 2k, \end{cases} \tag{3.4}$$

where the scaling factor 2.38^2 optimizes the mixing properties of the Metropolis search for the Gaussian proposals (Gelman, Roberts, and Gilks 1996). The AM proposal is a mixture of two normal distributions with a ratio parameter δ when the

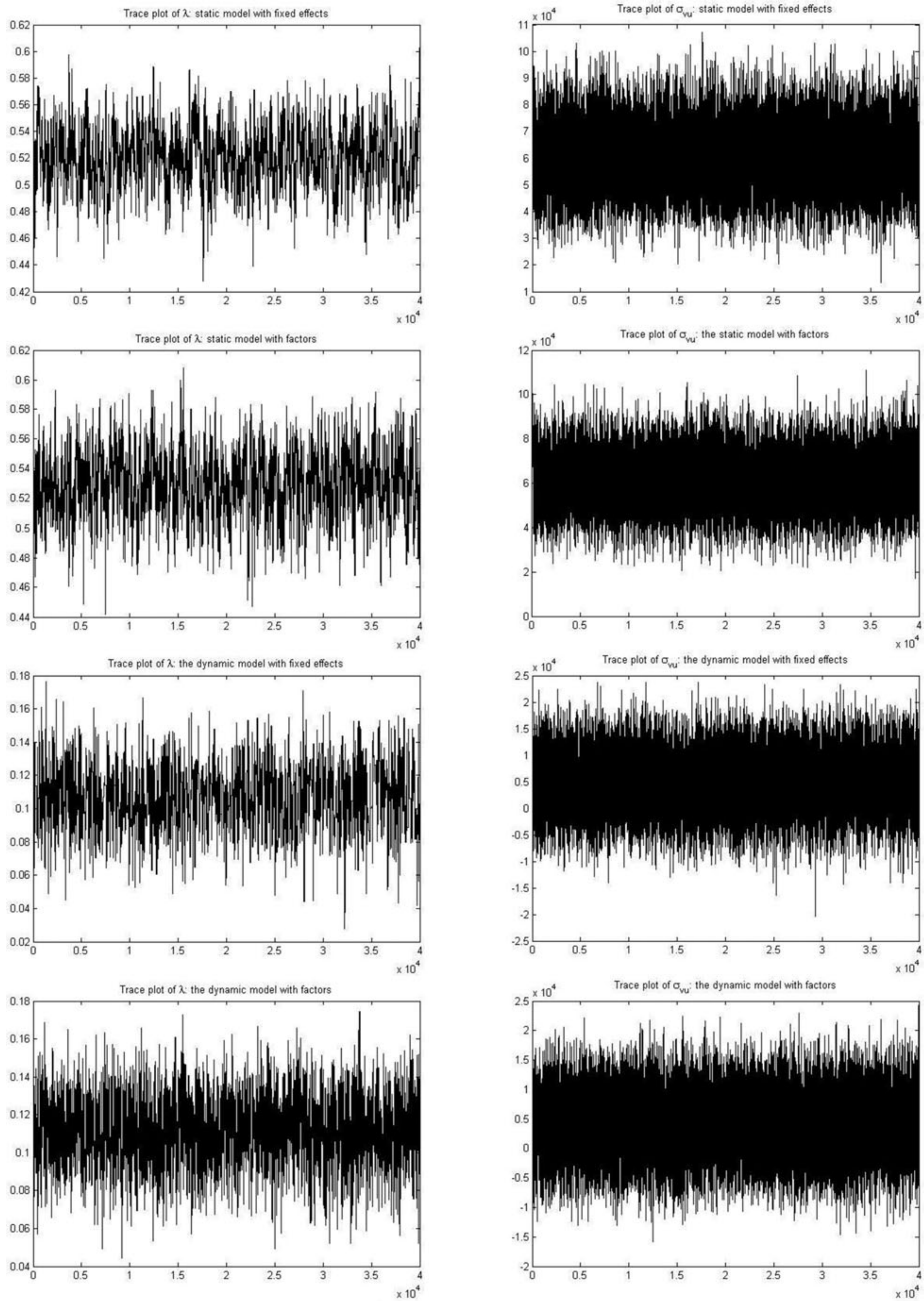

Figure 1. Graphs of simulation and empirical results (d) Trace plots of λ and σ_{vu} in empirical application: Medicaid related spending *(continued from previous page)*.

number of iterations exceed $2k$. The second component of the mixture $\mathcal{N}_k(\Psi^{(g-1)}, 0.1^2 I_k/k)$ would prevent us from generating singular covariance matrix due to some problematic values of $\mathrm{cov}(\Psi^{(0)}, \Psi^{(1)}, \ldots, \Psi^{(g-1)})$. δ is set to be 0.05 following Roberts and Rosenthal (2009). In our MCMC sampler, the AM step is only applied to the burn-in draws. After burn-in, we fix the value of the covariance matrix and use a normal proposal (with that covariance matrix) to continue the M-H step for Ψ.

Step 1: $\Psi|\{Y_{nt}\}, \{Z_{nt}\}, \Lambda, \Omega, \{f_t\}, \beta, \Sigma$.
By Bayes' theorem,

$$\pi(\Psi|\{Y_{nt}\}, \{Z_{nt}\}, \Lambda, \Omega, \{f_t\}, \beta, \Sigma) \propto \pi(\Psi) \times f(\{Y_{nt}\}|\{Z_{nt}\}, \Lambda, \Omega, \{f_t\}, \Psi, \beta, \Sigma), \tag{3.5}$$

where $f(\{Y_{nt}\}|\{Z_{nt}\}, \Lambda, \Omega, \{f_t\}, \Psi, \beta, \Sigma) \propto \prod_{t=1}^T |S_{nt}(\lambda)| \times \exp(-\frac{H'_{nt|f} H_{nt|f}}{2\sigma_\xi^2})$ with $H_{nt|f} = S_{nt}(\lambda) Y_{nt} - \overline{A}(\psi_1, \rho) Y_{n,t-1} - X_{n1t}\beta_1 - \Lambda f_t - (Z_{nt} - Y_{n,t-1}\psi_2 - X_{n2t}\beta_2 - \Omega \overline{f}_t)\eta$.
1.1. Propose $\tilde{\Psi} \sim \mathcal{N}_3(\Psi^{(j-1)}, A_\Psi)$, where $\Psi^{(j-1)}$ is the draw at the $(j-1)$th iteration, and the covariance matrix A_Ψ is determined using the AM algorithm.
1.2. With probability equal to

$$\Pr\left(\Psi^{(j-1)}, \tilde{\Psi}\right) = \min\left\{\frac{f(\{Y_{nt}\}|\{Z_{nt}\}, \Lambda^{(j-1)}, \Omega^{(j-1)}, \{f_t^{(j-1)}\}, \tilde{\Psi}, \beta^{(j-1)}, \Sigma^{(j-1)})}{f(\{Y_{nt}\}|\{Z_{nt}\}, \Lambda^{(j-1)}, \Omega^{(j-1)}, \{f_t^{(j-1)}\}, \Psi^{(j-1)}, \beta^{(j-1)}, \Sigma^{(j-1)})} \times \frac{\pi(\tilde{\Psi})}{\pi(\Psi^{(j-1)})}, 1\right\},$$

set $\Psi^{(j)}$ equal to $\tilde{\Psi}$, else set it equal to $\Psi^{(j-1)}$.

Step 2: $\beta|\{Y_{nt}\}, \{Z_{nt}\}, \Lambda, \Omega, \{f_t\}, \Psi, \Sigma$
Note that $\beta = (\beta_1', \psi_2, \mathrm{vec}(\beta_2)')'$. Let $\Omega_i' = (\Omega_{1,i}, \Omega_{2,i}, \ldots, \Omega_{p,i})'$ be the $p \times q$ collection of the ith row of Ω_l for $l = 1, 2, \ldots, p$. Let $y_{it|\beta f} = \begin{pmatrix} y_{it} - \lambda\sum_{j\neq i} w_{ij,t} y_{jt} - \psi_1 y_{i,t-1} - \rho\sum_{j\neq i} w_{ij,t-1} y_{j,t-1} - \Lambda_i' f_t \\ z_{it}' - \Omega_i' f_t \end{pmatrix}$ and $x_{it|\beta f} = \begin{pmatrix} x_{i1,t} & O & O \\ O & y_{i,t-1} I_p & I_p \otimes x_{i2,t} \end{pmatrix}$. By Bayes' theorem, $\pi(\beta|\{Y_{nt}\}, \{Z_{nt}\}, \Lambda, \Omega, \{f_t\}, \Psi, \Sigma) \sim \mathcal{N}_{k_1+p(1+k_2)}(T_{\beta|f}, \Sigma_{\beta|f})$, where $\Sigma_{\beta|f} = (B_O^{-1} + \sum_{t=1}^T \sum_{i=1}^n x_{it|\beta f}' \Sigma^{-1} x_{it|\beta f})^{-1}$ and $T_{\beta|f} = \Sigma_{\beta|f}(B_O^{-1}\beta_O + \sum_{t=1}^T \sum_{i=1}^n x_{it|\beta f}' \Sigma^{-1} y_{it|\beta f})$.

Step 3: $\Sigma|\{Y_{nt}\}, \{Z_{nt}\}, \Lambda, \Omega, \{f_t\}, \Psi, \beta$
By Bayes' theorem, $\pi(\Sigma|\{Y_{nt}\}, \{Z_{nt}\}, \Lambda, \Omega, \{f_t\}, \Psi, \beta) \sim \mathcal{IW}_{p+1}(\nu + nT, \sum_{t=1}^T \sum_{i=1}^n h_{it|f} h_{it|f}' + \nu R)$, where $h_{it|f} = y_{it|\beta f} - x_{it|\beta f}\beta$.

Step 4: $f_t|Y_{nt}, Z_{nt}, \Lambda, \Omega, \Psi, \beta, \Sigma$, for $t = 1, 2, \ldots, T$
Denote $y_{it|f} = \begin{pmatrix} y_{it} - \lambda\sum_{j\neq i} w_{ij,t} y_{jt} - \psi_1 y_{i,t-1} - \rho\sum_{j\neq i} w_{ij,t-1} y_{i,t-1} \\ -x_{i1t}\beta_1 \\ z_{it}' - y_{i,t-1}\psi_2' - (I_p \otimes x_{i2t})\mathrm{vec}(\beta_2) \end{pmatrix}$ and $\tilde{\Lambda}_i = (\Lambda_i, \Omega_i)'$. Applying Bayes' theorem, $\pi(f_t|Y_{nt}, Z_{nt}, \Lambda, \Omega, \Psi, \beta, \Sigma) \sim \mathcal{N}_q(T_{f_t}, \Sigma_{f_t})$ with $\Sigma_{f_t} = (I_q + \sum_{i=1}^n \tilde{\Lambda}_i' \Sigma^{-1} \tilde{\Lambda}_i)^{-1}$ and $T_{f_t} = \Sigma_{f_t} \sum_{i=1}^n \tilde{\Lambda}_i' \Sigma^{-1} y_{it|f}$.

