1-2016

Semiparametric estimation of partially linear dynamic panel data models with fixed effects

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Yonghui ZHANG
Singapore Management University, yhzhang@smu.edu.sg

Published in Essays in honor of Aman Ullah, Emerald, 2016, 137-204.
https://doi.org/10.1108/S0731-905320160000036014

Semiparametric Estimation of Partially Linear Dynamic Panel Data Models with Fixed Effects*

Liangjun Su[a] and Yonghui Zhang[b]
[a] School of Economics, Singapore Management University
[b] School of Economics, Renmin University of China

September 12, 2015

Abstract

In this paper, we study a partially linear dynamic panel data model with fixed effects, where either exogenous or endogenous variables or both enter the linear part, and the lagged dependent variable together with some other exogenous variables enter the nonparametric part. Two types of estimation methods are proposed for the first-differenced model. One is composed of a semiparametric GMM estimator for the finite dimensional parameter θ and a local polynomial estimator for the infinite dimensional parameter m based on the empirical solutions to Fredholm integral equations of the second kind, and the other is a sieve IV estimate of the parametric and nonparametric components jointly. We study the asymptotic properties for these two types of estimates when the number of individuals N tends to ∞ and the time period T is fixed. We also propose a specification test for the linearity of the nonparametric component based on a weighted square distance between the parametric estimate under the linear restriction and the semiparametric estimate under the alternative. Monte Carlo simulations suggest that the proposed estimators and tests perform well in finite samples. We apply the model to study the relationship between intellectual property right (IPR) protection and economic growth, and find that IPR has a nonlinear positive effect on the economic growth rate.

Key words: Additive structure, Fredholm integral equation, Generated covariate, GMM, Local polynomial regression, Partially linear model, Sieve method, Time effect

JEL Classification: C14, C33, C36.

Running head: Semiparametric Estimation of Partially Linear Dynamic Panel Data Models

*Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; E-mail: ljsu@smu.edu.sg, Phone: +65 6828 0386.

1 Introduction

Recently nonparametric panel models have received a lot of attention. The increasing popularity of nonparametric approach to panel data modelling largely comes from its flexibility in exploring hidden structures and its robustness to model misspecification which often occurs under a parametric setting. On the other hand, a fully nonparametric model usually suffers from the notorious problem of "curse of dimensionality" when the dimension of the nonparametric covariates is high, and it becomes very difficult to interpret a nonparametric estimator in empirical applications when the dimension is larger than two. To overcome these shortcomings, many semiparametric models have also been proposed in the panel data literature as a compromise between nonparametric and parametric specifications. By imposing different structures on the unknown functions, various semiparametric models such as additive models, partially linear models, single-index models, transformation models, and varying-coefficient models have been studied extensively in the literature. For example, Chen et al. (2012) consider partially linear panel data models where the time trend enters the nonparametric component; Chen et al. (2013a; 2013b) and Dong et al. (2014) study the estimation of panel data models with different single-index structures; Dong et al. (2015) consider a partially linear panel data model with cross-sectional dependence and non–stationarity; Feng et al. (2015) consider the estimation of varying-coefficient panel data models. For recent selective overviews, see Su and Ullah (2011), Chen et al. (2013), and Sun et al. (2015).

In this paper, we focus on the following partially linear dynamic panel data model with fixed effects

$$Y_{it} = Z_{it}'\theta_0 + m\left(Y_{i,t-1}, X_{it}\right) + \alpha_i + \varepsilon_{it}, \tag{1.1}$$

$i = 1, ..., N$, $t = 1, ..., T$, where Y_{it} is the scalar dependent variable for individual i at time period t, Z_{it} is a $d_z \times 1$ vector of regressors that enter the linear component of the model, θ_0 is an unknown parameter that takes value on the compact parameter space $\Theta \subset \mathbb{R}^{d_z}$, $m(\cdot)$ is an unknown smooth function defined on $\mathbb{R}^{d_x+1}$, X_{it} is a $d_x \times 1$ vector of regressors that enter the nonparametric component $m(\cdot)$ together with the lagged dependent variable $Y_{i,t-1}$, α_i's are unobserved individual effects, and ε_{it}'s are idiosyncratic error terms. The subscript "0" in θ_0 indicates the true parameter value. Clearly, $Y_{i,t-1}$ is correlated with the fixed effect α_i. We also allow Z_{it} and X_{it} to be correlated with α_i. As in Baltagi and D. Li (2002), Baltagi and Q. Li (2002), and Yao and Zhang (2015), we allow Z_{it} or a subset of Z_{it} to be endogenous and assume the existence of a $d_v \times 1$ vector of instrumental variables (IVs) V_{it} where $d_v \geq d_z$.[1] Also like them, we restrict X_{it} to be exogenous to avoid the ill-posed inverse problem. The latter problem can be addressed by extending the estimation procedure of Ai and Chen (2003), Chen and Pouzo (2012), or Florens et al. (2012) to the panel setting but is certainly beyond the scope of the current paper. We are interested in the estimation of θ_0 and $m(\cdot)$ under large N and small T.

Since Engle et al. (1986) and Robinson (1988), partially linear models have been widely studied and applied in the econometrics literature. In the panel framework, partially linear structures have also attracted a lot of attention. For example, Li and Stengos (1996) and Li and Ullah (1998) consider kernel estimation of partially linear panel data models with random effects where the endogenous variables appear in the linear component; Baltagi and D. Li (2002) and Baltagi and Q. Li (2002) respectively propose series and kernel estimation of partially linear dynamic panel models with fixed effects where the

[1] When some variables in Z_{it} are exogenous, we choose themselves as their IVs.

lagged dependent variables enter the model linearly; Su and Ullah (2006a) study profile likelihood kernel estimation of partially linear *static* panel data models with fixed effects; Baglan (2010) considers series estimation of partially linear dynamic panel data models with fixed effects where the lagged dependent variable enters the model linearly; Qian and Wang (2012) consider kernel estimation of nonparametric component in a fixed-effect partially linear *static* panel data model via marginal integration. To the best of our knowledge, so far endogeneity and lagged dependent variables have been allowed to enter the partially linear panel data models only through the linear component. It remains unclear whether one can allow both to enter the nonparametric component of the model. In this paper, our goal is less ambitious in the sense that we only allow lagged dependent variables to enter the nonparametric component. The endogenous variables, if present, only enter the linear component of the model. Despite this less ambitious goal, the inclusion of lagged dependent variables in the nonparametric component turns out to be sufficiently technically challenging and the introduction of endogeneity in the linear component further complicates the issue to a great deal.

Despite the existence of a large literature on nonparametric panel data models, there is a lack of satisfactory development in nonparametric and semiparametric panel data models where a lagged dependent variable enters the nonparametric component of the model. The few exceptions include Su and Lu (2013) and Lee (2014), who consider kernel and sieve estimation of nonparametric dynamic panel data models, respectively. Su and Lu (2013) also document the empirical evidence of nonlinear effects of lagged dependent variables in the economic growth literature. This paper contributes to the literature by allowing lagged dependent variables to enter the nonparametric component in panel data models.

For notational simplicity, let $\underline{\xi}_{it} \equiv \left(\xi_{it}, \xi_{i,t-1}, ..., \xi_{i1}\right)$, where $\xi = Y, X$, Z, or V. Following Su and Lu (2013), we assume that $E(\varepsilon_{it}|\underline{Y}_{i,t-1}, \underline{X}_{it}, \underline{V}_{it}) = 0$ and consider the first-differenced model

$$\Delta Y_{it} = \Delta Z_{it}'\theta_0 + m\left(Y_{i,t-1}, X_{it}\right) - m\left(Y_{i,t-2}, X_{i,t-1}\right) + \Delta\varepsilon_{it}, \tag{1.2}$$

where, e.g., $\Delta Y_{it} \equiv Y_{it} - Y_{i,t-1}$. Clearly, the first-differenced model in (1.2) exhibits some important features. First, (1.2) still has a partially linear structure with a linear component ($\Delta Z_{it}'\theta_0$) and two additive nonparametric components ($m\left(Y_{i,t-1}, X_{it}\right)$ and $m\left(Y_{i,t-2}, X_{i,t-1}\right)$). Second, the two additive nonparametric components share the same functional form which should be incorporated in the estimation procedure. In fact, as we show below, this feature, in conjunction with the martingale difference sequence (m.d.s.) type of conditions on ε_{it}, implies that $m\left(\cdot\right)$ implicitly solves a class of Fredholm integral equations of the second kind indexed by the finite dimensional parameter θ. Third, the error term $\Delta\varepsilon_{it}$ follows a *non-invertible* moving average process of order 1 (MA(1)) and is correlated with both $Y_{i,t-1}$ and ΔZ_{it} in general. Despite the fact that it is hard to apply this last feature to improve efficiency as in Xiao et al. (2003), Su and Ullah (2006b), and Gao et al. (2006), it invalidates the traditional kernel estimation based on either marginal integration or backfitting method because of the endogeneity issue in the first-differenced model. In addition, the presence of the linear component ΔZ_{it} in (1.2) makes our model different from that in Su and Lu (2013) and complicates our estimation procedures and asymptotic analysis substantially.

In this paper, we propose two estimation procedures, both of which take into account all the features mentioned above. The first one comprises the semiparametric GMM estimation of θ_0 and kernel estimation of $m\left(\cdot\right)$ based on the empirical solutions to Fredholm integral equations of the second kind. The second one estimates both the parametric and nonparametric component at a single step via the sieve IV method. As remarked later on, our methods can be used to estimate models with both individual and time fixed effects.

Under some fairly general and mild conditions that allow for the nonstationarity of $(Y_{i,t-1}, Z_{it}, V_{it}, X_{it}, \varepsilon_{it})$ along the time dimension and conditional heteroskedasticity among ε_{it}, we show that both types of estimates of the finite dimensional parameter have the usual parametric convergence rates and follow asymptotic normal distributions. We also derive the uniform convergence rates for the nonparametric estimates over a compact support and establish their asymptotic normal distributions by passing N to infinity and keeping T fixed as in typical micro panel data models. Based on our estimators, we also propose a nonparametric test for the linearity of the nonparametric component. We examine the finite sample performance of our estimators and tests through Monte Carlo simulations. We apply the proposed model to study the relationship between economic growth, its lagged value, and intellectual property right (IPR) protection. We find substantial nonlinearity in the relationship between a country's economic growth rate and its lagged value and a positive nonlinear relationship between economic growth rate and IPR protection.

The rest of the paper is organized as follows. In Section 2 we introduce the semiparametric GMM estimation of θ and kernel estimation of m based on a Fredholm integral equation of the second kind and study the asymptotic properties of these estimators. In Section 3 we discuss the sieve IV estimation of θ and m and derive the asymptotic properties of our proposed estimators. In Section 4, we propose a nonparametric test for the linearity of the nonparametric component. In Section 5 we conduct a small set of Monte Carlo simulations to evaluate the finite sample performance of our estimators and tests. We apply our method to a real data set in Section 6. Final remarks are contained in Section 7. All technical details are relegated to the Appendix.

