ARTICLE IN PRESS
REGEC-02729; No of Pages 17
Regional Science and Urban Economics xxx (2009) xxx–xxx
Contents lists available at ScienceDirect
Regional Science and Urban Economics
journal homepage: www.elsevier.com/locate/regec

Some recent developments in spatial panel data models

Lung-fei Lee [a,*], Jihai Yu [b]

[a] Department of Economics, Ohio State University, United States
[b] Department of Economics, University of Kentucky, United States

ARTICLE INFO

Article history:
Received 12 February 2009
Received in revised form 16 September 2009
Accepted 20 September 2009
Available online xxxx

JEL classification:
C13
C23
R15

Keywords:
Spatial econometrics
Panel data
Dynamic panel data
Fixed effects
Random effects
Testing
Market integration

ABSTRACT

Spatial econometrics has been an ongoing research field. Recently, it has been extended to panel data settings. Spatial panel data models can allow cross sectional dependence as well as state dependence, and can also enable researchers to control for unknown heterogeneity. This paper reports some recent developments in econometric specification and estimation of spatial panel data models. We develop a general framework and specialize it to investigate different spatial and time dynamics. Monte Carlo studies are provided to investigate finite sample properties of estimates and possible consequences of misspecifications. Two applications illustrate the relevance of spatial panel data models for empirical studies.

1. Introduction

Spatial econometrics consists of econometric techniques dealing with the interactions of economic units in space, where the space can be physical or economic in nature. For a cross sectional model, the spatial autoregressive (SAR) model by Cliff and Ord (1973) has received the most attention in economics. Spatial econometrics can be extended to panel data models (Anselin, 1988; Elhorst, 2003). Baltagi et al. (2003) consider the testing of spatial dependence in a panel model, where spatial dependence is allowed in the disturbances. In addition, Baltagi et al. (2007b) consider the testing of spatial and serial dependence in an extended model, where serial correlation over time is also allowed in the disturbances. Kapoor et al. (2007) provide theoretical analysis for a panel data model with SAR and error components disturbances. To allow different spatial effects in the random component and the disturbances terms, Baltagi et al. (2007a) generalize the panel regression model in Kapoor et al. (2007). Instead of the random effects specification of the above models, Lee and Yu (2008) investigate the asymptotic properties of the quasi-maximum likelihood estimators (QMLEs) for spatial panel data models with spatial lags, fixed effects and SAR disturbances. Mutl and Pfaffermayr (2008) consider the estimation of spatial panel data models with spatial lags under both fixed and random effects specifications, and propose a Hausman type specification test. These spatial panel data models have a wide range of applications. They can be applied to agricultural economics (Druska and Horrace, 2004), transportation research (Frazier and Kockelman, 2005), public economics (Egger et al., 2005), and good demand (Baltagi and Li, 2006), to name a few. The above panel models are static ones which do not incorporate time lagged dependent variables in the regression equation.

By allowing dynamic features in the spatial panel data models, Anselin (2001) and Anselin et al. (2008) divide spatial dynamic models into four categories, namely, "pure space recursive" if only a spatial time lag is included; "time–space recursive" if an individual time lag and a spatial time lag are included; "time–space simultaneous" if an individual time lag and a contemporaneous spatial lag are specified; and "time–space dynamic" if all forms of lags are included. Korniotis (forthcoming) investigates a time–space recursive model with fixed effects, and the model is applied to the growth of consumption in each state in the United States. As a recursive model, the parameters, including the fixed effects, can be estimated by OLS. Korniotis (forthcoming) has also considered a bias adjusted within estimator, which generalizes Hahn and Kuersteiner (2002). For a

* Corresponding author.
E-mail addresses: lflee@econ.ohio-state.edu (L. Lee), jihai.yu@uky.edu (J. Yu).

[1] Estimation and testing for spatial dependence in cross sectional data can be found in Anselin (1988, 1992), Kelejian and Robinson (1993), Cressie (1993), Anselin and Florax (1995), Anselin and Rey (1997), Anselin and Bera (1998), Kelejian and Prucha (1998, 2001, 2007) and Lee (2003, 2004, 2007), among others.

0166-0462/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.regsciurbeco.2009.09.002

dynamic panel data model with spatial error, Elhorst (2005) estimates the model with unconditional maximum likelihood method, and Mutl (2006) investigates the model using a three step generalized method of moments (GMM). Su and Yang (2007) derive the QMLEs of the above model under both fixed and random effects specifications. For the general "time–space dynamic" model, we term it the spatial dynamic panel data (SDPD) model to better link the terminology to the dynamic panel data literature (see, e.g., Hsiao, 1986; Alvarez and Arellano, 2003). Yu et al. (2007, 2008) and Yu and Lee (2007) study, respectively, the spatial cointegration, stable, and unit root models where the individual time lag, spatial time lag and contemporaneous spatial lag are all included. The SDPD models can be applied to the growth convergence of countries and regions (Baltagi et al., 2007c; Ertur and Koch, 2007), regional markets (Keller and Shiue, 2007), labor economics (Foote, 2007), public economics (Revelli, 2001; Tao, 2005; Franzese, 2007), and other fields.

The recent survey in Anselin et al. (2008) provides a list of spatial panel data models and presents the corresponding likelihood functions. It points out elementary aspects of the models and testing of spatial dependence via LM tests, but properties of estimation methods are left blank. This paper reports some recent developments in econometric specification and estimation of the spatial panel data models for both static and dynamic cases, investigates some finite sample properties of estimators, and illustrates their relevance for empirical research in economics with two applications. Section 2 gives a literature review of the static spatial panel data models with spatial lags. It discusses fixed and random effects specifications of the individual and time effects, and describes some estimation methods. In addition, the Hausman test procedure for the random specification is covered. Section 3 discusses SDPD models. Given different eigenvalue structures of the SDPD models, asymptotic properties of the estimates are different. Section 3 focuses mostly on QMLEs. Some Monte Carlo results on the estimates and two empirical illustrations are presented in Section 4. They demonstrate the importance of time effects for the accurate estimation of spatial interactions, and also show the use of the SDPD model to study market integration. Conclusions are in Section 5.

2. Static spatial panel data models

Panel regression models with SAR disturbances have recently been considered in the spatial econometrics literature. Anselin (1988) and Baltagi et al. (2003) have considered the model $Y_{nt}=X_{nt}\beta_0+\mathbf{c}_{n0}+U_{nt}$ and $U_{nt}=\lambda_0 W_n U_{nt}+V_{nt}$, $t=1, 2, ..., T$, where $Y_{nt}=(y_{1t}, y_{2t},..., y_{nt})'$ and $V_{nt}=(v_{1t}, v_{2t},..., v_{nt})'$ are $n\times 1$ (column) vectors, v_{it} is *i.i.d.* across i and t with zero mean and finite variance σ_0^2, and W_n is an $n\times n$ spatial weights matrix, which is predetermined and generates the spatial dependence among cross sectional units. Here, X_{nt} is an $n\times k$ matrix of nonstochastic time varying regressors, $\mathbf{c}_{n0}$ is an $n\times 1$ vector of individual random components, and the spatial correlation is in U_{nt}. Kapoor et al. (2007) consider a different specification with $Y_{nt}=X_{nt}\beta_0+U_{nt}^{+}$ and $U_{nt}^{+}=\lambda_0 W_n U_{nt}^{+}+\mathbf{d}_{n0}+V_{nt}, t=1, 2,...,T$, where $\mathbf{d}_{n0}$ is the vector of individual random components. Baltagi et al. (2007a) formulate a general model which allows for spatial correlations in both individual and error components with different spatial parameters. These panel models are different in terms of the variance matrices of the overall disturbances. The variance matrix in Baltagi et al. (2003, 2007a) is more complicated, and its inverse is computationally demanding[2] for a sample with a large n. For the model in Kapoor et al. (2007), spatial correlations in both the individual and error components have the same spatial effect parameter. As the variance matrix in Kapoor et al. (2007) has a special pattern, its inverse can be easier to compute.

[2] Both Baltagi et al. (2003, 2007a) have emphasized on the test of spatial correlation in their models.

The above static spatial panel data models can be generalized as

$$Y_{nt}=\lambda_{01}W_{n1}Y_{nt}+X_{nt}\beta_0+\mu_n+U_{nt},$$
$$\mu_n=\lambda_{03}W_{n3}\mu_n+c_{n0},\ \text{and}\ U_{nt}=\lambda_{02}W_{n2}U_{nt}+V_{nt}, \quad (1)$$

for $t=1,..., T$, where W_{nj} for $j=1,2,3$ are $n\times n$ spatial weights matrices and μ_n is an $n\times 1$ column vector of individual effects.[3] The Baltagi et al. (2007a) panel regression model is a special case of Eq. (1) under $\lambda_{01}=0$, i.e., without spatial lags in the main equation.

For the estimation, we may consider the fixed effects specification (where elements of μ_n are treated as fixed parameters) or the random effects specification (where μ_n is a random component). The random effects specification of μ_n in Eq. (1) can be assumed to be a SAR process. If the process of μ_n in Eq. (1) is correctly specified, estimates of the parameters can be more efficient than those of the fixed effects specification, as they utilize the variation of elements of μ_n across spatial units. On the other hand, the fixed effects specification is known to be robust against the possible correlation of μ_n with included regressors in the model. The fixed effects specification can also be robust against the spatial specification of μ_n. For example, the spatial panel model introduced in Kapoor et al. (2007) is equivalent to Eq. (1) with $W_{n3}=W_{n2}$ and $\lambda_{03}=\lambda_{02}$, but the model in Baltagi et al. (2007a) may not be so. However, with the fixed effects specification, all these panel models have the same representation. By the transformation $(I_n-\lambda_0 W_n)$, the data generating process (DGP) of Kapoor et al. (2007) becomes $Y_{nt}=X_{nt}\beta_0+\mathbf{c}_{n0}+U_{nt}$, where $\mathbf{c}_{n0}=(I_n-\lambda_0 W_n)^{-1}\mathbf{d}_{n0}$ can be regarded as a vector of unknown fixed effect parameters and $U_{nt}=\lambda_0 W_n U_{nt}+V_{nt}$ forms a SAR process.[4] Hence, these equations are identical to a linear panel regression with fixed effects and SAR disturbances, and the estimation of Eq. (1) with μ_n being fixed parameters can be robust under these different specifications. It can also be computationally simpler than some of the random component specifications.

In this section, we will consider several estimation methods for Eq. (1). Section 1 is for the direct estimation of the fixed individual effects. For the fixed effects model, when the time dimension T is small, we are likely to encounter the incidental parameter problem discussed in Neyman and Scott (1948). This is because the introduction of fixed individual effects increases the number of parameters to be estimated, and the time dimension does not provide enough information to consistently estimate those individual parameters. For simplicity, we first review the case with finite T, where the (possible) time effects can be treated as regressors. When T is large, we might also have the incidental parameter problem caused by the time effects; related issues on estimation will be discussed in Section 4. Section 2 covers the transformation approach which eliminates those fixed effects before the estimation. Both Sections 1 and 2 consider the fixed effects specification. Section 3 covers the random effects specification of the spatial panel models, and also discusses the testing issue. Section 4 considers the large T case, where we need to take care of the incidental parameter problem caused by the time effects.

2.1. Direct estimation of fixed effects

For the linear panel regression model with fixed effects, the direct maximum likelihood (ML) approach will estimate jointly the common parameters of interest and fixed effects. The corresponding ML estimates (MLEs) of the regression coefficients are known as the within estimates, which happen to be the conditional likelihood estimates conditional on the time means of the dependent variables. However, the MLE of the variance parameter is inconsistent when T is finite. For the spatial panel data models with individual effects, similar findings of the direct ML approach are found.

[3] When μ_n is treated as fixed effects, any time invariant regressors would be absorbed in μ_n.

[4] $U_{nt}=U_{nt}^{+}-(I_n-\lambda_0 W_n)^{-1}\mathbf{d}_{n0}$.

Please cite this article as: Lee, L., Yu, J., Some recent developments in spatial panel data models, Regional Science and Urban Economics (2009), doi:10.1016/j.regsciurbeco.2009.09.002

Denote $\theta=(\beta', \lambda_1, \lambda_2, \sigma^2)'$ and $\zeta=(\beta', \lambda_1, \lambda_2)'$. At the true value, $\theta_0=(\beta_0', \lambda_{01}, \lambda_{02}, \sigma_0^2)'$ and $\zeta_0=(\beta_0', \lambda_{01}, \lambda_{02})'$. Define $S_n(\lambda_1)=I_n-\lambda_1 W_{n1}$ and $R_n(\lambda_2)=I_n-\lambda_2 W_{n2}$ for any λ_1 and λ_2. At the true parameter, $S_n=S_n(\lambda_{01})$ and $R_n=R_n(\lambda_{02})$. The log likelihood function of Eq. (1), as if the disturbances were normally distributed, is

$$\ln L_{n,T}^d(\theta, \mathbf{c}_n) = -\frac{nT}{2}\ln(2\pi\sigma^2) + T[\ln|S_n(\lambda_1)| + \ln|R_n(\lambda_2)|] - \frac{1}{2\sigma^2}\sum_{t=1}^T V_{nt}'(\zeta, \mathbf{c}_n)V_{nt}(\zeta, \mathbf{c}_n), \quad (2)$$

where $V_{nt}(\zeta, \mathbf{c}_n)=R_n(\lambda_2)[S_n(\lambda_1)Y_{nt}-X_{nt}\beta-\mathbf{c}_n]$. If the disturbances in V_{nt} are normally distributed, the log likelihood (2) is the exact one. When V_{nt} is not really normally distributed, but its elements are i.i.d. $(0, \sigma_0^2)$, Eq. (2) is a quasi-likelihood function.[5] We can estimate $\mathbf{c}_{n0}$ directly from Eq. (2) and have the concentrated log likelihood function of θ. For notational purposes, we define $\tilde{Y}_{nt}=Y_{nt}-\overline{Y}_{nT}$ for $t=1, 2,\cdots, T$ where $\overline{Y}_{nT}=\frac{1}{T}\sum_{t=1}^T Y_{nt}$. Similarly, we define $\tilde{X}_{nt}=X_{nt}-\overline{X}_{nT}$ and $\tilde{V}_{nt}=V_{nt}-\overline{V}_{nT}$. Thus, the log likelihood function with $\mathbf{c}_n$ concentrated out is

$$\ln L_{n,T}^d(\theta) = -\frac{nT}{2}\ln(2\pi\sigma^2) + T[\ln|S_n(\lambda_1)| + \ln|R_n(\lambda_2)|] - \frac{1}{2\sigma^2}\sum_{t=1}^T \tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta), \quad (3)$$

where $\tilde{V}_{nt}(\zeta)=R_n(\lambda_2)[S_n(\lambda_1)\tilde{Y}_{nt}-\tilde{X}_{nt}\beta]$. This direct estimation approach will yield consistent estimates for the spatial and regression coefficients except for the variance parameter σ_0^2 when T is small (but n is large). Also, the estimator of σ_0^2 is consistent only when T is large. These conclusions can be easily seen by comparing the log likelihood in Eq. (3) with that in Section 2.2 (to be shown below).

2.2. Elimination of individual effects

Due to this undesirable property of the direct approach of the estimate of σ_0^2, we may eliminate the individual effects before estimation so as to avoid the incidental parameter problem. When an effective sufficient statistic can be found for each of the fixed effects, the method of conditional likelihood can be used. For the linear regression and logit panel models, the time average of the dependent variables provides the sufficient statistics (see Hsiao, 1986). For the spatial panel data models, we can use a data transformation, the deviation from the time mean operator (i.e., $J_T=I_T-\frac{1}{T}l_T l_T'$ where l_T is the vector of ones), to eliminate the individual effects. The transformed disturbances are uncorrelated, and the transformed equation can be estimated by the QML approach. The transformation approach for the model can be justified as a conditional likelihood approach (Kalbfleisch and Sprott, 1970; Cox and Reid, 1987; Lancaster, 2000).

The J_T eliminates the time invariant individual effects, and the transformed model consists of $\tilde{Y}_{nt}=\lambda_{01}W_{n1}\tilde{Y}_{nt}+\tilde{X}_{nt}\beta_0+\tilde{U}_{nt}$ and $\tilde{U}_{nt}=\lambda_{02}W_{n2}\tilde{U}_{nt}+\tilde{V}_{nt}$ where $\tilde{V}_{nt}=V_{nt}-\overline{V}_{nT}$. However, the resulting disturbances $\tilde{V}_{nt}$ would be linearly dependent over the time dimension because J_T is singular. To eliminate the individual fixed effects without creating linear dependence in the resulting disturbances, a better transformation can be based on the orthonormal matrix of J_T. Let $[F_{T,T-1}, \frac{1}{\sqrt{T}}l_T]$ be the orthonormal matrix of the eigenvectors of J_T, where $F_{T,T-1}$ is the $T\times(T-1)$ eigenvector matrix corresponding to the eigenvalues of 1. For any $n\times T$ matrix $[Z_{n1},\cdots, Z_{nT}]$, define the transformed $n\times(T-1)$ matrix $[Z_{n1}^*,\cdots, Z_{n,T-1}^*]=[Z_{n1},\cdots, Z_{nT}]F_{T,T-1}$ and define $X_{nt}^*=[X_{nt,1}^*, X_{nt,2}^*,\cdots, X_{nt,k}^*]$ accordingly. Then, Eq. (1) implies

$$Y_{nt}^* = \lambda_{01}W_{n1}Y_{nt}^* + X_{nt}^*\beta_0 + U_{nt}^*,\ U_{nt}^* = \lambda_{02}W_{n2}U_{nt}^* + V_{nt}^*,\ t = 1,\cdots, T-1. \quad (4)$$

After the transformation, the effective sample size is $n(T-1)$, and the elements v_{it}^*'s of V_{nt} are uncorrelated for all i and t (and independent under normality).

The log likelihood function of Eq. (4), as if the disturbances were normally distributed, is

$$\ln L_{n,T}(\theta) = -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)[\ln|S_n(\lambda_1)| + \ln|R_n(\lambda_2)|] - \frac{1}{2\sigma^2}\sum_{t=1}^T \tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta). \quad (5)$$

Lee and Yu (2008) show that the transformation approach will yield consistent estimators for all the common parameters including σ_0^2, when either n or T is large.