Step 5: $\Lambda_i|\{Y_{nt}\}, \{Z_{nt}\}, \Omega, \{f_t\}, \Psi, \beta, \Sigma$, for $i = 1, 2, 3, \ldots, n$
Let $y_{it|\Lambda} = \begin{pmatrix} y_{it} - \lambda\sum_{j\neq i} w_{ij,t} y_{jt} - \psi_1 y_{i,t-1} - \rho\sum_{j\neq i} w_{ij,t-1} y_{i,t-1} \\ -x_{i1t}\beta_1 \\ z_{it}' - y_{i,t-1}\psi_2' - (I_p \otimes x_{i2t})\mathrm{vec}(\beta_2) - \Omega_i' f_t \end{pmatrix}$ and $\tilde{f}_{t|\Lambda} = (f_t, 0)'$. By Bayes' theorem, $\pi(\Lambda_i|\{Y_{nt}\}, \{Z_{nt}\}, \Omega, \{f_t\}, \Psi, \beta, \Sigma) \sim \mathcal{N}_q(T_{\Lambda_i}, \Sigma_{\Lambda_i})$, with $\Sigma_{\Lambda_i} = (\frac{1}{\tau_1} I_q + \sum_{t=1}^T \tilde{f}_{t|\Lambda}' \Sigma^{-1} \tilde{f}_{t|\Lambda})^{-1}$ and $T_{\Lambda_i} = \Sigma_{\Lambda_i}(\frac{1}{\tau_1}\Lambda_O + \sum_{t=1}^T \tilde{f}_{t|\Lambda}' \Sigma^{-1} y_{it|\Lambda})$.

Step 6: $\Omega_{l,i}|\{Y_{nt}\}, \{Z_{nt}\}, \Lambda, \Omega_i^{(-l)}, \{f_t\}, \Psi, \beta, \Sigma$, for $i = 1, 2, 3, \ldots, n$ and $l = 1, 2, \ldots, p$
Let $\Omega_i^{(-l)} = (\Omega_{1,i}, \ldots, \Omega_{l-1,i}, 0, \Omega_{l+1,i}, \ldots, \Omega_{p,i})$ represent the collection of the ith rows of all Ω_l's except $\Omega_{l,i}$. Denote $y_{it|\Omega_l} = \begin{pmatrix} y_{it} - \lambda\sum_{j\neq i} w_{ij,t} y_{jt} - \psi_1 y_{i,t-1} - \rho\sum_{j\neq i} w_{ij,t-1} y_{i,t-1} - x_{i1t}\beta_1 - \Lambda_i' f_t \\ z_{it}' - y_{i,t-1}\psi_2' - (I_p \otimes x_{i2t})\mathrm{vec}(\beta_2) - \Omega_i^{(-l)'} f_t \end{pmatrix}$ and $\tilde{f}_{t|\Omega_l} = (0, \ldots, f_t, \ldots, 0)'$. By Bayes' theorem, $\pi(\Omega_{l,i}|\{Y_{nt}\}, \{Z_{nt}\}, \Lambda, \Omega_i^{(-l)}, \{f_t\}, \Psi, \beta, \Sigma) \sim \mathcal{N}_q(T_{\Omega_{l,i}}, \Sigma_{\Omega_{l,i}})$ with $\Sigma_{\Omega_{l,i}} = (\frac{1}{\tau_2} I_q + \sum_{t=1}^T \tilde{f}_{t|\Omega_l}' \Sigma^{-1} \tilde{f}_{t|\Omega_l})^{-1}$ and $T_{\Omega_{l,i}} = \Sigma_{\Omega_{l,i}}(\frac{1}{\tau_2}\Omega_O + \sum_{t=1}^T \tilde{f}_{t|\Omega_l}' \Sigma^{-1} y_{it|\Omega_l})$.

Note that Steps 4–6 can be simplified if one's main interest is not the identification of Λ, Ω, and f_t's. For instance, in Steps 4 and 5, instead of sampling Λ and f_t's, one can sample unit-specific random components $\alpha_{1it} = \Lambda_i' f_t$ for $i = 1, 2, \ldots n$ and $t = 1, 2, \ldots, T$. Hierarchical priors can be assumed for α_{1it}'s as $\alpha_{1it} \sim \mathcal{N}(\overline{\alpha}_1, A_1)$, $\overline{\alpha}_1 \sim \mathcal{N}(\alpha_{10}, A_{10})$, and $A_1 \sim \mathcal{IW}(g_1, G_1)$ and Gibbs sampling steps can be adopted to draw α_{1it}'s, $\overline{\alpha}_1$ and A_1.

The MCMC algorithm for restricted models, such as the static model with additive individual and time fixed effects or common factors, and the SDPD model with additive individual and time fixed effects can be similarly derived. Specifically, for the SDPD model with fixed effects, most of the sampling steps are similar to the steps outlined above, except to replace the simulation of factors and factor loadings with individual and time fixed effects. Let $c_i = (c_{i1}, c_{i2}')'$ be the $(p+1) \times 1$ vector of i's individual fixed effects and $\alpha_t = (\alpha_{1t}, \alpha_{2t}')'$ be the $(p+1) \times 1$ vector of time fixed effects. Assume $c_i \sim \mathcal{N}_{p+1}(c_O, C_O)$ for $i = 1, 2, \ldots, n$ and $\alpha_t \sim \mathcal{N}_{p+1}(\alpha_O, A_O)$ for $t = 1, 2, \ldots, T$. Denote $y_{it}^* = \begin{pmatrix} y_{it} - \lambda\sum_{j\neq i} w_{ij,t} y_{jt} - \psi_1 y_{j,t-1} - \rho\sum_{j\neq i} w_{ij,t-1} y_{j,t-1} \\ z_{it}' - y_{i,t-1}\psi_2' - (I_p \otimes x_{i2t})\mathrm{vec}(\beta_2) \end{pmatrix}$, and define $y_{it|c}$ and $y_{it|\alpha}$ as y_{it}^* subtracts $(c_{1i}, c_{2i}')'$ and $(\alpha_{1t}, \alpha_{2t}')'$'s, respectively. Then the conditional posterior distribution of c_i's and α_t's are, respectively, $\pi(c_i|\{Y_{nt}\}, \{Z_{nt}\}, \{\alpha_t\}, \Psi, \beta, \Sigma) \sim \mathcal{N}_{p+1}(T_{c_i}, \Sigma_{c_i})$, $i = 1, 2, \ldots, n$, and $\pi(\alpha_t|Y_{nt}, Z_{nt}, C_n, \Psi, \beta, \Sigma) \sim \mathcal{N}_{p+1}(T_{\alpha_t}, \Sigma_{\alpha_t})$, $t = 1, 2, \ldots, T$, where $\Sigma_{c_i} = (C_O^{-1} + \sum_{t=1}^T x_{it|c}' \Sigma^{-1} x_{it|c})^{-1}$, $T_{c_i} = \Sigma_{c_i}(C_O^{-1} c_O + \sum_{t=1}^T x_{it|c} \Sigma^{-1} y_{it|c})$, $\Sigma_{\alpha_t} = (A_O^{-1} + \sum_{i=1}^n x_{it|a}' \Sigma^{-1} x_{it|a})^{-1}$, and $T_{\alpha_t} = \Sigma_{\alpha_t}(A_O^{-1}\alpha_O + \sum_{i=1}^n x_{it|a}' \Sigma^{-1} y_{it|a})$. For the static models, by setting $\rho = \psi_1 = 0$ and $\psi_2 = 0$, one can follow the same sampling steps as the dynamic model.

The MCMC algorithms of the random coefficient model in (2.3) can be also derived based upon the above algorithm. Here we assume truncated normal priors between $(-1, 1)$ for λ_i's, ρ_i's, and ψ_{1i}'s. Unlike the above sampling step of Ψ, we do not impose the stability condition in Section 2.2 on them through those priors. The stability condition would be imposed at the AM step for $(\{\lambda_i\}, \{\rho_i\}, \{\psi_{1i}\})$. The means and variances of those truncated normal priors are assumed hierarchical normal and inverse-Wishart priors. The priors of $\beta_i = (\beta'_{1i}, \psi'_{2i,1}, \dots, \psi'_{2i,p}, \beta'_{2i,1}, \dots, \beta'_{2i,p})'$'s, Λ_i's, and $\Omega_{l,i}$'s are multivariate normal, with their means and covariance matrices further assumed, respectively, by normal and inverse-Wishart priors. The AM algorithm is used to sample $(\{\lambda_i\}, \{\rho_i\}, \{\psi_{1i}\})$ as well as their prior means and variances. Gibbs sampling steps are adopted to draw β_i's, Λ_i's, $\Omega_{l,i}$'s and their prior means and covariance matrices.

4. MODEL SELECTION

For models with common factors, the previous MCMC algorithm is based on the assumption that q, the number of factors, is correctly specified. So a fully Bayesian analysis would naturally call for model selection procedures to determine q. For model comparison with two competing models, one way is to compute a corresponding Bayes factor. Bayes factor can be applied to study cases where competing models are nested or nonnested, Kass and Raftery (1995). However, with many parameters in a model, it may be difficult to calculate the corresponding Bayes factor. An alternative approach is to adopt the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002). DIC is the Bayesian version of the Akaike information criterion (AIC). It can be directly calculated from the MCMC sampler and is almost trivial to obtain. Thus, we rely on DIC to detect the number of common factors. The implementation of DIC uses posterior MCMC draws. However, since the factors and factor loadings in our model are not separately identified, posterior draws of f_t's, Λ, and Ω_l's might suffer from poor-mixing. This would make estimates of their posterior means and standard deviations unreliable, Chan and Tobias (2015), and may cause computational problems of the above algorithm and criterion. Therefore, to do model selection, identification of f_t, Λ, and Ω_l's need to be adjusted first.