Throughout the paper, we restrict our attention to a balanced panel. We use $i = 1, \ldots, N$ to denote individuals and $t = 1, \ldots, T$ to denote time. All asymptotic theories are established by passing N to infinity and holding T as a fixed constant. For natural numbers a and b, we use $\mathbf{I}_a$ to denote an $a \times a$ identity matrix, $\mathbf{0}_{a\times b}$ an $a \times b$ matrix of zeros, and l_a an $a \times 1$ vector of ones. Let $T_l \equiv T - l$ for $l = 1, 2$. For conformable vectors u and v, we use u/v to denote element-wise division. For a real matrix B, let $\|B\| \equiv \sqrt{\mathrm{tr}\,(B'B)}$ denote its Frobenius norm and $\|B\|_{\mathrm{sp}} \equiv \sqrt{\lambda_{\max}\,(B'B)}$ its spectral norm where $\lambda_{\max}\,(\cdot)$ is the largest eigenvalue of "$\cdot$". Let $\mathbf{P}_B \equiv B(B'B)^{-}B$ and $\mathbf{M}_B \equiv \mathbf{I} - \mathbf{P}_B$ where $(\cdot)^{-}$ denotes the Moore-Penrose generalized inverse and $\mathbf{I}$ is a conformable identity matrix. Let $\overset{P}{\to}$ and $\overset{D}{\to}$ signify convergence in probability and distribution, respectively.

2 Semiparametric GMM estimation of θ and kernel estimation of m

In this section, we first outline the idea for the semiparametric GMM estimation of θ and kernel estimation of $m\,(\cdot)$ based on the empirical solution to a Fredholm integral equation of the second kind and then present the estimation details. We also derive the convergence rates for the proposed estimators and establish their asymptotic normal distributions.

2.1 Basic idea

Note that $E(\varepsilon_{it}|\underline{Y}_{i,t-1},\underline{X}_{it},\underline{V}_{it})=0$. By the law of iterated expectations, we have $E(\varepsilon_{it}|\underline{Y}_{i,t-1},\underline{X}_{it})$ $=0$. Based on this observation, we obtain the following conditional moment conditions

$$E\left[\Delta Y_{it}-\Delta Z_{it}'\theta_0-m\left(Y_{i,t-1},X_{it}\right)+m\left(Y_{i,t-2},X_{i,t-1}\right)|\underline{U}_{i,t-2}\right]=0, \tag{2.1}$$

$i=1,...,N$, $t=3,...,T$, where $\underline{U}_{i,t-2}\equiv(U_{i,t-2},U_{i,t-3},...,U_{i1})$ and $U_{it}\equiv\left(Y_{it},X_{i,t+1}'\right)'$. Clearly, for large t the conditioning information set $\underline{U}_{i,t-2}$ contains a large number of valid IVs for the local nonparametric identification of $m(\cdot)$. But for both technical reasons and practical concerns, it seems unrealistic to use all variables in $\underline{U}_{i,t-2}$ in nonparametric regressions. To avoid the curse of dimensionality problem, we consider only a small number of IVs that are measurable with respect to the σ-algebra generated by $\underline{U}_{i,t-2}$. In this paper, we only use $U_{i,t-2}$ and leave the optimal choice of IVs for the estimation of the nonparametric component as future research.

To proceed, we define some notation. Let $\mathcal{U}$ be a compact subset of $\mathbb{R}^{d_x+1}$.[2] We assume that $U_{i,t-2}$ has a positive density on $\mathcal{U}$ and denote the conditional probability density function (PDF) of $U_{i,t-1}$ given that $U_{i,t-2}$ lies in $\mathcal{U}$ as $f_{t-2}(\cdot)$. Similarly, we use $f_{t-1|t-2}(\cdot|\cdot)$ to denote the conditional PDF of $U_{i,t-1}$ given $U_{i,t-2}$, conditionally on $U_{i,t-2}\in\mathcal{U}$. Let

$$n\equiv\sum_{i=1}^{N}\sum_{t=3}^{T}\mathbf{1}\left(U_{i,t-2}\in\mathcal{U}\right)\text{ and }n_{t-2}\equiv\sum_{i=1}^{N}\mathbf{1}\left(U_{i,t-2}\in\mathcal{U}\right)\text{ for }t\in\{3,...,T\},$$

where $\mathbf{1}(\cdot)$ is the usual indicator function. By the Kinchin law of large numbers (LLN), $n_{t-2}/N\overset{P}{\to}$ $p_{t-2}\equiv P\left(U_{i,t-2}\in\mathcal{U}\right)$ and $n/N\overset{P}{\to}\sum_{t=3}^{T}p_{t-2}\equiv p$. Let $r_{t|t-2}^{[y]}(u)\equiv E(-\Delta Y_{it}|U_{i,t-2}=u)$ and $r_{t|t-2}^{[z]}(u)\equiv$ $E\left(-\Delta Z_{it}|U_{i,t-2}=u\right)$. Define

$$f(u)\equiv\sum_{t=3}^{T}\frac{p_{t-2}}{p}f_{t-2}(u),\ f(\bar{u}|u)\equiv\sum_{t=3}^{T}\frac{p_{t-2}}{p}f_{t-1|t-2}(\bar{u}|u), \tag{2.2}$$

$$r_y(u)\equiv\sum_{t=3}^{T}\frac{p_{t-2}}{p}r_{t|t-2}^{[y]}(u),\ r_z(u)\equiv\sum_{t=3}^{T}\frac{p_{t-2}}{p}r_{t|t-2}^{[z]}(u)\text{ and }r_\theta(u)\equiv r_y(u)-\theta'r_z(u) \tag{2.3}$$

where we suppress the dependence of p, $f(u)$, $f(\bar{u}|u)$, $r_y(u)$, $r_z(u)$ and $r_\theta(u)$ on T.

(2.1) implies that

$$\begin{aligned}m(u)&=E\left[m\left(U_{i,t-1}\right)|U_{i,t-2}=u\right]+E\left(-\Delta Y_{it}|U_{i,t-2}=u\right)-\theta_0'E\left(-\Delta Z_{it}|U_{i,t-2}=u\right),\\&=\int m(\bar{u})f_{t-1|t-2}(\bar{u}|u)\,d\bar{u}+r_{t|t-2}^{[y]}(u)-\theta_0'r_{t|t-2}^{[z]}(u)\text{ for }t=3,...,T.\end{aligned} \tag{2.4}$$

Multiplying both sides of (2.4) by p_{t-2}/p and summing up over $t=3,...,T$ yields

$$m(u)=\int m(\bar{u})f(\bar{u}|u)\,d\bar{u}+\left[r_y(u)-\theta_0'r_z(u)\right]. \tag{2.5}$$

Under certain regularity conditions, for any $\theta\in\Theta$ there exists a unique solution $m_\theta(u)$ to a Fredholm integral equation of the second kind in an infinite dimensional Hilbert space $\mathcal{L}_2(f)$:

$$m_\theta=\mathcal{A}m_\theta+r_\theta, \tag{2.6}$$

[2] The reason to introduce $\mathcal{U}$ is to handle the non-compact support of $U_{i,t-2}$. If one is willing to assume that $U_{i,t-2}$ has compact support, then one can take $\mathcal{U}$ as the support of $U_{i,t-2}$.

where $\mathcal{A}$: $\mathcal{L}_2(f) \to \mathcal{L}_2(f)$ is a bounded linear operator defined by

$$\mathcal{A}m(u) \equiv \int m(\bar{u}) f(\bar{u}|u) d\bar{u} \text{ for } u \in \mathcal{U} \tag{2.7}$$

and $\mathcal{L}_2(f)$ is a Hilbert space with norm $\|m\|_2 \equiv [\int_{\mathcal{U}} m(u)^2 f(u) du]^{1/2}$. Let $\mathcal{I}$ be the identity operator. Under some conditions to be specified later on, $\mathcal{I} - \mathcal{A}$ is invertible and

$$\sup_{\|m\|_2 \leq 1} \left\| (\mathcal{I} - \mathcal{A})^{-1} m \right\|_2 < \infty. \tag{2.8}$$

Then given θ, the unique solution to (2.6) is given by

$$m_\theta = (\mathcal{I} - \mathcal{A})^{-1} \left(r_y - \theta' r_z \right) = m_y - \theta' m_z, \tag{2.9}$$

where $m_y \equiv (\mathcal{I} - \mathcal{A})^{-1} r_y$ and $m_z \equiv (\mathcal{I} - \mathcal{A})^{-1} r_z$ are solutions to

$$m(u) = \mathcal{A}m(u) + r_y(u) \text{ and } m(u) = \mathcal{A}m(u) + r_z(u), \text{ respectively.} \tag{2.10}$$

To facilitate the theoretical study, we will consider the following auxiliary first-differenced models

$$\begin{aligned} \Delta Y_{it} &= m_y(U_{i,t-1}) - m_y(U_{i,t-2}) + \eta_{y,it}, \\ \Delta Z_{it,l} &= m_{z_l}(U_{i,t-1}) - m_{z_l}(U_{i,t-2}) + \eta_{z_l,it}, \ l = 1, ..., d_z, \end{aligned} \tag{2.11}$$

where $Z_{it,l}$ and z_l denote the l-th element in Z_{it} and z, respectively, the error terms $\eta_{y,it}$ and $\eta_{z_l,it}$ satisfy $E\left(\eta_{y,it}|U_{i,t-2}\right) = 0$ and $E\left(\eta_{z_l,it}|U_{i,t-2}\right) = 0$. It is easy to see that under these moment conditions, m_y and $m_z \equiv (m_{z_1}, ..., m_{z_{d_z}})'$ are the solutions to the Fredholm integral equations of the second kind in (2.10).

Apparently, m_θ is linear in θ. Given the estimates $\hat{m}_y(u)$ and $\hat{m}_z(u)$ of $m_y(u)$ and $m_z(u)$, we can estimate $m_\theta(u)$ by $\hat{m}_\theta(u) \equiv \hat{m}_y(u) - \theta' \hat{m}_z(u)$ for any $\theta \in \Theta$. Below we first introduce how one can obtain the kernel estimates $\hat{m}_y(u)$ and $\hat{m}_z(u)$ based on local polynomial regressions, and then study the semiparametric GMM estimation of θ based on some identification conditions.

2.2 Semiparametric GMM Estimation

2.2.1 Kernel estimation of $m_y(u)$ and $m_z(u)$

Now we consider how to estimate $m_y(u)$ and $m_z(u)$ as the solutions to the two equations in (2.10). Assume that nonparametric estimates of r_y, r_z and $\mathcal{A}$ are given by $\hat{r}_y$, $\hat{r}_z$ and $\widehat{\mathcal{A}}$, respectively. Then the plug-in estimators $\hat{m}_y$ and $\hat{m}_z$ are given by the solutions to

$$\hat{m}_y = \widehat{\mathcal{A}} \hat{m}_y + \hat{r}_y \text{ and } \hat{m}_z = \widehat{\mathcal{A}} \hat{m}_z + \hat{r}_z, \tag{2.12}$$

respectively.

