Estimation and Inference in Functional-Coefficient Spatial Autoregressive Panel Data Models with Fixed Effects

Sun, Yiguo and Malikov, Emir

University of Guelph, Auburn University

2017

Estimation and Inference in Functional-Coefficient Spatial Autoregressive Panel Data Models with Fixed Effects*

Yiguo Sun

Department of Economics and Finance, University of Guelph, Guelph, ON N1G2W1, Canada

Emir Malikov

Department of Agricultural Economics, Auburn University, Auburn, AL 36849, USA

First Draft: December 25, 2015
This Draft: July 15, 2017

Abstract

This paper develops an innovative way of estimating a functional-coefficient spatial autoregressive panel data model with unobserved individual effects which can accommodate (multiple) time-invariant regressors in the model with a large number of cross-sectional units and a fixed number of time periods. The methodology we propose removes unobserved fixed effects from the model by transforming the latter into a semiparametric additive model, the estimation of which however does not require the use of backfitting or marginal integration techniques. We derive the consistency and asymptotic normality results for the proposed kernel and sieve estimators. We also construct a consistent nonparametric test to test for spatial endogeneity in the data. A small Monte Carlo study shows that our proposed estimators and the test statistic exhibit good finite-sample performance.

Keywords: First Difference, Fixed Effects, Hypothesis Testing, Local Linear Regression, Nonparametric GMM, Sieve Estimator, Spatial Autoregressive, Varying Coefficient

JEL Classification: C12, C13, C14, C23

* *Email*: yisun@uoguelph.ca (Sun) and emalikov@auburn.edu (Malikov).

1 Introduction

Sun, Carroll & Li (2009) and Lin, Li & Sun (2014) study the following semiparametric functional-coefficient fixed-effects panel data model:

$$y_{it} = \mathbf{g}_i' \boldsymbol{\theta}(z_{it}) + \mathbf{x}_{it}' \boldsymbol{\beta}(z_{it}) + \mu_i + u_{it}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \tag{1.1}$$

where y_{it} is the (scalar) outcome variable of interest; $\mathbf{g}_i$ and $\mathbf{x}_{it}$ are the time-invariant and time-varying explanatory variables of dimensions d_g and d_x, respectively; z_{it} is a continuously distributed univariate random variable; and $\boldsymbol{\theta}(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ are the $d_g \times 1$ and $d_x \times 1$ vectors of unknown functions to be estimated. The unobserved fixed effects μ_i are allowed to correlate with the strictly exogenous covariates $\mathbf{g}_i$, $\mathbf{x}_{it}$ and z_{it}, but are assumed to be uncorrelated with the idiosyncratic error u_{it}, which is *i.i.d.* with zero mean and finite variance σ_u^2. Both Sun et al. (2009) and Lin et al. (2014) restrict their models to the case of $d_g \leq 1$.

The above semiparametric model has proven to be a popular specification among practitioners. Not only can the model in (1.1) be conveniently applied to reduce the "curse-of-dimensionality" problem, but it also nests purely nonparametric fixed-effects panel data models as well as partially linear fixed-effects panel data models studied by Henderson, Carroll & Li (2008), Qian & Wang (2012) and Li & Liang (2015), who all however focus on a rather restrictive case of $d_g = 0$.

In this paper, we seek to generalize model (1.1) further to the case with spatial dependence in the data. We do so by including the spatial lag of the outcome variable as an additional explanatory variable and allowing the corresponding spatial multiplier to vary with respect to the contextual covariate z_{it}. That is, we consider the following functional-coefficient spatial autoregressive (SAR) fixed-effects panel data model:

$$y_{it} = \rho(z_{it}) \sum_{j \neq i} w_{ij} y_{jt} + \mathbf{g}_i' \boldsymbol{\theta}(z_{it}) + \mathbf{x}_{it}' \boldsymbol{\beta}(z_{it}) + \mu_i + u_{it}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \tag{1.2}$$

where $\sum_{j \neq i} w_{ij} y_{jt}$ is called the "spatial lag" term; w_{ij} is the (i, j)-th element of an $n \times n$ predetermined non-stochastic time-invariant spatial weighting matrix $\mathbf{W}_0$ such that $w_{ii} = 0$ for all $i = 1, \ldots, n$; and y_{it} is spatially stationary.[1] Further, $\rho(\cdot)$ is usually referred to as the "spatial multiplier" or "spatial lag parameter", which is an unknown function to be estimated. For instance, when the SAR model is game-theoretically rationalized as a "reaction function" (e.g., Brueckner, 2003), the spatial multiplier $\rho(\cdot)$ can be conveniently interpreted as the "reaction" parameter, which our model permits to meaningfully vary with some contextual factor z_{it}. Since the spatial multiplier captures the direct impact of other units' actions/outcomes on the ith unit's action/outcome, extending (1.1) to model (1.2) enables us to test whether there is spatial/economic externality across individual units. Also note that, when $\rho(z) \equiv \rho_0$, $\boldsymbol{\theta}(z) \equiv \mathbf{0}_{d_g}$ and $\boldsymbol{\beta}(z) \equiv \boldsymbol{\beta}_0$, our semiparametric model (1.2) collapses to Lee & Yu's (2010a) fully parametric SAR fixed-effects panel data model.

Some potential applications of our model, for instance, include the estimation of growth models that explicitly account for technological interdependence between countries in the presence of spillover effects. Such a technological interdependence is usually formulated in the form of spatial externalities (e.g., see Ertur & Koch, 2007). However, the intensity of knowledge spillovers is naturally expected to greatly depend on institutional and cultural compatibility of neighboring countries (Kelejian, Murrell & Shepotylo, 2013). Our functional-coefficient model presents a practical method to allow for such indirect effects of institutions on the degree of spatial dependence in the cross-country conditional convergence regressions via a contextual variable z_{it}. The estimation

[1] We delay the discussion of the spatial stationarity condition to Section 2.

of hedonic house price functions is another application, where it is imperative to allow for potential spatial dependence in the data. House prices are widely believed to be spatially autoregressive because residential property values tend to reflect shared local amenities as well as observed and unobserved neighborhood characteristics. While these characteristics can be partly controlled for using locality fixed effects, such an approach may be unsatisfactory since it does not let characteristics of neighboring houses affect the price of a given house (Anselin & Lozano-Gracia, 2009). However, by including the spatial lag in a house pricing function, one is able to accommodate such cross-neighbor effects.

In recent decades, the econometric literature has seen a rapid development in the theory of nonparametric estimation and testing of fixed-effects panel data models. For instance, see Sun, Zhang & Li (2015) for an excellent survey on the nonparametric panel data analysis. However, the introduction of nonparametric structure to models with spatial dependence (and spatial autoregressive models, in particular) in the panel data setup still lacks enough attention and progress, although significantly more visible advances have been made in the development of parametric spatial models (e.g., Lee & Yu, 2010a,b, 2012, 2014; Yu, de Jong & Lee, 2012). Our work therefore aims to fill this research gap in the literature.

Few existing nonparametric studies, all of which focus on a purely cross-sectional setup, include the works of Su & Jin (2010), Su (2012) and Zhang (2013), who consider a Robinson-type partially linear semiparametric SAR model, whereas Sun, Yan, Zhang & Lu (2014) and Malikov & Sun (2017) study fully and/or partially linear functional-coefficient SAR models. The spatial autoregressive models in which spatial weights are specified in the form of unknown nonparametric functions of some geographic or economic distance are examined by Pinkse, Slade & Brett (2002) and Sun (2016).

For a large n and fixed T, we develop an innovative way of estimating model (1.2). We first propose a two-stage kernel estimation method to estimate $\rho(\cdot)$ and $\boldsymbol{\beta}(\cdot)$, after removing the unobserved fixed effects from the model via first differencing. Our approach transforms the model into a semiparametric additive panel data model from which a consistent estimator is usually constructed using either the backfitting (Henderson et al., 2008; Mammen, Støve & Tjøstheim, 2009; Li & Liang, 2015) or marginal integration techniques (Qian & Wang, 2012).[2] Unlike a more conventional model (1.1), our model of interest in (1.2) naturally suffers from the endogeneity problem due to the presence of the spatial lag term in the equation. We therefore resort to a nonparametric instrumental variable approach in order to construct consistent estimators of the unknown coefficient curves $\rho(\cdot)$ and $\boldsymbol{\beta}(\cdot)$. However, when based on both localized linear and quadratic moments, the nonparametric GMM estimator has no analytic expression. Consequently, both the backfitting and marginal integration techniques can be computationally challenging in the calculation of such an estimator for model (1.2). Therefore, we propose a new estimator that is significantly simpler to implement than the backfitting and marginal integration estimators.

Having consistently estimated $\rho(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ at the conventional nonparametric convergence rate in the second stage, we next propose a third-stage sieve estimator to consistently estimate unknown functional curves $\boldsymbol{\theta}(\cdot)$ for time-invariant regressors $\mathbf{g}_i$. Importantly, the estimator we propose in this paper can be used to estimate functional coefficients $\boldsymbol{\theta}(\cdot)$ even when the number of time-invariant regressors is greater than one, i.e., $d_g > 1$. The methodology that we develop can also be used to estimate the traditional functional-coefficient fixed-effects panel data models with time-invariant covariates like the one in (1.1). This makes a significant improvement over the existing estimation methods that are applicable to the case of $d_g \leq 1$ only (as in Sun et al., 2009).

[2]Where all these articles consider purely nonparametric fixed-effects panel data models with exogenous covariates and no spatial dependence.

Given that our semiparametric spatial autoregressive model (1.2) nests a more traditional functional-coefficient fixed-effects model in (1.1) as a special case, one may naturally wish to formally discriminate between the two models. Therefore, we also propose a consistent residual-based L_2-type test statistic to test for relevance of the spatial lag term in the model. The proposed is, essentially, the test for spatial endogeneity. Our specification test belongs to the family of similar nonparametric residual-based tests considered for independent data (e.g., Zheng, 1996; Li & Wang, 1998), weakly dependent time series data (e.g., Fan & Li, 1999; Li, 1999), integrated time series data (e.g., Wang & Phillips, 2012; Sun, Cai & Li, 2015) and, more recently, for spatial data (e.g., Su & Qu, 2017; Malikov & Sun, 2017).

The rest of the paper is organized as follows. Section 2 explains the model of interest along with the spatial stationarity condition. We derive the consistency and asymptotic normality results for the first-difference kernel estimator of $\rho(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ in Section 3, whereas the limiting results for a sieve estimator of $\boldsymbol{\theta}(\cdot)$ are discussed in Section 4. Section 5 contains a further discussion of the estimation issues. Section 6 presents a consistent nonparametric test statistic to test for the presence of spatial endogeneity in the model. Section 7 reports a small Monte Carlo simulation study to asses the small sample performance of our proposed estimators and the test statistic. We conclude in Section 8. All mathematical proofs are relegated to the Appendix.

Before anything else, we summarize our notation. Boldface letters are reserved for vectors and matrices. (i) Throughout this paper, we denote an $[n(T-1)] \times d_\omega$ matrix $\boldsymbol{\omega} = [\boldsymbol{\omega}_2', \ldots, \boldsymbol{\omega}_T']'$ with an $n \times d_\omega$ vector $\boldsymbol{\omega}_t = [\boldsymbol{\omega}_{1t}, \ldots, \boldsymbol{\omega}_{nt}]'$ for any $t = 2, \ldots, T$, where $\boldsymbol{\omega}_{it}$ is a $d_\omega \times 1$ vector. (ii) Let $\mathbf{i}_T$ be a $T \times 1$ vector of ones, $\mathbf{0}_q$ be a $q \times 1$ vector of zeros, $\mathbf{I}_m$ be an $m \times m$ identity matrix and $\mathbf{0}_{q \times p}$ be a $q \times p$ zero matrix. (iii) $\|\cdot\|$ refers to the Euclidian norm, and $\|\mathbf{A}\|_1 = \max_{1 \le j \le n} \sum_{i=1}^n |a_{ij}|$ and $\|\mathbf{A}\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^n |a_{ij}|$ are the column and row sum matrix norms, respectively. (iv) Let $\lambda_j(\mathbf{A})$, $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ respectively be the jth, smallest and largest eigenvalue of some $m \times m$ matrix $\mathbf{A} = (a_{ij})_{i,j=1}^m$, and $\|\mathbf{A}\|_{sp} = \max_{\|\omega\|=1, \omega \neq 0} \|\mathbf{A}\omega\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$ defines the spectral norm. For any vector $\mathbf{a}$, we see $\|\mathbf{a}\| = \|\mathbf{a}\|_{sp}$. (v) We denote $\mathbf{A}^s = \mathbf{A} + \mathbf{A}'$ for any square matrix $\mathbf{A}$. (vi) $\mathbf{A}_n = O_e(1)$ means that each and every element of a random matrix $\mathbf{A}_n$ is of order $O_p(1)$ not $o_p(1)$. (vii) $A_n \overset{d}{=} B_n$ means that A_n and B_n have the same distribution asymptotically. (viii) We use C to denote a generic constant that can take different values at different places.

2 The Model

We rewrite model (1.2) in matrix form:

$$\mathbf{y}_t = \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\mathbf{y}_t + \mathbf{g}'\boldsymbol{\theta}(\mathbf{z}_t) + \boldsymbol{\phi}_t'\boldsymbol{\beta}(\mathbf{z}_t) + \boldsymbol{\mu} + \mathbf{u}_t, \quad t = 1, \ldots, T, \tag{2.1}$$

where $\boldsymbol{\rho}(\mathbf{z}_t) = \mathrm{diag}\{\rho(z_{1t}), \ldots, \rho(z_{nt})\}$ is an $n \times n$ diagonal matrix of spatial autoregressive parameter functions; $\boldsymbol{\beta}(\mathbf{z}_t) = [\boldsymbol{\beta}(z_{1t})', \ldots, \boldsymbol{\beta}(z_{nt})']'$ and $\boldsymbol{\theta}(\mathbf{z}_t) = [\boldsymbol{\theta}(z_{1t})', \ldots, \boldsymbol{\theta}(z_{nt})']'$ are $(nd_x) \times 1$ and $(nd_g) \times 1$ vectors of functional coefficients, respectively; $\boldsymbol{\phi}_t = \mathrm{diag}\{\mathbf{x}_{1t}, \ldots, \mathbf{x}_{nt}\}$ is a $(nd_x) \times n$ matrix; $\mathbf{g} = \mathrm{diag}\{\mathbf{g}_1, \ldots, \mathbf{g}_n\}$ is a $(nd_g) \times n$ matrix; and $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]'$ is an $n \times 1$ vector of unobserved individual-specific fixed effects. The reduced form of model (2.1) is given by

$$\mathbf{y}_t = \mathbf{S}_n(\mathbf{z}_t)\left[\mathbf{g}'\boldsymbol{\theta}(\mathbf{z}_t) + \boldsymbol{\phi}_t'\boldsymbol{\beta}(\mathbf{z}_t) + \boldsymbol{\mu} + \mathbf{u}_t\right], \quad t = 1, \ldots, T \tag{2.2}$$

provided that $\mathbf{S}_n(\mathbf{z}_t) = [\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0]^{-1}$ exists.[3] This means that, if

$$\max_{1 \le j \le n, 1 \le t \le T} |\lambda_j\{\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\}| < 1 \tag{2.3}$$

[3] By Property 19.15 in Seber (2008, p.421), $\sum_{i=0}^\infty \mathbf{A}_n^i$ converges to $(\mathbf{I}_n - \mathbf{A}_n)^{-1}$ if $\lim_{k \to \infty} \mathbf{A}_n^k = 0$ or $\max_{1 \le j \le n} |\lambda_j(\mathbf{A}_n)| < 1$, where $\mathbf{A}_n$ is an $n \times n$ matrix.

holds almost surely, model (2.2) can be rewritten as

$$\mathbf{y}_t = \sum_{k=0}^{\infty} \left[\boldsymbol{\rho}\left(\mathbf{z}_t\right) \mathbf{W}_0\right]^k \left[\mathbf{g}'\boldsymbol{\theta}\left(\mathbf{z}_t\right) + \boldsymbol{\phi}_t'\boldsymbol{\beta}(\mathbf{z}_t) + \boldsymbol{\mu} + \mathbf{u}_t\right], \quad t = 1, \ldots, T. \tag{2.4}$$

For any given t, condition (2.3) implies that the spatial weight $\left[\boldsymbol{\rho}\left(\mathbf{z}_t\right) \mathbf{W}_0\right]^k$ becomes smaller in magnitude and less important as k increases. This is analogous to the time-series case of an AR(1) process, e.g., $s_t = \rho s_{t-1} + v_t$, becomes stationary if $|\rho| < 1$, under which condition $s_t = \sum_{k=0}^{\infty} \rho^k v_{t-k}$. Hence, we say that $\{y_{it}\}$ is *spatially stationary* if (2.3) holds true. Throughout this paper, we assume (2.3) holds. Further discussion of this condition is delayed until Section 5.2.

Next, we note that, in the presence of unobserved fixed effects $\boldsymbol{\mu}$, the functional coefficients $\boldsymbol{\theta}\left(z_{it}\right)$ of time-invariant regressors $\mathbf{g}_i$ are not identifiable from $\boldsymbol{\theta}_0 + \boldsymbol{\theta}_1\left(z_{it}\right)$, where $\boldsymbol{\theta}_0$ is a vector of constants. Therefore, we normalize these coefficients such that $\boldsymbol{\theta}(0) = \mathbf{0}_{d_g}$ holds true. This is a reasonable normalization, since $\mathbf{g}_i'\boldsymbol{\theta}_0$ can always be attributed to time-invariant fixed effects.

3 The First-Difference Kernel Estimator of $\rho(z)$ and $\boldsymbol{\beta}(z)$

We propose a two-stage kernel estimation method to estimate model (1.2), removing unobserved fixed effects from the model via the first-difference transformation. We opt to cancel fixed effects out by transforming the model as opposed to "concentrate" them out by employing Sun et al.'s (2009) smoothed dummy variable approach due to infeasibility of the latter in the GMM setup.[4]

Define two $(d_x+1)\times 1$ vectors: $\mathbf{m}_{it} = \left[\sum_{j\neq i} w_{ij} y_{jt}, \mathbf{x}_{it}'\right]'$ and $\boldsymbol{\gamma}(z_{it}) = [\rho(z_{it}), \boldsymbol{\beta}(z_{it})']'$. Then, applying the first-difference transformation to model (1.2) gives

$$\Delta y_{it} = \mathbf{g}_i'\left[\boldsymbol{\theta}\left(z_{it}\right) - \boldsymbol{\theta}\left(z_{i,t-1}\right)\right] + \mathbf{m}_{it}'\boldsymbol{\gamma}(z_{it}) - \mathbf{m}_{i,t-1}'\boldsymbol{\gamma}(z_{i,t-1}) + \Delta u_{it}, \quad i = 1, \ldots, n,\ t = 2, \ldots, T \tag{3.1}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$ and $\Delta u_{it} = u_{it} - u_{i,t-1}$.

Further, since the spatial lag term in (1.2) is endogenous, we assume there exist $d_q \geq 1$ valid instruments for $\sum_{j\neq i} w_{ij} y_{jt}$ denoted by $\mathbf{q}_{it}$ such that

$$\mathbb{E}\left[\mathbf{q}_{it} u_{is} | \mathbf{x}_i, \mathbf{z}_i, \mathbf{g}_i\right] = \mathbf{0}_{d_q} \quad \forall\ i, s, t \text{ almost surely}, \tag{3.2}$$

where $\mathbf{x}_i = [\mathbf{x}_{i1}', \ldots, \mathbf{x}_{iT}']'$ and $\mathbf{z}_i = [z_{i1}, \ldots, z_{iT}]'$, which implies that $\mathbf{q}_{it}$ is strictly exogenous.

Applying the Taylor expansion to $\boldsymbol{\theta}\left(z_{it}\right)$ and $\boldsymbol{\gamma}\left(z_{it}\right)$ at an interior point z_1 and to $\boldsymbol{\theta}\left(z_{i,t-1}\right)$ and $\boldsymbol{\gamma}\left(z_{i,t-1}\right)$ at an interior point $z_2 \neq z_1$, we approximate (3.1) by

$$\begin{aligned} \Delta y_{it} &\approx \mathbf{g}_i'\left[\boldsymbol{\theta}\left(z_1\right) - \boldsymbol{\theta}\left(z_2\right)\right] + \mathbf{m}_{it}'\boldsymbol{\gamma}(z_1) - \mathbf{m}_{i,t-1}'\boldsymbol{\gamma}(z_2) + \Delta u_{it} \\ &= \mathbf{g}_i'\dot{\boldsymbol{\theta}}\left(\mathbf{z}\right) + \boldsymbol{\xi}'\boldsymbol{m}_{it}'\boldsymbol{\gamma}\left(\mathbf{z}\right) + \Delta u_{it} \end{aligned} \tag{3.3}$$

for a given (i,t) such that $|z_{it} - z_1| = o(1)$ and $|z_{i,t-1} - z_2| = o(1)$, where $\mathbf{z} = [z_1, z_2]'$, $\boldsymbol{\xi} = [1, -1]'$, $\boldsymbol{m}_{it} = \operatorname{diag}\{\mathbf{m}_{it}, \mathbf{m}_{i,t-1}\}$, $\dot{\boldsymbol{\theta}}\left(\mathbf{z}\right) = \boldsymbol{\theta}\left(z_1\right) - \boldsymbol{\theta}\left(z_2\right)$ and $\boldsymbol{\gamma}\left(\mathbf{z}\right) = \left[\boldsymbol{\gamma}\left(z_1\right)', \boldsymbol{\gamma}\left(z_2\right)'\right]'$. Note that, due to the time invariance of $\mathbf{g}_i$, we can only identify $\dot{\boldsymbol{\theta}}\left(\mathbf{z}\right)$ and not $\boldsymbol{\theta}\left(z_1\right)$ and $\boldsymbol{\theta}\left(z_2\right)$ individually.

Equations (3.2) and (3.3) imply the following localized orthogonal moment conditions:

$$\mathbb{E}\left[\mathbf{Q}_{it}\left(\Delta y_{it} - \mathbf{g}_i'\dot{\boldsymbol{\theta}}\left(\mathbf{z}\right) - \boldsymbol{\xi}'\boldsymbol{m}_{it}'\boldsymbol{\gamma}\left(\mathbf{z}\right)\right) k_{it}(h, \mathbf{z})\right] \approx \mathbf{0}_d \tag{3.4}$$

[4]See the discussion in Section 5.3.

for $i = 1, \ldots, n$, $t = 2, \ldots, T$, where $d = d_g + 2(d_q + d_x)$, $\mathbf{Q}_{it} = \left[\mathbf{g}_i', \boldsymbol{\xi}'\check{\boldsymbol{m}}_{it}'\right]'$, $\check{\boldsymbol{m}}_{it} = \text{diag}\{\check{\mathbf{m}}_{it}, \check{\mathbf{m}}_{i,t-1}\}$, $\check{\mathbf{m}}_{it} = [\mathbf{q}_{it}', \mathbf{x}_{it}']'$, $k_{it}(h, \mathbf{z}) = k((z_{it} - z_1)/h)k((z_{i,t-1} - z_2)/h)$ with $k(\cdot)$ being a kernel function and h being the bandwidth. Note that we use a bivariate product kernel function because (3.3) involves a *two*-dimensional approximation, which we employ in order to avoid estimating $\boldsymbol{\theta}(\cdot)$ and $\boldsymbol{\gamma}(\cdot)$ via the backfitting (iterative) technique that would explicitly accommodate the additive structure of the first-differenced model. Thus, our methodology involves a two-dimensional semiparametric estimation which, expectedly, will be less efficient than iterative calculation methods. To improve the estimation accuracy, we therefore provide a second-stage estimator in Section 3.1.

If $\mathbf{x}_{it}$, $\mathbf{g}_i$ and z_{it} are all relevant in predicting y_{it}, a selection of linearly independent variables from $\mathbf{W}_0\mathbf{x}_t$, $\mathbf{W}_0\mathbf{z}_t, \mathbf{W}_0[\mathbf{g}_1, \ldots, \mathbf{g}_n]'$, $\mathbf{W}_0^2\mathbf{x}_t$, $\mathbf{W}_0^2\mathbf{z}_t$, $\mathbf{W}_0^2[\mathbf{g}_1, \ldots, \mathbf{g}_n]', \ldots$ will serve as a set of good instruments for the spatially endogenous variable $\mathbf{W}_0\mathbf{y}_t$ appearing in (2.1). Since we only seek to obtain a consistent nonparametric GMM estimator without pursuing the optimal estimator, we can use $\mathbf{q}_{it} = \sum_{j \neq i} w_{ij}[\mathbf{x}_{jt}', z_{jt}, \mathbf{g}_j']'$ as our instrument, having removed any redundant terms. However, if $\mathbf{x}_{it}$, $\mathbf{g}_i$ and z_{it} are all irrelevant or weak in predicting y_{it}, $\mathbf{q}_{it}$ is not going to be a good instrument. Without pre-testing the relevance of exogenous covariates in a purely cross-sectional version of model (1.2), Malikov & Sun (2017) show that combining both linear and quadratic moments can be used to consistently estimate unknown coefficient curves regardless of whether the exogenous covariates are relevant in predicting the dependent variable. We expect similar results to hold in our panel data setup.[5]

Different from parametric spatial panel data models with fixed effects, the first-differenced model in (3.1) and its local approximation in (3.3) are no longer SAR models. However, we are still able to construct quadratic moment conditions using $\mathbf{P}_{n,l} = \mathbf{I}_{T-1} \otimes \left[\mathbf{W}_0^l - n^{-1}\text{tr}\left\{\mathbf{W}_0^l\right\}\mathbf{I}_n\right]$ for $l = 1, 2, \ldots L$, where L is a finite integer. For an $[n(T-1)] \times 1$ vector of transformed errors $\Delta\mathbf{u} = [\Delta\mathbf{u}_2', \ldots, \Delta\mathbf{u}_T']'$, it is readily seen that $\mathbb{E}[\Delta\mathbf{u}'\mathbf{P}_{n,l}\Delta\mathbf{u}] = \text{tr}\{\mathbf{P}_{n,l}\mathbb{E}[\Delta\mathbf{u}\Delta\mathbf{u}']\} = 0$ because $\mathbb{E}[\Delta\mathbf{u}\Delta\mathbf{u}'] = \sigma_u^2\boldsymbol{\Sigma} \otimes \mathbf{I}_n$, where $\boldsymbol{\Sigma} = 2\mathbf{I}_{T-1} - \mathbf{J}_{T-1}(0) - \mathbf{J}_{T-1}(0)'$ is a $(T-1) \times (T-1)$ matrix, and $\mathbf{J}_{T-1}(0)$ defines a Jordan block matrix with zeros along the main diagonal and ones along the superdiagonal. Therefore, we obtain the following local quadratic moments:

$$\mathbb{E}\left[\left(\boldsymbol{\Delta}\mathbf{y} - \mathbf{M}\boldsymbol{\Theta}(\mathbf{z})\right)'\mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}\mathbf{K}_h(\mathbf{z})\left(\boldsymbol{\Delta}\mathbf{y} - \mathbf{M}\boldsymbol{\Theta}(\mathbf{z})\right)\right] \approx 0, \tag{3.5}$$

where $\boldsymbol{\Delta}\mathbf{y} = [\Delta\mathbf{y}_2', \ldots, \Delta\mathbf{y}_T']'$ is an $[n(T-1)] \times 1$ vector; $\mathbf{M} = [\mathbf{M}_2', \ldots, \mathbf{M}_T']'$ is a $[n(T-1)] \times [2(d_x+1) + d_g]$ data matrix with $\mathbf{M}_t = [\mathbf{M}_{1t}, \ldots, \mathbf{M}_{nt}]'$ and $\mathbf{M}_{it} = \left[\mathbf{g}_i', \boldsymbol{\xi}'\boldsymbol{m}_{it}'\right]'$; $\boldsymbol{\Theta}(\mathbf{z}) = \left[\dot{\boldsymbol{\theta}}(\mathbf{z})', \boldsymbol{\gamma}(\mathbf{z})'\right]'$ is of dimension $2(d_x+1) + d_g$; and $\mathbf{K}_h(\mathbf{z}) = \text{diag}\{\mathbf{K}_2(\mathbf{z}), \ldots, \mathbf{K}_T(\mathbf{z})\}$ is an $[n(T-1)] \times [n(T-1)]$ diagonal matrix of kernel weights with $\mathbf{K}_t(\mathbf{z}) = \text{diag}\{k_{1t}(h, \mathbf{z}), \ldots, k_{nt}(h, \mathbf{z})\}$.

