The Econometrics Journal

Econometrics Journal (2016), volume **19**, pp. 261–290.
doi: 10.1111/ectj.12069

Instrumental variable estimation of a spatial dynamic panel model with endogenous spatial weights when *T* is small

XI QU[†], XIAOLIANG WANG[†] AND LUNG-FEI LEE[‡]

[†]*Antai College of Economics and Management, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai, 200030, China.*
E-mail: xiqu@sjtu.edu.cn, wangxiaoliang2009@gmail.com

[‡]*Department of Economics, The Ohio State University, 410 Arps Hall, 1945 N. High St., Columbus, OH 43210, USA.*
E-mail: lee.1777@osu.edu

First version received: September 2015; final version accepted: July 2016

Summary The spatial dynamic panel data (SDPD) model is a standard tool for analysing data with both spatial correlation and dynamic dependences among economic units. Conventional estimation methods rely on the key assumption that the spatial weight matrix is exogenous, which would likely be violated in some empirical applications where spatial weights are determined by economic factors. In this paper, we propose an SDPD model with individual fixed effects in a short time dimension, where the spatial weights can be endogenous and time-varying. We establish the consistency and asymptotic normality of the two-stage instrumental variable (2SIV) estimator and we investigate its finite sample properties using a Monte Carlo simulation. When applying this model to study government expenditures in China, we find strong evidence of spatial correlation and time dependence in making spending decisions among China's provincial governments.

Keywords: *Endogenous spatial weights, Fixed effects, Public Expenditure, Spatial correlation, Spatial dynamic panel data model.*

1. INTRODUCTION

Spatial dynamic panel data (SDPD) models are techniques dealing with the spatial correlation and dynamic dependences of economic units. They are generalized from a cross-sectional spatial autoregressive (SAR) model proposed by Cliff and Ord (1973). Recently, there has been much progress in empirical and theoretical works on spatial panel data models. Static spatial panel data models can be applied to agricultural economics (Druska and Horrace, 2004), transportation research (Frazier and Kockelman, 2005), public economics (Egger et al., 2005), consumer demand (Baltagi and Li, 2006), to name a few. SDPD models can be applied to the growth convergence of countries and regions (Ertur and Koch, 2007), regional markets (Keller and Shiue, 2007), labour economics (Foote, 2007), public economics (Revelli, 2001, Brueckner, 2003, Tao, 2005, and Franzese and Hays, 2007), and some other fields. For the estimation and statistical inference, random effects and fixed effects spatial panel models are most commonly used. For the random effects model, Baltagi et al. (2003, 2007a,b), Mutl (2006) and Kapoor et al. (2007) investigate various specifications with error components. For the fixed effects model, Elhorst (2005), Korniotis (2010), Su and Yang (2015), Yu et al. (2008, 2012) and Lee and Yu (2010) study static or dynamic models under various spatial structures with the fixed effects specification. Mutl and Pfaffermayr (2010) and Lee and Yu (2012a) consider the estimation of spatial panel data models with both fixed and random effects specifications, and they propose Hausman-type specification tests. Lee and Yu (2012b) extend Yu et al. (2008) to allow for a time-varying spatial weights matrix. In these studies, the SDPD model with fixed effects gains increasing interest because it takes into account both the dynamic and spatial dependence and controls for the unobserved heterogeneity. Some related surveys can be found in Anselin et al. (2008), Elhorst (2010), Baltagi (2011) and Lee and Yu (2015).

In these spatial panel data models, the spatial weight matrices are usually specified to be exogenous and time invariant. This is plausible if spatial weight matrices are based on contiguity or geographic distances among regions. However, there are plenty of cases in which spatial weights are constructed using economic/socioeconomic distances. For example, Case et al. (1993) construct weights based on the difference in the percentage of the population that is black in order to study the state spending. Brueckner (1998) and Brueckner and Saavedra (2001) use weights according to population sizes in studies of local governments' competitions on economic policies. Moreover, Aiello and Cardamone (2008) construct their spatial weights using a variable that reflects firms' technological similarity and geographical proximity in order to study an R&D spillover in Italy. In Crabbé and Vandenbussche (2013), in addition to the physical distance, spatial weight matrices are constructed by inverse trade share and inverse distance between GDP per capita.

When elements of the spatial weights are constructed from economic/socioeconomic characteristics of regions (or districts), there can be some unobserved factors that affect both the final outcome and these economic characteristics, causing the spatial weights to be endogenous. However, most research still assumes that these weights are exogenous, with very few exceptions. Kelejian and Piras (2014) propose a two-stage least-squares (2SLS) estimation for a short spatial panel data model with endogenous regressors and endogenous weights, utilizing some high-level assumptions on the existence of relevant statistics and asymptotic convergence. Qu and Lee (2015) study a cross-sectional SAR model with endogenous spatial weights and find that ignoring the endogeneity in spatial weights matrices would have substantial consequences on estimates. In a panel setting, spatial weights can change over time if they are constructed from economic/socioeconomics characteristics. Also, with fixed effects to control for heterogeneity, one may wonder whether ignoring the endogeneity in spatial weights would have severe consequences in the panel setting, and whether the SDPD model with endogenous spatial weights can be easily handled and estimated. These motivate our investigation on the SDPD model with endogenous spatial weights.

Our SDPD model is in the setting of large n and small T, which fits the study of short panels. In a separate working paper, Qu et al. (2015) study the quasi-maximum likelihood estimation (QMLE) of the SDPD model with endogenous spatial weights when the time dimension T is large. Compared to the small T setting in this paper, their spatial weights have time dependences, so the asymptotic analysis is to explore the spatial-time near epoch dependence (NED) and it is much more complicated. They need to impose stronger restrictions on spatial weights because of the spatial-time dependence. In their large T setting, spatial weights matrices need to be sparse, i.e., the spatial weight $w_{ij} = 0$ if the physical distance between i and j is large. In this paper, we consider the large n and small T setting because many panel datasets available for empirical analyses have that feature. Also, more flexibly specified spatial weights matrices can be handled. In the small T setting, spatial weights can be distance decaying, i.e., w_{ij} is a non-negative decaying function of the physical distance between i and j. We focus on the 2SIV estimation method because it has a closed form expression and can be easily implemented. The asymptotic analysis is based on the spatial NED property similar to Qu and Lee (2015), with the additional complication coming from time-varying and time-differencing spatial weights due to the removal of individual fixed effects.

From the empirical aspect, we apply the 2SIV estimation of the SDPD model with endogenous spatial weights to study government expenditures in China. Spatial econometric models have valuable applications in the study of strategic interactions in public economics. See Brueckner (2003) for a review of empirical studies using cross-sectional spatial models. In the spatial panel framework, Redoano (2007) examines the fiscal interactive behaviour among European countries; Moscone et al. (2007) investigate determinants of local authority mental health expenditure in England; Foucault et al. (2008) study the public spending interactions among French municipalities; Yu et al. (2013) study the strategic interaction and the determinants of public health expenditures in China using a spatial Durbin panel model, to name a few. However, all of these studies are based on the restrictive assumption that the spatial weights matrix is exogenous, which might lead to a biased estimator. Using China's provincial expenditure data from 2007 to 2013, we find some solid evidence to support the endogeneity of spatial weights constructed from GDP, and our results indicate both significant spatial correlation and time dependence when China's provincial governments make spending decisions.

This paper is organized as follows. In Section 2, we introduce the SDPD model and specify the source of endogeneity in spatial weights. In Section 3, we establish the 2SIV estimation method and prove its asymptotic properties. Finite sample performances of our proposed 2SIV method are investigated using a Monte Carlo simulation in Section 4. In Section 5, we apply this model to study the spatial correlation and time dependences in making spending decisions among China's provincial governments. Section 6 concludes the paper. Some lemmata and proofs are collected in the Appendix.

2. THE MODEL

2.1. Model specification

Following Jenish and Prucha (2009, 2012), we consider spatial processes located on a (possibly) unevenly spaced lattice $D \subseteq R^d$, $d \geq 1$. The asymptotic methods we employ are increasing domain asymptotics: growth of the sample is ensured by an unbounded expansion of the sample region, as in Jenish and Prucha (2012).

Let $\{(\varepsilon'_{i,nt}, v_{i,nt}); i \in D_n, n \in N, t = 1, \ldots, T\}$ be a triangular double array of real random vectors defined on a probability space $(\Omega; F; P)$, where the index set $D_n \subset D$ is a finite set. In this paper, we consider an SDPD model with the specification,

$$Y_{nt} = \lambda_1 W_{nt} Y_{nt} + \rho_1 Y_{n,t-1} + \lambda_2 W_{n,t-1} Y_{n,t-1} + X_{1nt}\beta + \mathbf{c}_n + V_{nt}, \tag{2.1}$$

where $Y_{nt} = (y_{1,nt}, y_{2,nt}, \ldots, y_{n,nt})'$ and $V_{nt} = (v_{1,nt}, v_{2,nt}, \ldots, v_{n,nt})'$ are n-dimensional column vectors, and $v_{i,nt}$ are independent and identically distributed (i.i.d.) disturbances across individuals and time with zero mean and variance σ_v^2. The $X_{1nt} = (x_{1,nt}, x_{2,nt}, \ldots, x_{n,nt})'$ is an $n \times k_1$ matrix of individually and time-varying non-stochastic regressors. β is a k_1-dimensional vector of coefficients, and λ_1, λ_2 and ρ_1 are scalar coefficients. $\mathbf{c}_n$ is an n-dimensional vector of the individual fixed effects. The spatial weight matrix W_{nt} is an $n \times n$ matrix with each entry constructed by $(W_{nt})_{ij} = w_{ij,nt} = h_{ijt}(z_{i,nt}, z_{j,nt})$, where $z_{i,nt}$ is a p-dimensional vector. For any $i = 1, \ldots, n$, $z_{i,nt}$ satisfies the equation

$$z'_{i,nt} = z'_{i,n,t-1}\boldsymbol{\rho}_2 + x'_{i,2nt}\Gamma + d'_{i,n} + \varepsilon'_{i,nt},$$

where $\boldsymbol{\rho}_2$ is a $p \times p$ constant matrix, $x_{i,2nt}$ is a k_2-dimensional vector of individually and time-varying non-stochastic regressors, Γ is a $k_2 \times p$ matrix of coefficients, $d_{i,n}$ is a p-dimensional constant vector invariant over time, and $\varepsilon_{i,nt}$ is a p-dimensional disturbance term. Denote the $n \times k_2$ matrix $X_{2nt} = (x_{1,2nt}, \ldots, x_{n,2nt})'$, $n \times p$ matrices $Z_{nt} = (z_{1,nt}, \ldots, z_{n,nt})'$, $\mathbf{d}_n = (d_{1,n}, \ldots, d_{n,n})'$ and $\varepsilon_{nt} = (\varepsilon_{1,nt}, \ldots, \varepsilon_{n,nt})'$. Then, the equation can be written in matrix form as

$$Z_{nt} = Z_{n,t-1}\boldsymbol{\rho}_2 + X_{2nt}\Gamma + \mathbf{d}_n + \varepsilon_{nt}. \tag{2.2}$$

2.2. Model interpretation and source of endogeneity

We consider n agents in an area, each endowed with a predetermined location i. Because of competition or spillover effects, at period t, each agent i has an outcome $y_{i,nt}$ directly affected by its neighbours' current outcomes $y_{j,nt}$; because of the time dependence, $y_{i,nt}$ is also directly affected by its own outcome from last period $y_{i,n,t-1}$, and its neighbours' outcomes from the last period $y_{j,n,t-1}$. We refer to the first term $W_{nt}Y_{nt}$ on the right-hand side of equation (2.1) as the spatial lag, the second term $Y_{n,t-1}$ as the time lag and the third term $W_{n,t-1}Y_{n,t-1}$ as the spatial-time lag (or diffusion).

The spatial weight $w_{ij,nt}$ is a measure of relative strength of linkage between agents i and j at time t. However, this weight $w_{ij,nt}$ is not predetermined but depends on some observable variable Z_{nt}. We can think of $z_{i,nt}$ as some economic variables at location i and time t such as GDP, consumption, economic growth rate, etc., which influence the strength of links across units. Moreover, we make the following assumptions. For any random vector Y, $||Y||_p = [E[|Y|^p]]^{1/p}$ denotes its L_p-norm where $|Y|$ is the Euclidean norm of Y.

ASSUMPTION 2.1. *The lattice $D \subset R^{d_0}$, $d_0 \geq 1$, is infinitely countable. All elements in D are located at distances of at least $\rho_0 > 0$ from each other, i.e., $\forall i, j \in D : \rho_{ij} \geq \rho_0$, where ρ_{ij} is the distance between locations i and j; without loss of generality (w.l.o.g.), we assume that $\rho_0 = 1$.*

ASSUMPTION 2.2. *The error terms $v_{i,nt}$ and $\varepsilon_{i,nt}$ have a joint distribution, $(v_{i,nt}, \varepsilon'_{i,nt})' \sim i.i.d.(0, \Sigma_{v\varepsilon})$, where*

$$\Sigma_{v\varepsilon} = \begin{pmatrix} \sigma_v^2 & \sigma'_{v\varepsilon} \\ \sigma_{v\epsilon} & \Sigma_\varepsilon \end{pmatrix}$$

is positive definite, σ_v^2 is a scalar variance, covariance $\sigma_{v\varepsilon} = (\sigma_{v\varepsilon_1}, \ldots \sigma_{v\varepsilon_{p_2}})'$ is a p-dimensional vector and Σ_ε is a $p \times p$ matrix. $\sup_{i,n,t} ||v_{i,nt}||_{4+\delta_\epsilon}$ and $\sup_{i,n,t} ||\epsilon_{i,nt}||_{4+\delta_\epsilon}$ exist for some $\delta_\varepsilon > 0$. Furthermore, $E[v_{i,nt}|\varepsilon_{i,nt}] = \varepsilon'_{i,nt}\delta$ and $Var(v_{i,nt}|\varepsilon_{i,nt}) = \sigma_\xi^2$.

This is the same setting as Qu and Lee (2015). Assumption 2.1 is for increasing domain asymptotic. The endogeneity of W_{nt} comes from the correlation between $v_{i,nt}$ and $\varepsilon_{i,nt}$ in Assumption 2.2. If $\sigma_{v\varepsilon}$ is zero, the spatial weight matrix W_{nt} might be treated as exogenous and we can apply conventional methodologies of the SDPD models for estimation. However, if $\sigma_{v\varepsilon}$ is not zero, then W_{nt} becomes an endogenous spatial weights matrix. Then, a conventional instrumental variable (IV) method fails because $W_{nt}X_{nt}$ is no longer a valid IV for $W_{nt}Y_{nt}$ when W_{nt} is endogenous.