Here we consider the local polynomial estimates of r_y, r_z and $\mathcal{A}$. Let $u \equiv (y, x')' \equiv (u_0, u_1, ..., u_{d_x})'$ be a $(d_x + 1) \times 1$ vector, where x is $d_x \times 1$ and y is a scalar. Let $\boldsymbol{j} \equiv (j_0, j_1, ..., j_{d_x})'$ be a $(d_x + 1)$-vector of non-negative integers. Following Masry (1996), we adopt the notation: $u^{\boldsymbol{j}} \equiv \boldsymbol{\Pi}_{i=0}^{d_x} u_i^{j_i}$, $\boldsymbol{j}! \equiv \boldsymbol{\Pi}_{i=1}^{d_x} j_i!$,

$|\boldsymbol{j}| \equiv \sum_{i=0}^{d_x} j_i$, and $\sum_{0\leq|\boldsymbol{j}|\leq q} \equiv \sum_{k=0}^{q} \sum_{\substack{j_0=0 \\ j_0+j_1+\ldots+j_{d_x}=k}}^{k} \cdots \sum_{j_{d_x}=0}^{k}$. Let $g_l \equiv (l+d_x)!/(l!d_x!)$ be the number of distinct (d_x+1)-tuples $\boldsymbol{j}$ with $|\boldsymbol{j}|=l$. Let $G \equiv \sum_{l=0}^{q} g_l$. Let $\mu_h(\cdot) = \mu(\cdot/h)$, where μ is a stacking function such that $\mu_h(U_{i,t-2}-u)$ denotes a $G\times 1$ vector that stacks $((U_{i,t-2}-u)/h)^{\boldsymbol{j}}$, $0\leq|\boldsymbol{j}|\leq q$, in lexicographic order (e.g., $\mu_h(u) = (1,(u/h)')'$ when $q=1$).

To estimate $r_y(u)$, we consider the following minimization problem:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N}\sum_{t=3}^{T}\left[-\Delta Y_{it} - \sum_{0\leq|\boldsymbol{j}|\leq q} \beta_{\boldsymbol{j}}' \left((U_{i,t-2}-u)/h\right)^{\mathbf{j}}\right]^2 K_h(U_{i,t-2}-u)\,\mathbf{1}(U_{i,t-2}\in\mathcal{U}), \tag{2.13}$$

where $\boldsymbol{\beta}$ stacks the $\beta_{\boldsymbol{j}}$'s ($0\leq|\boldsymbol{j}|\leq q$) in lexicographic order, $K_h(u) = h_0^{-1}k(y/h_0)\Pi_{j=1}^{d_x} h_j^{-1}k(x_j/h_j)$ for $u\equiv(y,x')'$, k is a univariate PDF, and $h=(h_0,h_1,...,h_{d_x})'$ is a bandwidth sequence that shrinks to zero as $N\to\infty$. Note that in (2.13) we use an indicator function $\mathbf{1}(\cdot)$ to handle the non-compact support of $U_{i,t-2}$. Like Mammen et al. (1999), Mammen et al. (2009), and Su and Lu (2013), our estimators only use observations in the estimation of $r_y(u)$ if the covariates $U_{i,t-2}$ lie in a compact set $\mathcal{U}$ on $\mathbb{R}^{d_x+1}$. This device greatly facilitates the asymptotic analysis of our estimator in the case of infinite support for $U_{i,t-2}$; it is not needed if $U_{i,t-2}$ is compactly supported.

Let $\hat{r}_y(u)$ denotes the first element of the solution to the above minimization problem. Then

$$\hat{r}_y(u) = -e_1'[S_{NT}(u)]^{-1}\frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathbf{1}_{it}K_h(U_{i,t-2}-u)\,\mu_h(U_{i,t-2}-u)\,\Delta Y_{it} = \frac{-1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathcal{K}_{it}(u)\,\Delta Y_{it}$$

where

$$S_{NT}(u) \equiv \frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathbf{1}_{it}K_h(U_{i,t-2}-u)\,\mu_h(U_{i,t-2}-u)\,\mu_h(U_{i,t-2}-u)', \tag{2.14}$$

$$\mathcal{K}_{it}(u) \equiv \mathbf{1}_{it}e_1'S_{NT}(u)^{-1}\mu_h(U_{i,t-2}-u)\,K_h(U_{i,t-2}-u). \tag{2.15}$$

$e_1\equiv(1,0,...,0)'$ is a $G\times 1$ vector with 1 in the first position and zeros elsewhere, and $\mathbf{1}_{it}\equiv\mathbf{1}(U_{i,t-2}\in\mathcal{U})$. Similarly, we can estimate $\mathcal{A}m_y(u)$ by $\widehat{\mathcal{A}}m_y(u) = n^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathcal{K}_{it}(u)\,m_y(U_{i,t-1})$.

In terms of numerical algorithm, if (2.8) is satisfied, it is well known that the first part of (2.10) implies that $m_y = (\mathcal{I}-\mathcal{A})^{-1}r_y = \sum_{j=0}^{\infty}\mathcal{A}^j r_y$. Rust (2000) discusses several methods to solve an integral equation, including both iterative and non-iterative methods; see also Linton and Mammen (2005) and Darolles et al. (2011) for related discussions. The iterative method relies on the observation (e.g., Theorem 2.10 in Kress (1999)) that the sequence of approximations

$$m_y^{(l)} = \mathcal{A}m_y^{(l-1)} + r_y,\ \ l=1,2,..., \tag{2.16}$$

is ultimately close to the truth from any starting point $m_y^{(0)}$. As in Su and Lu (2013), the initial estimator can be constructed based on the sieve IV method detailed in Section 3. If $\widehat{\mathcal{A}}$ and $\hat{r}_y$ are sufficiently close to $\mathcal{A}$ and r_y respectively, then

$$\hat{m}_y^{(l)} = \widehat{\mathcal{A}}\hat{m}_y^{(l-1)} + \hat{r}_y,\ \ l=1,2,... \tag{2.17}$$

is close to m_y. The non-iterative method involves solving a linear system of equations. Using the local polynomial estimates to replace the unknown conditional expectations in (2.10) yields

$$\hat{m}_y(u) - \frac{1}{n}\sum_{j=1}^{N}\sum_{s=3}^{T}\mathcal{K}_{js}(u)\,\hat{m}_y(U_{j,s-1}) = -\frac{1}{n}\sum_{j=1}^{N}\sum_{s=3}^{T}\mathcal{K}_{js}(u)\,\Delta Y_{js}. \tag{2.18}$$

Evaluating (2.18) at $u = U_{i,t-1}$, $i = 1,...,N$, $t = 2,...,T$, yields the following linear system of equations with NT_2 equations and NT_2 unknowns:

$$\hat{\mathcal{M}}_y - \mathcal{K}\hat{\mathcal{M}}_y = -\mathcal{K}\mathcal{Y}, \tag{2.19}$$

where $\hat{\mathcal{M}}_y \equiv [\hat{m}_y(U_{12}),...,\hat{m}_y(U_{1,T-1}),...,\hat{m}_y(U_{N2}),...,\hat{m}_y(U_{N,T-1})]'$, $\mathcal{Y} \equiv (\Delta Y_{13},...,\Delta Y_{1T},...,\Delta Y_{N3},...,\Delta Y_{NT})'$, $\mathcal{Z} \equiv [\Delta Z_{13}',...,\Delta Z_{1T}',...,\Delta Z_{N3}',...\Delta Z_{NT}']'$ and

$$\mathcal{K} \equiv \frac{1}{n}\begin{bmatrix} \mathcal{K}_{13}(U_{12}) & \cdots & \mathcal{K}_{NT}(U_{12}) \\ \vdots & \ddots & \vdots \\ \mathcal{K}_{13}(U_{N,T-1}) & \cdots & \mathcal{K}_{NT}(U_{N,T-1}) \end{bmatrix}.$$

The solution to the above linear system of equations is given by $\hat{\mathcal{M}}_y = -(\mathbf{I}_{NT_2} - \mathcal{K})^{-}\mathcal{K}\mathcal{Y}$. Then we can obtain $\hat{m}_y(u)$ based on (2.18) for any $u \in \mathcal{U}$. The iterative and non-iterative estimators are asymptotically equivalent. Nevertheless, the non-iterative estimator involves the inversion of an $NT_2 \times NT_2$ matrix which may not be stable if NT_2 is large, say, $NT_2 \geq 1000$. For further discussions, see Section 3.3 in Linton and Mammen (2005).

By the same algorithms (iterative and noniterative), we can obtain the estimate $\hat{m}_z \equiv (\hat{m}_{z_1},...,\hat{m}_{z_{d_z}})$ of $m_z \equiv (m_{z_1},...,m_{z_{d_z}})$. Then we have the $\hat{m}_\theta \equiv \hat{m}_y - \theta'\hat{m}_z$ as the estimate of $m_\theta = m_y - \theta' m_z$ for any given $\theta \in \Theta$.

2.2.2 Semiparametric GMM estimation of θ

Now we turn to the estimation of θ. Let $\Delta\varepsilon_{it}(\theta) = \Delta Y_{it} - \Delta Z_{it}'\theta - [m_\theta(U_{i,t-1}) - m_\theta(U_{i,t-2})]$. Then by (2.9),

$$\begin{aligned} \Delta\varepsilon_{it}(\theta) &= [\Delta Y_{it} - m_y(U_{i,t-1}) + m_y(U_{i,t-2})] - \theta'[\Delta Z_{it} - m_z(U_{i,t-1}) + m_z(U_{i,t-2})] \\ &= \eta_{y,it} - \theta'\eta_{z,it}, \end{aligned} \tag{2.20}$$

where $\eta_{y,it} \equiv \Delta Y_{it} - \Delta m_{y,it}$, $\eta_{z,it} \equiv \Delta Z_{it} - \Delta m_{z,it}$, $\Delta m_{a,it} = m_{a,it} - m_{a,i,t-1}$, and $m_{a,it} \equiv m_a(U_{i,t-1})$ for $a = y$, z, and θ. Note that $\Delta\varepsilon_{it} = \Delta\varepsilon_{it}(\theta_0)$. In principle we can obtain an infeasible estimate of θ by considering the regression of $\eta_{y,it}$ on $\eta_{z,it}$. Difficulty arises because of the endogeneity issue. To see this, observe that

$$\begin{aligned} E\left(\Delta\varepsilon_{it}\eta_{z,it}\right) &= E(\varepsilon_{it}Z_{it}) + E(\varepsilon_{i,t-1}Z_{i,t-1}) - E(\varepsilon_{i,t-1}Z_{it}) + E[\varepsilon_{i,t-1}m_z(U_{i,t-1})] \\ &\quad - E(\varepsilon_{it}Z_{i,t-1}) - E[\varepsilon_{it}m_z(U_{i,t-2})] + E[\varepsilon_{i,t-1}m_z(U_{i,t-2})] - E[\varepsilon_{it}m_z(U_{i,t-1})] \\ &= E(\varepsilon_{it}Z_{it}) + E(\varepsilon_{i,t-1}Z_{i,t-1}) - E(\varepsilon_{i,t-1}Z_{it}) - E(\varepsilon_{it}Z_{i,t-1}) \\ &\quad + E[\varepsilon_{i,t-1}m_z(U_{i,t-1})], \end{aligned}$$

