Regional Science and Urban Economics 51 (2015) 37–46

Contents lists available at ScienceDirect

Regional Science and Urban Economics

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

Estimation of partially specified dynamic spatial panel data models with fixed-effects☆

Yuanqing Zhang [a,b,*], Yanqing Sun [c]

[a] School of Finance and Business, Shanghai Normal University, No. 100 Guilin Rd., Shanghai 200234, PR China
[b] Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, PR China
[c] Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

ARTICLE INFO

Article history:
Received 1 September 2014
Received in revised form 2 January 2015
Accepted 14 January 2015
Available online 23 January 2015

Keywords:
Spatial
Panel data
Partially linear
Dynamic

ABSTRACT

This paper studies estimation of a partially specified spatial dynamic panel data regression with fixed-effects. Under the assumption of strictly exogenous regressors and strictly exogenous spatial weighting matrix, the model is estimated by 2SLS method aided by the sieve method and through the instrumental variable. Under some sufficient conditions, the proposed estimator for the finite dimensional parameter is shown to be root-N consistent and asymptotically normally distributed. The proposed estimator for the unknown function is shown to be consistent and asymptotically distributed as well, though at a rate slower than root-N. Consistent estimators for the asymptotic variance–covariance matrices of both estimators are provided. The results can be generalized to several spatial weighting matrices and spatial matrix which vary with time. The simulation results suggest that the proposed approach has some practical value.

1. Introduction

Dynamic panel data can not only capture dynamics of economic activities but also enable researchers to control unobservable heterogeneity across units. So, the spatial econometrics literature has exhibited a growing interest in the specification and estimation of dynamic regression equations based on spatial panels in the last decade. Both Kukenova and Monteiro (2009) consider a dynamic spatial panel data model and find that the system GMM estimator substantially reduces the bias in the parameter estimate of the WY_t variable, and that the system GMM estimator outperforms the Arrelano and Bond difference GMM estimator. Yu et al. (2008) study the quasi maximum likelihood (QML) estimation for spatial dynamic panel data models. Korniotis (2010) constructs a bias-corrected LSDV estimator for a dynamic panel data model in the fixed effects setting. Lee and Yu (2010c) propose an optimal GMM estimator based on linear moment conditions, which are standard, and quadratic moment conditions and prove that this GMM estimator is consistent, also when T is small relative to N. These spatial panel data models have many applications, for instance, transportation research (Parent and LeSage, 2011), growth convergence of countries and regions (Baltagi et al., 2007; Ertur and Koch, 2007), regional markets (Keller and Shiue, 2007), labor economics (Foote, 2007), and economic growth (Elhorst, 2010), to name a few.

All of aforementioned studies specify a parametric regression function. Parametric regression models provide a parsimonious description of the relationship between the explained variables and the explanatory variables. However, a badly misspecified one may have large model approximation error leading to large modeling bias in practice. So an intermediate strategy is to apply a semiparametric form, among which partially linear models are widely used. Example studies include Cuzick (1992), Severini and Wong (1992), and Liang et al. (1999), among others. More methods and references can be found in the monograph by Härdle et al. (2000). On the application side, researchers have started addressing the importance of semiparametric modeling in spatial econometrics recently. For example, Basile and Gress (2005) propose a semiparametric spatial auto-covariance specification of the growth model for the European economy and find that nonlinearities are important in regional growth in Europe even when the spatial dependence is controlled. As a result, assuming a common linear relationship between economic growth and inputs is misleading.

To follow this line of literature, we wish to specify the following partially specified dynamic spatial panel regression (hereafter PSDSPR):

$$y_{it} = \gamma_0 y_{i,t-1} + \lambda_0 \sum_{j=1}^{N} w_{ij} y_{jt} + x_{it}'\beta_0 + g_0(z_{it}) + c_i + \varepsilon_{it}, \; i = 1, \dots, N, \;\; t = 2, \dots, T, \tag{1}$$

☆ Work of Zhang is supported by the National Natural Science Foundation of China (Grant Nos. 71371118, 71471117), National Social Science Fund of China (Grant No. 14BJY012), Program for Changjiang Scholars and Innovative Research Team in University (PCSIRTIRT13077), State Key Program of National Natural Science of China (Grant No. 71331006), Social Science Research Fund from Ministry of Education of China (Grant No. 11YJAZH044). The research of Yanqing Sun was also partially supported by the National Institutes of Health NIAID [grant number R37 AI054165] and the National Science Foundation [grant number DMS-1208978].

* Corresponding author at: School of Finance and Business, Shanghai Normal University, No. 100 Guilin Rd., Shanghai 200234, PR China.

http://dx.doi.org/10.1016/j.regsciurbeco.2015.01.003

where y_{it} denotes the dependent variable of individual i in period t, $x_{it1} = (x_{it1}, x_{it2}..., x_{itq})'$ and z_{it} denote explanatory variables, c_i denotes the unobserved and time invariant individual effect, w_{ij} denotes the spatial weight between individuals i and j, ε_{it} denotes random noise, and $\delta_0 = (\gamma_0, \lambda_0, \beta_0')'$ denotes the unknown true parameter value, where $g_0(.)$ is an unknown function.

