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

A R T I C L E   I N F O

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

A B S T R A C T

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.

© 2009 Elsevier B.V. All rights reserved.

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) also considered a bias adjusted within estimator, which generalizes Hahn and Kuersteiner (2002). For a

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

© 2009 Elsevier B.V. All rights reserved.

0166-0462/$ – see front matter © 2009 Elsevier B.V. All rights reserved.

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
dynamic panel data model with spatial error, Elhorst (2005) estimates the model with unconditional maximum likelihood method, and Murti (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 $Y_{nt} = X_{nt}' \alpha + \varepsilon_{nt} + U_{nt}$ and $U_{nt} = \lambda W_{nt} U_{nt} + V_{nt}$, where $\lambda = \lambda_0 + \lambda_1$, $\alpha$, $\varepsilon_{nt}$, and $V_{nt}$ are $n \times 1$ vectors, $W_{nt}$ is I.D. across $i$ and $t$ with zero mean and finite variance $\sigma^2$, and $U_{nt}$ 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, $\varepsilon_{nt}$ 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 $Y_{nt} = X_{nt}' \alpha + \varepsilon_{nt}$ and $U_{nt} = \lambda W_{nt} U_{nt} + d_{nt} + V_{nt}$, where $d_{nt}$ 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.

$^2$ 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

$$Y_{nt} = \lambda W_{nt} Y_{nt} + X_{nt}' \beta + \varepsilon_{nt} + U_{nt},$$

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

For the estimation, we may consider the fixed effects specification (where elements of $\mu_{nt}$ are treated as fixed parameters) or the random effects specification (where $\mu_{nt}$ is a random component). The random effects specification of $\mu_{nt}$ in Eq. (1) can be assumed to be a SAR process. If the process of $\mu_{nt}$ 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 $\mu_{nt}$ across spatial units. On the other hand, the fixed effects specification is known to be robust against the possible correlation of $\mu_{nt}$ with included regressors in the model. The fixed effects specification can also be robust against the spatial specification of $\mu_{nt}$. For example, the spatial panel model introduced in Kapoor et al. (2007) is equivalent to Eq. (1) with $W_{nt} = W_{nt}$ and $\lambda = \lambda_0$, 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 $(U_0 - \lambda_0 W_{nt})$, the data generating process (DGP) of Kapoor et al. (2007) becomes $Y_{nt} = X_{nt}' \beta + \varepsilon_{nt}$, where $\varepsilon_{nt} = (U_0 - \lambda_0 W_{nt})^{-1}d_{nt}$ can be regarded as a vector of unknown fixed effect parameters and $U_0 = \lambda_0 W_{nt} + V_{nt}$ 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 $\mu_{nt}$ 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 $\mu_{nt}$ is treated as fixed effects, any time invariant regressors would be absorbed in $\mu_{nt}$.

$^4$ $U_0 = U_0 - (U_0 - \lambda_0 W_{nt})^{-1}d_{nt}$. 

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, $\theta_0 = (\beta_{01}, \lambda_0, \lambda_2, \sigma_{0}^2)'$ and $\zeta_0 = (\beta_{01}, \lambda_0, \lambda_2)'$. Define $S_n(\lambda_1) = I_{n} - \lambda_1 W_{n1}$ and $R(\lambda_2) = I_{n} - \lambda_2 W_{n2}$ for any $\lambda_1$ and $\lambda_2$. At the true parameter, $S_0 = S_0(\lambda_0)$ and $R_0 = R_0(\lambda_0)$. The log likelihood function of Eq. (1), as if the disturbances were normally distributed, is

$$
\ln L_{n,t}^d(\theta, c_n) = -\frac{nT}{2} \ln (2\pi\sigma^2) + T \ln |S_n(\lambda_1)| + \ln |R_n(\lambda_2)|
$$

(2)

where $\nu_n(\zeta, c_n) = R_0(\lambda_2)[S_0(\lambda_1)\nu_{nt} - \chi_{nt}^T c_n]$. If the disturbances in $\nu_{nt}$ are normally distributed, the log likelihood (2) is the exact one. When $\nu_{nt}$ is not normally distributed, but its elements are i.i.d. $N(0, \sigma_0^2)$, Eq. (2) is a quasi-likelihood function. We can estimate $c_{0T}$ directly from Eq. (2) and have the concentrated log likelihood function of $\theta$. For notational purposes, we define $\tilde{Y}_{nt} = Y_{nt} - \tilde{V}_{nt}$ for $t = 1, 2, \cdots, T$ where $\tilde{Y}_{nt} = \frac{1}{T} \sum_{t=1}^{T} Y_{nt}$. Similarly, we define $\tilde{X}_{nt} = X_{nt} - \tilde{X}_{nt}$ and $\tilde{V}_{nt} = V_{nt} - \tilde{V}_{nt}$. Thus, the log likelihood function with $c_0$ concentrated out is

$$
\ln L_{n,t}^d(\theta) = -\frac{nT}{2} \ln (2\pi\sigma^2) + T \ln |S_n(\lambda_1)| + \ln |R_n(\lambda_2)|
$$

(3)

where $\tilde{V}_{nt}(\zeta) = R_0(\lambda_2)[S_0(\lambda_1)\tilde{V}_{nt} - \chi_{nt}^T].$ This direct estimation approach will yield consistent estimates for the spatial and regression coefficients except for the variance parameter $\sigma_0^2$ when $T$ is large (but $n$ is large). Also, the estimator of $\sigma_0^2$ 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 $\sigma_0^2$, 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 statistic (see Hsiao, 1986). For the spatial panel data models, we can use a data transformation, the deviation from the time mean operator (i.e., $I_{t}$ = $t - \frac{1}{T} \sum_{t=1}^{T} t$, where $t$ 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 $I_{t}$ eliminates the time invariant individual effects, and the transformed model consists of $Y_{nt} = \lambda_0 W_{n1} Y_{nt} + \tilde{X}_{nt} c_n + \tilde{U}_{nt}$ and $\tilde{U}_{nt} = \lambda_0 W_{n2} \tilde{U}_{nt} + \tilde{V}_{nt}$ where $\tilde{V}_{nt} = \tilde{V}_{nt} - \tilde{V}_{nt}$. However, the disturbing disturbances $\tilde{V}_{nt}$ would be linearly dependent over the time dimension because $I_{t}$ 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 $I_{t}$. Let $[F_{1}, F_{2}, \cdots, F_{T}]$ be the orthonormal matrix of the eigenvectors of $I_{t}$ where $F_{T+1} = I_{T+1}$ is the $T \times (T-1)$ eigenvector matrix corresponding to the eigenvalues of 1. For any $n \times T$ matrix $[Z_{n1}, \cdots, Z_{nT-1}]$, define the transformed $n \times (T-1)$ matrix $[Z_{n1}', \cdots, Z_{nT-1}'] = [Z_{n1}, \cdots, Z_{nT-1}]F_{T+1}$ and define $X_n' = [X_{n1}, X_{n2}, \cdots, X_{nT}]$ accordingly. Then, Eq. (1) implies

$$
V_{nt}^* = \lambda_0 W_{n1} Y_{nt}' + X_{nt} c_n + U_{nt}, U_{nt} = \lambda_0 W_{n2} U_{nt} + V_{nt}^*. t = 1, \cdots, T.
$$

(4)

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

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

$$
\ln L_{n,t}^d(\theta) = -\frac{n(T-1)}{2} \ln (2\pi\sigma^2) + (T-1) \ln |S_n(\lambda_1)| + \ln |R_n(\lambda_2)|
$$

(5)

Lee and Yu (2008) show that the transformation approach will yield consistent estimators for all the common parameters including $\sigma_0^2$, 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 $\beta$ out, Eq. (3) becomes

$$
\ln L_{n,t}^d(\lambda_1, \lambda_2) = -\frac{nT}{2} (\ln (2\pi \sigma^2) + 1) - \frac{nT}{2} \ln \sigma_0^2 + \frac{1}{2} \ln |S_n(\lambda_1)| + \ln |R_n(\lambda_2)|.
$$

(6)

and Eq. (5) becomes

$$
\ln L_{n,t}(\lambda_1, \lambda_2) = -\frac{n(T-1)}{2} (\ln (2\pi \sigma^2) + 1) - \frac{n(T-1)}{2} \ln \sigma_0^2 + \frac{1}{2} \ln |S_n(\lambda_1)| + \ln |R_n(\lambda_2)|.
$$

where $\sigma_0^2/\sigma^2 = \beta_0^2/\beta^2$ and $\sigma_0^2/\sigma^2 = \beta_0^2/\beta^2$. By comparing Eqs. (6) and (7), we see that they yield the same maximizer $(\hat{\lambda}_1, \hat{\lambda}_2)$.