where we use the fact that $E\left[\varepsilon_{it} m_z\left(U_{i,t-s}\right)\right] = 0$ for any $s \geq 1$ under the m.d.s. condition that $E(\varepsilon_{it}|\underline{Y}_{i,t-1}, \underline{X}_{it}, \underline{V}_{it}) = 0$. Note that each of the five terms in the last expression is typically nonvanishing if Z_{it} contains endogenous regressors. Even if Z_{it} is strictly exogenous, we have $E(\Delta\varepsilon_{it}\eta_{z,it}) = E\left[\varepsilon_{i,t-1} m_z\left(U_{i,t-1}\right)\right] \neq 0$ because of the dynamic nature of the model. Thus $E(\Delta\varepsilon_{it}\eta_{z,it}) \neq 0$ in general and we need to find a $d_w \times 1$ vector W_{it} of IVs to consistently estimate θ, where $d_w \geq d_z$. We need W_{it} to be orthogonal to $\Delta\varepsilon_{it}$ and correlated with $\eta_{z,it}$. Given the m.d.s. assumption, the set of valid IVs could be very large, which is particularly true if T is large. But it seems extremely difficult to address the optimal choice of IVs here. Given the IV V_{it} for Z_{it} such that $E\left(V_{it}\varepsilon_{it}\right) = 0$ and $E\left(V_{it}Z_{it}'\right) \neq 0$, in this paper we simply recommend choosing $W_{it} \equiv \left(V_{i,t-1}', U_{i,t-2}'\right)'$ as the IVs for $\eta_{z,it}$ and leave the efficient choice of IVs for future research.[3] Note that

$$E\left(\Delta\varepsilon_{it} W_{it}\right) = 0. \tag{2.21}$$

Let $\varsigma_{it} = (\Delta Y_{it}, \Delta Z_{it}', U_{i,t-1}', U_{i,t-2}', W_{it}')'$ and $\hat{\xi}(\cdot) = (\hat{m}_y(\cdot), \hat{m}_z(\cdot))$. Let $\hat{\eta}_{y,it} = \Delta Y_{it} - \hat{m}_{y,it} + \hat{m}_{y,i,t-1}$ and $\hat{\eta}_{z,it} = \Delta Z_{it} - \hat{m}_{z,it} + \hat{m}_{z,i,t-1}$, where $\hat{m}_{a,it} = \hat{m}_a\left(U_{i,t-1}\right)$ for $a = y$ and z. Let $\tilde{\mathbf{1}}_{it} = \mathbf{1}\left(U_{i,t-1} \in \mathcal{U}\right)\mathbf{1}\left(U_{i,t-2} \in \mathcal{U}\right)$ and $\tilde{W}_{it} = \tilde{\mathbf{1}}_{it} W_{it}$. By (2.20) and (2.21), we define the semiparametric GMM estimator $\hat{\theta}_{gmm}$ of θ_0 as the solution

$$\hat{\theta}_{gmm} = \arg\min_{\theta\in\Theta} \left\| Q_{NT}\left(\theta, \hat{\xi}\right)\right\|_{A_{NT}}$$

where $\|b\|_{A_{NT}} = b' A_{NT} b$ for a $d_w \times 1$ vector b, A_{NT} is a $d_w \times d_w$ matrix that is symmetric and asymptotically positive definite (p.d.), $Q_{NT}(\theta, \xi) = \frac{1}{\tilde{n}} \sum_{i=1}^N \sum_{t=3}^T q(\varsigma_{it}, \theta, \xi)$, $\xi(\cdot) = (m_y(\cdot), m_{z_1}(\cdot), ..., m_{z_{d_z}}(\cdot))$, $\tilde{n} = \sum_{i=1}^N \sum_{t=3}^T \tilde{\mathbf{1}}_{it}$, and

$$\begin{aligned} q\left(\varsigma_{it}, \theta, \xi\right) &= \left(\eta_{y,it} - \theta'\eta_{z,it}\right)\tilde{W}_{it} \\ &= \left\{\left[\Delta Y_{it} - m_y\left(U_{i,t-1}\right) + m_y\left(U_{i,t-2}\right)\right] - \theta'\left[\Delta Z_{it} - m_z\left(U_{i,t-1}\right) + m_z\left(U_{i,t-2}\right)\right)\right]\right\}\tilde{\mathbf{1}}_{it} W_{it}. \end{aligned}$$

Note that $q(\varsigma_{it}, \theta, \hat{\xi}) = \left(\hat{\eta}_{y,it} - \theta'\hat{\eta}_{z,it}\right)\tilde{W}_{it}$ and we have restricted our attention to those observations with $U_{i,t-1} \in \mathcal{U}$ and $U_{i,t-2} \in \mathcal{U}$ in the above procedure for some technical reasons. Apparently

$$\hat{\theta}_{gmm} = \left(\hat{\boldsymbol{\eta}}_z' \tilde{\mathbf{W}} A_{NT} \tilde{\mathbf{W}}' \hat{\boldsymbol{\eta}}_z\right)^{-1} \hat{\boldsymbol{\eta}}_z' \tilde{\mathbf{W}} A_{NT} \tilde{\mathbf{W}}' \hat{\boldsymbol{\eta}}_y, \tag{2.22}$$

where $\tilde{\mathbf{W}} \equiv (\tilde{W}_1', ..., \tilde{W}_N')'$, $\tilde{W}_i \equiv (\tilde{W}_{i3}, ..., \tilde{W}_{iT})'$, $\hat{\boldsymbol{\eta}}_a \equiv (\hat{\eta}_{a,1}', ..., \hat{\eta}_{a,N}')'$, and $\hat{\eta}_{a,i} \equiv (\hat{\eta}_{a,i3}, ..., \hat{\eta}_{a,iT})'$ for $a = y$ and z. In case $A_{NT} = (\frac{1}{\tilde{n}}\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}$, $\hat{\theta}_{gmm}$ becomes the two-stage least squares (2SLS) estimator:

$$\hat{\theta}_{2SLS} = \left(\hat{\boldsymbol{\eta}}_z' \mathbf{P}_{\tilde{W}} \hat{\boldsymbol{\eta}}_z\right)^{-1} \hat{\boldsymbol{\eta}}_z' \mathbf{P}_{\tilde{W}} \hat{\boldsymbol{\eta}}_y, \tag{2.23}$$

where $\mathbf{P}_{\tilde{\mathbf{W}}} \equiv \tilde{\mathbf{W}}(\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'$. Let $\eta_{a,i} \equiv (\eta_{a,i3}, ..., \eta_{a,iT})'$, and $\boldsymbol{\eta}_a \equiv (\eta_{a,1}', ..., \eta_{a,N}')'$ for $a = y$ and z. Then the infeasible semiparametric GMM estimator of θ_0 is given by

$$\tilde{\theta}_{gmm} = \left(\boldsymbol{\eta}_z' \tilde{\mathbf{W}} A_{NT} \tilde{\mathbf{W}}' \boldsymbol{\eta}_z\right)^{-1} \boldsymbol{\eta}_z' \tilde{\mathbf{W}} A_{NT} \tilde{\mathbf{W}}' \boldsymbol{\eta}_y. \tag{2.24}$$

Under some regularity conditions, both $\hat{\theta}_{gmm}$ and $\tilde{\theta}_{gmm}$ are asymptotic normal but with different asymptotic variances.

[3] Even if Z_{it} is strictly exogenous so that $V_{it} = Z_{it}$, we still need IVs for $\eta_{z,it}$ because of the appearance of the lagged dependent variable in $m_z\left(Y_{i,t-1}, X_{it}\right)$. In this case, we can use either $\left(Z_{it}', U_{i,t-2}'\right)'$ or $\left(\Delta Z_{it}', U_{i,t-2}'\right)'$ as IVs for $\eta_{z,it}$.

2.2.3 Final estimation of $m(u)$

Up to now, we have suppressed the dependence of all feasible estimates ($\hat{m}_y$, $\hat{m}_z$, $\hat{\theta}_{gmm}$) on the bandwidth vector $h = (h_0, h_1, ..., h_{d_x})$. As we shall see, to obtain the usual $\sqrt{\bar{n}}$-consistency of $\hat{\theta}_{gmm}$, we require the use of undersmoothing bandwidth (see Assumption A.6 below). For the estimation of the nonparametric object $m(u)$, we still follow the literature and recommend the use of the optimal rate of bandwidth sequence. To avoid confusion, we use $b = (b_0, b_1, ..., b_{d_x})$ to denote such a bandwidth sequence. Accordingly, we use $\hat{m}_y^b$ and $\hat{m}_z^b$ to denote $\hat{m}_y$ and $\hat{m}_z$ when the bandwidth h is replaced by b.

Plugging the estimator $\hat{\theta}_{gmm}$ into $\hat{m}_\theta(u)$, we obtain the estimate of $m(u)$ by

$$\hat{m}^b_{\hat{\theta}_{gmm}}(u) \equiv \hat{m}^b_y(u) - \hat{m}^b_z(u)' \hat{\theta}_{gmm}$$

for any $u \in \mathcal{U}$.

Note that $m(\cdot)$ is identified only upon to a location shift in (1.2). Under some assumptions (A.1(i)-(ii)) in the next subsection, we have $E[m(U_{i,t-1})] = E(Y_{it} - Z'_{it}\theta_0)$. This motivates us to recenter $\hat{m}^b_{\hat{\theta}_{gmm}}(u)$ to obtain

$$\hat{m}^b_{\hat{\theta}_{gmm}}(u) + \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}[Y_{it} - Z'_{it}\hat{\theta}_{gmm} - \hat{m}^b_{\hat{\theta}_{gmm}}(U_{i,t-1})].$$

2.3 Asymptotic properties of $\hat{\theta}_{IV}$ and $\hat{m}_{\hat{\theta}_{IV}}(u)$

Let $Y_i \equiv (Y_{i1}, ..., Y_{iT})'$. Similarly, we define Z_i, V_i, X_i or ε_i. Recall that $\eta_{y,it} = \Delta Y_{it} - \Delta m_{y,it}$, $E(\eta_{y,it}|U_{i,t-2}) = 0$, $\eta_{z,it} = \Delta Z_{it} - \Delta m_{z,it}$, and $E(\eta_{z,it}|U_{i,t-2}) = 0$. Let $\sigma^2_{\varepsilon,t-2}(u) \equiv \text{Var}(\Delta\varepsilon_{it}|U_{i,t-2} = u)$, $\sigma^2_{y,t-2}(u) \equiv \text{Var}(\eta_{y,it}|U_{i,t-2} = u)$, and $\sigma^2_{z_l,t-2}(u) \equiv \text{Var}(\eta_{z_l,it}|U_{i,t-2} = u)$ for $l = 1, ..., d_z$. Let $\bar{q} = q$ if q is odd and $q + 1$ if q is even. We make the following assumptions on $\{Y_{it}, X_{it}, Z_{it}, V_{it}, \alpha_i, \varepsilon_{it}\}$, the function of interest m, the kernel function k and the bandwidth h.

Assumptions.

A.1. (i) $(Y_i, X_i, Z_i, V_i, \alpha_i, \varepsilon_i)$, $i = 1, ..., N$, are IID. $E(\alpha_i) = 0$.

(ii) $E(\varepsilon_{it}|\underline{Y}_{i,t-1}, \underline{X}_{it}, \underline{V}_{it}) = 0$.

(iii) The PDF $f(\cdot)$ is uniformly bounded and bounded below from 0 on $\mathcal{U}$.

(iv) $\|m_y\|_2 \leq C_y$, $\max_{1\leq l\leq d_z}\|m_{z_l}\|_2 \leq C_z$ for some C_y, $C_z < \infty$, where m_{z_l} is the l-th element of m_z.

(v) $\int_{\mathcal{U}}[m(\bar{u}) - m(u)]^2 f(u) f(\bar{u}|u)\, d\bar{u}du > 0$ for all $m \in \mathcal{L}_2(f)$ with $m \neq 0$.

(vi) $\int_{\mathcal{U}}\int\left[\frac{f(\bar{u}|u)}{f(\bar{u})}\right]^2 f(\bar{u}) f(u)\, d\bar{u}du < \infty$.