The model (1) is a generalization of many usual parametric, semiparametric, panel data, spatial models. When $\gamma_0 \equiv 0$ and $T \equiv 1$, the model is the partially linear spatial autoregressive model study by Su and Jin (2010) and Su (2012). When $\gamma_0 \equiv 0$ and $\lambda_0 \equiv 0$, the model is the partially linear panel data model study by Li and Stengos (1996) and the book edited by Ai and Li (2008). When $\gamma_0 \equiv 0$, $g_0 \equiv 0$ and $T = 1$, the model is the spatial autoregressive model study by Kelejian and Prucha (1998, 1999). When $\gamma_0 \equiv 0$ and $g_0 \equiv 0$, the model is the spatial panel data model study which that Various versions of theirs have been studied extensively in a series papers, see the reviews in Anselin (2001), the works of Elhorst (2001, 2003), as well as the recent papers by Baltagi et al. (2003, 2006), Kapoor et al. (2007) and Lee and Yu (2010a,b), among others. When $\lambda_0 \equiv 0$ and $\gamma_0 \equiv 0$, the model is the dynamic panel data model, see Ahn and Schmidt (1995), for details. When $\gamma_0 \equiv 0$, the model is the partially linear spatial panel data model study by Ai and Zhang (2015). When $g_0 \equiv 0$, the model is the dynamic spatial panel data model study by Kukenova and Monteiro (2009), Korniotis (2010), Lee and Yu (2010c), and J. Paul Elhorst (2012). When $\lambda_0 \equiv 0$, the model is the partially linear dynamic panel data model study by Li and Kniesner (2002) and Park et al. (2007). Our estimation method proposed in the paper could be applied successfully to the above special cases.

To the best of our knowledge, this is the first work in which the estimating problem of modeling partially specified dynamic spatial panel data is investigated. The main objective of this paper is to propose an approach to estimate δ_0 and $g_0(.)$ consistently under the fixed-effects assumption which has the advantage of robustness in that the individual effects are allowed to correlate with included regressors in the model, to establish the asymptotic properties for the proposed estimators and to report on a simulation study.

The paper is organized as follows: Section 2 proposes an estimation of Eq. (1); Section 3 establishes the asymptotic properties of the proposed estimators; Section 4 presents some consistent covariance matrices; and Section 5 reports some Monte Carlo simulation results. All technical proofs are relegated to the appendices.

2. Estimation

Throughout the paper, we shall consider model (1) only for the case of small and fixed T but large N. The case of large T and large N will be dealt with in a separate study. There are three issues with estimation of model (1). First is the potential correlation between the unobserved individual effects c_i and the other explanatory variables $\{(x_{it}, z_{it}), t = 2, ..., T\}$ (i.e., the fixed-effects). Second is the correlation between the spatial term $\sum_{j=1}^{N} w_{ij}y_{jt}$ and the error term ε_{it} (i.e., endogeneity). Third is the infinite dimensionality of the unknown parameter $g_0(z_{it})$. We tackle the infinite dimensionality problem by applying the sieve method (see Ai and Chen, 2003). Specifically, let $p^K(z_{it}) = (p_1(z_{it}), ..., p^K(z_{it}))'$ denote a sequence of known basis functions that can approximate any measurable function arbitrarily well in the functional space endowed by a norm $||\cdot||_s$. Examples of sieves include polynomials, Fourier series, splines, and wavelets. For each $K = K(N)$, there exists some constant $\pi_0 = \pi_{0,k} = \pi_{0,K(N)}$, such that $g_0(z) = p^{K(N)}(z)'\pi_{0,K(N)} + v_0(z)$ with $v_0(z)$ as the approximation error satisfying $||v_0(z)||_s \to 0$ as $K \to \infty$. For simplicity, we denote $g_0(z) = p^K(z)'\pi_0 + v_0(z)$. Denote $v_{0it} = v_0(z_{it})$. We rewrite the model as

$$y_{it} = \gamma_0 y_{i,t-1} + \lambda_0 \sum_{j=1}^{N} w_{ij} y_{jt} + x_{it}'\beta_0 + p^K(z_{it})'\pi_0 + c_i + \varepsilon_{it}$$
$$+v_{0it}, \ i = 1, ..., N, \ \ t = 2, 3, ..., T.$$

Let Δ denote the first difference operator and denote $\Delta p_{it} = p^K(z_{it}) - p^K(z_{it-1})$. We tackle the fixed-effects issue by applying the first difference

$$\Delta y_{it} = \gamma_0 \Delta y_{i,t-1} + \lambda_0 \sum_{j=1}^{N} w_{ij} \Delta y_{jt} + \Delta x_{it}'\beta_0 + \Delta p_{it}'\pi_0 + \Delta\varepsilon_{it}$$
$$+\Delta v_{0it}, \ i = 1, ..., N, \ \ t = 3, 4, ..., T.$$