From the two conditional moments assumptions in Assumption 2.2, we have the p-dimensional vector $\delta = \Sigma_\varepsilon^{-1}\sigma_{v\varepsilon}$ and the scalar $\sigma_\xi^2 = \sigma_v^2 - \sigma_{v\varepsilon}'\Sigma_\varepsilon^{-1}\sigma_{v\varepsilon}$. Denote $\xi_{nt} = V_{nt} - \varepsilon_{nt}\delta$, then its conditional mean on ε_{nt} is zero and its conditional variance matrix is $\sigma_\xi^2 I_n$. In particular, ξ_{nt} is uncorrelated with all the regressors in the outcome equation (2.1). Then, (2.1) becomes

$$\begin{aligned} Y_{nt} &= \lambda_1 W_{nt}Y_{nt} + \rho_1 Y_{n,t-1} + \lambda_2 W_{n,t-1}Y_{n,t-1} + X_{1nt}\beta + \mathbf{c}_n \\ &\quad + (Z_{nt} - Z_{n,t-1}\boldsymbol{\rho}_2 - X_{2nt}\Gamma - \mathbf{d}_n)\delta + \xi_{nt}, \end{aligned} \tag{2.3}$$

where $E[\xi_{nt}|\varepsilon_{nt}] = 0$, $\mathrm{Var}(\xi_{nt}|\varepsilon_{nt}) = \sigma_\xi^2 I_n$, and $\xi_{i,nt}$ are i.i.d. across i and t. Our estimation will mainly rely on the differenced equation of (2.3), where the fixed effects cancel out and the differenced $(Z_{nt}, Z_{n,t-1}, X_{2nt})$ are control variables to control for the endogeneity of W_{nt}.

Notice that Assumption 2.2 is relatively general without imposing any specific distribution on disturbances, as it is based on only conditional moments restrictions. In the special case that $(v_{i,nt}, \varepsilon_{i,nt}')'$ has a jointly normal distribution, then $v_{i,nt}|\varepsilon_{i,nt} \sim$ i.i.d. $N(\varepsilon_{i,nt}'\Sigma_\varepsilon^{-1}\sigma_{v\varepsilon}, \sigma_v^2 - \sigma_{v\varepsilon}'\Sigma_\varepsilon^{-1}\sigma_{v\varepsilon})$ and ξ_{nt} is independent of ε_{nt} in (2.2).

3. ESTIMATION METHOD

3.1. 2SIV estimation

Because of the complication of the model, we use a two step procedure to estimate the parameter $(\lambda_1, \lambda_2, \rho_1, \beta', \delta')'$. For any $n \times m$ variable A_{nt}, denote $\Delta A_{nt} = A_{nt} - A_{n,t-1}$, and the $n(T-2) \times m$ matrices

$$\Delta A_n(T) = (\Delta A_{n,3}', \Delta A_{n,4}' \ldots, \Delta A_{n,T}')' \text{ and } \Delta A_n(T-1) = (\Delta A_{n,2}', \Delta A_{n,3}', \ldots, \Delta A_{n,T-1}')'.$$

In the first step, we estimate the differenced equation of (2.2)

$$\Delta Z_{nt} = \Delta Z_{n,t-1}\boldsymbol{\rho}_2 + \Delta X_{2nt}\Gamma + \Delta\varepsilon_{nt},$$

or in the matrix form

$$\Delta Z_n(T) = K_n\begin{pmatrix}\boldsymbol{\rho}_2\\ \Gamma\end{pmatrix} + \Delta\varepsilon_n(T), \text{ where } K_n = (\Delta Z_n(T-1), \Delta X_{2n}(T)).$$

A consistent estimator is derived from

$$\begin{pmatrix}\widehat{\boldsymbol{\rho}}_2\\ \widehat{\Gamma}\end{pmatrix} = (K_n' P_{1n} K_n)^{-1} K_n' P_{1n}\Delta Z_n(T),$$

where $P_{1n} = q_{1n}(q_{1n}'q_{1n})^{-1}q_{1n}'$ and $q_{1n} = (q_{1n}^*, \Delta X_{2n}(T))$ with q_{1n}^* being a valid instrument of $\Delta Z_n(T-1)$ in K_n.

Following Qu and Lee (2015), we use the control function approach to handle the case with the endogenous spatial weights matrix. Based on (2.3), we eliminate the fixed effect by first-order differencing. Let $\kappa = (\lambda_1, \lambda_2, \rho_1, \beta', \delta')'$ and

$$T_n = \big(\Delta(WY)_n(T), \Delta(WY)_n(T-1), \Delta Y_n(T-1), \Delta X_{1n}(T), \Delta\varepsilon_n(T)\big).$$

Then, in the second step, we want to estimate the following differenced equation:

$$\Delta Y_n(T) = \big(\Delta(WY)_n(T), \Delta(WY)_n(T-1), \Delta Y_n(T-1), \Delta X_{1n}(T), \Delta\varepsilon_n(T)\big)\kappa + \Delta\xi_n(T). \tag{3.1}$$

However, $\Delta\varepsilon_n(T)$ in T_n is unknown. To solve this, we use the regression residual from the first step $\Delta\widehat{\varepsilon}_n(T) = \Delta Z_n(T) - \Delta Z_n(T-1)\widehat{\boldsymbol{\rho}}_2 - \Delta X_{2n}(T)\widehat{\Gamma}$ as an approximation of $\Delta\varepsilon_n(T)$ and estimate κ by

$$\widehat{\kappa} = (\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\rho}_1, \widehat{\beta}', \widehat{\delta}')' = (\widehat{T}_n' P_{2n}\widehat{T}_n)^{-1}\widehat{T}_n' P_{2n}\Delta Y_n(T),$$

where

$$\widehat{T}_n = \big(\Delta(WY)_n(T), \Delta(WY)_n(T-1), \Delta Y_n(T-1), \Delta X_{1n}(T), \Delta\widehat{\varepsilon}_n(T)\big),$$

$P_{2n} = q_{2n}(q_{2n}'q_{2n})^{-1}q_{2n}'$ and $q_{2n} = (q_{2n}^*, \Delta X_{1n}(T), \Delta\widehat{\varepsilon}_n(T))$ with q_{2n}^* being a valid instrument of $[\Delta(WY)_n(T), \Delta(WY)_n(T-1), \Delta Y_n(T-1)]$ in $\widehat{T}_n$.

3.2. Asymptotic property

For the initial value of the stochastic process, we adopt a similar framework as Su and Yang (2015) where the starting position of data process can be either exogenous or endogenous.

(a) Data collection starts from the first period; the process starts from the $-t_0$th period, i.e., $t_0 + 1$ periods before the start of data collection, and then evolves according to the model specified by (2.1)and (2.2).

(b) The data starting period $0 \le t_0 < \infty$. The starting position of the process $Y_{n,-t_0}$ and $Z_{n,-t_0}$ can be treated as either exogenous or endogenous. If $Y_{n,-t_0}$ and $Z_{n,-t_0}$ are exogenous, then we treat them as constants; if $Y_{n,-t_0}$ and $Z_{n,-t_0}$ are endogenous, then they are assumed to have a cross-sectional specification

$$Y_{n,-t_0} = \lambda_1^o W_{n,-t_0} Y_{n,-t_0} + X_{1n,-t_0}\beta^o + \mathbf{c}_n^o + V_{n,-t_0}$$

with $Z_{n,-t_0} = X_{2n,-t_0}\Gamma^o + \mathbf{d}_n^o + \varepsilon_{n,-t_0}$. Here, λ_1^o, $\beta^{o\prime}$, Γ^o, $\mathbf{c}_n^o$ and $\mathbf{d}_n^o$ are the coefficients in the initial period, which can be different from λ_{10}, β_0', Γ_0, $\mathbf{c}_{0n}$ and $\mathbf{d}_{0n}$ in later periods, where the subscript 0 denotes the true value of parameters.

To analyse asymptotic properties of our 2SIV estimator $\widehat{\kappa}$ of the SDPD model, we need the following assumptions.

ASSUMPTION 3.1. *For any i, j, n and $-t_0 \le t \le T$, the spatial weight $w_{ij,nt} \ge 0$, $w_{ii,nt} = 0$, $\sup_{n,t} ||W_{nt}||_\infty = c_w < \infty$, $\sup_{n,t} ||W_{nt}||_1 = c_u < \infty$ and λ_{10} satisfies that $|\lambda_{10}|c_w < 1$.*

ASSUMPTION 3.2. *The spatial weight $w_{ij,nt}$ satisfies either 3.2(a) or 3.2(b): (a) the spatial weight $w_{ij,nt} = h_{ijt}(z_{i,nt}, z_{j,nt})$ for $i \ne j$, where $h_{ijt}(\cdot)$ are non-negative, uniformly bounded functions of some observable variable Z_{nt}, $0 \le w_{ij,nt} \le c_1\rho_{ij}{}^{-c_3 d_0}$ for some $0 \le c_1$ and $c_3 >$ 3 and, in any period t, there exist at most K ($K \ge 1$) columns of W_{nt} of which each column sum exceeds c_w, where K is a fixed number that does not depend n; (b) the spatial weight $w_{ij,nt} = 0$ if $\rho_{ij} > \rho_c$ (i.e., there exists a threshold $\rho_c > 1$ and if the geographic distance exceeds ρ_c, then the weight is zero) and for $i \ne j$, $w_{ij,nt} = h_{ijt}(z_{i,nt}, z_{j,nt})$ or*

$w_{ij,nt} = h_{ijt}(z_{i,nt}, z_{j,nt})/\sum_{\rho_{ik} \le \rho_c} h_{ikt}(z_{i,nt}, z_{k,nt})$, where $h_{ijt}(\cdot)$ are non-negative, uniformly bounded functions.

Assumption 3.1 is a declining distance function assumption in the spatial econometrics literature even with exogenous spatial weights. Assumption 3.2 is based on sparsity and near-neighbour features, which are adopted in Qu and Lee (2015) to show the consistency and asymptotic normality of the QMLE of a cross-sectional SAR model with an endogenous spatial weight matrix. To be more specific, it constrains the magnitudes of the endogenous spatial weights so that the estimator has the spatial NED property. In our SDPD model, as the time dimension T is small and the initial period is finitely far away, the asymptotic analysis is similar to the cross-sectional setting, but with additional complication from the time-varying and time-differencing spatial weights due to the removal of individual fixed effects.

The following notion of NED for random fields is from Jenish and Prucha (2012). Denote $\mathcal{F}_{i,n}(s)$ as a σ-field generated by the random vectors $\varsigma_{j,n}$ located within the ball $B_i(s)$, which is a ball centred at the location i with a radius s in a d_0-dimensional Euclidean space D. Let $|U|$ denote the cardinality of a finite subset $U \subset D$.

DEFINITION 3.1 (NED). *Let $H = \{H_{i,n}, i \in D_n, n \ge 1\}$ and $\varsigma = \{\varsigma_{i,n}, i \in D_n, n \ge 1\}$ be random fields with $||H_{i,n}||_p < \infty$, $p \ge 1$, where $D_n \subset D$ and $|D_n| \to \infty$ as $n \to \infty$, and let $d = \{d_{i,n}, i \in D_n, n \ge 1\}$ be an array of finite positive constants. Then, the random field H is said to be L_p-NED on the random field ς if $||H_{i,n} - E[H_{i,n}|\mathcal{F}_{i,n}(s)]||_p \le d_{i,n}\varphi(s)$ for some sequence $\varphi(s) \ge 0$ such that $\lim_{s\to\infty} \varphi(s) = 0$. The $\varphi(s)$, which is, w.l.o.g., assumed to be non-increasing, is called the NED coefficient, and the $d_{i,n}$ are called NED scaling factors.*

To show the consistency and asymptotic normality of the 2SIV estimator $\widehat{\kappa}$, in addition to the convergence of each separated term involved in $\widehat{\kappa}$, we need some rank conditions on relevant limiting matrices.

ASSUMPTION 3.3. *Assume that X_{1nt} and X_{2nt} are non-stochastic. The instrument variables q_{1n} and q_{2n} satisfy that $\operatorname*{plim}_{n\to\infty}(1/n)q_{2n}'K_n$ exists, $\operatorname*{plim}_{n\to\infty}(1/n)q_{1n}'q_{1n}$ and $\operatorname*{plim}_{n\to\infty}(1/n)q_{2n}'q_{2n}$ are non-singular, $\operatorname*{plim}_{n\to\infty}(1/n)q_{1n}'K_n$ and $\operatorname*{plim}_{n\to\infty}(1/n)q_{2n}'T_n$ have full column rank, $\operatorname*{plim}_{n\to\infty}(1/n)q_{2n}'\Delta\xi_n(T) = 0$, and $\operatorname*{plim}_{n\to\infty}(1/n)q_{1n}'\Delta\varepsilon_n(T) = 0$.*

Assumption 3.3 on deterministic X_{1nt} and X_{2nt} is for simplicity. The remaining assumption is a general assumption on the validity of instruments. The following claim gives a valid candidate of IV related to our model.

CLAIM 3.1. *Under some regularity conditions, if we choose, for example, $q_{1n} = [\Delta X_{2n}(T-1), \Delta X_{2n}(T)]$ and $q_{2n} = [\Delta(WX_1)_n(T), \Delta(WX_1)_n(T-1), \Delta X_{1n}(T-1), \Delta X_{1n}(T), \Delta\widehat{\varepsilon}_n(T)]$, then conditions in Assumption 3.3 are satisfied.*[1]

The proof of this claim and its detailed regularity conditions are given in Appendix C. The asymptotic properties of our 2SIV estimator are based on the following theorems.

THEOREM 3.1. *Under Assumptions 2.1–3.3, the 2SIV estimator $\widehat{\kappa}$ is a consistent estimator of κ_0.*

[1] These regularity conditions are listed in the proof of this claim in Appendix C. They are essentially rank conditions for IV estimation with those specific IVs.

THEOREM 3.2. *Under Assumptions 2.1–3.3,* $\sqrt{n}(\widehat{\kappa}-\kappa_0) \xrightarrow{d} N(0, \Sigma_{\kappa 0})$, *where*

$$\Sigma_{\kappa 0} = \operatorname*{plim}_{n\to\infty} \frac{1}{n} (T_n' P_{2n} T_n)^{-1} T_n' P_{2n} \Pi_n P_{2n} T_n (T_n' P_{2n} T_n)^{-1},$$

with

$$\Pi_n = \sigma_{\xi 0}^2 \Psi_0 \otimes I_n + K_n (K_n' P_{1n} K_n)^{-1} K_n' P_{1n} (\delta_0' \Sigma_{\varepsilon 0} \delta_0 \Psi_0 \otimes I_n) P_{1n} K_n (K_n' P_{1n} K_n)^{-1} K_n'$$

and

$$\Psi_0 = \begin{pmatrix} 2 & & & & \\ -1 & 2 & & & \\ 0 & -1 & 2 & & \\ 0 & 0 & \ddots & \ddots & \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix}_{(T-2)\times(T-2)}.$$

A consistent estimator of the variance-covariance matrix $\Sigma_{\kappa 0}$ is

$$\begin{aligned}\widehat{\Sigma}_\kappa &= (\widehat{T}_n' P_{2n} \widehat{T}_n)^{-1} \widehat{T}_n' P_{2n} K_n (K_n' P_{1n} K_n)^{-1} K_n' P_{1n} (\widehat{\delta}' \widehat{\Sigma}_\varepsilon \widehat{\delta} \Psi_0 \otimes I_n) P_{1n} K_n (K_n' P_{1n} K_n)^{-1} \\ &\quad \times K_n' P_{2n} \widehat{T}_n (\widehat{T}_n' P_{2n} \widehat{T}_n)^{-1} + (\widehat{T}_n' P_{2n} \widehat{T}_n)^{-1} \widehat{T}_n' P_{2n} (\widehat{\sigma}_\xi^2 \Psi_0 \otimes I_n) P_{2n} \widehat{T}_n (\widehat{T}_n' P_{2n} \widehat{T}_n)^{-1},\end{aligned}$$

where $\widehat{\Sigma}_\varepsilon$ and $\widehat{\sigma}_\xi^2$ are the consistent estimators of $\Sigma_{\varepsilon 0}$ and $\sigma_{\xi 0}^2$.