We may compare the estimates of the direct approach with those of the transformation approach. For the log likelihoods, the difference is in the use of T in Eq. (3) but $(T-1)$ in Eq. (5). If we further concentrate β out, Eq. (3) becomes

$$\ln L_{n,T}^d(\lambda_1, \lambda_2) = -\frac{nT}{2}(\ln(2\pi)+1) - \frac{nT}{2}\ln\hat{\sigma}_{nT}^{2d}(\lambda_1, \lambda_2) + T[\ln|S_n(\lambda_1)| + \ln|R_n(\lambda_2)|], \quad (6)$$

and Eq. (5) becomes

$$\ln L_{n,T}(\lambda_1, \lambda_2) = -\frac{n(T-1)}{2}(\ln(2\pi)+1) - \frac{n(T-1)}{2}\ln\hat{\sigma}_{nT}^2(\lambda_1, \lambda_2) + (T-1)[\ln|S_n(\lambda_1)| + \ln|R_n(\lambda_2)|], \quad (7)$$

where $\hat{\beta}_{nT}^d(\lambda_1, \lambda_2)=\hat{\beta}_{nT}(\lambda_1, \lambda_2)$ and $\hat{\sigma}_{nT}^2(\lambda_1, \lambda_2)$ are the generalized least square estimate of β and the MLE of σ^2 given values of λ_1 and λ_2, and $\hat{\sigma}_{nT}^{2d}(\lambda_1, \lambda_2)=\frac{T-1}{T}\hat{\sigma}_{nT}^2(\lambda_1, \lambda_2)$. By comparing Eqs. (6) and (7), we see that they yield the same maximizer $(\hat{\lambda}_{nT,1}, \hat{\lambda}_{nT,2})$. As $\hat{\beta}_{nT}^d(\lambda_1, \lambda_2)$ and $\hat{\beta}_{nT}(\lambda_1, \lambda_2)$ are identical, the QMLE of $\zeta_0=(\beta_0', \lambda_{01}, \lambda_{02})'$ from the direct approach will yield the same consistent estimate as the transformation approach. However, the estimation of σ_0^2 from the direct approach will be $\frac{T-1}{T}$ times the estimate from the transformation approach.

The transformation approach is a conditional likelihood approach when the disturbances are normally distributed. This is so as follows: Eq. (1) implies that $\overline{Y}_{nT}=\lambda_1 W_{n1}\overline{Y}_{nT}+\overline{X}_{nT}\beta_0+\mathbf{c}_{n0}+\overline{U}_{nT}$ with $\overline{U}_{nT}=\lambda_{02}W_{n2}\overline{U}_{nT}+\overline{V}_{nT}$, but $\mathbf{c}_{n0}$ does not appear in $\tilde{Y}_{nt}=\lambda_{01}W_{n1}\tilde{Y}_{nt}+\tilde{X}_{nt}\beta_0+\tilde{U}_{nt}$ with $\tilde{U}_{nt}=\lambda_{02}W_{n2}\tilde{U}_{nt}+\tilde{V}_{nt}$. Hence, $\overline{Y}_{nT}$ is a sufficient statistic for $\mathbf{c}_{n0}$. As $\tilde{V}_{nt}$, $t=1,\cdots,T$, are independent of $\overline{V}_{nT}$ under normality, the likelihood in Eq. (5) is a conditional likelihood of Y_{nt}, $t=1,...,T$, conditional on $\overline{Y}_{nT}$.

2.3. Random effects specification

In this section, we consider the random effects specification of the individual effects μ_n. When the individual effects are random and are independent of the exogenous regressors, the estimation under the random effects will be more efficient. The spatial effect in μ_n, if allowed, could be considered as the permanent spillover effects as described in Baltagi et al. (2007a). In a random effects model, the presence of time invariant regressors z_n can be allowed. Hence, the model is

$$Y_{nt} = l_n b_0 + z_n\eta_0 + \lambda_{01}W_{n1}Y_{nt} + X_{nt}\beta_0 + \mu_n + U_{nt},\ t = 1,\ldots,T,$$
$$\mu_n = \lambda_{03}W_{n3}\mu_n + \mathbf{c}_{n0},\ \text{and}\ U_{nt} = \lambda_{02}W_{n2}U_{nt} + V_{nt}, \quad (8)$$

[5] In some empirical papers, some authors seem to have the wrong impression for the estimation of a SAR model that: if the disturbances are not truly normally distributed, the MLE would be inconsistent. However, Lee (2004) has shown that the MLE can be consistent for the QML approach when the disturbances are i.i.d. $(0, \sigma_0^2)$ without normality.

Please cite this article as: Lee, L., Yu, J., Some recent developments in spatial panel data models, Regional Science and Urban Economics (2009), doi:10.1016/j.regsciurbeco.2009.09.002

where b_0 is the coefficient for the constant term, and η_0 is the parameter vector for the time invariant regressor z_n. Denote $C_n = I_n - \lambda_{03}W_{n3}$, $\mathbf{Y}_{nT} = (Y'_{n1}, Y'_{n2},...,Y'_{nT})'$ and $\mathbf{V}_{nT}$, $\mathbf{X}_{nT}$ similarly. The above equation in the vector form is

$$\mathbf{Y}_{nT} = l_T \otimes (l_n b_0 + z_n \eta_0) + \lambda_{01}(I_T \otimes W_{n1})\mathbf{Y}_{nT} + \mathbf{X}_{nT}\beta_0 + l_T \otimes C_n^{-1}\mathbf{c}_{n0} + (I_T \otimes R_n^{-1})\mathbf{V}_{nT}.$$

Under the assumptions that $\mathbf{c}_{n0}$ is $(0, \sigma_c^2 I_n)$, V_{nt} is $(0, \sigma_v^2 I_n)$, and they are uncorrelated, the variance matrix of $l_T \otimes C_n^{-1}\mathbf{c}_{n0} + (I_T \otimes R_n^{-1})\mathbf{V}_{nT}$ would be

$$\Omega_{nT} = \sigma_c^2 [l_T l_T' \otimes (C_n' C_n)^{-1}] + \sigma_v^2 [I_T \otimes (R_n' R_n)^{-1}].$$

From the likelihood function, ML random effects estimates can be obtained. By denoting $\mathbf{R}_{nT} = I_T \otimes R_n$ and $\mathbf{S}_{nT} = I_T \otimes S_n$, the log likelihood is

$$\ln L(\mathbf{Y}_{nT}) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_{nT}| + T\ln|S_n| - \frac{1}{2}\xi_{nT}'(\theta)\Omega_{nT}^{-1}\xi_{nT}(\theta),$$

where $\xi_{nT}(\theta) = \mathbf{S}_{nT}\mathbf{Y}_{nT} - \mathbf{X}_{nT}\beta - l_T \otimes (l_n b + z_n \eta)$. For the inverse and determinant of Ω_{nT}, the calculation can be reduced to that of an $n \times n$ matrix. By Lemma 2.2 in Magnus (1982), Baltagi et al. (2007a) show that

$$\Omega_{nT}^{-1} = \frac{1}{T} l_T l_T' \otimes [T\sigma_c^2 (C_n' C_n)^{-1} + \sigma_v^2 (R_n' R_n)^{-1}]^{-1} + J_T \otimes [(\sigma_v^2)^{-1}(R_n' R_n)],$$

and

$$|\Omega_{nT}| = |T\sigma_c^2 (C_n' C_n)^{-1} + \sigma_v^2 (R_n' R_n)^{-1}| \cdot |\sigma_v^2 (R_n' R_n)^{-1}|^{T-1}.$$

The above inverse and determinant can be simplified if $C_n = R_n$, which occurs in the panel model of Kapoor et al. (2007) specified as $Y_{nt} = X_{nt}\beta_0 + U_{nt}$ with $U_{nt} = \lambda_0 W_n U_{nt} + \varepsilon_{nt}$ and $\varepsilon_{nt} = \mu_n + V_{nt}$. This model specification implies that $W_{n2} = W_{n3}$ and $\lambda_{02} = \lambda_{03}$ in Eq. (8). The variance matrix of the error components is

$$\Omega_{nT}^{kkp} = (\sigma_c^2 l_T l_T' + \sigma_v^2 I_T) \otimes (R_n' R_n)^{-1},$$

and the inverse and determinant would be computationally simplified.

With linear and nonlinear moment conditions implied by the error components, Kapoor et al. (2007) propose a method of moments (MOM) estimation with the moment conditions in terms of (λ, σ_v^2, σ_1^2), where[6] $\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2$. The β can be consistently estimated by OLS for their regression equation. Denote $\bar{u}_{nT} = (I_T \otimes W_n)u_{nT}$, $\bar{\bar{u}}_{nT} = (I_T \otimes W_n)\bar{u}_{nT}$, and $\bar{\varepsilon}_{nT} = (I_T \otimes W_n)\varepsilon_{nT}$. Also, let $Q_{0,nT} = J_T \otimes I_n$ and $Q_{1,nT} = \frac{l_T l_T'}{T} \otimes I_n$. For $T \geq 2$, they suggest to use the moment conditions

$$E\begin{bmatrix} \frac{1}{n(T-1)}\varepsilon_{nT}' Q_{0,nT}\varepsilon_{nT} \\ \frac{1}{n(T-1)}\bar{\varepsilon}_{nT}' Q_{0,nT}\bar{\varepsilon}_{nT} \\ \frac{1}{n(T-1)}\bar{\varepsilon}_{nT}' Q_{0,nT}\varepsilon_{nT} \\ \frac{1}{n}\varepsilon_{nT}' Q_{1,nT}\varepsilon_{nT} \\ \frac{1}{n}\bar{\varepsilon}_{nT}' Q_{1,nT}\bar{\varepsilon}_{nT} \\ \frac{1}{n}\bar{\varepsilon}_{nT}' Q_{1,nT}\varepsilon_{nT} \end{bmatrix} = \begin{bmatrix} \sigma_v^2 \\ \sigma_v^2 \frac{1}{n} tr(W_n' W_n) \\ 0 \\ \sigma_1^2 \\ \sigma_1^2 \frac{1}{n} tr(W_n' W_n) \\ 0 \end{bmatrix}. \quad (9)$$

As $\varepsilon_{nT} = u_{nT} - \lambda_0 \bar{u}_{nT}$ and $\bar{\varepsilon}_{nT} = \bar{u}_{nT} - \lambda_0 \bar{\bar{u}}_{nT}$ because $u_{nT} = \lambda_0 (I_T \otimes W_n) u_{nT} + \varepsilon_{nT}$, we can substitute ε_{nT} and $\bar{\varepsilon}_{nT}$ into Eq. (9) and obtain a system of moments about u_{nT}, $\bar{u}_{nT}$ and $\bar{\bar{u}}_{nT}$. With estimates of (λ, σ_v^2, σ_1^2) available from the sample analogue of Eq. (9) based on the least squares residuals, a GLS estimation for β_0 can then be implemented as

$$\hat{\beta}_{GLS,n} = [\mathbf{X}_{nT}'(\Omega_{nT}^{kkp})^{-1}\mathbf{X}_{nT}]^{-1}[\mathbf{X}_{nT}'(\Omega_{nT}^{kkp})^{-1}\mathbf{Y}_{nT}].$$

The feasible GLS (FGLS) estimate can be obtained with (λ, σ_v^2, σ_1^2) in Ω_{nT}^{kkp} replaced by the estimates from the moment conditions in Eq. (9).

For the random effects specification of the linear panel data models, the GLS estimate is the weighted average of the within estimates and between estimates, as is shown in Maddala (1971). Such an interpretation can also be provided for the random effect estimate of the spatial panel model (1). Eq. (4) can be considered as the within equation, which is the deviation from the time average with the individual effects eliminated. On the other hand, the time mean equation

$$\bar{Y}_{nT} = l_n b_0 + z_n \eta_0 + \lambda_{01} W_{n1}\bar{Y}_{nT} + \bar{X}_{nT}\beta_0 + \mu_n + \bar{U}_{nT}, \quad (10)$$
$$\bar{U}_{nT} = \lambda_{02} W_{n2}\bar{U}_{nT} + \bar{V}_{nT},$$

captures the individual effects and can be considered as the between equation. By using $F_{T,T-1}' l_T = 0$, the errors V_{nt}^* and $\bar{V}_{nT}$ are uncorrelated (and independent under normality). Hence,

$$L(\mathbf{Y}_{nT}|\theta, \mu_n) = \left(\frac{1}{T}\right)^{n/2} L_1(Y_{n1}^*, \cdots, Y_{n,T-1}^*|\theta) \times L_2(\bar{Y}_{nT}|\theta, \mu_n), \quad (11)$$

where $(\frac{1}{T})^{n/2}$ is the Jacobian determinant, because $(Y_{n1}^{*\prime}, \cdots, Y_{n,T-1}^{*\prime}, \bar{Y}_{nT}')' = ((F_{T,T-1}, \frac{1}{T}l_T)' \otimes I_n)Y_{nT}$ and the determinant of $[F_{T,T-1}, \frac{1}{T}l_T]$ is $\frac{1}{\sqrt{T}}$. The likelihood L_1 in Eq. (11) for the within equation is in Eq. (5) and the likelihood L_2 for the between equation is

$$L_2(\bar{Y}_{nT}) = (2\pi)^{-n/2}|\Omega_n|^{-1/2} \times \times \exp\{-\frac{1}{2}[S_n\bar{Y}_{nT} - \bar{X}_{nT}\beta - l_n b - z_n\eta]' \Omega_n^{-1}[S_n\bar{Y}_{nT} - \bar{X}_{nT}\beta - l_n b - z_n\eta]\} \times |S_n|,$$

where $\Omega_n = E(\mu_n + \bar{U}_{nT})(\mu_n + \bar{U}_{nT})' = \sigma_\mu^2 (C_n' C_n)^{-1} + \frac{1}{T}\sigma_v^2 (R_n' R_n)^{-1}$. For each of the within and between equations, we may obtain, respectively, the within and between estimates.

With the likelihood decomposition for the spatial panel data model, the random effects ML estimate will be the weighted average of the within and between estimates. Denote $\mathbf{Y}_{n,T-1}^* = (Y_{n1}^{*\prime},...,Y_{n,T-1}^{*\prime})'$ as the sample observations for the within equation. In general, the parameter vector in the likelihood function of $\mathbf{Y}_{n,T-1}^*$ is a subset of that in $\mathbf{Y}_{nT}$ and/or $\bar{Y}_{nT}$. Let the common parameter vector be δ. Consider the concentrated likelihoods (denoted as L^c with a superscript c for a relevant likelihood L) of δ. For illustration, we assume that T is finite so that the within estimator δ_w would be $\sqrt{n}$-consistent. Its asymptotic distribution would be $\sqrt{n}(\hat{\delta}_w - \delta_0) = \left(-\frac{1}{n}\frac{\partial^2 \ln L_1(\mathbf{Y}_{n,T-1}^*)}{\partial\delta\partial\delta'}\right)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_1(\mathbf{Y}_{n,T-1}^*)}{\partial\delta} + o_p(1)$; that of the between estimator $\hat{\delta}_b$ is $\sqrt{n}(\hat{\delta}_b - \delta_0) = \left(-\frac{1}{n}\frac{\partial^2 \ln L_2^c(\bar{Y}_{nT})}{\partial\delta\partial\delta'}\right)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_2^c(\bar{Y}_{nT})}{\partial\delta} + o_p(1)$; and that of the ML estimator based on the likelihood $L^c(\mathbf{Y}_{nT})$ is $\sqrt{n}(\hat{\delta} - \delta_0) = \left(-\frac{1}{n}\frac{\partial^2 \ln L^c(\mathbf{Y}_{nT})}{\partial\delta\partial\delta'}\right)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L^c(\mathbf{Y}_{nT})}{\partial\delta} + o_p(1)$. By simple calculus from (11), $\frac{1}{\sqrt{n}}\frac{\partial \ln L^c(\mathbf{Y}_{nT})}{\partial\delta} = \frac{1}{\sqrt{n}}\frac{\partial \ln L_1(\mathbf{Y}_{n,T-1}^*)}{\partial\delta} + \frac{1}{\sqrt{n}}\frac{\partial \ln L_2^c(\bar{Y}_{nT})}{\partial\delta}$ and $\frac{1}{n}\frac{\partial^2 \ln L^c(\mathbf{Y}_{nT})}{\partial\delta\partial\delta'} = \frac{1}{n}\frac{\partial^2 \ln L_1(\mathbf{Y}_{n,T-1}^*)}{\partial\delta\partial\delta'} + \frac{1}{n}\frac{\partial^2 \ln L_2^c(\bar{Y}_{nT})}{\partial\delta\partial\delta'}$. Hence,

$$\sqrt{n}(\hat{\delta} - \delta_0) = \left(-\frac{1}{n}\frac{\partial^2 \ln L^c(\mathbf{Y}_{nT})}{\partial\delta\partial\delta'}\right)^{-1}\left(\frac{1}{\sqrt{n}}\frac{\partial \ln L_1(\mathbf{Y}_{n,T-1}^*)}{\partial\delta} + \frac{1}{\sqrt{n}}\frac{\partial \ln L_2^c(\bar{Y}_{nT})}{\partial\delta}\right) + o_p(1)$$
$$= A_{nT,1}\sqrt{n}(\hat{\delta}_w - \delta_0) + A_{nT,2}\sqrt{n}(\hat{\delta}_b - \delta_0) + o_p(1),$$

where $A_{nT,1} = \left(\frac{1}{n}\frac{\partial^2 \ln L^c(\mathbf{Y}_{nT})}{\partial\delta\partial\delta'}\right)^{-1}\frac{1}{n}\frac{\partial^2 \ln L_1(\mathbf{Y}_{n,T-1}^*)}{\partial\delta\partial\delta'}$ and $A_{nT,2} = \left(\frac{1}{n}\frac{\partial^2 \ln L^c(\mathbf{Y}_{nT})}{\partial\delta\partial\delta'}\right)^{-1}\frac{1}{n}\frac{\partial^2 \ln L_2^c(\bar{Y}_{nT})}{\partial\delta\partial\delta'}$. The $A_{nT,1}$ and $A_{nT,2}$ are weights because $A_{nT,1} + A_{nT,2} = I_{k_\delta}$

[6] Note that the σ_μ^2 will become σ_c^2 in Kapoor et al. (2007)'s specification.

Please cite this article as: Lee, L., Yu, J., Some recent developments in spatial panel data models, Regional Science and Urban Economics (2009), doi:10.1016/j.regsciurbeco.2009.09.002

where k_δ is the dimension of the common parameters. Thus, the random effects estimate is pooling the within and between estimates, which generalizes that of Maddala (1971) for the standard panel regression model.

The likelihood decomposition also provides a useful device to construct a Hausman type test of random effects specification against the fixed effects specification. Under the null hypothesis that the individual effects are independent of the regressors, the MLE $\hat{\theta}$ of the random effects model, and hence, $\hat{\delta}$, is consistent and asymptotically efficient. However, under the alternative hypothesis, $\hat{\theta}$ is inconsistent. The within estimator $\hat{\delta}_w$ is consistent under both the null and alternative hypotheses. Such a null hypothesis can be tested with a Hausman type statistic by comparing the two estimates $\hat{\delta}$ and $\hat{\delta}_w$ by $n(\hat{\delta}-\hat{\delta}_w)'\hat{\Omega}_n^+(\hat{\delta}-\hat{\delta}_w)$, where $\hat{\Omega}_n$ is a consistent estimate of the limiting variance matrix of $\sqrt{n}(\hat{\delta}-\hat{\delta}_w)$ under the null hypothesis, and $\hat{\Omega}_n^+$ is its generalized inverse. This test statistics will be asymptotically χ^2 distributed, and its degrees of freedom is the rank of Ω_n. Because $\hat{\delta}$ is asymptotically efficient, $[(-\frac{1}{n}\frac{\partial \ln L_1(\mathbf{Y}^*_{n,T-1})}{\partial\delta\partial\delta'})^{-1} - (-\frac{1}{n}\frac{\partial \ln L^c(\mathbf{Y}_{nT})}{\partial\delta\partial\delta'})^{-1}]$, evaluated at either $\hat{\delta}$ or $\hat{\delta}_w$, provides a consistent estimate of Ω_n under the null. By using the identity $B^{-1}-(B+C)^{-1}=B^{-1}(B^{-1}+C^{-1})^{-1}B^{-1}$ for any two positive definite matrices B and C, the preceding difference of the two information matrices is a positive definite matrix. Therefore, the generalized inverse is an inverse, and the degrees of freedom of the χ^2 test is the number of common parameters, i.e., the dimension of δ. Instead of the ML approach, if the main equation is estimated by the 2SLS method, Hausman test statistics can be constructed as in Mutl and Pfaffermayr (2008).