To simplify notation and to aid exposition, we stack the data in two ways and use whichever is convenient. One way is to stack data first by individual then by time. Specifically, define:

$$\Delta X_t = (\Delta x_{1t}, \Delta x_{2t}, ..., \Delta x_{Nt})' \ \text{ and } \ \Delta X = (\Delta X_3', \Delta X_4', ..., \Delta X_T')'.$$

Define ΔY_t, ΔY, ΔP_t, ΔP, $\Delta\varepsilon_t$, $\Delta\varepsilon$, ΔV_{0t}, and ΔV_0 analogously using Δy_{it}, Δp_{it}, $\Delta\varepsilon_{it}$, and Δv_{0it} respectively. Let $W = (w_{ij})$ denote the weighting matrix and let $D_w = diag\{W, ..., W\}$ denote the block diagonal matrix. With $\delta_0 = (\gamma_0, \lambda_0, \beta_0')'$, we write

$$\Delta Y_t = \gamma_0 \Delta Y_{t-1} + \lambda_0 W \Delta Y_t + \Delta X_t \beta_0 + \Delta P_t \pi_0 + \Delta\varepsilon_t + \Delta V_{0t}$$

or

$$\begin{aligned}\Delta Y &= \gamma_0 \Delta Y_{-1} + \lambda_0 D_w \Delta Y + \Delta X \beta_0 + \Delta P \pi_0 + \Delta\varepsilon + \Delta V_0 \\ &= \Delta B \delta_0 + \Delta P \pi_0 + \Delta\varepsilon + \Delta V_0,\end{aligned}$$

where

$$\Delta B = (\Delta B_1', \Delta B_2', ..., \Delta B_T')' \ \text{ and } \ \Delta B_t = (\Delta Y_{t-1}, W\Delta Y_t, \Delta X_t).$$

The second way of stacking data is to stack data first by time then by individual. We use a bar over the variable to distinguish this stacking from the first one. Specifically, define

$$\Delta\overline{X}_i = (\Delta x_{i2}, \Delta x_{i3}, ..., \Delta x_{iT}) \ \text{ and } \ \Delta\overline{X} = (\Delta\overline{X}_1', \Delta\overline{X}_2', ..., \Delta\overline{X}_N')'.$$

Define $\Delta\overline{Y}_i$, $\Delta\overline{Y}$, $\Delta\overline{P}_i$, $\Delta\overline{P}$, $\Delta\overline{B}_i$, $\Delta\overline{B}$, $s\Delta\overline{\varepsilon}_i$, and $\Delta\overline{\varepsilon}$ analogously.

Let $S = \Delta P(\Delta P'\Delta P)^{-1}\Delta P'$ denote the projection matrix onto the space spanned by ΔP. Define $\overline{S}$ using $\Delta\overline{P}$. Partialing out the sieve approximation, we obtain

$$(I-S)\Delta Y = (I-S)\Delta B\delta_0 + (I-S)\Delta\varepsilon + (I-S)\Delta V_0.$$

Finally, we tackle the endogeneity issue by applying the 2SLS. Let h_i denote a column vector of instrumental variables

$$h_i = \begin{pmatrix} [h_{i1}] & & & & 0 \\ & [h_{i2}] & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & \cdot & [h_{i,T-2}] \end{pmatrix}$$

where $h_{i1} = (y_{i1}, \sum_{j=1}^{N} w_{ij}x_{j1}, x_{i1}', x_{i2}')$; $h_{i2} = (y_{i1}, y_{i,2}, \sum_{j=1}^{N} w_{ij}x_{j1}, \sum_{j=1}^{N} w_{ij}x_{j,2}, x_{i1}', x_{i2}', x_{i,3}')$; and $h_{i,T-2} = (y_{i1}, ..., y_{i,T-2}, \sum_{j=1}^{N} w_{ij}x_{j1}, ..., \sum_{j=1}^{N} w_{ij}x_{j,T-2}, x_{i1}', ..., x_{i,T-1}')$. By computing we could obtain that the dimension of h_i is $(T-1) \times \frac{(T-2)(2qT+T-1)}{2}$ where q denotes the dimension of x_{it}.

Define $H = (H'_1, ..., H'_N)'$ and $M = H(H'H)^{-1}H'$. Define $\overline{H}_i$, $\overline{H}$, and $\overline{M}$ analogously. The 2SLS estimators for the unknown parameters δ_0, π_0, and $g_0(z)$ are respectively given by

$$\begin{aligned}\hat{\delta} &= \{\Delta B'(I-S)M(I-S)\Delta B\}^{-1}\Delta B'(I-S)M(I-S)\Delta Y \\ &= \left\{\Delta\overline{B}'\left(I-\overline{S}\right)\overline{M}\left(I-\overline{S}\right)\Delta\overline{B}\right\}^{-1}\Delta\overline{B}'\left(I-\overline{S}\right)\overline{M}\left(I-\overline{S}\right)\Delta\overline{Y};\end{aligned} \tag{2}$$