The following claim gives the detailed instruction to find valid candidates of $\widehat{\Sigma}_\varepsilon$ and $\widehat{\sigma}_\xi^2$, and hence a consistent estimator of $\Sigma_{\kappa 0}$.

CLAIM 3.2. *(a) Suppose* $\binom{\widehat{\rho}_2}{\widehat{\Gamma}}$ *is a consistent estimator of* $\binom{\rho_0}{\Gamma_0}$. *Let* $\Delta\widehat{\varepsilon}_n(T) = \Delta Z_n(T) - K_n \binom{\widehat{\rho}_2}{\widehat{\Gamma}}$. *Then,* $\widehat{\Sigma}_\varepsilon = (1/(2n(T-2))\Delta\widehat{\varepsilon}_n(T)'\Delta\widehat{\varepsilon}_n(T)$ *is a consistent estimator of* $\Sigma_{\varepsilon 0}$. *(b) Similarly, if* $\widehat{\kappa}$ *is a consistent estimator of* κ_0, *then* $\widehat{\sigma}_\xi^2 = (1/2n(T-2))\Delta\widehat{\xi}_n'(T)\Delta\widehat{\xi}_n(T)$ *is a consistent estimator of* $\sigma_{\xi 0}^2$, *where* $\Delta\widehat{\xi}_n = \Delta Y_n(T) - \widehat{T}_n\widehat{\kappa}$. *(c) Furthermore, if we replace* κ_0 *with a consistent estimator* $\widehat{\kappa}$, $\Delta\varepsilon_n(T)$ *with* $\Delta\widehat{\varepsilon}_n(T)$, $\Sigma_{\varepsilon 0}$ *with* $\widehat{\Sigma}_\varepsilon$, *and* $\sigma_{\xi 0}^2$ *with* $\widehat{\sigma}_\xi^2$, *to obtain the empirical estimate of* $\widehat{\Sigma}_\kappa$, *then* $\widehat{\Sigma}_\kappa \xrightarrow{p} \Sigma_{\kappa 0}$.

4. MONTE CARLO SIMULATION

4.1. Data-generating process

In this section, we conduct a Monte Carlo simulation to evaluate finite sample properties of our 2SIV estimator. The data-generating process (DGP) is

$$Y_{nt} = (I_n - \lambda_1 W_{nt})^{-1}(\lambda_2 W_{n,t-1} Y_{n,t-1} + \rho_1 Y_{n,t-1} + X_{1nt}\beta + \mathbf{c}_n + V_{nt}),$$

with the initial value

$$Y_{n,-t_0} = \lambda_1 W_{n,-t_0} Y_{n,-t_0} + X_{1n,-t_0}\beta + \mathbf{c}_n + V_{n,-t_0},$$

where $t_0 = 8$, i.e., eight periods before the zero period, $T = 10$ (i.e., the total observed periods are 10), $X_{1nt} \sim N(0, 1)$, $\lambda_1 = 0.5$, $\lambda_2 = 0.1$, $\rho_1 = 0.2$, $\beta = 1$, and $\mathbf{c}_n \sim U(0, 1)$. The endogenous, row-normalized $W_{nt} = (w_{ij,nt})$ is constructed as follows.

STEP 1. Generate bivariate normal random variables $(v_{i,nt}, \varepsilon_{i,nt})$ from

$$\text{i.i.d } N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

as disturbances in equations Y_{nt} and Z_{nt}.

STEP 2. Construct the spatial weights matrix as the Hadamard product $W_{nt} = W_n^d \circ W_{nt}^e$, i.e., $w_{ij,nt} = w_{ij,n}^d \cdot w_{ij,nt}^e$, where W_n^d is a predetermined matrix based on geographic distance (time invariant), $w_{ij,n}^d = 1$ if the two locations are neighbours and otherwise 0, W_{nt}^e is a time-varying matrix based on economic similarity, $w_{ij,nt}^e = 1/|z_{i,nt} - z_{j,nt}|$ if $i \neq j$ and $w_{ii,nt}^e = 0$, where elements of $z_{i,nt}$ is generated by $z_{i,nt} = z_{i,n,t-1}\rho_2 + x_{i,2nt}\gamma + d_{i,n} + \varepsilon_{i,nt}$ with the initial value $z_{i,n,-t_0} = x_{i,2n,-t_0}\gamma + d_{i,n} + \varepsilon_{n,-t_0}$, where $X_{2nt} = X_{1nt}$, $d_{i,n} \sim U(0, 1)$, $\rho_2 = 0.7$ and $\gamma = 0.8$.[2]

STEP 3. Row-normalize W_n.

For the predetermined W_n^d, we use four examples: (1) the US states spatial weights matrix $WS(49 \times 49)$, based on the contiguity of the 48 contiguous states and DC; (2) the Toledo spatial weight matrix $WO(98 \times 98)$, based on the five nearest neighbours of 98 census tracts in Toledo, Ohio; (3) the Iowa 'county' spatial weight matrix $WC(361 \times 361)$, based on whether the school districts are in the same county in Iowa in 2009; (4) the Iowa 'adjacency' spatial weight matrix $WA(361 \times 361)$, based on the adjacency of 361 school districts in Iowa in 2009.

In the simulation, we compare our 2SIV estimates with the conventional IV estimates, which refer to the case of treating W_{nt} as exogenous. Here the conventional IV only estimates the outcome equation (2.1) (the Z_{nt} equation is not estimated) as it treats W_{nt} as exogenous. For the simulation, we choose $\Delta(WX)_{nt}$ to instrument $\Delta(WY)_{nt}$, and $\Delta X_{n,t-1}$ to instrument $\Delta Y_{n,t-1}$. The 2SIV estimates both equations of Y_{nt} and Z_{nt}. In the first step, we use $\Delta X_{n,t-1}$ as the instrument for $\Delta Z_{n,t-1}$ to estimate the Z_{nt} equation, and in the second step, $\Delta(WX)_{nt}$, $\Delta(WX)_{n,t-1}$ and $\Delta X_{n,t-1}$ are used as instruments for $\Delta(WY)_{nt}$, $\Delta(WY)_{n,t-1}$ and $\Delta Y_{n,t-1}$ to estimate the Y_{nt} equation. Of special interest, we want to see how large the bias would be for this conventional estimation method when W_{nt} is indeed endogenous. To generate different degrees of endogeneity, we choose correlation coefficients $\rho = 0.2$, 0.5 and 0.8.[3]

4.2. Monte Carlo results

In this section, we report the finite sample performance of our 2SIV estimator. Tables 1–4 list the empirical mean, the empirical standard deviation of each estimator (in brackets) and the empirical mean of the estimated standard error based on the asymptotic variance–covariance matrix (in parentheses) from 1000 replications using WS, WO, WC or WA as the predetermined spatial weights matrix W_n^d. To see how the different estimation methods react to the magnitude

[2] In our model, parameters can be identified even if X_{1nt} and X_{2nt} are the same, so in the Monte Carlo simulation, we use $X_{2nt} = X_{1nt}$.

[3] We try DGPs with some other values of λ_1, λ_2, ρ_1, ρ_2, β and γ. These results can be found in online Appendix A and the patterns are similar.

Table 1. Estimates from the US states spatial weights matrix $WS(n = 49, T = 10)$.

	$\lambda_1 = 0.5$	$\lambda_2 = 0.1$	$\rho_1 = 0.2$	$\beta = 1$	$\rho_2 = 0.7$	$\gamma = 0.8$	$\delta = 0$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 0.8$
$\rho = 0$										
Conv. IV	0.49669	0.09392	0.20406	1.00201						
	(0.08052)	(0.10054)	(0.07753)	(0.06873)						
	[0.08138]	[0.10348]	[0.07906]	[0.06813]						
2SIV	0.49671	0.09365	0.20408	1.00201	0.71351	0.80631	0.00065			
	(0.08067)	(0.10059)	(0.07751)	(0.06874)	(0.15184)	(0.07793)	(0.06139)			
	[0.08141]	[0.10366]	[0.07902]	[0.06805]	[0.15459]	[0.08016]	[0.06062]			
$\rho = 0.2$										
Conv. IV	0.50693	0.08439	0.20476	0.99878						
	(0.08109)	(0.10252)	(0.07837)	(0.06916)						
	[0.08351]	[0.10667]	[0.08041]	[0.06876]						
2SIV	0.49536	0.09585	0.20433	1.00273	0.71300	0.80620		0.19893		
	(0.08019)	(0.10072)	(0.07822)	(0.06926)	(0.15180)	(0.07792)		(0.06269)		
	[0.08272]	[0.10495]	[0.08009]	[0.06841]	[0.15459]	[0.07951]		[0.06174]		
$\rho = 0.5$										
Conv. IV	0.52508	0.06849	0.20519	0.99222						
	(0.08127)	(0.10517)	(0.07951)	(0.06964)						
	[0.08341]	[0.10536]	[0.08109]	[0.06966]						
2SIV	0.49759	0.09985	0.20313	1.00265	0.71210	0.80598			0.49703	
	(0.07222)	(0.09195)	(0.07766)	(0.06894)	(0.15167)	(0.07789)			(0.06445)	
	[0.07458]	[0.09278]	[0.07869]	[0.06819]	[0.15452]	[0.07822]			[0.06280]	
$\rho = 0.8$										
Conv. IV	0.54319	0.04902	0.20644	0.98450						
	(0.08172)	(0.10781)	(0.08081)	(0.07006)						
	[0.08007]	[0.10938]	[0.08350]	[0.06998]						
2SIV	0.49942	0.09989	0.20215	1.00244	0.71098	0.80577				0.79500
	(0.05221)	(0.06702)	(0.07472)	(0.06697)	(0.15145)	(0.07781)				(0.05795)
	[0.05330]	[0.06948]	[0.07575]	[0.06533]	[0.15429]	[0.07686]				[0.05701]

Note: Estimated standard errors based on an asymptotic variance-covariance matrix are given in parentheses and empirical standard deviation is in brackets.

Table 2. Estimates from the Ohio State spatial weights matrix $WO(n = 98, T = 10)$.

	$\lambda_1 = 0.5$	$\lambda_2 = 0.1$	$\rho_1 = 0.2$	$\beta = 1$	$\rho_2 = 0.7$	$\gamma = 0.8$	$\delta = 0$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 0.8$
$\rho = 0$										
Conv. IV	0.49915	0.09504	0.20129	0.99900						
	(0.06102)	(0.07445)	(0.05421)	(0.04800)						
	[0.06041]	[0.07127]	[0.05182]	[0.04694]						
2SIV	0.49924	0.09500	0.20129	0.99900	0.70057	0.80302	−0.00058			
	(0.06122)	(0.07451)	(0.05420)	(0.04802)	(0.10448)	(0.05395)	(0.04353)			
	[0.06062]	[0.07116]	[0.05182]	[0.04700]	[0.10703]	[0.05555]	[0.04325]			
$\rho = 0.2$										
Conv. IV	0.51298	0.08122	0.20215	0.99500						
	(0.06155)	(0.07640)	(0.05487)	(0.04832)						
	[0.06088]	[0.07469]	[0.05281]	[0.04757]						
2SIV	0.49737	0.09390	0.20197	0.99939	0.70027	0.80265		0.20064		
	(0.06100)	(0.07508)	(0.05478)	(0.04842)	(0.10450)	(0.05394)		(0.04448)		
	[0.06046]	[0.07400]	[0.05259]	[0.04750]	[0.10591]	[0.05457]		[0.04542]		
$\rho = 0.5$										
Conv. IV	0.53595	0.06021	0.20294	0.98736						
	(0.06228)	(0.07920)	(0.05592)	(0.04880)						
	[0.06209]	[0.07636]	[0.05437]	[0.04833]						
2SIV	0.49693	0.09350	0.20200	0.99988	0.70020	0.80206			0.50203	
	(0.05556)	(0.06917)	(0.05463)	(0.04842)	(0.10463)	(0.05396)			(0.04565)	
	[0.05589]	[0.06887]	[0.05257]	[0.04757]	[0.10377]	[0.05288]			[0.04673]	
$\rho = 0.8$										
Conv. IV	0.56325	0.03726	0.20388	0.97669						
	(0.06269)	(0.08204)	(0.05706)	(0.04920)						
	[0.06310]	[0.08134]	[0.05626]	[0.04913]						
2SIV	0.49872	0.09594	0.20148	0.99986	0.70084	0.80144				0.80132
	(0.04030)	(0.05055)	(0.05255)	(0.04719)	(0.10491)	(0.05404)				(0.04094)
	[0.04080]	[0.05108]	[0.05080]	[0.04667]	[0.10131]	[0.05144]				[0.04016]

Note: Same as for Table 1.

Table 3. Estimates from the Iowa State county weights matrix $WC(n = 361, T = 10)$.

	$\lambda_1 = 0.5$	$\lambda_2 = 0.1$	$\rho_1 = 0.2$	$\beta = 1$	$\rho_2 = 0.7$	$\gamma = 0.8$	$\delta = 0$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 0.8$
$\rho = 0$										
Conv. IV	0.49891	0.10156	0.19936	0.99808						
	(0.02407)	(0.03201)	(0.02820)	(0.02507)						
	[0.02479]	[0.03113]	[0.02818]	[0.02442]						
2SIV	0.49893	0.10151	0.19937	0.99807	0.69944	0.80059	−0.00027			
	(0.02409)	(0.03201)	(0.02820)	(0.02508)	(0.05406)	(0.02789)	(0.02237)			
	[0.02483]	[0.03123]	[0.02818]	[0.02445]	[0.05477]	[0.02755]	[0.02272]			
$\rho = 0.2$										
Conv. IV	0.50539	0.09371	0.19959	0.99522						
	(0.02411)	(0.03236)	(0.02835)	(0.02512)						
	[0.02439]	[0.03110]	[0.02843]	[0.02447]						
2SIV	0.49833	0.10071	0.19964	0.9983	0.69906	0.80007		0.19971		
	(0.02378)	(0.03180)	(0.02832)	(0.02514)	(0.05406)	(0.02788)		(0.02278)		
	[0.02404]	[0.03078]	[0.02829]	[0.02437]	[0.05465]	[0.02749]		[0.02298]		
$\rho = 0.5$										
Conv. IV	0.51527	0.08273	0.19984	0.99063						
	(0.02419)	(0.03293)	(0.02859)	(0.02519)						
	[0.02485]	[0.03143]	[0.02863]	[0.02480]						
2SIV	0.49782	0.10032	0.19982	0.99850	0.69867	0.79929			0.50008	
	(0.02137)	(0.02888)	(0.02808)	(0.02496)	(0.05405)	(0.02786)			(0.02323)	
	[0.02175]	[0.02830]	[0.02800]	[0.02418]	[0.05443]	[0.02735]			[0.02271]	
$\rho = 0.8$										
Conv. IV	0.52487	0.06993	0.20066	0.98562						
	(0.02425)	(0.03350)	(0.02885)	(0.02526)						
	[0.02509]	[0.03185]	[0.02914]	[0.02509]						
2SIV	0.49826	0.09909	0.19998	0.99839	0.69859	0.79852				0.80044
	(0.01529)	(0.02112)	(0.02719)	(0.02432)	(0.05405)	(0.02784)				(0.02065)
	[0.01520]	[0.02103]	[0.02736]	[0.02358]	[0.05430]	[0.02714]				[0.01999]

Note: Same as for Table 1.