2.3. Random effects specification

In this section, we consider the random effects specification of the individual effects $\mu_t$. When the individual effects are random and independent of the exogenous regressors, the estimation under the random effects will be more efficient. The spatial effect in $\mu_t$, 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 $c_n$ can be allowed. Hence, the model is

$$
Y_{nt} = I_{t} b_0 + z_{n} \eta_{n} + \lambda_0 W_{n1} Y_{nt} + \lambda_0 W_{n2} \eta_{n} + \mu_t + U_{nt}, t = 1, \cdots, T.
$$

(8)

where $b_0$ is the coefficient for the constant term, and $\gamma$ is the parameter vector for the time invariant regressor $z_t$. Denote $C_n = I_n - \lambda_0W_{n_t}$, $Y_{n_t} = (Y_{n1}, Y_{n2}, \ldots, Y_{nT})'$ and $V_{n_t}$, $X_{n_t}$ similarly. The above equation in the vector form is

$$
Y_{n_t} = I_T \otimes (b_0 + z_t) + \lambda_0 (I_T \otimes W_{n_t}) Y_{n_t} + X_{n_t} + I_T \otimes C^{-1}_n e_{n_0} + (I_T \otimes R^{-1}_n) V_{n_t}.
$$

Under the assumptions that $e_{n_0} = (0, \omega_1^2, \omega_2^2, \ldots, \omega_T^2)$, $V_{n_t}$ is $\text{AR}(0)$, and they are uncorrelated, the variance matrix of $I_T \otimes C^{-1}_n e_{n_0} + (I_T \otimes R^{-1}_n) V_{n_t}$ would be

$$
\Omega_{n_t} = \omega_1^2 |I_T \otimes (C^{-1}_n e_{n_0})| + \omega_T^2 |I_T \otimes (R^{-1}_n) V_{n_t}|.
$$

From the likelihood function, ML random effects estimates can be obtained. By denoting $R_{n_t} = I_T \otimes R_n$ and $S_{n_t} = I_T \otimes S_n$, the log likelihood is

$$
\ln L(Y_{n_t}) = -\frac{n_T}{2} \ln(2\pi) - \frac{1}{2} \ln |\Omega_{n_t}| - \varnothing_n Y_{n_t} \Omega_{n_t}^{-1} \varnothing_n (0),
$$

where $\varnothing_n(0) = S_{n_t} Y_{n_t} - X_{n_t} + I_T \otimes (b_0 + z_t)$. For the inverse and determinant of $\Omega_{n_t}$, the calculation can be made to that of an $n \times n$ matrix. By Lemma 2.2 in Magnus (1982), Baltagi et al. (2007a) show that

$$
\Omega^{-1}_{n_t} = \frac{1}{|I_T \otimes C^{-1}_n e_{n_0}| + \omega_T^2 |I_T \otimes (R^{-1}_n) V_{n_t}|} |I_T \otimes (C^{-1}_n e_{n_0})|^{-1} - I_T \otimes \delta\theta_n^{-1} |I_T \otimes (R^{-1}_n) V_{n_t}|^{-1}.
$$

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 $Y_{n_t} = \lambda_0W_{n_t} + U_{n_t}$ with $U_{n_t} = \lambda_0W_{n_t} + \mu_0 + \eta_{n_t}$ and $\varnothing_n = \mu_n + V_{n_t}$. This model specification implies that $W_{n_t} = W_{n_0}$ and $\lambda_0 = \lambda_0$ in Eq. (8). The variance matrix of the error components is

$$
\Omega_{n_t}^{\text{hyp}} = (\omega_1^2 |I_T \otimes (C^{-1}_n e_{n_0})| + \omega_T^2 |I_T \otimes (R^{-1}_n) V_{n_t}|)^{-1},
$$

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 $(\lambda, \omega_1^2, \omega_T^2)$, where $\omega_1^2 = \omega_0^2 + \omega_T^2$. The $\beta$ can be consistently estimated by OLS for their regression equation. Denote $\hat{u}_{n_t} = I_T \otimes (W_{n_t}) \hat{u}_{n_t}$, $\hat{u}_{n_t} = I_T \otimes (W_{n_t}) \hat{u}_{n_t}$, and $\hat{e}_{n_t} = I_T \otimes (W_{n_t}) \hat{e}_{n_t}$. Also, let $Q_{0,0,0} = |I_T \otimes (W_{n_t})|$ and $Q_{1,1} = \frac{1}{\pi} \arccos^2$. For $T \geq 2$, they suggest to use the moment conditions

$$
E \left[ \begin{array}{c}
\frac{1}{n(T-1)} \hat{e}_{n_t} Q_{0,0,0} \hat{u}_{n_t} \\
\frac{1}{n(T-1)} \hat{u}_{n_t} Q_{0,0,0} \hat{e}_{n_t} \\
\frac{1}{n(T-1)} \hat{Q}_{0,0,0} \hat{e}_{n_t} \\
\frac{1}{n(T-1)} \hat{Q}_{0,0,0} \hat{u}_{n_t} \\
\frac{1}{n} \hat{Q}_{1,1} \hat{Q}_{0,0,0} \hat{e}_{n_t} \\
\frac{1}{n} \hat{Q}_{1,1} \hat{Q}_{0,0,0} \hat{u}_{n_t} \\
\frac{1}{n} \hat{Q}_{1,1} \hat{Q}_{1,1} \hat{e}_{n_t} \\
\frac{1}{n} \hat{Q}_{1,1} \hat{Q}_{1,1} \hat{u}_{n_t}
\end{array} \right] = \left[ \begin{array}{c}
\omega_0^2 \\
\omega_0^2 + \frac{1}{n} \text{tr}(W_{n_t}^2) \\
0 \\
0 \\
\omega_T^2 + \frac{1}{n} \text{tr}(W_{n_t}^2) \\
0 \\
\omega_T^2 + \frac{1}{n} \text{tr}(W_{n_t}^2) \\
0
\end{array} \right].
$$

As $e_{n_t} = u_{n_t} - \lambda_0 b_{n_t}$ and $e_{n_t} = \bar{u}_{n_t} - \lambda_0 b_{n_t}$ because $u_{n_t} = \lambda_0(I_T \otimes W_{n_t}) u_{n_t} + \bar{u}_{n_t} + \hat{e}_{n_t}$, we can substitute $\hat{u}_{n_t}$ and $\hat{e}_{n_t}$ into Eq. (9) and obtain a system of moments about $\hat{u}_{n_t}, \bar{u}_{n_t}$ and $\hat{e}_{n_t}$ with estimates of $(\lambda, \omega_0^2, \omega_T^2)$.

---

6 Note that the $\omega_T^2$ will become $\omega_T^2$ in Kapoor et al. (2007)’s specification.
where $k_{o}$ 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, $\delta$, is consistent and asymptotically efficient. However, under the alternative hypothesis, $\theta$ 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 $n \left( \hat{\delta} - \hat{\delta}_{w} \right) \Omega_{w}^{-1} \left( \hat{\delta} - \hat{\delta}_{w} \right)^{\prime}$, where $\Omega_{w}$ is a consistent estimate of the limiting variance matrix of $\sqrt{n} \left( \hat{\delta} - \hat{\delta}_{w} \right)$ under the null hypothesis, and $\Omega_{w}^{-1}$ is its generalized inverse. This test statistic will be asymptotically $\chi^{2}$ distributed, and its degrees of freedom is the rank of $\Omega_{w}$. Because $\hat{\delta}$ is asymptotically efficient, $\frac{\left[ \left( \frac{1}{n} \hat{\delta} \hat{\delta}^{\prime} \right)^{\prime} \Omega_{w} \left( \hat{\delta} \hat{\delta}^{\prime} \right)^{\prime} - \left( \frac{1}{n} \hat{\delta} \hat{\delta}^{\prime} \right)^{\prime} \left( \Omega_{w}^{-1} \Omega_{w} \right) \left( \frac{1}{n} \hat{\delta} \hat{\delta}^{\prime} \right)^{\prime} \right]}{\left( \Omega_{w}^{-1} \Omega_{w} \right)^{-1}}$, evaluated at either $\frac{1}{n} \hat{\delta}$ or $\hat{\delta}_{w}$, provides a consistent estimate of $\Omega_{w}$ under the null. By using the identity $B^{-1} - (B + C)^{-1} - B^{-1} \left( B^{-1} - C^{-1} \right)^{-2} \left( B^{-1} - C^{-1} \right)^{-1}$ for any two positive definite matrices $B$ and $C$, the preceding expression of the two information matrices is a positive definite matrix. Therefore, the generalized inverse is an inverse, and the degrees of freedom of the $\chi^{2}$ test is the number of common parameters, i.e., the dimension of $\delta$. Instead of the ML approach, if the main equation is estimated by the 2SLS method, Hausman test statistics can be constructed as in Mutluk and Palfenmayer (2008).

