Computational Statistics and Data Analysis 55 (2011) 475–490

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A space–time filter for panel data models containing random effects

Olivier Parent [a,*], James P. LeSage [b]

[a] Department of Economics, University of Cincinnati, Cincinnati, OH 45221, United States

[b] McCoy College of Business Administration, Department of Finance and Economics, Texas State University-San Marcos, San Marcos, TX 78666, United States

ARTICLE INFO

Article history:
Received 2 November 2009
Received in revised form 12 May 2010
Accepted 13 May 2010
Available online 1 June 2010

Keywords:
Spatial error correlation
Serial correlation
MCMC estimations

ABSTRACT

A space–time filter structure is introduced that can be used to accommodate dependence across space and time in the error components of panel data models that contain random effects. This specification provides insights regarding several space–time structures that have been used recently in the panel data literature. Markov Chain Monte Carlo methods are set forth for estimating the model which allow simple treatment of initial period observations as endogenous or exogenous. The performance of the approach is demonstrated using both Monte Carlo experiments and an applied illustration.

1. Introduction

A number of different specifications for error dependence in space–time panel data models have been introduced that control for heterogeneity as well as both time and spatial dependence (Baltagi et al., 2007; Elhorst, 2004; Yu et al., 2008). A specification based on a space–time filter is introduced here that results in space–time separability captured by a new component that reflects mixing of time and spatial dependence. Bayesian Markov Chain Monte Carlo (MCMC) estimation methods are set forth for the model. Advantages of MCMC estimation include simplification relative to maximum likelihood, where there is a need to optimize over three filter parameters subject to stability constraints (see Yu et al., 2008). In addition, estimates for the random effect parameters are produced along with measures of dispersion that can serve as the basis for inference, and finally treatment of the initial period observations as exogenous or endogenous is simplified. A variety of Bayesian hierarchical panel data specifications have been proposed to address heterogeneity in subject-specific coefficients in various interesting ways (see Chib, 2008, for an extensive review). However, our focus here is on the popular space–time specification from Baltagi et al. (2007) which assumes random effects uncorrelated with the disturbances.

Important theoretical and computational simplifications arise during estimation depending on the interaction between alternative implementation decisions made. The impact of three different implementation choices are considered here: (1) whether a restriction on the cross-product or space–time interaction term implied by the filter specification is imposed or not, (2) treatment of the first cross-sectional set of observations as endogenous or exogenous, and (3) whether the random effects parameters are marginalized allowing a block sampling scheme that conditions on other model parameters rather than simply sampling the effects from their conditional distributions. We carry out tests of these implementation choices in a Monte Carlo setting to explore how these decisions interact.

Our findings are that: (1) correct treatment of the initial period observations is important, especially in cases where T is small, (2) block sampling improved convergence of the MCMC sampler in cases where N and T were small, but not for large samples. This is fortuitous since blocking adds computational complexity in the case of large N and T, and (3) computational

* Corresponding author.
E-mail addresses: olivier.parent@uc.edu (O. Parent), jlesage@spatial-econometrics.com (J.P. LeSage).

0167-9473/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2010.05.016

simplifications arise from imposing the restriction implied by the filter on the cross-product space–time interaction term leads to a convenient factorization of the variance–covariance matrix.

An applied illustration based on a panel data set for US states over time is used to demonstrate a situation where a separability restriction on the space–time cross-product term implied by the filter specification is consistent with the sample data. We argue that both theoretical and computational benefits arise from imposing the restriction implied by the filter on the space–time cross-product term, making past space–time specifications that ignore this suspect.

In Section 2, we set forth the space–time filter specification for the panel data model disturbances. This includes subsections that discuss alternative specifications that arise from: treatment of the initial period observations as endogenous versus exogenous, and past practices that ignore imposition of the theoretical separability restriction on the parameters implied by the space–time filter. Throughout the paper we refer to our proposed specification as the space–time filter specification or filter model and use the term non-filter model or non-filter specification to reference models that do not rely on a separable space and time filter. We fully recognize that quite generally any specification taking the form $S(\rho, \phi)\varepsilon$ involving parameters for space and time dependence (ρ, ϕ) can be viewed as a space–time 'filter', but find the filter/non-filter moniker helpful in our discussion. The MCMC estimation method is set forth in Section 3, where the issue of block sampling for the random effects parameters is taken up. Results from a Monte Carlo experiment are presented in Section 4. An applied illustration based on a space–time panel of 49 US states over the period 1977–1997 is the subject of Section 5.

2. Model specification

The panel data model for $t = 1, \ldots, T$, can be expressed as:

$$\begin{aligned} y_t &= \iota_N \alpha + x_t \beta + \eta_t \\ \eta_t &= \mu + \varepsilon_t \end{aligned} \tag{1}$$

where $y_t = (y_{1t}, \ldots, y_{Nt})'$ is the $N \times 1$ vector of observations for period t, α is the intercept, ι_N is an $N \times 1$ column vector of ones, x_t denotes an $N \times k$ matrix of regressors and μ is an $N \times 1$ column vector of random effects. These are distributed $\mu_i \sim N(0, \sigma_\mu^2)$, and we assume that the vector μ is uncorrelated with ε_t. The disturbance vector $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ is assumed to exhibit spatial autoregressive dependence as well as first-order autoregressive time dependence. Before discussing other commonly used specifications, we will introduce a specific case of space–time dependence proposed by Baltagi et al. (2007). As described in (2), they develop a model based on a spatial autoregressive process governing the random component ε_t and introduce first-order autoregressive time dependence in the remainder error term:

$$\begin{aligned} \varepsilon_t &= \rho W \varepsilon_t + u_t \\ u_t &= \phi u_{t-1} + \nu_t, \end{aligned} \tag{2}$$

where the disturbance vector $\nu_t = (\nu_1, \ldots, \nu_N)'$ is i.i.d. with zero mean and variance $\sigma_\nu^2 I_N$. The scalar parameter ρ reflects the magnitude of spatial dependence, ϕ is the time dependence parameter and the $N \times N$ spatial weight matrix W is assumed known. This matrix defines the connectivity or dependence structure among cross-sectional units with zero elements on the diagonal to prevent dependence of a spatial unit on itself. It is conventional to row-normalize the matrix W, which leads to eigenvalues less than or equal to one. We assume all eigenvalues are real, which arise in the typical case where a symmetric connectivity matrix is row-normalized to produce W.[1]

It is important to note that the panel data model with random effect and space–time specification defined in (2) can be rewritten as

$$\begin{aligned} Y &= \iota_{NT} \alpha + X\beta + \eta \qquad (3) \\ \eta &= Q\mu + \varepsilon \\ (C \otimes B)\varepsilon &= \nu \end{aligned}$$

where $Y = (y_1', \ldots, y_T')'$, $X = (x_1', \ldots, x_T')'$, and $Q = \iota_T \otimes I_N$. The space–time filter corresponds to the Kronecker product of two matrices C and B where B represents the space filter defined as a nonsingular matrix $B = (I_N - \rho W)$ and $C = (I_T - \phi L)$ is the Prais–Winsten transformation shown in (4), which acts as a filter for time/serial dependence.

$$C = \begin{pmatrix} \psi & 0 & \cdots & 0 \\ -\phi & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\phi & 1 \end{pmatrix}. \tag{4}$$

The value of the (1, 1) element ψ in the matrix C depends on whether the first period is modeled as endogenous or assumed exogenous and known. We will refer to these alternative treatments of the initial period cross-sectional

[1] LeSage and Pace (2009) discuss the case where all eigenvalues are not real and show this has little impact in applications of the type considered here.

observations y_1 as conditional or unconditional using the terms *exogenous* and *endogenous specification*. This avoids double use of the word conditional, which we reserve for conditional distributions that make up our MCMC sampling scheme. In the next section we will show that when T is small, estimation results reveal an improvement when the first cross section is treated as endogenous. When modeling the first period cross section we must assume that the process is stationary which can be achieved by setting the (1, 1) element of the $T \times T$ matrix C described in (4) equal to:

$$\psi = \sqrt{1-\phi^2}. \tag{5}$$

For the case of the endogenous specification, applying the time filter to the error term η leads to a first period term equal to $\sqrt{1-\phi^2}\eta_1$, and the stationary condition will be satisfied if $|\phi| < 1$. For the exogenous specification we assume the likelihood is conditional on the first period observations y_1.

The aim of this study is to discuss the relevance of this space–time filter specification. It is interesting to note that the space–time filter corresponds to the Kronecker product of the matrices C and B leading to:

$$C \otimes B = I_{NT} - \rho I_T \otimes W - \phi L \otimes I_N + (\rho \times \phi) L \otimes W, \tag{6}$$

where L is the time-lag operator. This filter implies that spatial effects from the previous period produced by $(L \otimes W)$ have associated parameters $-\rho \times \phi$. A number of studies have relied on space–time specifications that can be viewed as arising from varying treatments of the product term $\rho \times \phi$. For example, the most commonly used space–time specification initially developed by Anselin (1988) who labelled this a *time–space dynamic model*, (see Yu et al., 2008; Parent and LeSage, 2010) introduces a third parameter (θ) to replace $-\rho \times \phi$, where θ is treated as an unknown parameter to be estimated. In essence, this approach relaxes the constraint $\theta = -\rho \times \phi$ implied by the space–time filter.