(vii) $\sup_{\|m\|_2\leq 1}\sup_{u\in\mathcal{U}}\int|m(\bar{u})| f(\bar{u}|u)\, d\bar{u} < \infty$.

A.2. (i) For $t = 3, ..., T$, $f_{t-2}(\cdot)$ has all $(\bar{q}+1)$-th partial derivatives which are uniformly continuous on $\mathcal{U}$.

(ii) $m_y(\cdot)$ and $m_z(\cdot)$ have all $(\bar{q}+1)$-th partial derivatives which are uniformly continuous on $\mathcal{U}$.

(iii) For $t = 3, ..., T$, $\sigma^2_{\varepsilon,t-2}(\cdot)$, $\sigma^2_{y,t-2}(\cdot)$, and $\sigma^2_{z_l,t-2}(\cdot)$, $l = 1, ..., d_z$, have all second order partial derivatives which are uniformly continuous on $\mathcal{U}$.

A.3. The kernel function $k: \mathbb{R} \to \mathbb{R}$ is a symmetric and continuous PDF that has a compact support.

A.4. Let $h! \equiv \Pi_{l=0}^{d_x} h_l$ and $\|h\|^2 \equiv \sum_{l=0}^{d_x} h_l^2$. As $N \to \infty$, T is fixed, $\|h\| \to 0$, $Nh!/\log N \to \infty$, $N\|h\|^{2(\bar{q}+1)} h! \to C \in [0, \infty)$.

A.1-A.4 parallel Assumptions A.1-A.4 in Su and Lu (2013). A.1(i) rules out cross sectional dependence among $\{X_i, Z_i, V_i, \alpha_i, \varepsilon_i\}$ but allow *nonstationarity* along the time dimension for the time series $\{Y_{it}, X_{it}, Z_{it}, V_{it}, \varepsilon_{it}\}_{t=1}^{T}$. The latter means that the observed data can have time-varying marginal or transition density functions. A.1(ii) imposes sequential exogeneity on (X_{it}, V_{it}). A.1(iii) requires that $f(\cdot)$ be well behaved on $\mathcal{U}$ as in typical local polynomial regressions. A.1(iv) imposes the finite second moment on $m(U_{i,t-1})$ so that $\mathcal{L}_2(f)$ is well defined. A.1(v) imposes assumptions on the functional forms of the regression function $m(\cdot)$ and the mixture densities $f(\cdot)$ and $f(\cdot|\cdot)$. It ensures that the operator $\mathcal{I} - \mathcal{A}$ is one-to-one and (2.8) is satisfied. A.1(vi) implies that there is not much dependence between $U_{i,t-1}$ and $U_{i,t-2}$ under the mixture transition density $f(\cdot|\cdot)$ and it ensures that the operator $\mathcal{A}$ is Hilbert-Schmidt and a fortiori compact (see Carrasco et al. (2007) for further discussions). A.1(vii) also imposes some restriction on the operator $\mathcal{A}$ and can be easily satisfied. For example, if $\int m(u)^2 f(u)\,du < \infty$ and $\sup_{u\in\mathcal{U}} \int f(\bar{u}|u)^2 / f(\bar{u})\,d\bar{u} < \infty$, A.1(vii) holds by Cauchy-Schwarz inequality.

A.2 mainly specifies the smoothness conditions on f_{t-2}, m_y, m_z, $\sigma^2_{\varepsilon,t-2}$, $\sigma^2_{y,t-2}$, and $\sigma^2_{z_l,t-2}$. A.3 requires that the kernel k be compactly supported. This assumption can be relaxed at the cost of lengthy arguments. A.4 specifies conditions on the choice of bandwidth sequences and the local polynomial order q. Note that we shall use the fact that the bias for the qth order local polynomial regression is of order $O(\|h\|^{q+1})$ if q is odd and $O(\|h\|^{q+2})$ if q is even when the kernel function k is symmetric. See Li and Racine (2007, pp. 90-91).

Let $Dm_{\theta,it}(u) \equiv m_\theta(U_{i,t-2}) - m_\theta(u) - \sum_{1\le|\mathbf{j}|\le q} \frac{1}{\mathbf{j}!} m_\theta^{(\mathbf{j})}(u)(U_{i,t-2}-u)^{\mathbf{j}}$. Define

$$B_{\theta,NT}(u) \equiv \frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T} \bar{\mathcal{K}}_{it}(u)\,Dm_{\theta,it}(u) \text{ and } V_{\theta,NT}(u) \equiv \frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T} \bar{\mathcal{K}}_{it}(u)\,\eta_{\theta,it}, \tag{2.25}$$

where

$$\bar{\mathcal{K}}_{it}(u) \equiv e_1' \left[\bar{S}_{NT}(u)\right]^{-1} \mathbf{1}_{it} K_h(U_{i,t-2}-u)\,\mu_h(U_{i,t-2}-u), \tag{2.26}$$

and $\bar{S}_{NT}(u) \equiv E[S_{NT}(u)]$. Note that the non-stochastic term $\bar{S}_{NT}(u)$ is used in the definition of $V_{\theta,NT}(u)$ and $B_{\theta,NT}(u)$ to facilitate the asymptotic analysis. Analogously, define $B_{a,NT}(u)$ and $V_{a,NT}(u)$ with θ in the definitions of $B_{\theta,NT}(u)$ and $V_{\theta,NT}(u)$ being replaced by a for $a = y, z_1, ..., z_{d_z}$. By the standard local polynomial regression theory (e.g., Masry, 1996; Hansen, 2008), we have $\sup_{\theta\in\Theta}\sup_{u\in\mathcal{U}} |V_{\theta,NT}(u)| = O_P[(\log n)^{1/2}(nh!)^{-1/2}]$ and $\sup_{\theta\in\Theta}\sup_{u\in\mathcal{U}} |B_{\theta,NT}(u)| = O_P(\|h\|^{\bar{q}+1})$. Similar results hold for $B_{a,NT}(u)$ and $V_{a,NT}(u)$ with $a = y, z_1, ..., z_{d_z}$.

The following theorem characterizes the Bahadur-type representations of $\hat{m}_a(u)$, $a = y, z_1, ..., z_{d_z}$, and $\hat{m}_\theta(u)$ with a uniform control on the higher order terms that are asymptotically negligible in later study.

Theorem 2.1 *Suppose Assumptions A.1-A.4 hold. Then*

(i) $\sup_{u\in\mathcal{U}} \left|\hat{m}_a(u) - m_a(u) - (\mathcal{I}-\mathcal{A})^{-1} V_{a,NT}(u) - (\mathcal{I}-\mathcal{A})^{-1} B_{a,NT}(u)\right| = O_P(\nu_n^2)$ *for* $a = y, z_1, ..., z_{d_z}$,

(ii) $\sup_{\theta\in\Theta}\sup_{u\in\mathcal{U}} \left|\hat{m}_\theta(u) - m_\theta(u) - (\mathcal{I}-\mathcal{A})^{-1} V_{\theta,NT}(u) - (\mathcal{I}-\mathcal{A})^{-1} B_{\theta,NT}(u)\right| = O_P(\nu_n^2)$,
where $\nu_n \equiv (nh!)^{-1/2}(\log n)^{1/2} + \|h\|^{\bar{q}+1}$.

Remark 2.1. The result in Theorem 2.1(i) is analogous to that in Theorem 2.1 in Su and Lu (2013) for the case $a = y$. In particular, the terms $(\mathcal{I} - \mathcal{A})^{-1} V_{y,NT}(u)$ and $(\mathcal{I} - \mathcal{A})^{-1} B_{y,NT}(u)$ signify the asymptotic variance and bias of $\hat{m}_y(u)$, respectively. In terms of our notation, Su and Lu (2013) show that

$$\sup_{u\in\mathcal{U}} \left|\hat{m}_y(u) - m_y(u) - V_{y,NT}(u) - (\mathcal{I} - \mathcal{A})^{-1} B_{y,NT}(u)\right| = O_P\left[n^{-1/2}(\log n)^{1/2} + \nu_n^2\right],$$

where the asymptotic bias term has the same structure as ours but the variance term is different from ours. Observing that $(\mathcal{I} - \mathcal{A})^{-1} = \mathcal{I} + \mathcal{A}(\mathcal{I} - \mathcal{A})^{-1}$, the variance term in Theorem 2.1(i) for the case $a = y$ can be decomposed into two terms

$$(\mathcal{I} - \mathcal{A})^{-1} V_{y,NT}(u) = V_{y,NT}(u) + \mathcal{A}(\mathcal{I} - \mathcal{A})^{-1} V_{y,NT}(u). \tag{2.27}$$

The first term $V_{y,NT}(u)$ stands for the usual variance term for local polynomial regression that is $O_P[(nh!)^{-1/2}]$ for each $u \in \mathcal{U}$ and $O_P[(nh!)^{-1/2}(\log n)^{1/2}]$ uniformly in $u \in \mathcal{U}$. The second term $\mathcal{A}(\mathcal{I} - \mathcal{A})^{-1}$ $\times V_{y,NT}(u)$ appears frequently in kernel estimation based on solving a Fredholm integral equation of the second kind (see, e.g., Linton and Mammen (2005) and Su and Lu (2013)), and it is $O_P(n^{-1/2})$ for each $u \in \mathcal{U}$ and $O_P[n^{-1/2}(\log n)^{1/2}]$ uniformly in $u \in \mathcal{U}$ (see also condition (B4b) in the proof of Theorem 2.1(i)). Apparently, the second term is of smaller order than the first term and has been ignored by Su and Lu (2013) in their study of kernel estimation of nonparametric dynamic panel data models. We keep the second term in (2.27) because it contributes to the asymptotic variance of our semiparametric GMM estimator $\hat{\theta}_{gmm}$ of θ despite the fact it does not contribute to the asymptotic variance of our nonparametric estimator $\hat{m}_{\hat{\theta}_{gmm}}(u)$ of $m(u)$. As in Su and Lu (2013), the asymptotic bias term in Theorem 2.1(i) reflects the fact that the bias accumulates during the iteration.

To study the asymptotic normality of $\hat{\theta}_{gmm}$, we need to introduce more notation. Let $Q_{NT,wz} \equiv \tilde{n}^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T} W_{it}\eta'_{z,it}\tilde{\mathbf{1}}_{it}$ and $Q_{wz} \equiv E(Q_{NT,wz})$. Define the operator $\mathcal{L}(\bar{u}, u)$ by

$$(\mathcal{I} - \mathcal{A})^{-1} m(u) = \int_{\mathcal{U}} \mathcal{L}(u, \bar{u})\, m(\bar{u})\, f(\bar{u})\, d\bar{u}. \tag{2.28}$$

Let $\bar{\mathcal{L}}(v, u) \equiv \int_{\mathcal{U}} \mathcal{L}(u, \bar{u})\, \mathbf{1}(v \in \mathcal{U})\, e_1'\bar{S}_{NT}(\bar{u})^{-1} \mu_h(v - \bar{u})\, K_h(v - \bar{u})\, d\bar{u}$. Note that $\bar{\mathcal{L}}(v, u) \equiv \int_{\mathcal{U}} \mathcal{L}(u, \bar{u})$ $\times \mathbf{1}(v \in \mathcal{U})\, K_h(v - \bar{u})\, d\bar{u}$ in case $q = 1$. Let $\chi_{a,i} = (U'_{i1}, ..., U'_{iT}, W'_{i3}, ..., W'_{iT}, \eta_{a,i3}, ..., \eta_{a,iT})'$. For $a = y, z_1, ..., z_{d_z}$, define

$$\varphi_1(\chi_{a,i}) = \frac{-N(N-1)}{2n\tilde{n}} \sum_{s=3}^{T}\sum_{t=3}^{T} E_j\left[\left[\bar{\mathcal{L}}(U_{it}, U_{j,s-1}) - \bar{\mathcal{L}}(U_{it}, U_{j,s-2})\right]\tilde{\mathbf{1}}_{js} W_{js}\right]\eta_{a,it} \tag{2.29}$$

where E_j denotes expectation with respect to variables indexed by j. Let $\varphi_{1,z}(\chi_{z,i}) = (\varphi_1(\chi_{z_1,i}), ..., \varphi_1(\chi_{z_{d_z},i}))$. Note that $\varphi_1(\chi_{y,i})$ and $\varphi_{1,z}(\chi_{z,i})$ are of dimensions $d_w \times 1$ and $d_w \times d_z$, respectively, and they reflect the estimation errors by replacing $\eta_{y,it}$ and $\eta_{z,it}$ in (2.20) by their respective kernel estimates $\hat{\eta}_{y,it}$ and $\hat{\eta}_{z,it}$.