Then, denoting

$$\boldsymbol{g}_n(\boldsymbol{\vartheta}) = \begin{bmatrix} \left(\boldsymbol{\Delta}\mathbf{y} - \mathbf{M}\boldsymbol{\vartheta}\right)'\mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,1}\mathbf{K}_h(\mathbf{z})\left(\boldsymbol{\Delta}\mathbf{y} - \mathbf{M}\boldsymbol{\vartheta}\right) \\ \vdots \\ \left(\boldsymbol{\Delta}\mathbf{y} - \mathbf{M}\boldsymbol{\vartheta}\right)'\mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,L}\mathbf{K}_h(\mathbf{z})\left(\boldsymbol{\Delta}\mathbf{y} - \mathbf{M}\boldsymbol{\vartheta}\right) \\ \mathbf{Q}'\mathbf{K}_h(\mathbf{z})\left(\boldsymbol{\Delta}\mathbf{y} - \mathbf{M}\boldsymbol{\vartheta}\right) \end{bmatrix} \tag{3.6}$$

for a $[2(d_x+1) + d_g] \times 1$ vector $\boldsymbol{\vartheta}$, where $\mathbf{Q} = [\mathbf{Q}_2', \ldots, \mathbf{Q}_T']'$ is an $[n(T-1)] \times d$ instrument matrix

[5] Motivated by the maximum likelihood method for the parametric SAR model, Lee (2007) shows that combining both the linear and quadratic moments could improve the GMM estimation efficiency. This idea is also applied in Kelejian & Prucha (1999) and Lee & Yu (2014).

with $\mathbf{Q}_t = [\mathbf{Q}_{1t}, \ldots, \mathbf{Q}_{nt}]'$, we construct our initial nonparametric GMM estimator, i.e.,

$$\widehat{\mathbf{\Theta}}(z) = \arg\min_{\mathbf{\Theta}(z)} \ \boldsymbol{g}_n\left(\mathbf{\Theta}(z)\right)' \boldsymbol{g}_n\left(\mathbf{\Theta}(z)\right). \tag{3.7}$$

Below, we list assumptions used to derive the limiting distribution of our proposed estimator.

Assumption 1 *$\{(\mathbf{g}_i, \mathbf{x}_{it}, z_{it}, u_{it})\}$ is i.i.d. across index i, y_{it} is generated from model* (1.2) *with $\mathbf{g}_i$, $\mathbf{x}_{it}$ and z_{it} being strictly exogenous and all these variables have finite second moments. Also,*

(i) $\mathbb{E}\left[u_{it}|\mathbf{x}_i = \mathbf{x}, \mathbf{z}_i = z, \mathbf{g}_i = \mathbf{g}\right] = 0$, $\mathbb{E}\left[u_{it}^2|\mathbf{x}_i = \mathbf{x}, \mathbf{z}_i = z, \mathbf{g}_i = \mathbf{g}\right] = \sigma_u^2 > 0$ *for any* $\mathbf{x} \in \mathcal{S}_x \subset R^{d_x}$, $z \in \mathcal{S}_z \subset R$ *and* $\mathbf{g} \in \mathcal{S}_g \subset R^{d_g}$ *and* $\sup_{x\in\mathcal{S}_x, z\in\mathcal{S}_z, g\in\mathcal{S}_g} \mathbb{E}\left[u_{it}^4|\mathbf{x}_i = \mathbf{x}, \mathbf{z}_i = z, g_i = \mathbf{g}\right] \le C < \infty$, *where $\mathcal{S}_z$ is a compact subset of R;*

(ii) For all i, $(z_{it}, z_{i,t-1})$ and $(z_{it}, z_{i,t-1}, z_{i,t-2})$ have a common joint pdf $f_{t,t-1}(z_1, z_2)$ and $f_{t,t-1,t-2}(z_1, z_2, z_3)$ with respect to the Lebesgue measure over their domains, respectively;

(iii) For any t, there exist a positive integer N and a constant $c_w \in (0,1)$ such that for all $n > N$, $\max_{1\le j\le n} |\lambda_j\{\boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\}| \le c_w$ *almost surely,* $\|\mathbf{W}_0\|_j \le C$ *and* $\left\|\left[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\right]^{-1}\right\|_j \le C$ *for $j = 1$ and ∞;*

(iv) $\mathbf{P}_{n,l} = \mathbf{I}_{T-1} \otimes P_{n,l}$ *with $P_{n,l}$ being an $n \times n$ matrix with finite row- and column-sum matrix norm and $tr\{P_{n,l}\} = 0$ for all $l = 1, \ldots, L$, where $L \ge 1$ is a finite positive integer. Also, $diag\{P_{n,l}\} \ne 0$ for at least one l.*

Assumption 2 *(i) In the neighborhood of an interior point $\mathbf{z} = [z_1, z_2]'$ with $z_1 \ne z_2$, $\boldsymbol{\beta}(\mathbf{z})$, $\boldsymbol{\rho}(\mathbf{z})$, $f_{t,t-1}(\mathbf{z})$, $\mathbb{E}\left[(\mathbf{g}_i'\mathbf{g}_i)^j|\mathbf{z}\right]$ for $j = 1, 2$, $\mathbb{E}[\mathbf{g}_i\check{\mathbf{m}}_{it}'|\mathbf{z}]$, $\mathbb{E}[\mathbf{g}_i'\mathbf{g}_i\check{\mathbf{m}}_{it}'\check{\mathbf{m}}_{it}|\mathbf{z}]$, $\mathbb{E}[\mathbf{m}_{is}\check{\mathbf{m}}_{it}'|\mathbf{z}]$, $\mathbb{E}[\check{\mathbf{m}}_{is}\check{\mathbf{m}}_{it}'|\mathbf{z}]$, $\mathbb{E}[\mathbf{m}_{is}'\mathbf{m}_{is}\check{\mathbf{m}}_{it}'\check{\mathbf{m}}_{it}|\mathbf{z}]$ and $\mathbb{E}[\check{\mathbf{m}}_{is}'\check{\mathbf{m}}_{is}\check{\mathbf{m}}_{it}'\check{\mathbf{m}}_{it}|\mathbf{z}]$ are all twice continuously differentiable for all t and s satisfying $0 \le |s-t| \le 1$, and $\mathbb{E}\left(\|\mathbf{x}_{it}\|^{(2+\delta_1)}|\mathbf{z}\right) < C$ and $\mathbb{E}\left(\|\mathbf{g}_i\|^{(2+\delta_1)}|\mathbf{z}\right) < C$ for some $\delta_1 \ge 2$, where $\mathbb{E}[\cdot|\mathbf{z}] = \mathbb{E}[\cdot|z_{it} = z_1, z_{i,t-1} = z_2]$;*

(ii) In the neighborhood of an interior point $\dot{\mathbf{z}} = [z_1, z_2, z_2]'$, $f_{t,t-1,t-2}(\dot{\mathbf{z}})$, $\mathbb{E}\left[(\mathbf{g}_i'\mathbf{g}_i)^j|\dot{\mathbf{z}}\right]$ for $j = 1, 2$, $\mathbb{E}[\mathbf{g}_i\check{\mathbf{m}}_{is}'|\dot{\mathbf{z}}]$, $\mathbb{E}[\check{\mathbf{m}}_{is}\check{\mathbf{m}}_{it}'|\dot{\mathbf{z}}]$ and $\mathbb{E}[\check{\mathbf{m}}_{is}'\check{\mathbf{m}}_{is}\check{\mathbf{m}}_{it}'\check{\mathbf{m}}_{it}|\dot{\mathbf{z}}]$ are all twice continuously differentiable for all t and s satisfying $0 \le |s-t| \le 1$, where $\mathbb{E}[\cdot|\dot{\mathbf{z}}] = \mathbb{E}[\cdot|z_{it} = z_1, z_{i,t-1} = z_2, z_{i,t-2} = z_2]$;

(iii) $\varkappa_B(h, \mathbf{z})$ is a non-singular matrix, where $\varkappa_B(h, \mathbf{z})$ is defined in Lemma 2 in the Appendix.

Assumption 3 *The kernel function $k(u)$ is a symmetric probability density function with a compact support $[-1, 1]$. Also, we denote $\upsilon_{i,j}(k) = \int k^i(u)\, u^j du$.*

Assumption 4 *As $n \to \infty$, $h \to 0$, and $\lim_{n\to\infty} nh^6 = c > 0$.*

Assumptions **1**–**4** contain regularity conditions, where the assumption of compactness of $\mathcal{S}_z$ in Assumption **1(i)** and the bounded support of the kernel function in Assumption **3** are not essential and are imposed to simplify our assumptions and mathematical proofs**.** Assumption **1(iii)** and the boundedness of $\boldsymbol{\rho}(\mathbf{z})$ in Assumption **2(i)** parallel Assumption 1(iii)–(iv) in Su (2012). These assumptions ensure spatial stationarity of the dependent variable and facilitate the limit result of our estimator.

The compactness of $\mathcal{S}_z$ and Assumption **2** ensure that the functions listed in Assumption **2** are all uniformly bounded for all t and s in the domain of $\mathbf{z}$. Relaxing the *i.i.d.* over i assumption about $\{(\mathbf{g}_i, \mathbf{x}_{it}, z_{it}, u_{it})\}$ to independence with heteroskedasticity in Assumption **1** does not shed extra light on our theory, so we maintain the current assumption to keep our formulae simple. In addition, since our paper considers the case when T is a finite number, we do not impose serial correlation assumptions on the panel data across time. Assumption **4** limits the speed at which the bandwidth h approaches to zero as the sample size n increases in order to balance the squared asymptotic bias and asymptotic variance of our estimator.

Theorem 1 *Under Assumptions 1–4, at an interior point* $\mathbf{z} = [z_1, z_2]'$*, we have*

$$\sqrt{nh^2}\left(\widehat{\boldsymbol{\Theta}}(\mathbf{z}) - \boldsymbol{\Theta}(\mathbf{z}) - \varkappa_B(h,\mathbf{z})^{-1}\varkappa_A(h,\mathbf{z})'\right) \xrightarrow{d} \mathbb{N}\left(\mathbf{0}_{2(d_x+1)+d_g}, \sigma_u^2 \upsilon_{2,0}(k)\varkappa_B(h,\mathbf{z})^{-1}\boldsymbol{\Omega}(\mathbf{z})\varkappa_B(h,\mathbf{z})^{-1}\right),$$

where $\varkappa_A(h,\mathbf{z}) = O_p\left(h^2\right)$, $\varkappa_B(h,\mathbf{z})$, $\boldsymbol{\Omega}(\mathbf{z})$ *is a finite p.d.f. matrix, and all are defined in Lemmas 1–3 in the Appendix.*

As noted earlier, $\mathbf{q}_{it}$ may be an invalid instrument when $\mathbf{x}_{it}$, z_{it} and $\mathbf{g}_i$ are irrelevant in predicting y_{it}. Under such circumstances, the use of local quadratic moments in (3.5) will ensure the non-singularity of $\varkappa_B(h,\mathbf{z})$ if $\mathbb{E}[\psi_{s,l}(h,\mathbf{z})] = 2\left(nh^2\right)^{-1}\sum_{i=1}^n\sum_{t=2}^T p_{l,ii}\mathbb{E}\left[a_{ii}(\mathbf{z}_{t-s})k_{it}^2(\mathbf{z})\Delta u_{it}u_{i,t-s}\right]$ converges to a non-zero constant for $s = 0, 1$, where $a_{ij}(\mathbf{z}_t)$ and $p_{l,ij}$ are the (i,j)th elements of $\mathbf{W}_0\mathbf{S}_n(\mathbf{z}_t)$ and $P_{n,l}$, respectively. Clearly, if one defines $\{P_{n,l}\}$ such that $\operatorname{diag}\{P_{n,l}\} = 0$ for all l, then $\psi_{s,l}(h,\mathbf{z}) = 0$ for $s = 0, 1$ implying that, in such a case, $\varkappa_B(h,\mathbf{z})$ will not be nonsingular in large samples as shown in the remark below Lemma 2 in the Appendix.

Theorem 1 indicates that $\widehat{\boldsymbol{\Theta}}(\mathbf{z}) - \boldsymbol{\Theta}(\mathbf{z}) = O_p\left(h^2 + \left(nh^2\right)^{-1/2}\right)$ which is in line with the conventional kernel estimation theory keeping in mind that first differencing transforms the one-dimensional estimation problem in (1.2) into a two-dimensional problem in (3.1). Evidently, the asymptotic variance term is too large. Using $\widehat{\boldsymbol{\Theta}}(\mathbf{z})$ as the initial consistent estimator of $\boldsymbol{\Theta}(\mathbf{z})$, we therefore construct the second-stage estimator of $\boldsymbol{\gamma}(z)$ in Section 3.1. We show that this estimator of $\boldsymbol{\gamma}(z)$ is more efficient than the first-stage estimator and reaches the conventional convergence rate of $O_p\left(h_0^2 + (nh_0)^{-1/2}\right)$, where h_0 is the bandwidth used in the second-stage estimation.

3.1 Second-Stage Estimator of $\boldsymbol{\gamma}(z)$

To derive the second-stage estimator of $\boldsymbol{\gamma}(z)$, we rewrite the model in (3.1) as follows:

$$\Delta y_{it}^\dagger = \mathbf{m}_{it}'\boldsymbol{\gamma}(z_{it}) + \Delta u_{it}, \quad i = 1, \ldots, n, \; t = 2, \ldots, T, \tag{3.8}$$

where $\Delta y_{it}^\dagger \equiv \Delta y_{it} - \mathbf{g}_i'\dot{\boldsymbol{\theta}}(z_{it}) + \mathbf{m}_{i,t-1}'\boldsymbol{\gamma}(z_{i,t-1})$. The matrix form of model (3.8) is given by

$$\Delta\mathbf{y}_t^\dagger = \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\mathbf{y}_t + \boldsymbol{\phi}_t'\boldsymbol{\beta}(\mathbf{z}_t) + \Delta\mathbf{u}_t, \quad t = 2, \ldots, T. \tag{3.9}$$

From (2.2) and (3.9), it is also easy to see that the endogeneity in the above model arises from

$$\mathbb{E}\left[\Delta\mathbf{u}_t'\mathbf{W}_0\mathbf{y}_t\right] = \sigma_u^2\operatorname{tr}\{\mathbf{W}_0\mathbf{S}_n(\mathbf{z}_t)\} \neq 0 \tag{3.10}$$

in general for $t = 2, \ldots, T$.

We first note that the error term $\Delta\mathbf{u}_t$ in (3.9) is not homoskedastic because $\mathbb{E}[\Delta\mathbf{u}\Delta\mathbf{u}'] = \sigma_u^2\boldsymbol{\Sigma}\otimes\mathbf{I}_n$. We can estimate (3.9) via a pooled local linear estimator without taking the serial correlation in $\{\Delta\mathbf{u}_t\}$ into consideration, or assuming "working independence". However, in the nonparametric literature on random-effects panel data models without the spatial lag term, it is well-known that the pooled local linear estimator is not asymptotically efficient in the presence of cross-sectionally and/or serially correlated errors (e.g., Martins-Filho & Yao, 2009; Su, Ullah & Wang, 2013). It then remains an open question whether we can improve the estimation efficiency by modifying our model so that its new error term is rid of dependence. More concretely, define this new error term as $\Delta\widetilde{\mathbf{u}} = \left(\boldsymbol{\Sigma}^{-1/2} \otimes \mathbf{I}_n\right) \Delta\mathbf{u}$. Then, $\{\Delta\widetilde{u}_{it}\}$ is *i.i.d.* in index i and serially uncorrelated in index t with zero mean and variance σ_u^2, where $\Delta\widetilde{u}_{it} = \sum_{s=2}^{T} \varphi_{ts}\Delta u_{is}$ and φ_{ts} is the $(t-1, s-1)$th element of $\boldsymbol{\Sigma}^{-1/2}$ for $t, s = 2, \ldots, T$. Motivated by Su et al. (2013), we modify model (3.8) as follows:

$$
\begin{aligned}
\mathbf{Y}_t \equiv & \sum_{s=2}^{T} \varphi_{ts}\Delta\mathbf{y}_s^{\dagger} + \sum_{s=2, s\neq t}^{T} \varphi_{ts}\left[\rho(\mathbf{z}_s)\mathbf{W}_0\mathbf{y}_s + \boldsymbol{\phi}_s'\boldsymbol{\beta}(\mathbf{z}_s)\right] \\
= & \varphi_{tt}\left[\rho(\mathbf{z}_t)\mathbf{W}_0\mathbf{y}_t + \boldsymbol{\phi}_t'\boldsymbol{\beta}(\mathbf{z}_t)\right] + \Delta\widetilde{\mathbf{u}}_t, \quad t = 2, \ldots, T,
\end{aligned}
\tag{3.11}
$$

where we move the regressors weighted by the off-diagonal elements of $\boldsymbol{\Sigma}^{-1/2}$ to the left-hand side of (3.11). Clearly, model (3.11) has a homoskedastic error. In addition, we can see that the modified model in (3.11) becomes model (3.9) when we set $\varphi_{tt} = 1$ and $\varphi_{ts} = 0$ for all $t \neq s$.

Applying the Taylor expansion for $\boldsymbol{\gamma}(z_{it})$ at an interior point z, we approximate (3.11) by

$$
\begin{aligned}
Y_{it} &\approx \varphi_{tt}\mathbf{m}_{it}'\left[\boldsymbol{\gamma}(z) + \triangledown\boldsymbol{\gamma}(z)\left(z_{it} - z\right)\right] + \Delta\widetilde{u}_{it} \\
&= \varphi_{tt}\mathbf{m}_{it}'\boldsymbol{\Phi}(z)\mathcal{Z}_{it}\left(z\right) + \Delta\widetilde{u}_{it}
\end{aligned}
\tag{3.12}
$$

for a given (i, t) such that $|z_{it} - z| = o\left(1\right)$, where $\mathcal{Z}_{it}\left(z\right) = \left[1, \left(z_{it} - z\right)/h_0\right]'$, $\boldsymbol{\Phi}(z) = \left[\boldsymbol{\gamma}(z), h_0\nabla\boldsymbol{\gamma}(z)\right]$, and $\nabla^j\boldsymbol{\gamma}(z) = \partial^j\boldsymbol{\gamma}(z)/\partial z^j$ denotes the jth partial derivative of $\boldsymbol{\gamma}(z)$ with respect to z. We then have the following local linear orthogonal moment conditions:

$$
\mathbb{E}\left[\mathbb{Q}_t\left(z\right)'\mathbf{K}_t(h_0, z)\left(\mathbf{Y}_t - \mathbb{M}_t\left(z\right)\mathrm{vec}\left\{\boldsymbol{\Phi}(z)\right\}\right)\right] \approx \mathbf{0}_{2(d_x+1)}
\tag{3.13}
$$

for $t = 2, \ldots, T$, where $\mathbb{M}_t\left(z\right) = \varphi_{tt}\left[\mathcal{Z}_{1t}\left(z\right) \otimes \mathbf{m}_{1t}, \ldots, \mathcal{Z}_{nt}\left(z\right) \otimes \mathbf{m}_{nt}\right]'$ is an $n \times \left[2\left(d_x + 1\right)\right]$ data matrix, $\mathbb{Q}_t\left(z\right) = \left[\mathcal{Z}_{1t}\left(z\right) \otimes \mathbf{q}_{1t}, \ldots, \mathcal{Z}_{nt}\left(z\right) \otimes \mathbf{q}_{nt}\right]'$ is an $n \times \left[2\left(d_x + d_q + 1\right)\right]$ instrument matrix with

$$
\mathbf{q}_t = \mathbf{W}_0\mathbf{S}_n\left(\mathbf{z}_t\right)\begin{bmatrix} \mathbf{g}_1' & \mathbf{x}_{1t}' & z_{1t} \\ \vdots & \vdots & \vdots \\ \mathbf{g}_n' & \mathbf{x}_{nt}' & z_{nt} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{q}_{1t}' \\ \vdots \\ \mathbf{q}_{nt}' \end{bmatrix},
$$

and $\mathbf{K}_t(h_0, z) = \mathrm{diag}\left\{k_{1t}(h_0, z), \ \ldots, k_{nt}(h_0, z)\right\}$ with kernel weights now redefined as $k_{it}(h_0, z) = k\left(\left(z_{it} - z\right)/h_0\right)$.

Next, motivated by the endogeneity relation in (3.10), we see that setting $\mathbf{P}_n = \mathrm{diag}\left\{P_2, \ldots, P_T\right\}$ with $P_t = \mathbf{W}_0\mathbf{S}_n\left(\mathbf{z}_t\right) - n^{-1}\mathrm{tr}\left\{\mathbf{W}_0\mathbf{S}_n\left(\mathbf{z}_t\right)\right\}\mathbf{I}_n$ implies that $\mathbb{E}\left[\Delta\widetilde{\mathbf{u}}'\mathbf{P}_n\Delta\widetilde{\mathbf{u}}\right] = 0$ because $\mathrm{tr}\left\{\mathbf{P}_n\right\} = 0$. With this, we construct the following local quadratic orthogonal moment condition:

$$
\mathbb{E}\left[\left(\mathbf{Y}_t - \mathbb{M}_t\left(z\right)\mathrm{vec}\left\{\boldsymbol{\Phi}(z)\right\}\right)'\mathbf{K}_t(h_0, z)P_t\mathbf{K}_t(h_0, z)\left(\mathbf{Y}_t - \mathbb{M}_t\left(z\right)\mathrm{vec}\left\{\boldsymbol{\Phi}(z)\right\}\right)\right] \approx 0
\tag{3.14}
$$

for $t = 2, \ldots, T$.

Since $\Delta y_{it}^{\dagger}$ is unknown, we replace it with $\Delta\widehat{y}_{it}^{\dagger} = \Delta y_{it} - \mathbf{g}_i'\widehat{\boldsymbol{\theta}}\left(z_{it}\right) + \mathbf{m}_{i,t-1}'\widehat{\boldsymbol{\gamma}}(z_{i,t-1})$, where $\widehat{\boldsymbol{\theta}}\left(z_{it}\right)$ and $\widehat{\boldsymbol{\gamma}}(z_{i,t-1})$ are calculated in the first stage via (3.7). Next, let $\widehat{\mathbf{S}}_n\left(\mathbf{z}_t\right)$ equal $\mathbf{S}_n\left(\mathbf{z}_t\right)$ with $\rho\left(z_{it}\right)$ being replaced with its first-stage estimate $\widehat{\rho}\left(z_{it}\right)$ and $\widehat{\mathbf{q}}_{it}$, $\widehat{\mathbf{q}}_t$ and $\widehat{P}_t$ respectively equal $\mathbf{q}_{it}$, $\mathbf{q}_t$ and P_t with $\mathbf{S}_n\left(\mathbf{z}_t\right)$ being replaced with $\widehat{\mathbf{S}}_n\left(\mathbf{z}_t\right)$. Also, $\widehat{Y}_{it} \equiv \sum_s \varphi_{ts}\Delta\widehat{y}_{is}^{\dagger} - \sum_{s\neq t}\varphi_{ts}\mathbf{m}_{is}'\widehat{\boldsymbol{\gamma}}(z_{is})$.

Lastly, we define $\widehat{\mathbf{P}}_n = \operatorname{diag}\left\{\widehat{P}_2, \ldots, \widehat{P}_T\right\}$, $\widehat{\mathbf{Y}} = \left[\widehat{\mathbf{Y}}_2', \ldots, \widehat{\mathbf{Y}}_T\right]'$, $\mathbb{M}(z) = \left[\mathbb{M}_2(z)', \ldots, \mathbb{M}_T(z)'\right]'$, $\widehat{\mathbb{Q}}(z) = \left[\widehat{\mathbb{Q}}_2(z)', \ldots, \widehat{\mathbb{Q}}_T(z)'\right]'$ and $\mathbf{K}_{h_0}(z) = \operatorname{diag}\{\mathbf{K}_2(h_0, z), \ldots, \mathbf{K}_T(h_0, z)\}$. We then construct our second-stage nonparametric GMM estimator as follows:

$$\widetilde{\boldsymbol{\Phi}}(z) = \arg\min_{\boldsymbol{\Phi}(z)} \; \boldsymbol{g}_n(\boldsymbol{\Phi}(z))' \boldsymbol{g}_n(\boldsymbol{\Phi}(z)), \tag{3.15}$$

where we define

$$\boldsymbol{g}_n(\boldsymbol{\vartheta}) = \begin{bmatrix} \left(\widehat{\mathbf{Y}} - \mathbb{M}(z)\mathrm{vec}\{\boldsymbol{\vartheta}\}\right)' \mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n\mathbf{K}_{h_0}(z)\left(\widehat{\mathbf{Y}} - \mathbb{M}(z)\mathrm{vec}\{\boldsymbol{\vartheta}\}\right) \\ \widehat{\mathbb{Q}}(z)'\mathbf{K}_{h_0}(z)\left(\widehat{\mathbf{Y}} - \mathbb{M}(z)\mathrm{vec}\{\boldsymbol{\vartheta}\}\right) \end{bmatrix}. \tag{3.16}$$

The limiting results for the second-stage estimator require the following additional assumptions.

Assumption 5 $\max_{\mathbf{z}\in\mathcal{S}_z\times\mathcal{S}_z} \left\|\widehat{\boldsymbol{\Theta}}(\mathbf{z}) - \boldsymbol{\Theta}(\mathbf{z})\right\| = O_p\left(h^2 + \sqrt{\ln n/(nh^2)}\right)$.

Assumption 6 *(i) z_{it} has a common pdf $f_t(z)$ with respect to the Lebesgue measure over $\mathcal{S}_z$; (z_{it}, z_{is}) has a common pdf $f_{t,s}(z_1, z_2)$ with respect to the Lebesgue measure over $\mathcal{S}_z \times \mathcal{S}_z$ for all t and s; (ii) in the neighborhood of an interior point z, $\boldsymbol{\beta}(z)$, $\rho(z)$, $f_t(z)$, $\mathbb{E}[\mathbf{g}_i\check{\mathbf{m}}_{it}'|z_{it} = z]$, $\mathbb{E}[\mathbf{g}_i'\mathbf{g}_i\check{\mathbf{m}}_{it}'\check{\mathbf{m}}_{it}|z_{it} = z]$, $\mathbb{E}[\mathbf{m}_{it}\check{\mathbf{m}}_{it}'|z_{it} = z]$, $\mathbb{E}[\mathbf{m}_{it}'\mathbf{m}_{it}\check{\mathbf{m}}_{it}'\check{\mathbf{m}}_{it}|z_{it} = z]$, $\mathbb{E}[\check{\mathbf{m}}_{it}\check{\mathbf{m}}_{it}'|z_{it} = z]$ and $\mathbb{E}[\check{\mathbf{m}}_{it}'\check{\mathbf{m}}_{it}\check{\mathbf{m}}_{it}'\check{\mathbf{m}}_{it}|z_{it} = z]$ are all twice continuously differentiable for all t; (iii) $f_{t,s}(z, z)$, $\mathbb{E}[\check{\mathbf{m}}_{it}\check{\mathbf{m}}_{is}'|z, z]$, and $\mathbb{E}[\check{\mathbf{m}}_{it}'\check{\mathbf{m}}_{it}\check{\mathbf{m}}_{is}'\check{\mathbf{m}}_{is}|z, z]$ are all twice continuously differentiable for all s and t, where $\mathbb{E}[\cdot|z, z] = \mathbb{E}[\cdot|z_{it} = z, z_{is} = z]$; (iv) $\varkappa_B(h_0, z)$ is a non-singular defined in Lemma 6 in the Appendix.*

Assumption 7 *As $n \to \infty$: (i) $h_0 \to 0$ and $\lim_{n\to\infty} nh_0^5 = c_0 > 0$; (ii) $h \to 0$, $nh^4 \to \infty$; (iii) $h/h_0 \to 0$ and $nh^2h_0^2/\ln n \to \infty$.*

Assumption **5** strengthens the pointwise convergence of $\widehat{\boldsymbol{\Theta}}(\mathbf{z})$ to a uniform convergence, which can be shown along the lines of Masry (1996). Assumption **6(i)–(iii)** are regularity conditions imposed in the local linear estimation, while Assumption **6(iv)** ensures the existence of the proposed estimator. Assumption **7(i)** implies that the second-step bandwidth h_0 is of order $n^{-1/5}$, Assumption **7(ii)** is required for the derivation of Assumption **5**, where the first-stage estimation has an asymptotically ignorable impact on the second-stage estimator under Assumption **7(iii)**, which implies that $h = cn^{-\alpha}$ with $1/5 < \alpha < 1/4$ if we set $h_0 = c_0 n^{-1/5}$.

Theorem 2 *Under Assumptions **1**–**3** and **5**–**7**, at an interior point z, we have*

$$\sqrt{nh_0}\left[\widetilde{\boldsymbol{\gamma}}(z) - \boldsymbol{\gamma}(z) - \varkappa_B(h_0, z)^{-1}\varkappa_A(h_0, z)'\right] \overset{d}{\to} \mathbb{N}\left(\mathbf{0}_{d_x+1}, \sigma_u^2\upsilon_{2,0}(k)\varkappa_B(h_0, z)^{-1}\boldsymbol{\Omega}(z)\varkappa_B(h_0, z)^{-1}\right),$$

where $\varkappa_A(h_0, z) = O_p(h^2)$, $\varkappa_B(h_0, z)$ and $\boldsymbol{\Omega}(z)$ are respectively defined in Lemmas 5–7 in the Appendix.

By the proof of Theorem **2**, we have $\widetilde{\boldsymbol{\gamma}}(z) - \boldsymbol{\gamma}(z) = O_p\left(h_0^2 + (nh_0)^{-1/2}\right)$. The impact of the initial first-stage estimator on the second-stage estimator vanishes asymptotically as $h/h_0 \to 0$ and $nh^2h_0^2/\ln n \to \infty$ hold. Importantly, Theorem **2** holds true for $d_g \geq 0$. Again, the localized quadratic moments have asymptotically non-ignorable contribution to $\varkappa_B(h_0, z)$ if $\operatorname{diag}\{\mathbf{P}_n\} \neq 0$.