The *time–space simultaneous* model also proposed in Anselin (1988) and used by Elhorst (2004), neglects the effect of the space–time interaction term which can be viewed as a specification based on the restriction that $\theta = 0$.

Applying the *time–space dynamic model* specification that introduces the parameter θ to the panel data model with random effects would alter the structure of space–time dependence for period t to take the form shown in (2)

$$\begin{aligned} \varepsilon_t &= \rho W \varepsilon_t + \phi \varepsilon_{t-1} + \theta W \varepsilon_{t-1} + u_t \\ u_t &= \phi u_{t-1} + \nu_t, \end{aligned} \tag{7}$$

where θ is a free parameter capturing the effect of the component mixing space and time dependence, usually interpreted as reflecting spatial diffusion that takes place over time.

Although this general specification is more commonly used, it will be shown that the restriction $\theta = -\rho \times \phi$ implied by the space–time filter specification greatly simplifies estimation of the associated model, since it produces a convenient factorization of the variance–covariance matrix. We show that use of the filter specification and the implied $\theta = -\rho \times \phi$ yields computational benefits suggesting that treating the space–time interaction term as a free parameter may be unwise. Parent and LeSage (2010) and Elhorst (2010) point to additional benefits from use of the filter when it comes to interpreting space–time diffusion dynamics in these models.

In Sections 2.1 and 2.2 we show that separability of the space–time filter interacts with treatment of the initial period observations in the endogenous versus exogenous specifications. We develop an MCMC estimation method for the case of the most general unrestricted model, but emphasize simplifications that arise from the restriction (6) on the space–time cross-product parameter.

2.1. Endogenous specification

The first cross section is endogenously determined if we assume that the within-sample observations and pre-sample values are generated by the same process. For this specification, we assume that the first element of the $T \times T$ time filter C described in (4) is equal to (5). Focusing on the space–time filter, we will show that the implied constraint $\theta = -\rho \times \phi$ on the parameter θ greatly reduces complexity of estimation for this specification.

2.1.1. Space–time filter

We consider the panel data model from (1) with the space–time filter applied to the disturbances ε, which results in (8) for each period $t = 2, \ldots, T$.

$$\begin{aligned} \varepsilon_t &= \rho W \varepsilon_t + \phi \varepsilon_{t-1} - \rho \times \phi W \varepsilon_{t-1} + \nu_t \\ B\varepsilon_t &= \phi B \varepsilon_{t-1} + \nu_t. \end{aligned} \tag{8}$$

A simplification for the likelihood function arises from the filter (6) in conjunction with the stationarity conditions required for endogenous treatment of the initial period observations. The stationarity conditions can be obtained by recursive substitution:

$$\varepsilon_t = \sum_{s=0}^{+\infty} \phi^s B^{-1} \nu_{t-s} \tag{9}$$

suggesting we require stationary conditions corresponding to $|\phi| < 1$ and $\lambda_{\min}^{-1} < \rho < \lambda_{\max}^{-1}$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of W respectively.

Thus, the constant variance, independent of time is equal to:

$$\mathrm{Var}(\varepsilon_t) = \sigma_\nu^2(1-\phi^2)^{-1}(B'B)^{-1}, \tag{10}$$

and the autocovariances depend only on the time lag between observations:

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-s}) = \sigma_\nu^2\phi^s(1-\phi^2)^{-1}(B'B)^{-1}. \tag{11}$$

The log-likelihood function of the complete sample size of T for the panel data model is defined in (12).

$$\begin{aligned}\ln L_T(\xi) = &-\frac{NT}{2}\ln(2\pi\sigma_\nu^2) + T\ln|B| + \frac{N}{2}\ln(1-\phi^2)\\ &-\frac{(1-\phi^2)}{2\sigma_\nu^2}(y_1 - x_1\beta - \iota_N\alpha - \mu)'B'B(y_1 - x_1\beta - \iota_N\alpha - \mu)\\ &-\frac{1}{2\sigma_\nu^2}\sum_{t=2}^{T}(y_t^\star - x_t^\star\beta - \iota_N^\star\alpha - \mu^\star)'(y_t^\star - x_t^\star\beta - \iota_N^\star\alpha - \mu^\star)\end{aligned} \tag{12}$$

where $\xi = (\beta', \alpha, \sigma_\nu^2, \phi, \rho, \mu)$, $y_t^\star = By_t - \phi By_{t-1}$, $x_t^\star = Bx_t - \phi Bx_{t-1}$, $\mu^\star = (1-\phi)B\mu$, and since we suppose that the spatial weight matrix W is row-normalized $\iota_N^\star = (1-\phi)B\iota_N = \iota_N(1-\phi)(1-\rho)$.

As we will see, Bayesian estimation methods do not require integration over the random effects that appear in the likelihood. However integration over these parameters can be useful if we want to introduce block sampling schemes that often improve convergence of the MCMC sampler by reducing serial dependence in the samples of parameters drawn (see Chib and Carlin, 1999). In this case, the log-likelihood will be equivalent to:

$$\begin{aligned}\ln L(\beta', \alpha, \sigma_\nu^2, \sigma_\mu^2, \phi, \rho) = &-\frac{NT}{2}\ln(2\pi\sigma_\nu^2) + (T-1)\ln|B|\\ &-\frac{1}{2}\ln\left\{[\alpha^2 + (T-1)](1-\phi)^2\sigma_\mu^2 I_N + \sigma_\nu^2(B'B)^{-1}\right\} - \frac{1}{2}u'\Omega^{\star-1}u,\end{aligned} \tag{13}$$

with $u = (C\otimes I_N)(Y - X\beta - \alpha\iota_{NT})$ and $\Omega^\star$ defined in (14).

$$\Omega = \sigma_\mu^2(J_T\otimes I_N) + \sigma_\nu^2\left[(C'C)^{-1}\otimes(B'B)^{-1}\right]. \tag{14}$$

An issue that needs to be explored is the computational efficiency of block sampling since this requires that we work with the inverse of $(B'B)$, which is a dense matrix rather than the sparse matrix product that does not involve inversion of the $N\times N$ matrix. In our small Monte Carlo study described in Section 4, we find that block sampling improves mixing of the sampler for small N and T, but not in cases where these are large. If more detailed study shows this to be the case, we could use blocking only for problems involving small N and T, where the added computational burden is small.

In contrast to this case where we benefit from separation of the space and time dimensions of the filter specification, we also present the general case where the disturbance covariance matrix takes a more complex form. This complexity arises from the inability to produce the separable factorization shown above.

2.1.2. The non-filter specification

In a maximum likelihood setting, Yu et al. (2008) proposed the general specification that arises from introducing an additional free parameter θ to replace the product $-\rho\times\phi$. The space–time filter (6) becomes: $I_{NT} - \rho I_T\otimes W - \phi L\otimes I_N - \theta L\otimes W$ in this case. The panel data model defined in (1) with this new structure applied to the disturbances ε, produces the model in (15) for each period t.

$$\begin{aligned}y_t &= \iota_N\alpha + x_t\beta + \eta_t\\ \eta_t &= \mu + \varepsilon_t\\ \varepsilon_t &= \rho W\varepsilon_t + \phi\varepsilon_{t-1} + \theta W\varepsilon_{t-1} + \nu_t.\end{aligned} \tag{15}$$

The stationary conditions in (16) for this new specification are defined by recursive substitution applied to (15), where we use the definitions: $A = (\phi I_N + \theta W)$ and $B = (I_N - \rho W)$.

$$B\varepsilon_t = \sum_{s=0}^{+\infty}(AB^{-1})^s\nu_{t-s}. \tag{16}$$

For this general specification, stationary conditions are satisfied only if $|AB^{-1}| < 1$, which requires:

$$\begin{array}{ll}\phi + (\rho+\theta)\lambda_{\max} < 1 & \text{if } \rho+\theta \geq 0\\ \phi + (\rho+\theta)\lambda_{\min} < 1 & \text{if } \rho+\theta < 0\\ \phi - (\rho-\theta)\lambda_{\max} > -1 & \text{if } \rho-\theta \geq 0\\ \phi - (\rho-\theta)\lambda_{\min} > -1 & \text{if } \rho-\theta < 0.\end{array} \tag{17}$$

Elhorst (2004) considers stationarity conditions for the simpler case where $\theta = 0$, concluding that stationarity requires: $\rho \geq 0$ and $|\phi| < 1 - \rho\lambda_{\max}$ or $\rho < 0$ and $|\phi| < 1 - \rho\lambda_{\min}$. Typically, the spatial weight matrix W is based on a symmetric matrix reflecting the neighbor relations between regions that is then row-normalized, leading to a non-symmetric matrix. We assume all eigenvalues λ_i ($i = 1, \ldots, n$) of the matrix W are real and $|\lambda_i| \leq 1$, where the latter condition follows from the row-standardized nature of W.[2] Yu et al. (2008) observed that y_t can have some nonstationary components if $\phi + \rho + \theta = 1$. However, we note that stationarity does not require that $|\phi| + |\rho| + |\theta| < 1$. To see this, we use the same eigenvalue decomposition as Yu et al. (2008) leading to $|AB^{-1}| < 1$ for stationarity which will be satisfied if $|\frac{\phi+\theta\lambda_i}{1-\rho\lambda_i}| < 1$ for all $i = 1, \ldots, n$. It is also noteworthy that the constraint $\theta = -\rho \times \phi$ leads to the specification in Baltagi et al. (2007) and a simplification in the condition for stationarity such that $|\frac{\phi+\theta\lambda_i}{1-\rho\lambda_i}| = |\phi|$. As we will see in the empirical illustration, it is often the case that $\rho > 0$ indicating positive spatial dependence in conjunction with positive time dependence, $\phi > 0$. This implies that θ must be negative when the data generating process is consistent with our space–time filter specification which implies $\theta = -\rho \times \phi$. Thus use of the filter where $\theta = -\rho \times \phi$ allows stationarity in situations where $|\phi| + |\rho| > 1$ with $|\phi| < 1$ and $\lambda_{\min}^{-1} < \rho < \lambda_{\max}^{-1}$.[3]