With the estimates of the spatial effect parameters $\lambda_{01}$ and $\lambda_{02}$, 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 $J_{n} = I_{n} - \frac{1}{T} J_{n}^{T}$ with the transformation from $J_{T}$ to eliminate both the individual and fixed effects. Let $(F_{T \rightarrow 1}, \frac{1}{\sqrt{T}} I_{n})$ be the orthonormal matrix of $J_{n}$, where $F_{T \rightarrow 1}$ corresponds to the eigenvalues of 1 and $\frac{1}{\sqrt{T}} I_{n}$ corresponds to the eigenvalue zero. The individual effects can be eliminated by $F_{T \rightarrow 1}$ as in Eq. (4), which yields

$$Y_{nt}^{\ast} = \lambda_{01} W_{nt}^{\ast} Y_{nt}^{\ast} + X_{nt}^{\ast} \beta_{0} + \alpha_{01} \beta_{n} + U_{nt}^{\ast},$$

(12)

$$U_{nt}^{\ast} = \lambda_{02} W_{nt} U_{nt}^{\ast} + V_{nt}^{\ast}, \quad t = 1, 2, \ldots, T - 1,$$

where $[\alpha_{11}^{T} \beta_{n}, \alpha_{21}^{T} \beta_{n}, \ldots, \alpha_{T - 1}^{T} \beta_{n}] = [\alpha_{11} \beta_{m}, \alpha_{21} \beta_{m}, \ldots, \alpha_{T - 1} \beta_{m}] | F_{T \rightarrow 1}^{-1}$ 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 $Y_{nt}^{\ast}$ as $Y_{nt}^{\ast} = F_{T \rightarrow 1}^{-1} Y_{nt}$. Such a transformation to $Y_{nt}^{\ast}$ can result in a well-defined spatial panel model when $W_{n1}$ and $W_{n2}$ are assumed to be row-normalized. Therefore, we have

$$Y_{nt}^{\ast\ast} = \lambda_{01} (F_{n1} W_{n1}^{\ast} F_{n1}^{T}) Y_{nt}^{\ast} + X_{nt}^{\ast} \beta_{0} + U_{nt}^{\ast\ast},$$

(13)

$$U_{nt}^{\ast\ast} = \lambda_{02} (F_{n1} W_{n1}^{\ast} F_{n1}^{T}) U_{nt}^{\ast} + V_{nt}^{\ast\ast},$$

for $t = 1, \ldots, T - 1$ where $X_{nt}^{\ast\ast} = F_{n1}^{T} X_{nt}^{\ast}$ and $V_{nt}^{\ast\ast} = F_{n1}^{T} V_{nt}^{\ast}$. 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

$$Y_{nt} = l_{n} \beta_{0} + z_{nt} \gamma_{0} + \lambda_{01} W_{nt} Y_{nt} + X_{nt} \beta_{0} + \mu_{n} + \alpha_{01} \beta_{n} + U_{nt},$$

(14)

where $\lambda_{01} = \lambda_{02} = \mu_{n} = \lambda_{02} W_{nt} U_{nt} + V_{nt},$ for $t = 1, \ldots, T$. In the vector form, it is

$$Y_{nt} = l_{n} \beta_{0} + z_{nt} \gamma_{0} + \lambda_{1} (l_{n} \beta_{0}) Y_{nt} + X_{nt} \beta_{0} + l_{n} \beta_{n}^{1} \gamma_{0} + \alpha_{01} \beta_{n} + U_{nt},$$

where $\gamma_{0} = (0, \alpha_{10}, \ldots, \alpha_{T - 1} \beta_{m})^{T}$. As $\alpha_{10} = (0, \alpha_{10}^{2}, \ldots, \alpha_{T - 1} \beta_{m})^{T}$, and $\gamma_{0} = (0, \alpha_{10}^{2}, \ldots, \alpha_{T - 1} \beta_{m})^{T}$, and they are uncorrelated, the variance matrix of the overall disturbances $l_{T} \beta_{0}^{C_{0}} \epsilon_{01}^{\ast} + \alpha_{01} \beta_{n}^{C_{0}} + (l_{T} \beta_{n}^{C_{0}}) | V_{nt}$ would be

$$\Omega_{nt} = \alpha_{10}^{2} [l_{T} \beta_{0}^{C_{0}}]^{1} + \alpha_{01}^{2} [l_{T} \beta_{n}^{C_{0}}] + \alpha_{01}^{2} [l_{T} \beta_{n}^{C_{0}}].$$

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

$$\ln L(Y_{nt}) = -\frac{n T}{2} \ln (2\pi) - \frac{1}{2} \ln | \Omega_{nt} | + T \ln | S_{0} | - \frac{1}{2} \lambda_{01}^{T} (\lambda_{01} \Omega_{nt}^{-1} \lambda_{01})^{1},$$

where $S_{0}(\theta) = S_{0} - X_{nt}^{\prime} \beta_{0} - \frac{1}{T} \sum_{t=1}^{T} l_{T} \beta_{0} + z_{nt} \gamma_{0}$. The calculation of the inverse and determinant of $\Omega_{nt}$ will involve essentially those of a $T \times T$ matrix as well as an $n \times n$ matrix. As a further generalization, $\alpha_{01}$ 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.

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 estimators, 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:

\[
Y_{nt} = \lambda_0 W_n Y_{nt-1} + \gamma_0 Y_{nt-1} + \rho_0 W_n Y_{nt-1} + X_{nt} \alpha_0 + \alpha_0 \delta_n + V_{nt},
\]

where \( \gamma_0 \) captures the pure dynamic effect and \( \rho_0 \) 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 \( A_n = S_n^{-1}(\gamma_0 \delta_n + \rho_0 W_n) \), Eq. (15) can be rewritten as

\[
Y_{nt} = A_n Y_{nt-1} + S_n^{-1} X_{nt} \alpha_0 + \alpha_0 \delta_n + V_{nt}.
\]

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 \( \sigma_\alpha = diag(\sigma_{\alpha_1}, \ldots, \sigma_{\alpha_n}) \) be the \( n \times n \) diagonal eigenvalue matrix of \( W_n \) such that \( W_n = \Gamma_n \sigma_\alpha \Gamma_n^{-1} \) where \( \Gamma_n \) is the corresponding eigenvector matrix. As \( A_n = S_n^{-1}(\gamma_0 \delta_n + \rho_0 W_n) \), the eigenvalue matrix of \( A_n \) is \( D_n = (I_m - \lambda_0 \sigma_\alpha)^{-1} (\gamma_0 \delta_n + \rho_0 \sigma_\alpha) \) 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 \( J_n = diag(1_{m_n}, 0, \ldots, 0) \) with 1_{m_n} being an \( m_n \times 1 \) vector of ones and \( \bar{D}_n = diag(0, \ldots, 0, d_n, m_{n+1}, \ldots, m_n) \), where \( d_n < 1 \) is assumed for \( i = m_n + 1, \ldots, n \). As \( J_n \bar{D}_n = 0 \), we have \( A_n^i = (\gamma_0 + \rho_0 \sigma_\alpha)^i \Gamma_n \sigma_\alpha^{-1} + B_n^i \) where \( B_n^i = \Gamma_n D_n^i \Gamma_n^{-1} \) for any \( h = 1, 2, \ldots \).