By the proofs given in Lemmas **5–7**, $\varkappa_A(h_0,z)$, $\varkappa_B(h_0,z)$ and $\mathbf{\Omega}(z)$ all depend on $\{\varphi_{tt}\}$, which means that the derived estimator has the asymptotic bias and variance different from those of the pooled local linear estimator. Unfortunately, it is difficult to conclude which estimator is better due to the complexity of our formulas. This result differs from Su et al.'s (2013) findings where, excluding the spatial lag term from models (3.9) and (3.11), they show that the local linear estimator from the modified model (3.11) has the same asymptotic bias but smaller asymptotic variance than the pooled local linear estimator. So far, our results indicate that, for the spatial autoregressive panel data model with fixed effects, Su et al.'s (2013) method may or may not be able to improve estimation efficiency over the pooled estimation under the "working independence" assumption when the nonparametric IV-based GMM estimation method is concerned.

4 Sieve Estimator of $\boldsymbol{\theta}(z)$

Having estimated $\rho(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ consistently at the conventional convergence rate in the second stage, we next discuss how to consistently estimate the unknown functional curves $\boldsymbol{\theta}(\cdot)$ in front of the time-invariant regressors $\mathbf{g}_i$. Specifically, we consider the following regression model:

$$\dot{y}_{it} = \mathbf{g}_i'\boldsymbol{\theta}(z_{it}) + \mu_i + u_{it}, \quad i = 1,\ldots,n,\ t = 1,\ldots,T,$$

where $\dot{y}_{it} \equiv y_{it} - \mathbf{m}_{it}'\boldsymbol{\gamma}(z_{it})$. Since $\dot{y}_{it}$ is unknown, we replace it with $\widetilde{\dot{y}}_{it} = y_{it} - \mathbf{m}_{it}'\widetilde{\boldsymbol{\gamma}}(z_{it})$, where $\widetilde{\boldsymbol{\gamma}}(z_{it})$ is the second-stage GMM estimator. Hence, we suggest estimating $\boldsymbol{\theta}(z)$ from

$$\widetilde{\dot{y}}_{it} \approx \mathbf{g}_i'\boldsymbol{\theta}(z_{it}) + \mu_i + u_{it}, \quad i = 1,\ldots,n,\ t = 1,\ldots,T. \tag{4.1}$$

We first note that applying the nonparametric smoothed least-squares dummy variable approach developed by Sun et al. (2009) is infeasible for the estimation of $\boldsymbol{\theta}(z)$; see the detailed discussion in Section 5.3. We therefore choose to employ a series approximation method. Specifically, letting $\{\phi_1(\cdot), \phi_2(\cdot), \ldots\}$ be a sequence of B-spline series, we approximate $\theta_l(z)$ by $\theta_l^*(z) = \boldsymbol{\psi}_l'\boldsymbol{\phi}_{L_n}(z)$ for $l = 1,\ldots,d_g$, where $\boldsymbol{\psi}_l$ is of dimension L_n and, for any integer $\kappa > 0$, we denote $\boldsymbol{\phi}_\kappa(v) = [\phi_1(v),\ldots,\phi_\kappa(v)]'$. Then, $\mathbf{g}_i'\boldsymbol{\theta}(z_{it})$ can be approximated by

$$\mathbf{g}_i'\boldsymbol{\theta}^*(z_{it}) = \left[\mathbf{g}_i \otimes \boldsymbol{\phi}_{L_n}(z_{it})\right]'\boldsymbol{\psi} \equiv \mathcal{X}_{it}'\boldsymbol{\psi},$$

where $\boldsymbol{\psi} = \left[\boldsymbol{\psi}_1',\ldots,\boldsymbol{\psi}_{d_g}'\right]'$ and $\mathcal{X}_{it} \equiv \mathbf{g}_i \otimes \boldsymbol{\phi}_{L_n}(z_{it})$ are both $(d_gL_n)\times 1$ vectors. Stacking up $\{\mathcal{X}_{it}'\}$ in the ascending order of index i first then index t gives an $(nT)\times(d_gL_n)$ data matrix $\mathcal{X}$. The series-based least-squares objective function is then given by

$$\min_{(\boldsymbol{\mu},\boldsymbol{\psi})} \left(\widetilde{\dot{\mathbf{y}}} - \mathcal{X}\boldsymbol{\psi} - \mathbf{D}\boldsymbol{\mu}\right)'\left(\widetilde{\dot{\mathbf{y}}} - \mathcal{X}\ \boldsymbol{\psi} - \mathbf{D}\boldsymbol{\mu}\right), \tag{4.2}$$

where $\mathbf{D} = \mathbf{I}_n \otimes \mathbf{i}_T$ is an $(nT)\times n$ matrix, $\boldsymbol{\mu}$ is as defined in Section 2, and $\widetilde{\dot{\mathbf{y}}}$ is an $(nT)\times 1$ vector that stacks up $\widetilde{\dot{y}}_{it}$ in the ascending order of index i first then index t. Applying the partitioned OLS yields

$$\widetilde{\boldsymbol{\psi}} = \left(\mathcal{X}'\mathbf{M}_{\mathbf{D}}\mathcal{X}\right)^{-1}\mathcal{X}'\mathbf{M}_D\widetilde{\dot{\mathbf{y}}}, \tag{4.3}$$

where $\mathbf{M}_{\mathbf{D}} = \mathbf{I}_{nT} - \mathbf{D}\left(\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{D}' = \mathbf{I}_{nT} - T^{-1}\mathbf{I}_n \otimes (\mathbf{i}_T\mathbf{i}_T')$. Hence, our (third-stage) series estimator of $\boldsymbol{\theta}(z)$ is given by

$$\widetilde{\boldsymbol{\theta}}(z) = \left[\widetilde{\boldsymbol{\psi}}_1'\boldsymbol{\phi}_{L_n}(z),\ldots,\widetilde{\boldsymbol{\psi}}_{d_g}'\boldsymbol{\phi}_{L_n}(z)\right]'. \tag{4.4}$$

Below are some regularity assumptions required for our limiting results.

Assumption 8 *For every L_n, there exist constants $\underline{c}$ and $\overline{c}$ such that* $0 < \underline{c} \leq \lambda_{\min}(\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}) \leq \lambda_{\max}(\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}) \leq \overline{c} < \infty$ *, where* $\mathbf{\Sigma}_{\boldsymbol{\phi\phi}} \equiv \sum_{t=1}^{T} \mathbb{E}\left[(\mathbf{g}_i\mathbf{g}_i') \otimes \left(\overline{\boldsymbol{\phi}}_{L_n}(z_{it})\,\overline{\boldsymbol{\phi}}_{L_n}(z_{it})'\right)\right]$ *and* $\overline{\boldsymbol{\phi}}_{L_n}(z_{it}) = \boldsymbol{\phi}_{L_n}(z_{it}) - T^{-1}\sum_{s=1}^{T} \boldsymbol{\phi}_{L_n}(z_{is})$.

Assumption 9 *For any $l = 1, \ldots, d_g$, there exists $\boldsymbol{\psi}_l$ such that*

$$\max_{1\leq l\leq d_g, z\in\mathcal{S}_z} \left|\theta_l(z) - \boldsymbol{\psi}_l'\boldsymbol{\phi}_{L_n}(z)\right| \leq CL_n^{-\xi} \tag{4.5}$$

for a sufficiently large L_n and $\xi > 2$.

Assumption 10 *As $n \to \infty$, $L_n \to \infty$, $L_n(\log L_n)/n \to 0$ and $nL_n^{-1-2\xi} \to 0$.*

Assumption **8** ensures that $\mathcal{X}'\mathbf{M_D}\mathcal{X}/n$ converges to a non-singular matrix $\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}$ for a sufficiently large n. Assumption **9** is similar to Assumption 3 in Newey (1997). In fact, if $\theta_l(\cdot)$'s and $\phi_l(\cdot)$'s are all ξ-smooth,[6] condition (4.5) holds over any compact support S_z by Theorem 1.1 in Dzyadyk & Shevchuk (2008, p. 381). If we set $L_n = cn^r$, Assumption **10** requires $r \in (1/(1+2\xi), 1)$.

Theorem 3 *Under Assumptions **1**–**10**, and if* $\max_{1\leq t\leq T} \mathbb{E}\left[|\mathbf{g}_i'\mathbf{g}_i|^{2+\delta}|u_{it}|^{2+\delta}\right] \leq C$ *for some $\delta > 0$, we have*

$$\sqrt{n}\mathbf{\Lambda}_n^{-1/2}(z)\left[\widetilde{\boldsymbol{\theta}}(z) - \boldsymbol{\theta}(z)\right] \xrightarrow{d} \mathbb{N}\left(\mathbf{0}_{d_g},\ \sigma_u^2\mathbf{I}_{d_g}\right), \tag{4.6}$$

where $\mathbf{\Lambda}_n(z) = \left[\mathbf{I}_{d_g} \otimes\ \boldsymbol{\phi}_{L_n}(z)\right]' \mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]$.

Theorem **3** indicates that the sieve estimator $\widetilde{\boldsymbol{\theta}}(z)$ is a consistent estimator of $\boldsymbol{\theta}(z)$ and that the first two stages of the estimation procedure have asymptotically negligible effects on the estimation of $\boldsymbol{\theta}(z)$. For any constant $\alpha \neq \mathbf{0}_{d_g}$, we have $|\alpha'\mathbf{\Lambda}_n(z)\alpha| \leq \lambda_{\max}^{-1}(\mathbf{\Sigma}_{\boldsymbol{\phi\phi}})\left\|\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]\alpha\right\| \leq CL_n$. Therefore, $\widetilde{\boldsymbol{\theta}}(z) - \boldsymbol{\theta}(z) = O_p\left(\sqrt{L_n/n}\right)$ under Assumption **10**.

5 Other Estimation Issues

This section provides brief arguments about an alternative estimator based on the within-groups transformation in Section 5.1, the spatial stationarity condition in Section 5.2 and the infeasibility of a nonparametric smoothed dummy variable approach in Section 5.3.

5.1 The Within-Groups Estimator of $\rho(z)$ and $\boldsymbol{\beta}(z)$

As an alternative to first-differencing transformation, one can remove the unobserved fixed effects from model (1.2) by applying the within-groups transformation which yields

$$\overline{y}_{it} = \mathbf{g}_i'\sum_{s=1}^{T}\xi_{st}\boldsymbol{\theta}(z_{is}) + \sum_{s=1}^{T}\xi_{st}\mathbf{m}_{is}'\boldsymbol{\gamma}(z_{is}) + \overline{u}_{it}, \quad i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{5.1}$$

[6]A function $h(\cdot)$ is called p-smooth for some real value $p > 0$ if it is $[p]$-times continuously differentiable and $\left|\nabla^{[p]}h(\omega) - \nabla^{[p]}h(\omega')\right| \leq M|\omega - \omega'|^{p-[p]}$ for any $\omega \in \mathcal{R}^+$ and $\omega' \in \mathcal{R}^+$, where $[p]$ is the largest positive integer less than p and $\nabla^{[p]}h(\omega)$ is the $[p]$th order derivative of $h(\cdot)$.

where we denote $\overline{a}_{it} = a_{it} - T^{-1}\sum_{s=1}^{T} a_{is}$ for $a = \{y, u\}$, and $\xi_{st} = 1 - T^{-1}$ if $s = t$ and $-T^{-1}$ otherwise. Following our first-differencing method, we notice that the first-stage estimation of model (5.1) suffers from a more severe "curse-of-dimensionality" problem than the first-differencing method when $T > 2$ as we need to approximate $\dot{\boldsymbol{\theta}}(\mathbf{z}_i) = \sum_{s=1}^{T} \xi_{st}\boldsymbol{\theta}(z_{is})$ and $\boldsymbol{\gamma}(\mathbf{z}_i) = \left[\boldsymbol{\gamma}(z_{i1})', \ldots, \boldsymbol{\gamma}(z_{iT})'\right]'$ in (5.1) by $\dot{\boldsymbol{\theta}}(\mathbf{z}) = \sum_{s=1}^{T} \xi_{st}\boldsymbol{\theta}(z_s)$ and $\boldsymbol{\gamma}(\mathbf{z}) = \left[\boldsymbol{\gamma}(z_1)', \ldots, \boldsymbol{\gamma}(z_T)'\right]'$ for i such that $\|\mathbf{z}_i - \mathbf{z}\| = o(1)$, where $\mathbf{z} = [z_1, \ldots, z_T]'$ be an interior point with $z_s \neq z_{s'}$ for $s \neq s'$. By the conventional nonparametric kernel estimation theory, we expect that the first-stage estimator based on the within-group transformation satisfies $MSE\left[\widehat{\boldsymbol{\Theta}}(\mathbf{z})\right] \approx O_p\left(h^2 + \left(nh^T\right)^{-1/2}\right)$, where the asymptotic variance is of order $\left(nh^T\right)^{-1}$ because the unknown curves to be estimated are functions of T arguments in the transformed model (5.1). In fact, as we show in Sun & Malikov (2015), the within-groups transformation method provides a consistent estimator only if $T < 3$ and $T < 5$ when the local constant and the local linear estimators are respectively applied in the first-stage estimation. Given finite samples, the less accurate first-stage estimator may reduce the estimation accuracy of the second-stage estimator. Therefore, this paper focuses on the limit results from the first-differencing method.[7]

5.2 Spatial Stationarity

The spatial stationarity of $\{y_{it}\}$ requires that $\max_{1\le j\le n, 1\le t\le T} |\lambda_j\{\boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\}| < 1$ almost surely. Following Kelejian & Prucha (2010), one can normalize the spatial weighting matrix $\mathbf{W}_0$ such that its largest eigenvalue in absolute value is no greater than one, so that the spatial stationarity condition holds if $|\rho(z)| < 1$ for any $z \in \mathcal{S}_z$. To impose this restriction on the spatial lag parameter function, we apply Hall & Huang's (2001) "tilting" procedure. The procedure essentially mutes or magnifies the impact of any given data point used in the estimation. This allows us to impose the spatial stationarity condition *post*-estimation via a quadratic programming technique. The idea is to slightly reweigh observations used in the estimation so that the spatial stationarity condition is satisfied in the local neighborhood of z. Since the estimator derived using both linear and quadratic orthogonality conditions does not have an analytical solution, the "tilting" procedure proposed by Hall & Huang (2001) does not apply. Here, we therefore limit our attention to a simpler estimator which makes use of linear moments only and hence is valid when $\boldsymbol{\beta}(z) \neq \mathbf{0}_{d_x}$ and $\boldsymbol{\theta}(z) \neq \mathbf{0}_{d_g}$ over at least one non-empty subset. It is noteworthy however that, when one does incorporate quadratic conditions during the estimation, the non-singularity condition can be imposed even more easily via box constraints on $\rho(z)$ during the numerical optimization.

Specifically, when using only linear moment conditions in (3.13), the second-stage estimator of $\rho(z)$ derived in Section 3.1 can be abbreviated as $\widetilde{\rho}(z) = \sum_{i=1}^{n}\sum_{t=2}^{T} \omega_{it}(\mathbf{G}, \mathbf{X}, z, h_0)\widehat{Y}_{it}$, where $\omega_{it}(\mathbf{G}, \mathbf{X}, z, h_0)$ is a local weight assigned to $\widehat{Y}_{it}$ and is the (it)th (column) element in the first row of $\left[\mathbb{M}(z)'\mathbf{K}_{h_0}(z)'\widehat{\mathbb{Q}}(z)\widehat{\mathbb{Q}}(z)'\mathbf{K}_{h_0}(z)\mathbb{M}(z)\right]^{-1}\mathbb{M}(z)'\mathbf{K}_{h_0}(z)'\widehat{\mathbb{Q}}(z)\widehat{\mathbb{Q}}(z)'\mathbf{K}_{h_0}(z)$. We then construct

$$\widetilde{\rho}(z|\boldsymbol{p}) = n(T-1)\sum_{i=1}^{n}\sum_{t=2}^{T} p_{it}\omega_{it}(\mathbf{G}, \mathbf{X}, z, h_0)\widehat{Y}_{it}, \tag{5.2}$$

where $\boldsymbol{p} = (p_{12}, \ldots, p_{1T}, \ldots p_{n2}, \ldots p_{nT})'$ is the sequence of additional weights such that $\sum_{i=1}^{n}\sum_{t=2}^{T} p_{it} = 1$. Note that p_{it} equals $1/[n(T-1)]$ (i.e., uniform weights) when the restriction is not imposed.

[7] We refer readers to Sun & Malikov (2015) for the detailed limit results of the within-group estimation method. This paper is available upon request from the authors.

If necessary, we can impose the non-singularity condition by selecting weights $\boldsymbol{p}$ that minimize the following L_2-metric:

$$\mathcal{D}(\boldsymbol{p}) = \Big([n\,(T-1)]^{-1}\, \mathbf{i}_{n(T-1)} - \boldsymbol{p} \Big)' \Big([n\,(T-1)]^{-1}\, \mathbf{i}_{n(T-1)} - \boldsymbol{p} \Big)$$

subject to $\mathbf{i}'_{n(T-1)}\boldsymbol{p} = 1$ and $-1 < \widetilde{\rho}(z_{it}|\boldsymbol{p}) < 1$ for any (i,t). Here, $\mathcal{D}(\boldsymbol{p})$ is the sum of squared deviations of p_j from the unrestricted value of $[n\,(T-1)]^{-1}$. In our choice of the distance metric, we follow Du, Parmeter & Racine (2013), which allows $\boldsymbol{p}$ to be both positive and negative.[8] The minimization problem is solved via a standard quadratic programming technique. Let $\widehat{\boldsymbol{p}}$ be the solution to this optimization problem. We then use $\widetilde{\rho}(z|\widehat{\boldsymbol{p}})$ as the final estimator for $\rho\,(z)$. Since the proofs of consistency and asymptotic normality of the "tilted" estimator are tedious and closely follow those given in Malikov & Sun (2017), we omit the details here.

5.3 Infeasibility of a Smoothed Dummy Variable Approach

As briefly mentioned earlier, an alternative approach to removing fixed effects from functional-coefficient models put forward in the literature is a nonparametric generalization of the so-called "dummy variables" approach by Sun et al. (2009). In the instance of a kernel-based *least-squares* estimator, Sun et al.'s (2009) method closely resembles a traditional least-squares dummy variables (LSDV) estimator of parametric fixed-effects panel data models. While this method can successfully be applied to functional-coefficient models with fixed effects that suffer from endogeneity due to selectivity (see Malikov, Kumbhakar & Sun, 2016), it however *cannot* be extended to varying coefficient panel data models subject to the general form of endogeneity stemming from the simultaneity of regressors. The latter is due to singularity in the first-order condition of the nonparametric GMM objective function, which we demonstrate below for the case when $d_g = 0$ (i.e., no time-invariant regressors) and for the local constant estimator, without loss of generality.

Approximating our model in (1.2) around z yields

$$y_{it} \approx \mathbf{m}'_{it}\boldsymbol{\gamma}(z) + \mu_i + u_{it} \tag{5.3}$$

for the (i,t)th observation with $|z_{it} - z| = o\,(1)$. The corresponding local moment condition is

$$\mathbb{E}\left[\mathbf{Q}'\mathbf{K}_h(z)\,(\mathbf{y} - \mathbf{M}\boldsymbol{\gamma}(z) - \mathbf{D}\boldsymbol{\mu})\right] \approx \mathbf{0}_d, \tag{5.4}$$

where $\mathbf{M}$, $\mathbf{Q}$ and $\mathbf{y}$ stack up $\mathbf{m}'_{it}$, $\breve{\mathbf{m}}'_{it}$ and y_{it} in the ascending order of index i first then index t, and a typical element of the diagonal matrix $\mathbf{K}_h(z)$ is $k\,((z_{it} - z)\,/h)$.

The kernel-weighted nonparametric GMM problem corresponding to the orthogonality condition in (5.4) is given by

$$\min_{\boldsymbol{\gamma}(z)} \; \left[\mathbf{Q}'\mathbf{K}_h(z)\,(\mathbf{y} - \mathbf{M}\boldsymbol{\gamma}(z) - \mathbf{D}\boldsymbol{\mu})\right]' \mathbf{Q}'\mathbf{K}_h(z)\,(\mathbf{y} - \mathbf{M}\boldsymbol{\gamma}(z) - \mathbf{D}\boldsymbol{\mu})\,. \tag{5.5}$$

The core idea of the "dummy variables" approach is to concentrate out the unknown fixed effects from the objective function. To do so, one needs to substitute the first-order condition for the optimization problem in (5.5) with respect to $\boldsymbol{\mu}$ back into the objective function. This first-order condition with respect to $\boldsymbol{\mu}$ is

$$\mathbf{D}'\mathbf{K}_h(z)\mathbf{Q}\mathbf{Q}'\mathbf{K}_h(z)\,(\mathbf{y} - \mathbf{M}\boldsymbol{\gamma}(z) - \mathbf{D}\boldsymbol{\mu}) = \mathbf{0}_n, \tag{5.6}$$

[8] Hall & Huang (2001) use a power divergence metric which has a rather complex form and is only valid for non-negative weights.

which can be manipulated to obtain

$$\widehat{\boldsymbol{\mu}} = \left[\mathbf{D}'\mathbf{K}_h(z)\mathbf{Q}\mathbf{Q}'\mathbf{K}_h(z)\mathbf{D}\right]^{-1}\mathbf{D}'\mathbf{K}_h(z)\mathbf{Q}\mathbf{Q}'\mathbf{K}_h(z)\left[\mathbf{y} - \mathbf{M}\boldsymbol{\gamma}(z)\right]. \tag{5.7}$$

However, note that the above first-order condition suffers from singularity. To see this, let $\mathbf{A} = \mathbf{Q}'\mathbf{K}_h(z)\mathbf{D}$. Then, $\mathbf{D}'\mathbf{K}_h(z)\mathbf{Q}\mathbf{Q}'\mathbf{K}_h(z)\mathbf{D} = \mathbf{A}'\mathbf{A}$. Since $\mathbf{A}$ is a $d \times n$ matrix, $\mathbf{A}'\mathbf{A}$ is a square matrix of dimension n. Given that $n > d$, we know that the rank of $\mathbf{A}'\mathbf{A}$ is no greater than d, rendering matrix $\mathbf{D}'\mathbf{K}_h(z)\mathbf{Q}\mathbf{Q}'\mathbf{K}_h(z)\mathbf{D}$ to be singular. Thus, $\widehat{\boldsymbol{\mu}}$ cannot be solved for and cannot be concentrated out of the GMM objective function (5.5). The "dummy variables" approach is infeasible. The above argument continues to hold if $d_g > 0$.

6 Test for Spatial Endogeneity

Given that our semiparametric spatial autoregressive model (1.2) nests the traditional functional-coefficient fixed-effects model in (1.1) as a special case, one may wish to formally discriminate between the two models. In this section, we are interested in testing for relevance of the spatial lag term in model (1.2), and the proposed is, essentially, a test for spatial endogeneity. Specifically, we consider the following null and alternative hypotheses:

$$H_0 : \Pr\{\rho(z) = 0\} = 1 \quad \text{vs.} \quad H_1 : \Pr\{\rho(z) = 0\} < 1.$$

That is, under H_0, our model (1.2) becomes model (1.1). Our proposed test statistic is based on a weighted squared distance between $\widetilde{\boldsymbol{\beta}}_0(z)$ and $\widetilde{\boldsymbol{\beta}}_1(z)$, which are the second-stage local-constant estimators of $\boldsymbol{\beta}(z)$ under the null and alternative hypotheses, respectively. Since the estimation of $\boldsymbol{\theta}(z)$ involves a three-step procedure, for the sake of simplicity and feasibility, we propose to construct our test statistic only from the estimators for $\boldsymbol{\beta}(z)$ under both hypotheses.

We start by defining $\mathcal{X}_j = \left[\mathcal{X}_{j,1}', \ldots, \mathcal{X}_{j,n}'\right]'$ for $j = 1, 2$, where $\mathcal{X}_{1,i} = [\mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}]'$ and $\mathcal{X}_{2,i} = \sum_{j \neq i} w_{ij}\left[y_{j2}, \ldots, y_{jT}\right]'$, and $\widehat{\boldsymbol{\Xi}}_{h_0}(z) \equiv \mathbf{K}_{h_0}(z)'\widehat{\mathcal{Q}}\widehat{\mathcal{Q}}'\mathbf{K}_{h_0}(z)$, where $\widehat{\mathcal{Q}} = \left[\widehat{\mathcal{Q}}_2', \ldots, \widehat{\mathcal{Q}}_T'\right]'$ with $\widehat{\mathcal{Q}}_t = [\widehat{\mathbf{q}}_{1t}, \ldots, \widehat{\mathbf{q}}_{nt}]'$. Next, we define $\mathbf{M}_2(z) = \mathbf{I}_{n(T-1)} - \mathcal{X}_2\left[\mathcal{X}_2{}'\widehat{\boldsymbol{\Xi}}_{h_0}(z)\mathcal{X}_2\right]^{-1}\mathcal{X}_2{}'\widehat{\boldsymbol{\Xi}}_{h_0}(z)$ and $\mathbf{S}_{h_0}(z) = \mathbf{M}_2(z)'\widehat{\boldsymbol{\Xi}}_{h_0}(z)\mathbf{M}_2(z)$, where we can show that $\mathbf{S}_{h_0}(z) = \mathbf{M}_2(z)'\widehat{\boldsymbol{\Xi}}_{h_0}(z) = \widehat{\boldsymbol{\Xi}}_{h_0}(z)\mathbf{M}_2(z)$. Then, the local constant estimator of $\boldsymbol{\beta}(z)$ for model (3.8) under H_1 is calculated as $\widetilde{\boldsymbol{\beta}}_1(z) = \left[\mathcal{X}_1{}'\mathbf{S}_{h_0}(z)\mathcal{X}_1\right]^{-1}\mathcal{X}_1{}'\mathbf{S}_{h_0}(z)\Delta\widehat{\mathbf{y}}^{\dagger}$, which is based on the local linear orthogonal moment condition only and is used to simply the formula of our test statistic. Given the above, we construct our test statistic as follows:

$$\begin{aligned}
\mathcal{T}_n &= \int \left[\widetilde{\boldsymbol{\beta}}_1(z) - \widetilde{\boldsymbol{\beta}}_0(z)\right]'\left[\mathcal{X}_1'\mathbf{S}_{h_0}(z)\mathcal{X}_1\right]^2\left[\widetilde{\boldsymbol{\beta}}_1(z) - \widetilde{\boldsymbol{\beta}}_0(z)\right]dz \\
&= \int \left[\Delta\widehat{\mathbf{y}}^{\dagger} - \mathcal{X}_1\widetilde{\boldsymbol{\beta}}_0(z)\right]'\mathbf{S}_{h_0}(z)'\mathcal{X}_1\mathcal{X}_1'\mathbf{S}_{h_0}(z)\left[\Delta\widehat{\mathbf{y}}^{\dagger} - \mathcal{X}_1\widetilde{\boldsymbol{\beta}}_0(z)\right]dz \\
&= \int \left[\Delta\widehat{\mathbf{y}}^{\dagger} - \mathcal{X}_1\widetilde{\boldsymbol{\beta}}_0(z)\right]'\widehat{\boldsymbol{\Xi}}_{h_0}(z)\mathbf{M}_2(z)\mathcal{X}_1\mathcal{X}_1'\mathbf{M}_2(z)'\widehat{\boldsymbol{\Xi}}_{h_0}(z)\left[\Delta\widehat{\mathbf{y}}^{\dagger} - \mathcal{X}_1\widetilde{\boldsymbol{\beta}}_0(z)\right]dz \geq 0,
\end{aligned}$$

where the typical element of $\Delta\widehat{\mathbf{y}}^{\dagger} - \mathcal{X}_1\widetilde{\boldsymbol{\beta}}_0(z)$ is given by

$$\begin{aligned}
\widetilde{\varepsilon}_{it}(z) &\equiv \Delta\widehat{y}_{it}^{\dagger} - \mathbf{x}_{it}'\widetilde{\boldsymbol{\beta}}_0(z) = \Delta y_{it} - \mathbf{g}_i'\widehat{\dot{\boldsymbol{\theta}}}_1(z_{it}) + \mathbf{m}_{i,t-1}'\widehat{\gamma}_1(z_{i,t-1}) - \mathbf{x}_{it}'\widetilde{\boldsymbol{\beta}}_0(z) \\
&\equiv \widehat{\epsilon}_{it} + \rho(z_{it})\sum_{j \neq i} w_{ij}y_{jt} + \mathbf{x}_{it}'\left[\boldsymbol{\beta}(z_{it}) - \widetilde{\boldsymbol{\beta}}_0(z)\right] + \Delta u_{it}, \quad i = 1, \ldots, n, \ t = 2, \ldots, T,
\end{aligned}$$

with $\widehat{\epsilon}_{it} \equiv \mathbf{g}_i' \left[\dot{\boldsymbol{\theta}}(z_{it}) - \widehat{\dot{\boldsymbol{\theta}}}_1(z_{it})\right] - \mathbf{m}_{i,t-1}' [\boldsymbol{\gamma}(z_{i,t-1}) - \widehat{\boldsymbol{\gamma}}_1(z_{i,t-1})]$, and $\widehat{\dot{\boldsymbol{\theta}}}_1(\cdot)$ and $\widehat{\boldsymbol{\gamma}}_1(\cdot)$ are estimators calculated under H_1.