4.1 Identification of Common Factors and Factor Loadings

4.1.1 Bayesian Literature on Identification of Factors and Factor Loadings. Common factors and factor loadings are not separately identified because $\Lambda f_t = \Lambda Q Q' f_t$ and $\Omega \overline{f}_t = \tilde{\Omega}_t$ from Equation (2.2), where $\tilde{\Omega}_t = (\Omega_1 Q Q' f_t, \Omega_2 Q Q' f_t, \dots, \Omega_p Q Q' f_t)$ and Q is an orthogonal matrix. To determine Q, $\frac{q(q-1)}{2}$ restrictions need be imposed on factor loadings, in addition to the normalization restriction on f_t's. To impose the $\frac{q(q-1)}{2}$ restrictions, the most popular approach in the Bayesian literature is to set the upper-diagonals of the first $q \times q$ submatrix in the loading matrix to zero, Geweke and Zhou (1996). However, this identification scheme requires the knowledge of some variable ordering, that is, which variable is only affected by the first factor and which variable is affected by the first two factors and so on, Bai and Ng (2013). So the corresponding estimation and inference might be influenced by the way that the cross-sectional units are ordered. Lopes and West (2004) found that the selection of number of factors is affected by the ordering of the cross-sectional units.

To tackle this issue, recently, researchers have tried to propose some "order-independent" identification schemes on factors and factor loadings. Kaufmann and Schumacher (2012) focused on sparse factor models, where certain factor loadings are set to be zeros. They attempted to identify zero rows in the factor loadings, with which those cross-sectional units are not affected by common factors. Strachan, Chan, and Leon-Gonzalez (2013) considered an "ex-ante" approach. They viewed the factor model as a "reduced-rank" model and imposed restrictions on priors of factors and factor loadings. By parameter expansion, they can obtain relatively simpler priors and derived the MCMC sampler. Aßmann, Boysen-Hogrefe, and Pape (2012) investigated an "ex-post" approach. Instead of imposing restrictions on priors, they borrowed the idea of the relabeling algorithm from Bayesian finite mixture models, Stephens (2000), and chose to post-screen the unconstrained MCMC sampler. Identification is achieved based on Procrustes transformation from Schönemann (1966). This is the approach we follow.

4.1.2 The Ex-Post Identification Scheme in Aßmann, Boysen-Hogrefe, and Pape (2012). Consider the identification of Λ and f_t's in the SAR equation. According to Aßmann, Boysen-Hogrefe, and Pape (2012), the unconstrained posterior sampler gives an orthogonally mixing samples of Λ and f_t's. Let $Q^{(s)}$ be the $q \times q$ unknown orthogonal transformation at iteration s. Also let $\Lambda^{(s)}$ be the draw of Λ at iteration s. Without restrictions, $\Lambda^{(s)}$ is subjected to an unknown orthogonal transformation of $\Lambda^{(s)} Q^{(s)}$. So as long as we can pin down all $Q^{(s)}$'s based upon some minimization criterion, identification is reached. The minimization problem considered is to determine a set of orthogonal matrices $Q^{(s)}$'s and a fixed point Λ^* (not dependent on s) as $\{\{Q^{(s)}\}, \Lambda^*\} =$ argmin $\sum_{s=1}^{S} \mathrm{tr}[(\Lambda^{(s)} Q^{(s)} - \Lambda^*)'(\Lambda^{(s)} Q^{(s)} - \Lambda^*)]$. Given an initial choice of Λ^*, the minimization is derived iteratively by a two-step optimization. Note that the above minimization problem gives the same solution if $\{Q^{(s)}\}_{s=1}^{S}$ and Λ^* are transformed by the same orthogonal matrix. But this orthogonal matrix would only change the orientations of $\{Q^{(s)}\}_{s=1}^{S}$ and Λ^* . Thus, identification can still be achieved up to orientation. With $Q^{(s)}$, the unconstrained draws of $\Lambda^{(s)}$, $f_t^{(s)}$'s, and $\Omega_l^{(s)}$'s can be transformed accordingly, to obtain identified draws as $\Lambda^{(s)} Q^{(s)}$, $Q^{(s)'} f_t^{(s)}$ for $t = 1, 2, \dots, T$ and $\Omega_l^{(s)} Q^{(s)}$ for $l = 1, 2, \dots p$. Let B be any $q \times q$ orthogonal matrix. Then the orientations of those identified draws can be changed to, $\Lambda^{(s)} Q^{(s)} B$, $B' Q^{(s)'} f_t^{(s)}$ for $t = 1, 2, \dots, T$, and $\Omega_l^{(s)} Q^{(s)} B$ for $l = 1, 2, \dots p$, by the same matrix B at each iteration. This orthogonal transformation would not affect the likelihood function because Λ, Ω_l's, and f_t's appear as products in it. The density of f_t's would not be influenced as well. The priors of Λ and Ω_l's are invariant under the transformation through B as long as their prior means are set to be zeros. Therefore, with prior means being zeros for Λ and Ω_l's, the orthogonal transformation by B would not give trouble to the computation of DIC. Below is the relabeling algorithm.

Step 0: Set the initial values of $\Lambda^{(s)}$, $f_t^{(s)}$, and $\Omega_l^{(s)}$ to be their unconstrained MCMC draws at iteration s. Let the initial value of Λ^* be the last draw of the unconstrained MCMC sample of Λ.

Step 1: Conditional on Λ^* and $\Lambda^{(s)}$, solve the following minimization problem for $Q^{(s)}$:

$$\min \operatorname{tr}\left[(\Lambda^{(s)}Q^{(s)} - \Lambda^*)'(\Lambda^{(s)}Q^{(s)} - \Lambda^*)\right], \quad \text{such that } Q^{(s)}Q^{(s)'} = I_q. \tag{4.1}$$

The solution of this orthogonal procrustes problem is discussed in Schönemann (1966). It involves the following sub-steps:

1.1. Let $A_2^{(s)} = \Lambda^{(s)'}\Lambda^*$.

1.2. Conduct the eigenvalue decomposition $A_2^{(s)'}A_2^{(s)} = K_s F_s K_s'$.

1.3. Conduct the eigenvalue decomposition $A_2^{(s)}A_2^{(s)'} = J_s F_s J_s'$.

1.4. Denote R_s as a $q \times q$ diagonal matrix, with $+1$ or -1 on the diagonal. Find the unique R_s, such that the q main diagonals of $(J_s R_s)' F_s K_s$ are nonnegative.

1.5. Derive the orthogonal transformation matrix $Q^{(s)} = (J_s R_s) K_s'$.

1.6. Transform $\Lambda^{(s)}$, $f_t^{(s)}$'s, and $\Omega_l^{(s)}$'s as $\tilde{\Lambda}^{(s)} = \Lambda^{(s)}Q^{(s)}$, $\tilde{f}_t^{(s)} = Q^{(s)'}f_t^{(s)}$ for $t = 1, 2, \ldots, T$, and $\tilde{\Omega}_l^{(s)} = \Omega_l^{(s)}Q^{(s)}$ for $l = 1, 2, \ldots, p$.

Step 2: Conditional on $Q^{(s)}$ and $\Lambda^{(s)}$, derive $\Lambda^* = (\Lambda_1^*, \Lambda_2^*, \ldots, \Lambda_n^*)'$ as $\Lambda_i^* = \frac{1}{S}\sum_{i=1}^{S}\tilde{\Lambda}_i^{(s)} = \frac{1}{S}\sum_{i=1}^{S}\Lambda_i^{(s)} Q^{(s)}$, $i = 1, 2, \ldots, n$. Set $\Lambda^{(s)} = \tilde{\Lambda}^{(s)}$, $f_t^{(s)} = \tilde{f}_t^{(s)}$ for $t = 1, 2, \ldots, T$, and $\Omega_l^{(s)} = \tilde{\Omega}_l^{(s)}$ for $l = 1, 2, \ldots, p$. Go back to Step 1.

These two steps goes iteratively until convergence to a fixed point Λ^* is reached. The corresponding $\Lambda^{(s)}$, $f_t^{(s)}$'s, and $\Omega_l^{(s)}$'s would become identified draws at iteration s. In particular, let $\Lambda_u^{(s)}$, $f_{t|u}^{(s)}$'s, and $\Omega_{l|u}^{(s)}$'s be the unconstrained MCMC draws at iteration s. If we run the above two steps for G times, let $Q_j^{(s)}$ be the derived orthogonal transformation matrix at time j, for iteration s. Then $\tilde{Q}^{(s)} = \prod_{j=1}^{G} Q_j^{(s)}$ would be the orthogonal transformation matrix at iteration s. Thus, $\Lambda_G^{(s)} = \Lambda_u^{(s)}\tilde{Q}^{(s)}$, $f_{t|G}^{(s)} = \tilde{Q}^{(s)'}f_{t|u}^{(s)}$ for $t = 1, 2, \ldots, T$ and $\Omega_{l|G}^{(s)} = \Omega_{l|u}^{(s)}\tilde{Q}^{(s)}$ for $l = 1, 2, \ldots, p$, with $\Lambda_G^{(s)}$, $f_{t|G}^{(s)}$'s, and $\Omega_{l|G}^{(s)}$'s being the identified draws at iteration s, after G times of the relabeling algorithm. In the following simulation study, trace plots of some elements of $\Lambda_u^{(s)}$, $f_{t|u}^{(s)}$'s, $\Omega_{l|u}^{(s)}$'s, $\Lambda_G^{(s)}$, $f_{t|G}^{(s)}$'s, and $\Omega_{l|G}^{(s)}$'s are depicted in Figure 1(b), to demonstrate that identification is indeed achieved.