We add the following assumptions.

A.5. (i) As $N \to \infty$, $A_{NT} \xrightarrow{P} A > 0$ and $Q_{NT,wz} \xrightarrow{P} Q_{wz}$. Q_{wz} has full rank d_z.

(ii) $\max_{3\leq t\leq T} E\left\|\tilde{W}_{it}\right\| < \infty$.

(iii) $\frac{1}{\sqrt{\tilde{n}}}\tilde{\mathbf{W}}'\Delta\boldsymbol{\varepsilon} + \frac{\sqrt{\tilde{n}}}{N}\sum_{i=1}^{N}[\varphi_1(\chi_{y,i}) - \varphi_{1,z}(\chi_{z,i})\theta_0] \xrightarrow{D} N(0, \Omega_0)$.

A.6. As $N \to \infty$, $N \|h\|^{2(\bar{q}+1)} \to 0$ and $N (h!)^2 / (\log N)^2 \to \infty$.

A.5(i) and (ii) are standard in the GMM literature. If one chooses $A_{NT} = (\tilde{n}^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}$, we could require $Q_{NT,w} \equiv \tilde{n}^{-1} \sum_{i=1}^{N} \sum_{t=3}^{T} W_{it} W_{it}' \tilde{\mathbf{1}}_{it} \xrightarrow{P} E(Q_{NT,w}) > 0$. A.5(iii) is a high level condition. Let $\xi_{i,N} = \sum_{t=3}^{T} \tilde{W}_{it} \boldsymbol{\Delta}\varepsilon_{it} + [\varphi_1(\chi_{y,i}) - \varphi_{1,z}(\chi_{z,i})\theta_0]$. Under our large N and fixed T framework, one set of sufficient conditions for A.5(iii) is: (i) $\xi_{i,N}$ are independent across i with $E(\xi_{i,N}) = 0$; (ii) $E\|\xi_i\|^{2+\delta} < C$ for some $C < \infty$ and some $\delta > 0$. The zero mean restriction can be easily verified. The main complication lies in the verification of the moment conditions on $\varphi_1(\chi_{y,i})$ and $\varphi_{1,z}(\chi_{z,i})$, where we find it is difficult to obtain more primitive conditions. The term $[\varphi_1(\chi_{y,i}) - \varphi_{1,z}(\chi_{z,i})\theta_0]$ indicates the effect of parameter estimation error (PEE) because of the use of $\hat{\eta}_{a,it}$, $a = y$ and z, in the semiparametric estimation of θ. The first part of A.6 requires the use of an undersmoothing bandwidth for the nonparametric estimates in order to eliminate the impact of their asymptotic biases on the second stage parametric estimates, and the second part is needed to ensure $\tilde{n}\nu_n^2 = o_P(1)$.

The following theorem establishes the asymptotic normality of $\hat{\theta}_{gmm}$.

Theorem 2.2 *Under Assumptions A.1-A.6, we have*

$$\sqrt{\tilde{n}}\left(\hat{\theta}_{gmm} - \theta_0\right) \xrightarrow{d} N\left(0, \left(Q_{wz}' A Q_{wz}\right)^{-1} Q_{wz}' A \Omega_0 A Q_{wz} \left(Q_{wz}' A Q_{wz}\right)^{-1}\right).$$

Remark 2.2. We prove the above theorem based on the explicit formula for the semiparametric GMM estimator $\hat{\theta}_{gmm}$ and Theorem 2.1. As expected, $\hat{\theta}_{gmm}$ converges to θ_0 at the usual parametric rate $\tilde{n}^{1/2}$ but with the asymptotic variance different from that of its infeasible version $\tilde{\theta}_{gmm}$ because it is standard to show that

$$\sqrt{\tilde{n}}\left(\tilde{\theta}_{gmm} - \theta_0\right) \xrightarrow{d} N\left(0, \left(Q_{wz}' A Q_{wz}\right)^{-1} Q_{wz}' A \Omega_0^{\dagger} A Q_{wz} \left(Q_{wz}' A Q_{wz}\right)^{-1}\right)$$

where $\Omega_0^{\dagger} \equiv \lim_{N\to\infty} \frac{1}{\tilde{n}} \mathrm{Var}(\tilde{\mathbf{W}}'\Delta\boldsymbol{\varepsilon})$. To make inference, we need to obtain consistent estimates of Q_{wz} and Ω_0. Apparently, we can estimate Q_{wz} by $Q_{NT,wz}$. The estimation of Ω_0 is quite involved. We follow Chen et al. (2003) and Mammen et al. (2015) and recommend the use of bootstrapping for practical purpose. The procedure is standard and detailed in these papers.

Remark 2.3. Alternatively, we can establish the above result based on the literature on semiparametric estimation with nonparametrically generated covariates; see Newey (1994), Ahn (1997), Chen et al. (2003), Ichimura and Lee (2010), Kong et al. (2010), Hahn and Ridder (2013), Escanciano et al. (2014), Mammen et al. (2015), among others. Under certain conditions, these authors show that the estimator $\hat{\theta}_{gmm}$ exhibits the following representation

$$\begin{aligned}\sqrt{\tilde{n}}\left(\hat{\theta}_{gmm} - \theta_0\right) &= -\left(Q_{NT}^{\theta\prime} A_{NT} Q_{NT}^{\theta}\right)^{-1} Q_{NT}^{\theta\prime} A_{NT} \left\{\frac{1}{\sqrt{\tilde{n}}} \sum_{i=1}^{N} \sum_{t=3}^{T} q(\varsigma_{it}, \theta_0, \xi_0) + \sqrt{\tilde{n}} Q_{NT}^{\xi}[\hat{\xi} - \xi_0]\right\} \\ &\quad + O_P\left(\left\|\hat{\xi} - \xi_0\right\|_{\Xi}^2\right) + o_P(1). \qquad (2.30)\end{aligned}$$

Here Q_{NT}^{θ} denotes the ordinary derivative of $Q_{NT}(\theta, \xi)$ with respect to θ evaluated at the true value (θ_0, ξ_0); $Q_{NT}^{\xi}[\bar{\xi}] \equiv Q_{NT}^{\xi}(\theta_0, \xi_0)[\bar{\xi}]$ denotes the pathwise derivative of $Q_{NT}(\theta_0, \xi)$ at ξ_0 in the direction $\bar{\xi}$, i.e., $Q_{NT}^{\xi}[\bar{\xi}] \equiv \lim_{\tau\to 0} [Q_{NT}(\theta_0, \xi_0 + \tau\bar{\xi}) - Q_{NT}(\theta_0, \xi_0)] / \tau$ for any $\bar{\xi}$ such that $\xi_0 + \tau\bar{\xi} \in \Xi$ for $|\tau|$

sufficiently small; $\|\cdot\|_\Xi$ denotes the pseudo-norm induced by the sup-norm on a suitable class of smooth functions Ξ. That is, for any $\xi = (m_y, m_{z_1}, ..., m_{z_{d_z}}) \in \Xi$,

$$\|\xi\|_\Xi \equiv \sup_{u\in\mathcal{U}} |m_y(u)| + \sup_{u\in\mathcal{U}} \sum_{j=1}^{d_z} \left|m_{z_j}(u)\right|.$$

Noting that $Q_{NT}^\theta = -Q_{NT,wz}$ and $\sum_{i=1}^N \sum_{t=3}^T q(\varsigma_{it}, \theta_0, \xi_0) = \tilde{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}$ in out setup, the representation in (2.30) implies that $\hat{\theta}_{gmm}$ is $\sqrt{\tilde{n}}$-consistent and asymptotically normal with asymptotic variance $(Q_{wz}'AQ_{wz})^{-1} Q_{wz}'A\Omega_0 AQ_{wz} (Q_{wz}'AQ_{wz})^{-1}$ provided

$$\left\|\hat{\xi} - \xi_0\right\|_\Xi = o_P\left(\tilde{n}^{-1/4}\right) \tag{2.31}$$

and

$$\frac{1}{\sqrt{\tilde{n}}}\tilde{\mathbf{W}}'\Delta\boldsymbol{\varepsilon} + \sqrt{\tilde{n}}Q_{NT}^\xi[\hat{\xi} - \xi_0] \xrightarrow{D} N(0, \Omega_0). \tag{2.32}$$

Verifying (2.31) is straightforward given the uniform convergence results in Theorem 2.1. To verify (2.32), we can readily calculate the path derivative of $Q_{NT}(\theta_0, \xi)$ at ξ_0 to obtain

$$\sqrt{\tilde{n}}Q_{NT}^\xi[\hat{\xi} - \xi_0] = \frac{1}{\sqrt{\tilde{n}}} \sum_{i=1}^N \sum_{t=1}^T \tilde{W}_{it} \left[(-\delta_{y,it} + \delta_{y,i,t-1}) - (-\delta_{z,it} + \delta_{z,i,t-1})'\theta_0\right] \tag{2.33}$$

where $\delta_{a,it} = \hat{m}_a(U_{i,t-1}) - m_a(U_{i,t-1})$ for $a = y$ and z. Under our conditions that include the use of undersmoothing bandwidth h, we can apply Theorem 2.1 and show that

$$\sqrt{\tilde{n}}Q_{NT}^\xi[\hat{\xi} - \xi_0] = \frac{\sqrt{\tilde{n}}}{N} \sum_{i=1}^N [\varphi_1(\chi_{y,i}) - \varphi_{1,z}(\chi_{z,i})\theta_0] + o_P(1)$$

where both $\varphi_1(\chi_{y,i})$ and $\varphi_{1,z}(\chi_{z,i})$ have zero mean and are square-integrable.

We did not use the above arguments in our proof for two reasons. First, we have explicit formula for our semiparametric GMM estimator and their asymptotic distribution can be established by following standard asymptotic tools without resorting to the empirical process theory. Second, to apply the results in papers such as Newey (1994) and Chen et al. (2003), we have to limit our attention to a particular class of smooth functions (e.g., Ξ is the popular nonparametric function class studied in detail in van der Vaart and Wellner (1996, p.154)), and verify that the corresponding nonparametric estimates, $\hat{m}_a$, $a = y, z_1, ..., z_d$, also belong to this class with probability tending to one. It is well known such a verification can be extremely difficult even in applications that are involved with standard kernel estimates (e.g., local polynomial quantile regression estimates). This is indeed the case here because $\hat{m}_a$'s do not possess a closed-form expression and we are unable to verify that they belong to the same function class as their population truth. See Escanciano et al. (2014) for further discussions.