Under H_0, we have $\widetilde{\varepsilon}_{it}(z) = \Delta u_{it} + o_p(1)$, whereas, under H_1, $\widetilde{\varepsilon}_{it}(z) = \Delta u_{it} + \rho(z_{it}) \sum_{j \neq i} w_{ij} y_{jt}$ $+\mathbf{x}_{it}' \left[\boldsymbol{\beta}(z_{it}) - \widetilde{\boldsymbol{\beta}}_0(z)\right] + o_p(1)$. Hence, intuitively, we expect $\mathcal{T}_n$ to go explosive at a faster speed under H_1 than under H_0.

We next describe how to calculate $\widetilde{\boldsymbol{\beta}}_0(z)$ under H_0. Since, when $\rho(z) = 0$, (3.3) becomes $\Delta y_{it} \approx \mathbf{g}_i' \dot{\boldsymbol{\theta}}(\mathbf{z}) + \boldsymbol{\xi}' \boldsymbol{x}_{it}' \boldsymbol{\beta}(\mathbf{z}) + \Delta u_{it}$, we consider the following kernel-weighted least-squares objective function:

$$\min_{\boldsymbol{\Theta}_0(\mathbf{z})} \left[\boldsymbol{\Delta}\mathbf{y} - \mathcal{M}\boldsymbol{\Theta}_0(\mathbf{z})\right]' \mathbf{K}_{\widetilde{h}}(\mathbf{z}) \left[\boldsymbol{\Delta}\mathbf{y} - \mathcal{M}\boldsymbol{\Theta}_0(\mathbf{z})\right], \tag{6.1}$$

where $\mathcal{M} = [\mathcal{M}_1', \ldots, \mathcal{M}_n']'$ is an $[n(T-1)] \times (2d_x + d_g)$ data matrix with $\mathcal{M}_i = [\mathcal{M}_{i2}, \ldots, \mathcal{M}_{iT}]'$ and $\mathcal{M}_{it} = \left[\mathbf{g}_i', \boldsymbol{\xi}' \boldsymbol{x}_{it}'\right]'$, $\boldsymbol{x}_{it} = \operatorname{diag}\{\mathbf{x}_{it}, \mathbf{x}_{i,t-1}\}$, and $\boldsymbol{\Theta}_0(\mathbf{z}) = \left[\dot{\boldsymbol{\theta}}_0(\mathbf{z})', \boldsymbol{\beta}_0(\mathbf{z})'\right]'$ is of dimension $(2d_x + d_g)$. The solution to (6.1) yields $\widehat{\boldsymbol{\Theta}}_0(\mathbf{z}) = \left[\mathcal{M}' \mathbf{K}_{\widetilde{h}}(\mathbf{z}) \mathcal{M}\right]^{-1} \mathcal{M}' \mathbf{K}_{\widetilde{h}}(\mathbf{z}) \boldsymbol{\Delta}\mathbf{y}$, which is the first-stage estimator of $\boldsymbol{\Theta}(\mathbf{z})$. The second-stage estimator $\widetilde{\boldsymbol{\beta}}_0(z)$ is obtained from $\Delta \widehat{y}_{it}^{\dagger} \approx \mathbf{x}_{it}' \boldsymbol{\beta}(z_{it}) + \Delta u_{it}$ via local constant estimator, where $\Delta \widehat{y}_{it}^{\dagger} = \Delta y_{it} - \mathbf{g}_i' \widehat{\dot{\boldsymbol{\theta}}}_0(z_{it}) + \mathbf{x}_{i,t-1}' \widehat{\boldsymbol{\beta}}_0(z_{i,t-1})$. Similar to the proof provided in Section 3, we can show that, under some regularity conditions, $\widetilde{\boldsymbol{\beta}}_0(z) - \boldsymbol{\beta}(z) = O_p\left(\widetilde{h}_0^2 + \left(n\widetilde{h}_0\right)^{-1/2}\right)$.

To simplify the test statistic, we replace $\widetilde{\varepsilon}_{it}(z)$ with $\widehat{\varepsilon}_{it} = \Delta \widehat{y}_{it}^{\dagger} - \mathbf{x}_{it}' \widetilde{\boldsymbol{\beta}}_0(z_{it})$, where $\Delta \widehat{y}_{it}^{\dagger} = \Delta y_{it} - \mathbf{g}_i' \widehat{\dot{\boldsymbol{\theta}}}_1(z_{it}) + \mathbf{x}_{i,t-1}' \widehat{\boldsymbol{\beta}}_1(z_{i,t-1})$ is calculated under the alternative hypothesis, and replace $\widehat{\boldsymbol{\Xi}}_{h_0}(z)$ with $\mathbf{K}_{h_0}(z)$ since $\widehat{\boldsymbol{\Xi}}_{h_0}(z)$ essentially serves as a local weight. In addition, since $\mathbf{M}_2(z) \mathcal{X}_1$ gives the estimated "residuals" from regressing $\mathcal{X}_1$ on $\mathcal{X}_2$ and because $\mathcal{X}_1$ is a strictly exogenous variable by Assumption **1**, it is reasonable to replace $\mathbf{M}_2(z) \mathcal{X}_1$ with $\mathcal{X}_1$. This replacement significantly simplifies our proof under H_0. Then, removing the center $i = j$ terms in $\mathcal{T}_n$, we obtain our modified test statistic, i.e.,

$$T_n = \frac{1}{n^2 h_0} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \widehat{\boldsymbol{\varepsilon}}_i' \int K_i(z) \mathcal{X}_{1,i} \mathcal{X}_{1,j}' K_j(z) \, dz \, \widehat{\boldsymbol{\varepsilon}}_j, \tag{6.2}$$

where $\widehat{\boldsymbol{\varepsilon}}_i = [\widehat{\varepsilon}_{i2}, \ldots, \widehat{\varepsilon}_{iT}]'$.

Applying simple algebra, we obtain

$$h_0^{-1} \int K_i(z) \mathcal{X}_{1,i} \mathcal{X}_{1,j}' K_j(z) \, dz = \begin{bmatrix} \bar{k}_{h_0}(z_{i2}, z_{j2}) \mathbf{x}_{i2}' \mathbf{x}_{j2} & \ldots & \bar{k}_{h_0}(z_{i2}, z_{jT}) \mathbf{x}_{i2}' \mathbf{x}_{jT} \\ \vdots & \ddots & \vdots \\ \bar{k}_{h_0}(z_{iT}, z_{j2}) \mathbf{x}_{iT}' \mathbf{x}_{j2} & \ldots & \bar{k}_{h_0}(z_{iT}, z_{jT}) \mathbf{x}_{iT}' \mathbf{x}_{jT} \end{bmatrix}, \tag{6.3}$$

where $\bar{k}_{h_0}(z_{it}, z_{js}) = h_0^{-1} \int k_{it}(z) k_{js}(z) \, dz = \int k((z_{it} - z_{js})/h_0 + \omega) \, d\omega$ effectively selects (i, t) and (j, s) such that $|z_{it} - z_{js}| = o_p(1)$. Therefore, without the loss of essence, we propose our final test statistic:

$$\widehat{T}_n = \frac{1}{n^2 h_0} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \widehat{\boldsymbol{\varepsilon}}_i' A_{ij} \widehat{\boldsymbol{\varepsilon}}_j, \tag{6.4}$$

where A_{ij} equals (6.3) with $\bar{k}_{h_0}(z_{it}, z_{js})$ replaced with $k_{h_0}(z_{it}, z_{js}) = k((z_{it} - z_{js})/h_0)$.

Let $\widetilde{h}$ and h be the bandwidths for the first-stage estimation and $\widetilde{h}_0$ and h_0 be the the bandwidths for the second-stage estimation under H_0 and H_1, respectively. To obtain consistent test statistic, we need the following restrictions on these bandwidths.

Assumption 11 *(i) As $n \to \infty$, $\widetilde{h}$, h, $\widetilde{h}_0$ and h_0 all converge to zero; (ii) $n\widetilde{h}^2/\ln n \to \infty$, $n\widetilde{h}_0 \to \infty$, $\widetilde{h}/\widetilde{h}_0 \to 0$; (iii) $nh^4\sqrt{h_0} \to 0$, $\sqrt{h_0}/h^2 \to 0$; (iv) $n\sqrt{h_0}\widetilde{h}_0^4 \to 0$, $\sqrt{h_0}/\widetilde{h}_0 \to 0$, $\sqrt{nh_0/\widetilde{h}_0}h^2 \to 0$, $\sqrt{nh_0}\widetilde{h}_0^2/h \to 0$, $\sqrt{h_0/\widetilde{h}_0}/h \to 0$; (v) $nh_0 \to \infty$.*

Assumption **11(i)** is a regularity condition. Assumption **11(ii)** ensures that the second-stage estimator of $\boldsymbol{\beta}(z)$ has an asymptotic bias term of order $O_p\left(\widetilde{h}_0^2\right)$ and an asymptotic variance term of order $O_p\left(\left(n\widetilde{h}_0\right)^{-1}\right)$ under H_0. Assumption **11(iii)** removes the asymptotic impact of the first-stage estimator (calculated under the alternative hypothesis) on the test statistic under both hypotheses. Assumption **11(iv)** makes asymptotically negligible the impact of the second-stage estimator (calculated under the null hypothesis) on the test statistic under H_0, while our test statistic explodes under Assumption **11(v).** The following theorem gives the limit result for the proposed test statistic $\widehat{T}_n$.

Theorem 4 *Under Assumptions **1–3** and **11**, we have under H_0 that*

$$J_n \equiv n\sqrt{h_0}\widehat{T}_n/\widehat{\sigma}_0 \overset{d}{\to} \mathbb{N}(0,1),$$

where $\widehat{\sigma}_0^2 = \left(n^2h_0\right)^{-1}\sum_{i=1}^n\sum_{j\neq i}^n(\widehat{\boldsymbol{\varepsilon}}_i'A_{ij}\widehat{\boldsymbol{\varepsilon}}_j)^2$ is a consistent estimator of

$$\sigma_0^2 = 2\sigma_u^4\upsilon_{2,0}(k)\sum_{t=2}^T\sum\nolimits_{0\leq|s-t|\leq1}c_{ts}\mathbb{E}\left[\mu(z_{2s})\,vec\left(\mathbf{x}_{2s}\mathbf{x}_{2s}'\right)f(z_{2s})\right] > 0,$$

$\mu(z) = \mathbb{E}\left[\mathbf{x}_{1t}'\otimes\mathbf{x}_{1t}'|z_{1t}=z\right]$, $c_{ts} = 4$ for $t = s$ and 1 otherwise. Under H_1, $\Pr[J_n > C_n] \to 1$ as $n \to \infty$, where C_n is some non-stochastic sequence such that $C_n = o\left(n\sqrt{h_0}\right)$.

Theorem **4** states that the test statistic $J_n = n\sqrt{h_0}\widehat{T}_n/\widehat{\sigma}_0$ provides a consistent test for testing H_0 against H_1. It is a one-sided test since, as we show in the Appendix, both $\widehat{T}_n$ and $\widehat{\sigma}_0^2$ converge to positive constants under H_1. If J_n is greater than the critical value from the standard normal distribution, we reject the null hypothesis at the corresponding significance level. If we set $h \sim n^{-r}$, $h_0 \sim n^{-\alpha}$, $\widetilde{h} \sim n^{-\widetilde{r}}$ and $\widetilde{h}_0 \sim n^{-\widetilde{\alpha}}$, and $\widetilde{\alpha} = 1/5$, Assumption **11** implies the following: $\widetilde{\alpha} < \widetilde{r} < 1/2$, $2/3 < \alpha < 1$ and $(1-\alpha/2)/4 < r < \alpha/4$.

Note that the test loses its power if $\widetilde{\boldsymbol{\beta}}_0(z)$ converges to $\boldsymbol{\beta}_0(z)$ in probability under the alternative hypothesis and the elements of the spatial weighting matrix are at most of order $O\left(\left(n\sqrt{h_0}\right)^{-1/2}\right)$. The estimator $\widetilde{\boldsymbol{\beta}}_0(z)$ is also consistent under the alternative hypothesis if the spatial lag term becomes less endogenous as the sample size increases.

7 Monte Carlo Simulations

In this section, we evaluate the finite-sample performance of our proposed estimators and the test statistic via Monte Carlo simulations. Section 7.1 reports the results for our proposed estimators, whereas the performance of our proposed test statistic is reported in Section 7.2.

7.1 Estimators

We study the finite sample performance of our proposed estimators in a small set of Monte Carlo experiments. Specifically, we consider the following data generating process with one time-invariant

Table 1. Simulation Results for the Estimator Using Linear and Quadratic Moments

	Working Independence	$T=2$ $n=98$	$T=2$ $n=147$	$T=2$ $n=196$	$T=3$ $n=98$	$T=3$ $n=147$	$T=3$ $n=196$
			Estimated MAE				
$\rho(z_{it})$							
1st Stage		0.4988	0.4661	0.4372	0.4227	0.3958	0.3638
2nd Stage	Yes	0.2971	0.2597	0.2259	0.2147	0.1833	0.1583
2nd Stage	No				0.2474	0.2127	0.1835
$\beta(z_{it})$							
1st Stage		0.2582	0.2230	0.2005	0.1935	0.1685	0.1514
2nd Stage	Yes	0.1560	0.1251	0.1114	0.1056	0.0874	0.0758
2nd Stage	No				0.1475	0.1295	0.1218
$\theta(z_{it})$							
3rd Stage	Yes	0.2115	0.1605	0.1363	0.1326	0.1031	0.0887
3rd Stage	No				0.1461	0.1094	0.0998
			Estimated RMSE				
$\rho(z_{it})$							
1st Stage		0.5973	0.5644	0.5351	0.5197	0.4904	0.4553
2nd Stage	Yes	0.3743	0.3284	0.2879	0.2728	0.2352	0.2059
2nd Stage	No				0.3194	0.2775	0.2418
$\beta(z_{it})$							
1st Stage		0.3477	0.2964	0.2645	0.2548	0.2192	0.1968
2nd Stage	Yes	0.2103	0.1695	0.1527	0.1425	0.1200	0.1054
2nd Stage	No				0.2052	0.1804	0.1693
$\theta(z_{it})$							
3rd Stage	Yes	0.2336	0.1773	0.1503	0.1465	0.1141	0.0976
3rd Stage	No				0.1608	0.1216	0.1094

regressor (i.e., $d_g = 1$):

$$y_{it} = \rho(z_{it}) \sum_{j \neq i} w_{ij} y_{jt} + g_i \theta(z_{it}) + x_{it}\beta(z_{it}) + \mu_i + u_{it}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T. \tag{7.1}$$

As in Lee (2007) and Liu, Lee & Bollinger (2010), rather than generating $\{w_{ij}\}$, we instead use the spatial weighting matrix from the crime study for 49 districts in Columbus, OH from Anselin (1988). The spatial weighting matrix is contiguity-based and uses the (first-order) queen definition for Columbus and corresponds to a sample of $n = 49$. To increase the sample size, we generate a block-diagonal spatial matrix with the original 49×49 Columbus matrix used as a diagonal block.

The data are generated as follows: $z_{it} \sim i.i.d.\ \mathbb{U}(0, \pi/2)$, $g_i \sim i.i.d.\ \mathbb{N}(0, 1)$, $u_{it} \sim i.i.d.\ \mathbb{N}(0, 0.5)$, and $x_{it} = 0.5x_{i,t-1} + \zeta_{it}$ with $\zeta_{it} \sim i.i.d.\ \mathbb{N}(0, 1)$. The fixed effects are constructed as $\mu_i = (\overline{z}_i + \overline{x}_i + g_i)/3 + \omega_i$, where $\omega_i \sim i.i.d.\ \mathbb{N}(0, 0.5)$. The coefficient functions are specified as follows: $\rho(z_{it}) = 0.5\exp(-2z_{it})$, $\theta(z_{it}) = 0.5\sin(\pi z_{it}/3)$, and $\beta(z_{it}) = (z_{it} - 0.75)^2 - 1$.[9]

We fix the length of panel at $T = 2, 3$ for each of which we let the sample size grow with the number of cross-sections $n = 98, 147, 196$. In the second-stage estimation, we use the rule-of-thumb bandwidth for the smoothing variable z_{it}, i.e., $h_0 = 1.06\widehat{\sigma}_z(n(T-1))^{-1/5}$, where $\widehat{\sigma}_z$ is the (pooled) sample standard deviation of z_{it}. We need to undersmooth in the first stage and, by Assumption **7**, the first-stage bandwidth $h \propto n^{-\alpha}$ with $\alpha \in (1/5, 1/4)$ given our choice of the second-stage bandwidth $h_0 \propto n^{-1/5}$. We set $h = n^{-1/21}h_0 \propto n^{-26/105}$ implying that $\alpha \approx 0.248$. Further, we use the following feasible instruments for $\sum_{j \neq i} w_{ij} y_{jt}$ in the first stage: $\mathbf{q}_{it} = \sum_{j \neq i} w_{ij}[x'_{jt}, z_{jt}, g'_j]'$

[9]Note that our choice of the $\theta(z_{it})$ functional coefficient is such that the $\theta(0) = 0$ normalization is satisfied.

Table 2. Simulation Results for the Estimator Using Linear Moments Only

	Working Independence	$T=2$			$T=3$		
		$n=98$	$n=147$	$n=196$	$n=98$	$n=147$	$n=196$
		Estimated MAE					
$\rho(z_{it})$							
1st Stage		0.7196	0.6172	0.5555	0.5205	0.4662	0.4132
2nd Stage	Yes	0.5572	0.4498	0.3606	0.3302	0.2702	0.2217
2nd Stage	No				0.3883	0.3190	0.2592
$\beta(z_{it})$							
1st Stage		0.2880	0.2408	0.2135	0.2043	0.1750	0.1554
2nd Stage	Yes	0.2069	0.1569	0.1350	0.1254	0.2217	0.0819
2nd Stage	No				0.1676	0.1392	0.1272
$\theta(z_{it})$							
3rd Stage	Yes	0.2910	0.2157	0.1746	0.1612	0.1251	0.1012
3rd Stage	No				0.1873	0.1402	0.1157
		Estimated RMSE					
$\rho(z_{it})$							
1st Stage		1.0649	0.8753	0.7985	0.7288	0.6418	0.5677
2nd Stage	Yes	0.8416	0.6699	0.5494	0.4871	0.4103	0.3380
2nd Stage	No				0.6068	0.5125	0.4114
$\beta(z_{it})$							
1st Stage		0.4157	0.3320	0.2930	0.2779	0.2330	0.2053
2nd Stage	Yes	0.3107	0.2304	0.2020	0.1811	0.1410	0.1201
2nd Stage	No				0.2453	0.2009	0.1817
$\theta(z_{it})$							
3rd Stage	Yes	0.3214	0.2370	0.1913	0.1778	0.1381	0.1113
3rd Stage	No				0.2063	0.1544	0.1269

when constructing the matrix of linear instruments $\mathbf{Q}$, and $\mathbf{P}_{n,l} = \mathbf{I}_{T-1} \otimes \left[\mathbf{W}_0^l - n^{-1}\text{tr}\left\{\mathbf{W}_0^l\right\}\mathbf{I}_n\right]$ for $l = 1, 2$ for quadratic moments. For the second-stage estimation, $\widehat{\mathbb{Q}}(z)$ and $\widehat{\mathbf{P}}_n$ are constructed using first-stage estimates of $\rho(z_{it})$ and $\beta(z_{it})$ as described in Section 3.1.

For each sample size, we simulate the model 500 times. For each simulation, we compute the mean absolute error (MAE) and the root mean squared error (RMSE) for each functional coefficient. Reported are their averages computed over 500 simulations. We report the results for both the first- and second-stage nonparametric GMM estimators fitted using two sets of orthogonality conditions: *(i)* linear and quadratic moments (Table 1) and *(ii)* linear moments only (Table 2). Also, for $T = 3$, we estimate the second-stage model twice: *(i)* accounting for "random effects" in Δu_{it} induced by first-differencing as outlined in Section 3.1 and *(ii)* ignoring them by applying the local Taylor approximation directly to (3.8) [as opposed to (3.11)]. In the second columns in both tables, "Yes" corresponds to case (ii) with the "working independence" assumption imposed, while "No" corresponds to case (i).

The two tables indicate that estimation of both $\rho(\cdot)$ and $\beta(\cdot)$ functional coefficients become more stable as the number of cross-sections increases. Both the MAE and RMSE decline significantly as n increases. Regardless of the instrument set, as expected, the second-stage estimator delivers a sizable improvement over its first-stage counterpart. This improvement is however far more significant when pooling the data by ignoring random effects in the second-stage estimation. We also observe that adding quadratic orthogonality conditions leads to a sizable increase in the estimation accuracy. In addition, our results indicate that the performance of our estimator improves significantly when working with longer panels. Comparing our second-stage estimator which

Table 3. Simulations Results for the Spatial Endogeneity Test

Signif. Level	$n = 100$	$n = 400$	$n = 600$	$n = 1,000$	$n = 1,00$	$n = 400$	$n = 600$	$n = 1,000$
	Estimated Size				Estimated Power			
$\alpha = 0.70;\ r = 0.168$								
1%	0.020	0.020	0.043	0.040	0.247	0.304	0.352	0.358
5%	0.071	0.070	0.089	0.084	0.378	0.408	0.482	0.532
10%	0.105	0.114	0.128	0.123	0.442	0.498	0.576	0.626
20%	0.181	0.213	0.213	0.211	0.542	0.601	0.666	0.704
$\alpha = 0.80;\ r = 0.175$								
1%	0.032	0.040	0.035	0.035	0.196	0.202	0.254	0.256
5%	0.080	0.098	0.107	0.087	0.322	0.343	0.402	0.400
10%	0.121	0.155	0.158	0.141	0.384	0.438	0.478	0.502
20%	0.207	0.247	0.234	0.226	0.504	0.524	0.574	0.604
$\alpha = 0.90;\ r = 0.181$								
1%	0.034	0.051	0.043	0.034	0.147	0.152	0.160	0.166
5%	0.098	0.117	0.107	0.096	0.274	0.272	0.282	0.294
10%	0.143	0.181	0.138	0.158	0.347	0.348	0.396	0.398
20%	0.213	0.258	0.235	0.258	0.479	0.470	0.502	0.506

accounts for random effects with its pooled local linear alternative, we observe that the estimation of $\beta(\cdot)$ benefits significantly from the "working independence" assumption, at least in our current data generating design.

Tables 1 and 2 also report the results for the third-stage sieve estimator of $\theta(\cdot)$. We use cubic B-splines to approximate the unknown functional coefficient. For simplicity, we set $L_n = 3$ in our experiments for all three different n's since the range of the sample size is not that large. Consistent with our theory, the sieve estimator of $\theta(\cdot)$ becomes more stable as the sample size grows.

Overall, simulation experiments lend support to asymptotic results for our proposed estimators.

7.2 Spatial Endogeneity Test

We next examine the small sample performance of our proposed residual-based test statistic for spatial endogeneity. To assess the size and power of our test statistic J_n, we consider the following two experimental designs for the data-generating process given in (7.1), where z_{it}, x_{it}, g_i, u_{it}, μ_i and the functional coefficients $\theta(\cdot)$ and $\beta(\cdot)$ are generated exactly as in Section 7.1. Following Kelejian & Prucha (1999) and Jin & Lee (2015), we choose a circular "1 ahead and 1 behind" structure of $\mathbf{W}_0 = \{w_{ij}\}_{i,j=1}^{n}$, where a given spatial unit is related to one neighbor immediately ahead and one neighbor immediately behind it in a row. Each of these two neighbors are assigned an equal non-zero weight of 0.5. We then specify the spatial lag functional parameter $\rho(\cdot)$ as follows:

(i) No spatial dependence: $\rho(z_{it}) = 0$ for all z_{it};

(ii) Spatial autoregression: $\rho(z_{it}) = 0.5 + 0.4\exp(-2z_{it})$.

We consider samples with $n = \{100,\ 400,\ 600,\ 1,000\}$ and $T = 3$. For each n, we simulate the model 500 times. We set $\widetilde{h}_0 = 1.06\widehat{\sigma}_z(n(T-1))^{-0.20}$, $\widetilde{h} = 1.06\widehat{\sigma}_z(n(T-1))^{-0.45}$, $h_0 = 1.06\widehat{\sigma}_z(n(T-1))^{-\alpha}$ and $h = 1.06\widehat{\sigma}_z(n(T-1))^{-r}$ with $2/3 < \alpha < 1$ and $(1-\alpha/2)/4 < r < \alpha/4$. To assess the sensitive of the results to the choice of bandwidths, we try different combinations of h and h_0. Specifically, in line with our assumptions, we set $(\alpha, r) = \{(0.70, 0.168), (0.80, 0.175), (0.90, 0.181)\}$.

Table 3 reports the estimated size under design (i) and power under design (ii) of our test statistic which are computed as rejection frequencies out of 500 simulations. Here, we use (asymptotic) standard normal critical values. We find that out test statistic J_n exhibits good power which increases with the sample size as anticipated, regardless of the choice of bandwidths. Also, the power is significantly better when we under-smooth in both stages under H_1. The size of the test also seems to be sensitive to the degree of smoothing under the alternative with the better results reported for the case when $\alpha = 0.70$ and $r = 0.168$, which imply stronger under-smoothing under H_1. Overall, the estimated size tends to be greater than the nominal level, which is quite expected given that we use asymptotic critical values and kernel-based nonparametric tests are known to be prone to finite-sample size distortions. In empirical applications, we certainly recommend using the bootstrap method.

8 Conclusion

This paper proposes an innovative way of estimating a functional-coefficient spatial autoregressive panel data model with unobserved individual fixed effects which can accommodate (multiple) time-invariant regressors in the model. The methodology we propose removes unobserved fixed effects from the model via the first-difference transformation. The estimation of the transformed nonparametric additive model however does not require the use of backfitting or marginal integration techniques. We derive the consistency and asymptotic normality results for the proposed kernel and sieve estimators. We also construct a consistent nonparametric test statistic to test for spatial endogeneity in the data. A small Monte Carlo study shows that both our proposed estimators and the test statistic exhibit good finite-sample performance.