With the estimates of the spatial effect parameters λ_{01} and λ_{02}, tests for the significance of these effects can be constructed by the Wald test. If the main interest is to test the existence of spatial effects, an alternative test strategy may be based on LM statistics (Baltagi et al., 2003, 2007a,b).

2.4. Large T Case

We can extend the model in Eq. (1) by including time effects. When T is short, the time effects can be treated as regressors. When T is large, the time effects might cause the incidental parameter problem.

Similar to Section 1, we can follow a direct estimation approach. With both individual and time effects, even when both n and T are large so that individual and time effects can be consistently estimated, the asymptotic distributions of common parameter estimates are not properly centered at the true parameter values. Hence, it is desirable to eliminate the time effects as well as the individual effects for estimation when they were assumed fixed. Thus, we can extend the transformation approach in Section 2. One may combine the transformation from $J_n = I_n - \frac{1}{n}l_n l_n'$ with the transformation from J_T to eliminate both the individual and time fixed effects. Let $(F_{n,n-1}, \frac{1}{\sqrt{n}}l_n)$ be the orthonormal matrix of J_n, where $F_{n,n-1}$ corresponds to the eigenvalues of 1 and $\frac{1}{\sqrt{n}}l_n$ corresponds to the eigenvalue zero. The individual effects can be eliminated by $F_{T,T-1}$ as in Eq. (4), which yields

$$Y^*_{nt} = \lambda_{01}W_{n1}Y^*_{nt} + X^*_{nt}\beta_0 + \alpha^*_{t0}l_n + U^*_{nt}, \tag{12}$$

$$U^*_{nt} = \lambda_{02}W_{n2}U^*_{nt} + V^*_{nt}, \quad t = 1,2,\ldots,T-1,$$

where $[\alpha^*_{10}l_n,\ \alpha^*_{20}l_n,\cdots,\ \alpha^*_{T-1,0}l_n] = [\alpha_{10}l_n,\ \alpha_{20}l_n,\cdots,\alpha_{T0}l_n]F_{T,T-1}$ are the transformed time effects. To eliminate the time effects, we can further transform the n-dimensional vector Y^*_{nt} to an $(n-1)$-dimensional vector Y^{**}_{nt} as $Y^{**}_{nt} = F'_{n,n-1}Y^*_{nt}$. Such a transformation to Y^{**}_{nt} can result in a well-defined spatial panel model when W_{n1} and W_{n2} are assumed to be row-normalized. [7] Therefore, we have

$$Y^{**}_{nt} = \lambda_{01}(F'_{n,n-1}W_{n1}F_{n,n-1})Y^{**}_{nt} + X^{**}_{nt}\beta_0 + U^{**}_{nt}, \tag{13}$$
$$U^{**}_{nt} = \lambda_{02}(F'_{n,n-1}W_{n2}F_{n,n-1})U^{**}_{nt} + V^{**}_{nt},$$

for $t=1,\ldots,T-1$ where $X^{**}_{nt} = F'_{n,n-1}X^*_{nt}$ and $V^{**}_{nt} = F'_{n,n-1}V^*_{nt}$. After the transformations, the effective sample size is $(n-1)(T-1)$. It can be shown that the common parameter estimates from the transformed approach are consistent when either n or T is large, and their asymptotic distributions are properly centered (Lee and Yu, 2008).

For the random effects specification with a large T, the model is

$$Y_{nt} = l_n b_0 + z_n\eta_0 + \lambda_{01}W_{n1}Y_{nt} + X_{nt}\beta_0 + \mu_n + \alpha_{t0}l_n + U_{nt}, \tag{14}$$
$$\mu_n = \lambda_{03}W_{n3}\mu_n + c_{n0},\ \text{and}\ U_{nt} = \lambda_{02}W_{n2}U_{nt} + V_{nt},$$

for $t=1,\ldots,T$. In the vector form, it is

$$\mathbf{Y}_{nT} = l_T\otimes(l_n b_0 + z_n\eta_0) + \lambda_1(I_T\otimes W_{n1})\mathbf{Y}_{nT} + \mathbf{X}_{nT}\beta_0 + l_T\otimes C_n^{-1}\mathbf{c}_{n0} + \alpha_{T0}\otimes l_n + (I_T\otimes R_n^{-1})\mathbf{V}_{nT},$$

where $\alpha_{T0} = (\alpha_{10},\ldots,\alpha_{T0})'$. As $\mathbf{c}_{n0}$ is $(0,\sigma_c^2 I_n)$, α_{T0} is $(0,\sigma_\alpha^2 I_T)$, V_{nt} is $(0,\sigma_v^2 I_n)$, and they are uncorrelated, the variance matrix of the overall disturbances $l_T\otimes C_n^{-1}\mathbf{c}_{n0} + \alpha_{T0}\otimes l_n + (I_T\otimes R_n^{-1})\mathbf{V}_{nT}$ would be

$$\Omega_{nT} = \sigma_c^2[l_T l_T'\otimes(C_n'C_n)^{-1}] + \sigma_\alpha^2[I_T\otimes l_n l_n'] + \sigma_v^2[I_T\otimes(R_n'R_n)^{-1}].$$

This is a generalized case of Baltagi et al. (2007a) where we have the spatial lag and time effects in the main equation, in addition to the spatial effect and the individual effects in the disturbances. The log likelihood function is

$$\ln L(\mathbf{Y}_{nT}) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_{nT}| + T\ln|S_n| - \frac{1}{2}\xi'_{nT}(\theta)\Omega_{nT}^{-1}\xi_{nT}(\theta),$$

where $\xi_{nT}(\theta) = \mathbf{S}_{nT}\mathbf{Y}_{nT} - \mathbf{X}_{nT}\beta - l_T\otimes(l_n b + z_n\eta)$. The calculation of the inverse and determinant of Ω_{nT} will involve essentially those of a $T\times T$ matrix as well as an $n\times n$ matrix. As a further generalization, α_{t0} may also be serially correlated, e.g., with an AR(1) process.

3. SDPD models

Spatial panel data models can include both spatial and dynamic effects to investigate the state dependence and serial correlations. To include the time dynamic features in the spatial panel data models, an immediate approach is to use the time lag term as an explanatory variable, which is the "time–space simultaneous" case in Anselin (2001). In a simple dynamic panel data model with fixed individual effects, the MLE of the autoregressive coefficient is biased and inconsistent when n tends to infinity but T is fixed (Nickell, 1981; Hsiao, 1986). By taking time differences to eliminate the fixed effects in the dynamic equation and by the construction of instrumental variables (IVs), Anderson and Hsiao (1981) show that IV methods can provide consistent estimates. When T is finite, additional IVs can improve the efficiency of the estimation. However, if the number of IVs is too large, the problem of many IVs arises as the asymptotic bias would increase with the number of IVs.

[7] When W_{n1} and W_{n2} are not row-normalized, we can still eliminate the transformed time effects; however, we will not have the presentation of (13). In that case, the likelihood function would not be feasible, and alternative estimation methods, such as the generalized method of moment, would be appropriate.

When both n and T go to infinity, the incidental parameter problem in the MLE becomes less severe as each individual fixed effect can be consistently estimated. However, Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003) have found the existence of asymptotic bias of order $O(1/T)$ in the MLE of the autoregressive parameter when both n and T tend to infinity with the same rate. In addition to the MLE, Alvarez and Arellano (2003) also investigate the asymptotic properties of the IV estimators in Arellano and Bond (1991). They have found the presence of asymptotic bias of a similar order to that of the MLE, due to the presence of many moment conditions. As the presence of asymptotic bias is an undesirable feature of these estimates, Kiviet (1995), Hahn and Kuersteiner (2002), and Bun and Carree (2005) have constructed bias corrected estimators for the dynamic panel data model by analytically modifying the within estimator. Hahn and Kuersteiner (2002) provide a rigorous asymptotic theory for the within estimator and the bias corrected estimator when both n and T go to infinity with the same rate. As an alternative to the analytical bias correction, Hahn and Newey (2004) have also considered the Jackknife bias reduction approach.

A general SDPD model can be specified as:

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt},$$
$$t = 1, 2, \ldots, T, \tag{15}$$

where γ_0 captures the pure dynamic effect and ρ_0 captures the spatial–time effect. Due to the presence of fixed individual and time effects, X_{nt} will not include any time invariant or individual invariant regressors. Section 3.1 classifies the above SDPD model into different cases depending on the structure of eigenvalue matrix of the reduced form of Eq. (15). Section 3.2 covers the asymptotic properties for the QMLEs of different cases when T is large. When T is fixed, we need to specify the initial condition if MLE is used.[8] Section 3.3 discusses the dynamic panel model with spatial correlated disturbances, which can be treated in some situations as a special case of the general SDPD model.

3.1. Classification of SDPD models

By denoting $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, Eq. (15) can be rewritten as

$$Y_{nt} = A_n Y_{n,t-1} + S_n^{-1} X_{nt}\beta_0 + S_n^{-1}\mathbf{c}_{n0} + \alpha_{t0} S_n^{-1} l_n + S_n^{-1} V_{nt}. \tag{16}$$

Depending on the eigenvalues of A_n, we might have different DGPs of the SDPD models. As is shown below, when all the eigenvalues of A_n are smaller than 1, we have the stable case. When some eigenvalues of A_n are equal to 1 (but not all), we have the spatial cointegration case. The pure unit root case corresponds to the situation in which all the eigenvalues are 1. When some of them are greater than 1, we have the explosive case.

Let $\varpi_n = diag\{\varpi_{n1},\ldots,\varpi_{nn}\}$ be the $n\times n$ diagonal eigenvalue matrix of W_n such that $W_n = \Gamma_n \varpi_n \Gamma_n^{-1}$ where Γ_n is the corresponding eigenvector matrix. As $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, the eigenvalue matrix of A_n is $D_n = (I_n - \lambda_0 \varpi_n)^{-1}(\gamma_0 I_n + \rho_0 \varpi_n)$ so that $A_n = \Gamma_n D_n \Gamma_n^{-1}$. When W_n is row-normalized, all the eigenvalues are less than or equal to 1 in absolute value, where it definitely has some eigenvalues equal to 1 (see Ord, 1975). Let m_n be the number of unit eigenvalues of W_n, and suppose that the first m_n eigenvalues of W_n are equal to 1. Hence, D_n can be decomposed into two parts, one corresponding to the unit eigenvalues of W_n, and the other corresponding to the eigenvalues of W_n smaller than 1. Define $\mathbb{J}_n = diag\{1'_{m_n}, 0, \ldots, 0\}$ with 1_{m_n} being an $m_n \times 1$ vector of ones and $\tilde{D}_n = diag\{0,\ldots,0,d_{n,m_n+1},\ldots,d_{nn}\}$, where $|d_{ni}|<1$ is assumed[9] for $i = m_n+1,\cdots,n$. As $\mathbb{J}_n \cdot \tilde{D}_n = 0$, we have $A_n^h = \left(\frac{\gamma_0+\rho_0}{1-\lambda_0}\right)^h \Gamma_n \mathbb{J}_n \Gamma_n^{-1} + B_n^h$ where $B_n^h = \Gamma_n \tilde{D}_n^h \Gamma_n^{-1}$ for any $h = 1,2,\cdots$.

Denote $W_n^u = \Gamma_n \mathbb{J}_n \Gamma_n^{-1}$. For $t \geq 0$, Y_{nt} can be decomposed into a sum of a possible stable part, a possible unstable or explosive part, and a time effect part:

$$Y_{nt} = Y_{nt}^u + Y_{nt}^s + Y_{nt}^\alpha, \tag{17}$$

where

$$Y_{nt}^s = \sum_{h=0}^{\infty} B_n^h S_n^{-1}(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h}),$$
$$Y_{nt}^u = W_n^u \left\{ \left(\frac{\gamma_0+\rho_0}{1-\lambda_0}\right)^{t+1} Y_{n,-1} + \frac{1}{(1-\lambda_0)}\left[\sum_{h=0}^{t}\left(\frac{\gamma_0+\rho_0}{1-\lambda_0}\right)^h \times (\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h})\right]\right\},$$
$$Y_{nt}^\alpha = \frac{1}{(1-\lambda_0)} l_n \sum_{h=0}^{t} \alpha_{t-h,0}\left(\frac{\gamma_0+\rho_0}{1-\lambda_0}\right)^h.$$

The Y_{nt}^u can be an unstable component when $\frac{\gamma_0+\rho_0}{1-\lambda_0} = 1$, which occurs when $\gamma_0+\rho_0+\lambda_0 = 1$ and $\lambda_0 \neq 1$. When $\gamma_0+\rho_0+\lambda_0 > 1$, it implies $\frac{\gamma_0+\rho_0}{1-\lambda_0} > 1$, and Y_{nt}^u can be explosive. The Y_{nt}^α can be complicated, as it depends on what the time dummies exactly represent. The Y_{nt}^α can be explosive when α_{t0} represents some explosive functions of t, even when $\frac{\gamma_0+\rho_0}{1-\lambda_0}$ is smaller than 1. Without an explicit specification for α_{t0}, it is desirable to eliminate this component for estimation. The Y_{nt}^s can be a stable component unless the sum $\gamma_0+\rho_0+\lambda_0$ is much larger than 1. If $\gamma_0+\rho_0+\lambda_0$ were too large, some of the eigenvalues d_{ni} in Y_{nt}^s might become larger than 1. Hence, depending on the value of $\frac{\gamma_0+\rho_0}{1-\lambda_0}$, we have three cases:

- Stable case when $\gamma_0+\rho_0+\lambda_0 < 1$.
- Spatial cointegration case when $\gamma_0+\rho_0+\lambda_0 = 1$ but $\gamma_0 \neq 1$.
- Explosive case when $\gamma_0+\rho_0+\lambda_0 > 1$.

For the stable case, Yu et al. (2008) consider the fixed effects specification with T going to infinity. The rates of convergence of QMLEs are $\sqrt{nT}$. For the spatial cointegration case where Y_{nt} and $W_n Y_{nt}$ are spatially cointegrated, it is shown in Yu et al. (2007) that the QMLEs are $\sqrt{nT}$ consistent and asymptotically normal. However, the presence of the unstable components will make the estimators' asymptotic variance matrix singular. Yu et al. (2007) show that the sum of the spatial and dynamic effects estimates converges at a higher rate. For the explosive case, the properties of the QMLEs remain unknown. However, the estimation of the explosive case becomes tractable when a transformation can reduce the explosive variables to be stable ones. The subsequent section presents more detailed discussions.

3.2. Stable, spatial cointegration, and explosive cases

For notational purposes, we define $\tilde{Y}_{nt} = Y_{nt} - \overline{Y}_{nT}$ and $\tilde{Y}_{n,t-1} = Y_{n,t-1} - \overline{Y}_{nT,-1}$ for $t = 1, 2,\cdots,T$ where $\overline{Y}_{nT} = \frac{1}{T}\sum_{t=1}^{T} Y_{nt}$ and $\overline{Y}_{nT,-1} = \frac{1}{T}\sum_{t=1}^{T} Y_{n,t-1}$. For the stable case and the spatial cointegration case below, we will focus on the model without the time effects. We then discuss the case where the time effects are included but eliminated by the transformations J_n or $I_n - W_n$.

3.2.1. Stable case

Denote $\theta = (\delta', \lambda, \sigma^2)'$ and $\zeta = (\delta', \lambda, \mathbf{c}_n')'$ where $\delta = (\gamma, \rho, \beta')'$. At the true value, $\theta_0 = (\delta_0', \lambda_0, \sigma_0^2)'$ and $\zeta_0 = (\delta_0', \lambda_0, \mathbf{c}_{n0}')'$ where $\delta_0 = (\gamma_0,$

[8] We may also consider the estimation by the generalized method of moments where lagged dependent variables can be used as IVs. Such an approach is under consideration.

[9] We note that $d_{ni} = (\gamma_0 + \rho_0\varpi_{ni})/(1-\lambda_0\varpi_{ni})$. Hence, if $\gamma_0+\lambda_0+\rho_0<1$, we have $d_{ni}<1$ as $|\varpi_{ni}| \leq 1$. Some additional conditions are needed to ensure that $d_{ni} > -1$. See Appendix A in Lee and Yu (2009).

Please cite this article as: Lee, L., Yu, J., Some recent developments in spatial panel data models, Regional Science and Urban Economics (2009), doi:10.1016/j.regsciurbeco.2009.09.002

$\rho_0, \beta_0')'$. By denoting $Z_{nt}=(Y_{n,t-1}, W_nY_{n,t-1}, X_{nt})$, the likelihood function of Eq. (15) is

$$\ln L_{n,T}(\theta, \mathbf{c}_n) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} V_{nt}'(\zeta)V_{nt}(\zeta), \tag{18}$$

where $V_{nt}(\zeta)=S_n(\lambda)Y_{nt}-Z_{nt}\delta-\mathbf{c}_n$. The QMLEs $\hat{\theta}_{nT}$ and $\hat{c}_{nT}$ are the extremum estimators derived from the maximization of Eq. (18), and $\hat{c}_{nT}$ can be consistently estimated when T goes to infinity.