As previously indicated, we can calculate the full variance matrix without integration over the random effects μ, which in the case of the non-filter specification leads to:

$$\mathrm{Var}(\varepsilon_t) = \sigma_\nu^2[B'B - B'AB^{-1}(B'AB^{-1})']^{-1} \tag{18}$$

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-s}) = \sigma_\nu^2(AB^{-1})^s[B'B - B'AB^{-1}(B'AB^{-1})']^{-1}. \tag{19}$$

Based on the stationary conditions (17), the vector process ε_t will have a constant variance and the autocovariances will only depend on the time separation s between the observations.

The first period contribution to the log-likelihood function takes the form:

$$\begin{aligned}\ln L_1(\zeta) = & -\frac{N}{2}\ln(2\pi\sigma_\nu^2) + \ln|B| + \frac{1}{2}\ln|I_n - AB^{-1}(AB^{-1})'| \\ & - \frac{1}{2\sigma_\nu^2}(y_1 - x_1\beta - \iota_N\alpha - \mu)'[B'B - B'AB^{-1}(B'AB^{-1})'](y_1 - x_1\beta - \iota_N\alpha - \mu),\end{aligned}$$

where $\zeta = (\beta', \alpha, \mu, \sigma_\mu^2, \phi, \rho, \theta)$. In this non-filter case, the space and time components cannot be separated and the likelihood expression for the first period cross section complicates estimation. Implementation for the remaining time periods is relatively straightforward, leading to a complete sample log-likelihood shown in (20).

$$\begin{aligned}\ln L_T(\zeta) = & -\frac{NT}{2}\ln(2\pi\sigma_\nu^2) + T\ln|B| + \frac{1}{2}\ln|I_N - AB^{-1}(AB^{-1})'| \\ & - \frac{1}{2\sigma_\nu^2}(y_1 - x_1\beta - \iota_N\alpha - \mu)'[B'B - B'AB^{-1}(B'AB^{-1})'](y_1 - x_1\beta - \iota_N\alpha - \mu) \\ & - \frac{1}{2\sigma_\nu^2}\sum_{t=2}^{T}(\tilde{y}_t - \tilde{x}_t\beta - \tilde{\iota}_N\alpha - \tilde{\mu})'(\tilde{y}_t - \tilde{x}_t\beta - \tilde{\iota}_N\alpha - \tilde{\mu})\end{aligned} \tag{20}$$

where $\tilde{y}_t = By_t - (\phi I_N + \theta W)y_{t-1}$, $\tilde{x}_t = Bx_t - (\phi I_N + \theta W)x_{t-1}$, $\tilde{\mu} = [(1-\phi)I_N - (\rho+\theta)W]\mu$ and $\tilde{\iota}_N = \iota_N(1-\phi-\rho-\theta)$.

The second line in the likelihood function (20) reflects the influence of the first period cross section, where these observations convey information about how the process has operated in the past.

For this non-filter specification, integration over the random effects presents several difficulties. Because of the first cross section, we cannot transform matrices to find idempotent counterparts as before. The calculations required to implement the MCMC estimation scheme for this model described in Section 3 are set forth in the Appendix.

2.2. Exogenous specification

Ignoring the first period cross-sectional observations represents a tradition associated with the well known Cochrane–Orcutt transformation. This approach ignores information contained in $\hat{\eta}_1$ by virtue of not modeling the initial period observation. This has been justified by arguing that when the time dimension is large, ignoring this information will have little effect on the estimates while greatly simplifying the computational complexity of estimation. In the following, the model specification that arises from this type of treatment is set forth.

2.2.1. Space–time filter

We start by applying the space–time filter in (6) to the error terms from the initial period regression (1) for $t = 2, \ldots, T$, and assume that the specification is conditional on y_1. Therefore the first cross section is taken as exogenous and is not

[2] See LeSage and Pace (2009), Chapter 4 for details regarding complex eigenvalues.

[3] Of course, for row-normalized W, $\lambda_{\max} = 1$.

modeled, leading to (21).

$$\begin{aligned} y_t &= \iota_N \alpha + x_t \beta + \eta_t \\ \eta_t &= \mu + \varepsilon_t \\ \varepsilon_t &= \rho W \varepsilon_t + \phi \varepsilon_{t-1} - \rho \times \phi W \varepsilon_{t-1} + \nu_t. \end{aligned} \tag{21}$$

For this specification, we set $y_t^\star = By_t - \phi By_{t-1}$ and $x_t^\star = Bx_t - \phi Bx_{t-1}$. Since the random effects and intercept do not vary across time, $\mu^\star = (1-\phi)B\mu$ and $\iota_N^\star = (1-\phi)(1-\rho)\iota_N$. The log-likelihood function for the complete sample size of $T-1$ takes the form in (22).

$$\begin{aligned} \ln L_{T-1}(\xi) = &-\frac{N(T-1)}{2}\ln(2\pi\sigma_\nu^2) + (T-1)\ln|B| \\ &- \frac{1}{2\sigma_\nu^2}\sum_{t=2}^{T}(y_t^\star - x_t^\star\beta - \iota_N^\star\alpha - \mu^\star)'(y_t^\star - x_t^\star\beta - \iota_N^\star\alpha - \mu^\star). \end{aligned} \tag{22}$$

We note that in this case we do not need to compute the inverse expression for the $N \times N$ matrix that arose previously, allowing samples involving large N and T to be handled easily.

As in the endogenous case, we can integrate the log-likelihood over the random effects μ which produces the expression in (23).

$$\ln L(\beta', \alpha, \sigma_\nu^2, \sigma_\mu^2, \phi, \rho) = -\frac{N(T-1)}{2}\ln(2\pi\sigma_\nu^2) + (T-1)\ln|B| - \frac{1}{2}\tilde{u}'\tilde{\Omega}^{-1}\tilde{u} \tag{23}$$

where: $\tilde{u} = (\tilde{C} \otimes I_N)(Y - X\beta - \alpha\iota_{NT})$, $\tilde{\Omega}$ is the $N(T-1) \times N(T-1)$ matrix $\tilde{\Omega} = \sigma_\mu^2(J_{T-1} \otimes I_N) + \sigma_\nu^2\left[I_{T-1} \otimes (B'B)^{-1}\right]$, and $\tilde{C}$ is defined in (24).

$$\tilde{C} = \begin{pmatrix} -\phi & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\phi & 1 \end{pmatrix}. \tag{24}$$

Because J_{T-1} is the idempotent counterpart of I_{T-1}, the inverse of $\tilde{\Omega}$ that appears in the likelihood is straightforward.

2.2.2. The non-filter specification

We consider the non-filter model specification which introduces the free parameter θ and relaxes the restriction implied by the space–time filter: $\theta = -\phi \times \rho$ for the case of the exogenous specification. For this model, we can write the disturbances from (3) as shown in (25).

$$\varepsilon_t = \rho W \varepsilon_t + (\phi I_N + \theta W)\varepsilon_{t-1} + \nu_t. \tag{25}$$

We set $\tilde{y}_t = By_t - (\phi I_N + \theta W)y_{t-1}$, $\tilde{x}_t = Bx_t - (\phi I_N + \theta W)x_{t-1}$, $\tilde{\mu} = [(1-\phi)I_N - (\rho+\theta)W]\mu$, and by virtue of a row-normalized spatial weight matrix W, $\tilde{\iota}_N = \iota_N(1-\phi-\rho-\theta)\alpha$. The log-likelihood function for the complete sample of size $T-1$ is:

$$\ln L_{T-1}(\zeta) = -\frac{N(T-1)}{2}\ln(2\pi\sigma_\nu^2) + (T-1)\ln|B| - \frac{1}{2\sigma_\nu^2}\sum_{t=2}^{T}(\tilde{y}_t - \tilde{x}_t\beta - \tilde{\mu} - \tilde{\iota}_N\alpha)'(\tilde{y}_t - \tilde{x}_t\beta - \tilde{\mu} - \tilde{\iota}_N\alpha). \tag{26}$$

Without the first period, the likelihood function takes a form that greatly simplifies estimation. As we will see in the next section when we discuss MCMC estimation, integrating over the random effects is also straightforward in this case when we do not impose the space–time filter restriction on θ. Ignoring the first period observations allows use of conventional matrix expressions and decompositions from the panel data literature that reduce the dimensionality of matrices requiring manipulation during estimation.