Table 4. Estimates from the Iowa State adjacency weights matrix $WA(n = 361, T = 10)$.

	$\lambda_1 = 0.5$	$\lambda_2 = 0.1$	$\rho_1 = 0.2$	$\beta = 1$	$\rho_2 = 0.7$	$\gamma = 0.8$	$\delta = 0$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 0.8$
$\rho = 0$										
Conv. IV	0.49815	0.10231	0.19922	0.99821						
	(0.03114)	(0.03832)	(0.02821)	(0.02502)						
	[0.03179]	[0.03766]	[0.02801]	[0.02425]						
2SIV	0.49814	0.10230	0.19923	0.99821	0.69944	0.80059	−0.00013			
	(0.03125)	(0.03836)	(0.02821)	(0.02504)	(0.05406)	(0.02789)	(0.02250)			
	[0.03179]	[0.03772]	[0.02800]	[0.02430]	[0.05477]	[0.02755]	[0.02290]			
$\rho = 0.2$										
Conv. IV	0.51381	0.08860	0.19969	0.99331						
	(0.03125)	(0.03912)	(0.02857)	(0.02519)						
	[0.03068]	[0.03756]	[0.02832]	[0.02423]						
2SIV	0.49805	0.10143	0.19942	0.99830	0.69906	0.80007		0.19981		
	(0.03105)	(0.03846)	(0.02852)	(0.02526)	(0.05406)	(0.02788)		(0.02303)		
	[0.03053]	[0.03700]	[0.02825]	[0.02414]	[0.05465]	[0.02749]		[0.02345]		
$\rho = 0.5$										
Conv. IV	0.53715	0.06783	0.20056	0.98460						
	(0.03156)	(0.04046)	(0.02916)	(0.02545)						
	[0.03132]	[0.03926]	[0.02889]	[0.02515]						
2SIV	0.49797	0.10078	0.19959	0.99838	0.69867	0.79929			0.50013	
	(0.02824)	(0.03537)	(0.02842)	(0.02523)	(0.05405)	(0.02786)			(0.02373)	
	[0.02841]	[0.03493]	[0.02828]	[0.02447]	[0.05443]	[0.02735]			[0.02355]	
$\rho = 0.8$										
Conv. IV	0.56100	0.04539	0.20189	0.97411						
	(0.03173)	(0.04181)	(0.02978)	(0.02567)						
	[0.03090]	[0.04072]	[0.03014]	[0.02523]						
2SIV	0.49808	0.09937	0.19989	0.99843	0.69859	0.79852				0.80055
	(0.02047)	(0.02583)	(0.02724)	(0.02451)	(0.05405)	(0.02784)				(0.02134)
	[0.02034]	[0.02597]	[0.02751]	[0.02350]	[0.05430]	[0.02714]				[0.02071]

Note: Same as for Table 1.

of endogeneity, we have four sets of results. In each table, the first panel shows the results with $\rho = 0$, i.e., there is no endogeneity in the spatial weights. Then, the second, third and fourth panels show the results with increasing endogeneity $\rho = 0.2$, 0.5 and 0.8. There are several patterns in the estimation results, as follows.

(a) For the bias, the more serious the endogeneity is (i.e., the larger ρ is), the more bias of the conventional IV method will create. Most biases come from the estimates of spatial correlation coefficients, i.e., the spatial lag λ_1 and the spatial-time lag λ_2. In general, estimators for β and ρ_1 are much less biased. In the current setting of DGP, $\widehat{\lambda_2}$ from the conventional IV suffers severe downward bias when ρ achieves 0.5, in some cases exceeding 100%. Correspondingly, $\widehat{\lambda_1}$ from the conventional IV has the upward bias. Although biases of β and ρ_1 in the conventional IV method can be negligible, interpretation of these coefficients significantly depends on spatial correlation coefficients λ. Also, the existence of spatial correlation indicates the multiplier effect. In all cases, our 2SIV has estimates very close to the true value. From this, we can conclude that with endogenous spatial weights matrices, the 2SIV estimation method is preferred.

(b) For the accuracy of estimation, we consider the empirical standard deviation or the estimated standard error. When there is no endogeneity in spatial weights (i.e., $\rho = 0$), both the conventional IV and 2SIV give estimates close to the true value, and the conventional IV is more accurate in terms of smaller variance. However, when there is endogeneity, estimates from the conventional IV have larger variances. Hence, the simulation results support our theory that the conventional IV is inconsistent when W_{nt} is truly endogenous, but more efficient with exogenous W_{nt}; 2SIV is consistent all the time, but might lose some efficiency when W_{nt} is exogenous.

(c) From Tables 1–4, as sample size increases while the number of neighbours for each agent grows at a slower rate, we observe that the bias and standard error of estimates decrease. When the sample size is 361, such as in the case of WA and WC, the estimates appear quite precise and close to the true parameters.

5. AN EMPIRICAL APPLICATION

In this section, we provide an illustration of using our method to investigate the spillover effects of government spending among the 30 provinces in China between the year 2007–2013.[4] Strategic interaction among governments has been studied in both theoretical and empirical frameworks in public economics. We believe that the government spending in one province depends on its neighbours' fiscal policies for the following reasons. First, as suggested by Baicker (2005) and Redoano (2007), there exist expenditure externalities that might generate 'a race to the bottom'. An example of this type of externality is the amount of public investments in infrastructures inside a province whose benefits can spill over into neighbouring provinces and lower the level of expenditures in the neighbouring provinces because of the free rider behaviour. From this aspect, there might exist strategic substitutes (i.e., the negative correlation) among the public expenditure from different provinces. Second, there might also be strategic complements

[4] For provinces, we exclude Tibet, Hong Kong and Macau due to data availability. These excluded places are also special in terms of administration, so they might behave differently from other provinces in China. For the year, we exclude data before 2007 because the National Bureau of Statistics of China has employed the new statistic caliber since 2007.

Table 5. Empirical results for the first step.

Variable	First-stage estimation	Std error
$\ln G_{i,t-1}$	0.4591***	(0.1726)
ln Population Density	0.1456	(0.1208)
ln City Population percentage	0.1673	(0.1189)
ln Under 14 Years Old percentage	0.1270	(0.1026)
ln Over 14 and Under 65 percentage	0.3681***	(0.1298)
ln Over 65 Years Old percentage	0.2352*	(0.1406)
ln Import to GDP Ratio	0.1908*	(0.1078)
ln High Way Density	0.1963*	(0.1139)

Note: Standard errors are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

(i.e., the positive correlation) among local governments, due to the competition effects. By improving the public service and providing better investment and working environments, the provinces compete with each other, especially their neighbours, to attract more capital and labour resources. Third, the promotion of local government officials in China is determined by the upper level government and is highly correlated with its GDP growth rate compared to the previous period and to its neighbours. Therefore, we can expect more competition effects between local governments in China. For these three reasons, it is necessary to take the interdependence among provinces into account when we try to evaluate the determination of the spending policy. In our empirical model, we include the lagged dependent variable because fiscal policies are serially correlated. Usually, a fiscal budget is based on the budget of last year. Any abrupt change would lead to a big cost in terms of paperwork in the fiscal system or objection from an interest group at the political level. We also want to investigate whether this serially correlated effect carries over neighbouring provinces, so we incorporate the lagged spatial term. To sum up, it is reasonable to employ the SDPD model in (2.1) to study spillover effects of government spending.

As suggested in Keen and Marchand (1997) and Borck et al. (2007), strategic interactions among communities would not only affect the level of public expenditure, but also the composition of expenditures. This implies the possible existence of different types of strategic interaction behaviour for different types of expenditures: strategic substitutes or strategic complements, so we want to study different types of expenditure separately. In this empirical application, we use four different types of expenditures as our dependent variable Y_n: the aggregated level of public expenditure per capita, education expenditure per capita, medical care expenditure per capita and social security expenditure per capita.

To measure the spatial dependence, based on Redoano (2007), that provinces affect and follow other provinces close to them both geographically and with similar economic structures, we construct spatial weights using both geographic contiguity and economic distance. Before row normalization, the spatial weights matrix $W_{nt}^* = W_n^d \circ W_{nt}^e$, where

$$(W_n^d)_{ij} = \begin{cases} 1, & \text{if } j \neq i, \text{ and locations } i \text{ and } j \text{ share a border} \\ 0, & \text{otherwise} \end{cases}$$

describes the geographic contiguity among the provinces in China and

$$(W_{nt}^e)_{ij} = \begin{cases} 1/|\ln G_{it} - \ln G_{jt}|, & \text{if } j \neq i \\ 0, & \text{if } j = i \end{cases}$$

Table 6. Empirical results for aggregate expenditure per capita.

Variable (ln)	2SIV	Conv. IV	2SIV with W_n^d
$W_{nt} \ln E_{nt}$	0.3440***	0.4091***	0.9041***
	(0.1065)	(0.0974)	(0.1435)
$W_{n,t-1} \ln E_{n,t-1}$	0.0573	0.0937	−0.4438***
	(0.0835)	(0.0879)	(0.1536)
$\ln E_{n,t-1}$	0.4205***	0.3594***	0.5007***
	(0.0999)	(0.1044)	(0.1185)
ln Population Density	−0.0223	−0.0392	−0.0326
	(0.0395)	(0.0397)	(0.0369)
ln City Population Percent	−0.0186	−0.0292	−0.0148
	(0.0416)	(0.0433)	(0.0370)
ln Under 14 Years Old Percent	−0.0088	−0.0106	−0.0051
	(0.0101)	(0.0104)	(0.0096)
ln Over 14 and Under 65 Percent	−0.0096	−0.0250	0.0150
	(0.0239)	(0.0258)	(0.0220)
ln Over 65 Years Old Percent	−0.0428**	−0.0599***	−0.0386**
	(0.0211)	(0.0218)	(0.0189)
City Unemployment Rate	−0.0351**	−0.0507***	−0.0388***
	(0.0165)	(0.0189)	(0.0148)
Primary Industry Ratio	0.0098	0.0083	0.0095
	(0.0130)	(0.0135)	(0.0118)
ln Import to GDP Ratio	0.0243	0.0255	0.0329*
	(0.0210)	(0.0216)	(0.0196)
ln Fixed Asset Investment to GDP Ratio	0.0197	0.0229	0.0325**
	(0.0146)	(0.0142)	(0.0126)
Residual from the First Step	0.2504**	NA	0.0662
	(0.1025)	NA	(0.0843)

Note: Standard errors are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

captures the economic distance, where G_{it} is GDP per capita in province i at time t. After row-normalizing W_{nt}^*, we have the W_{nt} used in our SDPD model, i.e.,

$$(W_{nt})_{ij} = \begin{cases} \dfrac{1}{|\ln G_{it} - \ln G_{jt}|} \Big/ \sum_j \dfrac{1}{|\ln G_{it} - \ln G_{jt}|} & \text{if } j \text{ and } i \text{ share a border and } j \neq i \\ 0, & \text{otherwise.} \end{cases}$$

To be more specific, for each of the 30 provinces, other provinces that share a common border are given the spatial weight of

$$\frac{1}{|\ln G_{it} - \ln G_{jt}|} \Big/ \sum_j \frac{1}{|\ln G_{it} - \ln G_{jt}|},$$

while those that do not are given the weight of zero. We provide some intuitions about the way how we construct such a weights matrix. On the one hand, we think each province can be only affected by its geographical neighbours. For each province, other provinces that do not share a common border are given the value of spatial weight as zero. On the other hand, we believe within neighbours (i.e., the provinces that share a common border), the magnitude of spatial weights depends on the economic structure. Provinces with more similar economic size have larger spatial weights. Thus, for the non-zero elements, we construct the weights matrix based on the inverse of the distance between GDP per capita. Intuitively, if the economics structures between provinces

Table 7. Empirical results for educational expenditure per capita.

Variable	2SIV	Conventional IV	2SIV with W_n^d
$W_{nt} \ln E_{nt}$	0.7376***	0.7119***	0.9648***
	(0.1057)	(0.1158)	(0.0864)
$W_{n,t-1} \ln E_{n,t-1}$	−0.1972	−0.1175	−0.2045
	(0.1832)	(0.1845)	(0.2214)
$\ln E_{n,t-1}$	0.3209**	0.2511	0.2294
	(0.1627)	(0.1665)	(0.2076)
ln Population Density	−0.0370	−0.0710	−0.0356
	(0.1014)	(0.1015)	(0.0940)
ln City Population Percent	−0.0308	−0.0530	0.0205
	(0.1094)	(0.1137)	(0.1037)
ln Under 14 Years Old Percent	0.0246	0.0243	0.0574
	(0.0517)	(0.0535)	(0.0491)
ln Over 14 and Under 65 Percent	0.2499*	0.1585	0.2216*
	(0.1335)	(0.1294)	(0.1251)
ln Over 65 Years Old Percent	0.0038	−0.0340	−0.0478
	(0.1107)	(0.1123)	(0.1051)
City Unemployment Rate	−0.0635	−0.0954*	−0.0159
	(0.0536)	(0.0558)	(0.0494)
ln Fixed Asset Investment to GDP Ratio	−0.1951*	−0.2309**	−0.1001
	(0.1027)	(0.1029)	(0.0919)
ln Road Density	−0.0747	−0.0828	−0.0552
	(0.0717)	(0.0739)	(0.0646)
Residual from First Step	0.3447**	NA	0.1309
	(0.1686)	NA	(0.1330)

Note: Standard errors are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

are more similar, the interaction of fiscal policy between these two provinces would be higher. In sum, we take both the physical distance and economic distance into consideration.[5]

The weights matrix is likely to be endogenous because of the potential endogeneity in GDP per capita. Here, we provide some intuitions and explanations. First, the GDP per capita and public expenditure might be affected by the same factors, such as the natural resource constraint and the ability of local governors, which are hard to observe or measure. Second, common macro shocks, such as the financial crisis and monetary policy shocks, might affect both GDP growth and public expenditure. As a result, it is very possible that the weights matrix is correlated with the error term in the expenditures. Therefore, we would expect a significant coefficient of the residual from our first-step estimation when we use the control function method to control for endogeneity. To verify this, we further present the empirical results with the pure geographic weight W_n^d. In other words, if the residual from the first step is insignificant with this exogenous weights, but indeed significant with our $W_{nt} = W_n^d \circ W_{nt}^e$, then it gives us strong support that economic distance is the source of the endogeneity of the weight matrix.