Using the first order condition for $\mathbf{c}_n$, the concentrated likelihood is

$$\ln L_{n,T}(\theta) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta), \tag{19}$$

where $\tilde{V}_{nt}(\zeta)=S_n(\lambda)\tilde{Y}_{nt}-\tilde{Z}_{nt}\delta$. The QMLE $\hat{\theta}_{nT}$ maximizes the concentrated likelihood function (19). As is shown in Yu et al. (2008), we have

$$\sqrt{nT}(\hat{\theta}_{nT}-\theta_0) + \sqrt{\frac{n}{T}}\varphi_{\theta_0,nT} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}\right)\right) \xrightarrow{d} N\left(0, \lim_{T\to\infty}\Sigma_{\theta_0,nT}^{-1}(\Sigma_{\theta_0,nT}+\Omega_{\theta_0,nT})\Sigma_{\theta_0,nT}^{-1}\right), \tag{20}$$

where $\varphi_{\theta_0,nT}$ is the leading bias term of order $O(1)$, $\Sigma_{\theta_0,nT}$ is the information matrix, and $\Omega_{\theta_0,nT}$ captures the non-normality feature of the disturbances. For the leading bias term, $\varphi_{\theta_0,nT}=\Sigma_{\theta_0,nT}^{-1}\varphi_1$ where

$$\varphi_1 = \begin{pmatrix} \frac{1}{n}tr((\sum_{h=0}^{\infty}A_n^h)S_n^{-1}) \\ \frac{1}{n}tr(W_n(\sum_{h=0}^{\infty}A_n^h)S_n^{-1}) \\ 0_{k\times 1} \\ \frac{1}{n}\gamma_0 tr(G_n(\sum_{h=0}^{\infty}A_n^h)S_n^{-1}) + \frac{1}{n}\rho_0 tr(G_nW_n(\sum_{h=0}^{\infty}A_n^h)S_n^{-1}) + \frac{1}{n}trG_n \\ \frac{1}{2\sigma_0^2} \end{pmatrix}, \tag{21}$$

and

$$\Sigma_{\theta_0,nT} = \frac{1}{\sigma_0^2}\begin{pmatrix} E\mathcal{H}_{nT} & * \\ 0_{1\times(k+3)} & 0 \end{pmatrix} + \begin{pmatrix} 0_{(k+2)\times(k+2)} & * & * \\ 0_{1\times(k+2)} & \frac{1}{n}[tr(G_n'G_n)+tr(G_n^2)] & * \\ 0_{1\times(k+2)} & \frac{1}{\sigma_0^2 n}tr(G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix} + O\left(\frac{1}{T}\right),$$

with $G_n\equiv W_nS_n^{-1}$ and $\mathcal{H}_{nT}=\frac{1}{nT}\sum_{t=1}^{T}(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)'(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)$. Hence, for distribution of the common parameters, when T is large relative to n , the estimators are $\sqrt{nT}$ consistent and asymptotically normal, with the limiting distribution centered around 0; when n is asymptotically proportional to T, the estimators are $\sqrt{nT}$ consistent and asymptotically normal, but the limiting distribution is not centered around 0; and when n is large relative to T, the estimators are T consistent, and have a degenerate limiting distribution.

3.2.2. Spatial cointegration case

The log likelihood function of the spatial cointegration model is the same as the stable case. However, the properties of the estimators are not the same. We have

$$\sqrt{nT}(\hat{\theta}_{nT}-\theta_0) + \sqrt{\frac{n}{T}}\varphi_{\theta_0,nT} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}\right)\right) \xrightarrow{d} N(0, \lim_{T\to\infty}\Sigma_{\theta_0,nT}^{-1}(\Sigma_{\theta_0,nT}+\Omega_{\theta_0,nT})\Sigma_{\theta_0,nT}^{-1}), \tag{22}$$

where $\varphi_{\theta_0,nT}\equiv\Sigma_{\theta_0,nT}^{-1}\cdot\varphi_2$ is the leading bias term of order $O(1)$ and

$$\varphi_2 = a_{\theta_0,n}^s + \frac{m_n}{n}a_{\theta_0,T}^u \tag{23}$$

with

$$a_{\theta_0,n}^s = \begin{pmatrix} \frac{1}{n}tr((\sum_{h=0}^{\infty}B_n^h)S_n^{-1}) \\ \frac{1}{n}tr(W_n(\sum_{h=0}^{\infty}B_n^h)S_n^{-1}) \\ 0_{k\times 1} \\ \frac{1}{n}\gamma_0 tr(G_n(\sum_{h=0}^{\infty}B_n^h)S_n^{-1}) + \frac{1}{n}\rho_0 tr(G_nW_n(\sum_{h=0}^{\infty}B_n^h)S_n^{-1}) + \frac{1}{n}trG_n \\ \frac{1}{2\sigma_0^2} \end{pmatrix},$$

$$a_{\theta_0,T}^u = T\cdot\frac{1}{2(1-\lambda_0)}\cdot(1,1,0_{1\times k},1,0)'.$$

The distinctive feature of the spatial cointegration case is that $\lim_{T\to\infty}\Sigma_{\theta_0,nT}^{-1}$ exists but is singular. This indicates that some linear combinations may have higher rates of convergence. Indeed, we have

$$\sqrt{nT^3}(\hat{\lambda}_{nT}+\hat{\gamma}_{nT}+\hat{\rho}_{nT}-1) + \sqrt{\frac{n}{T}}b_{\theta_0,nT} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}\right)\right) \xrightarrow{d} N(0, \lim_{T\to\infty}\sigma_{1,nT}^2).$$

Here, $\sigma_{1,nT}^2=\lim_{T\to\infty}\omega_{nT}^{-1}+\lim_{T\to\infty}T^2(1,1,0_{1\times k},1,0)(\lim_{T\to\infty}\Sigma_{\theta_0,nT}^{-1}\Omega_{\theta_0,nT}\Sigma_{\theta_0,nT}^{-1})(1,1,0_{1\times k},1,0)'$ is a positive scalar variance where $\omega_{nT}=\frac{1}{nT^3}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}^{u'}\tilde{Y}_{n,t-1}^{u}$, and $b_{\theta_0,nT}=T\cdot(1,1,0_{1\times k},1,0)\cdot\varphi_{\theta_0,nT}$ is $O(1)$.

The spatial cointegration model is related to the cointegration literature. Here, the unit roots are generated from the mixed time and spatial dimensions. The cointegration matrix is (I_n-W_n), and its rank is the number of eigenvalues of W_n being less than 1 in absolute value. Compared to conventional cointegration in time series literature, the cointegrating space is completely known and is determined by the spatial weights matrix; while in the conventional time series, it is the main object of inference. Also, in the conventional cointegration, the dimension of VAR is fixed and relatively small while the spatial dimension in the SDPD model is large. The spatial cointegration features of this case can be seen as follows. Denote the time difference as $\Delta Y_{nt}=Y_{nt}-Y_{n,t-1}$, we have, from Eq. (16),

$$\Delta Y_{nt} = (A_n-I_n)Y_{n,t-1} + S_n^{-1}(X_{nt}\beta_0+\mathbf{c}_{n0}+V_{nt}).$$

As $\gamma_0+\rho_0+\lambda_0=1$, it follows that $A_n-I_n=(I_n-\lambda_0W_n)^{-1}(\gamma_0I_n+\rho_0W_n)-I_n=(1-\gamma_0)(I_n-\lambda_0W_n)^{-1}(W_n-I_n)$. Thus, the error correction model (ECM) form is

$$\Delta Y_{nt} = (1-\gamma_0)(I_n-\lambda_0W_n)^{-1}(W_n-I_n)Y_{n,t-1} + S_n^{-1}(X_{nt}\beta_0+\mathbf{c}_{n0}+V_{nt}).$$

As $W_n=\Gamma_n\varpi_n\Gamma_n^{-1}$ and $M_n=\Gamma_n\mathbb{J}_n\Gamma_n^{-1}$, it follows that $(I_n-W_n)M_n=\Gamma_n(I_n-\varpi_n)\mathbb{J}_n\Gamma_n^{-1}=0$. Hence, $(I_n-W_n)Y_{nt}^u=0$, and also $(I_n-W_n)Y_{nt}=(I_n-W_n)Y_{nt}^s$, which depends only on the stationary component.

Therefore, Y_{nt} is spatially cointegrated. The matrix $I_n - W_n = \Gamma_n(I_n - \varpi_n)\Gamma_n^{-1}$ has its rank equal to $n - m_n$, which is the number of eigenvalues of W_n that are smaller than 1 — the cointegration rank.

3.2.3. Transformation approach of J_n: the case with time dummies

When we have time effects included in the SDPD model, the direct estimation method above will yield a bias of order $O(\max(1/n, 1/T))$ for the common parameters.[10] In order to avoid the bias of the order $O(1/n)$, we may use a data transformation approach, while the resulting estimator may have the same asymptotic efficiency as the direct QML estimator. This transformation procedure is particularly useful when $n/T \to 0$ where the estimates of the transformed approach will have a faster rate of convergence than that of the direct estimates. Also, when $n/T \to 0$, the estimates under the direct approach have a degenerate limit distribution, but the estimates under the transformation approach are properly centered and asymptotically normal.

With the transformation J_n, when $W_n l_n = l_n$, i.e., W_n is a row-normalized matrix, $J_n W_n = J_n W_n (J_n + \frac{1}{n} l_n l_n') = J_n W_n J_n$ because $J_n W_n l_n = J_n l_n = 0$. Hence,

$$(J_n Y_{nt}) = \lambda_0 (J_n W_n)(J_n Y_{nt}) + \gamma_0 (J_n Y_{n,t-1}) + \rho_0 (J_n W_n)(J_n Y_{n,t-1}) + (J_n X_{nt})\beta_0 + (J_n \mathbf{c}_{n0}) + (J_n V_{nt}), \quad (24)$$

which does not involve the time effects, and $J_n \mathbf{c}_{n0}$ can be regarded as the transformed individual effects. With the additional transformation $F_{n,n-1}$, by denoting $Y_{nt}^* = F'_{n,n-1} J_n Y_{nt} = F'_{n,n-1} Y_{nt}$, which is of the dimension $(n-1)$, we have

$$Y_{nt}^* = \lambda_0 W_n^* Y_{nt}^* + \gamma_0 Y_{n,t-1}^* + \rho_0 W_n^* Y_{n,t-1}^* + X_{nt}^* \beta_0 + \mathbf{c}_{n0}^* + V_{nt}^*, \quad (25)$$

where $W_n^* = F'_{n,n-1} W_n F_{n,n-1}$, $X_{nt}^* = F'_{n,n-1} X_{nt}$, $\mathbf{c}_{n0}^* = F'_{n,n-1} \mathbf{c}_{n0}$ and $V_{nt}^* = F'_{n,n-1} V_{nt}$. The V_{nt}^* is an $(n-1)$ dimensional disturbance vector with zero mean and variance matrix $\sigma_0^2 I_{n-1}$. Eq. (25) is in the format of a typical SDPD model, where the number of observations is $T(n-1)$, reduced from the original sample observations by one for each period. Eq. (25) is useful because a likelihood function for Y_{nt}^* can be constructed. Such a likelihood function is a partial likelihood — a terminology introduced in Cox (1975). If V_{nt} is normally distributed $N(0, \sigma_0^2 I_n)$, the transformed V_{nt}^* will be $N(0, \sigma_0^2 I_{n-1})$. Thus, the log likelihood function of Eq. (25) can be written as

$$\ln L_{n,T}(\theta, \mathbf{c}_n) = -\frac{(n-1)T}{2} \ln 2\pi - \frac{(n-1)T}{2} \ln \sigma^2 - T \ln(1-\lambda) + T \ln |I_n - \lambda W_n| - \frac{1}{2\sigma^2} \sum_{t=1}^{T} V_{nt}'(\theta) J_n V_{nt}(\theta). \quad (26)$$

As is shown in Lee and Yu (2007), the QMLE from the above maximization is free of $O(1/n)$ bias.

3.2.4. Explosive case

When some eigenvalues of A_n are greater than 1, it might be difficult to obtain the estimates in our experience. Furthermore, asymptotic properties of the QML estimates of such a case are unknown. However, the explosive feature of the model can be avoided by the data transformation $I_n - W_n$. The transformation $I_n - W_n$ can eliminate not only time dummies but also the unstable component. Hence, we end up with the following equation after the $I_n - W_n$ transformation:

$$(I_n - W_n) Y_{nt} = \lambda_0 W_n (I_n - W_n) Y_{nt} + \gamma_0 (I_n - W_n) Y_{n,t-1} + \rho_0 W_n (I_n - W_n) Y_{n,t-1} + (I_n - W_n) X_{nt} \beta_0 + (I_n - W_n) \mathbf{c}_{n0} + (I_n - W_n) V_{nt}. \quad (27)$$

This transformed equation has fewer degrees of freedom than n. Denote the degrees of freedom of Eq. (27) as n^*. Then, n^* is the rank of the variance matrix of $(I_n - W_n)V_{nt}$, which is the number of non-zero eigenvalues of $(I_n - W_n)(I_n - W_n)'$. Hence, $n^* = n - m_n$ is also the number of non-unit eigenvalues of W_n. The transformed variables do not have time effects and can be stable even when $\gamma_0 + \rho_0 + \lambda_0$ is equal to or greater than 1.

The variance of $(I_n - W_n)V_{nt}$ is $\sigma_0^2 \sum_n$, where $\sum_n = (I_n - W_n)(I_n - W_n)'$. Let $[F_n, H_n]$ be the orthonormal matrix of eigenvectors and Λ_n be the diagonal matrix of nonzero eigenvalues of $\sum_n$ such that $\sum_n F_n = F_n \Lambda_n$ and $\sum_n H_n = 0$. That is, the columns of F_n consist of eigenvectors of non-zero eigenvalues, and those of H_n are for zero-eigenvalues of $\sum_n$. The F_n is an $n \times n^*$ matrix, and Λ_n is an $n^* \times n^*$ diagonal matrix. Denote $W_n^* = \Lambda_n^{-1/2} F_n' W_n F_n \Lambda_n^{1/2}$ which is an $n^* \times n^*$ matrix. We have

$$Y_{nt}^* = \lambda_0 W_n^* Y_{nt}^* + \gamma_0 Y_{n,t-1}^* + \rho_0 W_n^* Y_{n,t-1}^* + X_{nt}^* \beta_0 + \mathbf{c}_{n0}^* + V_{nt}^*, \quad (28)$$

where $Y_{nt}^* = \Lambda_n^{-1/2} F_n'(I_n - W_n) Y_{nt}$ and other variables are defined accordingly. Note that this transformed Y_{nt}^* is an n^*-dimensional vector. The eigenvalues of W_n^* are exactly those eigenvalues of W_n less than 1 in absolute value. It follows that the eigenvalues of $A_n^* = (I_{n^*} - \lambda_0 W_n^*)^{-1}(\gamma_0 I_{n^*} + \rho_0 W_n^*)$ are all less than 1 in absolute values even when $\gamma_0 + \rho_0 + \lambda_0 = 1$ with $|\lambda_0| < 1$ and $|\gamma_0| < 1$. For the explosive case with $\gamma_0 + \rho_0 + \lambda_0 > 1$, all the eigenvalues of A_n^* can be less than 1 only if $\frac{\rho_0 + \lambda_0}{1 - \gamma_0} < \frac{1}{\varpi_{max}}$, where ϖ_{max} is the maximum positive eigenvalue of W_n less than 1. Hence, the transformed model (28) is a stable one as long as $\gamma_0 + \rho_0 + \lambda_0$ is not too much larger than 1.

For the concentrated log likelihood of Eq. (28), it is

$$\ln L_{n,T}(\theta) = -\frac{n^* T}{2} \ln 2\pi - \frac{n^* T}{2} \ln \sigma^2 - (n - n^*) T \ln(1-\lambda) + T \ln |I_n - \lambda W_n| - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \tilde{V}_{nt}'(\theta)(I_n - W_n)' \sum\nolimits_n^{+} (I_n - W_n) \tilde{V}_{nt}(\theta), \quad (29)$$

where $\tilde{V}_{nt}(\theta) = S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta$. From Lee and Yu (2009), we have similar results to those of Yu et al. (2008) for the stable model, where the bias term and the variance term would involve only the stable component that is left after the $I_n - W_n$ transformation.[11]

Therefore, we can use the spatial difference operator, $I_n - W_n$, which may eliminate not only the time effects, but also the possible unstable or explosive components that are generated from the spatial cointegration or explosive roots. This implies that the spatial difference transformation can be applied to DGPs with stability, spatial cointegration or explosive roots. The asymptotics of the resulting estimates can then be easily established for these DGPs. Thus, the transformation $I_n - W_n$ provides a unified estimation procedure for the estimation of the SDPD models.

[10] This bias has been worked out for the stable case in Lee and Yu (2007). For the spatial cointegration case, Yu et al. (2007) have not considered the model with time dummies. However, we would expect the presence of such a bias order for the spatial cointegration case.

[11] We note that the spatial difference operator $I_n - W_n$ can also be applied to cross sectional units. However, its function is different from the time difference operator for a time series. The spatial difference operator does not eliminate the pure time series unit roots or explosive roots.

Please cite this article as: Lee, L., Yu, J., Some recent developments in spatial panel data models, Regional Science and Urban Economics (2009), doi:10.1016/j.regsciurbeco.2009.09.002

3.2.5. Bias correction

For each case, we may propose a bias correction for the estimators, which would be valuable for moderately large T. For the stable model with only individual effects, the bias is $\varphi_{\theta_0,nT}=\Sigma_{\theta_0,nT}^{-1}\varphi_1$ where φ_1 is in Eq. (21); for the spatial cointegration case, the bias is $\varphi_{\theta_0,nT}=\Sigma_{\theta_0,nT}^{-1}\varphi_2$ where φ_2 is in Eq. (23). For the stable case with the transformation J_n, the bias is $\varphi_{\theta_0,nT}=\Sigma_{\theta_0,nT}^{-1}\varphi_3$ where

$$\varphi_3=\begin{pmatrix} \frac{1}{n-1}tr((J_n\sum_{h=0}^{\infty}A_n^h)S_n^{-1}) \\ \frac{1}{n-1}tr(W_n(J_n\sum_{h=0}^{\infty}A_n^h)S_n^{-1}) \\ 0_{k\times 1} \\ \frac{1}{n-1}\gamma_0 tr(G_n(J_n\sum_{h=0}^{\infty}A_n^h)S_n^{-1})+\frac{1}{n-1}\rho_0 tr(G_nW_n(J_n\sum_{h=0}^{\infty}A_n^h)S_n^{-1})+\frac{1}{n-1}tr(J_nG_n) \\ \frac{1}{2\sigma_0^2} \end{pmatrix}. \tag{30}$$

For the unified transformation approach, the bias is $\varphi_{\theta_0,nT}=\Sigma_{\theta_0,nT}^{-1}\varphi_4$ and

$$\varphi_4=\begin{pmatrix} \frac{1}{n^*}tr((J_n^*\sum_{h=0}^{\infty}B_n^h)S_n^{-1}) \\ \frac{1}{n^*}tr(W_n(J_n^*\sum_{h=0}^{\infty}B_n^h)S_n^{-1}) \\ 0_{k\times 1} \\ \frac{1}{n^*}\gamma_0 tr(G_n(J_n^*\sum_{h=0}^{\infty}B_n^h)S_n^{-1})+\frac{1}{n^*}\rho_0 tr(G_nW_n(J_n^*\sum_{h=0}^{\infty}B_n^h)S_n^{-1})+\frac{1}{n^*}trG_n^* \\ \frac{1}{2\sigma_0^2} \end{pmatrix}, \tag{31}$$

where $J_n^*=(I_n-W_n)'\Sigma_n^{+}(I_n-W_n)$.

Hence, the QMLE $\hat{\theta}_{nT}$ has the bias $-\frac{1}{T}\varphi_{\theta_0,nT}$ and the confidence interval is not centered when $\frac{n^*}{T}\to c$ where n^* is the corresponding degrees of freedom in each model for some finite positive constant c. Furthermore, when T is small relative to n in the sense that $\frac{n}{T}\to\infty$, the presence of $\varphi_{\theta_0,nT}$ causes $\hat{\theta}_{nT}$ to have the slower T-rate of convergence. An analytical bias reduction procedure is to correct the bias $B_{nT}=-\varphi_{\theta_0,nT}$ by constructing an estimate $\hat{B}_{nT}$. The bias corrected estimator is

$$\hat{\theta}_{nT}^1=\hat{\theta}_{nT}-\frac{\hat{B}_{nT}}{T}. \tag{32}$$

We may choose[12]

$$\hat{B}_{nT}=\left[\left(E\left(\frac{1}{nT}\frac{\partial^2\ln L_{n,T}(\theta)}{\partial\theta\partial\theta'}\right)\right)^{-1}\varphi_i(\theta)\right]\Bigg|_{\theta=\hat{\theta}_{nT}}, \tag{33}$$

where $i=1, 2, 3, 4$ corresponds to stable, spatial cointegration, J_n-transformed and (I_n-W_n)-transformed models. When T grows faster than $n^{*1/3}$, the correction will eliminate the bias of order $O(T^{-1})$ and yield a properly centered confidence interval.