3. Bayesian MCMC estimation

Even in this more complicated non-filter model specification MCMC holds some advantages over maximum likelihood or Quasi-Maximum Likelihood (QML) estimation. For a discussion of issues involved in maximizing a concentrated version of the likelihood function that involves trivariate optimization over the parameters ρ, ϕ and θ subject to the stationarity restrictions, see Su and Yang (2007) and Yu et al. (2008). It should be clear that this type of constrained optimization may lead to local optima that may produce misleading inferences. The ability of Bayesian MCMC methods to directly sample from the posterior avoids these problems (LeSage and Pace, 2009), and allows us to sample from simpler expressions reflecting the conditional distribution of each model parameter. Moreover, we propose a Bayesian updating scheme that greatly simplifies the computational complexity of estimation. In practice our approach does not require integration over the random effects that appear in the likelihood, but for completeness we consider this issue.

We implement a fully hierarchical Bayesian framework where the posterior distributions of the parameters are derived using conjugate and non-informative priors. The model set forth here contains only proper prior distributions leading to a joint posterior distribution that is proper.

Bayesian inference is based on the joint posterior distribution of the parameters given the data. Due to the hierarchical structure of the most general model likelihood in (20), the un-normalized posterior takes the form in (27), where the expression $p(y|.)$ is the likelihood function and $p(.)$ represent prior distributions.

$$p(\rho, \alpha, \beta, \sigma_\nu^2, \mu, \sigma_\mu^2, \phi, \theta|y) \propto p(y|\mu, \alpha, \beta, \sigma_\nu^2, \phi, \theta, \rho)p(\mu|\sigma_\mu^2)p(\sigma_\mu^2)p(\sigma_\nu^2)p(\alpha)p(\beta)p(\rho, \phi, \theta). \tag{27}$$

Direct evaluation of the joint posterior distribution involves multidimensional numerical integration and is not computationally feasible. We use MCMC sampling methods which involve generating sequential samples from the complete set of conditional posterior distributions.

We suppose that the prior distributions for the individual parameters $p(\alpha, \beta, \sigma_\nu^2, \sigma_\mu^2)$ are all a priori independent. We separately estimate the intercept term α and parameters β assuming a non-hierarchical prior of the independent Normal–Gamma variety shown in (28).

$$\begin{aligned} &\alpha \sim N(\alpha_0, M_\alpha^{-1}) \\ &\beta \sim N(\beta_0, M_\beta^{-1}) \\ &\sigma_\nu^{-2} \sim G(v_0/2, S_0/2) \\ &\sigma_\mu^{-2} \sim G(v_1/2, S_1/2). \end{aligned} \tag{28}$$

Om way to satisfy the stationarity assumption defined in (17), we use a joint uniform prior for the parameters ρ, ϕ and θ whose support is restricted over the stationarity region.

In the Appendix we set forth an MCMC sampling scheme which involves generating sequential samples from the complete set of conditional posterior distributions for all parameters in the model. We make a number of points regarding implementation issues that should be of interest to practitioners. These arise from the interaction of modeling decisions related to: assuming an exogenous versus endogenous specification, block sampling of the random effects parameters, and imposition of the parameter restriction $\theta = -\rho \times \phi$ implied by the space–time filter.

4. Monte Carlo results

We conduct a Monte Carlo experiment to evaluate: (1) the performance of our MCMC sampling method with and without blocking, (2) to consider the impact of incorrect treatment of the initial period observations as exogenous versus endogenous, and (3) to examine performance under varying sample dimensions for N and T.

The data generating process shown in (29), where the spatial weight matrix W was generated using random points in conjunction with the MATLAB Delaunay routine to produce spatial contiguity weight matrices that were then row-normalized (see Chapter 4 in LeSage and Pace, 2009).

$$\begin{aligned} y_t &= \iota_N \alpha + x_t \beta + \eta_t \\ \eta_t &= \mu + \varepsilon_t \\ \varepsilon_t &= \rho W \varepsilon_t + \phi \varepsilon_{t-1} + \theta W \varepsilon_{t-1} + v_t. \end{aligned} \tag{29}$$

We implement different settings for $\mu_0 = (\rho, \phi, \theta, \sigma_\nu^2, \sigma_\mu^2, \beta, \alpha)$ allowing us to relax the restriction implied by the space–time filter $\theta = -\rho \times \phi$. Values for x_{it} were generated using independent normal ($N(0, 4)$) distributions. We varied $T = 10, 50$ and $N = 50, 200$. For each set of generated sample observations, we generated MCMC estimates for the parameters in the vector μ_0.

Before discussing results from these general Monte Carlo experiments, we present evidence regarding the role of block sampling the effects parameters. As noted, integration over the random effects is not required to implement our MCMC method, but it is generally believed to improve mixing of the chain of parameter draws constructed from the conditional distributions. For this first experiment we only consider the endogenous model specification, which correctly matches the data generating process. Table 1 presents posterior means, standard deviations and numerical standard error (NSE) measures of the accuracy associated with estimates based on integration versus non-integration of the effects parameters.

Block sampling relies on the specification derived by integrating over the random effects to generate draws for the intercept α and parameters β. The reduction in NSE's associated with these parameter estimates based on integration of the random effects and block sampling are clear for the case of $T = 10$, $N = 50$, but less clear for the larger sample case where $T = 50$, $N = 80$. Since there are computational costs associated with marginalizing over the random effects parameters when T and N are large, these results may be fortuitous. Based on these results, the more general experimental results were produced using integration over the random effects only for cases involving small N and T.

Table 1
Monte Carlo simulations for NSE.

T	N	Param. (= true value)	Integration			No integration		
			Post. mean	S.d.	NSE	Post. mean	S.d.	NSE
5	50	ρ (=0.7)	0.6971	0.0097	0.0004	0.7032	0.0075	0.0004
		ϕ (=0.8)	0.8031	0.0120	0.0004	0.8084	0.0073	0.0004
		θ (= − 0.56)	–	–	–	–	–	–
		σ_v^2 (=0.5)	0.5119	0.0098	0.0003	0.5239	0.0080	0.0003
		σ_μ^2 (=0.5)	0.4491	0.0526	0.0019	0.4145	0.0985	0.0031
		β (=0.5)	0.5050	0.0064	0.0001	0.5092	0.0090	0.0002
		α (=5.0)	5.0088	0.0849	0.0025	5.0102	0.1605	0.0053
50	80	ρ (=0.7)	0.6986	0.0033	0.0001	0.6988	0.0028	0.0001
		ϕ (=0.8)	0.7983	0.0017	0.0001	0.7986	0.0020	0.0001
		θ (= − 0.56)	–	–	–	–	–	–
		σ_v^2 (=0.5)	0.5056	0.0023	0.0001	0.5055	0.0021	0.0001
		σ_μ^2 (=0.5)	0.5084	0.0266	0.0008	0.5088	0.0312	0.0010
		β (=0.5)	0.4988	0.0061	0.0001	0.4977	0.0059	0.0001
		α (=5.0)	4.9963	0.0396	0.0013	4.9992	0.0401	0.0013

Table 2
Monte Carlo simulations, $T = 5$, $N = 50$.

Parameter (= true value)	Post. mean	Std. dev.	Lower 0.05	Upper 0.95
	Endogenous and no restriction			
ρ (=0.7)	0.6962	0.0108	0.6783	0.7133
ϕ (=0.8)	0.7967	0.0154	0.7723	0.8248
θ (= − 0.56)	−0.5758	0.0279	−0.6178	−0.5270
σ_v^2 (=0.5)	0.5242	0.0136	0.4985	0.5441
σ_μ^2 (=0.5)	0.4353	0.0798	0.2854	0.5611
β (=0.5)	0.5052	0.0093	0.4885	0.5206
α (=5.0)	4.9874	0.0980	4.8222	5.1509
	Endogenous and restriction			
ρ (=0.7)	0.6971	0.0097	0.6809	0.7127
ϕ (=0.8)	0.8031	0.0120	0.7832	0.8222
θ (= − 0.56)	–	–	–	–
σ_v^2 (=0.5)	0.5119	0.0098	0.4953	0.5262
σ_μ^2 (=0.5)	0.4491	0.0526	0.3645	0.5511
β (=0.5)	0.5050	0.0064	0.4944	0.5157
α (=5.0)	5.0088	0.0849	4.8753	5.1450
	Exogenous and restriction			
ρ (=0.7)	0.5351	0.0112	0.5145	0.5527
ϕ (=0.8)	0.5537	0.0108	0.5376	0.5718
θ (= − 0.56)	–	–	–	–
σ_v^2 (=0.5)	0.4895	0.0190	0.4587	0.5224
σ_μ^2 (=0.5)	0.5906	0.1030	0.4034	0.7689
β (=0.5)	0.4941	0.0096	0.4792	0.5108
α (=5.0)	4.8700	0.1027	4.6998	5.0409

Table 2 presents results that shed light on misspecification of the generating process used to produce initial period observations. These were produced for the case of y_1 endogenous by assuming that pre-sample iterations were generated by the same process as the within-sample values. This was accomplished by generating a pre-sample of 1000 iterations, but using the last T periods as our sample data. If y_1 is taken as exogenous, the pre-sample observations are excluded and the model is treated as conditional on this first period. We run for each case simulations with 2000 iterations after a burn-in period of 1000. Posterior distributions are summarized in Table 2.