4.2 Model Selection by Deviance Information Criterion (DIC)

The deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002) can be interpreted as the Bayesian version of Akaike information criterion (AIC) in Akaike (1974). Like AIC, it trades off a measure of model fit against a measure of complexity. Let $\mathbf{Y}_{nt}$ represent the data and Θ represent the "model parameter vector." For models with fixed effects, $\Theta = (\theta, C_n, \{\alpha_t\})$. For models with common factors, $\Theta = (\theta, \Lambda, \{\Omega_l\}, \{f_t\})$ and with random coefficients $\Theta = (\{\lambda_i\}, \{\psi_{1i}\}, \{\rho_i\}, \overline{\beta}_1, \overline{\beta}_2, \Sigma, \Lambda, \{\Omega_l\}, \{f_t\})$. By treating latent factors f_t's as parameters, we are actually using the "conditional DIC" in Celeux et al. (2006). The deviance in Spiegelhalter et al. (2002) is given by $D(\Theta) = -2\ln f(\{\mathbf{Y}_{nt}\}|\Theta) + 2\ln g(\{\mathbf{Y}_{nt}\})$, where $g(\{\mathbf{Y}_{nt}\})$ is some fully specified standardizing term, which is a function of $\{\mathbf{Y}_{nt}\}$. We follow Berg, Meyer, and Yu (2004) to set $g(\{\mathbf{Y}_{nt}\}) = 1$, so $D(\Theta)$ is simply the minus 2 log-likelihood function. DIC has two components: $\mathrm{DIC}_1 = \overline{D} + P_D$. The first component, $\overline{D} = E_{\Theta|\{\mathbf{Y}_{nt}\}}[D(\Theta)] = E_{\Theta|\{\mathbf{Y}_{nt}\}}(-2\ln f(\{\mathbf{Y}_{nt}\}|\Theta))$ is defined as a Bayesian measure of model fit. According to Gelman et al. (2014), it is "reasonable" to use the expected log-likelihood to measure the overall model fit. The log-likelihood (log predictive density) has connections with the Kullback–Leibler information measure. When sample size goes to infinity, the model with the lowest Kullback–Leibler information measure has the highest expected log-likelihood. So it will also have the highest posterior probability. Hence, the better the model fit the data, the larger the value of the expected log-likelihood, and the smaller $\overline{D}$ is. The second component $P_D = \overline{D} - D(\overline{\Theta}) = E_{\Theta|\{\mathbf{Y}_{nt}\}}(-2\ln f(\{\mathbf{Y}_{nt}\}|\Theta)) + 2\ln f(\{\mathbf{Y}_{nt}\}|\overline{\Theta})$, with $\overline{\Theta}$ being the posterior mean of Θ, is the effective number of parameters and a measure of model complexity. According to Berg, Meyer, and Yu (2004), $-2\ln f(\{\mathbf{Y}_{nt}\}|\Theta)$ is the residual information in the data, conditional on Θ; hence it can be regarded as a measure of uncertainty. Then P_D can be interpreted as the reduction in uncertainty due to estimation. Note that P_D might be negative when there were substantial conflict between the prior and the data, or poor estimates of posterior mean, Spiegelhalter et al. (2002). An alternative measure of effective number of parameters, advocated by Gelman et al. (2014), is $P_{DV} = 2 \times \mathrm{var}_{\Theta|\{\mathbf{Y}_{nt}\}}(\ln f(\{\mathbf{Y}_{nt}\}|\Theta))$, which is just two times the variance of posterior draws of the log-likelihood function. Compared with P_D, it has the advantage of always being positive. Thus, an alternative version of DIC is $\mathrm{DIC}_2 = \overline{D} + P_{DV}$. As with AIC, the smaller DIC, the better the model is. The model with a smaller DIC is the one that gives a better prediction of future data, generated by the same mechanism that gave rise to the observed data, Spiegelhalter et al. (2002). Compared with the Bayes factor, DIC is almost trivial to compute with MCMC draws and does not involve the priors. So it is very useful for models with many parameters or latent variables. Applications of DIC in model selection issues can be found, for example, in Huang and Yu (2010) and Berg, Meyer, and Yu (2004). Here, we apply DIC to select the number of f_t's. Additionally, it is used to select various specifications of W_{nt} in the empirical application. For the model with common factors, to calculate DIC_1, we may use identified posterior draws of f_t's, Λ, and Ω_l's to evaluate $D(\overline{\Theta})$. Alternatively, one can use the unidentified draws of f_t's, Λ, and Ω_l's to calculate the posterior mean of the common component Λf_t and $\Omega \overline{f}_t$, which are invariant to factor rotation, to evaluate $D(\overline{\Theta})$. Similarly, for DIC_2, we do not need to use the identified draws because only the log-likelihood function is present in its calculation and, factor and factor loadings always show up as products in the likelihood function. This further simplifies the model selection procedure.

5. SIMULATION STUDY

5.1 Monte Carlo Simulation Setup

We apply the Bayesian estimation algorithm and model selection procedure outlined above to simulated datasets. The study consists of two parts. In the first part, we focus on the performance of the proposed Bayesian MCMC algorithm for SAR models with an endogenous W_{nt} and common factors. In the second part, we use DIC to deal with model selection with different numbers of common factors. The models considered are the static model with common factors

$$Y_{nt} = \lambda W_{nt} Y_{nt} + X_{n1t}\beta_1 + X_{n2t}\beta_2 + \Lambda f_t + V_{nt},$$
$$Z_{nt} = X_{n1t}\beta_3 + \Omega f_t + U_{nt},\ t = 1, 2, \ldots, T, \tag{5.1}$$

and that with time and effect effects, that is, $\Lambda f_t = C_{n1} + l_n \alpha_{1t}$ and $\Omega f_t = l_n' \otimes C_{n2} + l_n \alpha_{2t}$. The column dimension of Z_{nt}, p, is set to be 1. The column dimensions of X_{n1t} and X_{n2t} are also 1. In both models, the only exogenous regressor in the entry equation is X_{n1t}. We also investigate the case where the entry equation contains both X_{n1t} and X_{n2t}. The estimation results do not change much, so not reported here. The number of repetitions for all experiments is 50. In each repetition, the number of cross-sectional units is set to be 50 and the time length is 18. The W_{nt} is constructed as

$$w_{ij,t} = \frac{w_{ij,t}^b}{\sum_{k=1}^n w_{ik,t}^b},\ i \neq j,\ w_{ii,t} = 0,\ w_{ij,t}^b = \gamma_{ij} \times \tilde{w}_{ij,t}^b,$$
$$\tilde{w}_{ij,t}^b = d_{ij}^{-\phi_0} \times E_{ij,t}^{-\phi_1},\ t = 1, 2, \ldots, T, \tag{5.2}$$

where d_{ij} refers to the geographical distance and $E_{ij,t} = |z_{it} - z_{jt}|$ represents the economic distance. Specifically, d_{ij} is taken as $d_{ij} = \frac{1}{6} \times \tilde{d}_{ij},\ \tilde{d}_{ij} < 1;\ d_{ij} = \tilde{d}_{ij},\ 1 < \tilde{d}_{ij} < 2;\ d_{ij} = 6 \times \tilde{d}_{ij},\ \tilde{d}_{ij} > 2$, where $\tilde{d}_{ij} = \sqrt{(xc_i - xc_j)^2 + (yc_i - yc_j)^2}$ with (xc_i, yc_i) and (xc_j, yc_j) being the coordinates of i and j. Moreover, γ_{ij} is a prespecified $0-1$ indicator. All γ_{ij}'s are generated by the function "makeneighborsw," taken from LeSage's matlab codes for spatial econometrics. This function generates a row-normalized matrix Γ_n, with the (i,j)th element $\gamma_{ij} = 1$, based upon the coordinates of i and j being the m-nearest-neighbors; $\gamma_{ij} = 0$ otherwise. The number of nearest neighbors m is 5. Furthermore, for prespecified ϕ_0 and ϕ_1 in the construction of $\tilde{w}_{ij,t}^b$ in Section 2, three values are considered: DGP 1: $\phi_0 = 1;\ \phi_1 = 1;$ DGP 2: $\phi_0 = 2;\ \phi_1 = 1;$ DGP 3: $\phi_0 = 1;\ \phi_1 = 2$. The true values of parameters are $\lambda = 0.6; \beta_1 = 1; \beta_2 = 1; \beta_3 = 1$. The bivariate disturbances (v_{it}, u_{it}) are generated from iid $N(0, (\begin{smallmatrix}\sigma_v^2 & \sigma_{vu}' \\ \sigma_{vu} & \sigma_u^2\end{smallmatrix}))$, in which $\sigma_v^2 = 1$ and $\sigma_u^2 = 1$. To have different levels of endogeneity, $\sigma_{vu} = 0.3$, 0.5, and 0.8 are set. X_{n1t} and X_{n2t} are generated from $\mathcal{N}(0, 2)$. The time fixed effects α_{1t} and α_{2t} are generated from $\mathcal{N}(0, 1)$ for all t. The individual fixed effects C_{n1} and C_{n2} are generated according to Mundlak (1978), namely, $c_{i1} = 0.3\overline{X}_{i1.} + 0.3\overline{X}_{i2.} + \epsilon_1,\ c_{i2} = 0.3\overline{X}_{i1.} + \epsilon_2$, where $\overline{X}_{i1.}$ and $\overline{X}_{i2.}$ represent, respectively, the empirical time-averages of X_{i1t}'s and X_{i2t}'s; ϵ_1 and ϵ_2 are generated from $\mathcal{N}(0, 0.05)$.

For models with common factors, the dimension of factor q is set to be 2. With $p = 1$, the factor loading Ω in Equation (5.1) is an $n \times 2$ matrix. f_t's are generated independently from $\mathcal{N}(0, I_2)$. Elements of Λ and Ω are generated as $\Lambda_{i,j} = 4\overline{X}_{i1.} + 4\overline{X}_{i2.} + \epsilon_\Lambda,\ \Omega_{i,j} = 2\overline{X}_{i1.} + \epsilon_\Omega$, for $j = 1, 2$, where ϵ_Λ and ϵ_Ω are independently sampled from $\mathcal{N}(0, 0.05)$.

The Bayesian MCMC algorithm outlined in Section 3 is implemented. The values of prior parameters are: $\beta_O = 0,\ B_O = 10 \times I_{k_1+pk_2};\ \nu = 8,\ R = I_{p+1}, c_O = 0,\ C_O = 10 \times I_{p+1};\ \alpha_O = 0,\ A_O = 10 \times I_{p+1}$, and $\Lambda_O = 0,\ \tau_1 = 10;\ \Omega_O = 0,\ \tau_2 = 10$. The length of the Markov chain is 20,000 for all models. We apply the method in Raftery and Lewis (1992) to determine the adequate length of our MCMC sampler. The first 50% draws of each chain are burned in. Some trace plots of the Bayesian estimates of λ, β_1 σ_{vu} are in Figures 1(c), to demonstrate the convergence of the MCMC sampler.

When conducting model selection, we focus on DGP1 for W_{nt}, where $\phi_0 = 1$ and $\phi_1 = 1$. The data-generating process of most parameters, fixed effects, and common factors are the same as above, except that, for factor loadings, the variances of ϵ_Λ and ϵ_Ω are set, respectively, to be 4 or 0.05. The number of repetitions is still 50 for all experiments. To select the number of common factors, the cases with $\sigma_{vu} = 0.3$ and $\sigma_{vu} = 0.5$ are investigated. We rely on DIC_2 because for DIC_1, we encounter some negative P_D's when the DGP has fewer true factors but estimated with more factors. As suggested by Spiegelhalter et al. (2002), substantial conflict between the prior and the data could lead to negative P_D's. Therefore, only DIC_2 is adopted. The mean and standard deviation of DIC_2 and the model frequencies in which the true models are selected are reported.