3.3. Dynamic panel data models with SAR disturbances

Elhorst (2005), Su and Yang (2007), and Yu and Lee (2007) consider the estimation of a dynamic panel data model with spatial disturbances

$$\begin{aligned} Y_{nt}&=\gamma_0Y_{n,t-1}+X_{nt}\beta_0+z_n\eta_0+U_{nt}, t=1,\dots,T, \\ U_{nt}&=\mu_n+\varepsilon_{nt},\ \text{and}\ \varepsilon_{nt}=\lambda_0W_n\varepsilon_{nt}+V_{nt}. \end{aligned} \tag{34}$$

When T is moderate, this model with $|\gamma_0|<1$ can be estimated by the methods discussed in Section 2, because the dynamic specification in Eq. (34) can be transformed to $Y_{nt}=\lambda_0W_nY_{nt}+\gamma_0Y_{n,t-1}-\gamma_0\lambda_0W_nY_{n,t-1}+X_{nt}\beta_0-W_nX_{nt}\lambda_0\beta_0+\mathbf{c}_{n0}+V_{nt}$. This corresponds to an SDPD model with transformed individual effects $\mathbf{c}_{n0}=(I_n-W_n)^{-1}\mu_n$, nonlinear constraints $\rho_0=-\gamma_0\lambda_0$, and $\mathbb{X}_{nt}\beta_0=X_{nt}\beta_0-W_nX_{nt}\lambda_0\beta_0$ with $\mathbb{X}_{nt}=[X_{nt}, W_nX_{nt}]$ and $\beta_0=[\beta_0, -\lambda_0\beta_0]$. The case $\gamma_0=1$ is special in the sense that the model is a pure unit root case in the time dimension with spatial disturbances. We shall discuss the estimation of such a case in a subsequent paragraph.

Elhorst (2005) and Su and Yang (2007) have focused on estimating the short panel case, i.e., n is large but T is fixed. Elhorst (2005) uses the first difference to eliminate the fixed individual effects in μ_n, and Su and Yang (2007) derive the asymptotic properties of QMLEs using both the random and fixed effects specifications. As T is fixed and we have the dynamic feature, the specification of the initial observation Y_{n0} is important. When Y_{n0} is assumed to be exogenous, the likelihood function can be obtained easily, either for the random effects specification, or for the fixed effects specification where the first difference is made to eliminate the individual effects. When Y_{n0} is assumed to be endogenous, Y_{n0} will need to be generated from a stationary process, or its distribution will be approximated. With the corresponding likelihood, QMLE can be obtained.[13]

3.3.1. Pure unit root case

In Yu and Lee (2007) for the SDPD model, when $\gamma_0=1$ and $\rho_0+\lambda_0=0$, we have $A_n=I_n$ in Eq. (16). Hence, the eigenvalues of A_n have no relation with the eigenvalues of W_n because all of them are equal to 1. We term this model a unit root SDPD model. This model includes the unit root panel model with SAR disturbances in Eq. (34) as a special case. The likelihood of the unit root SDPD model without imposing the constraints $\gamma_0=1$ and $\rho_0+\lambda_0=0$ is similar to the stable case in Eq. (19), but the asymptotic distributions of the estimates are different.

For the unit root SDPD model, the estimate of the pure dynamic coefficient γ_0 is $\sqrt{nT^3}$ consistent and the estimates of all the other parameters are $\sqrt{nT}$ consistent; and they are asymptotically normal. Also, the sum of the contemporaneous spatial effect estimate of λ_0 and the dynamic spatial effect estimate of ρ_0 will converge at $\sqrt{nT^3}$ rate. The rates of convergence of the estimates can be compared with those of the spatial cointegration case in Yu et al. (2007). For the latter, all the estimates of parameters including γ_0 are $\sqrt{nT}$ consistent; only the sum of the pure dynamic and spatial effects estimates is convergent at the faster $\sqrt{nT^3}$ rate. Also, there are differences in the bias orders of estimates. For the spatial cointegration case, the biases of all the estimates have the order $O(1/T)$. But for the unit root SDPD model, the bias of the estimate of γ_0 is of the smaller order $O(1/T^2)$, while the order of biases for all the other estimates have the same $O(1/T)$ order. These differences are due to different asymptotic behaviors of the two models, even though both models involve unit eigenvalues in A_n. The unit eigenvalues of the unit root SDPD model are not linked to the eigenvalues of the spatial weights matrix. On the contrary, for the spatial cointegration model, the unit eigenvalues correspond exactly to the unit eigenvalues of the spatial weights matrix via a well defined relation. For the unit roots SDPD model, the outcomes of different spatial units do not show co-movements. For the spatial cointegration model, the outcomes of different spatial units can be cointegrated with a reduced rank, where the rank is the number of eigenvalues of W_n different from 1.

[12] An asymptotically equivalent alternative way is to replace $\Sigma_{\theta_0,nT}^{-1}$ by the empirical Hessian matrix of the concentrated log likelihood function.

[13] Mutl (2006) suggests feasible generalized 2SLS approach for the estimation of dynamic panel data model with fixed effects and SAR disturbances after first-difference of data. His feasible 2SLS is based on three steps, which extends the three steps feasible GLS approach in Kapoor et al. (2007) for the panel regression model with random component and SAR disturbances to the estimation of dynamic panel model. Tao (2005) considers the SDPD model with fixed effects where the disturbances are i.i.d. and suggests the use of 2SLS for the estimation. His 2SLS is also applied to the equation after first-difference.

Please cite this article as: Lee, L., Yu, J., Some recent developments in spatial panel data models, Regional Science and Urban Economics (2009), doi:10.1016/j.regsciurbeco.2009.09.002

3.3.2. Random effects specification with a fixed T

For Eq. (34) under the random effects specification, as shown in Su and Yang (2007), the variance matrix of the disturbances is $\sigma_v^2\Omega_{nT}=\sigma_v^2[\phi_\mu(l_Tl_T'\otimes I_n)+I_T\otimes(S_n'S_n)^{-1}]$ where $\phi_\mu=\frac{\sigma_\mu^2}{\sigma_v^2}$. There are two cases under this specification.

Case 1. Y_{n0} is exogenous. Let $\theta=(\beta',\eta',\gamma)'$, $\delta=(\lambda,\phi_\mu)'$ and $\varsigma=(\theta',\sigma_v^2,\delta')'$. The log likelihood is

$$\ln L(\varsigma)=-\frac{nT}{2}\ln(2\pi)-\frac{nT}{2}\ln(\sigma_v^2)-\frac{1}{2}\ln|\Omega_{nT}|-\frac{1}{2\sigma_v^2}\mathbf{u}_{nT}'(\theta)\Omega_{nT}^{-1}\mathbf{u}_{nT}(\theta),$$

where $\mathbf{u}_{nT}(\theta)=\mathbf{Y}_{nT}-\gamma\mathbf{Y}_{nT,-1}-\mathbf{X}_{nT}\beta-l_T\otimes z_n\eta$ with $\mathbf{Y}_{nT}=(Y_{n1}',...,Y_{nT}')'$ and other variables in the vector form are similarly defined. By concentration, we can work on the log likelihood with δ

$$\ln L(\delta)=-\frac{nT}{2}(\ln(2\pi)+1)-\frac{nT}{2}\ln[\hat{\sigma}_v^2(\delta)]-\frac{1}{2}\ln|\Omega_{nT}|,$$

where $\hat{\sigma}_v^2(\delta)=\frac{1}{nT}\tilde{\mathbf{u}}_{nT}'(\delta)\Omega_{nT}^{-1}\tilde{\mathbf{u}}_{nT}(\delta)$, $\tilde{\mathbf{u}}_{nT}(\delta)=\mathbf{Y}_{nT}-\mathbf{Z}_{nT}\hat{\theta}(\delta)$ with $Z_{nT}=(X_{nT},l_T\otimes z_n,\mathbf{Y}_{nT,-1})$ and $\hat{\theta}=[Z_{nT}'\Omega_{nT}^{-1}\mathbf{Z}_{nT}]^{-1}\mathbf{Z}_{nT}'\Omega_{nT}^{-1}\mathbf{Y}_{nT}$.

Case 2. Y_{n0} is endogenous. Eq. (34) implies that $Y_{n0}=\tilde{Y}_{n0}+\zeta_{n0}$ where $\tilde{Y}_{n0}$ is the exogenous part of Y_{n0} and ζ_{n0} is the endogenous part. The exogenous part $\tilde{Y}_{n0}$ is $\sum_{j=0}^{\infty}\gamma_0^jX_{n,t-j}\beta_0+\frac{z_n\eta_0}{1-\gamma_0}$, and the endogenous part ζ_{n0} is $\frac{\mu_n}{1-\gamma_0}+\sum_{j=0}^{\infty}\gamma_0^jS_n^{-1}V_{n,t-j}$. The difficulty to use this directly is due to the missing observations X_{nt} for $t<0$. Under this situation, Su and Yang (2007) suggest the use of the Bhargava and Sargan (1983) approximation where the initial value is specified as $Y_{n0}=X_{nT}\pi+\epsilon_n$ with $X_{nT}=[l_n,\mathbb{X}_{n,T+1},z_n]$, $\mathbb{X}_{n,T+1}=[X_{n0},...,X_{nT}]$ and $\pi=(\pi_0,\pi_1',\pi_2)'$, or, $X_{nT}=[l_n,\bar{\mathbb{X}}_{n,T+1},z_n]$ and $\bar{\mathbb{X}}_{n,T+1}=\frac{1}{T}\sum_{t=0}^{T}X_{nt}$. The disturbances of the initial period are specified as $\epsilon_n=\zeta_n+\zeta_{n0}=\zeta_n+\frac{\mu_n}{1-\gamma_0}+\sum_{j=0}^{\infty}\gamma_0^jS_n^{-1}V_{n,t-j}$ where ζ_n is $(0,\sigma_\zeta^2I_n)$. The ϵ_n has mean zero, its variance matrix is $E(\epsilon_n\epsilon_n')=\sigma_\zeta^2I_n+\frac{\sigma_\mu^2}{(1-\gamma_0)^2}I_n+\frac{\sigma_v^2}{1-\gamma_0^2}(S_n'S_n)^{-1}$, and its covariance with $\mathbf{u}_{nT}$ is $E(\epsilon_n\mathbf{u}_{nT}')=\frac{\sigma_\mu^2}{1-\gamma_0}l_T'\otimes I_n$. The motivation is that $X_{nT}\pi+\zeta_n$ approximates $\tilde{Y}_{n0}$. Hence, the disturbances vector would be $\mathbf{u}_{n,T+1}^*=(\epsilon_n',\mathbf{u}_{nT}')'$ where $\mathbf{u}_{nT}$ is from Case 1. Its variance matrix is $\sigma_v^2\Omega_{n,T+1}^*$ with the dimension $n(T+1)\times n(T+1)$ where

$$\sigma_v^2\Omega_{n,T+1}^*=\begin{pmatrix}\sigma_\zeta^2I_n+\frac{\sigma_\mu^2}{(1-\gamma_0)^2}I_n+\frac{\sigma_v^2}{1-\gamma_0^2}(S_n'S_n)^{-1} & \frac{\sigma_\mu^2}{1-\gamma_0}l_T'\otimes I_n\\ \frac{\sigma_\mu^2}{1-\gamma_0}l_T\otimes I_n & \sigma_v^2\Omega_{nT}\end{pmatrix}.$$

Let $\theta=(\beta',\eta',\pi')'$, $\delta=(\gamma,\lambda,\phi_\mu,\sigma_\zeta^2)'$ and $\varsigma=(\theta',\sigma_v^2,\delta')'$. The log likelihood is

$$\ln L(\varsigma)=-\frac{n(T+1)}{2}\ln(2\pi)-\frac{n(T+1)}{2}\ln(\sigma_v^2)-\frac{1}{2}\ln|\Omega_{n,T+1}^*|-\frac{1}{2\sigma_v^2}\mathbf{u}_{n,T+1}^{*\prime}(\theta)\Omega_{n,T+1}^{*-1}\mathbf{u}_{n,T+1}^*(\theta).$$

3.3.3. Fixed effects specification with a fixed T

As is discussed in Elhorst (2005) and Su and Yang (2007), the model may also be first differenced to eliminate the individual effects. Thus, we have

$$\Delta Y_{nt}=\gamma_0\Delta Y_{n,t-1}+\Delta X_{nt}\beta_0+S_n^{-1}\Delta V_{nt},$$

for $t=2,...,T$, and the difference of the first two periods is specified to be $\Delta Y_{n1}=\Delta X_{nT}\pi+e_n$, where $\Delta X_{nT}=[l_n,X_{n1}-X_{n0},...,X_{nT}-X_{n,T-1}]$ or $\Delta X_{nT}=[l_n,\frac{1}{T}\sum_{t=1}^{T}(X_{nt}-X_{n,t-1})]$. Here, e_n is specified as $(\xi_{n1}-E(\xi_{n1}|\Delta X_{nT}))+\sum_{j=0}^{m}(\gamma_0^jS_n^{-1}\Delta V_{n,1-j})$ where $\xi_{n1}-E(\xi_{n1}|\Delta X_{nT})$ is assumed to be $(0,\sigma_e^2I_n)$. With this specification, we have $E(e_n|\Delta X_{nT})=0$ and $E(e_ne_n')=\sigma_e^2I_n+\sigma_v^2c_m(S_n'S_n)^{-1}$, where σ_e^2 and c_m are parameters to be estimated. Also, for the correlation of e_n with $\Delta u_{nt}=S_n^{-1}\Delta V_{nt}$ for $t=2,...,T$, we have $E(e_n\Delta u_{n2}')=-\sigma_v^2(S_n'S_n)^{-1}$ and $E(e_n\Delta u_{nt}')=0$ for $t\geq3$. Therefore, the variance matrix of the disturbances vector $\Delta\mathbf{u}_{nT}=(e_n',\Delta u_{n2}',...,\Delta u_{nT}')'$ is

$$\mathrm{var}(\Delta\mathbf{u}_{nT})=\sigma_v^2(I_T\otimes S_n^{-1})H_E(I_T\otimes S_n'^{-1})\equiv\sigma_v^2\Omega_{nT},$$

where

$$H_E=\begin{pmatrix}E_n & -I_n & 0 & \cdots & 0\\ -I_n & 2I_n & -I_n & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & -I_n\\ 0 & \cdots & 0 & -I_n & 2I_n\end{pmatrix},$$

and $E_n=\frac{\sigma_e^2}{\sigma_v^2}(I_n+c_m(S_n'S_n)^{-1})$. The log likelihood is

$$\ln L(\varsigma)=-\frac{nT}{2}\ln(2\pi)-\frac{nT}{2}\ln(\sigma_v^2)-\frac{1}{2}\ln|\Omega_{nT}|-\frac{1}{2\sigma_v^2}\Delta\mathbf{u}_{nT}'(\theta)\Omega_{nT}^{-1}\Delta\mathbf{u}_{nT}(\theta),$$

where

$$\Delta\mathbf{u}_{nT}(\theta)=\begin{pmatrix}\Delta Y_{n1}-\Delta X_{nT}\pi\\ \Delta Y_{n2}-\rho\Delta Y_{n1}-\Delta X_{n2}\beta\\ \vdots\\ \Delta Y_{nT}-\rho\Delta Y_{nT}-\Delta X_{nT}\beta\end{pmatrix}.$$

As is shown in Su and Yang (2007), the ML estimates under both random and fixed effects specifications are consistent and asymptotically normally distributed, under the assumption that the specification of ΔY_{n1} is correct. In principle, one could show that the estimates would not be consistent for a short panel if the initial specification were misspecified. Elhorst (2005) and Su and Yang (2007) have provided some Monte Carlo results to demonstrate that their proposed approximation could be valuable.

4. Monte Carlo and empirical illustrations

4.1. Monte Carlo

We report a small scale Monte Carlo experiment on the performance of estimates under different settings and consequences of possible model misspecifications.

4.1.1. Static spatial panel data models

For the static spatial panel model, we will generate the data according to

$$Y_{nt}=\lambda_0W_nY_{nt}+X_{nt}\beta_0+\mu_n+\alpha_tl_n+U_{nt},\; U_{nt}=\rho_0W_nU_{nt}+V_{nt},$$
$$t=1,2,...,T. \quad (35)$$

The direct approach and the transformation approach will be compared. We also check the consequence of omitting time effects when the DGP has them. The results are summarized in Table 1. We use $T=10$, 50, $n=16$, 49 and $\theta_0=(1,0.2,0.5,1)'$ where $\theta_0=(\beta_0',\lambda_0,\rho_0,\sigma_0^2)'$. The X_{nt}, μ_n, $\alpha_T=(\alpha_1,\alpha_2,\cdots,\alpha_T)$ and V_{nt} are generated from independent standard normal distributions, and the spatial weights matrix W_n is a rook matrix. For each set of generated sample observations, we calculate the ML estimator $\hat{\theta}_{nT}$ and evaluate the bias $\hat{\theta}_{nT}-\theta_0$. We do this 1000 times to have $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_{nT}-\theta_0)_i$ as the bias. For each case, we report bias (Bias), empirical standard deviation (E-SD) and root mean square error (RMSE).

Please cite this article as: Lee, L., Yu, J., Some recent developments in spatial panel data models, Regional Science and Urban Economics (2009), doi:10.1016/j.regsciurbeco.2009.09.002

Table 1
Static spatial panel data models.