Before focusing on the misspecification that could be implied if we ignore the first cross section, we underline the importance of the restriction $\theta = -\rho \times \phi$ implied by the space–time filter. Results presented in the second part of Table 2 represent the case where the estimation procedure is based on this restriction and the initial period is endogenously determined. The first part of Table 2 relaxes this filter restriction and allows for estimation of the parameter θ. Of course in both cases, mean estimates match the true value, but the standard deviations are slightly higher when we do not take the space–time filter restriction on the parameter θ into account. The problem in this overidentified case is that posterior distributions for most of the other parameters are conditional on θ. Introducing extra variation at each iteration by relaxing the assumption $\theta = -\rho \times \phi$ does not impact the posterior means, but increases dispersion as well as correlation across the draws which can be avoided by imposing the relevant restriction from the space–time filter that decreases the number of parameters estimated.

Table 3
Monte Carlo simulations, $T = 5$, $N = 50$.

Parameter (= true value)	Post. mean	Std. dev.	Lower 0.05	Upper 0.95
	Endogenous and no restriction			
ρ (=0.7)	0.7149	0.0093	0.6928	0.7255
ϕ (=0.8)	0.8002	0.0037	0.7951	0.8046
θ (= − 0.75)	−0.7529	0.0220	−0.7810	−0.7273
σ^2_v (=0.5)	0.4985	0.0030	0.4941	0.5009
σ^2_μ (=0.5)	0.4822	0.0693	0.3675	0.5886
β (=0.5)	0.4965	0.0043	0.4889	0.5001
α (=5.0)	4.9998	0.0298	4.9496	5.0478
	Endogenous and restriction			
ρ (=0.7)	0.7491	0.0081	0.7290	0.7574
ϕ (=0.8)	0.6934	0.0287	0.6747	0.7139
θ (= − 0.75)	–	–	–	–
σ^2_v (=0.5)	0.4698	0.0102	0.4623	0.4863
σ^2_μ (=0.5)	0.9875	0.0862	0.7998	1.0471
β (=0.5)	0.4937	0.0015	0.4916	0.4963
α (=5.0)	4.9355	0.0705	4.8138	5.0474
	Exogenous and no restriction			
ρ (=0.7)	0.5828	0.0204	0.5486	0.6150
ϕ (=0.8)	0.8502	0.0089	0.8362	0.8656
θ (= − 0.75)	−0.6154	0.0129	−0.6365	−0.5924
σ^2_v (=0.5)	0.4813	0.0117	0.4631	0.5015
σ^2_μ (=0.5)	0.4608	0.1346	0.2672	0.7090
β (=0.5)	0.5059	0.0353	0.4475	0.5636
α (=5.0)	4.9062	0.0815	4.8245	4.9835

In order to emphasize the impact of incorrect treatment of initial period observations, we consider a third experiment that assumes the data generating process started at the first period, ignoring the information conveyed by pre-sample values. Estimation results are presented in the third part of Table 2. The estimation of this exogenous specification performs badly, especially for the constant term and the parameters ϕ and ρ that measure the strength of time and space dependence. We also see a substantial increase in the standard deviation of the estimates for these parameters.

Similar conclusions can be drawn for the case where we relax the space–time filter restriction in the data generating process. Of course, for data generated without the restriction imposed, we would not expect a model that imposes this restriction during estimation to perform well (see Table 3). The endogenous specification based on a relaxation of the space–time filter restriction produces slightly better estimation results than those for the exogenous case. Ignoring the first period cross section greatly alters the likelihood function, and we observed the Metropolis–Hastings step used to generate draws for the parameters ϕ, ρ and θ performing poorly in cases where T was small.

As shown in Table 4, the estimates exhibit an increase in precision as the number of periods T and the number of spatial units N increase. For all estimation procedures, estimates perfectly match the true values. As expected, increasing the number of time periods reduces the impact of changes seen in approaches that vary with regard to treatment of the first period cross section. These results are consistent with findings of Su and Yang (2007) using a dynamic framework. They report results showing that incorrect treatment of the initial period observations can lead to biased estimates when T is small. Based on our simulations, when the estimation procedure use an incorrect model specification, the 0.05 and 0.95 intervals do not cover the true parameter values.

5. An applied example

As an applied illustration, we use the extended Solow model with physical and human capital (see Mankiw et al., 1992). This model is similar to a growth accounting exercise that decomposes output growth into growth in capital, labor and technological change. Based on the extended Solow model (see Barro and Sala-i-Martin, 2004), we consider the Cobb–Douglas production function that uses physical capital K, human capital H, and raw labor L shown in (30).

$$Y = CK^{\alpha}H^{\beta}(AL)^{1-\alpha-\beta}, \tag{30}$$

where C is a constant term and A is the labor-augmenting technological progress growing at a constant rate.

Traditionally, A is a Hicks-neutral productivity term and $\dot{A}/A$ is commonly referred to as total factor productivity.[4] A is usually measured as a residual (Solow, 1957), reflecting not only technology but all factors that determine output for a given amount of human and physical capital. We adopt a narrower view here and consider the level of technology A as

4 Growth accounting is usually done with a function such as $Y = AF(K, L)$.

Table 4
Monte Carlo simulations, $T = 50$, $N = 200$.

Parameter (= true value)	Post. mean	Std. dev.	Lower 0.05	Upper 0.95
	Endogenous and no restriction			
ρ (=0.7)	0.6999	0.0017	0.6969	0.7029
ϕ (=0.8)	0.7995	0.0019	0.7965	0.8029
θ (= − 0.56)	−0.5610	0.0025	−0.5649	−0.5569
σ_v^2 (=0.5)	0.5022	0.0019	0.4992	0.5053
σ_μ^2 (=0.5)	0.5131	0.0196	0.4793	0.5470
β (=0.5)	0.4973	0.0019	0.4941	0.5005
α (=5.0)	4.9940	0.0475	4.9197	5.0727
	Endogenous and restriction			
ρ (=0.7)	0.7001	0.0013	0.6978	0.7020
ϕ (=0.8)	0.7996	0.0013	0.7975	0.8018
θ (= − 0.56)	–	–	–	–
σ_v^2 (=0.5)	0.5023	0.0018	0.4995	0.5054
σ_μ^2 (=0.5)	0.5155	0.0200	0.4834	0.5494
β (=0.5)	0.4972	0.0017	0.4941	0.5001
α (=5.0)	4.9965	0.0389	4.9378	5.0568
	Exogenous and restriction			
ρ (=0.7)	0.6917	0.0067	0.6796	0.7024
ϕ (=0.8)	0.7973	0.0016	0.7948	0.8001
θ (= − 0.56)	–	–	–	–
σ_v^2 (=0.5)	0.5041	0.0026	0.4998	0.5082
σ_μ^2 (=0.5)	0.5047	0.0215	0.4692	0.5406
β (=0.5)	0.4969	0.0022	0.4934	0.5004
α (=5.0)	4.9995	0.0346	4.9647	5.0368

driven only by knowledge. In fact, a number of theoretical and applied studies assume that technological change is driven by the knowledge–production sector. Dividing the production function (30) by AL we obtain the following output per unit of effective labor:

$$\widehat{y} = C\widehat{k}^{\alpha}\widehat{h}^{\beta} \tag{31}$$

where $\widehat{y} = Y/(AL)$, $\widehat{h} = H/(AL)$ and $\widehat{k} = K/(AL)$.

Then, differentiation of the logged production function results in the key identity from growth accounting shown in (32).

$$\frac{\dot{\widehat{y}}}{\widehat{y}} = \alpha\frac{\dot{\widehat{k}}}{\widehat{k}} + \beta\frac{\dot{\widehat{h}}}{\widehat{h}}. \tag{32}$$

Expression (32) states that the growth rate of output per unit of effective labor is determined by the growth rate of human and physical capital per unit of effective labor.

We empirically implement a model that measures the influence of employment, capital stock and gross investment on gross state product over the period 1969–1997 for the 48 contiguous US states and District of Columbia. We estimate the panel data model based on Eq. (32). Technological innovation can be measured using proxy variables like patent citations or patents granted (see Audretsch and Feldman, 2004 for a survey). We use the number of US patents granted by the US Patent and Trademark Office during the period 1969–1997.[5] Since the Bureau of Economic Analysis (BEA) only provides capital stock estimates for the US as a whole, not at the state level, we use a measure of capital introduced by Garofalo and Yamarik (2002) based on regional personal income. They allocate capital in the manufacturing sector according to a state's proportion of personal income in each sector at the national level. This method has the advantage that data are available for long periods of time. Garofalo and Yamarik (2002) constructed US state-level capital stocks from 1947 to 1996.[6] Basically, they allocate the national capital stock for each industry according to the relative income generated within each state using the nine one-digit BEA industries. Each state's total capital stock estimate is given by the sum of the industry estimates.

The measure of output is Gross State Product (GSP) in manufacturing from the BEA Website. Extension of the estimates beyond 1997 is not possible because of the 1997 change to the North American Industrial Classification System (NAICS) that took place. The measure for labor comes from the US Department of Labor's Bureau of Labor Statistics (BLS) and includes statewide data from 1969 to 1997. The REIS database reports full time and part time private employment in terms of the number of total workers. To measure the stock of human capital, we used Census estimates for the number of individuals over age 25 that have a bachelor's degree or higher educational attainment level.[7]

5 Data available at the US Department of Commerce, US Patent and Trademark Office. *Patent Counts By Country/State And Year*: *Utility Patents*: *January* 1, 1963–*December* 31, 2006.