Denote \( W_n = \Gamma_n J_n \Gamma_n^{-1} \). 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:

\[
Y_{nt} = Y_{nt}^m + Y_{nt}^s + Y_{nt}^e,
\]

where

\[
Y_{nt}^m = \sum_{h=0}^n \Gamma_n B_n^h \sigma_\alpha^{-1} \Gamma_n^{-1},
Y_{nt}^s = \sum_{h=0}^n J_n \sigma_\alpha^{(h+1)} (\gamma_0 + \rho_0 \sigma_\alpha)^h \sigma_\alpha^{-1} \Gamma_n^{-1},
Y_{nt}^e = \sum_{h=0}^n J_n \sigma_\alpha^{(h+1)} \Gamma_n D_n \sigma_\alpha^{-1} \Gamma_n^{-1}.
\]

The \( Y_{nt}^m \) can be an unstable component when \( \gamma_0 + \rho_0 = 1 \), 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 \( \gamma_0 + \rho_0 > 1 \), and \( Y_{nt}^m \) can be explosive. The \( Y_{nt}^m \) can be complicated, as it depends on what the time dummies exactly represent. The \( Y_{nt}^s \) can be explosive when \( \alpha_\delta \) represents some explosive functions of \( t \), even when \( \gamma_0 + \rho_0 = 1 \). Without an explicit specification for \( \alpha_\delta \), it is desirable to eliminate this component for estimation. The \( Y_{nt}^e \) can be a stable component unless the sum \( \gamma_0 + \rho_0 + \lambda_0 \) is much larger than 1. If \( \gamma_0 + \rho_0 + \lambda_0 > 1 \), some of the eigenvalues of \( d_n \) in \( Y_{nt}^e \) might become larger than 1.

### 3.2. Stable, spatial cointegration, and explosive cases

For notational purposes, we define \( \bar{Y}_{nt} = Y_{nt} - \bar{Y}_{nt} \) and \( \bar{Y}_{nt-1} = Y_{nt-1} - \bar{Y}_{nt} \) for \( t = 1, 2, \ldots, T \) where \( \bar{Y}_{nt} = \frac{1}{T} \sum_{s=n}^{T} Y_{st} \) and \( \bar{Y}_{nt-1} = \frac{1}{T} \sum_{s=n-1}^{T} Y_{st-1} \). 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 \( \theta = (\delta^2, \lambda') \) and \( \zeta = (\delta', \lambda, \sigma_\delta^2) \) where \( \delta = (\gamma, \rho, \beta)' \). At the true value, \( \theta_0 = (\delta_0, \lambda_0, \sigma_\delta^2) \) and \( \zeta_0 = (\delta_0, \lambda_0, \sigma_\delta^2) \) where \( \delta_0 = (\gamma_0, \rho_0, \lambda_0, \sigma_\delta^2) \).

\[ d_n = (\gamma_0 + \rho_0 \sigma_\alpha)/(1 - \lambda_0 \sigma_\alpha). \]

Hence, if \( \gamma_0 + \rho_0 \sigma_\alpha > 1 \), we have \( d_n < 1 \). Some additional conditions are needed to ensure that \( d_n > -1 \). See Appendix A in Lee and Yu (2009).

---

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 assume that \( |d_n| \leq 1 \).
\( \rho \), \( \beta \)'s. By denoting \( Z_{nt} = (Y_{nt} - \mu_1, W_{nt}, \ldots, W_{nt}) \), the likelihood function of Eq. (15) is

\[
\ln L_{T}(\theta, \phi_n) = \frac{nT}{2} \ln 2\pi - \frac{nT}{2} \ln \sigma^2 + T \ln S_{n}(\lambda) \nonumber
\]

where \( \bar{Y}_{nt}(\lambda) = S_{n}(\lambda)\bar{Y}_{nt} = Z_{nt} - \delta_{n} \). The QMLEs \( \hat{\theta}_n \) and \( \hat{\phi}_n \) are the extremum estimators derived from the maximization of Eq. (18), and \( \hat{\phi}_n \) can be consistently estimated when \( T \) goes to infinity.

Using the first order condition for \( \phi_n \), the concentrated likelihood is

\[
\ln L_{T}(\theta) = \frac{nT}{2} \ln 2\pi - \frac{nT}{2} \ln \sigma^2 + T \ln S_{n}(\lambda) \nonumber
\]

where \( \bar{Y}_{nt}(\lambda) = S_{n}(\lambda)\bar{Y}_{nt} = Z_{nt} - \delta_{n} \). The QMLE \( \hat{\theta}_n \) maximizes the concentrated likelihood function (19). As is shown in Yu et al. (2008), we have

\[
\sqrt{nT}(\hat{\theta}_n - \theta_0) + \sqrt{n}(\hat{\phi}_n + \phi_0) \left( \max \left( \frac{n}{nT}, \frac{1}{T} \right) \right)
\]

and

\[
\left. \frac{1}{\sqrt{nT}} \right| \frac{n}{nT} \rightarrow \frac{1}{T} \chi^2(1) \text{ and } \frac{n}{nT} \rightarrow \frac{1}{T} \chi^2(3),
\]

and

\[
\Sigma_{n,nt} = \frac{1}{n} \begin{pmatrix}
EH_{nt} & 0 \\
0 & \lambda \end{pmatrix}
\]

with \( G_n \equiv W_nS_n^{-1} \) and \( H_{nt} = \frac{1}{T} \sum_{t=1}^{T} (\bar{Z}_{nt}, \bar{G}_n, \bar{G}_n^{-1})(\bar{Z}_{nt}, \bar{G}_n, \bar{G}_n^{-1}). \)

\[ \text{Hence, for distribution of common parameters, when } T \text{ is large relative to } n, \text{ the estimators are } \sqrt{nT} \text{ consistent and asymptotically normal, with the limiting distribution centered around } 0; \text{ when } n \text{ is asymptotically proportional to } T, \text{ the estimators are } \sqrt{nT} \text{ consistent and asymptotically normal, but the limiting distribution is not centered around } 0; \text{ and when } n \text{ is large relative to } T, \text{ the estimators are } T \text{ consistent, and have a degenerate limiting distribution.} \]

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

\[
\sqrt{nT}(\hat{\theta}_n - \theta_0) + \sqrt{n}(\hat{\phi}_n + \phi_0) \left( \max \left( \frac{n}{nT}, \frac{1}{T} \right) \right)
\]

and

\[ N(0, \lim_{T \rightarrow \infty} \Sigma_{n,nt}^{-1}(\Sigma_{n,nt} + \Omega_nnt)^{-1}) \Sigma_{n,nt}^{-1} \]

where \( \Sigma_{n,nt} \equiv \sum_{n} \hat{\phi}_n + \phi_0 = \text{the leading bias term of order } O(1) \text{ and } \phi_2 = \phi_n + \phi_0 \).

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

\[
\sqrt{nT}(\hat{\theta}_n - \theta_0) + \sqrt{n}(\hat{\phi}_n + \phi_0) \left( \max \left( \frac{n}{nT}, \frac{1}{T} \right) \right)
\]

and

\[ N(0, \lim_{T \rightarrow \infty} \Sigma_{n,nt}^{-1}) \]

Here, \( \Sigma_{n,nt} = \sum_{n} \hat{\phi}_n + \phi_0 = \text{the leading bias term of order } O(1) \text{ and } \phi_2 = \phi_n + \phi_0 \).

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 \( (I_n - W) \), and its rank is the number of eigenvalues of \( W \) 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 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),

\[ \Delta Y_{nt} = (A_n - I_n)Y_{n,t-1} + S_n^{-1}(X_{n,t-1} + \phi_n + V_{nt}) \]

As \( \gamma_0 + \rho_0 + \lambda_0 = 1 \), it follows that \( A_n = (I_n - \lambda_0W) \) and \( \Delta Y_{nt} = (I_n - \gamma_0W)Y_{n,t-1} + S_n^{-1}(X_{n,t-1} + \phi_n + V_{nt}) \). Hence, \( Y_{nt} = (I_n - \lambda_0W)Y_{n,t-1} + S_n^{-1}(X_{n,t-1} + \phi_n + V_{nt}) \).

As \( W_n = \Gamma_n \sigma_n \Gamma_n^{-1} + \text{and } M_n = \Gamma_n \sigma_n \Gamma_n^{-1} \), it follows that \( \Delta Y_{nt} = (I_n - \lambda_0W)Y_{n,t-1} + S_n^{-1}(X_{n,t-1} + \phi_n + V_{nt}) \), which depends only on the stationary component.
Therefore, $Y_{nt}$ is spatially cointegrated. The matrix $I_n - W_n = F_n(I_n - \sigma_{22}I_n^{-1})$ 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 $I_n - W_n$ the case with time dummys

When we have time effects included in the SDPD model, the direct estimation procedure above will yield a bias of order $O(\max(1/n, 1/T))$ for the common parameters. 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 will have a degenerate limit distribution, but the estimates under the transformation approach are properly centered and asymptotically normal.