Given the asymptotic normality result of $\hat{\theta}_{gmm}$ in the above theorem, we establish the uniform convergence and point-wise asymptotic normality of $\hat{m}^b_{\hat{\theta}_{gmm}}(u)$ in the following theorem.

Theorem 2.3 *Suppose Assumptions A.1-A.6 hold. Suppose that Assumption A.4 is also satisfied with h replaced by $b = (b_0, b_1, ..., b_{d_x})$. Then*

(i) $\sup_{u\in\mathcal{U}}\left|\hat{m}^{b}_{\hat{\theta}_{gmm}}(u)-m(u)\right|=O_P((nb!)^{-1/2}(\log n)^{1/2}+\|b\|^{\bar{q}+1})$;
(ii) for any $u\in$*interior*$(\mathcal{U})$,

$$\sqrt{nb!}\left[\hat{m}^{b}_{\hat{\theta}_{gmm}}(u)-m(u)-(\mathcal{I}-\mathcal{A})^{-1}B_0(u)\right]\xrightarrow{d}N\left(0,\frac{\sigma_0^2(u)}{f(u)}e_1'\mathbb{S}^{-1}\mathbb{K}\mathbb{S}^{-1}e_1\right),$$

where $B_0(u)\equiv e_1'\mathbb{S}^{-1}\sum_{|\mathbf{j}|=\bar{q}+1}(\mathbf{j}!)^{-1}m^{(\mathbf{j})}(u)\int K(w)\mu_h(w)(w\odot b)^{\mathbf{j}}dw$, $\mathbb{S}\equiv[f(u)]^{-1}\lim_{N\to\infty}E[\bar{S}_{NT}(u)]$, $\mathbb{K}\equiv\int K(\bar{u})^2\mu(\bar{u})\mu(\bar{u})'d\bar{u}$, *and* $\sigma_0^2(u)=\sum_{t=3}^{T}(p_{t-2}/p)f_{t-2}(u)\sigma^2_{\varepsilon,t-2}(u)$.

Remark 2.4. The above results are as expected. Since the parametric estimate $\hat{\theta}_{gmm}$ converges to θ_0 at the usual parametric rate, it has no asymptotic impact on the estimation of the nonparametric component $m(u)$. As in Su and Lu (2013), the asymptotic bias and variance formulae in Theorem 2.3 exhibit complicated forms because of the allowance of a general order of local polynomial regressions and the use of different bandwidths for different covariates. In the special case where $q=1$, one can easily verify that

$$\mathbb{S}=\begin{pmatrix}1 & \mathbf{0}_{1\times(d_x+1)}\\ \mathbf{0}_{(d_x+1)\times 1} & \mathbf{I}_{d_x+1}\int v^2k(v)dv\end{pmatrix},\text{ and }\mathbb{K}=\begin{pmatrix}\left[\int k(v)^2dv\right]^{d_x+1} & \mathbf{0}_{1\times(d_x+1)}\\ \mathbf{0}_{(d_x+1)\times 1} & \mathbf{I}_{d_x+1}\left[\int v^2k(v)^2dv\right]^{d_x+1}\end{pmatrix}.$$

Then the asymptotic variance simplifies to $\frac{\sigma_0^2(u)}{f(u)}[\int k(v)^2dv]^{d_x+1}$ and $B_0(u)=\frac{1}{2}\sum_{l=0}^{d_x}b_l^2\frac{\partial^2 m(u)}{\partial u_l^2}\int v^2k(v)dv$.
Remark 2.5. From Theorem 2.3, we can see that the asymptotic variance of the estimator $\hat{m}^{b}_{\hat{\theta}_{gmm}}(u)$ shares the same structure as that of a typical local polynomial estimator of either m in the model

$$\Delta Y_{it}-\theta_0'\Delta Z_{it}=m(U_{i,t-1})-m(U_{i,t-2})+\Delta\varepsilon_{it} \tag{2.34}$$

by pretending the other one and the finite dimensional parameter θ_0 are known. Nevertheless, the asymptotic bias of $\hat{m}^{b}_{\hat{\theta}_{gmm}}(u)$ is different from the case where one of the two m's and θ_0 are known in (2.34) since the operator $(\mathcal{I}-\mathcal{A})^{-1}$ signifies the accumulated bias. Since the error term $\Delta\varepsilon_{it}$ in (2.34) follows an MA(1) process, it is interesting to explore such an MA(1) structure and propose a more efficient estimate of $m(u)$. But because the process $\{\Delta\varepsilon_{it},\ t\geq 2\}$ is not invertible, the techniques developed in Xiao et al. (2003) and Su and Ullah (2006b) are not applicable here.
Remark 2.6. Interestingly, our method can also be applied to estimate the partially linear model with both individual and time fixed effects:

$$Y_{it}=Z_{it}'\theta_0+m(Y_{i,t-1},X_{it})+\alpha_i+\omega_t+\varepsilon_{it}, \tag{2.35}$$

where the new term ω_t indicates unobserved time effect. The first-differenced model now becomes

$$\begin{aligned}\Delta Y_{it} &= \Delta Z_{it}^{\dagger\prime}\theta_0^{\dagger}+D_t'\omega+m(Y_{i,t-1},X_{it})-m(Y_{i,t-2},X_{i,t-1})+\Delta\varepsilon_{it}\\ &= \Delta Z_{it}^{\dagger\prime}\theta_0^{\dagger}+m(Y_{i,t-1},X_{it})-m(Y_{i,t-2},X_{i,t-1})+\Delta\varepsilon_{it}\end{aligned} \tag{2.36}$$

where $\omega=(\omega_1,...,\omega_T)'$, $\theta_0^{\dagger}=(\theta_0^{\dagger\prime},\omega')'$, $\Delta Z_{it}^{\dagger}=(\Delta Z_{it}',D_t')'$, and D_t is a $T\times 1$ vector with 1 and -1 in its t and $(t-1)$th positions, respectively, for $t=2,...,T$. Then one can apply the method proposed in this section to estimate both $m(\cdot)$ and $\theta_0^{\dagger}$ jointly.

3 Sieve IV estimation

In this section, we consider sieve IV estimation of θ and m.

3.1 Estimation

Since $m(\cdot)$ is unknown, we propose to estimate $m(\cdot)$ and θ jointly by the method of sieves (see, e.g., Chen (2007)). To proceed, let $\{p_l(u),\ l=1,2,\ldots\}$ denote a sequence of known basis functions that can well approximate any square-integrable function of u. Let $L \equiv L_N$ be some integer such that $L \to \infty$ as $N \to \infty$. Let $p^L(u) \equiv (p_1(u), p_2(u), \ldots, p_L(u))'$ be the $L \times 1$ vector of basis functions. Let $p_{i,t-1} \equiv p^L(U_{i,t-1})$, $\Delta p_{i,t-1} \equiv p_{i,t-1} - p_{i,t-2}$, $\Delta p_i \equiv (\Delta p_{i2}, \ldots, \Delta p_{i,T-1})'$, and $\mathbf{\Delta p} \equiv (\Delta p_1', \Delta p_2', \ldots, \Delta p_N')'$. Obviously, we suppress the dependence of p_{it}, Δp_{it}, Δp_i and $\mathbf{\Delta p}$ on L, N, or T. In particular, Δp_i and $\mathbf{\Delta p}$ are of dimension $T_2 \times L$ and $NT_2 \times L$, respectively.

Under fairly weak conditions, we can approximate $m(U_{i,t-1}) - m(U_{i,t-2})$ in (2.1) by $\beta_m' \Delta p_{i,t-1}$ for some $L \times 1$ vector β_m. This motivates us to consider the following model

$$\Delta Y_{it} = \Delta Z_{it}'\theta_0 + \beta_m' \Delta p_{i,t-1} + \Delta\varepsilon_{it} + R_{it}, \tag{3.1}$$

where $R_{it} \equiv m(U_{i,t-1}) - m(U_{i,t-2}) - \beta_m' \Delta p_{i,t-1}$ signifies the sieve approximation error. Note that $\Delta p_{i,t-1}$ is correlated with $\Delta\varepsilon_{it}$ and that ΔZ_{it} may be correlated with $\Delta\varepsilon_{it}$ too unless Z_{it} is strictly exogenous. To estimate θ and β_m in the above model consistently, we run the regression of ΔY_{it} on ΔZ_{it} and $\Delta p_{i,t-1}$ by using a $d_{\bar{w}} \times 1$ vector $\bar{W}_{it}$ as the IV for $\Delta p_{i,t-1}$ and ΔZ_{it}, where $d_{\bar{w}} \geq d_z + L$. Note that a variety of measurable functions of $\underline{U}_{i,t-2}$ can serve as valid instruments for $\Delta p_{i,t-1}$. Following the lead of Anderson and Hsiao (1981), in the simulations and applications below we choose $\bar{W}_{it} \equiv (V_{i,t-1}', p_{i,t-2}')'$ when Z_{it} is endogenous and can be implemented by V_{it} and $\bar{W}_{it} \equiv \left(\Delta Z_{it}', p_{i,t-2}'\right)'$ when Z_{it} is strictly exogenous. We assume that $d_{\bar{w}} \leq cL$ for some $c > 1$.

Let $\bar{W}_i \equiv (\bar{W}_{i3}, \ldots, \bar{W}_{iT})'$, $\bar{\mathbf{W}} \equiv (\bar{W}_1', \bar{W}_2', \ldots, \bar{W}_N')'$, $\Delta Y_i \equiv (\Delta Y_{i3}, \ldots, \Delta Y_{iT})'$, and $\mathbf{\Delta Y} \equiv (\Delta Y_1', \Delta Y_2', \ldots, \Delta Y_N')'$. Similarly define R_i and $\mathbf{R}$. Then the sieve IV/2SLS estimate of $\left(\theta', \beta_m'\right)'$ is given by[4]

$$\left(\hat{\theta}_{sieve}',\ \hat{\beta}_m'\right)' \equiv \left[(\mathbf{\Delta Z}, \mathbf{\Delta p})' \mathbf{P}_{\bar{\mathbf{W}}} (\mathbf{\Delta Z}, \mathbf{\Delta p})\right]^{-} (\mathbf{\Delta Z}, \mathbf{\Delta p})' \mathbf{P}_{\bar{\mathbf{W}}} \mathbf{\Delta Y}$$

where $\mathbf{P}_{\bar{\mathbf{W}}} \equiv \bar{\mathbf{W}}(\bar{\mathbf{W}}'\bar{\mathbf{W}})^{-}\bar{\mathbf{W}}'$. Let $\mathbf{Y}_w \equiv \mathbf{P}_{\bar{\mathbf{W}}}\mathbf{\Delta Y}$, $\mathbf{Z}_w \equiv \mathbf{P}_{\bar{\mathbf{W}}}\mathbf{\Delta Z}$, and $\mathbf{p}_w \equiv \mathbf{P}_{\bar{\mathbf{W}}}\mathbf{\Delta p}$. By the formula for partitioned regressions, we have

$$\hat{\beta}_m = \left(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\right)^{-1}\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{Y}_w \text{ and } \hat{\theta}_{sieve} = \left(\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{Z}_w\right)^{-1}\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{Y}_w.$$

Then we can estimate $m(u)$ by $\hat{m}_{sieve}(u) \equiv \hat{\beta}_m' p^L(u)$. We recenter $\hat{m}_{sieve}(u)$ as follows

$$\hat{m}_{sieve}(u) + \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[Y_{it} - Z_{it}'\hat{\theta}_{sieve} - \hat{m}_{sieve}(U_{i,t-1})\right].$$

[4] More generally, we can consider the sieve GMM estimate defined by: $(\hat{\theta}_{sieve}', \hat{\beta}_m')' = [(\mathbf{\Delta Z}, \mathbf{\Delta p})' \bar{\mathbf{W}}\bar{A}_{NT}\bar{\mathbf{W}}' (\mathbf{\Delta Z}, \mathbf{\Delta p})]^{-} (\mathbf{\Delta Z}, \mathbf{\Delta p})' \bar{\mathbf{W}}\bar{A}_{NT}\bar{\mathbf{W}}'\mathbf{\Delta Y}$, $\bar{A}_{NT}$ is a $d_{\bar{w}} \times d_{\bar{w}}$ weight matrix that is symmetric and asymptotically positive definite. The asymptotic properties are similar to the 2SLS estimator but the notation is slightly more complicated. So we decide to focus on the sieve 2SLS estimation here.