5.2 Results

Table 1 summarizes the estimation results for both the static model with fixed effects and common factors. As shown, the Bayesian estimates of λ, all β's, σ_v^2, σ_{vu}, and σ_u^2 are close to their true values with small standard deviations. The MCMC sampler is able to recover the spatial parameter λ, with different levels of endogeneity for W_{nt}.

Table 2 collects the model selection results for detecting the number of common factors, by DIC_2. When the true model has more than one common factors and variations in Λ and Ω are small, DIC_2 tend to favor the model with fewer common factors than the true one. However, as variations in Λ and Ω become larger, the performance of DIC_2 turns remarkably good in all cases.

6. EMPIRICAL APPLICATION

We apply the model to study spillover effects of state Medicaid related spending. We are interested in two possible sources of those spillovers: welfare motivated move and yardstick competition. Geographical distance is related to welfare motivated move while economic distance is related to yardstick competition. So three kinds of spatial weights matrices are explored: one constructed from geographical distance, one constructed from economic distance, and the other constructed from both distances. We would like to see whether state governments respond to their geographically close or economically similar neighbors or not.

6.1 Data

Data on the 48 contiguous states from 1983 to 2002 are used. The primary focus of our study is the spillovers of the Medicaid spending across states. The dependent variable Y_{nt} in the

Table 1. Estimation of the static model

		$\sigma_{vu}=0.3$				$\sigma_{vu}=0.5$				$\sigma_{vu}=0.8$			
		Fixed effects		Time factors		Fixed effects		Time factors		Fixed effects		Time factors	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
DGP1	λ	0.59	0.01	0.60	0.01	0.59	0.01	0.60	0.01	0.59	0.01	0.60	0.01
	β_1	1.01	0.02	1.00	0.02	1.01	0.02	1.00	0.02	1.00	0.02	1.00	0.02
	β_2	1.00	0.03	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.01
	β_3	1.00	0.03	1.00	0.02	1.00	0.03	1.00	0.02	1.00	0.03	1.00	0.02
	σ_v^2	1.02	0.04	1.01	0.05	1.01	0.05	1.02	0.05	1.01	0.05	1.02	0.05
	σ_{vu}	0.31	0.04	0.31	0.04	0.50	0.04	0.51	0.04	0.80	0.04	0.81	0.04
	σ_u^2	1.00	0.05	1.01	0.05	1.00	0.05	1.01	0.05	1.01	0.05	1.01	0.05
DGP2	λ	0.59	0.01	0.60	0.01	0.60	0.01	0.60	0.01	0.60	0.01	0.60	0.01
	β_1	1.01	0.02	1.00	0.03	1.00	0.03	1.00	0.03	1.00	0.02	1.00	0.02
	β_2	1.00	0.03	1.00	0.03	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.01
	β_3	1.00	0.03	1.00	0.02	1.00	0.03	1.00	0.02	1.00	0.03	1.00	0.02
	σ_v^2	1.01	0.04	1.01	0.05	0.99	0.04	1.02	0.05	1.01	0.05	1.02	0.05
	σ_{vu}	0.31	0.04	0.29	0.03	0.49	0.03	0.51	0.04	0.80	0.04	0.80	0.04
	σ_u^2	1.00	0.05	1.00	0.05	0.99	0.05	1.01	0.05	1.00	0.05	1.01	0.05
DGP3	λ	0.59	0.01	0.60	0.01	0.59	0.01	0.60	0.01	0.60	0.01	0.60	0.00
	β_1	1.01	0.02	1.00	0.03	1.01	0.02	1.00	0.02	1.00	0.02	1.00	0.02
	β_2	1.00	0.03	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.01
	β_3	1.00	0.03	1.00	0.02	1.00	0.03	1.00	0.02	1.00	0.03	1.00	0.02
	σ_v^2	1.02	0.04	1.02	0.05	1.02	0.05	1.02	0.05	1.02	0.05	1.02	0.05
	σ_{vu}	0.31	0.04	0.31	0.04	0.51	0.04	0.51	0.04	0.80	0.05	0.80	0.04
	σ_u^2	1.00	0.05	1.01	0.05	1.01	0.05	1.01	0.05	1.00	0.05	1.01	0.05

NOTE: $(\lambda, \beta_1, \beta_2, \beta_3, \sigma_v^2, \sigma_u^2) = (0.6, 1, 1, 1, 1, 1)$; DGP1: $(\phi_0, \phi_1) = (1, 1)$; DGP2: $(\phi_0, \phi_1) = (2, 1)$; DGP3:$(\phi_0, \phi_1) = (1, 2)$.

SAR outcome equation is vendor payments for medical care for each state (named "Medicaid related spending" hereafter), which consists mostly of Medicaid spending and spending on state children's health insurance program. Data on vendor payments for medical care comes from the Survey and Census of Government Finances and is classified as a sub-category under public welfare spending. The dependent variable Z_{nt} in the entry equation is state personal income, which can be obtained from the website of the Bureau of Economic Activity (BEA). All dependent variables are expressed in real per capita, with the base year 1992.

Variables in X_{n1t} are: population, population density, percentage of population under age 16 and over 65, per capita federal grants, lagged unemployment rate, and poverty rate. Population, population density, and percentage of population under age 16 and over 65 are included to control for the demographic characteristics of a state while per capita federal grant aims to capture the resource available for state spending. Lagged unemployment rate and poverty rate are used to control for state specific economic conditions. Many empirical literatures on spillovers of state spending, that is, Baicker (2005), also include Z_{nt} as an explanatory variable in the SAR equation. Variables included in X_{n2t} to explain state personal income Z_{nt} are percentage of population under age 16, between age 16 and 65, and 1-year lagged state net capital stock. As for the data source, per capita federal grant comes from the Survey and Census of Government Finances. Population data are obtained from the BEA. Unemployment rate and poverty rate are collected from the websites of The University of Kentucky Center for Poverty Research. Finally, state net capital stock data comes from the estimates of Garofalo and Yamarik (2002) and Yamarik (2013). Table 3 presents summary statistics of the data.

6.2 Empirical Specification

The spatial weights matrix W_{nt} is specified as

$$w_{ij,t} = \frac{w^b_{ij,t}}{\sum_{k=1}^n w^b_{ik,t}};\ w^b_{ij,t} = \gamma_{ij} \times d_{ij}^{-\phi_0} \times E_{ij,t}^{-\phi_1},\ i \neq j;$$
$$w_{ii,t} = 0,\ t = 1, 2, \ldots, T. \tag{6.1}$$

The specifications of the binary indicator γ_{ij}, the geographical distance d_{ij}, and economic distance $E_{ij,t}$ all follow the common practice in the literature. Two ways of defining γ_{ij} are considered. One corresponds to geographic contiguity: if i and j are bordering states, $\gamma_{ij} = 1$; otherwise 0. The other is simply set all $\gamma_{ij} = 1$ as long as $i \neq j$. d_{ij} is the geographical distance between the capital cities of state i and state j. The economic distance $E_{ij,t} = |z_{i,t} - z_{j,t}|$ is the difference between states i and j's income per capita at time t. The prespecified ϕ_0 and ϕ_1 are the weights of geographical and economic distances. In other words, they are the weights of two possible sources of strategic interactions among governments: welfare motivated move and yardstick competition. Three sets of values for ϕ_0 and ϕ_1 are considered: $(\phi_0, \phi_1) = (1, 1)$, $(0, 1)$, and $(1, 0)$. According

Table 2. Detecting the number of common factors by DIC_2

True model	Model	$\sigma_{vu}=0.3$ Frequency by DIC_2 Small variation		Large variation		$\sigma_{vu}=0.5$ Frequency by DIC_2 Small variation		Large variation	
SAR_1	SAR_1	1		1		1		1	
	SAR_2	0		0		0		0	
	SAR_3	0		0		0		0	
		Mean of DIC_2	(S.D.)	Mean of DIC_2	(S.D.)	Mean of DIC_2	(S.D.)	Mean of DIC_2	(S.D.)
	SAR_1	3651.9	(57.83)	3647.2	(60.71)	3474.1	(58.81)	3477.2	(59.15)
	SAR_2	3720.0	(56.52)	3714.8	(59.81)	3541.2	(57.15)	3543.9	(57.58)
	SAR_3	3796.9	(56.95)	3793.2	(58.81)	3617.5	(59.02)	3620.9	(62.44)
		Frequency by DIC_2 Small variation		Large variation		Frequency by DIC_2 Small variation		Large variation	
SAR_2	SAR_1	0.8		0		0.66		0	
	SAR_2	0.2		1		0.34		1	
	SAR_3	0		0		0		0	
		Mean of DIC_2	(S.D.)	Mean of DIC_2	(S.D.)	Mean of DIC_2	(S.D.)	Mean of DIC_2	(S.D.)
	SAR_1	3746.3	(53.63)	6423.9	(376.98)	3590.8	(55.42)	6409.7	(365.85)
	SAR_2	3777.1	(53.29)	3796.8	(60.09)	3601.5	(54.29)	3622.0	(60.02)
	SAR_3	3859.4	(55.71)	3880.1	(61.67)	3683.4	(57.83)	3705.2	(64.02)
		Frequency by DIC_2 Small variation		Large variation		Frequency by DIC_2 Small variation		Large variation	
SAR_3	SAR_1	0.64		0		0.44		0	
	SAR_2	0.34		0		0.54		0	
	SAR_3	0.02		1		0.02		1	
		Mean of DIC_2	(S.D.)	Mean of DIC_2	(S.D.)	Mean of DIC_2	(S.D.)	Mean of DIC_2	(S.D.)
	SAR_1	3836.4	(70.88)	7438.0	(387.50)	3695.5	(73.89)	7438.7	(387.31)
	SAR_2	3843.3	(66.72)	6070.0	(406.84)	3684.3	(68.81)	6068.6	(404.61)
	SAR_3	3917.4	(61.90)	3960.8	(62.56)	3747.0	(61.69)	3790.0	(61.86)

NOTE: $(\lambda, \beta_1, \beta_2, \beta_3, \sigma_v^2, \sigma_u^2) = (0.6, 1, 1, 1, 1, 1)$; $(\phi_0, \phi_1) = (1, 1)$; Small variation: $\mathrm{var}(\epsilon_\Lambda) = 0.05$ and $\mathrm{var}(\epsilon_\Omega) = 0.05$; Large variation: $\mathrm{var}(\epsilon_\Lambda) = 4$ and $\mathrm{var}(\epsilon_\Omega) = 4$; SAR_i: The SAR model with i common factor, $i = 1, 2, 3$.