	T	n		β	λ	ρ	σ²	
DGP with no time effect, direct approach (transformation approach)								
(1a)	10	49	Bias	−0.0005	0.0040	−0.0110	−0.1104	(−0.0116)
			E-SD	0.0492	0.0948	0.0939	0.0633	(0.0704)
			RMSE	0.0492	0.0949	0.0945	0.1273	(0.0713)
(1b)	50	16	Bias	−0.0010	0.0021	−0.0050	−0.0278	(−0.0079)
			E-SD	0.0380	0.0692	0.0660	0.0525	(0.0536)
			RMSE	0.0380	0.0692	0.0662	0.0594	(0.0542)
(1c)	50	49	Bias	−0.0009	−0.0011	−0.0004	−0.0224	(−0.0025)
			E-SD	0.0220	0.0405	0.0401	0.0298	(0.0305)
			RMSE	0.0220	0.0405	0.0401	0.0373	(0.0306)
DGP with time effect, direct approach								
(2a)	10	49	Bias	0.0038	0.0241	−0.0779	−0.1151	
			E-SD	0.0488	0.0856	0.0910	0.0623	
			RMSE	0.0489	0.0889	0.1198	0.1308	
(2b)	50	16	Bias	0.0038	0.0262	−0.1964	−0.0608	
			E-SD	0.0377	0.0496	0.0551	0.0498	
			RMSE	0.0379	0.0561	0.2040	0.0786	
(2c)	50	49	Bias	0.0030	0.0195	−0.0671	−0.0272	
			E-SD	0.0217	0.0365	0.0385	0.0291	
			RMSE	0.0219	0.0413	0.0774	0.0398	
DGP with time effect, transformation approach								
(3a)	10	49	Bias	−0.0001	0.0056	−0.0137	−0.0124	
			E-SD	0.0500	0.0986	0.1031	0.0706	
			RMSE	0.0500	0.0988	0.1040	0.0717	
(3b)	50	16	Bias	−0.0011	0.0019	−0.0046	−0.0093	
			E-SD	0.0393	0.0755	0.0845	0.0540	
			RMSE	0.0393	0.0755	0.0846	0.0548	
(3c)	50	49	Bias	−0.0009	−0.0011	−0.0002	−0.0026	
			E-SD	0.0222	0.0422	0.0434	0.0305	
			RMSE	0.0222	0.0423	0.0434	0.0306	
DGP with time effect, omitted in the estimation, direct (transformation)								
(4a)	10	49	Bias	−0.0582	−0.0890	0.1850	−0.1359	(−0.0399)
			E-SD	0.0567	0.2887	0.2910	0.0757	(0.0841)
			RMSE	0.0812	0.3021	0.3448	0.1556	(0.0931)
(4b)	50	16	Bias	−0.0585	−0.1612	0.2747	−0.0517	(−0.0324)
			E-SD	0.0406	0.1073	0.0945	0.0570	(0.0582)
			RMSE	0.0712	0.1937	0.2905	0.0770	(0.0666)
(4c)	50	49	Bias	−0.0745	−0.2231	0.3226	−0.0746	(−0.0557)
			E-SD	0.0239	0.0695	0.0610	0.0333	(0.0339)
			RMSE	0.0782	0.2337	0.3283	0.0817	(0.0652)

Note: $\theta_0 = (1, 0.2, 0.5, 1)'$.

For the DGP with only individual effects, from item (1a)–(1c), we see that both approaches provide the same estimate of $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$ while the estimator of σ_0^2 by the direct approach has a larger bias. When T is small, the transformation approach yields a consistent estimator of σ_0^2 while the direct approach does not. The Biases, E-SDs, RMSEs for the estimators of ζ_0 are small when either n or T is large. Also, when T is larger, the bias of the estimator of σ_0^2 by the direct approach decreases. For the DGP with both individual and time effects, from (3a)–(3c), we see that the bias of the transformation approach is small when either n or T is large. For the direct approach, from (2a)–(2c), the bias for the common parameter ζ_0 is small when n is large, and is large when n is small and T might be large; while the bias for the estimate of σ_0^2 is small only when both n and T are large. Also, from (4a)–(4c), when we omit the time effects in the regression, we have much larger bias for the spatial effects coefficients λ_0 and ρ_0 from both the direct and transformation approaches. The biases for λ_0 are downward but those for ρ_0 are upward, and the absolute biases increase as T increases.

4.1.2. SDPD models

We also run simulations to check the performance of the SDPD estimators. The true DGP is a stable SDPD model with time effects

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt}, \tag{36}$$

Table 2
Stable SDPD models: before bias correction.

T	n		γ	ρ	β	λ	σ²
(1) Direct estimation							
20	54	Bias	−0.0286	0.0083	−0.0010	−0.0381	−0.0696
		E-SD	0.0213	0.0453	0.0305	0.0376	0.0401
		RMSE	0.0390	0.0627	0.0418	0.0615	0.0853
(2) Transformation by $F_{n,n-1}$							
20	54	Bias	−0.0302	−0.0018	−0.0015	−0.0034	−0.0538
		E-SD	0.0215	0.0458	0.0307	0.0383	0.0420
		RMSE	0.0402	0.0622	0.0420	0.0525	0.0759
(3) W_nY_{nt} omitted; transformation by $F_{n,n-1}$							
20	54	Bias	−0.0217	0.0628	0.0017	–	−0.0302
		E-SD	0.0217	0.0446	0.0311	–	0.0421
		RMSE	0.0351	0.0870	0.0426	–	0.0649
(4) $W_nY_{n,t-1}$ omitted; transformation by $F_{n,n-1}$							
20	54	Bias	−0.0535	–	−0.0184	0.4551	0.1744
		E-SD	0.0175	–	0.0342	0.0181	0.1388
		RMSE	0.0595	–	0.0497	0.4554	0.2257
(5) $Y_{n,t-1}$ omitted; transformation by $F_{n,n-1}$							
20	54	Bias	–	−0.0696	−0.0265	0.4523	0.2154
		E-SD	–	0.0236	0.0347	0.0188	0.1393
		RMSE	–	0.0830	0.0530	0.4528	0.2589
(6) Both W_nY_{nt} and $W_nY_{n,t-1}$ omitted; transformation by $F_{n,n-1}$							
20	54	Bias	0.0024	–	0.0017	–	0.0038
		E-SD	0.0217	–	0.0316	–	0.0436
		RMSE	0.0309	–	0.0433	–	0.0622
(7) $Y_{n,t-1}$ and $W_nY_{n,t-1}$ omitted; direct							
20	54	Bias	–	–	-0.0156	0.0350	0.0088
		E-SD	–	–	0.0316	0.0358	0.0467
		RMSE	–	–	0.0458	0.0650	0.0643
(8) α_t omitted							
20	54	Bias	−0.0569	−0.1902	−0.0183	0.4511	0.1726
		E-SD	0.0230	0.0307	0.0342	0.0187	0.1376
		RMSE	0.0624	0.1948	0.0496	0.4515	0.2236
(9) Transformation by $I_n - W_n$							
20	49	Bias	−0.0306	−0.0034	−0.0023	−0.0092	−0.0561
		E-SD	0.0249	0.0936	0.0334	0.0767	0.0439
		RMSE	0.0438	0.1266	0.0460	0.1066	0.0808

Note: $\theta_0 = (0.2, 0.2, 1, 0.2, 1)'$ where $\gamma_0 + \rho_0 + \lambda_0 = 0.6$.

using $\theta_0 = (\gamma_0, \rho_0, \beta_0', \lambda_0, \sigma_0^2)' = (0.2, 0.2, 1, 0.2, 1)'$. We estimate the model with the direct approach, the transformation approaches with $F_{n,n-1}$ and $(I_n - W_n)$, and several misspecifications of the model where some spatial effects or time dynamics are omitted. The spatial weights matrix is a block diagonal matrix formed by a row-normalized queen matrix, where we have 6 blocks of a 9×9 queen matrix. Hence, the number of the unit roots in W_n is 6. Due to space limitations, we will present the case with $n = 54$ and $T = 20$.[14] The results are in Tables 2 and 3. From items (1) and (2), we can see that both the direct and the transformation approaches yield consistent estimates. In the simulation, as n is large, the $O(1/n)$ bias of the estimates from the direct approach in item (1) is not obvious. If we have some omitted spatial or dynamic explanatory variables in Eq. (36), the bias of the estimates might be large, regardless of the bias correction procedure. In item (3), the spatial lag is omitted, which results in a larger bias in $\hat{\rho}_{nT}$, and the bias correction makes the bias even larger. In items (4) and (5) where the spatial time lag or the time lag is omitted, the resulting biases in $\hat{\lambda}_{nT}$ and $\hat{\sigma}^2_{nT}$ are so large that the estimates are not informative at all. In items (6) and (7), we have two such explanatory variables omitted, and the biases are mild. As we can see from item (8), the omission of the time effects will cause a large bias in the estimates of the included spatial effects λ_0 and ρ_0, which calls for inclusion of time effects in the model. Also, from item (9), we see that the $I_n - W_n$ transformation performs well.

We also present the simulation of the SDPD model that is not stable in Tables 4 and 5. The DGP is a spatial cointegration case from Eq. (36)

[14] We generated the data with $20 + T$ periods and then took the last T periods as the sample. The initial value is generated as $N(0, I_n)$ in the simulation.

Table 3
Stable SDPD models: after bias correction.

T	n		γ	ρ	β	λ	σ^2
(1) Direct estimation							
20	54	Bias	−0.0002	0.0006	−0.0007	−0.0045	−0.0078
		E-SD	0.0220	0.0470	0.0315	0.0369	0.0428
		RMSE	0.0302	0.0639	0.0426	0.0512	0.0598
(2) Transformation by $F_{n,n-1}$							
20	54	Bias	−0.0005	−0.0012	0.0004	−0.0028	−0.0065
		E-SD	0.0220	0.0473	0.0315	0.0384	0.0409
		RMSE	0.0302	0.0642	0.0426	0.0526	0.0583
(3) W_nY_{nt} omitted; transformation by $F_{n,n-1}$							
20	54	Bias	0.1888	−0.1512	0.0135	–	0.0198
		E-SD	0.0230	0.0471	0.0329	–	0.0477
		RMSE	0.1958	0.1694	0.0464	–	0.0655
(4) W_nY_{nt-1} omitted; transformation by $F_{n,n-1}$							
20	54	Bias	−0.0285	–	−0.0162	0.4530	0.2344
		E-SD	0.0181	–	0.0350	0.0182	0.1383
		RMSE	0.0438	–	0.0498	0.4534	0.2741
(5) $Y_{n,t-1}$ omitted; transformation by $F_{n,n-1}$							
20	54	Bias	–	−0.0571	−0.0266	0.4533	0.2756
		E-SD	–	0.0243	0.0356	0.0188	0.1389
		RMSE	–	0.0764	0.0537	0.4538	0.3104
(6) Both W_nY_{nt} and W_nY_{nt-1} omitted; transformation by $F_{n,n-1}$							
20	54	Bias	0.0337	–	0.0037	–	0.0538
		E-SD	0.0217	–	0.0317	–	0.0437
		RMSE	0.0439	–	0.0434	–	0.0781
(7) $Y_{n,t-1}$ and W_nY_{nt-1} omitted; transformation							
20	54	Bias	–	–	−0.0159	0.0442	0.0837
		E-SD	–	–	0.0328	0.0386	0.0530
		RMSE	–	–	0.0463	0.0701	0.1044
(8) α_t omitted							
20	54	Bias	−0.0247	−0.2059	−0.0162	0.4509	0.2313
		E-SD	0.0236	0.0316	0.0350	0.0189	0.1371
		RMSE	0.0397	0.2087	0.0497	0.4514	0.2709
(9) Transformation by $I_n - W_n$							
20	49	Bias	−0.0010	−0.0031	−0.0005	−0.0087	−0.0088
		E-SD	0.0255	0.0961	0.0341	0.0735	0.0428
		RMSE	0.0347	0.1302	0.0465	0.1042	0.0635

Note: $\theta_0 = (0.2, 0.2, 1, 0.2, 1)'$ where $\gamma_0 + \rho_0 + \lambda_0 = 0.6$.

Table 4
Non-stable SDPD models: before bias correction.

T	n		γ	ρ	β	λ	σ^2
(1) Direct estimation							
20	54	Bias	−0.0302	0.0517	−0.0006	−0.0335	−0.0673
		E-SD	0.0205	0.0343	0.0306	0.0301	0.0404
		RMSE	0.0394	0.0662	0.0420	0.0507	0.0839
(2) Transformation by $F_{n,n-1}$							
20	49	Bias	−0.0332	0.0273	−0.0025	−0.0048	−0.0545
		E-SD	0.0206	0.0349	0.0308	0.0309	0.0481
		RMSE	0.0416	0.0528	0.0421	0.0422	0.0804
(3) W_nY_{nt} omitted; transformation by $F_{n,n-1}$							
20	49	Bias	0.0032	0.3695	0.0236	–	0.1022
		E-SD	0.0221	0.0242	0.0332	–	0.0479
		RMSE	0.0292	0.3703	0.0503	–	0.1171
(4) W_nY_{nt-1} omitted; transformation by $F_{n,n-1}$							
20	49	Bias	−0.1351	–	−0.0390	0.3234	0.1628
		E-SD	0.0137	–	0.0341	0.0134	0.2063
		RMSE	0.1359	–	0.0584	0.3237	0.2663
(5) $Y_{n,t-1}$ omitted; transformation by $F_{n,n-1}$							
20	49	Bias	–	−0.1652	−0.0602	0.3427	0.3631
		E-SD	–	0.0156	0.0368	0.0146	0.2209
		RMSE	–	0.1661	0.0745	0.3430	0.4284
(6) Both W_nY_{nt} and W_nY_{nt-1} omitted; transformation by $F_{n,n-1}$							
20	49	Bias	0.4668	–	0.0491	–	0.6864
		E-SD	0.0122	–	0.0410	–	0.0733
		RMSE	0.4670	–	0.0716	–	0.6903
(7) $Y_{n,t-1}$ and W_nY_{nt-1} omitted; direct							
20	49	Bias	–	–	−0.0686	0.4520	0.4019
		E-SD	–	–	0.0378	0.0108	0.8609
		RMSE	–	–	0.0822	0.4522	0.9718
(8) α_t omitted							
20	49	Bias	−0.0685	−0.2956	−0.0368	0.3472	0.1253
		E-SD	0.0223	0.0261	0.0336	0.0140	0.2147
		RMSE	0.0725	0.2968	0.0567	0.3475	0.2531
(9) Transformation by $I_n - W_n$							
20	49	Bias	−0.0374	0.0009	−0.0034	−0.0110	−0.0573
		E-SD	0.0229	0.0876	0.0331	0.0740	0.0423
		RMSE	0.0465	0.1190	0.0457	0.1058	0.0803

Note: $\theta_0 = (0.4, 0.2, 1, 0.4, 1)'$ where $\gamma_0 + \rho_0 + \lambda_0 = 1$.

with $\theta_0 = (0.4, 0.2, 1, 0.4, 1)'$. Most of the MC results are similar to the above stable SDPD case except for some model misspecifications. For the misspecifications of the general model as a time–space recursive model, a pure dynamic panel model, or a static spatial panel model, we have large biases for the estimates. This difference between Tables 2 and 3 and Tables 4 and 5 might be due to the nonstability of the DGP. In Tables 6 and 7, we run an intermediate case with $\theta_0 = (0.4, 0.2, 1, 0.3, 1)'$, which implies $\gamma_0 + \rho_0 + \lambda_0 = 0.9$, and we have intermediate magnitude of the bias for items (3), (6) and (7).

Because the unified transformation method will lose more degrees of freedom than the other methods, we expect less precision for the estimates from the unified transformation approach. It is of interest to see that the estimators by the unified transformation method perform well. They are slightly worse than the corresponding estimators in the loss of precision. All its estimates have small biases.

4.2. Empirical illustrations

In this section, we provide two empirical illustrations of the estimation of SDPD models. The first illustrates the importance of accounting for time effects in estimation. The second provides an empirical example for the possible spatial cointegration.

4.2.1. Dynamic demand for cigarettes

Baltagi and Levin (1986, 1992) investigate the dynamic demand for cigarette consumption by using the panel data of 46 states over the periods 1963–1980 and 1963–1988, respectively. The main findings of Baltagi and Levin (1986, 1992) are a significant price elasticity. For the income elasticity, it is insignificant in Baltagi and Levin (1986), and it is significant but small in Baltagi and Levin (1992). Also, the "bootlegging" effect is found to be significant so that the minimum price of neighboring states influences the cigarette consumption in a state. However, this bootlegging specification ignores the possibility that cross border shopping can take place in different neighboring states, but not just the minimum price of neighboring states. To partially overcome this problem, Elhorst (2005) specifies a spatial process in the disturbances so that the equation for estimation is

$$\ln C_{nt} = \gamma_0 \ln C_{n,t-1} + \beta_{01} \ln P_{nt} + \beta_{02} \ln D_{nt} + \beta_{03} \ln P_{mt} + \mu_n + \alpha_t l_n + U_{nt}, U_{nt} = \lambda_0 W_n U_{nt} + V_{nt},$$

where C_{nt} is the per capita consumption of cigarettes by persons of smoking age (14 years and older), P_{nt} is the real price of cigarettes, D_{nt} is the real disposable income per capita, P_{mt} is the minimum price of neighboring states, μ_n is the vector of individual effects and α_t is a time effect. Elhorst (2005) estimates the model with fixed effects μ_n eliminated by time differencing. Yang et al. (2006) also use the same data to illustrate the estimation of the dynamic panel with spatial errors in a random component setting.

Instead of the above models, the SDPD model can be considered that takes into account possible contemporaneous and time lagged regional spillovers (Case, 1991; Case et al., 1993). In order to be comparable with and nest Elhorst's spatial disturbance specification, we extend the SDPD model with the inclusion of W_nX_{nt} as extra regressors. The specification in Elhorst (2005) with spatial disturbances can be regarded as a special

Table 5
Non-stable SDPD models: after bias correction.

T	n		γ	ρ	β	λ	σ^2
(1) Direct estimation							
20	49	Bias	−0.0033	−0.0168	−0.0011	−0.0146	−0.0063
		E-SD	0.0212	0.0361	0.0316	0.0290	0.0431
		RMSE	0.0444	0.1713	0.0578	0.0791	0.0644
(2) Transformation by $F_{n,n-1}$							
20	49	Bias	0.0006	0.0118	0.0014	−0.0031	−0.0075
		E-SD	0.0211	0.0361	0.0316	0.0309	0.0472
		RMSE	0.0294	0.0492	0.0427	0.0423	0.0633
(3) W_nY_{nt} omitted; transformation by $F_{n,n-1}$							
20	49	Bias	0.0992	0.2730	0.0305	–	0.1522
		E-SD	0.0223	0.0244	0.0335	–	0.0487
		RMSE	0.1017	0.2742	0.0534	–	0.1610
(4) W_nY_{nt-1} omitted; transformation by $F_{n,n-1}$							
20	49	Bias	−0.1079	–	−0.0339	0.3208	0.2229
		E-SD	0.0146	–	0.0349	0.0133	0.2060
		RMSE	0.1092	–	0.0564	0.3211	0.3063
(5) $Y_{n,t-1}$ omitted; transformation by $F_{n,n-1}$							
20	49	Bias	–	−0.1468	−0.0589	0.3450	0.4290
		E-SD	–	0.0166	0.0377	0.0145	0.2205
		RMSE	–	0.1482	0.0741	0.3453	0.4854
(6) Both W_nY_{nt} and W_nY_{nt-1} omitted; transformation by $F_{n,n-1}$							
20	49	Bias	0.5313	–	0.0583	–	0.7364
		E-SD	0.0123	–	0.0413	–	0.0757
		RMSE	0.5315	–	0.0779	–	0.7403
(7) $Y_{n,t-1}$ and W_nY_{nt-1} omitted; transformation							
20	49	Bias	–	–	−0.0691	0.4671	0.4857
		E-SD	–	–	0.0533	0.0457	1.8126
		RMSE	–	–	0.0917	0.4695	1.9012
(8) α_t omitted							
20	49	Bias	−0.0305	−0.3137	−0.0324	0.3481	0.1808
		E-SD	0.0228	0.0271	0.0344	0.0140	0.2144
		RMSE	0.0417	0.3149	0.0551	0.3485	0.2845
(9) Transformation by $I_n - W_n$							
20	49	Bias	−0.0020	−0.0028	−0.0006	−0.0098	−0.0099
		E-SD	0.0235	0.0900	0.0338	0.0709	0.0412
		RMSE	0.0325	0.1230	0.0462	0.1034	0.0621

Note: $\theta_0 = (0.4, 0.2, 1, 0.4, 1)'$ where $\gamma_0 + \rho_0 + \lambda_0 = 1$.