With the transformation $J_n$, when $W_nJ_n = I_n$, i.e., $W_n$ is a row-normalized matrix, $J_nW_n(J_n + 1/nI_nJ_n) = J_nW_n$ because $J_nW_nI_n = J_nI_n = 0$. Hence,

\[ (J_nY_{nt}) = \lambda_0(J_nW_n)(J_nY_{nt}) + \gamma_0(J_nY_{n,t-1}) + \rho_0(J_nW_n)(J_nY_{n,t-1}) \]

\[ + (J_nW_n)c_0 + (J_nV_{nt}), \tag{24} \]

which does not involve the time effects, and $J_nW_n$ can be regarded as the transformed individual effects. With the additional transformation $F_{n,t-1}$, by denoting $Y_{nt}^* = F_{n,t-1}Y_{nt} = F_{n,t-1} - V_{nt}$ which is of the dimension $(n-1)$, we have

\[ Y_{nt}^* = \lambda_0W_{nt}Y_{nt}^* + \gamma_0Y_{n,t-1}^* + \rho_0W_{nt}Y_{n,t-1}^* + \epsilon_{n0}^* + V_{nt}^*. \tag{25} \]

where $W_{nt}^* = F_{n,t-1}W_{nt}F_{n,t-1}$, $X_{nt}^* = F_{n,t-1}X_{nt}$, $\epsilon_{n0}^* = F_{n,t-1}c_0$, and $V_{nt}^* = F_{n,t-1}V_{nt}$. The $Y_{nt}^*$ is an $(n-1)$ dimensional disturbance vector with zero mean and variance matrix $\sigma_{Y^2}^* = \sigma_{Y}/\sigma_{X}$. 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 $N(0, \sigma^2Y^2)$, the transformed $Y_{nt}^*$ will be $N(0, \sigma^2Y^2)$. Thus, the log likelihood function of Eq. (25) can be written as

\[ \ln L_{n,t}(\theta, \epsilon_{n0}^*) = -\frac{(n-1)T}{2}\ln 2\pi - \frac{(n-1)T}{2}\ln \sigma^2 - T\ln(1-\lambda) \]

\[ + T\ln |I_n - \lambda W_n| - \frac{T}{2\sigma^2} \sum_{i=1}^{T} V_{nt,i}^*(\theta)J_nV_{nt,i}^*(\theta). \tag{26} \]

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 dummys but also the unstable component. Hence, we end up with the following equation after the $I_n - W_n$ transformation:

\[ (I_n - W_n)Y_{nt} = \lambda_0W_{nt}(I_n - W_n)Y_{nt} + \gamma_0(I_n - W_n)Y_{n,t-1}^* \]

\[ + \rho_0W_{nt}(I_n - W_n)Y_{n,t-1}^* + (I_n - W_n)c_0 + (I_n - W_n)V_{nt}. \tag{27} \]

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 $(I_n - W_n)U_{nt}$ which is the number of non-zero eigenvalues of $(I_n - W_n)(I_n - W_n)^\dagger$. 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 $\sigma^2_0 \sum_n$, where $\sum_n = (I_n - W_n)(I_n - W_n)^\dagger$. Let $[F_n, H_n]$ be the orthonormal matrix of eigenvectors and $A_n$ be the diagonal matrix of nonzero eigenvalues of $\sum_n$ such that $\sum_n F_n = F_nA_n$ 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 $\lambda_n$ is an $n^* \times n^*$ diagonal matrix. Denote $W_{nt} = A_n^{-1/2}F_nW_{nt}A_n^{1/2}$ which is an $n^* \times n^*$ matrix. We have

\[ Y_{nt}^* = \lambda_0W_{nt}Y_{nt}^* + \gamma_0Y_{n,t-1}^* + \rho_0W_{nt}Y_{n,t-1}^* + X_{nt}^*c_0 + \epsilon_{n0}^* + V_{nt}^*. \tag{28} \]

where $Y_{nt}^* = A_n^{-1/2}F_n(I_n - W_n)Y_{nt}$ and other variables are defined accordingly. Note that this transformed $Y_{nt}^*$ is an $n^*-dimensional$ vector. The eigenvalues of $W_{nt}^*$ are exactly those eigenvalues of $W_{nt}$ less than 1 in absolute value. It follows that the eigenvalues of $A_{nt}^* = (I_n - \lambda_0W_{nt}^{-1})(\gamma_0 + \rho_0 + \lambda_0)^{-1}$ are all less than 1 in absolute value even when $\gamma_0 + \rho_0 + \lambda_0 = 1$ with $|\lambda_0| < 1$. Hence, the transformed model (28) is 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

\[ \ln L_{n,t}(\theta) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln \sigma^2 - (n - n^*)T\ln(1-\lambda) + T\ln |I_n - \lambda W_n| \]

\[ - \frac{T}{2\sigma^2} \sum_{i=1}^{T} V_{nt,i}^*(\theta)J_nV_{nt,i}^*(\theta), \tag{29} \]

where $V_{nt,i}^*(\theta) = S_0(\lambda)Y_{nt,i}^* - Z_{nt,i}^\dagger$. 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. 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 dummys. 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.

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 $\hat{\theta}_n = \sum_{n \in \mathbb{N}} \varphi_n$ where $\varphi_n$ is in Eq. (21); for the spatial cointegration case, the bias is $\hat{\theta}_n = \sum_{n \in \mathbb{N}} \varphi_n$ where $\varphi_n$ is in Eq. (23). For the stable case with the transformation $J_n$, the bias is $\hat{\theta}_n = \sum_{n \in \mathbb{N}} \varphi_n$ where

$$
\varphi_n = \begin{pmatrix}
\frac{1}{n^{\alpha}} \nu \left( I_n, J_n, \sum_n \gamma\vartheta_n \right) \\
\frac{1}{n^{\alpha}} \nu \left( W_n, J_n, \sum_n \gamma\vartheta_n \right)
\end{pmatrix}.
$$

For the unified transformation approach, the bias is $\hat{\theta}_n = \sum_{n \in \mathbb{N}} \varphi_n$ and

$$
\varphi_n = \begin{pmatrix}
\frac{1}{n^{\alpha}} \nu \left( I_n, J_n, \sum_n \gamma\vartheta_n \right) \\
\frac{1}{n^{\alpha}} \nu \left( W_n, J_n, \sum_n \gamma\vartheta_n \right)
\end{pmatrix}.
$$

(30)

where $\gamma\vartheta_n = \left( J_n - W_n \right) \sum_n \gamma\vartheta_n$.

Hence, the QMLE $\hat{\theta}_n$ has the bias $-\hat{\theta}_n \hat{\theta}_n$ and the confidence interval is not centered when $n^{\alpha} - c$ where $n^{\alpha}$ 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 $T^{-\alpha}$, the presence of $\hat{\theta}_n$ causes $\hat{\theta}_n$ to have the smaller $T$-rate of convergence. An analytical bias reduction procedure is correct to the bias $B_n = -\hat{\theta}_n \hat{\theta}_n$ by constructing an estimate $\hat{B}_n$. The bias corrected estimator is

$$
\hat{\theta}_n = \hat{\theta}_n - \hat{B}_n.
$$

(32)

We may choose

$$
\hat{B}_n = \left[ \left( E \left( \frac{1}{n^{\alpha}} \nu \left( I_n, J_n, \sum_n \gamma\vartheta_n \right) \right) \right) \right]^{-1} \nu_1 = \hat{\theta}_n \hat{\theta}_n.
$$

(33)

where $i = 1, 2, 3, 4$ corresponds to stable, spatial cointegration, $J_n$-transformed and $(L_n - W_n)$-transformed models. When $T$ grows faster than $n^{\alpha/3}$, the correction will eliminate the bias at 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

$$
Y_{nt} = \gamma_0 Y_{nt-1} + X_{nt}\rho_{nt} + Z_{nt}\epsilon_{nt} + U_{nt}, \quad t = 1, ..., T,
$$

$$
U_{nt} = \mu_0 + \epsilon_{nt}, \quad \text{and} \quad \epsilon_{nt} = \lambda_0 W_n \epsilon_{nt} + \epsilon_{nt}.
$$

(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 $Y_{nt} = \lambda_0 W_n Y_{nt-1} + \gamma_0 Y_{nt-1} + X_{nt}\rho_{nt} - W_n X_{nt}\lambda_0 Y_{nt-1} + \epsilon_{nt}$. This corresponds to an SDPD model with transformed individual effects $c_0 = (J_n - W_n)^{-1} \mu_0$, nonlinear constraints $\rho_0 = -\lambda_0$, and $X_{nt}\rho_{nt} = X_{nt}\lambda_0 Y_{nt-1}$ with $x_{nt} = [X_{nt}, W_n Y_{nt}]$ and $\beta_k = [\beta_k, -\lambda_0 \beta_k]$. 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 $\mu_0$, 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 $Y_{nt}$ is important. When $Y_{nt}$ 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 $Y_{nt}$ is assumed to be endogenous, $Y_{nt}$ will need to be generated from a stationary process, or its distribution will be approximated. With the corresponding likelihood, QMLE can be obtained.