to prespecified values of ϕ_0, ϕ_1, and γ_{ij}'s, we have

$$
\begin{aligned}
&\phi_0 = 1, \phi_1 = 1,\ W_{nt|1}:\ \gamma_{ij} = 1 \text{ for bordering states } i \text{ and } j;\\
&\qquad W_{nt|2}:\ \gamma_{ij} = 1 \text{ for } i \neq j,\\
&\phi_0 = 0, \phi_1 = 1,\ W_{nt|3}:\ \gamma_{ij} = 1 \text{ for bordering states } i \text{ and } j;\\
&\qquad W_{nt|4}:\ \gamma_{ij} = 1 \text{ for } i \neq j,\\
&\phi_0 = 1, \phi_1 = 0,\ W_{nt|5}:\ \gamma_{ij} = 1 \text{ for bordering states } i \text{ and } j;\\
&\qquad W_{nt|6}: \gamma_{ij} = 1 \text{ for } i \neq j.
\end{aligned} \tag{6.2}
$$

$W_{nt|5}$ and $W_{nt|6}$ are only related to welfare motivated move because only geographical distance matters for their specification. By construction, they are time-invariant and exogenous spatial weights matrices. On the other hand, $W_{nt|3}$ and $W_{nt|4}$ are specified only based on economic distance. They are related to yardstick competition. $W_{nt|4}$ represents a spatial dependence structure, which does not restrict to bordering states. Even if two states are far from each other, as long as they are economically similar, there can be some strategic interactions between them. Finally, $W_{nt|1}$ and $W_{nt|2}$ control for effects of both distances. They assign larger weights to geographically closer and economically more similar neighbors. So they are related to both welfare motivated move and yardstick competition. For spatial weights involving economic distance, they can be time-varying and endogenous. We estimate various empirical models. The most complex one is the dynamic SAR models with random effects:

$$
\begin{aligned}
Y_{nt} &= \mathcal{L}W_{nt|i}Y_{nt} + \Psi_1 Y_{n,t-1} + \mathcal{R}W_{n,t-1|i}Y_{n,t-1} + \overline{X}_{n1t}\overline{\beta}_1\\
&\quad + \Lambda f_t + V_{nt},\ i = 1, 2, \ldots, 6,\ t = 1, 2, \ldots, T,\\
Z_{nt} &= \overline{X}_{n2t}\overline{\beta}_2 + \Omega f_t + U_{nt},\ t = 1, 2, \ldots, T.
\end{aligned} \tag{6.3}
$$

It follows by the model with random coefficients of the SAR Y_{nt} equation but common coefficients for Z_{nt} equation. More restricted models are the dynamic SDPD models with time factors or additive individual and time effects, as well as static panels with time factors or additive individual and time effects. The Bayesian MCMC sampler is applied to estimate all these models. The values of prior parameters are: $\beta_O = 0$, $B_O = 10 \times I_{k_1+pk_2}$; $\nu = 8$, $R = 10 \times I_{p+1}$, $c_O = 0$, $C_O = 10 \times I_{p+1}$; $\alpha_O = 0$, $A_O = 10 \times I_{p+1}$, and $\Lambda_O = 0$, $\tau_1 = 10$; $\Omega_O = 0$, $\tau_2 = 10$ for relevant restricted models with fixed parameters. Some trace plots of λ and σ_{vu} are depicted in Figure 1(d) to show the convergence of the MCMC sampler. For the random coefficient models, the prior mean of λ_i's is assumed a hierarchical prior as $\mathcal{N}(0, 1)$ and the prior variance is assumed a prior as $\mathcal{IW}(1, 1)$, similarly for the prior means and variances of ρ_i's and ψ_{1i}'s. The prior means of $\beta_i = (\beta'_{1i}, \psi'_{2i}, \beta'_{2i})'$'s, Λ_i's, and Ω_i's are assumed a prior $\mathcal{N}_m(0, I_m)$, with m being their dimensions and the prior covariances are assumed $\mathcal{IW}(m, mI_m)$. Meanwhile, the "restricted" estimation of these models, by treating $W_{nt|i}$'s as exogenous is also investigated, for $i = 1, 2, 3, 4$. For models with time factors, we estimate them with numbers of common factors equal either to 1, 2, or 3. DIC_2 is used to select the proper number q. Furthermore, we compare all six different $W_{nt|i}$'s for all models, based upon DIC_1 and DIC_2, to see which $W_{nt|i}$ is most compatible with the data. For dynamic models with common factors, in the context of Medicaid related spending, we got negative P_D's when computing DIC_1. So only DIC_2 is used. Lastly, model comparison among models with fixed effects, common factors, and random coefficients, based upon DIC_2, are conducted.

Table 3. Summary statistics

Variable	Mean	Maximum	Minimum	S.D.
Medicaid related spending	321.93	973.56	41.74	169.81
Federal grants	754.47	2129.81	309.24	290.71
Income	20320.40	36597.00	11337.40	4097.60
Population (millions)	5.35	34.87	0.45	5.66
Population density (people/square mile)	174.27	1153.06	4.67	240.79
Percentage under age 16	23.66	35.94	8.92	2.67
Percentage over age 65	12.59	20.34	7.73	1.76
Lagged unemployment rate (percent)	5.89	17.40	2.30	2.16
Lagged poverty rate (percent)	13.25	27.20	2.90	4.11
Percentage between age 16 and 65	63.64	78.46	56.33	2.34
Lagged state net capital stock	24678.62	45664.54	14659.54	5471.59

NOTE: Sample is 48 contiguous states, from 1983 to 2002. Dollar amounts are in real per capita terms, with base year 1992.

6.3 Empirical Results

The results on selection of numbers of common factors are summarized in Table 4. In the context of Medicaid related spending, q is selected to be 1 in all cases for both the static and dynamic SAR panel models. However, when it comes to random coefficients, the model with $q = 3$ factors are selected. Table 5 summarizes the estimation results of both the static and

Table 4. Detecting the number of common factors by DIC_2: Empirical application on Medicaid related spending

Model specification				DIC_2					
Setting	Spatial weights	Random effect	Factor	$W_{nt\|1}$	$W_{nt\|2}$	$W_{nt\|3}$	$W_{nt\|4}$	$W_{nt\|5}$	$W_{nt\|6}$
Static	Endogenous	N.A.	1	27,979	28,106	28,004	28,201	N.A.	N.A.
			2	28,007	28,320	28,041	28,214		
			3	28,037	28,343	28,080	28,230		
	Exogenous	N.A.	1	28,010	28,126	28,033	28,230	28,009	27,968
			2	28,036	28,141	28,067	28,241	28,033	28,039
			3	28,065	28,162	28,104	28,257	28,062	28,107
Dynamic	Endogenous	N.A.	1	26,559	26,687	26,556	26,725	N.A.	N.A.
			2	26,636	26,761	26,643	26,793		
			3	26,707	26,802	26,714	26,825		
	Exogenous	N.A.	1	26,549	26,683	26,544	26,711	26,542	26,520
			2	26,648	26,760	26,636	26,778	26,642	26,609
			3	26,689	26,803	26,696	26,833	26,684	26,670
Dynamic	Endogenous	SAR equation	1	26,053	26,115	26,052	26,074	N.A.	N.A.
			2	26,096	26,074	26,098	26,079		
			3	25,989	26,169	26,028	26,027		
	Exogenous	SAR equation	1	26,040	26,106	26,068	26,073	26,061	26,064
			2	26,128	26,104	26,048	26,101	26,052	26,111
			3	26,013	26,065	26,055	26,023	26,050	25,972
Dynamic	Endogenous	both equations	1	22,603	22,559	22,537	22,637	N.A.	N.A.
			2	21,913	21,950	21,849	21,922		
			3	21,929	21,839	21,983	21,727		
	Exogenous	both equations	1	22,941	23,166	22,523	22,635	22,663	22,536
			2	21,910	21,957	21,974	21,946	21,823	21,888
			3	21,754	21,800	21,751	21,749	21,848	21,686

Table 5. Bayesian estimation of the static and dynamic model with $W_{nt|1}$: Medicaid related spending

	Static				Dynamic			
	Fixed effects		With time factors		Fixed effects		With time factors	
Independent variable The SAR equation	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Per capita federal grant	0.24	0.01	0.23	0.01	0.04	0.01	0.04	0.01
Population	3.61	0.46	3.56	0.46	0.51	0.26	0.50	0.25
Population density	−0.00	0.01	−0.01	0.01	−0.01	0.01	−0.01	0.01
Lagged unemployment rate	−1.53	1.46	−1.58	1.46	−0.03	0.87	−0.03	0.85
Lagged poverty rate	2.29	0.80	2.36	0.80	0.96	0.45	0.97	0.44
Percentage under age 16	−3.13	0.60	−3.09	0.60	−1.11	0.34	−1.12	0.33
Percentage over age 65	0.69	1.04	0.63	1.03	0.53	0.60	0.58	0.58
Entry equation								
Percentage under 16	3.78	3.14	3.81	3.11	4.98	3.12	5.00	3.10
Percentage between age 16 and 65	18.88	3.06	18.87	3.06	22.44	3.04	22.49	3.02
Lagged per capita state net capital stock	0.76	0.01	0.76	0.01	0.61	0.01	0.60	0.01
Key parameters								
λ	0.52	0.02	0.53	0.02	0.11	0.02	0.11	0.02
ρ	N.A.	N.A.	N.A.	N.A.	−0.00	0.01	−0.00	0.01
ψ_1	N.A.	N.A.	N.A.	N.A.	0.86	0.02	0.86	0.02
ψ_2	N.A.	N.A.	N.A.	N.A.	11.79	0.60	11.79	0.60
σ_v^2	6.31×10^3	311.02	6.43×10^3	317.20	1.64×10^3	77.56	1.70×10^3	87.35
σ_{vu}	6.07×10^4	1.12×10^4	6.10×10^4	1.11×10^4	4.84×10^3	5.00×10^3	4.05×10^3	4.9×10^3
σ_u^2	1.26×10^7	5.93×10^5	1.26×10^7	5.92×10^5	8.79×10^6	4.20×10^5	8.78×10^6	4.16×10^5

NOTE: Number of iterations is 50,000. We burn in the first 20% draws. $q = 1$ for models with common factors.