Appendix. Brief Mathematical Proofs

Proof of Theorem 1. Denote $\boldsymbol{\vartheta}_n = \zeta_n \left[\begin{array}{cc} \widehat{\dot{\boldsymbol{\theta}}}(\mathbf{z})' - \dot{\boldsymbol{\theta}}(\mathbf{z})' & \widehat{\boldsymbol{\gamma}}(\mathbf{z})' - \boldsymbol{\gamma}(\mathbf{z})' \end{array} \right]'$, $\triangle y_{it}^* = \triangle y_{it} - \mathbf{g}_i' \dot{\boldsymbol{\theta}}(\mathbf{z}) - \boldsymbol{\xi}' \mathbf{m}_{it}' \boldsymbol{\gamma}(\mathbf{z})$ and $\varepsilon_{it}(\boldsymbol{\vartheta}) = \triangle y_{it}^* - \zeta_n^{-1} \left[\mathbf{g}_i', \boldsymbol{\xi}' \mathbf{m}_{it}' \right] \boldsymbol{\vartheta}$, where $\{\zeta_n\}$ is a sequence of positive constants such that $0 < C_1 < \|\boldsymbol{\vartheta}_n\| < C_2 < \infty$ for all n and T. We then rewrite (3.6) as

$$\boldsymbol{g}_n(\boldsymbol{\vartheta}) = \left[\begin{array}{c} \boldsymbol{\varepsilon}(\boldsymbol{\vartheta})' \mathbf{K}_h(\mathbf{z}) \mathbf{P}_{n,1} \mathbf{K}_h(\mathbf{z}) \boldsymbol{\varepsilon}(\boldsymbol{\vartheta}) \\ \vdots \\ \boldsymbol{\varepsilon}(\boldsymbol{\vartheta})' \mathbf{K}_h(\mathbf{z}) \mathbf{P}_{n,L} \mathbf{K}_h(\mathbf{z}) \boldsymbol{\varepsilon}(\boldsymbol{\vartheta}) \\ \mathbf{Q}' \mathbf{K}_h(\mathbf{z}) \boldsymbol{\varepsilon}(\boldsymbol{\vartheta}) \end{array} \right], \tag{A.1}$$

where $\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})$ is an $[n(T-1)] \times 1$ vector stacking up $\{\varepsilon_{it}(\boldsymbol{\vartheta})\}$ in the ascending order of index t first then index i, and obtain the following:

$$\frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}'} = -\zeta_n^{-1} \left[\begin{array}{c} \boldsymbol{\varepsilon}(\boldsymbol{\vartheta})' \mathbf{K}_h(\mathbf{z}) \mathbf{P}_{n,1}^s \mathbf{K}_h(\mathbf{z}) \mathbf{M} \\ \vdots \\ \boldsymbol{\varepsilon}(\boldsymbol{\vartheta})' \mathbf{K}_h(\mathbf{z}) \mathbf{P}_{n,L}^s \mathbf{K}_h(\mathbf{z}) \mathbf{M} \\ \mathbf{Q}' \mathbf{K}_h(\mathbf{z}) \mathbf{M} \end{array} \right].$$

Minimizing the objective function in (3.7) is equivalent to minimizing $\Lambda_n(\boldsymbol{\vartheta}) = \boldsymbol{g}_n(\boldsymbol{\vartheta})' \boldsymbol{g}_n(\boldsymbol{\vartheta})$

in $\boldsymbol{\vartheta} \in \mathbb{S}$, a compact subset of $R^{2(d_x+1)+d_g}$. Since $\boldsymbol{\vartheta}_n$ minimizes $\Lambda_n(\boldsymbol{\vartheta}) = \boldsymbol{g}_n(\boldsymbol{\vartheta})' \boldsymbol{g}_n(\boldsymbol{\vartheta})$, we have

$$\mathbf{0}_{2(d_x+1)+d_g} = \frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \boldsymbol{\vartheta}} \boldsymbol{g}_n(\boldsymbol{\vartheta}_n) = \frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \boldsymbol{\vartheta}} \left[\boldsymbol{g}_n(\mathbf{0}) + \frac{\partial \boldsymbol{g}_n\left(\widetilde{\boldsymbol{\vartheta}}_n\right)}{\partial \boldsymbol{\vartheta}'} \boldsymbol{\vartheta}_n \right],$$

where $\widetilde{\boldsymbol{\vartheta}}_n$ lies between $\boldsymbol{\vartheta}_n$ and $\mathbf{0}_{2(d_x+1)+d_g}$, and hence:

$$\boldsymbol{\vartheta}_n = -\left[\frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \boldsymbol{\vartheta}} \frac{\partial \boldsymbol{g}_n\left(\widetilde{\boldsymbol{\vartheta}}_n\right)}{\partial \boldsymbol{\vartheta}'} \right]^{-1} \frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \boldsymbol{\vartheta}} \boldsymbol{g}_n(\mathbf{0}).$$

Specifically, denoting $\boldsymbol{\Xi}_h(\mathbf{z}) \equiv \mathbf{K}_h(\mathbf{z})\mathbf{Q}\mathbf{Q}'\mathbf{K}_h(\mathbf{z})$, we have

$$\begin{aligned} A_n(\mathbf{z}) &= -\frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \boldsymbol{\vartheta}} \boldsymbol{g}_n(\mathbf{0}) \\ &= \frac{1}{2\zeta_n} \sum_{l=1}^{L} \mathbf{M}'\mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}^s \mathbf{K}_h(\mathbf{z}) \boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_n) \Delta\mathbf{y}^{*\prime} \mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}^s \mathbf{K}_h(\mathbf{z}) \Delta\mathbf{y}^* + \frac{1}{\zeta_n} \mathbf{M}'\boldsymbol{\Xi}_h(\mathbf{z})\Delta\mathbf{y}^* \end{aligned}$$

and

$$\begin{aligned} B_n(\mathbf{z}) &= \frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \boldsymbol{\vartheta}} \frac{\partial \boldsymbol{g}_n\left(\widetilde{\boldsymbol{\vartheta}}_n\right)}{\partial \boldsymbol{\vartheta}'} \\ &= \frac{1}{\zeta_n^2} \sum_{l=1}^{L} \left\{ \boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_n)' \mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}^s \mathbf{K}_h(\mathbf{z})\mathbf{M} \right\}' \boldsymbol{\varepsilon}\left(\widetilde{\boldsymbol{\vartheta}}_n\right)' \mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}^s \mathbf{K}_h(\mathbf{z})\mathbf{M} + \frac{1}{\zeta_n^2} \mathbf{M}'\boldsymbol{\Xi}_h(\mathbf{z})\mathbf{M}. \end{aligned}$$

For each (i,t), we have $\triangle y_{it}^* = \triangle u_{it} + c_{it}(\mathbf{z})$, where $c_{it}(\mathbf{z}) = \mathbf{g}_i' [\boldsymbol{\theta}(z_{it}) - \boldsymbol{\theta}(z_1)] - \mathbf{g}_i' [\boldsymbol{\theta}(z_{i,t-1}) - \boldsymbol{\theta}(z_2)]$ $+\mathbf{m}_{it}' [\boldsymbol{\gamma}(z_{it}) - \boldsymbol{\gamma}(z_1)] - \mathbf{m}_{i,t-1}' [\boldsymbol{\gamma}(z_{i,t-1}) - \boldsymbol{\gamma}(z_2)]$. Stacking up $\{c_{it}(\mathbf{z})\}$ in the ascending order of index t first then index i gives an $[n(T-1)] \times 1$ vector $\mathbf{C}(\mathbf{z})$. We also denote $\Gamma_{1,l} = \triangle\mathbf{u}'\mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}^s\mathbf{K}_h(\mathbf{z})\mathbf{C}(z)$, $\Gamma_{2,l} = \mathbf{C}(z)' \mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}^s\mathbf{K}_h(\mathbf{z})\mathbf{C}(z)$, $\Gamma_{3,l} = \triangle\mathbf{u}'\mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}^s\mathbf{K}_h(\mathbf{z})\Delta\mathbf{u}$, $\Psi_{1,l} = \triangle\mathbf{u}'\mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}^s\mathbf{K}_h(\mathbf{z})\mathbf{M}$, $\Psi_{2,l} = \mathbf{C}(z)' \mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}^s\mathbf{K}_h(\mathbf{z})\mathbf{M}$ and $\Psi_{3,l} = \mathbf{M}'\mathbf{K}_h(\mathbf{z})\mathbf{P}_{n,l}^s\mathbf{K}_h(\mathbf{z})\mathbf{M}$ for $l = 1, \ldots, L$. Then, we have $A_n(\mathbf{z}) = A_{n1}(\mathbf{z}) + A_{n2}(\mathbf{z}) - A_{n3}(\mathbf{z})$ with

$$\begin{aligned} A_{n1}(\mathbf{z}) &= \frac{1}{2\zeta_n} \sum_{l=1}^{L} (2\Gamma_{1,l} + \Gamma_{2,l} + \Gamma_{3,l})(\Psi_{1,l} + \Psi_{2,l})' + \frac{1}{\zeta_n} \mathbf{M}'\boldsymbol{\Xi}_h(z)\mathbf{C}(\mathbf{z}), \\ A_{n2}(\mathbf{z}) &= \frac{1}{\zeta_n} \mathbf{M}'\boldsymbol{\Xi}_h(\mathbf{z})\Delta\mathbf{u}, \\ A_{n3}(\mathbf{z}) &= \frac{1}{2\zeta_n^2} \sum_{l=1}^{L} (2\Gamma_{1,l} + \Gamma_{2,l} + \Gamma_{3,l}) \Psi_{3,l} \boldsymbol{\vartheta}_n \end{aligned}$$

and

$$B_n(\mathbf{z}) = \frac{1}{\zeta_n^2} \sum_{l=1}^{L} \left(\Psi_{1,l} + \Psi_{2,l} - \frac{1}{\zeta_n} \boldsymbol{\vartheta}_n' \Psi_{3,l} \right)' \left(\Psi_{1,l} + \Psi_{2,l} - \frac{1}{\zeta_n} \widetilde{\boldsymbol{\theta}}_n' \Psi_{3,l} \right) + \frac{1}{\zeta_n^2} \mathbf{M}'\boldsymbol{\Xi}_h(\mathbf{z})\mathbf{M}.$$

By Lemmas **1–3** below, we have

$$\left(nh^2\right)^{-2} \zeta_n A_{n1}(\mathbf{z}) = \varkappa_A(h, \mathbf{z}) + o_p\left(h^2\right), \tag{A.2}$$

$$\left(nh^2\right)^{-3/2}\zeta_n A_{n2}(\mathbf{z}) \xrightarrow{d} \mathbb{N}\left(\mathbf{0}_{2(d_x+1)+d_g}, \sigma_u^2 \upsilon_{2,0}(k)\,\boldsymbol{\Omega}(\mathbf{z})\right),$$
$$\left(nh^2\right)^{-2}\zeta_n^2 A_{n3}(\mathbf{z}) = O_p\left(h^2 + n^{-1/2}\right),$$
$$\left(nh^2\right)^{-2}\zeta_n^2 B_n(\mathbf{z}) = \varkappa_B(h,\mathbf{z}) + O_p\left(\zeta_n^{-1} + \zeta_n^{-2}\right) + o_p(1),$$

where the last equation holds uniformly over all $\widetilde{\boldsymbol{\theta}}_n \in \mathbb{S}$ and $\varkappa_A(h,\mathbf{z}) = O_p\left(h^2\right)$ and $\varkappa_B(h,\mathbf{z}) = O(1)$. In addition, we have

$$\begin{aligned}
&\zeta_n\left\{\left[\begin{array}{cc}\widehat{\dot{\boldsymbol{\theta}}}(\mathbf{z})' - \dot{\boldsymbol{\theta}}(\mathbf{z})' & \widehat{\boldsymbol{\gamma}}(\mathbf{z})' - \boldsymbol{\gamma}(\mathbf{z})'\end{array}\right]' - \left[\left(nh^2\right)^{-2}\zeta_n^2 B_n(\mathbf{z})\right]^{-1}\left(nh^2\right)^{-2}\zeta_n A_{n1}(\mathbf{z})\right\} \\
= \;& \frac{\zeta_n}{\sqrt{nh^2}}\left[\frac{\zeta_n^2 B_n(\mathbf{z})}{\left(nh^2\right)^2}\right]^{-1}\frac{\zeta_n A_{n2}(\mathbf{z})}{\left(nh^2\right)^{3/2}} + \left[\frac{\zeta_n^2 B_n(\mathbf{z})}{\left(nh^2\right)^2}\right]^{-1}\frac{\zeta_n^2 A_{n3}(\mathbf{z})}{\left(nh^2\right)^2}.
\end{aligned} \tag{A.3}$$

Combining (A.2)–(A.3) with the fact that $0 < C_1 < \|\boldsymbol{\vartheta}_n\| < C_2 < \infty$ for all n, we can deduce that ζ_n must be of order $\sqrt{nh^2}$. The logic is explained below.

(i) If $\zeta_n/\sqrt{nh^2} \to \infty$, it implies that $\sqrt{nh^2}\left[\begin{array}{cc}\widehat{\dot{\boldsymbol{\theta}}}(\mathbf{z})' - \dot{\boldsymbol{\theta}}(\mathbf{z})' & \widehat{\boldsymbol{\gamma}}(\mathbf{z})' - \boldsymbol{\gamma}(\mathbf{z})'\end{array}\right]' \to \mathbf{0}$ as $n \to \infty$. By (A.2)–(A.3) and Assumption **4**, we obtain

$$\begin{aligned}
&\sqrt{nh^2}\left\{\left[\begin{array}{cc}\widehat{\dot{\boldsymbol{\theta}}}(\mathbf{z})' - \dot{\boldsymbol{\theta}}(\mathbf{z})' & \widehat{\boldsymbol{\gamma}}(\mathbf{z})' - \boldsymbol{\gamma}(\mathbf{z})'\end{array}\right]' - \varkappa_B(h,\mathbf{z})^{-1}\varkappa_A(h,\mathbf{z})\right\} \\
&\overset{d}{=} \mathbb{N}\left(\mathbf{0}_{2(d_x+1)+d_g}, \sigma_u^2\upsilon_{2,0}(k)\,\varkappa_B(h,\mathbf{z})^{-1}\boldsymbol{\Omega}(\mathbf{z})\,\varkappa_B(h,\mathbf{z})^{-1}\right) + O_p\left(h^2 + n^{-1/2}\right).
\end{aligned}$$

Since the first term is of order $O_e(1)$, a contradiction occurs.

(ii) Now, suppose that $\zeta_n/\sqrt{nh^2} \to 0$ holds true. By (A.2)–(A.3) and Assumption **4**, we have $\boldsymbol{\vartheta}_n = o_p(1)$ which contradicts the fact that $\|\boldsymbol{\vartheta}_n\|$ is uniformly bounded and positive.

Therefore, applying the exclusion method, we have shown that ζ_n must be of order $\sqrt{nh^2}$ exactly, which gives

$$\begin{aligned}
&\sqrt{nh^2}\left\{\left[\begin{array}{cc}\widehat{\dot{\boldsymbol{\theta}}}(\mathbf{z})' - \dot{\boldsymbol{\theta}}(\mathbf{z})' & \widehat{\boldsymbol{\gamma}}(\mathbf{z})' - \boldsymbol{\gamma}(\mathbf{z})'\end{array}\right]' - \varkappa_B(h,\mathbf{z})^{-1}\varkappa_A(h,\mathbf{z})\right\} \\
&\overset{d}{=} \mathbb{N}\left(\mathbf{0}_{2(d_x+1)+d_g}, \sigma_u^2\upsilon_{2,0}(k)\,\varkappa_B(h,\mathbf{z})^{-1}\boldsymbol{\Omega}(\mathbf{z})\,\varkappa_B(h,\mathbf{z})^{-1}\right).
\end{aligned}$$

This completes the proof of this theorem. ■

In the following three lemmas, we define $\varkappa_j(h,\mathbf{z}) = \varkappa_{j,Q}(\mathbf{z}) + \varkappa_{j,P}(h,\mathbf{z})$ for $j = A$, B, where subscripts Q and P mean that the variable results from the use of $\mathbf{Q}$ and $\{\mathbf{P}_{n,l}\Delta\mathbf{u}\}$ as the instrument, respectively.

Lemma 1 *Under Assumptions **1**–**4**, we obtain* $\left(nh^2\right)^{-2}\zeta_n A_{n1}(\mathbf{z}) = \varkappa_A(h,\mathbf{z}) + o_p\left(h^2\right)$*, where*

$$\begin{aligned}
\varkappa_{A,Q}(\mathbf{z}) &= h^2\upsilon_{1,2}(k)\,\mathbf{E}_1(\mathbf{z})\,\mathbf{E}_2(\mathbf{z}), \\
\varkappa_{A,P}(h,\mathbf{z}) &= \sum_{l=1}^{L}\left[F_{1,l}(h,\mathbf{z}) + \frac{1}{2}F_{2,l}(h,\mathbf{z})\right]\mathbb{E}\left[\psi_l(h,\mathbf{z})\right]' = O_p\left(h^2 + n^{-1/2}\right)
\end{aligned}$$

and $\mathbf{E}_1(\mathbf{z})$*,* $\mathbf{E}_2(\mathbf{z})$*,* $\psi_l(h,\mathbf{z})$ *and* $F_{j,l}$ $(j = 1,2,3)$ *are defined in the proof below.*

Proof. Under Assumptions **1–4** and denoting $\mathbb{E}[\cdot|\mathbf{z}] = \mathbb{E}[\cdot|z_{it} = z_1, z_{i,t-1} = z_2]$, we obtain

$$\frac{1}{nh^2}\mathbf{M}'\mathbf{K}_h(\mathbf{z})\mathbf{Q} = \frac{1}{nh^2}\sum_{i=1}^{n}\sum_{t=2}^{T} k_{it}(\mathbf{z}) \begin{bmatrix} \mathbf{g}_i\mathbf{g}_i' & \mathbf{g}_i\boldsymbol{\xi}'\breve{\boldsymbol{m}}_{it}' \\ \boldsymbol{m}_{it}\boldsymbol{\xi}\mathbf{g}_i' & \boldsymbol{m}_{it}\boldsymbol{\xi}\boldsymbol{\xi}'\breve{\boldsymbol{m}}_{it}' \end{bmatrix} = \mathbf{E}_1(\mathbf{z}) + O_p\left(h^2 + \left(nh^2\right)^{-1/2}\right) \tag{A.4}$$

and

$$\frac{1}{nh^2}\mathbf{Q}'\mathbf{K}_h(\mathbf{z})\mathbf{C}(\mathbf{z}) = \frac{1}{nh^2}\sum_{i=1}^{n}\sum_{t=2}^{T} k_{it}(\mathbf{z}) \begin{bmatrix} \mathbf{g}_i \\ \breve{\boldsymbol{m}}_{it}\boldsymbol{\xi} \end{bmatrix} c_{it}(\mathbf{z}) = h^2\upsilon_{1,2}(k)\,\mathbf{E}_2(\mathbf{z}) + O_p\left(h^4 + n^{-1/2}\right), \tag{A.5}$$

where

$$\begin{aligned}
\mathbf{E}_1(\mathbf{z}) &= \sum_{t=2}^{T} f_{t,t-1}(\mathbf{z}) \begin{bmatrix} \mathbb{E}[\mathbf{g}_i\mathbf{g}_i'|\mathbf{z}] & \mathbb{E}\left[\mathbf{g}_i\boldsymbol{\xi}'\breve{\boldsymbol{m}}_{it}'|\mathbf{z}\right] \\ \mathbb{E}[\boldsymbol{m}_{it}\boldsymbol{\xi}\mathbf{g}_i'|\mathbf{z}] & \mathbb{E}\left[\boldsymbol{m}_{it}\boldsymbol{\xi}\boldsymbol{\xi}'\breve{\boldsymbol{m}}_{it}'|\mathbf{z}\right] \end{bmatrix}, \\
\mathbf{E}_2(\mathbf{z}) &= \sum_{t=2}^{T}\sum_{s=1}^{2}\xi_s\left\{ \begin{bmatrix} \frac{\partial\left(\mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'|\mathbf{z}\right]f_{t,t-1}(\mathbf{z})\right)}{\partial z_s} & \frac{\partial\left(\mathbb{E}\left[\mathbf{g}_i\mathbf{m}_{i,t+1-s}'|\mathbf{z}\right]f_{t,t-1}(\mathbf{z})\right)}{\partial z_s} \\ \frac{\partial\left(\mathbb{E}\left[\breve{\boldsymbol{m}}_{it}\boldsymbol{\xi}\mathbf{g}_i'|\mathbf{z}\right]f_{t,t-1}(\mathbf{z})\right)}{\partial z_s} & \frac{\partial\left(\mathbb{E}\left[\breve{\boldsymbol{m}}_{it}\boldsymbol{\xi}\mathbf{m}_{i,t+1-s}'|\mathbf{z}\right]f_{t,t-1}(\mathbf{z})\right)}{\partial z_s} \end{bmatrix} \begin{bmatrix} \triangledown\boldsymbol{\theta}(z_s) \\ \triangledown\boldsymbol{\gamma}(z_s) \end{bmatrix} + \right. \\
&\left. \frac{f_{t,t-1}(\mathbf{z})}{2} \begin{bmatrix} \mathbb{E}[\mathbf{g}_i\mathbf{g}_i'|\mathbf{z}] & \mathbb{E}\left[\mathbf{g}_i\mathbf{m}_{i,t+1-s}'|\mathbf{z}\right] \\ \mathbb{E}[\breve{\boldsymbol{m}}_{it}\mathbf{g}_i'|\mathbf{z}] & \mathbb{E}\left[\breve{\boldsymbol{m}}_{it}\mathbf{m}_{i,t+1-s}'|\mathbf{z}\right] \end{bmatrix} \begin{bmatrix} \triangledown^2\boldsymbol{\theta}(z_s) \\ \triangledown^2\boldsymbol{\gamma}(z_s) \end{bmatrix} \right\}.
\end{aligned}$$

This gives $\varkappa_{A,Q}(\mathbf{z}) = h^2\upsilon_{1,2}(k)\,\mathbf{E}_1(\mathbf{z})\,\mathbf{E}_2(\mathbf{z})$.

Next, let $a_{ij}(\mathbf{z}_t)$ be the (i,j)th element of the $n \times n$ matrix $\mathbf{W}_0\mathbf{S}_n(\mathbf{z}_t)$ and, by (2.2), we have $\mathbf{W}_0\mathbf{y}_t = \bar{\mathbf{y}}_t + \bar{\mathbf{u}}_t$, where the typical element of $\bar{\mathbf{y}}_t$ and $\bar{\mathbf{u}}_t$ are $\bar{y}_{it} = \sum_{j=1}^{n} a_{ij}(\mathbf{z}_t)\left[\mathbf{g}_j'\boldsymbol{\theta}(z_{jt}) + \mathbf{x}_{jt}'\boldsymbol{\beta}(z_{jt}) + \mu_j\right]$ and $\bar{u}_{it} = \sum_{j=1}^{n} a_{ij}(\mathbf{z}_t)\,u_{jt}$, respectively. Correspondingly, we also denote $\mathbf{m}_{it} = \bar{\mathbf{m}}_{it} + \widetilde{\mathbf{m}}_{it}$ with $\bar{\mathbf{m}}_{it} = [\bar{y}_{it}, \mathbf{x}_{it}']'$ and $\widetilde{\mathbf{m}}_{it} = \left[\bar{u}_{it}, \mathbf{0}_{d_x}'\right]'$, $c_{it}(\mathbf{z}) = \bar{c}_{it}(\mathbf{z}) + \widetilde{c}_{it}(\mathbf{z})$ with $\bar{c}_{it}(\mathbf{z}) = \mathbf{g}_i'[\boldsymbol{\theta}(z_{it}) - \boldsymbol{\theta}(z_1)] - \mathbf{g}_i'[\boldsymbol{\theta}(z_{i,t-1}) - \boldsymbol{\theta}(z_2)] + \bar{\mathbf{m}}_{it}'[\boldsymbol{\gamma}(z_{it}) - \boldsymbol{\gamma}(z_1)] - \bar{\mathbf{m}}_{i,t-1}'[\boldsymbol{\gamma}(z_{i,t-1}) - \boldsymbol{\gamma}(z_2)]$ and $\widetilde{c}_{it}(\mathbf{z}) = \bar{u}_{it}[\rho(z_{it}) - \rho(z_1)]$ $-\bar{u}_{i,t-1}[\rho(z_{i,t-1}) - \rho(z_2)]$, and $\mathbf{M}_{it} = \bar{\mathbf{M}}_{it} + \widetilde{\mathbf{M}}_{it}$ with $\bar{\mathbf{M}}_{it} = \left[\mathbf{g}_i', \boldsymbol{\xi}'\bar{\boldsymbol{m}}_{it}'\right]'$ and $\widetilde{\mathbf{M}}_{it} = \left[\mathbf{0}_{d_g}', \boldsymbol{\xi}'\widetilde{\boldsymbol{m}}_{it}'\right]' = \left[\mathbf{0}_{d_g}', \widetilde{\mathbf{m}}_{it}', -\widetilde{\mathbf{m}}_{i,t-1}\right]'$. Under Assumption **1(iii)**, we have $\|\mathbf{W}_0\mathbf{S}_n(\mathbf{z}_t)\|_j < C < \infty$ for $j = 1, \infty$. Applying straightforward calculations gives

$$\begin{aligned}
\frac{\Psi_{1,l}}{nh^2} &= \frac{1}{nh^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=2}^{T} p_{l,ij}^{s} k_{it}(\mathbf{z})\,k_{jt}(\mathbf{z})\,\Delta u_{it}\mathbf{M}_{jt}' \\
&= \frac{2}{nh^2}\sum_{i=1}^{n}\sum_{t=2}^{T} p_{l,ii} k_{it}^2(\mathbf{z})\,\Delta u_{it}\widetilde{\mathbf{M}}_{it}' + O_p\left(h^2 + \frac{1}{\sqrt{n}h}\right) \\
&\equiv \psi_l(h,\mathbf{z}) + O_p\left(h^2 + \left(\sqrt{n}h\right)^{-1}\right) = \mathbb{E}[\psi_l(h,\mathbf{z})] + o_p(1),
\end{aligned} \tag{A.6}$$

where $p_{l,ij}^{s} = p_{l,ij} + p_{l,ji}$, $\psi_l(h,\mathbf{z}) = \left[\mathbf{0}_{d_g}', \psi_{0,l}(h,\mathbf{z}), \mathbf{0}_{d_x}', -\psi_{1,l}(h,\mathbf{z}), \mathbf{0}_{d_x}'\right]$ with $\psi_{s,l}(h,\mathbf{z}) = 2\left(nh^2\right)^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} p_{l,i}$ for $s = 0, 1$. As

$$\sup_i \mathbb{E}\left[h^{-1}\sum_{t=2}^{T} p_{l,ii}a_{ii}(\mathbf{z}_{t-s})\,k_{it}^2(\mathbf{z})\,\Delta u_{it}u_{i,t-s}\right]^2 \le C < \infty,$$

applying the Chebyshev inequality we can show that $\psi_{s,l}(h,\mathbf{z}) = \mathbb{E}[\psi_{s,l}(h,\mathbf{z})] + o_p(1) = 2\sigma_u^2\upsilon_{2,0}(k)\,n^{-1}\sum_{i=1}^{n} p_{l,ii}$ $\times \sum_{t=2}^{T}(-1)^s\,\mathbb{E}[a_{ii}(\mathbf{z}_{t-s})|\mathbf{z}]\,f_{t,t-1}(\mathbf{z}) + o_p(1) = O_p(1)$.

Similarly, we obtain

$$\frac{\Psi_{2,l}}{nh^2} = \frac{1}{nh^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=2}^{T} p_{l,ij}^{s} k_{it}(\mathbf{z}) k_{jt}(\mathbf{z}) c_{it}(\mathbf{z}) \mathbf{M}'_{jt} = O_p\left(h^2+n^{-1/2}\right), \tag{A.7}$$

$$\begin{aligned}
\frac{\Gamma_{1,l}}{nh^2} &= \frac{1}{nh^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=2}^{T} p_{l,ij}^{s} k_{it}(\mathbf{z}) k_{jt}(\mathbf{z}) \Delta u_{it} c_{jt}(\mathbf{z}) \\
&= \frac{2}{nh^2}\sum_{i=1}^{n}\sum_{t=2}^{T} p_{l,ii} k_{it}^2(\mathbf{z}) \Delta u_{it} \widetilde{c}_{it}(\mathbf{z}) + O_p\left(h^4+\frac{1}{\sqrt{n}}\right) \\
&\equiv F_{1,l}(h,\mathbf{z}) + O_p\left(h^4+\frac{1}{\sqrt{n}}\right) = O_p\left(h^2+\frac{1}{\sqrt{n}}\right), \qquad \text{(A.8)}\\
\frac{\Gamma_{2,l}}{nh^2} &= \frac{1}{nh^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=2}^{T} p_{l,ij}^{s} k_{it}(\mathbf{z}) k_{jt}(\mathbf{z}) c_{it}(\mathbf{z}) c_{jt}(\mathbf{z}) \\
&= \frac{2}{nh^2}\sum_{i=1}^{n}\sum_{t=2}^{T} p_{l,ii} k_{it}^2(\mathbf{z}) c_{it}^2(\mathbf{z}) + O_p\left(h^4\right) \\
&\equiv F_{2,l}(h,\mathbf{z}) + O_p\left(h^4\right) = O_p\left(h^2+hn^{-1/2}\right), \qquad \text{(A.9)}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\Gamma_{3,l}}{nh^2} &= \frac{1}{nh^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=2}^{T} p_{l,ij}^{s} k_{it}(\mathbf{z}) k_{jt}(\mathbf{z}) \Delta u_{it}\Delta u_{jt} \\
&= \frac{2}{nh^2}\sum_{i=1}^{n}\sum_{t=2}^{T} p_{l,ii} k_{it}^2(\mathbf{z}) (\Delta u_{it})^2 + O_p\left(n^{-1/2}\right) = O_p\left(h^2+\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

since $\operatorname{tr}\{P_{n,l}\}=0$ for all l. It then follows that

$$\frac{\Gamma_{3,l}}{nh^2}\frac{\Psi_{2,l}}{nh^2} = O_p\left(h^2+\frac{1}{\sqrt{n}}\right) O_p\left(h^2+n^{-1/2}\right) = O_p\left(h^4+\frac{h^2}{\sqrt{n}}+\frac{1}{n}\right).$$

Also, we have

$$\begin{aligned}
\frac{\Gamma_{3,l}}{nh^2}\psi_{s,l}(h,\mathbf{z}) &\approx \frac{4}{n^2h^4}\sum_{i=1}^{n}\sum_{t=2}^{T} p_{l,ii} k_{it}^2(\mathbf{z}) (\Delta u_{it})^2 \sum_{i=1}^{n}\sum_{t=2}^{T} p_{l,ii} a_{ii}(\mathbf{z}_{t-s}) k_{it}^2(\mathbf{z}) \Delta u_{it} u_{i,t-s} \\
&= O_p\left(\frac{1}{nh^2}+\frac{1}{n}\right) = O_p\left(\frac{1}{nh^2}\right),
\end{aligned}$$

so that we obtain

$$\frac{\Gamma_{3,l}}{nh^2}\frac{\Psi_{1,l}}{nh^2} = O_p\left(\frac{1}{nh^2}+\frac{1}{\sqrt{n}}\right).$$

Combining all the above results gives

$$\begin{aligned}
&\left(nh^2\right)^{-2}\zeta_n A_{n1}(\mathbf{z}) \\
= \; &\frac{1}{2\left(nh^2\right)^2}\sum_{l=1}^{L}\left(2\Gamma_{1,l}+\Gamma_{2,l}\right)\Psi'_{1,l} + \frac{1}{\left(nh^2\right)^2}\mathbf{M}'\boldsymbol{\Xi}_h(z)\mathbf{C}(\mathbf{z}) + O_p\left(h^4+\left(nh^2\right)^{-1}+n^{-1/2}\right)
\end{aligned}$$

$$= \sum_{l=1}^{L}\left[F_{1,l}(h,\mathbf{z})+\frac{1}{2}F_{2,l}(h,\mathbf{z})\right]\psi_l(h,\mathbf{z})'+h^2\upsilon_{1,2}(k)\,\mathbf{E}_1(\mathbf{z})\,\mathbf{E}_2(\mathbf{z})+O_p\left(h^4+\left(nh^2\right)^{-1}+n^{-1/2}\right),$$

which completes the proof of this lemma. ■

Remark 1 If $\boldsymbol{\theta}(z)$ and $\boldsymbol{\gamma}(z)$ are both constant, we have $\Gamma_{1,l}=\Gamma_{2,l}=0$ and $\Psi_{2,l}=\mathbf{0}'_{2(d_x+1)+d_g}$, from which it follows that $\left(nh^2\right)^{-2}\zeta_n A_{n1}(\mathbf{z})=\left(2nh^2\right)^{-2}\sum_{l=1}^{L}\Gamma_{3,l}\Psi'_{1,l}\;=O_p\left(n^{-1/2}\right)$ under Assumption **4**.