3.3.1. Pure unit root case

In Yu and Lee (2007) for the SDPD model, when $\gamma_0 = 1$ and $\rho_0 = 0$, we have $A_n = 0$. 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 $\rho_0 = 0$ 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 $\gamma_0$ is $\sqrt{nT}$ 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 $\lambda_0$ and the dynamic spatial effect estimate of $\rho_0$ will converge at $\sqrt{nT}$ 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 $\gamma_0$ are $\sqrt{nT}$ consistent; only the sum of the pure dynamic and spatial effects estimates is convergent at the faster $\sqrt{nT}$ 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 $\gamma_0$ 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.

---

13 Murri (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.
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 $\sigma^2_0|\Omega_T| + \Theta_0|\Sigma_S|^{-1}$ where $\Theta_0 = \sigma^2_0/\sigma^2_c$. There are two cases under this specification.

**Case 1. $Y_{it}$ is exogenous.** Let $\theta = (j', \eta', \gamma')$, $\delta = (\lambda, \phi, \gamma)$ and $\varsigma = (\theta', \sigma^2_0, \sigma^2_1')$. The log likelihood is

$$
\ln L(\varsigma) = -\frac{nT}{2} \ln(2\pi) - \frac{nT}{2} \ln(\sigma^2_0) - \frac{1}{2} \ln|\Omega_T| - \frac{1}{2 \sigma^2_0} \hat{u}_{it}^T(0) \Omega_{it}^{-1} \hat{u}_{it}(0),
$$

where $\hat{u}_{it}(\theta) = Y_{it} - \Theta_0|\Sigma_S|^{-1} - X_{it}\hat{\eta}_0 - 1_1 \otimes \sigma_{\nu}^2$ with $\Theta_0 = (Y_{it} - \Theta_0|\Sigma_S|^{-1})$ and other variables in the vector form are similarly defined. By concentration, we can work on the log likelihood with $\delta$.

$$
\ln L(\delta) = -\frac{nT}{2} \ln(2\pi) - \frac{nT}{2} \ln(\sigma^2_0|\Omega_T|) - \frac{1}{2} \ln|\Omega_T|,
$$

where $\hat{\sigma}^2_0(\delta) = \frac{1}{\sum_i\hat{u}_{it}(\delta)\hat{u}_{it}(\delta)^T}$, $\hat{u}_{it}(\delta) = Y_{it} - Z_{it}\theta(\delta)$ with $Z_{it} = (X_{it}, \Theta_0|\Sigma_S|^{-1})$ and $\theta = (Z_{it}|\Sigma_S|^{-1} - Z_{it}\theta(\delta))^T Z_{it}|\Sigma_S|^{-1} = \hat{\sigma}^2_0(\delta)\Omega_{it}^{-1}$. The disturbances of the initial period are specified as $\sigma_{\nu}^2 = \sigma_{\nu}^2 = \sigma_{\nu}^2 + \sum_j\frac{\sigma_{\nu}^2}{\sum_i\hat{u}_{it}(\delta)\hat{u}_{it}(\delta)^T}$. The variance matrix is $E(\epsilon_{it}^2) = \sigma^2_{it} + \frac{\sigma^2_0}{1 - \gamma^2}$, where $\sigma_{\nu}^2$ has mean zero, its variance matrix is $E(\epsilon_{it}^2) = \sigma^2_{it} + \frac{\sigma^2_0}{1 - \gamma^2}$, and its covariance with $\hat{u}_{it}$ is $E(\epsilon_{it}\hat{u}_{it}) = \frac{\sigma^2_0}{1 - \gamma^2} \delta_i$. The log likelihood is

$$
\ln L(\delta) = -\frac{nT}{2} \ln(2\pi) - \frac{nT}{2} \ln(\sigma^2_0|\Omega_T|) - \frac{1}{2} \ln|\Omega_T|.
$$

**Case 2. $Y_{it}$ is endogenous.** Eq. (34) implies that $Y_{it} = \hat{Y}_{it} + \xi_{it}$ where $\hat{Y}_{it}$ is the exogenous part of $Y_{it}$ and $\xi_{it}$ is the endogenous part. The exogenous part $\hat{Y}_{it} = \sum_j\gamma_j Y_{i(t-j)} + \sum_j\kappa_j Y_{i(t-j)} + \sum_j\lambda_j X_{i(t-j)} + \sum_j\phi_j X_{i(t-j)}$, and the endogenous part $\xi_{it} = \sum_j\gamma_j Y_{i(t-j)} + \sum_j\kappa_j Y_{i(t-j)} + \sum_j\lambda_j X_{i(t-j)} + \sum_j\phi_j X_{i(t-j)}$. The difficulty to use this directly is due to the missing observations $X_{it}$ 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 $Y_{it} = X_{it}\eta + \epsilon_{it}$ with $X_{it} = [1, X_{i(t-1)}, X_{i(t-2)}, \ldots, X_{i(t-n)}]$ and $n=\{n_0, n_1, n_2\}$, or, $X_{it} = [1, \hat{Y}_{it}, X_{i(t-1)}, X_{i(t-2)}]$. The disturbances of the initial period are specified as $\sigma_{\nu}^2 = \sum_j\gamma_j^2 \sigma_Y + \sum_j\kappa_j^2 \sigma_Y + \sum_j\lambda_j^2 \sigma_X + \sum_j\phi_j^2 \sigma_X + \sum_j\gamma_j^2 \sigma_Y + \sum_j\kappa_j^2 \sigma_Y + \sum_j\lambda_j^2 \sigma_X + \sum_j\phi_j^2 \sigma_X$, where $\gamma_j^2$ is a constant. The variance matrix is $E(\epsilon_{it}^2) = \sigma^2_{it} + \frac{\sigma^2_0}{1 - \gamma^2}$, and its covariance with $\hat{u}_{it}$ is $E(\epsilon_{it}\hat{u}_{it}) = \frac{\sigma^2_0}{1 - \gamma^2} \delta_i$. The log likelihood is

$$
\ln L(\delta) = -\frac{nT}{2} \ln(2\pi) - \frac{nT}{2} \ln(\sigma^2_0|\Omega_T|) - \frac{1}{2} \ln|\Omega_T|.
$$

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 $\Delta Y_{it}$ 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 $Y_{it} = \lambda_0 Y_{it-1} + X_{it}\eta_0 + \xi_{it} + 1_1 \otimes \sigma_{\nu}^2$, where $Y_{it}$ is a rook matrix. For each set of generated sample observations, we calculate the ML estimates under different settings and consequences of model misspecifications.

(35)
For the DGP with only individual effects, from item (1a)--(1c), we see that both approaches provide the same estimate of \( \gamma_0 = (\lambda_0, \rho_0, \sigma_0^2) \). While the estimator of \( \sigma_0^2 \) by the direct approach has a larger bias. When \( T \) is small, the transformation approach yields a consistent estimator of \( \sigma_0^2 \) while the direct approach does not. The Biases, E-SDs, RMSEs for the estimators of \( \gamma_0 \) are small when either \( n \) or \( T \) is large. Also, when \( T \) is larger, the bias of the estimator of \( \sigma_0^2 \) 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. Also, when \( T \) is larger, the bias of the estimator of \( \sigma_0^2 \) by the direct approach decreases. The Biases, E-SDs, RMSEs for the estimators of \( \gamma_0 \) are small when either \( n \) or \( T \) is large. Also, when \( T \) is larger, the bias of the estimator of \( \sigma_0^2 \) by the direct approach decreases.