[5] We use the same specification here as our Monte Carlo simulation, but our model can apply to other cases of W_{nt}, as long as Assumption 3.2 is satisfied. For example, if W_{nt} is sparse, i.e., $W_{nt} = W_n^d \circ W_{nt}^e$, then $w_{nt,ij}$ can be any bounded function of z_{it} and z_{jt}, e.g., $w_{nt,ij}^e = z_{it}/z_{jt}$ or $w_{nt,ij}^e = z_{it} \cdot z_{jt}$ if z are bounded. If W_{nt} is not sparse, it needs to have a distance decaying spatial weights matrix, e.g., $w_{nt,ij} = z_{it} \cdot z_{jt}/\rho_{ij}^{\alpha}$, where ρ_{ij} is the physical distance between i and j, and $\alpha >$ some positive number. However, the physical distance plays an important role in our asymptotic analysis, so we cannot allow for spatial weights constructed by pure economics factors.

Table 8. Empirical results for medical care expenditure per capita.

Variable	2SIV	Conventional IV	2SIV with W_n^d
$W_{nt} \ln E_{nt}$	0.6690***	0.7347***	0.9257***
	(0.1166)	(0.1187)	(0.1110)
$W_{n,t-1} \ln E_{n,t-1}$	0.2579*	0.3592**	−0.1654
	(0.1415)	(0.1457)	(0.1727)
$\ln E_{n,t-1}$	−0.1112	−0.1627	0.2230
	(0.1639)	(0.1734)	(0.1722)
ln Population Density	−0.0115	−0.0138	0.0145
	(0.0526)	(0.0548)	(0.0448)
ln Under 14 Years Old Percent	−0.0093	−0.0108	−0.0177
	(0.0155)	(0.0160)	(0.0132)
ln Over 14 and Under 65 Percent	0.0075	0.0119	−0.0057
	(0.0399)	(0.0423)	(0.0293)
ln Over 65 Years Old Percent	0.0661***	0.0527*	0.0887***
	(0.0248)	(0.0276)	(0.0200)
Primary Industry Ratio	0.0470**	0.0435*	0.0538***
	(0.0224)	(0.0240)	(0.0185)
ln Import to GDP Ratio	0.0384	0.0229	0.0270
	(0.0243)	(0.0253)	(0.0220)
ln Fixed Asset Investment to GDP Ratio	−0.0241	−0.0143	0.0144
	(0.0249)	(0.0250)	(0.0220)
Residual from First Step	0.4161**	NA	0.1431
	(0.1695)	NA	(0.1315)

Note: Standard errors are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

To handle the possibly endogenous weights matrix, we estimate the spillover effect of public expenditures by the 2SIV method in Section 3. By noting the economic distance in the weights matrix and (2.2), we study the log GDP per capita in the first step. Like Islam (1995) and Caselli et al. (1996) who study the economic growth using dynamic panel data models, we consider the following regression model,

$$\ln G_{it} = \eta_i + \rho_2 \ln G_{i,t-1} + \gamma' x_{1it} + \varepsilon_{it}, \qquad t = 1, 2, \ldots, 7,$$

where G_{it} denotes the real GDP per capita in province i in the tth year from 2007 to 2013 and x_{1it} is a list of province i's characteristics at time period t including variables measuring population, imports and highways.

Based on the residual $\widehat{\varepsilon}_{it}$ from this first-step estimation and noting (2.3), we consider the following SDPD model in the second-step estimation:

$$\ln E_{it} = c_i + \rho_1 \ln E_{i,t-1} + \lambda_1 \sum_{j\neq i} w_{ijt} \ln E_{jt} + \lambda_2 \sum_{j\neq i} w_{ij,t-1} \ln E_{j,t-1} + x'_{2it}\beta + \delta\widehat{\varepsilon}_{it} + v_{it},$$

where E_{it} can be the aggregated level of public expenditure per capita or separate spending for education, medical care and social security per capita as mentioned above. Here, c_i is the individual fixed effect, $\ln E_{i,t-1}$ is the time lag, $\sum_{j\neq i} w_{ijt} \ln E_{jt}$ is the spatial lag and $\sum_{j\neq i} w_{ij,t-1} \ln E_{j,t-1}$ is the spatial-time lag. For x_{2it}, following Borck et al. (2007) and Shelton (2007), we choose variables that are traditionally used in the determination of the

Table 9. Empirical results for social security expenditure per capita.

Variable	2SIV	Conventional IV	2SIV with W_n^d
W_{nt} ln E_{nt}	0.5285***	0.5086***	0.9814***
	(0.1823)	(0.1860)	(0.1664)
$W_{n,t-1}$ ln $E_{n,t-1}$	0.3709***	0.4299**	0.4918***
	(0.1332)	(0.1262)	(0.1725)
ln $E_{n,t-1}$	−0.3827**	−0.3053**	−0.5744***
	(0.1663)	(0.1643)	(0.1523)
ln Population Density	−0.0160	−0.0371	0.0784
	(0.1379)	(0.1379)	(0.1281)
ln Under 14 Years Old Percent	0.2629**	0.2388**	0.2372**
	(0.1300)	(0.1330)	(0.1094)
ln Over 14 and Under 65 Percent	−0.0912	−0.0999	−0.0815
	(0.0762)	(0.0752)	(0.0707)
ln Over 65 Years Old Percent	0.1153	0.0684	0.0826
	(0.0946)	(0.0850)	(0.0842)
City Unemployment Rate	0.0653	0.0114	0.0889
	(0.0732)	(0.0735)	(0.0654)
ln Import to GDP Ratio	0.1423*	0.0945	0.1196*
	(0.0775)	(0.0664)	(0.0680)
ln High Way Density	−0.2535***	−0.2874***	−0.2755***
	(0.0852)	(0.0831)	(0.0793)
Residual from First Step	0.4868**	NA	0.0700
	(0.2353)	NA	(0.1807)

Note: Standard errors are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

province expenditure, e.g., population density, population percentage of different age groups, unemployment rate, primary industry ratio and import to GDP ratio.

Table 5 reports the first-step estimation results. Tables 6–9 report the second-step estimation results, using different types of government expenditures as the dependent variable. From these estimates, we have some interesting findings, as follows.

(a) In all four tables, Tables 6–9, $\hat{\delta}$, the coefficient associated with the residual from the first-step estimation, is significant at the 5% level when we use W_n and insignificant when we use the geographic contiguity weight matrix W_n^d. This supports our model in that the weights matrix using economic variables is endogenous.

(b) In most cases, coefficients of the spatial lag, the time lag and spatial-time lag terms are significant, indicating the spatial and time dependences of the provincial government spending in China, and it is reasonable to employ the SDPD model proposed in this paper to study the spillover effect of the government expenditure.

(c) It is worth pointing out that the spatial lag terms are significant and have positive signs in all cases. From this, we can conclude that the strategic complement plays a major role in the determination of the local fiscal policy. In other words, we believe that a specific province would increase its government expenditure to respond to its neighbours' expansion of government expenditure.

(d) Compared with results from the conventional method that ignores the possible endogeneity in W_n, we find that the signs of significant variables are consistent, but their magnitudes are changed. Therefore, if we use the endogenous spatial weights matrix but treat it as exogenous, we can still confirm the existence of spatial spillover effects, but the magnitude of these effects might be biased.

(e) From $\hat{\rho}_2 = 0.4591 < 1$ in the first-step estimation, we find evidence to support the hypothesis of economic growth convergence across regions within a certain country, which relevant theory is considered in Chapter 11 of Barro and Sala-i-Martin (2004).

6. CONCLUSION

This paper presents an SDPD model with endogenous spatial weights in the setting of large n and small T. We propose a 2SIV estimation method based on the control function method and we establish the consistency and asymptotic normality of our estimator. Monte Carlo experiments show that the proposed 2SIV method has a satisfactory finite sample performance, while the conventional IV method, which regards spatial weights as exogenous, is biased when spatial weights are endogenous. This model is applied to study government spending in China and we find there does exist strategic interaction, especially expenditure competition among China's provincial governments.

ACKNOWLEDGEMENTS

We are grateful to the co-editor Victor Chernozhukov and to two anonymous referees for their constructive suggestions and comments. Previous versions of this paper have been presented at the 11th International Symposium on Econometric Theory and Applications (SETA) in Tokyo, the 2nd Meeting of Young Econometricians in Asian-Pacific Region (YEAP) in Beijing, and seminars in Shanghai University of Finance and Economics (SUFE) and Tsinghua University. We also thank the participants of these conferences for their helpful discussions. The first author gratefully acknowledges that the paper is sponsored by the Chenguang Programme supported by the Shanghai Education Development Foundation and Shanghai Municipal Education Commission under the grant number 15CG09.

REFERENCES

Aiello, F. and P. Cardamone (2008). R&D spillovers and firms' performance in Italy: evidence from a flexible production function. *Empirical Economics 34*, 143–66.

Anselin, L., J. Le Gallo and H. Jayet (2008). Spatial panel econometrics. In L. Matyas and P. Sevestre (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, 625–60. Berlin: Springer.

Baicker, K. (2005). The spillover effects of state spending. *Journal of Public Economics 89*, 529–44.

Baltagi, B. H. (2011). Spatial panels. In A. Ullah and D. E. A. Giles (Eds.), *Handbook on Empirical Economics and Finance*, 397–434. New York, NY: Taylor & Francis.

Baltagi, B. and D. Li (2006). Prediction in the panel data model with spatial correlation: the case of liquor. *Spatial Economic Analysis 1*, 175–85.

Baltagi, B., S. H. Song and W. Koh (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics 117*, 123–50.

Baltagi, B., P. Egger and M. Pfaffermayr (2007a). A generalised spatial panel data model with random effects. Working paper, Syracuse University.

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

Barro, R. J. and X. Sala-i-Martin (2004). *Economic Growth* (2nd ed.). Cambridge, MA: MIT Press.

Borck, R., M. Caliendo and V. Steiner (2007). Fiscal competition and the composition of public spending: theory and evidence. *FinanzArchiv/Public Finance Analysis 63*, 264–77.

Brueckner, J. K. (1998). Testing for strategic interaction among local governments: the case of growth controls. *Journal of Urban Economics 44*, 438–67.

Brueckner, J. K. (2003). Strategic interaction among governments: an overview of empirical studies. *International Regional Science Review 26*, 175–88.

Brueckner, J. K. and L. A. Saavedra (2001). Do local governments engage in strategic property tax competition? *National Tax Journal 54*, 203–29.

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

Caselli, F., G. Esquivel and F. Lefort (1996). Reopening the convergence debate: a new look at cross-country growth empirics. *Journal of Economic Growth 1*, 363–89.

Cliff, A. D. and J. K. Ord (1973). *Spatial Autocorrelation*. London: Pion Ltd.

Crabbé, K. and H. Vandenbussche (2013). Are your firm's taxes set in Warsaw? Spatial tax competition in Europe. *FinanzArchiv/Public Finance Analysis 69*, 317–37.

Druska, V. and W. C. Horrace (2004). Generalised moments estimation for spatial panel data: Indonesian rice farming. *American Journal of Agricultural Economics 86*, 185–98.

Egger, P., M. Pfaffermayr and H. Winner (2005). An unbalanced spatial panel data approach to US state tax competition. *Economics Letters 88*, 329–35.

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

Elhorst, J. P. (2010). Spatial panel data models. In M. M. Fischer and A. Getis (Eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, 377–407. Berlin: Springer.

Ertur, C. and W. Koch (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics 22*, 1033–62.

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

Foucault, M., T. Madies and S. Paty (2008). Public spending interactions and local politics: empirical evidence from French municipalities. *Public Choice 137*, 57–80.

Franzese, R. J. and J. C. Hays (2007). Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. *Political Analysis 15*, 140–64.

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

Jenish, N. and I. R. Prucha (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics 150*, 86–98.

Jenish, N. and I. R. Prucha (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics 170*, 178–90.

Islam, N. (1995). Growth empirics: a panel data approach. *Quarterly Journal of Economics 110*, 1127–170.

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

Keen, M. and M. Marchand (1997). Fiscal competition and the pattern of public spending. *Journal of Public Economics 66*, 33–53.

Kelejian, H. H. and G. Piras (2014). Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics 46*, 140–49.

Keller, W. and C. H. Shiue (2007). The origin of spatial interaction. *Journal of Econometrics 140*, 304–32.

Korniotis, G. M. (2010). Estimating panel models with internal and external habit formation. *Journal of Business and Economic Statistics 28*, 145–58.

Lee, L. F. and J. Yu (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics 154*, 165–85.

Lee, L. F. and J. Yu (2012a). Spatial panels: random random components vs. fixed effects. *International Economic Review 53*, 1369–412.

Lee, L. F. and J. Yu (2012b). QML estimation of spatial dynamic panel data models with time varying spatial weights matrices. *Spatial Economic Analysis 7*, 31–74.

Lee, L. F. and J. Yu (2015). Spatial panel data models. In B. Baltagi (Eds.), *Oxford Handbook of Panel Data*, 363–401. Oxford: Oxford University Press.

Moscone, F., M. Knapp and E. Tosetti (2007). Mental health expenditure in England: a spatial panel approach. *Journal of Health Economics 26*, 842–64.

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

Mutl, J. and M. Pfaffermayr (2011). The Hausman test in a Cliff and Ord panel model. *Econometrics Journal 14*, 48–76.

Qu, X. and L. F. Lee (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics 184*, 209–32.

Qu, X., L. F. Lee and J. Yu (2015). QML Estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. Working paper, Shanghai Jiao Tong University.

Redoano, M. (2007). Fiscal interactions among European countries: does the EU matter? Working paper, University of Warwick.

Revelli, F. (2001). Spatial patterns in local taxation: taxmimicking or errormimicking? *Applied Economics 33*, 1101–107.

Shelton, C. A. (2007). The size and composition of government expenditure. *Journal of Public Economics 91*, 2230–260.

Su, L. and Z. Yang (2015). QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics 185*, 230–58.

Tao, J. (2005). Analysis of local school expenditures in a dynamic fame. Working paper, Shanghai University of Finance and Economics.

Yu, J., R. de Jong and L. F. Lee (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, *146*, 118–34.

Yu, J., R. de Jong and L. F. Lee (2012). Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration. *Journal of Econometrics*, *167*, 16–37.