Table 6. Bayesian estimation results of key parameters: Medicaid related spending

(a) Bayesian estimation of λ for the static and dynamic models

		Static				Dynamic			
		Fixed effects		Time factors		Fixed effects		Time factors	
Spatial weights	Parameter	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
$W_{nt\|1}$	λ_{ED}	0.52	0.02	0.53	0.02	0.11	0.02	0.11	0.02
	λ_{EX}	0.55	0.02	0.56	0.02	0.11	0.02	0.11	0.02
$W_{nt\|2}$	λ_{ED}	0.51	0.03	0.52	0.03	0.06	0.02	0.07	0.02
	λ_{EX}	0.56	0.03	0.56	0.03	0.06	0.02	0.07	0.02
$W_{nt\|3}$	λ_{ED}	0.52	0.02	0.53	0.02	0.11	0.02	0.11	0.02
	λ_{EX}	0.55	0.02	0.56	0.02	0.11	0.02	0.11	0.02
$W_{nt\|4}$	λ_{ED}	0.42	0.03	0.43	0.03	0.06	0.02	0.06	0.02
	λ_{EX}	0.48	0.03	0.48	0.03	0.06	0.02	0.06	0.02
$W_{nt\|5}$	λ_{EX}	0.56	0.02	0.57	0.02	0.11	0.02	0.11	0.02
$W_{nt\|6}$	λ_{EX}	0.74	0.03	0.74	0.02	0.12	0.02	0.12	0.02

(b) Some Bayesian estimation results of the dynamic random coefficient model with $W_{nt|1}$

	Random coefficients (RC) for both equations	
Key parameter	Mean	S.D.
RC mean of λ	0.08	0.02
RC variance of λ	0.03	0.01
RC mean of ρ	0.01	0.02
RC variance of ρ	0.03	0.01
RC mean of ψ_1	0.54	0.03
RC variance of ψ_1	0.03	0.01
RC mean of ψ_2	1.51	0.96
RC variance of ψ_2	7.07	2.61
σ_v^2	984.43	50.92
σ_{vu}	214.55	338.59
σ_u^2	6.28×10^4	3.98×10^3

NOTE: Number of iterations is 50,000 for (a) and 100,000 for (b), with the first 20% and 70% burned in; $q = 1$ for (a) and $q = 3$ for (b); λ_{ED} (λ_{EX}): estimates of λ with an endogenous (exogenous) W_{nt}.

Table 7. Model selection by DIC_1 and DIC_2: Medicaid related spending

(a) Model selection for proper spatial weight matrices: model with fixed coefficients

	Static								Dynamic					
	Fixed effects				1-factor				Fixed effects				1-factor	
Model	DIC_1	Rank_1	DIC_2	Rank_2	DIC_1	Rank_1	DIC_2	Rank_2	DIC_1	Rank_1	DIC_2	Rank_2	DIC_2	Rank_2
$W_{nt\|1}^{\mathrm{ED}}$	27,858	2	27,959	2	27,872	2	27,979	2	26,176	4	26,320	3	26,559	6
$W_{nt\|2}^{\mathrm{ED}}$	27,991	7	28,089	7	28,013	7	28,106	7	26,201	8	26,346	7	26,687	8
$W_{nt\|3}^{\mathrm{ED}}$	27,875	3	27,976	3	27,888	3	28,004	3	26,176	4	26,320	3	26,556	5
$W_{nt\|4}^{\mathrm{ED}}$	28,088	9	28,183	9	28,110	9	28,201	9	26,203	10	26,347	9	26,725	10
$W_{nt\|1}^{\mathrm{EX}}$	27,887	4	27,987	4	27,902	4	28,010	5	26,176	4	26,321	5	26,549	4
$W_{nt\|2}^{\mathrm{EX}}$	28,012	8	28,108	8	28,033	8	28,126	8	26,200	7	26,346	7	26,683	7
$W_{nt\|3}^{\mathrm{EX}}$	27,903	6	28,002	6	27,917	6	28,033	6	26,175	3	26,321	5	26,544	3
$W_{nt\|4}^{\mathrm{EX}}$	28,120	10	28,216	10	28,141	10	28,230	10	26,202	9	26,350	10	26,711	9
$W_{nt\|5}^{\mathrm{EX}}$	27,890	5	27,991	5	27,904	5	28,009	4	26,173	2	26,178	2	26,542	2
$W_{nt\|6}^{\mathrm{EX}}$	27,811	1	27,909	1	27,825	1	27,968	1	26,155	1	26,176	1	26,520	1

(b) Model selection for proper spatial weight matrices: model with random coefficients and three factors

	RC for SAR equation		RC for both equations	
Model	DIC_2	Rank_2	DIC_2	Rank_2
$W_{nt\|1}^{\mathrm{ED}}$	25,989	2	21,929	9
$W_{nt\|2}^{\mathrm{ED}}$	26,169	10	21,839	7
$W_{nt\|3}^{\mathrm{ED}}$	26,028	6	21,983	10
$W_{nt\|4}^{\mathrm{ED}}$	26,027	5	21,727	2
$W_{nt\|1}^{\mathrm{EX}}$	26,013	3	21,754	5
$W_{nt\|2}^{\mathrm{EX}}$	26,065	9	21,800	6
$W_{nt\|3}^{\mathrm{EX}}$	26,055	8	21,751	4
$W_{nt\|4}^{\mathrm{EX}}$	26,023	4	21,749	3
$W_{nt\|5}^{\mathrm{EX}}$	26,050	7	21,848	8
$W_{nt\|6}^{\mathrm{EX}}$	25,972	1	21,686	1

NOTE: $W_{nt|i}^{\mathrm{ED}}$: models with endogenous $W_{nt|i}$'s, $i = 1, 2, 3, 4$; $W_{nt|i}^{\mathrm{EX}}$: models with exogenous $W_{nt|i}$'s, $i = 5, 6$.

dynamic models, with spatial weight matrix $W_{nt|1}$. We rely on standard deviations (SD) of estimates as well as Bayesian 95% confidence intervals to decide whether the estimated coefficients are significantly different from zero or not. To save space for presentation, only SDs are presented as they provide similar conclusions. For the static models, the Bayesian estimates of σ_{vu} are significantly different from 0 with 5% level of significance, suggesting that the endogeneity of $W_{nt|1}$ might matter. The estimates of λ are positive and significant. For example, the estimate of λ for the static model with common factors is 0.53, which implies that a dollar's increase in a state's neighbors' Medicaid related spendings will increase its own Medicaid related spending by about 53 cents. As $W_{nt|1}$ is specified based upon both geographical distance and economic distance, the significant estimate of λ suggests that state might respond to both its geographically close and economically similar neighbors. However, when it comes to dynamic model, estimation results turn different. First, estimates of σ_{vu} are not significantly different from 0. Also, the estimates of λ are still significant but become much smaller at around 0.11. Additionally, the estimates of ρ are not significantly different from 0, which indicates that the spatial diffusion effect is weak. Finally, the estimates of ψ_1 are significant and over 0.8, implying that state governments are quite persistent when designing their Medicaid policies. Notice that the estimate of ψ_2, which captures the effect of $Y_{n,t-1}$ on Z_{nt}, are positive and significant too. Thus lagged Medicaid spending $Y_{n,t-1}$ can affect both Medicaid spending Y_{nt} and per capita income Z_{nt} (W_{nt}). It is the source of endogeneity of spatial weights matrices in static models as it is omitted in those models. Once it is controlled in the dynamic models, spatial weights matrices become exogenous.

Table 6(a) collects the estimates of λ for all models with six different spatial weight matrices, $W_{nt|i}, i = 1, \ldots, 6$. The estimates of λ are positive with different magnitudes and small SDs. This further confirms that state governments do respond to their geographically close and/or economics similar neighbors. Moreover, for static models, there are some differences in the estimates for models with endogenous or exogenous W_{nt}'s. Specifically, restricted (treating W_{nt} as exogenous) estimates of λ tend to be larger than the unrestricted (allowing W_{nt} to be endogenous) estimates. But for dynamic models, the differences between the restricted and unrestricted estimates are very small. Additionally, with $W_{nt|5}$ and $W_{nt|6}$ constructed only based on geographical distances and hence exogenous, estimates of λ are larger than those with W_{nt}'s involving economic distance. This result is consistent with Baicker (2005), where she finds the largest spillover effects from models with W_{nt} specified from geographical distances and interstate mobility.

Table 6(b) provides some estimation results of the model with random coefficients for $W_{nt|1}$. As shown in the table, states do exhibit heterogenous responses to the average spending of their neighbors. The estimate of the prior mean of λ_i's is 0.08, with a prior variance estimate 0.03. The effect of lagged spending $Y_{n,t-1}$ on states' income per capita Z_{nt} is also heterogenous, with an estimated prior variance of 7.07. Moreover, the estimate of σ_{uv} is not significant after $Y_{n,t-1}$ is controlled in both equations. This confirms in the context of Medicaid spending, $Y_{n,t-1}$ is the source of endogeneity of $W_{nt|1}$, even for the model with random coefficients.

Table 7(a) and (b) summarizes model selections by DIC_1 and DIC_2, regarding spatial weights matrices. From Table 7(a), $W_{nt|6}$ is the spatial weights matrix mostly compatible with the data for model with additive fixed effects and common factors. Results from Table 7(b) also pick $W_{nt|6}$ as the proper spatial weights matrix for models with random coefficients. This indicates compared with yardstick competition, welfare motivated move turns out to be a more important driving force for the interdependence of states' Medicaid spending policies. Moreover, if we compare the two tables, we find that models with random coefficients are preferred over models with fixed coefficients. This implies states do exhibit heterogenous responses when devising their Medicaid spending policies.

7. CONCLUSION

This article considers the specification and estimation of SAR panel models with endogenous spatial weights matrices and common factors. Also random coefficients of interactions and regressors are allowed. We combine these features in both static and dynamic SAR models. For estimation, the Bayesian MCMC method is developed and supported by simulation results. Identification of factors and factor loadings, and various model selection issues by DIC are explored.

Various spatial panel models, corresponding Bayesian MCMC algorithms and model selection procedures are applied to study spillover effects of state Medicaid related spending. The empirical results show positive and significant spillover effects in all cases, after controlling for the endogeneity of spatial weights matrices and the effect of common factors. This implies in the context of Medicaid spending, state governments do respond to their geographically close and/or economically similar neighbors. Both welfare-motivated move and yardstick competition might be possible sources of strategic interactions among state governments. Furthermore, the lagged Medicaid spending $Y_{n,t-1}$ may be the source of the endogeneity of W_{nt} if it were not included in the SAR equation or the entry equations. However, the model selection results about various spatial weights indicate that compared with yardstick competition, welfare motivated move turns out to be a more important driving force for the interdependence of state's Medicaid spending policies; states exhibit heterogenous responses when devising their policies.