For the DGP with only individual effects, from item (1a)--(1c), we see that both approaches provide the same estimate of \( \gamma_0 = (\lambda_0, \rho_0, \sigma_0^2) \). While the estimator of \( \sigma_0^2 \) by the direct approach has a larger bias. When \( T \) is small, the transformation approach yields a consistent estimator of \( \sigma_0^2 \) while the direct approach does not. The Biases, E-SDs, RMSEs for the estimators of \( \gamma_0 \) are small when either \( n \) or \( T \) is large. Also, when \( T \) is larger, the bias of the estimator of \( \sigma_0^2 \) 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. Also, when \( T \) is larger, the bias of the estimator of \( \sigma_0^2 \) by the direct approach decreases.

### 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

\[
Y_{nt} = \lambda_0 W_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_{n,t-1} + \sigma_0 \epsilon_{nt} + \alpha_0 \eta_{nt} + \nu_{nt},
\]

(36)

using \( \theta_0 = (\gamma_0, \lambda_0, \rho_0, \sigma_0^2) = (0.2, 0.2, 1, 0.2, 1)' \). We estimate the model with the direct approach, the transformation approaches with \( F_{n-1} \) 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 x 9 queen matrix. Hence, the number of the unit roots in \( W_n \) is 6. Due to space limitations, we will present the results as tables in Results 1 and 2. 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 \( \rho \) and the bias correction makes the bias even larger. In items (4) and (5) where the spatial time lag or the spatial lag is omitted, the bias and the estimates in \( \lambda_0, \rho_0, \sigma_0^2 \) 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 \( \lambda_0, \rho_0 \), 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 coinintegration case from Eq. (36)
with $h_0 = (0.4, 0.2, 1, 0.2, 1')$. 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 $h_0 = (0.4, 0.2, 1, 0.3, 1')$, which implies $h_0 + \beta_0 + \lambda_0 = 0.9$, 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 estimates 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

$$
\ln C_{nt} = \gamma_0 + \beta_0 D_{nt} + \beta_1 P_{nt} + \beta_2 P_{mt} + \mu_n + \alpha_n I_n + U_{nt} + V_{nt},
$$

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, $\mu_n$ is the vector of individual effects and $\alpha_n$ is a time effect. Elhorst (2005) estimates the model with fixed effects $\mu_n$ 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 extend 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...
case of the SDPD model with nonlinear restrictions across coefficients. By premultiplying both sides with \((I_n - \lambda_n W_n)\), the transformed equation is reduced to

\[
\ln C_{nt} = \lambda_0 W_n \ln C_{nt} + \gamma_0 \ln C_{n-1} + \rho_0 W_n \ln C_{n-1} + X_{nt}\beta_0 + \mu_n^* + \alpha_n^* l_n + V_n. 
\]

with \(\rho_0 = -\lambda_0 \gamma_0\), \(\phi_0 = -\lambda_0 \beta_0\), and \(\mu^*_n\) are transformed individual effects and time effects. Here, \(X_{nt} = [\ln l_{nt}, \ln In_{nt}, \ln P_{nt}^D] \) and \(\rho_0 = (\beta_0, \beta_2, \beta_3)\). 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 \(l_n\), 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 \(l_n - W_n\). 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 a bias for the estimate of \(\rho_0\).

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 \(\lambda_0\) and \(\rho_0\).

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)).
where \( Y_{it} \) is the selling price of mid-quality rice. Keller and Shiue (2007) argue that different weights matrices could be used. Denote \( d_{it} \) 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 \( d_{it} \leq 300 \); (2) \( W_{n}^{(2)} \), where prefectures are neighbors if the \( d_{it} \leq 600 \); (3) \( W_{n}^{(3)} \), where \( W_{n}^{(3)} = 1 \) if \( d_{it} \leq 300, \) \( W_{n}^{(3)} = 0.5 \) if \( 300 < d_{it} \leq 600 \), and \( W_{n}^{(3)} = 0 \) if \( d_{it} > 600 \); and (4) \( W_{n}^{(4)} \), where \( w_{ij} = \exp(\theta_{d_{ij}}) \) with \( D_{ij} = d_{ij} \) and a larger absolute value of a negative \( \theta_{d} \) 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 \( \lambda_0 \) are about 0.8 or slightly larger; those of \( \gamma_0 \) are about 0.5; those for \( \rho_0 \) 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 \( \theta_{d} \) in \( w_{ij} = \exp(-1.2d_{ij}) \). 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 \( \lambda_0 \) are about 0.8 or slightly larger; those of \( \gamma_0 \) are about 0.5; those for \( \rho_0 \) 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 \( \theta_{d} \) in \( w_{ij} = \exp(-1.2d_{ij}) \). 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 \( \lambda_0 \) are about 0.8 or slightly larger; those of \( \gamma_0 \) are about 0.5; those for \( \rho_0 \) 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 \( \theta_{d} \) in \( w_{ij} = \exp(-1.2d_{ij}) \). 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 \( \lambda_0 \) are about 0.8 or slightly larger; those of \( \gamma_0 \) are about 0.5; those for \( \rho_0 \) 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.

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

### Table 7

<table>
<thead>
<tr>
<th>T</th>
<th>n</th>
<th>( \gamma )</th>
<th>( \rho )</th>
<th>( \beta )</th>
<th>( \lambda )</th>
<th>( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>54</td>
<td>Bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>54</td>
<td>Transformation by ( F_{n-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>54</td>
<td>( W_{n}Y_{n-1} ) omitted; transformation by ( F_{n-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>54</td>
<td>( W_{n}Y_{n-1} ) omitted; transformation by ( F_{n-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>54</td>
<td>( W_{n}Y_{n-1} ) omitted; transformation by ( F_{n-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>54</td>
<td>Both ( W_{n}Y_{n-1} ) and ( W_{n}Y_{n-1} ) omitted; transformation by ( F_{n-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7)</td>
<td>54</td>
<td>( W_{n}Y_{n-1} ) omitted; transformation by ( F_{n-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8)</td>
<td>54</td>
<td>( W_{n}Y_{n-1} ) omitted; transformation by ( F_{n-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9)</td>
<td>54</td>
<td>Transformation by ( W_{n} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: \( \theta_{0} = \{0.4, 0.2, 1, 0.3, 1\} \) where \( \gamma_0 + \rho_0 + \lambda_0 = 0.9 \).
Table 8
Estimation results for the cigarettes demand.

<table>
<thead>
<tr>
<th>Estimates and t-statistics</th>
<th>Direct</th>
<th>1n</th>
<th>1n - W</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ (InC, t−1)</td>
<td>0.8651</td>
<td>[0.672425]</td>
<td>0.8643</td>
</tr>
<tr>
<td>ρ (Wn InC, t−1)</td>
<td>−0.0145</td>
<td>[−0.0336]</td>
<td>−0.0177</td>
</tr>
<tr>
<td>β₁ (InPn)</td>
<td>−0.2619</td>
<td>[−10.6646]</td>
<td>−0.2621</td>
</tr>
<tr>
<td>β₃ (InYn)</td>
<td>0.0997</td>
<td>[3.3481]</td>
<td>0.0994</td>
</tr>
<tr>
<td>φ₀ (Wn InPn)</td>
<td>0.0873</td>
<td>[0.2000]</td>
<td>0.0674</td>
</tr>
<tr>
<td>φ₄ (Wn InYn)</td>
<td>−0.0256</td>
<td>[−0.6443]</td>
<td>−0.0228</td>
</tr>
<tr>
<td>λ (Wn InCn)</td>
<td>−0.0220</td>
<td>[−0.4362]</td>
<td>−0.0240</td>
</tr>
<tr>
<td>Tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ = −λγ (X₄₀ = 3.8)</td>
<td>5.8042</td>
<td>5.6227</td>
<td>0.2634</td>
</tr>
<tr>
<td>φ = −λφ (X₄₀ = 7.8)</td>
<td>8.9087</td>
<td>9.2183</td>
<td>8.8058</td>
</tr>
<tr>
<td>Joint above (X₄₀ = 9.4)</td>
<td>10.5028</td>
<td>10.6685</td>
<td>8.8228</td>
</tr>
</tbody>
</table>

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

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

![Graph](image)

Table 10
SDPD models, August prices, W₀ = exp(−1.2D₀) with row-normalization.