6 A revised and extended series through 2001 are available online on the author's website.

7 This variable is available online at http://www.census.gov/population/www/socdemo/education.htm.

Table 5
Estimation results for the non-filter specification.

Parameters	Posterior mean	Std. dev	Lower 0.05	Upper 0.95
Part A: Exogenous specification				
Constant	0.0418	0.0026	0.0376	0.0461
α	0.5243	0.0379	0.4613	0.5866
β	0.4845	0.0382	0.4221	0.5501
ρ	0.6530	0.0211	0.6171	0.6879
ϕ	0.1123	0.0313	0.0621	0.1640
θ	−0.0652	0.0394	−0.1282	−0.0027
σ_μ^2	0.0001	1.57×10^{-5}	0.47×10^{-4}	0.0001
σ_ν^2	0.0006	2.16×10^{-5}	0.0005	0.0006
Part B: Endogenous specification				
Constant	0.0404	0.0026	0.0361	0.0448
α	0.5339	0.0373	0.4716	0.5943
β	0.4753	0.0376	0.4114	0.5319
ρ	0.6555	0.0193	0.6240	0.6871
ϕ	0.1090	0.0312	0.0570	0.1592
θ	−0.0579	0.0394	−0.1228	0.0052
σ_μ^2	0.0001	1.60×10^{-5}	0.46×10^{-4}	0.0001
σ_ν^2	0.0006	2.24×10^{-5}	0.0005	0.0006

We focus on the following specification that includes both spatial and time dependent error components as well as random effects parameters:

$$g_y(t) = \iota_N c + \alpha g_k(t) + \beta g_h(t) + \eta_t \tag{33}$$

$$\eta_t = \mu + \varepsilon_t$$

$$\varepsilon_t = \rho W \varepsilon_t + \phi \varepsilon_{t-1} + \theta W \varepsilon_{t-1} + \nu_t, \tag{34}$$

where $g_x(t)$ represents the growth rate of the variable x at time t.

We note that the constant term is not part of the theoretical model, but including this component seems reasonable if we assume that output growth rates across the sample of US states have a common component. This seems a realistic assumption since the underlying economic environment facing states is influenced by common central bank and national fiscal policies. Estimation results are presented in Table 5, based on a sample of 100,000 draws collected after a burn-in period of 90,000 draws. Part A of the table corresponds to estimates based on the exogenous specification, whereas Part B presents estimates based on endogenous treatment of the initial period vector y_1.

The results appear to provide strong support for models from the endogenous growth literature. In the neighborhood of the steady state, the speed of convergence will be exactly equal to 0. The convergence coefficient corresponds to $\beta^\star = (1 - \alpha - \beta)(\delta + n + g)$, where δ is the capital depreciation rate, n is the population growth rate and g is the exogenous rate of technological progress. As explained by Mankiw et al. (1992) the fact that $\alpha + \beta = 1$ implies there are constant returns to scale for the set of all reproducible factors of production. This result is important because conditional convergence appears to be an empirical regularity in the literature. Jones (2002) explains that including human capital and externalities could raise the physical and human capital shares $\alpha + \beta$ to 4/5 but "there is little evidence to suggest the coefficient is one". The results from our empirical illustration show that ignoring information about innovative activities in growth theory can be misleading. Despite the assumption of exogenous technical progress, and use of a proxy measure for technical progress based on patents, we find that *ideas* have a large role to play in the growth process. A related point is that technological progress is usually measured as a residual from a model that maintains independence between regions and ignores spatial dependence. In order to test for constant returns to scale we set the following hypotheses, $H_1 : \alpha + \beta = 1$ versus $H_2 : \alpha + \beta \neq 1$. We compute the Bayes factor between these two models. Noting that H_1 nests H_0, we use the density ratio of Dickey (1971) to simplify the computation of the Bayes factor:

$$BF_{12} = \frac{P(\alpha + \beta = 1 | y, H_2)}{P(\alpha + \beta = 1 | H_2)}. \tag{35}$$

Using a proper informative prior normally distributed with a mean of zero and a variance of 10, we obtain a Bayes factor of 68.93, showing a strong evidence against H_1, consistent with the hypothesis of constant returns to scale. Here we show that allowing for both spatial and serial correlation in the disturbance process produces potentially interesting results regarding the role of input factors in growth accounting.

Some other points to note regarding the estimates. Based on the stationarity conditions defined in (17) since $\rho + \theta \geq 0$, $\phi + \rho + \theta < 1$, $\rho - \theta \geq 0$ and $\phi - \rho + \theta > -1$, the estimated values suggest the process is stationary. For both sets of estimates, the strength of spatial dependence is greater than that of time dependence, which is consistent with possible localized spillovers across space and time. Positive estimates for the spatial dependence parameter indicate a strong effect

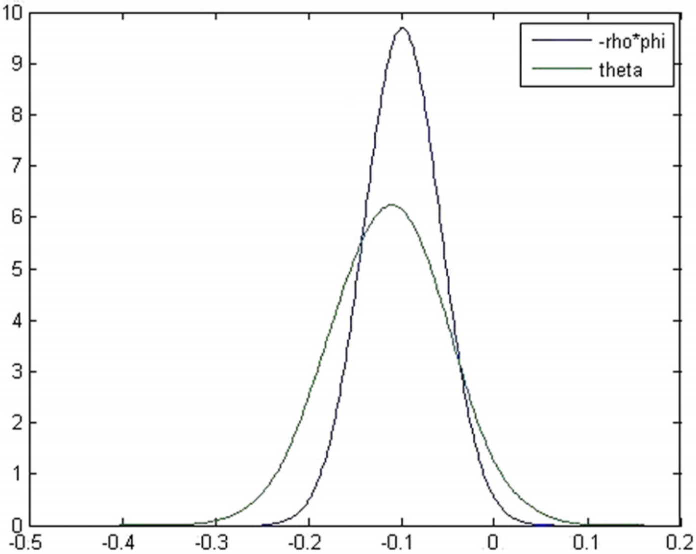

Fig. 1. θ versus $-\phi \times \rho$ (space–time filter constraint).

from geographically adjacent areas on production growth rates. This points to spatial/regional clustering of growth rates in productive activities in contrast to the conventional assumption of statistical independence between regions.

Practitioners may be interested in whether the restriction ($\theta = -\rho \times \phi$) implied by the space–time filter is consistent with a specific set of sample data. One approach might be to visually compare kernel density estimates of the posterior distributions for the parameter θ and the product of draws $-\rho \times \phi$ as shown in Fig. 1. A formal test can be constructed using MCMC draws for θ from the non-filter model and for ρ, ϕ from a model estimated based on the filter specification. Standard Bayesian model comparison criterion such as Bayes Factors or posterior model probabilities can be constructed based on the marginal likelihoods for the two models. Unfortunately, formal investigation of this question requires the non-filter model be estimated which we have noted is more computationally challenging.

To illustrate this we estimated the following model based on the space–time filter specification in (37) which produced the results shown in Table 6.

$$g_y(t) = \iota_N c + \alpha g_k(t) + \beta g_h(t) + \eta_t \tag{36}$$

$$\eta_t = \mu + \varepsilon_t$$

$$\varepsilon_t = \rho W \varepsilon_t + \phi \varepsilon_{t-1} - \phi W \varepsilon_{t-1} + \nu_t. \tag{37}$$

The results in Table 6 appear similar to those shown in Table 5.[8] The second part of the table corresponds to estimates that treat y_1 endogenously, whereas the first part corresponds to the exogenous specification for y_1. Consistent with intuition based on the large number of time periods, the method for treating the first period cross section does not greatly affect the estimates.

Calculating the ratio of marginal likelihoods from the filter and non-filter specifications (based on exogenous treatment of y_1) using (log) marginal likelihood values constructed with the method proposed by Chib and Jeliazkov (2001) produced a Bayes factors lower that 1.5, consistent with similar estimates and inferences from these two specifications.

6. Conclusion

We introduced a space–time filter specification for the disturbances in a panel data model. The specification can be viewed as a separable space–time filter that includes a cross-product term that mixes dependence across both space and time. One focus of our investigation was on a restriction implied by the filter that has not been imposed in previous space–time panel data models.

Bayesian Markov Chain Monte Carlo (MCMC) estimation methods were set forth, and we discuss a number of advantages to this type of estimation. For example, simplification relative to maximum likelihood where there is a need to simultaneously optimize over space and time dependence parameters that are subject to stability constraints, along with a number of other model parameters. Posterior estimates for the random effects parameters are produced along with measures of dispersion that can serve as the basis for inference. Finally, MCMC estimation allows simple treatment of the initial period observations as exogenous or endogenous.

8 These estimates were based on an MCMC sample of 100,000 iterations collected after a burn-in period of 90,000 draws.

Table 6
Estimation results for the space–time filter specification.