ACKNOWLEDGMENTS

The authors thank the co-editor, the guest editor, and an anonymous referee for helpful comments.

[Received January 2015. Revised November 2015.]

REFERENCES

Aguilar, O., and West, M. (2000), "Bayesian Dynamic Factor Models and Portfolio Allocation," *Journal of Business & Economic Statistics*, 18, 338–357. [643]

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716–723. [652]

Anselin, L. (1980), *Estimation Methods for Spatial Autoregressive Structure, Regional Science Dissertation and Monograph Series # 8*, Ithaca, NY: Cornell University. [642]

Aßmann, C., Boysen-Hogrefe, J., and Pape, M. (2012), "The Directional Identification Problem in Bayesian Factor Analysis: An Ex-Post Approach," Working Paper 1799, Kiel Institute for the World Economy, Kiel, Germany. [651]

Bai, J. S. (2009), "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279. [643]

Bai, J. S., and Li, K. P. (2014), "Spatial Panel Data Models With Common Shocks," Munich Personal RePEc Archive Paper No. 52786, Munich University Library, Munich, Germany. [643]

Bai, J. S., and Ng, S. (2013), "Principal Components Estimation and Identification of Static Factors," *Journal of Econometrics*, 176, 18–29. [651]

Bai, J. S., and Wang, P. (2015), "Identification and Bayesian Estimation of Dynamc Factor Models," *Journal of Business and Economic Statistics*, 33, 221–240. [643]

Baicker, K. (2005), "The Spillover Effects of State Spending," *Journal of Public Economics*, 89, 529–544. [642,644,654,659]

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2013), "A Generalized Spatial Data Model With Random Effects," *Econometric Reviews*, 32, 650–685. [642]

Baltagi, B. H., Song, S. H., Jung, B. C., and Koh, W. (2003), "Testing Panel Data Regression Models With Spatial Error Correlation," *Journal of Econometrics*, 117, 123–150. [642]

——— (2007), "Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data," *Journal of Econometrics*, 140, 5–51. [642]

Berg, A., Meyer, R., and Yu, J. (2004), "Deviance Information Criterion for Comparing Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 22, 107–120. [652]

Besley, T., and Case, A. (1995), "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition," *American Economic Review*, 85, 25–45. [643]

Campos, J., Ericsson, R. N., and Hendry, F. D. (2005), *General-to-Specific Modelling*, Cheltenham, UK: Edward Elgar. [643,644]

Case, A., Hines, Jr, J., and Rosen, H. (1993), "Budget Spillovers and Fiscal Policy Independence: Evidence From the States," *Journal of Public Economics*, 52, 285–307. [642]

Celeux, G., Forbes, F., Robert, C., and Titterington, D. (2006), "Deviance Information Criterion for Missing Data Models," *Bayesian Analysis*, 1, 651–674. [652]

Chan, J., and Tobias, J. (2015), "Priors and Posterior Computation in Linear Endogenous Variable Models With Imperfect Instruments," *Journal of Applied Econometrics*, 30, 301–319. [651]

Corrado, L., and Fingleton, B. (2012), "Where is the Economics in Spatial Econometrics?," *Journal of Regional Science*, 52, 210–239. [644]

Doornik, A. J., and Hendry, F. D. (2014), *Statistical Model Selection With 'Big Data'*, Discussion Paper Series 735, Department of Economics, University of Oxford. [643,644]

Ertur, C., and Koch, W. (2007), "Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence," *Journal of Applied Econometrics*, 22, 1033–1062. [642]

Fingleton, B. (2001), "Equilibrium and Economic Growth: Spatial Econometric Models and Simulations," *Journal of Regional Science*, 41, 117–147. [642]

Garofalo, G., and Yamarik, S. (2002), "Regional Convergence: Evidence From a New State-By-State Capital Stock Series," *The Review of Economics and Statistics*, 84, 316–323. [654]

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2014), *Bayesian Data Analysis*, Boca Raton, FL: Taylor and Francisl/CRC. [652]

Gelman, A., Roberts, G., and Gilks, W. (1996), "Efficient Metropolis Jumping Rules," *Bayesian Statistics*, 5, 599–607. [648]

Geweke, J., and Zhou, G. (1996), "Measuring the Pricing Error of the Arbitrage Pricing Theory," *The Review of Financial Studies*, 9, 557–587. [643,651]

Haario, H., Saksman, E., and Tamminen, J. (2001), "An Adaptive Metropolis Algorithm," *Bernoulli*, 7, 223–242. [648]

Holley, S., Pesaran, M., and Yamagata, T. (2010), "A Spatio-Temporal Model of House Prices in the USA," *Journal of Econometrics*, 158, 160–173. [643]

Horn, R., and Johnson, C. (1985), *Matrix Analysis*, New York: Cambridge University Press. [645]

Hsiao, C. (1983), "Identification," in *Handbook of Econometrics*, eds. Z. Griliches, and M. Intriligator, Amsterdam: North-Holland, pp. 224–283. [645]

Hsieh, C.-S., and Lee, L.-f. (2016), "A Social Interactions Model With Endogenous Friendship Formation and Selectivity," *Journal of Applied Econometrics*, 31, 301–319. [643]

Huang, S., and Yu, J. (2010), "Bayesian Analysis of Structural Credit Risk Models With Microstructure Noises," *Journal of Economic Dynamics and Control*, 34, 2259–2272. [652]

Kadane, J. (1975), "The Role of Identification in Bayesian Theory," in *Studies in Bayesian Econometrics and Statistics*, eds. S. Fienberg, and A. Zellner, Amsterdam: North-Holland, pp. 175–191. [645]

Kapoor, M., Kelejian, H., and Prucha, I. (2007), "Panel Data Models With Spatially Correlated Error Components," *Journal of Econometrics*, 140, 97–130. [642]

Kass, R., and Raftery, A. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795. [651]

Kaufmann, S., and Schumacher, C. (2012), "Finding Relevant Variables in Sparse Bayesian Factor Models: Economic Applications and Simulation Results," Deutsche Bundesbank Discussion Paper. No. 29/2012, Deutsche Bundes-bank Research Centre. [651]

Kelejian, H. H., and Piras, G. (2014), "Estimation of Spatial Models With Endogenous Weighting Matrices, and an Application to a Demand Model for Cigarettes," *Regional Science and Urban Economics*, 46, 140–149. [643]

Kelejian, H. H., and Prucha, I. R. (1998), "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model With Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17, 99–121. [642]

——— (1999), "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40, 509–533. [642]

Lee, L.-F. (2004), "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72, 1899–1925. [642]

——— (2007), "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models," *Journal of Econometrics*, 137, 489–514. [642]

Lee, L.-F., and Yu, J. (2010), "Estimation of Spatial Autoregressive Panel Data Model With Fixed Effects," *Journal of Econometrics*, 154, 165–185. [642]

——— (2012), "QML Estimation of Spatial Dynamic Panel Data Models With Time Varying Spatial Weights Matrices," *Spatial Economic Analysis*, 7, 31–74. [645]

——— (2015), "Spatial Panels," in *The Oxford Handbook of Panel Data*, ed. B. Baltagi, Northamptonshire, UK: Oxford University Press, pp. 363–401. [642]

LeSage, J., and Pace, R. K. (2009), *Introduction to Spatial Econometrics*, Boca Raton, FL: CRC Press. [642]

Lin, X. (2010), "Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models With Group Unobservables," *Journal of Labor Economics*, 28, 825–860. [642]

Lopes, H., and West, M. (2004), "Bayesian Model Assessment in Factor Analysis," *Statistica Sinica*, 14, 41–67. [643,644,651]

Moon, H., and Weidner, M. (2009), "Likelihood Expansion for Panel Regression Models With Factors," unpublished manuscript, Department of Economics, University of Southern California. [643]

Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69–85. [653]

Ord, K. (1975), "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, 120–126. [642]

Parent, O., and LeSage, J. P. (2012), "Spatial Dynamic Panel Data Models With Random Effects," *Regional Science and Urban Economics*, 42, 727–738. [642,648]

Pesaran, M. H. (2006), "Estimation and Inference in Large Heterogeneous Panels With a Multifactor Error Structure," *Econometrica*, 74, 967–1012. [643]

Pesaran, M. H., and Tosetti, E. (2011), "Large panels With Common Factors and Spatial Correlation," *Journal of Econometrics*, 161, 182–202. [643]

Poirier, D., and Tobias, J. (2003), "On the Predictive Distribution of Outcome Gains in the Presence of an Unidentified Parameter," *Journal of Business and Economic Statistics*, 21, 258–268. [645]

Qu, X., and Lee, L.-f. (2015), "Estimating a Spatial Autoregressive Model With an Endogenous Spatial Weight Matrix," *Journal of Econometrics*, 184, 209–232. [643]

Raftery, A. E., and Lewis, S. M. (1992), "How Many Iterations in the Gibbs Sampler?" in *Bayesian Statistics*, eds. J. M. Bernardo, J.O. Berger, A.P. Dawid, and A. Smith, Northamptonshire, UK: Oxford University Press, pp. 763–773. [653]

Roberts, G., and Rosenthal, J. S. (2009), "Examples of Adaptive MCMC," *Journal of Computational and Graphical Statistics*, 18, 349–367. [648,650]

Schönemann, P. H. (1966), "A Generalized Solution to the Orthogonal Procrustes Problem," *Psychometrika*, 31, 1–10. [651,652]

Spiegelhalter, D. J., Best, N., Carlin, B., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society*, Series B, 64, 583–639. [651,652,653]

Stephens, M. (2000), "Dealing With Label Switching in Mixture Models," *Journal of the Royal Statistical Society*, Series B, 62, 795–809. [651]

Strachan, R., Chan, J., and Leon-Gonzalez, R. (2013), "Invariant Inference and Efficient Computation in the Static Factor Model," Working Paper 2013-32, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University. [651]

Yamarik, S. (2013), "State-Level Capital and Investment: Updates and Implications," *Contemporary Economic Policy*, 31, 62–72. [654]

Yu, J., de Jong, R., and Lee, L.-F. (2008), "Quasi-Maximum Likelihood Estimators For Spatial Dynamic Panel Data With Fixed Effects When Both n and T are large," *Journal of Econometrics*, 146, 118–134. [642]

——— (2012), "Quasi-Maximum Likelihood Estimators For Spatial Dynamic Panel Data With Fixed Effects When Both n and T are Large: A Nonstationary Case," *Journal of Econometrics*, 67, 16–37. [642]