<table>
<thead>
<tr>
<th>Models</th>
<th>I</th>
<th>II (a)</th>
<th>II (b)</th>
<th>II (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before bias correction estimates</td>
<td>Yₙ₋₀</td>
<td>0.5279</td>
<td>0.5276</td>
<td>0.5272</td>
</tr>
<tr>
<td></td>
<td>W₀Yₙ₋₀</td>
<td>−0.4112</td>
<td>−0.3708</td>
<td>−0.3958</td>
</tr>
<tr>
<td></td>
<td>W₀Yₙ₋₀</td>
<td>0.8520</td>
<td>0.7960</td>
<td>0.8359</td>
</tr>
<tr>
<td></td>
<td>σ²</td>
<td>0.0044</td>
<td>0.0044</td>
<td>0.0044</td>
</tr>
<tr>
<td>Tests (Wald χ² statistics)</td>
<td>ρ = −γλ</td>
<td>19.9968</td>
<td>15.1323</td>
<td>12.8392</td>
</tr>
<tr>
<td></td>
<td>ρ + γ + λ = 1</td>
<td>13.5655</td>
<td>13.7998</td>
<td>6.5918</td>
</tr>
<tr>
<td>Value of</td>
<td>0.9686</td>
<td>0.9528</td>
<td>0.9673</td>
<td>0.9783</td>
</tr>
<tr>
<td>InL</td>
<td>10.164</td>
<td>10.142</td>
<td>10.199</td>
<td>10.198</td>
</tr>
</tbody>
</table>

Table 11
SDPD models, February and August Prices, W₀ = exp(−1.2D₀) with row-normalization.

<table>
<thead>
<tr>
<th>Models</th>
<th>I</th>
<th>II (a)</th>
<th>II (b)</th>
<th>II (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Bias Correction Estimates</td>
<td>Yₙ₋₀</td>
<td>0.6646</td>
<td>0.6637</td>
<td>0.6634</td>
</tr>
<tr>
<td></td>
<td>W₀Yₙ₋₀</td>
<td>−0.5138</td>
<td>−0.4651</td>
<td>−0.4998</td>
</tr>
<tr>
<td></td>
<td>W₀Yₑ</td>
<td>0.8240</td>
<td>0.7730</td>
<td>0.8180</td>
</tr>
<tr>
<td></td>
<td>σ²</td>
<td>0.0036</td>
<td>0.0036</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>ρ + γ + λ = 1</td>
<td>23.3447</td>
<td>13.1849</td>
<td>5.5853</td>
</tr>
<tr>
<td>Value of</td>
<td>0.9748</td>
<td>0.9716</td>
<td>0.9816</td>
<td>0.9893</td>
</tr>
<tr>
<td>InL</td>
<td>21.985</td>
<td>21.944</td>
<td>22.032</td>
<td>22.005</td>
</tr>
</tbody>
</table>

Table 12
SDPD models, February and August Prices, W₀ = exp(−1.2D₀) with row-normalization.

<table>
<thead>
<tr>
<th>Models</th>
<th>I</th>
<th>II (a)</th>
<th>II (b)</th>
<th>II (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Bias Correction Estimates</td>
<td>Yₙ₋₀</td>
<td>0.6804</td>
<td>0.6793</td>
<td>0.6790</td>
</tr>
<tr>
<td></td>
<td>W₀Yₙ₋₀</td>
<td>−0.5273</td>
<td>−0.5518</td>
<td>−0.5121</td>
</tr>
<tr>
<td></td>
<td>W₀Yₑ</td>
<td>0.8240</td>
<td>0.8055</td>
<td>0.8181</td>
</tr>
<tr>
<td></td>
<td>σ²</td>
<td>0.0036</td>
<td>0.0036</td>
<td>0.0036</td>
</tr>
<tr>
<td>Tests (Wald χ² statistics)</td>
<td>ρ = −γλ</td>
<td>38.2825</td>
<td>32.7226</td>
<td>30.0950</td>
</tr>
<tr>
<td></td>
<td>ρ + γ + λ = 1</td>
<td>19.2219</td>
<td>4.5782</td>
<td>3.6844</td>
</tr>
<tr>
<td>Value of</td>
<td>0.9771</td>
<td>0.9830</td>
<td>0.9850</td>
<td>0.9924</td>
</tr>
</tbody>
</table>

Note: The numbers in the (·) are the standard deviations.
Table 12
SDPD models, February and August Prices, \( w_{ij} = \exp(\theta_{ij}d_{ij}) \) with row-normalization.

<table>
<thead>
<tr>
<th>Models</th>
<th>( \theta_{ij} = -0.7 )</th>
<th>( \theta_{ij} = -1.4 )</th>
<th>( \theta_{ij} = -2.8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_{it} )</td>
<td>( Y_{it} )</td>
<td>( Y_{it} )</td>
<td>( Y_{it} )</td>
</tr>
<tr>
<td>( W_{it}Y_{it} - 1 )</td>
<td>( -0.0646 (0.0151) )</td>
<td>( -0.5859 (0.0474) )</td>
<td>( -0.4705 (0.0125) )</td>
</tr>
<tr>
<td>( W_{it}Y_{it} )</td>
<td>( 0.7970 (0.0081) )</td>
<td>( 1.0000 (0.0552) )</td>
<td>( 0.7750 (0.0094) )</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>( 0.0038 (0.0001) )</td>
<td>( 0.0038 (0.0002) )</td>
<td>( 0.0035 (0.0001) )</td>
</tr>
</tbody>
</table>

**Tests (Wald \( \chi^2 \) statistics)**

- \( \rho = -\lambda \gamma \)
  - \( \rho = -\lambda \gamma = 1 \)
  - Value of \( \rho + \gamma + \lambda \)
  - Int.

| \( Y_{it} \) | \( 0.6602 \) | \( 24.8034 \) | \( 32.9799 \) | \( 21.9052 \) | \( 56.524 \) | \( 18.7204 \) |
| \( W_{it}Y_{it} - 1 \) | \( 15.7321 \) | \( 19.1677 \) | \( 19.5359 \) | \( 7.4900 \) | \( 204.27 \) | \( 128.8964 \) |
| \( W_{it}Y_{it} \) | \( 1.0391 \) | \( 1.0810 \) | \( 0.9671 \) | \( 0.9664 \) | \( 0.9058 \) | \( 0.8592 \) |
| \( \sigma^2 \) | \( 21.823 \) | \( 21.854 \) | \( 22.021 \) | \( 21.994 \) | \( 21.640 \) | \( 21.407 \) |

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

<table>
<thead>
<tr>
<th>Models</th>
<th>( W_{ij}^{(1)} )</th>
<th>( W_{ij}^{(2)} )</th>
<th>( W_{ij}^{(3)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_{it} )</td>
<td>( Y_{it} )</td>
<td>( Y_{it} )</td>
<td>( Y_{it} )</td>
</tr>
<tr>
<td>( W_{it}Y_{it} - 1 )</td>
<td>( 0.6538 (0.0067) )</td>
<td>( 0.6561 (0.0067) )</td>
<td>( 0.7062 (0.0063) )</td>
</tr>
<tr>
<td>( W_{it}Y_{it} )</td>
<td>( -0.4958 (0.0013) )</td>
<td>( -0.4859 (0.00332) )</td>
<td>( -0.6542 (0.00211) )</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>( 0.8140 (0.0103) )</td>
<td>( 0.8160 (0.0394) )</td>
<td>( 0.7970 (0.0162) )</td>
</tr>
</tbody>
</table>

**Tests (Wald \( \chi^2 \) statistics)**

- \( \rho = -\lambda \gamma \)
  - \( \rho = -\lambda \gamma = 1 \)
  - Value of \( \rho + \gamma + \lambda \)
  - Int.

| \( Y_{it} \) | \( 29.3264 \) | \( 23.6320 \) | \( 4.3833 \) | \( 3.9623 \) | \( 12.3065 \) | \( 9.2139 \) |
| \( W_{it}Y_{it} - 1 \) | \( 3.5519 \) | \( 0.1217 \) | \( 2.1682 \) | \( 1.6010 \) | \( 128.8964 \) | \( 116.9106 \) |
| \( W_{it}Y_{it} \) | \( 1.0391 \) | \( 1.0810 \) | \( 0.9671 \) | \( 0.9664 \) | \( 0.9058 \) | \( 0.8592 \) |
| \( \sigma^2 \) | \( 332.8166 \) | \( 23.6530 \) | \( 32.8950 \) | \( 21.8267 \) | \( 54.713 \) | \( 18.6606 \) |

**Acknowledgements**

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

**References**


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


Please cite this article as: Lee, L., Yu, J., Some recent developments in spatial panel data models, Regional Science and Urban Economics xxx (2009) xxx–xxx