Lemma 2 *Under Assumptions **1**–**4**, we obtain*

$$\left(nh^2\right)^{-2}\zeta_n^2 B_n(z)=\varkappa_B(h,\mathbf{z})+O_p\left({\zeta_n}^{-1}+{\zeta_n}^{-2}\right)+o_p(1)\,, \tag{A.10}$$

where $\varkappa_{B,Q}(\mathbf{z})=\mathbf{E}_1(\mathbf{z})\,\mathbf{E}_1(z)'$ *and* $\varkappa_{B,P}(h,\mathbf{z})=\sum_{l=1}^{L}\mathbb{E}\left[\psi_l(h,\mathbf{z})\right]'\mathbb{E}\left[\psi_l(h,\mathbf{z})\right]$.

Proof. By (A.4), we obtain $\varkappa_{B,Q}(\mathbf{z})=\mathbf{E}_1(\mathbf{z})\,\mathbf{E}_1(z)'$. In addition, we have

$$\begin{aligned}\frac{\Psi_{3,l}}{nh^2} &= \frac{1}{nh^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=2}^{T}p^s_{l,ij}k_{it}(\mathbf{z})\,k_{jt}(\mathbf{z})\,\mathbf{M}_{it}\mathbf{M}'_{jt}\\ &= \frac{2}{nh^2}\sum_{i=1}^{n}\sum_{t=2}^{T}p_{l,ii}k_{it}(\mathbf{z})\,\mathbf{M}_{it}\mathbf{M}'_{it}+O_p\left(h^2\right)=O_p(1)\,.\end{aligned} \tag{A.11}$$

Combining this result with (A.6) and (A.7) yields

$$\begin{aligned}&\left(nh^2\right)^{-2}\zeta_n^2 B_n(\mathbf{z})\\ =\;& \frac{1}{\left(nh^2\right)^2}\sum_{l=1}^{L}\Psi'_{1,l}\Psi_{1,l}+\frac{1}{\left(nh^2\right)^2}\mathbf{M}'\boldsymbol{\Xi}_h(\mathbf{z})\mathbf{M}+O_p\left(\zeta_n^{-1}+\zeta_n^{-2}+h^2+n^{-1/2}\right)\\ =\;& \varkappa_{B,P}(h,\mathbf{z})+\varkappa_{B,Q}(\mathbf{z})+O_p\left(\zeta_n^{-1}+\zeta_n^{-2}\right)+O_p\left(h^2+\left(nh^2\right)^{-1/2}\right),\end{aligned}$$

which completes the proof of this lemma. ■

Remark 2 When model (1.2) becomes a pure spatial autoregressive panel data model with random effects, we have

$$\mathbb{E}\left[\boldsymbol{m}_{it}\boldsymbol{\xi}\boldsymbol{\xi}'\check{\boldsymbol{m}}'_{it}|\mathbf{z}\right]=\begin{bmatrix}\mathbf{0}'_{d_q+d_x} & \mathbf{0}'_{d_q+d_x}\\ \mathbb{E}\left[\mathbf{x}_{it}\check{\mathbf{m}}'_{it}|\mathbf{z}\right] & -\mathbb{E}\left[\mathbf{x}_{it}\check{\mathbf{m}}'_{i,t-1}|\mathbf{z}\right]\\ \mathbf{0}'_{d_q+d_x} & \mathbf{0}'_{d_q+d_x}\\ -\mathbb{E}\left[\mathbf{x}_{i,t-1}\check{\mathbf{m}}'_{i,t}|\mathbf{z}\right] & \mathbb{E}\left[\mathbf{x}_{i,t-1}\check{\mathbf{m}}'_{i,t-1}|\mathbf{z}\right]\end{bmatrix}\text{ and }\mathbb{E}\left[\boldsymbol{m}_{it}\boldsymbol{\xi}\mathbf{g}'_i|\mathbf{z}\right]=\begin{bmatrix}\mathbf{0}'_{d_g}\\ \mathbb{E}\left[\mathbf{x}_{it}\mathbf{g}'_i|\mathbf{z}\right]\\ \mathbf{0}'_{d_g}\\ -\mathbb{E}\left[\mathbf{x}_{i,t-1}\mathbf{g}'_i|\mathbf{z}\right]\end{bmatrix}$$

for all t, which implies that $\varkappa_{B,Q}(\mathbf{z})=\mathbf{E}_1(\mathbf{z})\,\mathbf{E}_1(z)'$ becomes a singular matrix. Consequently, applying only the linear moments, (3.4), will not give a consistent estimator due to the singularity of $\varkappa_{B,Q}(\mathbf{z})$. However, $\varkappa_B(h,\mathbf{z})$ still can be a non-singular matrix if $\psi_l(h,\mathbf{z})=O_e(1)$ hold for some l. Therefore, the nonparametric GMM estimator using both linear and quadratic moments is recommended.

Lemma 3 *Under Assumptions **1**–**4**, we obtain*

$$\left(nh^2\right)^{-3/2}\zeta_n A_{n,2}(\mathbf{z})\xrightarrow{d}\mathbb{N}\left(\mathbf{0}_{2(d_x+1)+d_g},\sigma_u^2\upsilon_{2,0}(k)\,\boldsymbol{\Omega}(\mathbf{z})\right), \tag{A.12}$$

where $\boldsymbol{\Omega}(z)=\mathbf{E}_1(\mathbf{z})\,\mathbf{E}_3(\mathbf{z})\,\mathbf{E}_1(\mathbf{z})'$ *and* $\mathbf{E}_3(\mathbf{z})=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}f_{t,t-1}(\mathbf{z})\,\mathbb{E}\left[\mathbf{Q}_{it}\mathbf{Q}'_{it}|\mathbf{z}\right]$.

Proof. By definition, $\left(nh^2\right)^{-3/2}\zeta_n A_{n,2}(\mathbf{z}) = \left[\left(nh^2\right)^{-1}\mathbf{M}'\mathbf{K}_h(\mathbf{z})\mathbf{Q}\right]\left(nh^2\right)^{-1/2}\mathbf{Q}'\mathbf{K}_h(\mathbf{z})\Delta\mathbf{u}$, where $\left(nh^2\right)^{-1}\mathbf{M}'\mathbf{K}_h(\mathbf{z})\mathbf{Q} = \mathbf{E}_1(\mathbf{z}) + O_p\left(h^2 + \left(nh^2\right)^{-1/2}\right)$ by Lemma **1**. Below, we show that

$$\mathbf{\Delta}_n \equiv \frac{1}{\sqrt{nh^2}}\mathbf{Q}'\mathbf{K}_h(\mathbf{z})\Delta\mathbf{u} \ = \frac{1}{\sqrt{nh^2}}\sum_{i=1}^{n}\mathbf{Q}_i'\mathbf{K}_i(\mathbf{z})\Delta\mathbf{u}_i$$

converges to a normal distribution, where $\mathbf{Q}_i = [\mathbf{Q}_{i2},\ldots,\mathbf{Q}_{iT}]'$, $\Delta\mathbf{u}_i = [\Delta u_{i2},\ldots,\Delta u_{iT}]'$ and $\mathbf{K}_i(\mathbf{z}) = \operatorname{diag}\{k_{i2}(\mathbf{z}),\ldots,k_{iT}(\mathbf{z})\}$.

Let $\boldsymbol{\alpha}\neq\mathbf{0}$ be a $[2(d_q+d_x)+d_g]\times 1$ vector and we also denote $\chi_{n,i} = \left(nh^2\right)^{-1/2}\boldsymbol{\alpha}'\mathbf{Q}_i'\mathbf{K}_i(z)\Delta\mathbf{u}_i$. Next, we construct a scalar $\Delta_{n,\alpha} \equiv \boldsymbol{\alpha}'\mathbf{\Delta}_n = \boldsymbol{\alpha}'\sum_{i=1}^n\chi_{n,i}$. Let $\mathcal{F}_{n,i} = \sigma\left(\Delta\mathbf{u}_j, \mathbf{x}_{j'}, \mathbf{z}_{j'}, \mathbf{g}_{j'}, 1\leq j\leq i,\right.$ $\left. j' = 1,\ldots,n\right)$ be the smallest sigma field containing all the information on $\Delta\mathbf{u}_j$ for $1\leq j\leq i$ and $\left(\mathbf{x}_{j'},\mathbf{z}_{j'},\mathbf{g}_{j'}\right)$ for $j' = 1,\ldots,n$. Under Assumption **1**, $\{(\chi_{n,i},\mathcal{F}_{n,i})\}$ forms a martingale difference sequence. In order to apply Hall & Heyde's (1980, Corollary 3.1) martingale difference central limit theorem, we need to verify that, for any small $\varepsilon > 0$:

$$\sum_{i=1}^{n}\mathbb{E}\left[\chi_{n,i}^2\mathbb{I}\left(|\chi_{n,i}|>\varepsilon\right)\middle|\mathcal{F}_{n,i-1}\right] \ \xrightarrow{p}\ 0 \tag{A.13}$$

and

$$\sum_{i=1}^{n}\mathbb{E}\left[\chi_{n,i}^2\middle|\mathcal{F}_{n,i-1}\right] \ \xrightarrow{p}\ \sigma_u^2\upsilon_{2,0}(k)\,\boldsymbol{\alpha}'\mathbf{E}_3(\mathbf{z})\,\boldsymbol{\alpha} > 0, \tag{A.14}$$

where (A.14) holds true because $\sum_{i=1}^n\mathbb{E}\left[\chi_{n,i}^2\middle|\mathcal{F}_{n,i-1}\right] = \left(nh^2\right)^{-1}\sigma_u^2\sum_{i=1}^n\boldsymbol{\alpha}'\mathbf{Q}_i'\mathbf{K}_i(\mathbf{z})\mathbf{\Sigma}\mathbf{K}_i(\mathbf{z})\mathbf{Q}_i\boldsymbol{\alpha}$ with $\mathbf{\Sigma} = 2\mathbf{I}_{T-1} - \mathbf{J}_{T-1}(0) - \mathbf{J}_{T-1}(0)'$ and

$$\begin{aligned}\frac{1}{nh^2}\sum_{i=1}^{n}\mathbf{Q}_i'\mathbf{K}_i(\mathbf{z})\mathbf{\Sigma}\mathbf{K}_i(\mathbf{z})\mathbf{Q}_i &= \frac{1}{nh^2}\sum_{i=1}^{n}\sum_{t=1}^{T}k_{it}^2(\mathbf{z})\,\mathbf{Q}_{it}\mathbf{Q}_{it}' + O_p\left(h^2\right)\\ &= \frac{\upsilon_{2,0}(k)}{nh^2}\sum_{i=1}^{n}\sum_{t=1}^{T}f_{t,t-1}(\mathbf{z})\,\mathbb{E}\left[\mathbf{Q}_{it}\mathbf{Q}_{it}'|\mathbf{z}\right] + O_p\left(h^2 + \left(\sqrt{n}h\right)^{-1/2}\right).\end{aligned}$$

Next, we verify (A.13). Under Assumptions **1**–**6**, we have

$$\begin{aligned}&\sum_{i=1}^{n}\mathbb{E}\left[\chi_{n,i}^2\mathbb{I}\left(|\chi_{n,i}|>\varepsilon\right)\middle|\mathcal{F}_{n,i-1}\right]\\ \leq\ & \varepsilon^{-2}\sum_{i=1}^{n}\mathbb{E}\left[\chi_{n,i}^4|\mathcal{F}_{n,i-1}\right] = \frac{1}{\left(nh^2\right)^2\varepsilon^2}\sum_{i=1}^{n}\mathbb{E}\left[\left(\sum_{t=2}^{T}k_{it}(\mathbf{z})\,\Delta u_{it}\boldsymbol{\alpha}'\mathbf{Q}_{it}\right)^4\middle|\mathcal{F}_{n,i-1}\right]\\ \leq\ & \frac{T^3}{\left(nh^2\right)^2\varepsilon^2}\sum_{i=1}^{n}\sum_{t=2}^{T}\mathbb{E}\left[k_{it}^4(\mathbf{z})\left(\Delta u_{it}\right)^4\left[\boldsymbol{\alpha}'\mathbf{Q}_{it}\right]^4\middle|\mathcal{F}_{n,i-1}\right]\\ \leq\ & \frac{C}{\left(nh^2\right)^2\varepsilon^2}\sum_{i=1}^{n}\sum_{t=1}^{T}k_{it}^4(\mathbf{z})\left[\mathbf{Q}_{it}'\mathbf{Q}_{it}\right]^2\left(\boldsymbol{\alpha}'\boldsymbol{\alpha}\right)^2\\ =\ & O_p\left(\frac{1}{nh^2\varepsilon^2}\right) = o_p(1)\ \text{ as } nh^2\to\infty \text{ when } n\to\infty \text{ for any small } \varepsilon>0,\end{aligned}$$

since $\left(\sum_{t=2}^{T} a_t\right)^4 \leq T^3 \sum_{t=2}^{T} a_t^4$, where $\mathbb{E}\left[\mathbf{Q}_{it}'\mathbf{Q}_{it}\right]^2$ is uniformly bounded over (i,t) under Assumption **1** and $\mathbb{E}\left(\|\mathbf{x}_{it}\|^4\right)$, $\mathbb{E}\left(\|\mathbf{g}_i\|^4\right)$ and $\mathbb{E}\left(|z_{it}|^4\right)$ are all uniformly bounded over (i,t).

Therefore, applying Corollary 3.1 of Hall & Heyde (1980) and the Cramér's Wold device completes the proof of this lemma. ■

Proof of Theorem 2. Define a $(d_x+1)\times 2$ matrix $\boldsymbol{\vartheta}_n = \zeta_n \left[\begin{array}{cc} \widetilde{\boldsymbol{\gamma}}(z) - \boldsymbol{\gamma}(z) & h_0\left[\triangledown\widetilde{\boldsymbol{\gamma}}(z) - \triangledown\boldsymbol{\gamma}(z)\right] \end{array}\right]$ as well as $\widehat{Y}_{it}^* = \widehat{Y}_{it} - \varphi_{tt}\mathbf{m}_{it}'\boldsymbol{\gamma}(z) - \varphi_{tt}\mathbf{m}_{it}'\nabla\boldsymbol{\gamma}(z)(z_{it}-z)$ and $\varepsilon_{it}(\boldsymbol{\vartheta}) = \widehat{Y}_{it}^* - \zeta_n^{-1}\varphi_{tt}\mathbf{m}_{it}'\boldsymbol{\vartheta}\mathcal{Z}_{it}(z)$, where $\{\zeta_n\}$ is a sequence of positive constants such that $0 < C_1 < \|\boldsymbol{\vartheta}_n\| < C_2 < \infty$ for all n. Then, we can rewrite (3.16) as

$$\boldsymbol{g}_n(\boldsymbol{\vartheta}) = \left[\begin{array}{c} \boldsymbol{\varepsilon}(\boldsymbol{\vartheta})'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n\mathbf{K}_{h_0}(z)\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}) \\ \widehat{\mathbb{Q}}(z)'\mathbf{K}_{h_0}(z)\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}) \end{array}\right], \tag{A.15}$$

where $\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})$ is an $[n(T-1)]\times 1$ vector stacking up $\{\varepsilon_{it}(\boldsymbol{\vartheta})\}$ in the ascending order of index t first then index i. In addition, we have

$$\frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta})}{\partial \text{vec}(\boldsymbol{\vartheta})'} = -\zeta_n^{-1}\left[\begin{array}{c} \boldsymbol{\varepsilon}(\boldsymbol{\vartheta})'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z) \\ \widehat{\mathbb{Q}}(z)'\mathbf{K}_{h_0}(z)\mathbb{M}(z) \end{array}\right].$$

Minimizing the objective function in (3.15) is equivalent to minimizing $\Lambda_n(\boldsymbol{\vartheta}) = \mathbf{g}_n(\boldsymbol{\vartheta})'\mathbf{g}_n(\boldsymbol{\vartheta})$ in $\boldsymbol{\vartheta}\in\mathbb{S}$, a compact subset of $R^{(d_x+1)}\times R^2$. Since $\boldsymbol{\vartheta}_n$ minimizes $\Lambda_n(\boldsymbol{\vartheta}) = \boldsymbol{g}_n(\boldsymbol{\vartheta})'\boldsymbol{g}_n(\boldsymbol{\vartheta})$, we have

$$\mathbf{0}_{2(d_x+1)} = \frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \text{vec}(\boldsymbol{\vartheta})}\boldsymbol{g}_n(\boldsymbol{\vartheta}_n) = \frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \text{vec}(\boldsymbol{\vartheta})}\left[\boldsymbol{g}_n(\mathbf{0}) + \frac{\partial \boldsymbol{g}_n\left(\widetilde{\boldsymbol{\vartheta}}_n\right)}{\partial \text{vec}(\boldsymbol{\vartheta})'}\text{vec}(\boldsymbol{\vartheta}_n)\right],$$

where $\widetilde{\boldsymbol{\vartheta}}_n$ lies between $\boldsymbol{\vartheta}_n$ and $\mathbf{0}_{2(d_x+1)}$, so that we have

$$\text{vec}(\boldsymbol{\vartheta}_n) = -\left[\frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \text{vec}(\boldsymbol{\vartheta})}\frac{\partial \boldsymbol{g}_n\left(\widetilde{\boldsymbol{\vartheta}}_n\right)}{\partial \text{vec}(\boldsymbol{\vartheta})'}\right]^{-1}\frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \text{vec}(\boldsymbol{\vartheta})}\boldsymbol{g}_n(\mathbf{0}).$$

Specifically, denoting $\widehat{\boldsymbol{\Xi}}_{h_0}(z) \equiv \mathbf{K}_{h_0}(z)\widehat{\mathbb{Q}}(z)\widehat{\mathbb{Q}}(z)'\mathbf{K}_{h_0}(z)$, we have

$$\begin{aligned} \widehat{A}_n(z) &= -\frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \text{vec}(\boldsymbol{\vartheta})}\boldsymbol{g}_n(\mathbf{0}) \\ &= \frac{1}{2\zeta_n}\mathbb{M}(z)'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_n)\widehat{\mathbf{Y}}^{*\prime}\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\widehat{\mathbf{Y}}^* + \frac{1}{\zeta_n}\mathbb{M}(z)'\widehat{\boldsymbol{\Xi}}_{h_0}(z)\widehat{\mathbf{Y}}^* \end{aligned}$$

and

$$\begin{aligned} \widehat{B}_n(z) &= \frac{\partial \boldsymbol{g}_n(\boldsymbol{\vartheta}_n)'}{\partial \text{vec}(\boldsymbol{\vartheta})}\frac{\partial \boldsymbol{g}_n\left(\widetilde{\boldsymbol{\vartheta}}_n\right)}{\partial \text{vec}(\boldsymbol{\vartheta})'} \\ &= \frac{1}{\zeta_n^2}\left[\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_n)'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)\right]'\boldsymbol{\varepsilon}\left(\widetilde{\boldsymbol{\vartheta}}_n\right)'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z) + \frac{1}{\zeta_n^2}\mathbb{M}(z)'\widehat{\boldsymbol{\Xi}}_{h_0}(z)\mathbb{M}(z). \end{aligned}$$

For each (i,t), we define a $(d_x+1)\times 1$ vector $\mathbf{\Pi}(z_{it}^*)$, whose lth element equals $\Pi_l(\mathbf{z}_{it}^*) = \nabla^2\gamma_l(z_{it}^*)(z_{it}-z)^2$, and z_{it}^* lies between z_{it} and z for $l=1,\ldots,(d_x+1)$. Then, we have $\widehat{Y}_{it}^* = \varphi_{tt}\mathbf{m}_{it}'\mathbf{\Pi}(z_{it}^*)/2+\widehat{v}_{it}$ with $\widehat{v}_{it}=\widehat{\pi}_{it}+\Delta\widetilde{u}_{it}$, $\Delta\widetilde{u}_{it}=\sum_{s=2}^T\varphi_{ts}\Delta u_{is}$ and

$$\widehat{\pi}_{it}=\sum_{s=2}^{T}\varphi_{ts}\left\{\mathbf{g}_i'\left[\dot{\boldsymbol{\theta}}\left(z_{is},z_{i,s-1}\right)-\widehat{\dot{\boldsymbol{\theta}}}\left(z_{is},z_{i,s-1}\right)\right]+\mathbf{m}_{is}'\left[\boldsymbol{\gamma}\left(z_{is}\right)-\widehat{\boldsymbol{\gamma}}\left(z_{is}\right)\right]-\mathbf{m}_{i,s-1}'\left[\boldsymbol{\gamma}\left(z_{i,s-1}\right)-\widehat{\boldsymbol{\gamma}}\left(z_{i,s-1}\right)\right]\right\}. \tag{A.16}$$

Further, we define an $[n(T-1)]\times 1$ vector $\mathbf{C}(z)$ whose typical term equals $\varphi_{tt}\mathbf{m}_{it}'\mathbf{\Pi}(z_{it}^*)/2$. Note that $\max_{i,t}|\widehat{\pi}_{it}|=O_p\left(h^2+\sqrt{\ln n/(nh^2)}\right)$ holds under Assumptions **1** and **5**. Also, let $\widehat{\mathbf{\Pi}}$ be an $[n(T-1)]\times 1$ vector stacking up $\widehat{\pi}_{it}$ in the ascending order of index t first.

Closely following the proof of Theorem **1**, we denote

$$\begin{aligned}
\widehat{A}_{n1}(z) &= \frac{1}{2\zeta_n}\left(2\widehat{\Gamma}_1+\widehat{\Gamma}_2+\widehat{\Gamma}_3\right)\left(\widehat{\Psi}_1+\widehat{\Psi}_2\right)'+\frac{1}{\zeta_n}\mathbb{M}(z)'\,\widehat{\mathbf{\Xi}}_{h_0}(z)\left[\mathbf{C}(z)+\widehat{\mathbf{\Pi}}\right],\\
\widehat{A}_{n2}(z) &= \frac{1}{\zeta_n}\mathbb{M}(z)'\widehat{\mathbf{\Xi}}_{h_0}(z)\Delta\widetilde{\mathbf{u}},\\
\widehat{A}_{n3}(z) &= \frac{1}{2\zeta_n^2}\left(2\widehat{\Gamma}_1+\widehat{\Gamma}_2+\widehat{\Gamma}_3\right)\widehat{\Psi}_3\text{vec}\left(\boldsymbol{\vartheta}_n\right)
\end{aligned}$$

and

$$\widehat{B}_n(z)=\frac{1}{\zeta_n^2}\left(\widehat{\Psi}_1+\widehat{\Psi}_2-\frac{1}{\zeta_n}\left[\text{vec}\left(\boldsymbol{\vartheta}_n\right)\right]'\widehat{\Psi}_3\right)\left(\widehat{\Psi}_1+\widehat{\Psi}_2-\frac{1}{\zeta_n}\left[\text{vec}\left(\widetilde{\boldsymbol{\vartheta}}_n\right)\right]'\widehat{\Psi}_3\right)\frac{1}{\zeta_n^2}\mathbb{M}(z)'\widehat{\mathbf{\Xi}}_{h_0}(z)\mathbb{M}(z),$$

where, denoting $\widehat{\mathbf{v}}=[\widehat{\mathbf{v}}_2',\ldots,\widehat{\mathbf{v}}_T']'$ with $\widehat{\mathbf{v}}_t=[\widehat{v}_{1t},\ldots,\widehat{v}_{nt}]'$, $\widehat{\Gamma}_1=\widehat{\mathbf{v}}'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbf{C}(z)$, $\widehat{\Gamma}_2=\mathbf{C}(z)'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbf{C}(z)$, $\widehat{\Gamma}_3=\widehat{\mathbf{v}}'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\widehat{\mathbf{v}}$, $\widehat{\Psi}_1=\widehat{\mathbf{v}}'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)$, $\widehat{\Psi}_2=\mathbf{C}(z)'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)$ and $\widehat{\Psi}_3=\mathbb{M}(z)'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)$.

By Lemmas **5–7** given below, we have

$$(nh_0)^{-2}\zeta_n\widehat{A}_{n1}(z)=\varkappa_A(h_0,z)+o_p\left(h_0^2\right), \tag{A.17}$$

$$(nh_0)^{-3/2}\zeta_n\widehat{A}_{n2}(z)\xrightarrow{d}\mathbb{N}\left(\mathbf{0}_{2(d_x+1)},\sigma_u^2\begin{bmatrix}\upsilon_{2,0}(k) & 0\\ 0 & \upsilon_{2,2}(k)\,\upsilon_{1,2}^2(k)\end{bmatrix}\otimes\mathbf{\Omega}(z)\right),$$

$$(nh_0)^{-2}\zeta_n^2\widehat{A}_{n3}(z)=O_p\left(h_0^2\right),$$

$$(nh_0)^{-2}\zeta_n^2\widehat{B}_n(z)=\varkappa_B(h_0,z)+O_p\left(\zeta_n^{-1}+\xi_n^{-2}\right)+o_p(1), \tag{A.18}$$

where the last line holds uniformly over all $\widetilde{\boldsymbol{\vartheta}}_n\in\mathbb{S}$. Following exactly the same argument as in the proof of Theorem **1**, we obtain $\zeta_n=\sqrt{nh_0}$, which completes the proof of this theorem. ■

In the following three lemmas, we define $\varkappa_j(h_0,z)=\varkappa_{j,Q}(h_0,z)+\varkappa_{j,P}(h_0,z)$ for $j=A$, B, where subscripts Q and P mean that the variable results from the use of localized linear and quadratic moments, respectively.

Lemma 4 *Under Assumptions **1**–**5**, we obtain* $\left\|\widehat{\mathbf{S}}_n(\mathbf{z}_t)-\mathbf{S}_n(\mathbf{z}_t)\right\|_{sp}=O_p\left(h^2+\sqrt{(\ln n)/(nh^2)}\right)$ *and* $n^{-1}tr\left\{\mathbf{W}_0\left[\widehat{\mathbf{S}}_n(\mathbf{z}_t)-\mathbf{S}_n(\mathbf{z}_t)\right]\right\}=O_p\left(h^2+\sqrt{(\ln n)/(nh^2)}\right)$.