Yu, Y., L. Zhang, F. Li and X. Zheng (2013). Strategic interaction and the determinants of public health expenditures in China: a spatial panel perspective. *Annals of Regional Science*, *50*, 203–21.

APPENDIX A: SOME USEFUL CLAIMS

The first three claims are in, or are implied by results in, Jenish and Prucha (2012). Claim A.1 is trivial. Claim A.2 is a law of large numbers (LLN) and Claim A.3 is a central limit theorem (CLT).

CLAIM A.1. *For any distance ρ, there are at most $c_5\rho^{d_0}$ points in $B_i(\rho)$ and at most $c_4\rho^{d_0-1}$ points in the space $B_i(\rho+1)\backslash B_i(\rho)$, where c_4 and c_5 are positive constants.*

CLAIM A.2. *Under Assumption 2.1, if the random field* $\{T_{i,n}, i \in D_n, n \geq 1\}$ *is* L_1*-NED, the base* $\{\varsigma_{i,n}\}$ *are i.i.d., and* $\{T_{i,n}\}$ *are uniformly* L_p *bounded for some* $p > 1$*, then* $(1/n)\sum_{i=1}^{n}(T_{i,n} - ET_{i,n}) \xrightarrow{L_1} 0$.

CLAIM A.3. *Let* $\{T_{i,n}, i \in D_n, n \geq 1\}$ *be a random field that is* L_2*-NED on an i.i.d. random field* ς*. Then,* $\sigma_n^{-1}\sum_{i=1}^{n}(T_{i,n} - ET_{i,n}) \xrightarrow{d} N(0,1)$ *if Assumption 2.1 and the following conditions are met: (a)* $\{T_{i,n}, i \in D_n, n \geq 1\}$ *is uniformly* $L_{2+\delta}$*-bounded for some* $\delta > 0$*; (b)* $\inf_n (1/n)\sigma_n^2 > 0$ *where* $\sigma_n^2 = \mathrm{Var}(\sum_{i=1}^{n} T_{i,n})$*; (c) NED coefficients satisfy* $\sum_{r=1}^{\infty} r^{d_0-1}\varphi(r) < \infty$*; (d) NED scaling factors satisfy* $\sup_{n,i\in D} d_{i,n} < \infty$.

For the NED properties in this paper, let $\mathcal{F}_{i,n}(s)$ be the σ-field generated by $\varsigma_{j,n}$ centred at location i with a radius of s, where $\varsigma_{i,n} = (\varepsilon_{in,-t_0}, \ldots, \varepsilon_{i,nT}, \xi_{in,-t_0}, \ldots \xi_{i,nT})$. We note that $\varsigma_{i,n}$ is purely determined by the location i, independent of error terms associated with any other places. The following claim indicates that two NED processes are also an NED process. Its proof can be found in the online Appendix.

CLAIM A.4. *If, for any* $t = 1, \ldots T$, $\sup_{i,n,t} ||Q_{1i,nt} - E[Q_{1i,nt}|\mathcal{F}_{i,n}(s)]||_4 \leq C_1\varphi_1(s)$ *and* $\sup_{i,n,t} ||Q_{2i,nt} - E[Q_{2i,nt}|\mathcal{F}_{i,n}(s)]||_4 \leq C_2\varphi_2(s)$*, with* $\max_{i,n,t}(||Q_{1i,nt}||_4, ||Q_{2i,nt}||_4) \leq C$*, then for any* $t_1, t_2 = 1, \ldots T$,

$$\sup_{i,n,t_1,t_2} ||Q_{1i,nt_1}Q_{2i,nt_2} - E[Q_{1i,nt_1}Q_{2i,nt_2}|\mathcal{F}_{i,n}(s)]||_2 \leq 2C(C_1 + C_2)\varphi(s),$$

where $\varphi(s) = \max(\varphi_1(s), \varphi_2(s))$.

APPENDIX B: NED PROPERTIES UNDER ASSUMPTION 3.2

This appendix establishes NED properties for some relevant statistics from our model.

Some claims are generalizations of cross-sectional results established in Qu and Lee (2015) to the panel setting with small T. Their proofs can be straightforward extensions from those in Qu and Lee (2015) and so are omitted here.[6]

CLAIM B.1. *Under Assumptions 2.1, 3.1 and 3.2(a), for any n and positive integer q,* $||\prod_{l=1}^{q} W_{n,t-l}||_1 \leq (q-1)c_u(t_0+T)Kc_w^{q-1} + c_u c_w^{q-1} \leq qc_u(t_0+T)Kc_w^{q-1}$*, where* $c_u = \sup_{n,t}||W_{nt}||_1$ *and* $c_w = \sup_{n,t}||W_{nt}||_\infty$.

CLAIM B.2. *Suppose* W_{nt} *is an* $n \times n$ *square matrix which can be decomposed into the sum of two* $n \times n$ *matrices such that* $W_{nt} = A_{nt} + B_{nt}$*. Denote* $|A|_{\max} = \sup_t \max\{|a_{ij,nt}| : i, j = 1, \ldots, n\}$*. Then, for any positive integer k and any* $i, j = 1, \ldots, n$,

$$\left(\prod_{l=1}^{k} W_{n,t-l} - \prod_{l=1}^{k} B_{n,t-l}\right)_{ij} \leq |A|_{\max}\sum_{m=0}^{k-1}||\prod_{l_1=1}^{m} B_{n,t-l_1}||_\infty \cdot ||\prod_{l_2=m+2}^{k} W_{n,t-l_2}||_1.$$

CLAIM B.3. *Let* $Q_{i,nt}(m)$ *be the ith element of the vector* $(\prod_{l=1}^{m} W_{n,t-l})\varsigma_{n,t-m}^{*}a$*, where* $\varsigma_{i,nt}^{*} = f_i(\varsigma_{i,nt}, X_{nt})$ *with* $\varsigma_{nt} = (\varepsilon_{nt}, \xi_{nt})$ *is a vector-valued function and a is any conformable vector of constants. Under Assumptions 2.1, 3.1 and 3.2(a), suppose* $\sup_{i,n,t}||\varsigma_{i,nt}^{*}||_p < \infty$*, then* $\sup_{i,n,t}||Q_{i,nt}(m)||_p \leq m^{c_3d_0+2}c_w^m C_{ap}$ *and* $\sup_{i,n,t}||Q_{i,nt}(m) - E[Q_{i,nt}(m)|\mathcal{F}_{i,n,T}(s)]||_p \leq C_{ap}c_w^m m^{3+c_3d_0}s^{(2-c_3)d_0}$ *with* C_{ap} *being a finite constant.*

CLAIM B.4. *Let* $g_{i,nt}(m) = e_{i,n}'(\prod_{l=1}^{m} G_{n,t-l}(\lambda_1))\varsigma_{n,t-m}^{*}a$*, where* ς_{nt}^{*} *and a are the same as in Claim B.3, and* $G_{nt}(\lambda_1) = W_{nt}(I_n - \lambda_1 W_{nt})^{-1}$*. Under Assumptions 2.1, 3.1 and 3.2(a), suppose* $\sup_{i,n,t}||\varsigma_{i,nt}^{*}||_p < \infty$*, then* $\sup_{i,n,t}||g_{i,nt}(m)||_p < \infty$ *and* $\sup_{i,n,t}||g_{i,nt}(m) - E[g_{i,nt}(m)|\mathcal{F}_{i,n}(s)]||_p \leq C_{apm}s^{(2-c_3)d_0}$ *with* C_{apm} *being a finite constant.*

CLAIM B.5. *Under Assumptions 2.1 and 3.2(b), if the i, jth element of* $\prod_{l=1}^{m} W_{n,t-l}$ *is not zero, then it must be* $\rho_{ij} \leq m\rho_c$.

[6] For interested readers, detailed proofs are available in online Appendix E.

CLAIM B.6. *Under Assumptions 2.1, 3.1 and 3.2(b), for any positive integer q,* $\sup_{n,t} ||\prod_{l=1}^{q} W_{n,t-l}||_1 \leq c_w^q c_5 (q\rho_c)^{d_0}$.

CLAIM B.7. *For any positive integer p and $0 < q < 1$, if $s \geq p/(-\ln q) + 1$, then there exists a finite constant c such that $\sum_{l=[s]} l^p q^l < cs^p q^s$, where $[s]$ denotes the largest integer less than or equal to s.*

CLAIM B.8. *Let $Q_{i,nt}(m) = e'_{i,n}(\prod_{l=1}^{m} W_{n,t-l})\varsigma^*_{n,t-m}a$, where ς^*_{nt} and a are the same as in Claim B.3. Under Assumptions 2.1, 3.1 and 3.2(b), suppose $\sup_{i,n,t} ||\varsigma^*_{i,nt}||_p < \infty$, then $\sup_{i,n,t} ||Q_{i,nt}(m)||_p \leq C_{ap} m^{d_0} c_w^m$ and $\sup_{i,n,t} ||Q_{i,n}(m) - E[Q_{i,n}(m)|\mathcal{F}_{i,n}(s)]||_p \leq C_{ap1}\varphi(s)$ with C_{ap} and C_{ap1} being positive constants; $\varphi(s) = 1$ if $s \leq m\rho_c$ and $\varphi(s) = 0$ if $s > m\rho_c$.*

CLAIM B.9. *Denote $g_{i,nt}(m) = e'_{i,n}(\prod_{l=1}^{m} G_{n,t-l}(\lambda_1))\varsigma^*_{n,t-m}a$, where $G_{nt}(\lambda_1)$, ς^*_{nt}, and a are the same as in Claim B.3. Under Assumptions 2.1, 3.1 and 3.2(b), suppose $\sup_{i,n,t} ||\varsigma^*_{i,nt}||_p < \infty$, then $\sup_{i,n,t} ||g_{i,nt}(m)||_p < \infty$ and $\sup_{i,n,t} ||g_{i,nt}(m) - E[g_{i,nt}(m)|\mathcal{F}_{i,n}(s)]||_p \leq C_{apm}\varphi(s)$ with C_{apm} being a finite constant; $\varphi(s) = 1$ if $s \leq m\rho_c$ and $\varphi(s) = s^{d_0+m}|\lambda c_w|^{s/\rho_c}$ if $s > m\rho_c$.*

Denote $B_{nt}^{(h)} = \prod_{k=1}^{h}(\rho_0 S_{n,t-k}^{-1} + \lambda_{20} G_{n,t-k})$ with $B_{nt}^{(0)} = I_n$, $S_{n,t}(\lambda_1) = I_n - \lambda_1 W_{nt}$, $S_{nt} = S_{nt}(\lambda_{10})$ and $G_{nt} = W_{nt} S_{nt}^{-1}$. Then, this $B_{nt}^{(h)}$ is directly linked to our SDPD model. The following two claims have new features that are not covered in Qu and Lee (2015) because of this $B_{nt}^{(h)}$.

CLAIM B.10. *Under Assumptions 2.1, 2.2, 3.1 and 3.2(a), for any finite H, we have $\sup_{i,n,t} ||e'_{i,n} G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h}||_p < \infty$ and*

$$\sup_{i,n,t} ||e'_{i,n} G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h} - E[e'_{i,n} G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h} |\mathcal{F}_{i,n}(s)]||_p \leq d\varphi(s)$$

such that $\lim_{s\to\infty} \varphi(s) = 0$.

Proof: First, we show that $\sup_{i,n,t,H} ||e'_{i,n} G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h} - E[e'_{i,n} G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h}|\mathcal{F}_{i,n}(s)]||_p \leq d_1\varphi(s)$ holds for any $h = 0, \ldots H$. Denote

$$\prod_{k=1}^{h\backslash l} A_{n,k} = \prod_{k=1}^{l-1} A_{n,k} \prod_{k=l+1}^{h} A_{n,k},$$

where $1 \leq l \leq h$,

$$\prod_{k=1}^{h\backslash l_1,l_2} A_{n,k} = \prod_{k=1}^{l_1-1} A_{n,k} \prod_{k=l_1+1}^{l_2-1} A_{n,k} \prod_{k=l_2+1}^{h} A_{n,k},$$

where $1 \leq l_1 \leq l_2 \leq h$, and so on. As

$$\begin{aligned} B_{nt}^{(h)} &= \prod_{k=1}^{h}(\rho_0 S_{n,t-k}^{-1} + \lambda_{20} G_{n,t-k}) = \prod_{k=1}^{h}\Big(\rho_0 I_n + (\rho_0\lambda_{10} + \lambda_{20}) G_{n,t-k}\Big) \\ &= (\rho_0\lambda_{10} + \lambda_{20})^h \prod_{k=1}^{h} G_{n,t-k} + (\rho_0\lambda_{10} + \lambda_{20})^{h-1}\rho_0 \sum_{l=1}^{h} \prod_{k=1}^{h\backslash l} G_{n,t-k} \\ &\quad + (\rho_0\lambda_{10} + \lambda_{20})^{h-2}\rho_0^2 \sum_{l_1=1}^{h} \sum_{l_2\neq l_1} \prod_{k=1}^{h\backslash l_1,l_2} G_{n,t-k} \\ &\quad + \cdots + (\rho_0\lambda_{10} + \lambda_{20})\rho_0^{h-1} \sum_{k=1}^{h} G_{n,t-k} + \rho_0^h I_n, \end{aligned}$$

we have

$$
\begin{aligned}
&\sup_{i,n,t,H} ||e_{i,n}' G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h} - E[e_{i,n}' G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h} | \mathcal{F}_{i,n}(s)]||_p \\
&\quad \leq |\rho_0\lambda_{10} + \lambda_{20}|^h C_{ap(h+1)} s^{(2-c_3)d_0} + h|\rho_0\lambda_{20} + \lambda_{20}|^{h-1} |\rho_0| C_{aph} s^{(2-c_3)d_0} \\
&\qquad + \binom{h}{2} |\rho_0\lambda_{20} + \lambda_{20}|^{h-2} |\rho_0|^2 C_{ap(h-1)} s^{(2-c_3)d_0} \\
&\qquad + \cdots + \binom{h}{h-1} |\rho_0\lambda_{20} + \lambda_{20}| \cdot |\rho_0|^{h-1} C_{ap(2)} s^{(2-c_3)d_0} + |\rho_0|^h C_{ap(1)} s^{(2-c_3)d_0} \\
&\quad \leq \sum_{q=0}^{h} \binom{h}{q} \cdot |\rho_0|^q |\rho_0\lambda_{20} + \lambda_{20}|^{h-q} \cdot C_{apH} s^{(2-c_3)d_0} \\
&\quad = (|\rho_0| + |\rho_0\lambda_{10} + \lambda_{20}|)^h C_{apH} s^{(2-c_3)d_0}, \qquad \text{(B.1)}
\end{aligned}
$$

where $C_{apH} = \max_{0\leq h\leq H} C_{aph} < \infty$ is a finite constant not depending on n. The first inequality is from Claim B.4 that

$$
\begin{aligned}
&\sup_{i,n,t} ||e_{i,n}' \left(\prod_{l=1}^{m} G_{n,t-l}(\lambda_1) \right) \varsigma_{n,t-m}^* a - E[e_{i,n}' \left(\prod_{l=1}^{m} G_{n,t-l}(\lambda_1) \right) \varsigma_{n,t-m}^* a | \mathcal{F}_{i,n,T}(s)]||_p \\
&\qquad \leq C_{apm} s^{(2-c_3)d_0}.
\end{aligned}
$$

As (B.1) holds for any $h = 0, \ldots H$, we have

$$
\begin{aligned}
&\sup_{i,n,t} ||e_{i,n}' G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h} - E[e_{i,n}' G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h} | \mathcal{F}_{i,n}(s)]||_p \\
&\quad \leq \sup_{i,n,t} \sum_{h=0}^{H} ||e_{i,n}' G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h} - E[e_{i,n}' G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h} | \mathcal{F}_{i,n}(s)]||_p \\
&\quad \leq C_{apH} s^{(2-c_3)d_0} \sum_{h=0}^{H} (|\rho_0| + |\rho_0\lambda_{20} + \lambda_{20}|)^h \leq C_{p,H} s^{(2-c_3)d_0},
\end{aligned}
$$

for some finite constant $C_{p,H}$. Similarly,

$$
\begin{aligned}
\sup_{i,n,t} ||e_{i,n}' G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h}||_p &\leq \sup_{i,n,t} \sum_{h=0}^{H} ||e_{i,n}' G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h}||_p \\
&\leq C_{apH} \sum_{h=0}^{H} (|\rho_0| + |\rho_0\lambda_{20} + \lambda_{20}|)^h \leq C_{p,H}.
\end{aligned}
$$

□

CLAIM B.11. *Under Assumptions 2.1, 2.2, 3.1 and 3.2(b), for any finite H, we have $\sup_{i,n,t} ||e_{i,n}' G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h}||_p < \infty$ and*

$$
\sup_{i,n,t} ||e_{i,n}' G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h} - E[e_{i,n}' G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h} | \mathcal{F}_{i,n}(s)]||_p \leq d\varphi(s),
$$

such that $\varphi(s) = 1$ if $s \leq (H+1)\rho_c$ and $\varphi(s) = s^{d_0+H} |\lambda_1 c_w|^{s/\rho_c}$ if $s > (H+1)\rho_c$.