Table 6
Stable SDPD models: before bias correction.

T	n		γ	ρ	β	λ	σ^2
(1) Direct estimation							
20	54	Bias	−0.0347	0.0187	−0.0022	−0.0392	−0.0696
		E-SD	0.0204	0.0401	0.0305	0.0343	0.0402
		RMSE	0.0425	0.0571	0.0419	0.0583	0.0854
(2) Transformation by $F_{n,n-1}$							
20	54	Bias	−0.0368	−0.0000	−0.0032	−0.0079	−0.0552
		E-SD	0.0206	0.0407	0.0308	0.0350	0.0440
		RMSE	0.0442	0.0546	0.0422	0.0479	0.0780
(3) W_nY_{nt} omitted; transformation by $F_{n,n-1}$							
20	54	Bias	−0.0188	0.1850	0.0071	–	−0.0105
		E-SD	0.0212	0.0352	0.0318	–	0.0439
		RMSE	0.0331	0.1890	0.0442	–	0.0631
(4) $W_nY_{n,t-1}$ omitted; transformation by $F_{n,n-1}$							
20	54	Bias	−0.1302	–	−0.0307	0.3732	0.1835
		E-SD	0.0162	–	0.0343	0.0160	0.1629
		RMSE	0.1313	–	0.0545	0.3736	0.2483
(5) $Y_{n,t-1}$ omitted; transformation by $F_{n,n-1}$							
20	54	Bias	–	−0.1934	−0.0525	0.3912	0.3902
		E-SD	–	0.0203	0.0371	0.0172	0.1734
		RMSE	–	0.1946	0.0695	0.3916	0.4291
(6) Both W_nY_{nt} and $W_nY_{n,t-1}$ omitted; transformation by $F_{n,n-1}$							
20	54	Bias	0.0833	–	0.0068	–	0.1321
		E-SD	0.0202	–	0.0336	–	0.0492
		RMSE	0.0874	–	0.0467	–	0.1449
(7) $Y_{n,t-1}$ and $W_nY_{n,t-1}$ omitted; direct							
20	54	Bias	–	–	−0.0496	0.1878	0.2808
		E-SD	–	–	0.0357	0.0272	0.0848
		RMSE	–	–	0.0664	0.1907	0.2938
(8) α_t omitted							
20	54	Bias	−0.0696	−0.3063	−0.0286	0.3982	0.1470
		E-SD	0.0222	0.0281	0.0339	0.0165	0.1663
		RMSE	0.0735	0.3076	0.0530	0.3986	0.2259
(9) Transformation by $I_n - W_n$							
20	49	Bias	−0.0373	−0.0009	−0.0036	−0.0120	−0.0575
		E-SD	0.0233	0.0896	0.0332	0.0752	0.0428
		RMSE	0.0467	0.1216	0.0459	0.1066	0.0809

Note: $\theta_0 = (0.4, 0.2, 1, 0.3, 1)'$ where $\gamma_0 + \rho_0 + \lambda_0 = 0.9$.

case of the SDPD model with nonlinear restrictions across coefficients. By premultiplying both sides with $(I_n - \lambda_0 W_n)$, the transformed equation is reduced to

$$\ln C_{nt} = \lambda_0 W_n \ln C_{nt} + \gamma_0 \ln C_{n,t-1} + \rho_0 W_n \ln C_{n,t-1} + X_{nt}\beta_0 + W_n X_{nt}\phi_0 + \mu_n^* + \alpha_t^* l_n + V_{nt},$$

with $\rho_0 = -\lambda_0\gamma_0$, $\phi_0 = -\lambda_0\beta_0$, and μ_n^*, α_t^* are transformed individual effects and time effects. Here, $X_{nt} = [\ln P_{nt}, \ln D_{nt}, \ln P_{nt}]$ and $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03})'$. Thus, the modified equation can be estimated as an SDPD model.

We first estimate the model by directly estimating the individual effects and time effects. In the SDPD model, this direct estimation will cause biases for estimates of order $O(\max(1/n, 1/T))$. By using the eigenvector matrix of J_n, we then estimate the model where time effects are eliminated and make bias correction to the estimates. Finally, we estimate the model with the robust transformation $I_n - W_n$. The results are summarized in Table 8, where the hypotheses of $\rho_0 = -\lambda_0\gamma_0$ and $\phi_0 = -\lambda_0\beta_0$ are also tested.

From Table 8, we can see that the price elasticity is significant, which is consistent with Baltagi and Levin (1986). However, the income elasticity is significant, and the bootlegging effect is insignificant which are different from Baltagi and Levin (1986). These differences might be explained by the inclusion of the spatial effects. As we can see from item (3) in Tables 2 and 3 for the Monte Carlo study, omitting the spatial effect will lead to bias for the estimate of ρ_0.

In Elhorst (2005), the price elasticity and income elasticity are significant, and the bootlegging effect is insignificant. These are the same as the SDPD estimation results. In fact, the magnitudes of his estimates are similar to the results in Table 8. For the Wald tests of constrained coefficients implied by the spatial correlated disturbances, they are rejected near the 5% critical value. Therefore, the spatial lag specification in the main equation seems more appropriate than the specification of spatial correlated disturbances. In Yang et al. (2006), the regressors and the regressant are different. They use nominal data, where the individual invariant consumer price index (CPI) is included as a regressor, and time effects are not specified. In Yang et al. (2006), all the effects of interest, namely the price effect, the income effect and bootlegging effect, are significant. A possible explanation for the difference of Elhorst's and the results here with those in Yang et al. (2006) could be the omission of the time effects in Yang et al. (2006). While the CPI is included as a regressor which captures some time effects, there might be other important time variables missing. With time effects omitted as a misspecification, the spatial effects might capture a part of them. This can be seen from item (8) in Tables 2 and 3 for the Monte Carlo study, where the omission of time effects will cause biases for estimates, in particular, those of λ_0 and ρ_0.

4.2.2. Market integration

Keller and Shiue (2007) use historical data of the price of rice in China to study the role of spatial features in the expansion of interregional trade and market integration. The data are available for $n = 121$ prefectures (from 10 provinces) and $T = 108$ periods, where we have 54 years in the mid-Qing (Qing Dynasty, 1644–1912), and the months of February and August are recorded (other months have the missing data problem as is pointed out by Keller and Shiue (2007);

Table 7
Stable SDPD models: after bias correction.

T	n		γ	ρ	β	λ	σ^2
(1) Direct estimation							
20	54	Bias	−0.0004	0.0050	0.0009	−0.0056	−0.0093
		E-SD	0.0221	0.0418	0.0315	0.0334	0.0428
		RMSE	0.0293	0.0575	0.0426	0.0466	0.0601
(2) Transformation by $F_{n,n-1}$							
20	54	Bias	−0.0008	0.0024	0.0006	−0.0051	−0.0081
		E-SD	0.0211	0.0424	0.0315	0.0350	0.0430
		RMSE	0.0293	0.0579	0.0427	0.0477	0.0600
(3) W_nY_{nt} omitted; transformation by $F_{n,n-1}$							
20	54	Bias	0.0597	0.1103	0.0130	–	0.0605
		E-SD	0.0213	0.0354	0.0320	–	0.0445
		RMSE	0.0637	0.1181	0.0453	–	0.0828
(4) W_nY_{nt-1} omitted; transformation by $F_{n,n-1}$							
20	54	Bias	−0.0953	–	−0.0266	0.3728	0.2428
		E-SD	0.0170	–	0.0352	0.0157	0.1625
		RMSE	0.0982	–	0.0534	0.3732	0.2943
(5) $Y_{n,t-1}$ omitted; transformation by $F_{n,n-1}$							
20	54	Bias	–	−0.1761	−0.0522	0.3930	0.4582
		E-SD	–	0.0212	0.0380	0.0172	0.1729
		RMSE	–	0.1778	0.0699	0.3934	0.4916
(6) Both W_nY_{nt} and W_nY_{nt-1} omitted; transformation by $F_{n,n-1}$							
20	54	Bias	0.1254	–	0.0110	–	0.1821
		E-SD	0.0203	–	0.0337	–	0.0494
		RMSE	0.1275	–	0.0474	–	0.1889
(7) $Y_{n,t-1}$ and W_nY_{nt-1} omitted; transformation							
20	54	Bias	–	–	−0.0497	0.2038	0.3705
		E-SD	–	–	0.0370	0.0305	0.1084
		RMSE	–	–	0.0671	0.2067	0.3867
(8) α_t omitted							
20	54	Bias	−0.0312	−0.3214	−0.0249	0.3987	0.2041
		E-SD	0.0228	0.0292	0.0347	0.0165	0.1659
		RMSE	0.0423	0.3228	0.0521	0.3990	0.2661
(9) Transformation by $I_n - W_n$							
20	49	Bias	−0.0020	−0.0024	−0.0007	−0.0109	−0.0102
		E-SD	0.0238	0.0920	0.0339	0.0721	0.0417
		RMSE	0.0329	0.1256	0.0464	0.1041	0.0627

Note: $\theta_0 = (0.4, 0.2, 1, 0.3, 1)'$ where $\gamma_0 + \rho_0 + \lambda = 0.9$.

for the information on the data collection, [15] see Shiue (2002)). Table 9 is the plot for the mid-price of the cross sectional average. It seems that there is a time trend which could be explained by the spatial cointegrated DGP, explosive DGP, or some time factors.

From the estimates in Keller and Shiue (2007), the spatial features are important as the geographical distances influence the trade and possible arbitrage. The spatial effect, dynamic effect and spatial time effect are found to be significant. However, even the data are in the form of a panel, their estimation is based on annual cross section SAR models with (or without) the lagged price variables $Y_{n,t-1}$ and $W_nY_{n,t-1}$ as explanatory variables. Their reported estimates are the average from 53 (54) years. With panel data, it may be more desirable to formulate the SDPD model and estimate it with techniques as in Section 3. A panel model can control more explicitly both regional fixed effects and unobserved time effects. Therefore, the SDPD model with time effects and individual effects is specified for the price equation. Compared to Keller and Shiue (2007), the weather indicators are not included as exogenous variables due to the data availability. However, as those weather regressors are insignificant in Keller and Shiue (2007), the omission would not be controversial. Hence, the estimated equation is

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt},$$

$$t = 1, 2, ..., T,$$

where Y_{nt} is the selling price of mid-quality rice. Keller and Shiue (2007) argue that different weights matrices could be used. Denote d_{ij}s as the distances among the capitals of prefectures ranging from 10 to 1730 km. Examples of the spatial weights matrices would be (1) $W_n^{(1)}$, where prefectures are neighbors if the $d_{ij} \le 300$; (2) $W_n^{(2)}$, where prefectures are neighbors if the $d_{ij} \le 600$; (3) $W_n^{(3)}$, where $w_{ij}^{(3)} = 1$ if $d_{ij} \le 300$, $w_{ij}^{(3)} = 0.5$ if $300 < d_{ij} \le 600$ and $w_{ij}^{(3)} = 0$ if $d_{ij} > 600$; and (4) $W_n^{(4)}$, where $w_{ij} = \exp\{\theta_d D_{ij}\}$ with $D_{ij} = \frac{d_{ij}}{100}$ and a larger absolute value of a negative θ_d denotes a more rapid decline in the size of the weights when d_{ij} increases. All these weights matrices are row-normalized as in practice. Keller and Shiue (2007) state that the specification (4) with $\theta_d = -1.4$ fits the data well. By the criterion of log likelihood value, we find that $\theta_d = -1.2$ can be better than -1.4. We use different specifications of the SDPD model and estimate them with different methods.

Model I: use the SDPD model without time effects in Yu et al. (2008).

Model II (a): use the SDPD model with time effects, and use the direct estimation in Lee and Yu (2007).

Model II (b): use the SDPD model with time effects, and use the transformation in Lee and Yu (2007).

Model II (c): use the SDPD model with time effects, and use the robust transformation in Lee and Yu (2009).

The results are in Tables 10 and 11 where we use $W_n^{(4)}$ with $w_{ij} = \exp\{-1.2D_{ij}\}$. Table 10 uses the August data which is the same as Keller and Shiue (2007) with $T = 54$. We can see that all the effects are significant under different estimation methods. The estimates of λ_0 are about 0.8 or slightly larger; those of γ_0 are about 0.5; those for ρ_0 are around -0.4. For the test of $\gamma_0 + \rho_0 + \lambda_0 = 1$, it is rejected under Model I and Model II (a) but not rejected under Model II (c). It is rejected at 5% significance level but not at 1% significance level under Model II (b). For the log likelihood, we can see that the transformation methods II (b) and II (c) yield higher values. This indicates that Model II (b) and Model II (c) might be better fitted; hence, there may be spatial cointegration in the DGP. Table 11 uses the February and August data together so that $T = 108$. We can see that the results are similar to Table 10.

Table 12 presents the results using the February and August data with different values of θ_d in $w_{ij} = \exp\{\theta_d D_{ij}\}$, specifically $\theta_d = -0.7$, -1.4 and -2.8 where -1.4 is used in Keller and Shiue (2007). We see when $\theta_d = -0.7$ so that distant neighbors still receive non-neglectible weights, $\gamma_0 + \rho_0 + \lambda_0$ could be larger than 1, which implies an explosive DGP. For the case $\theta_d = -1.4$ and -2.8, $\gamma_0 + \rho_0 + \lambda_0$ is close to but smaller than 1. The tests of $\gamma_0 + \rho_0 + \lambda_0 = 1$ are all rejected for above weights matrix specifications. We also present the results with $W_n^{(1)}$, $W_n^{(2)}$ and $W_n^{(3)}$ in Table 13. All the effects are significant under these three specifications. Under $W_n^{(1)}$ so that only prefectures within 300 km are considered as neighbors, $\gamma_0 + \rho_0 + \lambda_0$ is close to 1 and the spatial cointegration is not rejected. However, under $W_n^{(2)}$ and $W_n^{(3)}$, $\gamma_0 + \rho_0 + \lambda_0$ is greater than 1; the spatial cointegration is rejected under $W_n^{(3)}$ but not rejected under $W_n^{(2)}$.

Therefore, all the spatial and dynamic effects are significant under different weights matrix specifications and estimation methods. The sum of the estimates of λ_0, γ_0, and ρ_0 is close to 1 even though their sum of being 1 is statistically rejected under some specifications. We may conclude that the markets are overall integrated or nearly integrated.

5. Conclusion

This paper has presented some recent developments in the specification and estimation of spatial panel data models. For the static case, we can use the direct or transformation approaches under the fixed effects specification, while we have various frameworks of the error components under the random effects specification. For the dynamic

[15] We have the minimum price and the maximum price for each prefecture, where the prices are collected from counties of each prefecture. Similar to Keller and Shiue (2007), the (log) mid-price is constructed and used for the estimation.

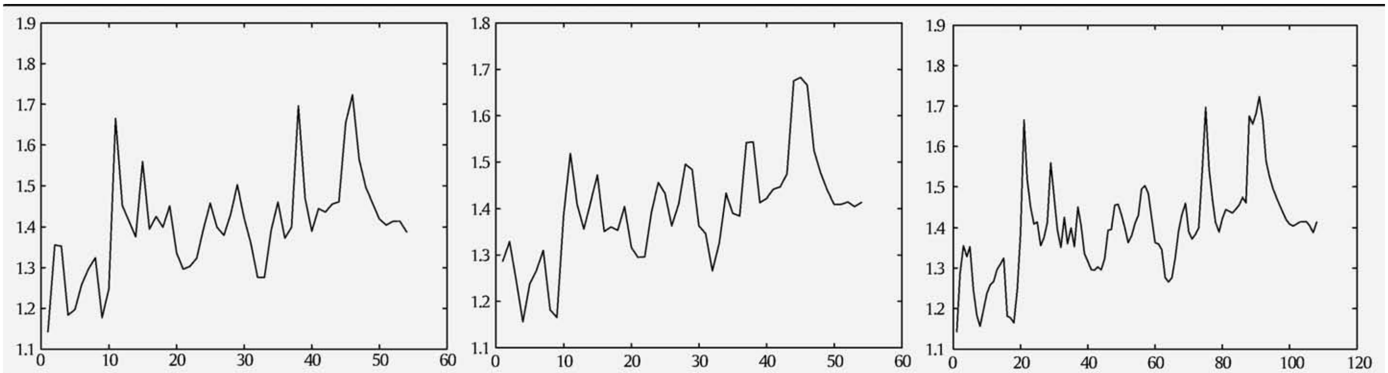
Please cite this article as: Lee, L., Yu, J., Some recent developments in spatial panel data models, Regional Science and Urban Economics (2009), doi:10.1016/j.regsciurbeco.2009.09.002

Table 8
Estimation results for the cigarettes demand.

	Direct		J_n		I_n-W_n	
Estimates and t-statistics						
γ ($\ln C_{n,t-1}$)	0.8651	[67.2425]	0.8643	[67.1020]	0.8577	[62.3541]
ρ ($W_n\ln C_{n,t-1}$)	−0.0145	[−0.3364]	−0.0177	[−0.4455]	−0.0258	[−0.4646]
β_1 ($\ln P_{nt}$)	−0.2619	[−10.6646]	−0.2621	[−10.6649]	−0.2619	[−10.2456]
β_2 ($\ln Y_{nt}$)	0.0997	[3.3481]	0.0994	[3.3359]	0.1026	[3.4068]
β_3($\ln P_{mt}$)	0.0073	[0.2000]	0.0074	[0.2031]	−0.0142	[−0.3592]
ϕ_1 ($W_n\ln P_{nt}$)	0.1671	[3.1780]	0.1700	[3.2364]	0.1772	[2.8543]
ϕ_2 ($W_n\ln Y_{nt}$)	−0.0256	[−0.6443]	−0.0228	[−0.5764]	−0.0231	[−0.4252]
ϕ_3 ($W_n\ln P_{mt}$)	−0.0220	[−0.4362]	−0.0240	[−0.4782]	−0.0705	[−1.0845]
λ ($W_n\ln C_{nt}$)	−0.0757	[2.0611]	0.0784	[2.0668]	0.0472	[0.8726]
Tests						
$\rho=-\lambda\gamma$ ($\chi^2_{1,0.05}=3.8$)	5.8042		5.6227		0.2634	
$\phi=-\lambda\beta$ ($\chi^2_{3,0.05}=7.8$)	8.9087		9.2183		8.8058	
Joint above ($\chi_{4,0.05}=9.4$)	10.5028		10.6685		8.8228	

Note: The numbers in the [·] are the t-statistics.

Table 9
Average of 121 mid-prices of February, August and combined.



Note: 1. From the first column to the third column are February, August and combined prices.

Table 10
SDPD models, August prices, $w_{ij}=\exp\{-1.2D_{ij}\}$ with row-normalization.