Proof. First, we have

$$\left\|\widehat{\mathbf{S}}_n(\mathbf{z}_t)-\mathbf{S}_n(\mathbf{z}_t)\right\|_{sp} = \left\|\left[\mathbf{I}_n-\widehat{\boldsymbol{\rho}}(\mathbf{z}_t)\mathbf{W}_0\right]^{-1}-\left[\mathbf{I}_n-\boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\right]^{-1}\right\|_{sp}$$

$$\leq \left\| \left[\mathbf{I}_n - \widehat{\boldsymbol{\rho}}(\mathbf{z}_t)\mathbf{W}_0\right]^{-1} \right\|_{sp} \left\| \left[\widehat{\boldsymbol{\rho}}(\mathbf{z}_t) - \boldsymbol{\rho}(\mathbf{z}_t)\right]\mathbf{W}_0 \right\|_{sp} \left\| \left[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\right]^{-1} \right\|_{sp},$$

where $\left\| \left[\widehat{\boldsymbol{\rho}}(\mathbf{z}_t) - \boldsymbol{\rho}(\mathbf{z}_t)\right]\mathbf{W}_0 \right\|_{sp} \leq \left\| \widehat{\boldsymbol{\rho}}(\mathbf{z}_t) - \boldsymbol{\rho}(\mathbf{z}_t) \right\|_{sp} \left\| \mathbf{W}_0 \right\|_{sp} = O_p\left(h^2 + \sqrt{(nh^2)^{-1}\ln n}\right)$ under Assumptions **1** and **5.**

Next, Letting $A = \left[\mathbf{I}_n - \widehat{\boldsymbol{\rho}}(\mathbf{z}_t)\mathbf{W}_0\right]\left[\mathbf{I}_n - \widehat{\boldsymbol{\rho}}(\mathbf{z}_t)\mathbf{W}_0\right]'$ and $B = \left[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\right]\left[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\right]'$, we have $A - B = \left[\boldsymbol{\rho}(\mathbf{z}_t) - \widehat{\boldsymbol{\rho}}(\mathbf{z}_t)\right]\mathbf{W}_0 + \mathbf{W}_0'\left[\boldsymbol{\rho}(\mathbf{z}_t) - \widehat{\boldsymbol{\rho}}(\mathbf{z}_t)\right] + \mathbf{W}_0\left[\widehat{\boldsymbol{\rho}}^2(\mathbf{z}_t) - \boldsymbol{\rho}^2(\mathbf{z}_t)\right]\mathbf{W}_0'$. Applying Weyl's theorem and Property 4.67 (e) in Seber (2008) gives $\left|\lambda_{\min}(A) - \lambda_{\min}(B)\right| \leq \left(\|A-B\|_1 \|A-B\|_\infty\right)^{1/2}$ $= O_p\left(h^2 + \sqrt{(\ln n)/(nh^2)}\right)$ under Assumptions **1** and **5**. In addition, we have $\left\| \left[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\right]^{-1} \right\|_{sp} \leq$ C under Assumption **1(iii).** We therefore obtain

$$\left\| \left[\mathbf{I}_n - \widehat{\boldsymbol{\rho}}(\mathbf{z}_t)\mathbf{W}_0\right]^{-1} \right\|_{sp}^2 = \lambda_{\min}^{-1}\left\{\left[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\right]\left[\mathbf{I}_n - \boldsymbol{\rho}(\mathbf{z}_t)\mathbf{W}_0\right]'\right\} + o_p(1),$$

from which it follows

$$\left\| \widehat{\mathbf{S}}_n(\mathbf{z}_t) - \mathbf{S}_n(\mathbf{z}_t) \right\|_{sp} = O_p\left(h^2 + \sqrt{\frac{\ln n}{nh^2}}\right). \tag{A.19}$$

Now, by (A.19), we obtain

$$\max_{i,j}\left|\widehat{a}_{ij}(\mathbf{z}_t) - a_{ij}(\mathbf{z}_t)\right| \leq \left\|\mathbf{W}_0\right\|_{sp}\left\| \widehat{\mathbf{S}}_n(\mathbf{z}_t) - \mathbf{S}_n(\mathbf{z}_t) \right\|_{sp} = O_p\left(h^2 + \sqrt{\frac{\ln n}{nh^2}}\right), \tag{A.20}$$

where $\widehat{a}_{ij}(\mathbf{z}_t)$ and $a_{ij}(\mathbf{z}_t)$ are the (i,j)th element of $\mathbf{W}_0\widehat{\mathbf{S}}_n(\mathbf{z}_t)$ and $\mathbf{W}_0\mathbf{S}_n(\mathbf{z}_t)$, respectively. Consequently, we have

$$\left| n^{-1}\mathrm{tr}\left\{\mathbf{W}_0\left[\widehat{\mathbf{S}}_n(\mathbf{z}_{nt}) - \mathbf{S}_n(\mathbf{z}_{nt})\right]\right\}\right| \leq n^{-1}\sum_{i=1}^n\sum_{j\neq i}|w_{ij}|\left|\widehat{a}_{ij}(\mathbf{z}_t) - a_{ij}(\mathbf{z}_t)\right| = O_p\left(h^2 + \sqrt{\frac{\ln n}{nh^2}}\right), \tag{A.21}$$

which completes the proof of this lemma. ■

Lemma 5 *Under Assumptions **1**–**7**, we obtain* $(nh_0)^{-2}\zeta_n\widehat{A}_{n1}(z) = \varkappa_A(h_0, z) + o_p\left(h_0^2\right)$, *where*

$$\begin{aligned}
\varkappa_{A,Q}(h_0,z) &= \frac{1}{2}\upsilon_{1,2}(k)h_0^2\begin{bmatrix} 1 & 0 \end{bmatrix}' \otimes \left[\mathbf{E}_1(z)'\mathbf{E}_1(z)\nabla^2\boldsymbol{\gamma}(z)\right], \\
\varkappa_{A,P}(h_0,z) &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} F_1(h_0,z)\psi_1(h_0,z) \\ \mathbf{0}_{d_x} \end{bmatrix} = O_p\left(h_0^2\right),
\end{aligned}$$

where $\mathbf{E}_1(z) \equiv \lim_{n\to\infty} n^{-1}\sum_{i=1}^n\sum_{t=2}^T \varphi_{tt} f_t(z)\mathbb{E}\left(\mathbf{q}_{it}\mathbf{m}_{it}'|\mathbf{z}_{it} = z\right)$, *and* $F_1(h_0,z)$ *and* $\psi_1(h_0,z)$ *are defined in the proof below.*

Proof. First, we have

$$\begin{aligned}
\frac{1}{nh_0}\widehat{\mathbb{Q}}(z)'\mathbf{K}_{h_0}(z)\mathbb{M}(z) &= \frac{1}{nh_0}\sum_{i=1}^n\sum_{t=2}^T \varphi_{tt}k_{it}(h_0,z)\left[\mathcal{Z}_{it}(z)\mathcal{Z}_{it}(z)'\right] \otimes \left(\mathbf{q}_{it}\mathbf{m}_{it}'\right) + O_p\left(h^2 + \sqrt{\frac{\ln n}{nh^2}}\right) \\
&= \begin{bmatrix} 1 & 0 \\ \mathbf{0} & \upsilon_{1,2}(k) \end{bmatrix} \otimes \mathbf{E}_1(z) + O_p\left(h_0^2 + h^2 + \frac{1}{\sqrt{nh_0}} + \sqrt{\frac{\ln n}{nh^2}}\right), \qquad \text{(A.22)}
\end{aligned}$$

$$
\begin{aligned}
\frac{1}{nh_0}\mathbf{C}(z)'\mathbf{K}_{h_0}(z)\widehat{\mathbb{Q}}(z) &= \frac{1}{2nh_0}\sum_{i=1}^{n}\sum_{t=2}^{T}\varphi_{tt}k_{it}\left(h_0,z\right)\mathbf{m}_{it}'\mathbf{\Pi}(\mathbf{z}_{it}^*)\left[\mathcal{Z}_{it}(z)'\otimes\mathbf{q}_{it}'\right]+O_p\left(\left(h^2+\sqrt{\frac{\ln n}{nh^2}}\right)h_0^2\right)\\
&= \frac{1}{2}\upsilon_{1,2}(k)h_0^2\left[\begin{array}{cc}1 & 0\end{array}\right]\otimes\left[\mathbf{E}_1\left(z\right)\nabla^2\gamma(z)\right]\left[1+o_p\left(1\right)\right],
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{nh_0}\widehat{\mathbf{\Pi}}'\mathbf{K}_{h_0}(z)\widehat{\mathbb{Q}}(z) &= \frac{1}{nh_0}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=2}^{T}k_{it}\left(h_0,z\right)\widehat{\pi}_{it}\left[\mathcal{Z}_{it}(z)'\otimes\mathbf{q}_{it}'\right]+O_p\left(h^4+\frac{\ln n}{nh^2}\right)\\
&= O_p\left(h^2+\frac{\ln n}{nh^2}\right). \qquad \text{(A.23)}
\end{aligned}
$$

Second, letting $p_{t,ij}$ be the (i,j)th element of $P_t=\mathbf{W}_0\mathbf{S}_n\left(\mathbf{z}_t\right)-n^{-1}\mathrm{tr}\{\mathbf{W}_0\mathbf{S}_n\left(\mathbf{z}_t\right)\}\mathbf{I}_n$ and by Lemma **4**, we obtain

$$
\begin{aligned}
\frac{\widehat{\Psi}_1}{nh_0} &= \frac{1}{nh_0}\widehat{\mathbf{v}}'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)\\
&= \frac{1}{nh_0}\Delta\widetilde{\mathbf{u}}'\mathbf{K}_{h_0}(z)\mathbf{P}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)+\frac{1}{nh_0}\widehat{\mathbf{\Pi}}'\mathbf{K}_{h_0}(z)\mathbf{P}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)+\\
&\quad \frac{1}{nh_0}\widehat{\mathbf{v}}'\mathbf{K}_{h_0}(z)\left(\widehat{\mathbf{P}}_n-\mathbf{P}_n\right)^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)\\
&= \frac{2}{nh_0}\sum_{i=1}^{n}\sum_{t=2}^{T}\varphi_{tt}p_{t,ii}k_{it}^2\left(h_0,z\right)\Delta\widetilde{u}_{it}\mathcal{Z}_{it}(z)\otimes\left[\bar{u}_{it},\mathbf{0}_{d_x}'\right]'+O_p\left(h_0+h^2+\frac{1}{\sqrt{nh_0}}+\sqrt{\frac{\ln n}{nh^2}}\right)\\
&\equiv \left[\begin{array}{c}1\\0\end{array}\right]\otimes\left[\begin{array}{c}\psi_1\left(h_0,z\right)\\\mathbf{0}_{d_x}\end{array}\right]+o_p\left(1\right), \qquad \text{(A.24)}
\end{aligned}
$$

where $\psi_1\left(h_0,z\right)=2\left(nh_0\right)^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}\varphi_{tt}\mathbb{E}\left[p_{t,ii}k_{it}^2\left(h_0,z\right)\Delta\widetilde{u}_{it}\bar{u}_{it}\right]=O\left(1\right)$.

Similarly, we can show that

$$
\begin{aligned}
\frac{\widehat{\Psi}_2}{nh_0} &= \frac{1}{nh_0}\mathbf{C}\left(z\right)'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)\\
&= \frac{1}{nh_0}\mathbf{C}\left(z\right)'\mathbf{K}_{h_0}(z)\mathbf{P}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)+\frac{1}{nh_0}\mathbf{C}\left(z\right)'\mathbf{K}_{h_0}(z)\left(\widehat{\mathbf{P}}_n-\mathbf{P}_n\right)^s\mathbf{K}_{h_0}(z)\mathbb{M}(z)\\
&= O_p\left(h_0^2+\left(nh_0\right)^{-1/2}\right), \qquad \text{(A.25)}\\
\frac{\widehat{\Gamma}_1}{nh_0} &= \frac{1}{nh_0}\widehat{\mathbf{v}}'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbf{C}(z)\\
&= \frac{2}{nh_0}\sum_{i=1}^{n}\sum_{t=2}^{T}\varphi_{tt}p_{t,ii}k_{it}^2\left(h_0,z\right)\Delta\widetilde{u}_{it}\widetilde{c}_{it}(z)+O_p\left(h_0^3+h_0\left(h^2+\sqrt{\frac{\ln n}{nh^2}}\right)\right)\\
&\equiv F_1\left(h_0,z\right)+O_p\left(h_0^3+h_0^2\left(h^2+\sqrt{\frac{\ln n}{nh^2}}\right)\right)=O_p\left(h_0^2\right), \qquad \text{(A.26)}\\
\frac{\widehat{\Gamma}_2}{nh_0} &= \frac{1}{nh_0}\mathbf{C}\left(z\right)'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbf{C}(z)\\
&= \frac{2}{nh_0}\sum_{i=1}^{n}\sum_{t=2}^{T}\varphi_{tt}^2p_{t,ii}k_{it}^2\left(h_0,z\right)c_{it}^2(z)+O_p\left(h_0^4\left(h^2+\sqrt{\frac{\ln n}{nh^2}}\right)\right)=O_p\left(h_0^4\right), \qquad \text{(A.27)}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\widehat{\Gamma}_3}{nh_0} &= \frac{1}{nh_0}\widehat{\mathbf{v}}'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\widehat{\mathbf{v}} \\
&= \frac{2}{nh_0}\sum_{i=1}^n\sum_{t=2}^T p_{t,ii}k_{it}^2(h_0,z)(\Delta\widetilde{u}_{it})^2 + O_p\left(\left(h^2+\sqrt{\frac{\ln n}{nh^2}}\right)\left(h_0+h^2+\frac{1}{\sqrt{nh_0}}+\sqrt{\frac{\ln n}{nh^2}}\right)\right) \\
&= O_p\left(h_0^2+\frac{1}{\sqrt{nh_0}}\right)+O_p\left(\left(h^2+\sqrt{\frac{\ln n}{nh^2}}\right)\left(h^2+h_0+\sqrt{\frac{\ln n}{nh^2}}+\frac{1}{\sqrt{nh_0}}\right)\right), \qquad \text{(A.28)}
\end{aligned}
$$

since $\mathrm{tr}\{P_t\}=0$ for all t. It then follows that $(nh_0)^{-2}\widehat{\Gamma}_3\widehat{\Psi}_2 = o_p(h_0^2)$. In addition, we obtain

$$
\frac{2}{nh_0}\sum_{i=1}^n\sum_{t=2}^T p_{t,ii}k_{it}^2(h_0,z)(\Delta\widetilde{u}_{it})^2\psi_1(h_0,z) = O_p\left(\frac{1}{nh_0}+\frac{1}{n}\right),
$$

so that $(nh_0)^{-2}\widehat{\Gamma}_3\widehat{\Psi}_1 = o_p(h_0^2)$ under Assumption **7**. This completes the proof of this lemma. ■

Lemma 6 *Under Assumptions **1**–**7**, we obtain*

$$
(nh_0)^{-2}\zeta_n^2\widehat{B}_n(z) = \varkappa_B(h_0,z)+O\left(\zeta_n^{-1}+\zeta_n^{-2}\right)+o_p(1), \qquad \text{(A.29)}
$$

where

$$
\varkappa_{B,Q}(z) = \begin{bmatrix} 1 & 0 \\ 0 & \upsilon_{1,2}^2(k)\end{bmatrix}\otimes\left[\mathbf{E}_1(z)'\mathbf{E}_1(z)\right] \text{ and } \varkappa_{B,P}(h_0,z) = \begin{bmatrix} 1 & 0 \\ 0 & 0\end{bmatrix}\otimes\begin{bmatrix}\psi_1^2(h_0,z) & \mathbf{0}_{d_x}' \\ \mathbf{0}_{d_x} & \mathbf{0}_{d_x\times d_x}\end{bmatrix}.
$$

Proof. By (A.22), we have

$$
\frac{1}{(nh_0)^2}\mathbb{M}(z)'\widehat{\mathbf{\Xi}}_{h_0}(z)\mathbb{M}(z) = \varkappa_{B,Q}(z)(1+o_p(1)).
$$

In addition, following the proof of Lemma **5**, we have

$$
\begin{aligned}
\frac{\widehat{\Psi}_{3,l}}{nh_0} &= \frac{1}{nh_0}\mathbb{M}(z)'\mathbf{K}_{h_0}(z)\widehat{\mathbf{P}}_n^s\mathbf{K}_{h_0}(z)\mathbb{M}(z) \\
&= \frac{2}{nh_0}\sum_{i=1}^n\sum_{t=2}^T\varphi_{tt}^2p_{t,ii}k_{it}^2(z)\left[\mathcal{Z}_{it}(z)\mathcal{Z}_{it}(z)'\right]\otimes\left(\mathbf{m}_{it}\mathbf{m}_{it}'\right)+O_p\left(h^2+\sqrt{\frac{\ln n}{nh^2}}\right) = O_p(1),
\end{aligned}
$$

which, when combined with (A.24) and (A.25), gives

$$
\begin{aligned}
(nh_0)^{-2}\zeta_n^2\widehat{B}_n(z) &= \frac{1}{(nh_0)^2}\widehat{\Psi}_1'\widehat{\Psi}_1+\frac{1}{(nh_0)^2}\mathbb{M}(z)'\widehat{\mathbf{\Xi}}_{h_0}(z)\mathbb{M}(z)+O_p\left(\zeta_n^{-1}+\zeta_n^{-2}+h_0^2+(nh_0)^{-1/2}\right) \\
&= \varkappa_{B,P}(h_0,z)+\varkappa_{B,Q}(z)+o_p(1)+O_p\left(\zeta_n^{-1}+\zeta_n^{-2}\right).
\end{aligned}
$$

This completes the proof of this lemma. ■

Remark 3 If model (1.2) becomes a parametric spatial autoregressive panel data model with random effects, $\mathbf{E}_1(z) \equiv \left[\mathbf{0}_{d_q}, n^{-1}\sum_{i=1}^n\sum_{t=2}^T\varphi_{tt}f_t(z)\mathbb{E}\left(\mathbf{q}_{it}\mathbf{x}_{it}'|\mathbf{z}_{it}=z\right)\right]$ causing $\varkappa_{B,Q}(z)$ to be singular. However, $\varkappa_B(h_0,z)$ can still be a nonsingular matrix so long as $\psi_1(h_0,z)=O_e(1)$.

Lemma 7 *Under Assumptions **1**–**7**, we obtain*

$$(nh_0)^{-3/2}\,\zeta_n \widehat{A}_{n,2}(z) \xrightarrow{d} \mathbb{N}\Big(\mathbf{0}, \sigma_u^2 \begin{bmatrix} \upsilon_{2,0}(k) & 0 \\ 0 & \upsilon_{2,2}(k)\,\upsilon_{1,2}^2(k) \end{bmatrix} \otimes \mathbf{\Omega}(z)\Big), \tag{A.30}$$

where $\mathbf{\Omega}(z) = \mathbf{E}_1(z)'\,\mathbf{E}_3(z)\,\mathbf{E}_1(z)$ *with* $\mathbf{E}_3(z) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \sum_{t=1}^T f_t(z)\,\mathbb{E}[\mathbf{q}_{it}\mathbf{q}_{it}'|z]$.

Proof. By definition, $(nh_0)^{-3/2}\,\zeta_n \widehat{A}_{n,2}(z) = \Big[(nh_0)^{-1}\,\mathbb{M}(z)'\,\mathbf{K}_{h_0}(z)\widehat{\mathbb{Q}}(z)\Big]\,(nh_0)^{-1/2}\,\widehat{\mathbb{Q}}(z)'\,\mathbf{K}_{h_0}(z)\Delta\widetilde{\mathbf{u}}$. We need to show that

$$\frac{1}{\sqrt{nh_0}}\widehat{\mathbb{Q}}(z)'\,\mathbf{K}_{h_0}(z)\Delta\widetilde{\mathbf{u}} = \frac{1}{\sqrt{nh_0}}\sum_{i=1}^n \mathbb{Q}_i(z)\,\mathbf{K}_i(\mathbf{z})\,\Delta\widetilde{\mathbf{u}}_i + o_p(1)$$

converges to a normal distribution, where $\mathbb{Q}_i = [\mathbb{Q}_{i2}, \ldots, \mathbb{Q}_{iT}]'$, $\Delta\widetilde{\mathbf{u}}_i = [\Delta\widetilde{u}_{i2}, \ldots, \Delta\widetilde{u}_{iT}]'$ and $\mathbf{K}_i(z) = \operatorname{diag}\{k_{i2}(h_0, z), \ldots, k_{iT}(h_0, z)\}$.

Since the proof of this result closely follows that of Lemma **3**, we only show the following:

$$\begin{aligned} \frac{1}{nh_0}\sum_{i=1}^n \mathbb{Q}_i(z)\,\mathbf{K}_i^2(\mathbf{z})\,\mathbb{Q}_i(z) &= \frac{1}{nh_0}\sum_{i=1}^n\sum_{t=1}^T k_{it}^2(h_0, z)\left[\mathcal{Z}_{it}(z)\,\mathcal{Z}_{it}(z)'\right] \otimes \left(\mathbf{q}_{it}\mathbf{q}_{it}'\right) \\ &= \begin{bmatrix} \upsilon_{2,0}(k) & 0 \\ 0 & \upsilon_{2,2}(k) \end{bmatrix} \otimes \frac{1}{n}\sum_{i=1}^n\sum_{t=1}^T f_t(z)\,\mathbb{E}\left[q_{it}q_{it}'|\mathbf{z}\right] + O_p\left(h_0^2 + \frac{1}{\sqrt{nh_0}}\right). \end{aligned}$$

Combing above results with (A.22) completes the proof of this lemma. ■

Proof of Theorem 3. Under Assumption **8**, we have

$$\widetilde{\boldsymbol{\theta}}(z) - \boldsymbol{\theta}(z) = \widetilde{\boldsymbol{\theta}}(z) - \boldsymbol{\theta}^*(z) + \boldsymbol{\theta}^*(z) - \boldsymbol{\theta}(z) = \left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]'\left(\widetilde{\boldsymbol{\psi}} - \boldsymbol{\psi}\right) + O\left(L_n^{-\xi}\right).$$

By (4.4), we have $\widetilde{\boldsymbol{\psi}} = (\mathcal{X}'\mathbf{M_D}\mathcal{X})^{-1}\,\mathcal{X}'\mathbf{M_D}\widetilde{\dot{\mathbf{y}}} = \boldsymbol{\psi} + \boldsymbol{\Delta}_{n1} + \boldsymbol{\Delta}_{n2}$ since $\mathbf{M_D}\mathbf{D} = \mathbf{0}_{nT}$, where $\boldsymbol{\Delta}_{n1} \equiv (\mathcal{X}'\mathbf{M_D}\mathcal{X})^{-1}\,\mathcal{X}'\mathbf{M_D}\mathbf{\Pi}$, $\boldsymbol{\Delta}_{n2} \equiv (\mathcal{X}'\mathbf{M_D}\mathcal{X})^{-1}\,\mathcal{X}'\mathbf{M_D}\mathbf{u}$, and $\mathbf{\Pi} = [\mathbf{\Pi}_1', \ldots, \mathbf{\Pi}_n']'$ is of dimension nT with $\mathbf{\Pi}_i = [\Pi_{i1}, \ldots, \Pi_{iT}]'$ and $\Pi_{it} = \mathbf{m}_{it}'\left[\boldsymbol{\gamma}(z_{it}) - \widetilde{\boldsymbol{\gamma}}(z_{it})\right] + \mathbf{g}_i'\left[\boldsymbol{\theta}(z_{it}) - \boldsymbol{\theta}^*(z_{it})\right]$.

First, we show that $\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]'\boldsymbol{\Delta}_{n2} = O_p\left(\sqrt{L_n/n}\right)$ has an asymptotic normal distribution. Letting $\boldsymbol{\alpha} \neq \mathbf{0}$ be any finite vector of dimension d_g, we have $\Gamma_n \equiv \boldsymbol{\alpha}'\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]'\boldsymbol{\Delta}_{n2} \equiv \Gamma_{n,1} + \Gamma_{n,2}$, where $\Gamma_{n,1} \equiv \boldsymbol{\alpha}'\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]'\boldsymbol{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\mathcal{X}'\mathbf{M_D}\mathbf{u}/n$ and $\Gamma_{n,2} \equiv \boldsymbol{\alpha}'\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]'\left(\boldsymbol{\Sigma}_{n,\boldsymbol{\phi\phi}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\right) \times \mathcal{X}'\mathbf{M_D}\mathbf{u}/n$ with $\boldsymbol{\Sigma}_{n,\boldsymbol{\phi\phi}} \equiv n^{-1}\mathcal{X}'\mathbf{M_D}\mathcal{X}$. Since $\Gamma_{n,2}$ is a vector, we have

$$\begin{aligned} \|\Gamma_{n,2}\| &= n^{-1}\left\|\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]'\left(\boldsymbol{\Sigma}_{n,\boldsymbol{\phi\phi}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\right)\mathcal{X}'\mathbf{M_D}\mathbf{u}\right\|_{sp} \\ &\leq n^{-1}\left\|\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]'\right\|_{sp}\left\|\boldsymbol{\Sigma}_{n,\boldsymbol{\phi\phi}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\right\|_{sp}\left\|\mathcal{X}'\mathbf{M_D}\mathbf{u}\right\|_{sp} \\ &= O_p\left(\frac{\sqrt{L_n}}{n}\sqrt{\frac{L_n \log L_n}{n}}\sqrt{L_n}\right) = o_p\left(\frac{L_n}{n}\right) \end{aligned}$$

under Assumptions **7** and **9**, where also, under Assumption **1**, we have

$$\mathbb{E}\left[\left\|\mathcal{X}'M_{\mathbf{D}}\mathbf{u}\right\|^2\right] = \mathbb{E}\left[\mathbf{u}'\mathbf{M_D}\mathcal{X}\left(\mathcal{X}'\mathbf{M_D}\mathcal{X}\right)^{-1}\mathcal{X}'\mathbf{M_D}\mathbf{u}\right]$$

$$
\begin{aligned}
&= \sigma_u^2 \mathbb{E}\left[\operatorname{tr}\left(\mathbf{M_D}\mathcal{X}\left(\mathcal{X}'\mathbf{M_D}\mathcal{X}\right)^{-1}\mathcal{X}'\mathbf{M_D}\right)\right] \\
&= \sigma_u^2 \operatorname{tr}\left(\mathbf{I}_{d_g L_n}\right) = \sigma_u^2 d_g L_n,
\end{aligned}
$$

so that $\|\mathcal{X}'\mathbf{M_D}\mathbf{u}\|_{sp} = \|\mathcal{X}'\mathbf{M_D}\mathbf{u}\| = O_p(L_n)$ by Markov's inequality. In addition,

$$
\begin{aligned}
\left\|\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}}^{-1} - \mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\right\|_{sp} &= \left\|\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}}^{-1}\left(\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}} - \mathbf{\Sigma}_{\boldsymbol{\phi\phi}}\right)\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\right\|_{sp} \le \left\|\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}}^{-1}\right\|_{sp}\left\|\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}} - \mathbf{\Sigma}_{\boldsymbol{\phi\phi}}\right\|_{sp}\left\|\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\right\|_{sp} \\
&= O_p(1)\left\|\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}}^{-1}\right\|_{sp}\left\|\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}} - \mathbf{\Sigma}_{\boldsymbol{\phi\phi}}\right\| = O_p\left(\sqrt{\frac{L_n \log L_n}{n}}\right), \qquad \text{(A.31)}
\end{aligned}
$$

since $\left\|\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}}^{-1}\right\|_{sp} = \lambda_{\min}^{-1}\left(\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}}\right) = \lambda_{\min}^{-1}\left(\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}\right) + O\left(\|\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}} - \mathbf{\Sigma}_{\boldsymbol{\phi\phi}}\|_1\right)$ by Weyl's theorem in Seber (2008, p.117), $\|\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}} - \mathbf{\Sigma}_{\boldsymbol{\phi\phi}}\| = O_p\left(\sqrt{L_n(\log L_n)/n}\right)$ by Lemma 6.2 in Belloni, Chernozhukov, Chetverikov & Kato (2015), and the B-spline basis functions are uniformly bounded over the compact domain of z.

Next, we consider

$$
\begin{aligned}
\sqrt{n}\Gamma_{n,1} &\equiv n^{-1/2}\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)\right]'\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\mathcal{X}'\mathbf{M_D}\mathbf{u} \\
&= n^{-1/2}\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)\right]'\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\sum_{i=1}^{n}\mathcal{X}_i'\left(\mathbf{I}_T - T^{-1}\mathbf{i}_T\mathbf{i}_T'\right)\mathbf{u}_i \\
&= n^{-1/2}\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)\right]'\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\mathbf{g}_i\otimes\overline{\boldsymbol{\phi}}_{L_n}(z_{it})\right]u_{it},
\end{aligned}
$$

where

$$
\mathbb{V}\mathrm{ar}\left(\sqrt{n}\Gamma_{n,1}|\mathbf{G},\mathbf{Z}\right) = \sigma_u^2\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)\right]'\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\mathbf{\Sigma}_{n,\boldsymbol{\phi\phi}}\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)\right]\boldsymbol{\alpha} = \sigma_n^2\left(1+o_p(1)\right)
$$

and $\sigma_n^2 \equiv \sigma_u^2\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)\right]'\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)\right]\boldsymbol{\alpha} \ge \sigma_u^2\lambda_{\max}^{-1}\left(\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}\right)\left\|\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)\right]\right\| \ge CL_n$.

Under Assumption **1**, $\mathbb{E}[\Gamma_{n,1}] = 0$ and, by Minkowski's inequality, we have for some $\delta > 0$:

$$
\begin{aligned}
&\mathbb{E}\left|\sigma_n^{-1}\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)\right]'\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\sum_{t=1}^{T}\left[\mathbf{g}_i\otimes\overline{\boldsymbol{\phi}}_{L_n}(z_{it})\right]u_{it}\right|^{2+\delta} \\
\le\ & \frac{1}{\sigma_n^{2+\delta}}\left\{\sum_{t=1}^{T}\left(\mathbb{E}\left[\left|\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)\right]'\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\ \left[\mathbf{g}_i\otimes\overline{\boldsymbol{\phi}}_{L_n}(z_{it})\right]\right|^{2+\delta}|u_{it}|^{2+\delta}\right]\right)^{1/(2+\delta)}\right\}^{2+\delta} \\
\le\ & \frac{1}{\sigma_n^{2+\delta}}\lambda_{\min}^{-(2+\delta)}\left(\mathbf{\Sigma}_{\boldsymbol{\phi\phi}}\right)\left\|\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g}\otimes\boldsymbol{\phi}_{L_n}(z)'\right]\right\|^{2+\delta}\left\{\sum_{t=1}^{T}\left(\mathbb{E}\left[\left\|\mathbf{g}_i\otimes\overline{\boldsymbol{\phi}}_{L_n}(z_{it})\right\|^{2+\delta}|u_{it}|^{2+\delta}\right]\right)^{1/(2+\delta)}\right\}^{2+\delta} \\
\le\ & C\,\frac{L_n^{2+\delta}}{\sigma_n^{2+\delta}}\left\{\sum_{t=1}^{T}\left(\mathbb{E}\left[\left|\mathbf{g}_i'\mathbf{g}_i\right|^{2+\delta}|u_{it}|^{2+\delta}\right]\right)^{1/(2+\delta)}\right\}^{2+\delta} \le C
\end{aligned}
$$

if $\max_{1\le t\le T}\mathbb{E}\left[\left|\mathbf{g}_i'\mathbf{g}_i\right|^{2+\delta}|u_{it}|^{2+\delta}\right] \le C < \infty$, since $\max_{z\in\mathcal{S}_z}\sum_{j=1}^{L_n}\boldsymbol{\phi}_j^2(z) \le ML_n$ for B-spline basis functions and $\left\|\mathbf{g}_i\otimes\overline{\boldsymbol{\phi}}_{L_n}(z_{it})\right\| = \sqrt{\mathbf{g}_i'\mathbf{g}_i\overline{\boldsymbol{\phi}}_{L_n}(z_{it})'\overline{\boldsymbol{\phi}}_{L_n}(z_{it})}$. Applying Liapounonv's central limit theorem we obtain that $\sqrt{n}\Gamma_{n,1}/\sigma_n \xrightarrow{d} \mathbb{N}(0,1)$, which implies that $\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]' \boldsymbol{\Delta}_{n2} = O_p\left(\sqrt{L_n/n}\right)$.