Proof: From Claim B.9,

$$\sup_{i,n,t} ||e'_{i,n} G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h}||_p \leq \sum_{h=0}^{H} ||e'_{i,n} G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h}||_p$$

$$\leq C_{(H+1)} \sum_{h=0}^{H} (|\rho_0| + |\rho_0 \lambda_{20} + \lambda_{20}|)^h \leq C_{p,H+1},$$

for some constants $C_{(H+1)}$ and $C_{p,H+1}$.

If $s \leq (h+1)\rho_c \leq (H+1)\rho_c$, then

$$\sup_{i,n,t} ||e'_{i,n} G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h} - E[e'_{i,n} G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h} | \mathcal{F}_{i,n,T}(s)]||_p$$

$$\leq C|\rho_0\lambda_{10} + \lambda_{20}|^h + hC|\rho_0\lambda_{10} + \lambda_{20}|^{h-1}|\rho_0| + \binom{h}{2} |\rho_0\lambda_{20} + \lambda_{20}|^{h-2}|\rho_0|^2 C$$

$$+ \cdots + \binom{h}{h-1} |\rho_0\lambda_{10} + \lambda_{20}||\rho_0|^{h-1} C + |\rho_0|^h C$$

$$\leq \sum_{q=0}^{h} \binom{h}{q} |\rho_0|^q |\rho_0\lambda_{10} + \lambda_{20}|^{h-q} C = (|\rho_0| + |\rho_0\lambda_{10} + \lambda_{20}|)^h C$$

and

$$\sup_{i,n,t} ||e'_{i,n} G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h} - E[e'_{i,n} G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h} | \mathcal{F}_{i,n}(s)]||_p$$

$$\leq \sum_{h=0}^{H} (|\rho_0| + |\rho_0\lambda_{10} + \lambda_{20}|)^h C = C_H$$

for some finite constants C and C_H.

If $s > (H+1)\rho_c$, then we have

$$\sup_{i,n,t} ||e'_{i,n} G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h} - E[e'_{i,n} G_{nt} B_{nt}^{(h)} \varsigma_{1n,t-h} | \mathcal{F}_{i,n}(s)]||_p$$

$$\leq |\rho_0\lambda_{10} + \lambda_{20}|^h C s^{d_0+h} |\lambda_1 c_w|^{s/\rho_c} + h|\rho_0\lambda_{10} + \lambda_{20}|^{h-1}|\rho_0| C s^{d_0+h-1} |\lambda_1 c_w|^{s/\rho_c}$$

$$+ \binom{h}{2} |\rho_0\lambda_{10} + \lambda_{20}|^{h-2}|\rho_0|^2 C s^{d_0+h-2} |\lambda_1 c_w|^{s/\rho_c} + \cdots + \binom{h}{h-1} |\rho_0\lambda_{10}$$

$$+ \lambda_{20}| \cdot |\rho_0|^{h-1} C s^{d_0+1} |\lambda_1 c_w|^{s/\rho_c} + |\rho_0|^h C s^{d_0} |\lambda_1 c_w|^{s/\rho_c}$$

$$\leq (|\rho_0| + |\rho_0\lambda_{10} + \lambda_{20}|)^h C s^{d_0+h} |\lambda_1 c_w|^{s/\rho_c} \tag{B.2}$$

and

$$\sup_{i,n,t} ||e'_{i,n} G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h} - E[e'_{i,n} G_{nt} \sum_{h=0}^{H} B_{nt}^{(h)} \varsigma_{1n,t-h} | \mathcal{F}_{i,n}(s)]||_p$$

$$\leq \sum_{h=0}^{H} (|\rho_0| + |\rho_0\lambda_{10} + \lambda_{20}|)^h C s^{d_0+h} |\lambda_1 c_w|^{s/\rho_c} \leq C_H s^{d_0+H} |\lambda_1 c_w|^{s/\rho_c}.$$

□

APPENDIX C: PROOFS OF THE MAIN RESULTS

Proof of Claim 3.1: We want to verify that $q_{1n}(T) = [\Delta X_{2n}(T-1), \Delta X_{2n}(T))]$ and $q_{2n}(T) = [\Delta(WX)_n(T), \Delta(WX)_n(T-1), \Delta X_{1n}(T-1), \Delta X_{1n}(T), \Delta\widehat{\varepsilon}_n(T)]$ are valid instrument variables. First, we show that

$$\frac{1}{n}q_{2n}'T_n - E\left[\frac{1}{n}q_{2n}'T_n\right],$$

$$\frac{1}{n}q_{2n}'q_{2n} - E\left[\frac{1}{n}q_{2n}'q_{2n}\right],$$

$$\frac{1}{n}q_{1n}'K_n - E\left[\frac{1}{n}q_{1n}'K_n\right],$$

$$\frac{1}{n}q_{2n}'K_n - E\left[\frac{1}{n}q_{2n}'K_n\right],$$

$$\frac{1}{n}q_{2n}'\Delta\xi_n(T) - E\left[\frac{1}{n}q_{2n}'\Delta\xi_n(T)\right]$$

and

$$\frac{1}{n}q_{1n}'\Delta\varepsilon_n(T) - E\left[\frac{1}{n}q_{1n}'\Delta\varepsilon_n(T)\right]$$

are all $o_p(1)$. As T is small and the data starting period $-t_0$ is finite, all these terms can be expressed as the finite summation of a general form $(1/n)U_{1nt_1,h_1}'U_{2nt_2,h_2}$. Also, the LLN of this general term is proven by Lemma C.1.

Then, we need to check properties of these limiting expectations. Apparently,

$$E[q_{2n}'\Delta\xi_n(T)] = E[q_{2n}'E[\Delta\xi_n(T)|\varepsilon_n(T)]] = 0 \quad \text{and} \quad E[q_{1n}'\Delta\varepsilon_n(T)] = 0.$$

It remains to check that $\lim_{n\to\infty}(1/n)E[q_{1n}'q_{1n}]$ and $\lim_{n\to\infty}(1/n)E[q_{2n}'q_{2n}]$ are non-singular, that $\lim_{n\to\infty}(1/n)E[q_{2n}'T_n]$ and $\lim_{n\to\infty}(1/n)E[q_{1n}'K_n]$ have full column rank, and that $\lim_{n\to\infty} E[(1/n)q_{2n}'K_n] = O(1)$. In our setting, the reduced form is

$$Z_{nt} = \sum_{h=0}^{t+t_0-1} \rho_{20}^h(X_{2n,t-h}\Gamma_0 + \mathbf{d}_{n0} + \varepsilon_{n,t-h}) + \rho_{20}^{t+t_0}Z_{n,-t_0};$$

$$Y_{nt} = S_{nt}^{-1}\left(\sum_{h=0}^{t+t_0-1} B_{nt}^{(h)}(X_{1n,t-h}\beta_0 + \mathbf{c}_{n0} + \varepsilon_{n,t-h}\delta_0 + \xi_{n,t-h}) + B_{nt}^{(t+t_0-1)}(\rho_{10}I_n + \lambda_{20}W_{n,-t_0})Y_{n,-t_0}\right),$$

where $B_{nt}^{(h)} = \prod_{k=1}^{h}(\rho_0 S_{n,t-k}^{-1} + \lambda_{20}G_{n,t-k})$ with $B_{nt}^{(0)} = I_n$, $S_{n,t}(\lambda_1) = I_n - \lambda_1 W_{nt}$, $S_{nt} = S_{nt}(\lambda_{10})$ and $G_{nt} = W_{nt}S_{nt}^{-1}$, which are the same as in Appendix B, and $Y_{n,-t_0}$ and $Z_{n,-t_0}$ can be treated, either exogenous or endogenous, as

$$Y_{n,-t_0} = (I_n - \lambda_1^o W_{n,-t_0})^{-1}(X_{1n,-t_0}\beta^o + \mathbf{c}_{n0}^o + \varepsilon_{n,-t_0}\delta_0 + \xi_{n,-t_0})$$

and

$$Z_{n,-t_0} = X_{2n,-t_0}\Gamma_0^o + \mathbf{d}_{n0}^o + \varepsilon_{n,-t_0}.$$

Then

$$E[\Delta Z_{n,t-1}] = X_{2n,t-1}\Gamma_0 + (\rho_{20}-1)\sum_{h=0}^{t+t_0-3}\rho_{20}^h(X_{2n,t-2-h}\Gamma_0) + \rho_{20}^{t+t_0-2}(\mathbf{d}_{n0} + (\rho_{20}-1)Z_{n,-t_0}^{(x)}),$$

where $Z_{n,-t_0}^{(x)} = X_{2n,-t_0}\Gamma_0^o + \mathbf{d}_{n0}^o$ or $Z_{n,-t_0}^{(x)} = Z_{n,-t_0}$ if $Z_{n,-t_0}$ is exogenous. Let $E[A_{nt}|\varepsilon]$ be the expectation of A_{nt} conditional on $\vec{\varepsilon}_t$ from $\varepsilon_{n,-t_0}$ to ε_{nt}. Then, $E[\Delta(Y)_{nt}|\vec{\varepsilon}_t] = E[Y_{nt}|\vec{\varepsilon}_t] - E[Y_{n,t-1}|\vec{\varepsilon}_t]$ and $E[\Delta(WY)_{nt}|\vec{\varepsilon}_t] = W_{nt}E[Y_{nt}|\vec{\varepsilon}_t] - W_{n,t-1}E[Y_{n,t-1}|\vec{\varepsilon}_t]$ with

$$E[Y_{nt}|\vec{\varepsilon}_t] = S_{nt}^{-1}\left(\sum_{h=0}^{t+t_0-1}B_{nt}^{(h)}(X_{1n,t-h}\beta_0 + \mathbf{c}_{n0} + \varepsilon_{n,t-h}\delta_0) + B_{nt}^{(t+t_0-1)}(\rho_{10}I_n + \lambda_{20}W_{n,-t_0})Y_{n,-t_0}^{(x)}\right),$$

where $Y_{n,-t_0}^{(x)} = (I_n - \lambda_1^o W_{n,-t_0})^{-1}(X_{1n,-t_0}\beta^o + \mathbf{c}_{n0}^o + \varepsilon_{n,-t_0}\delta_0)$ or $Y_{n,-t_0}^{(x)} = Y_{n,-t_0}$ if $Y_{n,-t_0}$ is exogenous. Let $M_n = I_{n(T-2)} - K_n(K_n'P_{1n}K_n)^{-1}K_n'P_{1n}$. Then, $\Delta\widehat{\varepsilon}_n(T) = M_n\Delta\varepsilon_n(T)$ and

$$\begin{aligned}
\frac{1}{n}E[q_{1n}'q_{1n}] &= \frac{1}{n}[\Delta X_{2n}(T-1), \Delta X_{2n}(T)]'[\Delta X_{2n}(T-1), \Delta X_{2n}(T)],\\
\frac{1}{n}E[q_{1n}'K_n] &= \frac{1}{n}[\Delta X_{2n}(T-1), \Delta X_{2n}(T)]'[E[\Delta Z_n(T-1)], \Delta X_{2n}(T)],\\
\frac{1}{n}E[q_{2n}'q_{2n}] &= \frac{1}{n}E[\Delta(WX)_n(T), \Delta(WX)_n(T-1), \Delta X_{1n}(T-1), \Delta X_{1n}(T), M_n\Delta\varepsilon_n(T)]'\\
&\quad\times[\Delta(WX)_n(T), \Delta(WX)_n(T-1), \Delta X_{1n}(T-1), \Delta X_{1n}(T), M_n\Delta\varepsilon_n(T)],\\
\frac{1}{n}E[q_{2n}'T_n] &= \frac{1}{n}E[\Delta(WX)_n(T), \Delta(WX)_n(T-1), \Delta X_{1n}(T-1), \Delta X_{1n}(T), M_n\Delta\varepsilon_n(T)]'\\
&\quad\times[E[\Delta(WY)_n(T)|\vec{\varepsilon}_t], E[\Delta(WY)_n(T-1)|\vec{\varepsilon}_t], E[\Delta Y_n(T-1)|\vec{\varepsilon}_t], \Delta X_{1n}(T), M_n\Delta\varepsilon_n(T)].
\end{aligned}$$

When the following conditions hold, we can verify that this example of q_{1n} and q_{2n} are valid instruments:

(a) $\Delta X_{2n}(T-1)$ and $\Delta X_{2n}(T)$ are linearly independent such that the limit of $[\Delta X_{2n}(T-1), \Delta X_{2n}(T)][\Delta X_{2n}(T-1), \Delta X_{2n}(T)]'/n$ has full column rank;
(b) the limit of $[\Delta X_{2n}(T-1), \Delta X_{2n}(T)]'[E[\Delta Z_n(T-1)], \Delta X_{2n}(T)]/n$ has full column rank;
(c) the limit of $E[[\Delta(WX)_n(T), \Delta(WX)_n(T-1), M_n\Delta\varepsilon_n(T)]'\cdot[\Delta(WX)_n(T), \Delta(WX)_n(T-1), M_n\Delta\varepsilon_n(T)]]/n$ is non-singular;
(d) the limit of $E[[\Delta(WX)_n(T), \Delta(WX)_n(T-1), \Delta X_{1n}(T-1), \Delta X_{1n}(T), M_n\Delta\varepsilon_n(T)]'\cdot[E[\Delta(WY)_n(T)|\vec{\varepsilon}_t], E[\Delta(WY)_n(T-1)|\vec{\varepsilon}_t], E[\Delta Y_n(T-1)|\vec{\varepsilon}_t], \Delta X_{1n}(T), M_n\Delta\varepsilon_n(T)]]\ /n$ has full column rank. □

LEMMA C.1. *Denote $U_{1nt,h} = A_{1nt}B_{nt}^{(h)}(\theta)\varsigma_{1n,t-h}$ and $U_{2nt,h} = A_{2nt}B_{nt}^{(h)}(\theta)\varsigma_{2n,t-h}$, where A_{1nt} and A_{2nt} can be either S_{nt}^{-1}, G_{nt} or I_n, and $B_{nt}^{(h)}(\theta) = \prod_{k=1}^{h}(\rho S_{n,t-k}^{-1}(\lambda_1) + \lambda_2 G_{n,t-k}(\lambda_1))$. Under Assumptions 2.1–3.2 and data starting from $-t_0$ period, for any $h_1 = 0,\ldots,t_1+t_0$ with $t_1 = 1,\ldots,T$ and $h_2 = 0,\ldots,t_2+t_0$ with $t_2 = 1,\ldots,T$,*

$$\frac{1}{n}U_{1nt_1,h_1}'U_{2nt_2,h_2} = O_p(1) \quad \textit{and} \quad \frac{1}{n}\Big[U_{1nt_1,h_1}'U_{2nt_2,h_2} - E[U_{1nt_1,h_1}'U_{2nt_2,h_2}]\Big] = o_p(1).$$

Proof: Let $q_{i,n} = e'_{i,n} A_{1nt_1} B_{nt_1}^{(h_1)} \varsigma_{1n,t_1-h_1} e'_{i,n} A_{2nt_2} B_{nt_2}^{(h_2)} \varsigma_{2n,t_2-h_2}$. Then we have

$$\frac{1}{n}\Big[U'_{1nt_1,h_1} U_{2nt_2,h_2} - E[U'_{1nt_1,h_1} U_{2nt_2,h_2}]\Big] = \frac{1}{n}\sum_{i=1}^{n}[q_{i,n} - E[q_{i,n}]].$$

Under Assumption 3.2(a), $\sup_{i,n,t,h} ||u_{1i,nt,h} - E[u_{1i,nt,h}|\mathcal{F}_{i,n}(s)]||_4 \le C_{4,T+t_0} s^{(2-c_3)d_0}$ and $\sup_{i,n,t,h} ||u_{2i,nt,h} - E[u_{2i,nt,h}|\mathcal{F}_{i,n}(s)]||_4 \le C_{4,T+t_0} s^{(2-c_3)d_0}$. Together with $\max_{i,n,t,h}(||u_{1i,nt,h}||_4, ||u_{2i,nt,h}||_4) \le C_1$ from Claim B.10, we have $||q_{i,n} - E[q_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \le 16 C_1 C_{4,T+t_0} s^{(2-c_3)d_0}$ from Claim A.4 and $||q_{i,n}||_2 \le 4C_1^2 T^2$.

Under Assumption 3.2(b), we have $\sup_{i,n,t,h} ||u_{1i,nt,h} - E[u_{1i,nt,h}|\mathcal{F}_{i,n}(s)]||_4 \le C\varphi(s)$ and $\sup_{i,n,t,h} ||u_{2i,nt,h} - E[u_{2i,nt,h}|\mathcal{F}_{i,n}(s)]||_4 \le C\varphi(s)$ with $\varphi(s) = 1$ if $s \le (T+t_0)\rho_c$ and $\varphi(s) = s^{d_0+T+t_0}|\lambda_1 c_w|^{s/\rho_c}$ if $s > (T+t_0)\rho_c$. Also, $\max_{i,n,t,h}(||u_{1i,nt,h}||_4, ||u_{2i,nt,h}||_4) \le C_1$ from Claim B.11, $||q_{i,n} - E[q_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \le 16 C_1 s^{d_0+T+t_0}|\lambda_1 c_w|^{s/\rho_c}$ from Claim A.4 and $||q_{i,n}||_2 \le 4C_1^2 T^2$. Therefore, $(1/n)\sum_{i=1}^{n}[q_{i,n} - E[q_{i,n}]] \xrightarrow{L_1} 0$ as conditions of the LLNs under the spatial NED in Jenish and Prucha (2012) hold. □

Proof of Theorem 3.1: As $\Delta\widehat{\varepsilon}_n(T) = \Delta\varepsilon_n(T) - K_n(K'_n P_{1n} K_n)^{-1} K'_n P_{1n} \Delta\varepsilon_n(T)$, we have $T_n - \widehat{T}_n = [0, \Delta\varepsilon_n(T) - \Delta\widehat{\varepsilon}_n(T)] = [0, K_n(K'_n P_{1n} K_n)^{-1} K'_n P_{1n} \Delta\varepsilon_n(T)]$ and

$$\begin{aligned}\widehat{\kappa} - \kappa_0 &= (\widehat{T}'_n P_{2n} \widehat{T}_n)^{-1} \widehat{T}'_n P_{2n} \Delta Y_n(T) - \kappa_0 \\ &= (\widehat{T}'_n P_{2n} \widehat{T}_n)^{-1} \widehat{T}'_n P_{2n} \Delta\xi_n(T) \\ &\quad + (\widehat{T}'_n P_{2n} \widehat{T}_n)^{-1} \widehat{T}'_n P_{2n} K_n (K'_n P_{1n} K_n)^{-1} K'_n P_{1n} \Delta\varepsilon_n(T)\delta_0.\end{aligned}$$

For consistency, we need $(q'_{2n} q_{2n}/n)^{-1}$, $(1/n) q'_{2n} K_n$, $(q'_{1n} q_{1n}/n)^{-1}$ to be $O_p(1)$, $(1/n) q'_{2n}\ \Delta\xi_n(T)$ and $(1/n) q'_{1n} \Delta\varepsilon_n(T)$ to be $o_p(1)$, and $\text{plim}(q'_{2n}\widehat{T}_n/n)$ and $\text{plim}(q'_{1n} K_n/n)$ to have full column rank. These are the conditions for the valid IV in Assumption 3.3, except for the term $q'_{2n}\widehat{T}_n/n$. Because $\widehat{T}_n = T_n - K_n(K'_n P_{1n} K_n)^{-1} K'_n P_{1n} \Delta\varepsilon_n(T)$,

$$\begin{aligned}\operatorname*{plim}_{n\to\infty} \frac{1}{n} q'_{2n}\widehat{T}_n &= \operatorname*{plim}_{n\to\infty} \frac{1}{n} q'_{2n} T_n - \operatorname*{plim}_{n\to\infty} \frac{1}{n} q'_{2n} K_n \operatorname*{plim}_{n\to\infty}\left(\frac{K'_n P_{1n} K_n}{n}\right)^{-1} \operatorname*{plim}_{n\to\infty} \frac{K'_n P_{1n} \Delta\varepsilon_n(T)}{n} \\ &= \operatorname*{plim}_{n\to\infty} \frac{1}{n} q'_{2n} T_n.\end{aligned}$$

By Assumption 3.3, $\operatorname*{plim}_{n\to\infty}(1/n) q'_{2n} T_n$ has the full column rank, so $\operatorname*{plim}_{n\to\infty}(1/n) q'_{2n}\widehat{T}_n$ also has the full column rank and hence $\text{plim}\,\widehat{\kappa} = \kappa_0$. □

Proof of Theorem 3.2: For the asymptotic distribution of $\widehat{\kappa}$, as

$$\begin{aligned}\widehat{\kappa} - \kappa_0 &= (\widehat{T}'_n P_{2n} \widehat{T}_n)^{-1} \widehat{T}'_n P_{2n} \Delta\xi_n(T) \\ &\quad + (\widehat{T}'_n P_{2n} \widehat{T}_n)^{-1} \widehat{T}'_n P_{2n} K_n (K'_n P_{1n} K_n)^{-1} K'_n P_{1n} \Delta\varepsilon_n(T)\delta_0,\end{aligned}$$

$\sqrt{n}(\widehat{\kappa} - \kappa_0)$ can be written as $R_n + o_p(1)$, where $R_n = (1/\sqrt{n})\sum_{i=1}^{n} r_{i,n}$ with $r_{i,n} = \sum_{t=3}^{T}(C_2 q_{2i,nt}\Delta\xi_{i,nt} + C_1 q_{1i,nt}\Delta\varepsilon_{i,nt})$ for some constant matrices C_1 and C_2. With the NED property of $r_{i,n}$, we have the CLT of our 2SIV. Furthermore, $E[\Delta\xi_n(T)\Delta\xi_n(T)'] = \sigma_{\xi 0}^2 \Psi_0 \otimes I_n$ and $E[\Delta\varepsilon_n(T)\delta_0\delta'_0\Delta\varepsilon_n(T)'] = (\delta'_0\Sigma_{\varepsilon 0}\delta_0)\Psi_0 \otimes I_n$. Hence,

$$\sqrt{n}(\widehat{\kappa} - \kappa_0) \xrightarrow{d} N(0, \Sigma_{\kappa 0}),$$

where

$$\Sigma_{\kappa 0} = \operatorname*{plim}_{n\to\infty} \frac{1}{n}(T'_n P_{2n} T_n)^{-1} T'_n P_{2n} \Pi_n P_{2n} T_n (T'_n P_{2n} T_n)^{-1},$$

with

$$\Pi_n = \sigma_{\xi 0}^2 \Psi_0 \otimes I_n + K_n (K_n' P_{1n} K_n)^{-1} K_n' P_{1n} (\delta_0' \Sigma_{\varepsilon 0} \delta_0 \Psi_0 \otimes I_n) P_{1n} K_n (K_n' P_{1n} K_n)^{-1} K_n'.$$

□

Proof of Claim 3.2: Because

$$\Delta\widehat{\varepsilon}_n(T) = \Delta Z_n(T) - K_n \begin{pmatrix} \widehat{\rho}_2 \\ \widehat{\Gamma} \end{pmatrix} = \Delta\varepsilon_n(T) + K_n \left[\begin{pmatrix} \rho_{20} \\ \Gamma_0 \end{pmatrix} - \begin{pmatrix} \widehat{\rho}_2 \\ \widehat{\Gamma} \end{pmatrix} \right],$$

we have

$$\frac{1}{n} \Delta\widehat{\varepsilon}_n(T)' \Delta\widehat{\varepsilon}_n(T) \xrightarrow{p} \lim_{n\to\infty} \frac{1}{n} \Delta\varepsilon_n(T)' \Delta\varepsilon_n(T) = 2(T-2)\Sigma_{\varepsilon 0}.$$

Similarly, the expression of $\Delta\widehat{\xi}_{nt}$ is

$$\begin{aligned} \Delta\widehat{\xi}_{nt}(T) &= \Delta Y_{nt}(T) - \widehat{T}_{nt}\widehat{\kappa} = \Delta\xi_{nt}(T) + \Delta(WY)_{nt}(\lambda_{10} - \widehat{\lambda}_1) + \Delta(WY)_{n,t-1}(\lambda_{20} - \widehat{\lambda}_2) \\ &\quad + \Delta Y_{n,t-1}(\rho_{10} - \widehat{\rho}_1) + \Delta X_{1nt}(\beta_0 - \widehat{\beta}) + \Delta\varepsilon_{nt}\delta_0 - \Delta\widehat{\varepsilon}_{nt}(T)\widehat{\delta} \\ &= \Delta\xi_{nt}(T) + \Delta(WY)_{nt}(\lambda_{10} - \widehat{\lambda}_1) + \Delta(WY)_{n,t-1}(\lambda_{20} - \widehat{\lambda}_2)\Delta Y_{n,t-1}(\rho_{10} - \widehat{\rho}_1) \\ &\quad + \Delta X_{1nt}(\beta_0 - \widehat{\beta}) + \Delta\varepsilon_{nt}(\delta_0 - \widehat{\delta}) - [\Delta Z_{n,t-1}(\rho_{20} - \widehat{\rho}_2) + \Delta X_{2nt}(\Gamma_0 - \widehat{\Gamma})]\delta_0 \\ &\quad + [\Delta Z_{n,t-1}(\rho_{20} - \widehat{\rho}_2) + \Delta X_{2nt}(\Gamma_0 - \widehat{\Gamma})](\delta_0 - \widehat{\delta}). \end{aligned}$$

Therefore,

$$\frac{1}{n} \Delta\widehat{\xi}_n(T)' \Delta\widehat{\xi}_n(T) \xrightarrow{p} \lim_{n\to\infty} \frac{1}{n} E[\Delta\xi_{nt}(T)' \Delta\xi_{nt}(T)] = \sigma_{\xi 0}^2 tr(\Psi_0 \otimes I_n) = 2(T-2)\sigma_{\xi 0}^2.$$

Because

$$\widehat{\Sigma}_\varepsilon = \frac{1}{2n(T-2)} [\Delta Z_n(T) - K_n \begin{pmatrix} \widehat{\rho} \\ \widehat{\Gamma} \end{pmatrix}]' [\Delta Z_n(T) - K_n \begin{pmatrix} \widehat{\rho} \\ \widehat{\Gamma} \end{pmatrix}]$$

and

$$\widehat{\sigma}_\xi^2 = \frac{1}{2n(T-2)} \Delta\widehat{\xi}_n'(T) \Delta\widehat{\xi}_n(T)$$

are consistent estimators of $\Sigma_{\varepsilon 0}$ and $\sigma_{\xi 0}^2$, by Slutsky's theorem, the matrix

$$\begin{aligned} \widehat{\Sigma}_\kappa &= (\widehat{T}_n' P_{2n} \widehat{T}_n)^{-1} \widehat{T}_n' P_{2n} K_n (K_n' P_{1n} K_n)^{-1} K_n' P_{1n} (\widehat{\delta}' \widehat{\Sigma}_\varepsilon \widehat{\delta} \Psi_0 \otimes I_n) P_{1n} K_n (K_n' P_{1n} K_n)^{-1} \\ &\quad \times K_n' P_{2n} \widehat{T}_n (\widehat{T}_n' P_{2n} \widehat{T}_n)^{-1} + (\widehat{T}_n' P_{2n} \widehat{T}_n)^{-1} \widehat{T}_n' P_{2n} (\widehat{\sigma}_\xi^2 \Psi_0 \otimes I_n) P_{2n} \widehat{T}_n (\widehat{T}_n' P_{2n} \widehat{T}_n)^{-1} \end{aligned}$$

is a consistent estimator of $\Sigma_{\kappa 0}$. □

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendix D
Online Appendix E
Replication files