Parameters	Posterior mean	Std. dev	Lower 0.05	Upper 0.95
Part A: Exogenous specification				
Constant	0.0418	0.0025	0.0376	0.0459
α	0.5241	0.0369	0.4623	0.5832
β	0.4890	0.0371	0.4289	0.5508
ρ	0.6557	0.0207	0.6204	0.6900
ϕ	0.1128	0.0291	0.0680	0.1593
σ_μ^2	0.0001	1.61×10^{-5}	0.47×10^{-4}	0.0001
σ_ν^2	0.0005	2.21×10^{-5}	0.0005	0.0006
Part B: Endogenous specification				
Constant	0.0406	0.0062	0.0305	0.0504
α	0.5298	0.0354	0.4731	0.5867
β	0.4818	0.0355	0.4231	0.5392
ρ	0.6605	0.0219	0.6259	0.6980
ϕ	0.1149	0.0292	0.0652	0.1605
σ_μ^2	0.0001	1.58×10^{-5}	0.46×10^{-4}	0.0001
σ_ν^2	0.0005	2.23×10^{-5}	0.0005	0.0006

The conditional distributions required for MCMC sampling were set forth under a number of differing implementation decisions that might be made. Specifically, we considered: (1) cases involving imposition or relaxation of a theoretical restriction on the cross-product or space–time interaction term by the space–time filter specification, (2) treatment of the first period cross-sectional set of observations as endogenous or exogenous, and (3) marginalizing the random effects parameters and block sampling versus sampling these parameters from their conditional distributions. A small set of misspecification tests were used in a Monte Carlo setting to explore how these alternative implementation decisions interact.

Our findings were that: (1) correct treatment of the initial period observations is important, especially in cases where T is small, (2) block sampling improved convergence of the MCMC sampler in cases where N and T were small, and (3) computational simplifications arise from imposing a restriction on the cross-product space–time interaction term.

Appendix

An MCMC estimation procedure is set forth here for the non-filter specification where y_1 is treated as endogenous and θ is treated as a parameter to be estimated. This is the most general specification, so the MCMC procedures for other specifications reflect special cases of this most general situation. In order to apply the Prais–Winsten transformation to the initial model from (3), we define the matrix (38), where $S = [I_N - AB^{-1}(AB^{-1})']^{1/2}B$ and $A = (\phi I_N + \theta W)$.

$$P = \begin{pmatrix} S & & & \mathbf{0} \\ -(\phi I_N + \theta W) & B & & \\ & \ddots & \ddots & \\ \mathbf{0} & & -(\phi I_N + \theta W) & B \end{pmatrix}. \tag{38}$$

Having the posterior distribution of the parameters β associated with the explanatory variables conditional on the random effects parameter vector μ is not desirable. This is because these two sets of parameters tend to be highly correlated which can create problems with mixing of the draws for these parameters during sampling. Chib and Carlin (1999) suggest first sampling the parameters β marginalized over the vector μ and then sampling the vector μ conditioned on β. Setting $\tilde{Y} = PY$, $\tilde{X} = PX$ and $\tilde{\alpha} = \alpha P\iota_{NT}$, we can integrate the likelihood function defined by (20) over μ, and obtain

$$\begin{aligned} \tilde{Y} &= \tilde{\alpha} + \tilde{X}\beta + \eta \\ \eta &= P(\iota_T \otimes I_N)\mu + \nu \\ \eta &\sim N(0, \tilde{\Omega}). \end{aligned} \tag{39}$$

Setting $\eta = (\eta_1', \eta_{-1}')' = (\eta_1', \eta_2', \dots, \eta_T')'$, we can derive the $NT \times NT$ variance–covariance matrix $\tilde{\Omega}$:

$$\tilde{\Omega} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & \Omega \end{pmatrix} \tag{40}$$

where Ω, w_{11} and $w_{12} = w_{21}'$, are defined in (41), with $\bar{J}_{T-1} = (\iota_{T-1}\iota_{T-1}')/(T-1)$.

$$\begin{aligned} w_{11} &= E(\eta_1\eta_1') = \sigma_\mu^2 SS' + \sigma_\nu^2 I_N \\ &= \sigma_\mu^2[BB' - B'AB^{-1}(B'AB^{-1})'] + \sigma_\nu^2 I_N \end{aligned}$$

$$w_{21} = E(\eta_{-1}\eta_1') = \sigma_\mu^2 \iota_{T-1} \otimes [(B-A)S']$$

$$\Omega = E(\eta_{-1}\eta_{-1}') = (\bar{J}_{T-1}) \otimes [I_N\sigma_\nu^2 + (T-1)(B-A)(B-A)'\sigma_\mu^2] + [I_{T-1} - \bar{J}_{T-1}] \otimes (I_N\sigma_\nu^2). \tag{41}$$

In this non-filter specification, it is not possible to separate the spatial and time filters because of the first period cross-sectional observations present in the likelihood. So, block sampling imposes some computational costs in this situation. The exact nature of the trade-off between faster convergence due to blocking and the increased computational effort that slows calculations is unclear, and should be the subject of a careful investigation. However, in a small set of Monte Carlo experiments (that we describe and report in Section 4) we find that for small N and T blocking is beneficial but not for large N and T. These results would be fortuitous since the computational burden from blocking is greatest in large N and T problems where it may not be useful.

We digress slightly to demonstrate the simplification that arises from the exogenous specification that eliminates initial period observations. This allows us to apply the decomposition in (42) proposed by Baltagi et al. (2007) to the matrix Ω^{-1}.

$$\Omega^{-1} = (\bar{J}_{T-1}) \otimes [I_N\sigma_\nu^2 + (T-1)(B-A)(B-A)'\sigma_\mu^2]^{-1} + [I_{T-1} - \bar{J}_{T-1}] \otimes (I_N\sigma_\nu^{-2}) \tag{42}$$

where we note that this precision matrix is slightly different from Baltagi et al. (2007) because we premultiply the model by the $NT \times NT$ matrix P instead of $(C \otimes I_N)$. Magnus (1982) shows that the determinant is equal to:

$$|\Omega| = |I_N\sigma_\nu^2 + (T-1)(B-A)(B-A)'\sigma_\mu^2|\sigma_\nu^{2N(T-2)}. \tag{43}$$

Returning to the endogenous specification, we can use expressions for the partitioned matrix inverse to obtain the (sparse) precision matrix:

$$\tilde{\Omega}^{-1} = \begin{pmatrix} l_{11} & -w_{11}^{-1}w_{12}\Lambda \\ -\Lambda w_{21}w_{11}^{-1} & \Lambda \end{pmatrix}, \tag{44}$$

$$l_{11} = [w_{11} - w_{12}\Omega^{-1}w_{21}]^{-1} \tag{45}$$

$$\Lambda = [\Omega - w_{21}w_{11}^{-1}w_{12}]^{-1}. \tag{46}$$

Because of the special structure of Ω^{-1}, the different blocks of the precision matrix $\tilde{\Omega}^{-1}$ are equal to:

$$\begin{aligned} \Lambda &= (\bar{J}_{T-1}) \otimes \left\{I_N\sigma_\nu^2 + \sigma_\mu^2(T-1)(B-A)(I_N - \sigma_\mu^2 S'w_{11}^{-1}S)(B-A)'\right\}^{-1} + [I_{T-1} - \bar{J}_{T-1}] \otimes (I_N\sigma_\nu^{-2}) \\ l_{11} &= w_{11}^{-1} - (T-1)w_{11}^{-1}\sigma_\mu^4(T-1)S(B-A)' \\ &\quad \times \left\{I_N\sigma_\nu^2 + \sigma_\mu^2(T-1)(B-A)(I_N - \sigma_\mu^2 S'w_{11}^{-1}S)(B-A)'\right\}^{-1}(B-A)S'w_{11}^{-1}. \end{aligned} \tag{47}$$

This MCMC updating scheme greatly simplifies the algorithm and provides better mixing properties. Integrating over the random effects μ and relying on conditional distributions that treat first period observations as given allows us to eliminate the non-sparse full variance–covariance matrix from (14).

We also note that use of the space–time filter restriction $\theta = -\rho \times \phi$, further simplifies the variances to $w_{11} = \sigma_\nu^2(1-\phi^2)^{-1}(B'B)^{-1} + \sigma_\mu^2 J_T$, leading to (48), where $J_T = \iota_T\iota_T'$ and $\tilde{C}$ is defined in (24).

$$\Omega^\star = \sigma_\mu^2(J_{T-1} \otimes I_N) + \sigma_\nu^2\left[(\tilde{C}'\tilde{C})^{-1} \otimes (B'B)^{-1}\right]. \tag{48}$$

Based on the model developed in (39) and the block sampling scheme proposed by Chib and Carlin (1999), the conditional posterior distribution of β (conditioned on $\sigma_\nu^2, \phi, \rho, \theta, \alpha$ and marginalized over the random effects μ) is given by

$$\begin{aligned} p(\beta|y, \sigma_\nu^2, \sigma_\mu^2, \phi, \rho, \theta, \alpha) &\propto \exp\left\{-\frac{1}{2}(\tilde{Y} - \tilde{X}\beta - \tilde{\alpha})'\tilde{\Omega}^{-1}(\tilde{Y} - \tilde{X}\beta - \tilde{\alpha})\right\}\exp\left\{-\frac{1}{2}(\beta - \beta_0)'M_\beta(\beta - \beta_0)\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\beta - \beta_1)'(\tilde{X}'\tilde{\Omega}^{-1}\tilde{X} + M_\beta)(\beta - \beta_1)\right\} \end{aligned}$$

where $\tilde{\Omega}^{-1}$ is defined in (44) and $\beta_1 = [\tilde{X}'\tilde{\Omega}^{-1}\tilde{X} + M_\beta]^{-1}[\tilde{X}'\tilde{\Omega}^{-1}\tilde{X}\widehat{\beta} + M_\beta\beta_0]$ and $\widehat{\beta} = (\tilde{X}'\tilde{\Omega}^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Omega}^{-1}(\tilde{Y} - \tilde{\alpha})$. That is $p(\beta|y, \sigma_\nu^2, \sigma_\mu^2, \phi, \rho, \theta, \alpha) \sim N(\beta_1, [\tilde{X}'\tilde{\Omega}^{-1}\tilde{X} + M_\beta]^{-1})$.