Second, we consider $\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]' \boldsymbol{\Delta}_{n1}$. Applying a method similar to the one used above, we obtain that $\Lambda_n \equiv \boldsymbol{\alpha}'\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]' \boldsymbol{\Delta}_{n1} = \boldsymbol{\alpha}'\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]' (\mathcal{X}'\mathbf{M_D}\mathcal{X})^{-1} \mathcal{X}'\mathbf{M_D}\boldsymbol{\Pi} = \Lambda_{n,1}\big(1 + o_p(1)\big)$, where $\Lambda_{n,1} \equiv \boldsymbol{\alpha}'\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]' \boldsymbol{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\mathcal{X}'\mathbf{M_D}\boldsymbol{\Pi}/n$ and

$$\begin{aligned}
\|\Lambda_{n,1}\| &= \|\Lambda_{n,1}\|_{sp} \leq \left\|\boldsymbol{\alpha}'\left[\mathbf{I}_{d_g} \otimes \boldsymbol{\phi}_{L_n}(z)\right]'\right\|_{sp} \left\|\boldsymbol{\Sigma}_{\boldsymbol{\phi\phi}}^{-1}\right\|_{sp} \left\|\mathcal{X}'\mathbf{M_D}\boldsymbol{\Pi}/n\right\|_{sp} \\
&\leq Cn^{-1/2}L_n^{1/2}\left(\boldsymbol{\Pi}'\mathbf{M_D}\mathcal{X}\mathcal{X}'\mathbf{M_D}\boldsymbol{\Pi}/n\right)^{1/2} \\
&\leq Cn^{-1/2}L_n^{1/2}\left(\boldsymbol{\Pi}'\boldsymbol{\Pi}\right)^{1/2}\lambda_{\max}^{1/2}\left(\boldsymbol{\Sigma}_{n,\phi\phi}\right) \\
&= O_p\left(\sqrt{\frac{L_n}{n}}\left(h_0^2 + \sqrt{\frac{\ln n}{nh_0}} + L_n^{-\xi}\right)\right)
\end{aligned}$$

under Assumptions **7** and **8**. We extend the pointwise convergence result in Theorem **2** to the uniform convergence result $\max_{z \in \mathcal{S}_z} \|\widetilde{\boldsymbol{\gamma}}(z) - \boldsymbol{\gamma}(z)\| = O_p\left(h_0^2 + \sqrt{(\ln n)/(nh_0)}\right)$, following the proof in Masry (1996). This completes the proof of this theorem. ■

Proof of Theorem 4. Begin by defining $\boldsymbol{\Delta}_i = [\Delta_{i2}, \ldots, \Delta_{iT}]'$, where $\Delta_{it} = \mathbf{x}_{it}'\left[\boldsymbol{\beta}(z_{it}) - \widetilde{\boldsymbol{\beta}}_0(z_{it})\right] + (\mathbf{Wy})_{it}\,\rho(z_{it})$. Hence, $\widehat{\varepsilon}_{it} = \widehat{\epsilon}_{it} + \Delta_{it} + \Delta u_{it}$, where $\widehat{\epsilon}_{it} \equiv \mathbf{g}_i'\left[\dot{\boldsymbol{\theta}}(z_{it}) - \widehat{\dot{\boldsymbol{\theta}}}_1(z_{it})\right] - \mathbf{m}_{i,t-1}'\left[\boldsymbol{\gamma}(z_{i,t-1}) - \widehat{\boldsymbol{\gamma}}_1(z_{i,t-1})\right]$ and $\widehat{\dot{\boldsymbol{\theta}}}_1(\cdot)$ and $\widehat{\boldsymbol{\gamma}}_1(\cdot)$ are the first-stage estimator under H_1. Applying simple algebra, we obtain

$$\begin{aligned}
\widehat{T}_n &= \frac{1}{n^2h_0}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\widehat{\boldsymbol{\varepsilon}}_i' A_{ij}\widehat{\boldsymbol{\varepsilon}}_j \\
&= \frac{1}{n^2h_0}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\boldsymbol{\Delta}_i' A_{ij}\boldsymbol{\Delta}_j + \frac{2}{n^2h_0}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\boldsymbol{\Delta}_i' A_{ij}\widehat{\boldsymbol{\epsilon}}_j + \frac{2}{n^2h_0}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\boldsymbol{\Delta}_i' A_{ij}\Delta\mathbf{u}_j \; + \\
&\quad \frac{2}{n^2h_0}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\widehat{\boldsymbol{\epsilon}}_i' A_{ij}\Delta\mathbf{u}_j + \frac{1}{n^2h_0}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\widehat{\boldsymbol{\epsilon}}_i' A_{ij}\widehat{\boldsymbol{\epsilon}}_j + \frac{1}{n^2h_0}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\Delta\mathbf{u}_i' A_{ij}\Delta\mathbf{u}_j \\
&\equiv \widehat{T}_{n,1} + 2\left(\widehat{T}_{n,2} + \widehat{T}_{n,3} + \widehat{T}_{n,4}\right) + \widehat{T}_{n,5} + \widehat{T}_{n,6},
\end{aligned}$$

where we denote each term appearing in the second equation as $\widehat{T}_{n,j}$ $(j = 1, \ldots, 6)$ in the order of their appearance. In what follows, we show that (i) $n\sqrt{h_0}\widehat{T}_{n,6}$ converges to a normal distribution under both hypotheses, (ii) $n\sqrt{h_0}\widehat{T}_{n,j} = o_p(1)$ under both hypotheses for $j = 4$ and 5, and (iii) $n\sqrt{h_0}\widehat{T}_{n,j} = o_p(1)$ under H_0 and is explosive under H_1 for $j = 1, 2, 3$.

First, we consider $n\sqrt{h_0}\widehat{T}_{n,6} = n^{-2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} H_n\left(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j\right)$, where we denote $H_n\left(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j\right) = h_0^{-1/2}\Delta\mathbf{u}_i' A_{ij}\Delta\mathbf{u}_j$ and $\boldsymbol{\chi}_i = (\Delta\mathbf{u}_i, \mathcal{X}_i)$. Evidently, under Assumption **1**, $H_n\left(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j\right)$ is a symmetric matrix with $\mathbb{E}\left[H_n\left(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j\right)|\boldsymbol{\chi}_i\right] = 0$ almost surely for all $i \neq j$. It is readily seen that

$$\frac{\mathbb{E}\left[G_n^2\left(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2\right)\right] + n^{-1}\mathbb{E}\left[H_n^4\left(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2\right)\right]}{\left(\mathbb{E}\left[H_n^2\left(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2\right)\right]\right)^2} = \frac{O_p\left(h_0^2\right) + O_p\left((nh_0)^{-1}\right)}{O_p(1)} = o_p(1)$$

if $h_0 \to 0$ and $nh_0 \to \infty$ as $n \to \infty$, where $G_n(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2) = \mathbb{E}_{\boldsymbol{\chi}_i}[H_n(\boldsymbol{\chi}_1, \boldsymbol{\chi}_i) H_n(\boldsymbol{\chi}_2, \boldsymbol{\chi}_i)]$. Specifically, we have

$$\begin{aligned} \mathbb{E}\left[H_n^2(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2)\right] &= h_0^{-1}\mathbb{E}\left[\Delta\mathbf{u}_1' A_{12}\Delta\mathbf{u}_2\Delta\mathbf{u}_2' A_{12}\Delta\mathbf{u}_1\right] \\ &= \frac{\sigma_u^2}{h_0}\mathbb{E}\left[\Delta\mathbf{u}_1' A_{12}\boldsymbol{\Sigma} A_{12}\Delta\mathbf{u}_1\right] = \frac{\sigma_u^4}{h_0^2}\mathrm{tr}\left\{E\left[A_{12}\boldsymbol{\Sigma} A_{12}\boldsymbol{\Sigma}\right]\right\} \\ &= \frac{\sigma_u^4}{h_0}\sum_{t=2}^{T}\sum\nolimits_{0\le|s-t|\le 1} c_{ts}\mathbb{E}\left[k^2\left(\frac{z_{1t}-z_{2s}}{h_0}\right)\left(\mathbf{x}_{1t}'\mathbf{x}_{2s}\right)^2\right] + O(h_0) \\ &= \sigma_u^4 \upsilon_{2,0}(k)\sum\sum\nolimits_{0\le|s-t|\le 1} c_{ts}\mathbb{E}\left[\mu(z_{2s})\, vec\left(\mathbf{x}_{2s}\mathbf{x}_{2s}'\right) f(z_{2s})\right] + O(h_0) \\ &\equiv \sigma_0^2/2 + O(h_0), \end{aligned}$$

where $\boldsymbol{\Sigma}$ is defined in Section **3**, and $\mu(z_{2s})$ and c_{ts} are defined in Theorem **4**. Applying Hall's (1984) central limit theorem gives $n\sqrt{h_0}\widehat{T}_{n,6} \overset{d}{\to} \mathbb{N}\left(0, \sigma_0^2\right)$.

Second, we verify $n\sqrt{h_0}\widehat{T}_{n,5} = O_p\left(nh^4\sqrt{h_0}\right)$ under both hypotheses, where

$$n\sqrt{h_0}\widehat{T}_{n,5} = \frac{1}{n\sqrt{h_0}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{\epsilon}_{it}\widehat{\epsilon}_{js}k\left(\frac{z_{it}-z_{js}}{h_0}\right)\mathbf{x}_{it}'\mathbf{x}_{js}.$$

Following the proof in Masry (1996), we obtain the uniform convergence result for the first-stage estimator such that $\left[\mathbf{E}_1(\mathbf{z})\mathbf{E}_1(\mathbf{z})'\right]^{-1}\mathbf{E}_1(\mathbf{z})\left[h^2\upsilon_{1,2}(k)\mathbf{E}_2(\mathbf{z}) + \left(nh^2\right)^{-1}\mathbf{Q}'\mathbf{K}_h(\mathbf{z})\Delta\mathbf{u}\right]$ is the leading term of $\widehat{\boldsymbol{\Theta}}_1(\mathbf{z}) - \boldsymbol{\Theta}(\mathbf{z})$ over $\mathbf{z} \in \mathcal{S}_z \times \mathcal{S}_z$, where $\mathbf{E}_1(\mathbf{z})$, $\mathbf{E}_2(\mathbf{z})$, $\mathbf{Q}$ and $\mathbf{K}_h(\mathbf{z})$ are as defined in Theorem **1.** Let $\mathbf{S}_1$ be the first d_g columns of the identity matrix $\mathbf{I}_{d_g+2(d_x+1)}$ and $\mathbf{S}_2$ contain the (d_g+d_x+1)th to the $[d_g+2(d_x+1)]$th columns of matrix $\mathbf{I}_{d_g+2(d_x+1)}$. Then, we have that

$$\begin{aligned} \widehat{\epsilon}_{it} &= \mathbf{g}_i'\left[\dot{\boldsymbol{\theta}}(z_{it}) - \widehat{\dot{\boldsymbol{\theta}}}_1(z_{it})\right] - \mathbf{m}_{i,t-1}'\left[\boldsymbol{\gamma}(z_{i,t-1}) - \widehat{\boldsymbol{\gamma}}_1(z_{i,t-1})\right] \\ &= \left(-\mathbf{g}_i'\mathbf{S}_1' + \mathbf{m}_{i,t-1}'\mathbf{S}_2'\right)\boldsymbol{\chi}(\mathbf{z}_i)\left[h^2\upsilon_{1,2}(k)\mathbf{E}_2(\mathbf{z}_i) + \left(nh^2\right)^{-1}\mathbf{Q}'\mathbf{K}_h(\mathbf{z}_i)\Delta\mathbf{u}\right] \times [1 + o_p(1)] \end{aligned}$$

holds uniformly over all i and t, where $\boldsymbol{\chi}(\mathbf{z}_i) \equiv \left[\mathbf{E}_1(\mathbf{z}_i)\mathbf{E}_1(\mathbf{z}_i)'\right]^{-1}\mathbf{E}_1(\mathbf{z}_i)$. Thus, we have

$$\begin{aligned} n\sqrt{h_0}\widehat{T}_{n,5} &\approx \frac{1}{n\sqrt{h_0}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} k\left(\frac{z_{it}-z_{js}}{h_0}\right)\mathbf{x}_{it}'\mathbf{x}_{js} \\ &\quad \times \left(-\mathbf{g}_i'\mathbf{S}_1' + \mathbf{m}_{i,t-1}'\mathbf{S}_2'\right)\boldsymbol{\chi}(\mathbf{z}_i)\left[h^2\upsilon_{1,2}(k)\mathbf{E}_2(\mathbf{z}_i) + \left(nh^2\right)^{-1}\mathbf{Q}'\ \mathbf{K}_h(\mathbf{z}_i)\Delta\mathbf{u}\right] \\ &\quad \times \left(-\mathbf{g}_j'\mathbf{S}_1' + \mathbf{m}_{j,t-1}'\mathbf{S}_2'\right)\boldsymbol{\chi}(\mathbf{z}_j)\left[h^2\upsilon_{1,2}(k)\mathbf{E}_2(\mathbf{z}_j) + \left(nh^2\right)^{-1}\mathbf{Q}'\ \mathbf{K}_h(\mathbf{z}_j)\Delta\mathbf{u}\right] \\ &= O_p\left(nh^4\sqrt{h_0} + h^2\sqrt{nh_0} + \sqrt{h_0}/h^2\right). \end{aligned} \tag{A.32}$$

Third, we verify that $n\sqrt{h_0}\widehat{T}_{n,4} = O_p\left(h^2\sqrt{nh_0} + \sqrt{h_0}/h\right)$ under both hypotheses, where

$$\begin{aligned} n\sqrt{h_0}\widehat{T}_{n,4} &= \frac{1}{n\sqrt{h_0}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{\epsilon}_{it}k\left(\frac{z_{it}-z_{js}}{h_0}\right)\mathbf{x}_{it}'\mathbf{x}_{js}\Delta u_{js} \\ &\approx \frac{1}{n\sqrt{h_0}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} k\left(\frac{z_{it}-z_{js}}{h_0}\right)\mathbf{x}_{it}'\mathbf{x}_{js}\Delta u_{js} \end{aligned}$$

$$\times\left(-\mathbf{g}_i'\mathbf{S}_1' + \mathbf{m}_{i,t-1}'\mathbf{S}_2'\right)\boldsymbol{\chi}\left(\mathbf{z}_i\right)\left[h^2\upsilon_{1,2}\left(k\right)\mathbf{E}_2\left(\mathbf{z}_i\right) + \left(nh^2\right)^{-1}\mathbf{Q}'\,\mathbf{K}_h\left(\mathbf{z}_i\right)\Delta\mathbf{u}\right]$$
$$= O_p\left(h^2\sqrt{nh_0} + \sqrt{h_0}/h + \sqrt{h_0}\right). \tag{A.33}$$

Fourth, we consider $n\sqrt{h_0}\widehat{T}_{n,3}$. Similar to the proof for $n\sqrt{h_0}\widehat{T}_{n,4}$, we obtain that

$$\begin{aligned} n\sqrt{h_0}\widehat{T}_{n,3} &= \frac{1}{n\sqrt{h_0}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}k\left(\frac{z_{it}-z_{js}}{h_0}\right)\mathbf{x}_{it}'\mathbf{x}_{js}\Delta u_{js}\mathbf{x}_{it}'\left[\boldsymbol{\beta}\left(z_{it}\right)-\widetilde{\boldsymbol{\beta}}_0\left(z_{it}\right)\right] \\ &= O_p\left(\widetilde{h}_0^2\sqrt{nh_0} + \sqrt{h_0/\widetilde{h}_0}\right) \text{ under } H_0, \end{aligned} \tag{A.34}$$

and

$$\begin{aligned} n\sqrt{h_0}\widehat{T}_{n,3} &= \frac{1}{n\sqrt{h_0}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}k\left(\frac{z_{it}-z_{js}}{h_0}\right)\mathbf{x}_{it}'\mathbf{x}_{js}\Delta u_{js} \\ &\quad\times\left(\mathbf{x}_{it}'\left[\boldsymbol{\beta}\left(z_{it}\right)-\widetilde{\boldsymbol{\beta}}_0\left(z_{it}\right)\right] + \left(\mathbf{W}\mathbf{y}\right)_{it}\rho(z_{it})\right) \\ &= O_p\left(\sqrt{nh_0}\right) \text{ under } H_1. \end{aligned} \tag{A.35}$$

Fifth, we have

$$\begin{aligned} n\sqrt{h_0}\widehat{T}_{n,2} &\approx \frac{1}{n\sqrt{h_0}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}k\left(\frac{z_{it}-z_{js}}{h_0}\right)\mathbf{x}_{it}'\mathbf{x}_{js} \\ &\quad\times\left(\mathbf{x}_{it}'\left[\boldsymbol{\beta}\left(z_{it}\right)-\widetilde{\boldsymbol{\beta}}_0\left(z_{it}\right)\right] + \left(\mathbf{W}\mathbf{y}\right)_{it}\rho(z_{it})\right)\left(-\mathbf{g}_j'\mathbf{S}_1' + \mathbf{m}_{j,t-1}'S_2'\right) \\ &\quad\times\boldsymbol{\chi}\left(\mathbf{z}_j\right)\left[h^2\upsilon_{1,2}\left(k\right)\mathbf{E}_2\left(\mathbf{z}_j\right) + \left(nh^2\right)^{-1}\mathbf{Q}'\mathbf{K}_h\left(\mathbf{z}_j\right)\Delta\mathbf{u}\right]. \end{aligned}$$

Applying the method similar to that used above, we obtain

$$n\sqrt{h_0}\widehat{T}_{n,2} = O_p\left(n\sqrt{h_0}\left(h^2 + \left(nh^2\right)^{-1/2}\right)\left(\widetilde{h}_0^2 + \left(n\widetilde{h}_0\right)^{-1/2}\right)\right) \text{ under } H_0, \tag{A.36}$$
$$n\sqrt{h_0}\widehat{T}_{n,2} = O_p\left(n\sqrt{h_0}h^2 + h^{-1}\sqrt{nh_0}\right) \text{ under } H_1. \tag{A.37}$$

Lastly, we consider

$$\begin{aligned} n\sqrt{h_0}\widehat{T}_{n,1} &= \frac{1}{n\sqrt{h_0}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}k\left(\frac{z_{it}-z_{js}}{h_0}\right)\mathbf{x}_{it}'\mathbf{x}_{js} \\ &\quad\times\left(\mathbf{x}_{it}'\left[\boldsymbol{\beta}\left(z_{it}\right)-\widetilde{\boldsymbol{\beta}}_0\left(z_{it}\right)\right] + \left(\mathbf{W}\mathbf{y}\right)_{it}\rho(z_{it})\right) \\ &\quad\times\left(\mathbf{x}_{js}'\left[\boldsymbol{\beta}\left(z_{js}\right)-\widetilde{\boldsymbol{\beta}}_0\left(z_{js}\right)\right] + \left(\mathbf{W}\mathbf{y}\right)_{js}\rho(z_{js})\right). \end{aligned}$$

Similar to the proof for $n\sqrt{h_0}\widehat{T}_{n,5}$, under H_0, we obtain $n\sqrt{h_0}\widehat{T}_{n,1} = O_p\left(n\sqrt{h_0}\left(\widetilde{h}_0^4 + \left(n\widetilde{h}_0\right)^{-1}\right)\right)$, and under H_1,

$$n\sqrt{h_0}\widehat{T}_{n,1} \approx \frac{1}{n\sqrt{h_0}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}k\left(\frac{z_{it}-z_{js}}{h_0}\right)\mathbf{x}_{it}'\mathbf{x}_{js}$$

$$\times \left\{ \mathbf{x}_{it}' \left[\boldsymbol{\beta}\left(z_{it}\right) - b\left(z_{it}\right) \right] + \left(\mathbf{W}\mathbf{y}\right)_{it} \rho(z_{it}) \right\}$$
$$\times \left\{ \mathbf{x}_{js}' \left[\boldsymbol{\beta}\left(z_{js}\right) - b\left(z_{js}\right) \right] + \left(\mathbf{W}\mathbf{y}\right)_{js} \rho(z_{js}) \right\}$$
$$= O_p\left(n\sqrt{h_0}\right), \tag{A.38}$$

where $b\left(z_{it}\right)$ is the leading term of $\mathbb{E}\left[\widetilde{\boldsymbol{\beta}}_0\left(z\right)|\mathbf{G},\mathbf{X},\mathbf{Z}\right]$ at point $z = z_{it}$, and denoting $\mu_2\left(z\right) = \mathbb{E}\left[\mathbf{x}_{it}'\left[\boldsymbol{\beta}\left(z_{it}\right) - b\left(z_{it}\right)\right]\mathbf{x}_{it}|z_{it} = z\right]$, we have

$$\frac{1}{n^2 h_0}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} k\left(\frac{z_{it}-z_{js}}{h_0}\right)\mathbf{x}_{it}'\mathbf{x}_{js}\left\{\mathbf{x}_{it}'\left[\boldsymbol{\beta}\left(z_{it}\right)-b\left(z_{it}\right)\right]\mathbf{x}_{js}'\left[\boldsymbol{\beta}\left(z_{js}\right)-b\left(z_{js}\right)\right]\right\}$$
$$= \sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[f\left(z_{js}\right)\mu_2\left(z_{js}\right)'\mu_2\left(z_{js}\right)\right]\left(1+o_p\left(1\right)\right) > 0.$$

Under H_1, we show that $\widehat{T}_{n,1}/\left(n\sqrt{h_0}\right)$ converges to a positive constant.

Next, we consider $\widehat{\sigma}_0^2 = \left(n^2 h\right)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\widehat{\boldsymbol{\varepsilon}}_i' A_{ij}\widehat{\boldsymbol{\varepsilon}}_j)^2$. Under H_0, we can show that

$$\widehat{\sigma}_0^2 \approx \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\Delta\mathbf{u}_i' A_{ij}\Delta\mathbf{u}_j)^2 = 2\mathbb{E}\left[H_n^2\left(\boldsymbol{\chi}_1,\boldsymbol{\chi}_2\right)\right]\left(1+o_p\left(1\right)\right) \overset{p}{\to} \sigma_0^2,$$

while, under H_1, the leading term of $\widehat{\sigma}_0^2$ is $\left(n^2 h\right)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\boldsymbol{\Delta}_i' A_{ij}\boldsymbol{\Delta}_j')^2 = O_p\left(1\right)$ which converges to a positive constant. Combining the above results completes the proof of this theorem. ■

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.

Anselin, L. & Lozano-Gracia, N. (2009). Spatial hedonic models. In K. Patterson & T. C. Mills (Eds.), *Palgrave Handbook of Econometrics, 2*. Basingstoke: Palgrave Macmillan.

Belloni, A., Chernozhukov, V., Chetverikov, D., & Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, *186*, 345–366.

Brueckner, J. K. (2003). Strategic interaction among governments: An overview of empirical studies. *International Regional Science Review*, *26*, 175–188.

Du, P., Parmeter, C. F., & Racine, J. S. (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statistica Sinica*, *23*(3), 1347–1371.

Dzyadyk, V. K. & Shevchuk, I. A. (2008). *Theory of Uniform Approximation of Functions by Polynomials*. New York: Walter de Gruyter & GmbH Co.

Ertur, C. & Koch, W. (2007). Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics*, *22*, 1033–1062.

Fan, Y. & Li, Q. (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics*, *10*, 245–271.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Annals of Statistics*, *14*, 1–16.

Hall, P. & Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. New York: Academic Press.

Hall, P. & Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics*, *29*(3), 624–647.

Henderson, D. J., Carroll, R. J., & Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, *144*(1), 257–275.

Jin, F. & Lee, L.-f. (2015). On the bootstrap for Moran's I test for spatial dependence. *Journal of Econometrics*, *184*, 295–314.

Kelejian, H. H., Murrell, P., & Shepotylo, O. (2013). Spatial spillovers in the development of institutions. *Journal of Development Economics*, *101*, 297–315.

Kelejian, H. H. & Prucha, I. R. (1999). A generalized moment estimator for the autoregressive parameter in a spatial model. *International Economic Review*, *40*, 509–533.

Kelejian, H. H. & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, *157*, 53–67.

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

Lee, L.-F. & Yu, J. (2010a). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, *154*, 165–185.

Lee, L.-F. & Yu, J. (2010b). Some recent developments in spatial panel data models. *Regional Science and Urban Economics*, *40*, 255–271.

Lee, L.-F. & Yu, J. (2012). Spatial panels: Random components versus fixed effects. *International Economic Review*, *53*, 1369–1412.

Lee, L.-F. & Yu, J. (2014). Efficient GMM estimation of spacial dynamic panel data models with fixed effects. *Journal of Econometrics*, *180*, 174–197.

Li, C. & Liang, Z. (2015). Asymptotics for nonparametric and semiparametric fixed effects panel models. *Journal of Econometrics*, *185*, 420–434.

Li, Q. (1999). Consistent model specification tests for time series econometric models. *Journal of Econometrics*, *92*, 101–147.

Li, Q. & Wang, S. (1998). A simple consistent bootstrap test for a parametric regression functional form. *Journal of Econometrics*, *87*, 145–165.

Lin, Z., Li, Q., & Sun, Y. (2014). A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *Journal of Econometrics*, *178*, 167–179.

Liu, X., Lee, L.-f., & Bollinger, C. T. (2010). An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics*, *159*, 303–319.

Malikov, E., Kumbhakar, S., & Sun, Y. (2016). Varying coefficient panel data model in the presence of endogenous selectivity and fixed effects. *Journal of Econometrics*, *190*(2), 233–251.

Malikov, E. & Sun, Y. (2017). Semiparametric estimation and testing of smooth coefficient spatial autoregressive models. *Journal of Econometrics*, *199*(1), 13–34.

Mammen, E., Støve, B., & Tjøstheim, D. (2009). Nonparametric additive models for panels of time series. *Econometric Theory*, *25*, 442–481.

Martins-Filho, C. & Yao, F. (2009). Nonparametric regression estimation with general parametric error covariance. *Journal of Multivariate Analysis*, *100*, 309–333.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis*, *17*, 571–599.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, *79*, 147–168.

Pinkse, J., Slade, M. E., & Brett, C. (2002). Spatial price competition: A semiparametric approach. *Econometrica*, *70*, 1111–1153.

Qian, J. & Wang, L. (2012). Estimating semiparametric panel data models by marginal integration. *Journal of Econometrics*, *167*, 483–493.

Seber, G. A. F. (2008). *A Matrix Handbook for Statisticians*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.

Su, L. (2012). Semiparametric GMM estimation of spatial autoregressive models. *Journal of Econometrics*, *167*, 543–560.

Su, L. & Jin, S. (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics*, *157*, 18–33.

Su, L. & Qu, X. (2017). Specification test for spatial autoregressive models. *Journal of Business and Economic Statistics (online)*. forthcoming.

Su, L., Ullah, A., & Wang, Y. (2013). Nonparametric regression estimation with general parametric error covariance: A more efficient two-step estimator. *Empirical Economics*, *45*, 1009–1024.

Sun, Y. (2016). Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *Journal of Econometrics*, *195*, 134–153.

Sun, Y., Cai, Z., & Li, Q. (2015). A consistent nonparametric test on semiparametric smooth coefficient models with integrated time series. *Econometric Theory*. forthcoming.

Sun, Y., Carroll, R. J., & Li, D. (2009). Semiparametric estimation of fixed-effects panel data varying coefficient models. *Advances in Econometrics*, *25*, 101–129.

Sun, Y. & Malikov, E. (2015). Estimation and inference in functional-coefficient spatial autoregressive panel data models with fixed effects. *Manuscript.*

Sun, Y., Yan, H., Zhang, W., & Lu, Z. (2014). A semiparametric spatial dynamic model. *Annals of Statistics*, *42*, 700–727.

Sun, Y., Zhang, Y., & Li, Q. (2015). Nonparametric panel data regression models. In B. H. Baltagi (Ed.), *The Oxford Handbook on Panel Data* chapter 10. Oxford University Press.

Wang, Q. & Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, *40*, 727–758.

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

Zhang, Z. (2013). A pairwise difference estimator for partially linear spatial autoregressive models. *Spatial Economic Analysis*, *8*, 176–194.

Zheng, J. X. (1996). A consistent test of function form via nonparametric estimation techniques. *Journal of Econometrics*, *75*, 263–289.