Models	I	II (a)	II (b)	II (c)
Before bias correction estimates				
$Y_{n,t-1}$	0.5279	0.5276	0.5272	0.5266
	(0.0108)	(0.0109)	(0.0068)	(0.0109)
$W_nY_{n,t-1}$	−0.4112	−0.3708	−0.3958	−0.3943
	(0.0154)	(0.0185)	(0.0107)	(0.0407)
W_nY_{nt}	0.8520	0.7960	0.8359	0.8640
	(0.0090)	(0.0111)	(0.0085)	(0.0556)
σ^2	0.0044	0.0044	0.0044	0.0044
	(0.0003)	(0.0001)	(0.0001)	(0.0003)
Tests (Wald χ^2 statistics)				
$\rho=-\lambda\gamma$	19.9968	15.1323	12.8392	8.9456
$\rho+\gamma+\lambda=1$	13.5655	13.7998	6.5918	0.7384
Value of $\rho+\gamma+\lambda$	0.9686	0.9528	0.9673	0.9783
$\ln L$	10,164	10,142	10,199	10,198
After bias correction estimates				
$Y_{n,t-1}$	0.5568	0.5563	0.5560	0.5555
	(0.0109)	(0.0110)	(0.0110)	(0.0110)
$W_nY_{n,t-1}$	−0.4354	−0.4132	−0.4193	−0.4179
	(0.0156)	(0.0185)	(0.0193)	(0.0416)
W_nY_{nt}	0.8520	0.8273	0.8361	0.8461
	(0.0090)	(0.0099)	(0.0126)	(0.0545)
σ^2	0.0045	0.0045	0.0045	0.0045
	(0.0003)	(0.0001)	(0.0002)	(0.0003)
Tests (Wald χ^2 statistics)				
$\rho=-\lambda\gamma$	20.1311	13.8543	13.0395	9.1086
$\rho+\gamma+\lambda=1$	9.7487	5.4837	4.5497	0.4716
Value of $\rho+\gamma+\lambda$	0.9734	0.9705	0.9728	0.9837

Note: The numbers in the (·) are the standard deviations.

Table 11
SDPD models, February and August Prices, $w_{ij}=\exp\{-1.2D_{ij}\}$ with row-normalization.

Models	I	II (a)	II (b)	II (c)
Before Bias Correction Estimates				
$Y_{n,t-1}$	0.6646	0.6637	0.6634	0.6629
	(0.0067)	(0.0068)	(0.0068)	(0.0068)
$W_nY_{n,t-1}$	−0.5138	−0.4651	−0.4998	−0.5006
	(0.0105)	(0.0125)	(0.0130)	(0.0316)
W_nY_{nt}	0.8240	0.7730	0.8180	0.8270
	(0.0072)	(0.0084)	(0.0095)	(0.0382)
σ^2	0.0036	0.0036	0.0035	0.0035
	(0.0002)	(0.0000)	(0.0001)	(0.0002)
Tests (Wald χ^2 statistics)				
$\rho=-\lambda\gamma$	39.6135	36.2013	29.5725	20.0357
$\rho+\gamma+\lambda=1$	23.3447	13.1849	5.5853	0.6525
Value of $\rho+\gamma+\lambda$	0.9748	0.9716	0.9816	0.9893
$\ln L$	21,985	21,944	22,032	22,005
After bias correction estimates				
$Y_{n,t-1}$	0.6804	0.6793	0.6790	0.6786
	(0.0068)	(0.0068)	(0.0068)	(0.0068)
$W_nY_{n,t-1}$	−0.5273	−0.5018	−0.5121	−0.5133
	(0.0106)	(0.0123)	(0.0131)	(0.0319)
W_nY_{nt}	0.8240	0.8055	0.8181	0.8271
	(0.0072)	(0.0076)	(0.0094)	(0.0378)
σ^2	0.0036	0.0036	0.0036	0.0036
	(0.0002)	(0.0000)	(0.0001)	(0.0002)
Tests (Wald χ^2 statistics)				
$\rho=-\lambda\gamma$	38.2825	32.7226	30.0950	20.1723
$\rho+\gamma+\lambda=1$	19.2219	4.7582	3.6844	0.3529
Value of $\rho+\gamma+\lambda$	0.9771	0.9830	0.9850	0.9924

Note: The numbers in the (·) are the standard deviations.

Table 12
SDPD models, February and August Prices, $w_{ij} = \exp\{\theta_d D_{ij}\}$ with row-normalization.

	$\theta_d = -0.7$		$\theta_d = -1.4$		$\theta_d = -2.8$	
Models	II (b)	II (c)	II (b)	II (c)	II (b)	II (c)
Before bias correction estimates						
$Y_{n,t-1}$	0.6707 (0.0067)	0.6669 (0.0068)	0.6626 (0.0068)	0.6626 (0.0068)	0.6658 (0.0068)	0.6665 (0.0068)
$W_nY_{n,t-1}$	−0.6046 (0.0151)	−0.5859 (0.0474)	−0.4705 (0.0125)	−0.4602 (0.0300)	−0.3299 (0.0107)	−0.2523 (0.0204)
W_nY_{nt}	0.9730 (0.0081)	1.0000 (0.0552)	0.7750 (0.0094)	0.7640 (0.0369)	0.5700 (0.0085)	0.4450 (0.0244)
σ^2	0.0038 (0.0001)	0.0038 (0.0002)	0.0035 (0.0001)	0.0035 (0.0002)	0.0037 (0.0001)	0.0035 (0.0001)
Tests (Wald χ^2 statistics)						
$\rho = -\lambda\gamma$	22.6602	24.8034	32.9797	21.9052	56.524	18.7204
$\rho + \gamma + \lambda = 1$	15.7321	19.1677	19.5539	7.0490	204.27	128.8964
Value of $\rho + \gamma + \lambda$	1.0391	1.0810	0.9671	0.9664	0.9058	0.8592
ln$L$	21,823	21,854	22,021	21,994	21,640	21,407
After bias correction estimates						
$Y_{n,t-1}$	0.6602 (0.0067)	0.6828 (0.0068)	0.6783 (0.0068)	0.6783 (0.0068)	0.6816 (0.0069)	0.6822 (0.0068)
$W_nY_{n,t-1}$	−0.4113 (0.0186)	−0.6045 (0.0476)	−0.4825 (0.0126)	−0.4722 (0.0303)	−0.3393 (0.0108)	−0.2591 (0.0207)
W_nY_{nt}	0.9822 (0.0064)	1.0008 (0.0546)	0.7750 (0.0094)	0.7640 (0.0365)	0.5697 (0.0085)	0.4449 (0.0241)
σ^2	0.0038 (0.0001)	0.0038 (0.0002)	0.0036 (0.0001)	0.0036 (0.0002)	0.0037 (0.0001)	0.0036 (0.0001)
Tests (Wald χ^2 statistics)						
$\rho = -\lambda\gamma$	332.8166	23.6530	32.8950	21.8267	54.713	18.6606
$\rho + \gamma + \lambda = 1$	390.7173	19.1778	15.4715	5.9502	178.91	116.9106
Value of $\rho + \gamma + \lambda$	1.2311	1.0792	0.9707	0.9701	0.9120	0.8681

Note: The numbers in the $(\cdot)$ are the standard deviations.

Table 13
SDPD models, February and August Prices, $W_n = W_n^{(i)}$, $i = 1, 2, 3$ with row-normalization.

	$W_n^{(1)}$		$W_n^{(2)}$		$W_n^{(3)}$	
Models	II (b)	II (c)	II (b)	II (c)	II (b)	II (c)
Before bias correction estimates						
$Y_{n,t-1}$	0.6658 (0.0067)	0.6651 (0.0067)	0.7062 (0.0063)	0.7062 (0.0063)	0.6921 (0.0065)	0.6921 (0.0065)
$W_nY_{n,t-1}$	−0.4958 (0.0138)	−0.4859 (0.0332)	−0.6542 (0.0211)	−0.6543 (0.0433)	−0.6415 (0.0186)	−0.6415 (0.0393)
W_nY_{nt}	0.8140 (0.0103)	0.8160 (0.0394)	0.9700 (0.0162)	0.9700 (0.0494)	0.9950 (0.0112)	0.9950 (0.0394)
σ^2	0.0038 (0.0001)	0.0038 (0.0002)	0.0047 (0.0001)	0.0047 (0.0002)	0.0044 (0.0001)	0.0044 (0.0002)
Tests (Wald χ^2 statistics)						
$\rho = -\lambda\gamma$	29.3264	23.3260	4.3833	3.9623	12.7703	9.2133
$\rho + \gamma + \lambda = 1$	3.5519	0.1217	2.1682	1.6010	12.3065	11.5299
Value of $\rho + \gamma + \lambda$	0.9840	0.9952	1.0219	1.0220	1.0456	1.0456
ln$L$	21,641	21,616	20,589	20,589	21,039	21,039
After bias correction estimates						
$Y_{n,t-1}$	0.6815 (0.0067)	0.6808 (0.0067)	0.7223 (0.0063)	0.7223 (0.0063)	0.7061 (0.0065)	0.7061 (0.0065)
$W_nY_{n,t-1}$	−0.5078 (0.0139)	−0.4986 (0.0335)	−0.6695 (0.0213)	−0.6695 (0.0437)	−0.6167 (0.0199)	−0.6167 (0.0454)
W_nY_{nt}	0.8142 (0.0103)	0.8161 (0.0390)	0.9702 (0.0162)	0.9702 (0.0489)	0.9976 (0.0109)	0.9976 (0.0376)
σ^2	0.0039 (0.0001)	0.0039 (0.0002)	0.0048 (0.0001)	0.0048 (0.0002)	0.0044 (0.0001)	0.0044 (0.0002)
Tests (Wald χ^2 statistics)						
$\rho = -\lambda\gamma$	30.0780	23.3616	4.5045	4.0701	40.5591	18.8330
$\rho + \gamma + \lambda = 1$	2.0656	0.0153	2.3925	1.8648	44.2386	40.0265
Value of $\rho + \gamma + \lambda$	0.9878	0.9983	1.0231	1.0231	1.0870	1.0870

Note: The numbers in the $(\cdot)$ are the standard deviations.

case, we review the estimation and asymptotic properties of various SDPD models depending on the eigenvalue structure, as well as the dynamic panel data model with spatial disturbances. We provide some Monte Carlo studies on misspecifications when restricted models are estimated. We find that the omission of time effects can have important consequences in the estimation of spatial effects. This issue is illustrated with an empirical application. We also illustrate the possibility of spatial cointegration due to market integration.

Many extensions of the SDPD model and related estimation issues are of interest for future research. Models of simultaneous equations with spatial and dynamic structures are important ones for future consideration, and so are SDPD models with common shocks and factors for cross-sectional dependence. Common factor models with spatial disturbances have already received attention in the work of Pesaran and Tosetti (2007).

Acknowledgements

We would like to thank an anonymous referee, Professor Harry Kelejian, and the editor, Professor Daniel McMillen, for their helpful comments.

References

Alvarez, J., Arellano, M., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. Econometrica 71, 1121–1159.
Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error components. Journal of the American Statistical Association 76, 598–606.
Anselin, L., 1988. Spatial econometrics: methods and models. Kluwer Academic, The Netherlands.
Anselin, L., 1992. Space and applied econometrics. In: Anselin (Ed.), Special Issue. Regional Science and Urban Economics, vol. 22.

Anselin, L., 2001. Spatial econometrics. In: Baltagi, Badi H. (Ed.), A companion to theoretical econometrics. Blackwell Publishers Lte, Massachusetts.

Anselin, L., Bera, A.K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A., Giles, D.E.A. (Eds.), Handbook of Applied Economics Statistics. Marcel Dekker, New York.

Anselin, L., Florax, R., 1995. New directions in spatial econometrics. Springer-Verlag, Berlin.

Anselin, L., Rey, S., 1997. Spatial econometrics. In: Anselin, L., Rey, S. (Eds.), Special Issue International Regional Science Review, vol. 20.

Anselin, L., Le Gallo, J, Jayet, J., 2008. Spatial panel econometrics. The econometrics of panel data: fundamentals and recent developments in theory and practice. Springer, Berlin Heidelberg.

Arellano, M., Bond, O., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277–297.

Baltagi, B., Levin, D., 1986. Estimating dynamic demand for cigarettes using panel data: the effects of bootlegging, taxation and advertising reconsidered. Review of Economics and Statistics 48, 148–155.

Baltagi, B., Levin, D., 1992. Cigarette taxation: raising revenues and reducing consumptions. Structural Change and Economic Dynamics 3, 321–335.

Baltagi, B., Li, D., 2006. Prediction in the panel data model with spatial correlation: the case of liquor. Spatial Economic Analysis 1, 175–185.

Baltagi, B., Song, S.H., Koh, W., 2003. Testing panel data regression models with spatial error correlation. Journal of Econometrics 117, 123–150.

Baltagi, B., Egger, P., Pfaffermayr, M., 2007a. A generalized spatial panel data model with random effects. Syracuse University, Working Paper.

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

Baltagi, B., Bresson, G., Pirotte, A., 2007c. Panel unit root tests and spatial dependence. Journal of Applied Econometrics 22, 339–360.

Bhargava, A., Sargan, J.D., 1983. Estimating dynamic random effects models from panel data covering short time periods. Econometrica 51, 1635–1659.

Bun, M., Carree, M., 2005. Bias-corrected estimation in dynamic panel data models. Journal of Business & Economic Statistics 3 (2), 200–210 (11).

Case, A., 1991. Spatial patterns in household demand. Econometrica 59, 953–965.

Case, A., Hines, J.R., Rosen, H.S., 1993. Budget spillovers and fiscal policy interdependence: evidence from the states. Journal of Public Economics 52, 285–307.

Cliff, A.D., Ord, J.K., 1973. Spatial autocorrelation. Pion Ltd., London.

Cox, D.R., 1975. Partial likelihood. Biometrika 62, 269–276.

Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference. Journal of the Royal Statistical Society. Series B (Methodological) 49, 1–39.

Cressie, N., 1993. Statistics for spatial data. Wiley, New York.

Druska, V., Horrace, W.C., 2004. Generalized moments estimation for spatial panel data: Indonesian rice farming. American Journal of Agricultural Economics 86, 185–198.

Egger, P., Pfaffermayr, M., Winner, H., 2005. An unbalanced spatial panel data approach to US state tax competition. Economics Letters 88, 329–335.

Ertur, C., Koch, W., 2007. Growth, technological interdependence and spatial externalities: theory and evidence. Journal of Applied Econometrics 22, 1033–1062.

Elhorst, J.P., 2003. Specification and estimation of spatial panel data models. International Regional Science Review 26, 244–268.

Elhorst, J.P., 2005. Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. Geographical Analysis 37, 85–106.

Foote, C.L., 2007. Space and time in macroeconomic panel data: young workers and state-level unemployment revisited. Working Paper No. 07-10, Federal Reserve Bank of Boston.

Franzese, R.J., 2007. Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. Political Analysis 15, 140–164.

Frazier, C., Kockelman, K.M., 2005. Spatial econometric models for panel data: incorporating spatial and temporal data. Transportation Research Record: Journal of the Transportation Research Board 1902 (2005), 80–90.

Hahn, J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. Econometrica 70, 1639–1657.

Hahn, J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. Econometrica 72, 1295–1319.

Hsiao, C., 1986. Analysis of Panel Data. Cambridge University Press.

Kalbfleisch, J.D., Sprott, D.A., 1970. Application of likelihood methods to models involving large numbers of parameters. Journal of the Royal Statistical Society. Series B (Methodological) 32, 175–208.

Kapoor, M., Kelejian, H.H., Prucha, I.R., 2007. Panel data models with spatially correlated error components. Journal of Econometrics 140, 97–130.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. Journal of Real Estate Finance and Economics 17 (1), 99–121.

Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219–257.

Kelejian, H.H., Prucha, I.R., 2007. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, Forthcoming in Journal of Econometrics.

Kelejian, H.H., Robinson, D., 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. Papers in Regional Science 72, 297–312.

Keller, W., Shiue, C.H., 2007. The origin of spatial interaction. Journal of Econometrics 140, 304–332.

Kiviet, J., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of Econometrics 68, 81–126.

Korniotis, G.M., forthcoming. Estimating panel models with internal and external habit formation. Journal of Business and Economic Statistics.

Lancaster, T., 2000. The incidental parameter problem since 1948. Journal of Econometrics 95, 391–413.

Lee, L.F., 2003. Best spatial two-stage least squares estimator for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22, 307–335.

Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. Econometrica 72, 1899–1925.

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

Lee, L.F., Yu, J., 2007. A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory (forthcoming). doi:10.1017/S0266466609100099.

Lee, L.F., Yu, J., 2008. Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics (forthcoming). doi:10.1016/j.jeconom.2009.08.001.

Lee, L.F., Yu, J., 2009. A unified transformation approach for the estimation of spatial dynamic panel data models: stability, spatial cointegration and explosive roots. Manuscript. Ohio State University. http://gatton.uky.edu/faculty/yu/Research/Unified-Transformation-0211-09.pdf.

Maddala, G.S., 1971. The use of variance components models in pooling cross section and time series data. Econometrica 39, 341–358.

Magnus, J.R., 1982. Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. Journal of Econometrics 19, 239–285.

Mutl, J., 2006. Dynamic panel data models with spatially correlated disturbances. PhD thesis, University of Maryland, College Park.

Mutl, J., Pfaffermayr, M., 2008. The spatial random effects and the spatial fixed effects model: the Hausman test in a Cliff and Ord panel model. Manuscript, Institute for Advanced Studies, Vienna.

Neyman, J., Scott, E., 1948. Consistent estimates based on partially consistent observations. Econometrica 16, 1–32.

Nickell, S.J., 1981. Biases in dynamic models with fixed effects. Econometrica 49, 1417–1426.

Ord, J.K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120–297.

Pesaran, M.H., Tosetti, E., 2007. Large panels with common factors and spatial correlations. Working paper. Cambridge University.

Revelli, F., 2001. Spatial patterns in local taxation: tax mimicking or error mimicking? Applied Economics 33, 1101–1107.

Shiue, C.H., 2002. Transport costs and the geography of arbitrage in eighteen-century China. American Economic Review 92, 1406–1419.

Su, L., Yang, Z., 2007. QML estimation of dynamic panel data models with spatial errors. Singapore Management University, Manuscript.

Tao, J., 2005. Spatial econometrics: models, methods and applications. PhD thesis, Ohio State University.

Yang, Z., Li, C., Tse, Y.K., 2006. Functional form and spatial dependence in dynamic panels. Economic Letters 91, 138–145.

Yu, J., Lee, L.F., 2007. Estimation of unit root spatial dynamic panel data models. Econometric Theory, forthcoming.

Yu, J., de Jong, R., Lee, L.F., 2007. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large: a nonstationary case. Manuscript. Ohio State University. http://www.econ.ohio-state.edu/lee/wp/NonStationary-Spatial-Panel-0825.pdf.

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.

Please cite this article as: Lee, L., Yu, J., Some recent developments in spatial panel data models, Regional Science and Urban Economics (2009), doi:10.1016/j.regsciurbeco.2009.09.002