We adopt the same procedure to draw from the conditional posterior distribution of the intercept term, resulting in $p(\alpha|y, \beta, \sigma_\nu^2, \sigma_\mu^2, \phi, \rho, \theta) \sim N(\alpha_1, [(P\iota_{NT})'\tilde{\Omega}^{-1}(P\iota_{NT}) + M_\alpha]^{-1})$, where $\alpha_1 = [(P\iota_{NT})'\tilde{\Omega}^{-1}(P\iota_{NT}) + M_\alpha]^{-1}[(P\iota_{NT})'\tilde{\Omega}^{-1}(\tilde{Y} - \tilde{X}\beta) + M_\alpha\alpha_0]$.

To estimate the random effects, we introduce a hierarchical prior and assume that $\mu_i \sim N(0, \sigma_\mu^2)$, for $i = 1, \ldots, N$, with μ_i and μ_j independent for $i \neq j$. A hierarchical structure for the prior arises if we treat σ_μ^2 as an unknown parameter that requires a prior distribution. If we specify a Gamma distribution for the precision parameter, $\sigma_\mu^{-2} \sim G(S_1, v_1)$, the conditional distribution for the effects parameter vector can be derived in a fashion similar to that for α and β leading to

$p(\mu|y, \beta, \alpha, \sigma_\nu^2, \theta, \phi, \rho) \sim N(m_1, [\sigma_\nu^{-2}(PQ)'(PQ) + \sigma_\mu^{-2}I_N]^{-1})$, where $m_1 = [\sigma_\nu^{-2}(PQ)'(PQ) + \sigma_\mu^{-2}I_N]^{-1}[\sigma_\nu^{-2}(PQ)'(\tilde{Y} - \tilde{\alpha} - \tilde{X}\beta)]$.

Estimation for the remaining parameters is based on the likelihood function developed in (20) since there is no need to integrate over the random effects μ in the case of these. We define the error terms as an $NT \times 1$ vector:

$$e = \tilde{Y} - \tilde{\alpha} - \tilde{Q}\mu - \tilde{X}\beta, \tag{49}$$

where $\tilde{Q} = P(\iota_T \otimes I_N)$.

This notation allows us to express the conditional posterior for σ_ν^{-2} as:

$$\begin{aligned} &\sigma_\nu^{-2}|y, \alpha, \beta, \mu, \rho, \phi, \theta \sim G(\overline{\tau}, \overline{H}) \\ &\overline{\tau} = (NT + v_0)/2 \\ &\overline{H} = (e'e + S_0)/2. \end{aligned} \tag{50}$$

For σ_μ^{-2} we have

$$\begin{aligned} p(\sigma_\mu^{-2}|y, \mu,) &\propto (\sigma_\mu^{-2})^{N/2} \exp\left\{-\frac{\sigma_\mu^{-2}}{2}\mu'\mu\right\} (\sigma_\mu^{-2})^{v_1/2-1} \exp\left\{-\frac{\sigma_\mu^{-2}}{2}S_1\right\} \\ &\propto (\sigma_\mu^{-2})^{(v_1+N)/2-1} \exp\left\{-\frac{\sigma_\mu^{-2}}{2}(\mu'\mu + S_1)\right\} \end{aligned}$$

that is $\sigma_\mu^{-2}|y, \mu \sim G([v_1 + N]/2, [\mu'\mu + S_1]/2)$.

Finally, we adopt proper priors for the parameters ρ, ϕ and θ, assuming that the joint distribution of these parameters is uniformly distributed over the stationary system defined in (17).

$$P(\rho|\alpha, \beta, \mu, \sigma_\nu^2, \phi, \theta, y) \propto |\sigma_v|^{-NT} \prod_{i=1}^{N} (1 - \rho\lambda_i)^T \prod_{i=1}^{N} \left[1 - \left(\frac{\phi + \theta\lambda_i}{1 - \rho\lambda_i}\right)^2\right] \exp\left(-\frac{1}{2\sigma_\nu^2}e'e\right) \tag{51}$$

where λ_i, for $i = 1, \dots, N$, corresponds to the ith eigenvalue of W.

This is not reducible to a standard distribution, so we adopt a Metropolis–Hastings (MH) step during the MCMC sampling procedures that relies on a random walk proposal with normally distributed increments, $\rho_{new} = \rho_{old} + \gamma N(0, 1)$. The acceptance probability is calculated as the ratio of (51) evaluated at the new proposal value (ρ_{new}) and the current draw (ρ_{old}). The proposal tuning parameter γ was systematically incremented or decremented when the acceptance rate moved below 0.40 or above 0.60, which lead to a stable acceptance rate close to 0.50 after a burn-in period. We compute the log-determinant term using a direct sparse matrix LU decomposition approach described in Pace and Barry (1997) that produces a vectorized grids of values for this over the domain of support for ρ, θ and ϕ.

We draw ϕ and θ using the same MH sampling procedure, where we rely on the posterior distributions:

$$p(\phi|\alpha, \beta, \mu, \sigma_\nu^2, \rho, \theta, y) \propto |\sigma|^{-NT} \prod_{i=1}^{N} \left[1 - \left(\frac{\phi + \theta\lambda_i}{1 - \rho\lambda_i}\right)^2\right] \exp\left(-\frac{1}{2\sigma_\nu^2}e'e\right) \tag{52}$$

$$p(\theta|\alpha, \beta, \mu, \sigma_\nu^2, \rho, \phi, y) \propto |\sigma|^{-NT} \prod_{i=1}^{N} \left[1 - \left(\frac{\phi + \theta\lambda_i}{1 - \rho\lambda_i}\right)^2\right] \exp\left(-\frac{1}{2\sigma_\nu^2}e'e\right). \tag{53}$$

The stationarity constraints needed to produce parameters ϕ, ρ and θ consistent with $|AB^{-1}| < 1$ are imposed using rejection sampling.

References

Anselin, L., 1988. Spatial Econometrics, Methods and Models. Kluwer Academic, Boston.

Audretsch, D.B., Feldman, M., 2004. Knowledge Spillovers and the Geography of Innovation. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Urban and Regional Economics, vol. 4. Elsevier, Amsterdam, pp. 2713–2739.

Baltagi, B.H., Song, S.H., Jung, B.C., Koh, W., 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data. Journal of Econometrics 140, 5–51.

Barro, R., Sala-i-Martin, X., 2004. Economic Growth. MIT Press, Cambridge, MA.

Chib, S., 2008. Panel data modeling and inference: a Bayesian primer. In: Matyas, L., Sevestre, P. (Eds.), The Econometrics of Panel Data, 3rd ed. Springer-Verlag, Berlin, Heidelberg, pp. 479–515.

Chib, S., Carlin, B.P., 1999. On MCMC sampling in hierarchical longitudinal models. Statistics and Computing 9, 17–26.

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

Dickey, J., 1971. The weighted likelihood ratio, linear hypothesis on normal location parameters. Annals of Mathematical Statistics 42, 204–223.

Elhorst, P., 2004. Serial and spatial error dependence in space–time models. In: Getis, A., Mur, J., Zoller, H.G. (Eds.), Spatial Econometrics and Spatial Statistics. Palgrave-MacMillan, pp. 176–193.

Elhorst, P., 2010. Applied spatial econometrics: raising the bar. Spatial Economic Analysis 5, 9–28.

Garofalo, G., Yamarik, S., 2002. Regional convergence: evidence from a new state-by-state capital stock series. Review of Economics and Statistics 84, 316–323.
Jones, C.I., 2002. Sources of U.S. economic growth in a world of ideas. American Economic Review 92, 220–239.
LeSage, J., Pace, R.K., 2009. An Introduction to Spatial Econometrics. CRC Press, Boca Raton, FL.
Magnus, J.R., 1982. Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. Journal of Econometrics 19, 239–285.
Mankiw, N.G., Romer, D., Weil, D.N., 1992. A contribution to the empirics of economic growth. Quarterly Journal of Economics 107, 407–437.
Pace, R.K., Barry, R., 1997. Quick computation of spatial autoregressive estimators. Geographical Analysis 29, 232–246.
Parent, O., LeSage, J., 2010. A spatial dynamic panel model with random effects applied to commuting times. Transportation Research Part B 44, 633–645.
Solow, R.M., 1957. Technical change and the aggregate production function. Review of Economics and Statistics 39, 312–320.
Su, L., Yang, Z., 2007. QML estimation of dynamic panel data models with spatial errors. Singapore Management University, Manuscript.
Yu, J., de Jong, R., 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.