3.2 Asymptotic properties of $\hat{\theta}_{sieve}$ and $\hat{m}_{sieve}$

To apply the method of sieves, we assume that $m(u)$ satisfies some smoothness conditions. Let $\mathcal{X} \equiv \mathcal{Y} \times \mathcal{X}_1 \subset \mathbb{R} \times \mathbb{R}^{d_x}$ be the support of $(Y_{i,t-1}, X_{it})$. Early literature (e.g., Newey (1997) and de Jong (2002)) requires compact support implicitly or explicitly. To allow for the unboundedness of $\mathcal{X}$, we follow Chen et al. (2005), Su and Jin (2012), and Lee (2014) and use a weighted sup-norm metric defined as

$$\|m\|_{\infty,\varpi} \equiv \sup_{u \in \mathcal{X}} |m(u)| \left[1 + \|u\|^2\right]^{-\varpi/2} \text{ for some } \varpi \geq 0. \tag{3.2}$$

If $\varpi = 0$, the norm defined in (3.2) is the usual sup-norm that is suitable for the case of compact support.

Recall that a typical smoothness assumption requires that a function $m : \mathcal{X} \to \mathbb{R}$ belong to a Hölder space. Let $\boldsymbol{\alpha} \equiv (\alpha_1, \ldots, \alpha_{d_x+1})'$ denote a $(d_x + 1)$-vector of non-negative integers and $|\boldsymbol{\alpha}| \equiv \sum_{l=1}^{d_x+1} \alpha_l$. For any $u = (u_1, \ldots, u_{d_x+1})$, the $|\boldsymbol{\alpha}|$th derivative of $m : \mathcal{X} \to \mathbb{R}$ is denoted as $\nabla^{\boldsymbol{\alpha}} m(u) \equiv \partial^{|\boldsymbol{\alpha}|} m(u)/(\partial u_1^{\alpha_1} \ldots \partial u_{d_x+1}^{\alpha_d})$. The Hölder space $\Lambda^\gamma(\mathcal{X})$ of order $\gamma > 0$ is a space of functions $m : \mathcal{X} \to \mathbb{R}$ such that the first $\lceil \gamma \rceil$ derivatives are bounded, and the $\lceil \gamma \rceil$th derivatives are Hölder continuous with the exponent $\gamma - \lceil \gamma \rceil \in (0, 1]$. Define the Hölder norm:

$$\|m\|_{\Lambda^\gamma} \equiv \sup_{u \in \mathcal{X}} |m(u)| + \max_{|\boldsymbol{\alpha}| = \lceil \gamma \rceil} \sup_{u \neq u^*} \frac{|\nabla^{\boldsymbol{\alpha}} m(u) - \nabla^{\boldsymbol{\alpha}} m(u^*)|}{\|u - u^*\|^{\gamma - \lceil \gamma \rceil}}.$$

The following definition is adopted from Chen et al. (2005).

Definition 1. *Let $\Lambda^\gamma(\mathcal{X}, \varpi) \equiv \{m : \mathcal{X} \to \mathbb{R} \text{ such that } m(\cdot)[1 + ||\cdot||^2]^{-\varpi/2} \in \Lambda^\gamma(\mathcal{X})\}$ denote a weighted Hölder space of functions. A weighted Hölder ball with radius c is*

$$\Lambda_c^\gamma(\mathcal{X}, \varpi) \equiv \left\{ m \in \Lambda^\gamma(\mathcal{X}, \varpi) : \left\| m(\cdot)[1 + ||\cdot||^2]^{-\varpi/2} \right\|_{\Lambda^\gamma} \leq c < \infty \right\}.$$

Function $m(\cdot)$ is said to be $H(\gamma, \varpi)$-smooth on $\mathcal{X}$ if it belongs to a weighted Hölder ball $\Lambda_c^\gamma(\mathcal{X}, \varpi)$ for some $\gamma > 0$, $c > 0$ and $\varpi \geq 0$.

To proceed, we define some additional notation. Let $Q_{NT,\bar{w}z} \equiv N^{-1} T_2^{-1} \sum_{i=1}^N \sum_{t=3}^T \bar{W}_{it} \Delta Z_{it}'$, $Q_{NT,\bar{w}} \equiv N^{-1} T_2^{-1} \sum_{i=1}^N \sum_{t=3}^T \bar{W}_{it} \bar{W}_{it}'$, and $Q_{NT,\bar{w}p} \equiv N^{-1} T_2^{-1} \sum_{i=1}^N \sum_{t=3}^T \bar{W}_{it} \Delta p_{it}'$. Let $Q_{\bar{w}z} \equiv E(Q_{NT,\bar{w}z})$, $Q_w \equiv E(Q_{NT,\bar{w}})$, and $Q_{\bar{w}p} \equiv E(Q_{NT,\bar{w}p})$. Define

$$\begin{aligned}
Q_1 &\equiv Q_{\bar{w}z}' Q_{\bar{w}}^{-1} Q_{\bar{w}z} - Q_{\bar{w}z}' Q_{\bar{w}}^{-1} Q_{\bar{w}p} \left(Q_{\bar{w}p}' Q_{\bar{w}}^{-1} Q_{\bar{w}p} \right)^{-1} Q_{\bar{w}p}' Q_{\bar{w}}^{-1} Q_{\bar{w}z}, \\
Q_2 &\equiv Q_{\bar{w}z}' Q_{\bar{w}}^{-1} - Q_{\bar{w}z}' Q_{\bar{w}}^{-1} Q_{\bar{w}p} \left(Q_{\bar{w}p}' Q_{\bar{w}}^{-1} Q_{\bar{w}p} \right)^{-1} Q_{\bar{w}p}' Q_{\bar{w}}^{-1}, \\
Q_3 &\equiv Q_{\bar{w}p}' Q_{\bar{w}}^{-1} Q_{\bar{w}p} - Q_{\bar{w}p}' Q_{\bar{w}}^{-1} Q_{\bar{w}z} \left(Q_{\bar{w}z}' Q_{\bar{w}}^{-1} Q_{\bar{w}z} \right)^{-1} Q_{\bar{w}z}' Q_{\bar{w}}^{-1} Q_{\bar{w}p}, \text{ and} \\
Q_4 &\equiv Q_{\bar{w}p}' Q_{\bar{w}}^{-1} - Q_{\bar{w}p}' Q_{\bar{w}}^{-1} Q_{\bar{w}z} \left(Q_{\bar{w}z}' Q_{\bar{w}}^{-1} Q_{\bar{w}z} \right)^{-1} Q_{\bar{w}z}' Q_{\bar{w}}^{-1}.
\end{aligned}$$

Let $Q_{pp,\alpha} \equiv \int_{u \in \mathcal{X}} p^L(u) p^L(u)' \alpha(u) du$ where $\alpha(u)$ is a nonnegative weight function.

To establish the large sample properties of $\hat{\theta}_{sieve}$, we need the following assumptions.

A.7. (i) $m(\cdot)$ is $H(\gamma, \varpi)$-smooth on $\mathcal{X}$ for some $\gamma > (d_x + 1)/2$ and $\varpi \geq 0$.

(ii) For any $H(\gamma, \varpi)$-smooth function $m(u)$, there exists a linear combination of basis functions $\Pi_{\infty,L} m \equiv \beta_m' p^L(\cdot)$ in the sieve space $\mathcal{G}_L \equiv \{m(\cdot) = \alpha' p^L(\cdot)\}$ such that $\|m - \Pi_{\infty,L} m\|_{\infty,\bar{\varpi}} = O\left(L^{-\gamma/(d_x+1)}\right)$ for some $\bar{\varpi} > \varpi + \gamma$.

(iii) $\text{plim}_{(N,T)\to\infty} (NT_1)^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} \left(1 + ||U_{i,t-1}||^2\right)^{\bar{\varpi}} \alpha(U_{i,t-1}) < \infty$.

(iv) There are a sequence of constants $\zeta_0(L)$ and compact sets $\mathcal{U}_N$ satisfying that $\sup_{u\in\mathcal{U}_N} \|u\| = O(\zeta_0(L)^{1/\bar{\varpi}})$, $\sup_{u\in\mathcal{U}_N} \left\|p^L(u)\right\| \leq \zeta_0(L)$, and $\zeta_0(L)^2 L/N \to 0$ as $N \to \infty$.

A.8. (i) $\sup_{2\leq t\leq T} E\|\chi_{it}\|^4 \leq C < \infty$ for $\chi_{it} = Z_{it}$, $\bar{W}_{it}$, and p_{it}.

(ii) $Q_{NT,\bar{w}} \xrightarrow{P} Q_{\bar{w}} > 0$, $Q_{NT,\bar{w}z} \xrightarrow{P} Q_{\bar{w}z}$, and $Q_{NT,\bar{w}p} \xrightarrow{P} Q_{\bar{w}p}$. $(Q_{\bar{w}z}, Q_{\bar{w}p})$ has full rank $d_z + L$.

(iii) $\Omega_1 \equiv \text{Var}[T_2^{-1/2} \sum_{t=3}^{T} \bar{W}_{1t}\Delta\varepsilon_{1t}] > 0$.

(iv) There exist positive constants $\underline{c}_{Q_1}$, $\bar{c}_{Q_1}$, $\underline{c}_{Q_3}$ and $\bar{c}_{Q_3}$ such that $0 < \underline{c}_{Q_l} \leq \lambda_{\min}(Q_l) \leq \bar{c}_{Q_l} < \infty$ for $l = 1, 3$.

(v) $\lambda_{\max}(Q_{pp,\alpha}) < \infty$.

A.9. As $N \to \infty$, $L^3/N \to 0, \sqrt{N}L^{-\gamma/(d_x+1)} \to 0$.

Assumptions A.7(i)-(iii) are widely assumed in the literature on sieve estimation when an infinite support is allowed; see Chen et al. (2005), Su and Jin (2012), Lee (2014), and Su and Zhang (2015). A.7(iv) is needed to obtain the uniform convergence rate for the sieve estimate $\hat{m}_{sieve}(u)$ over a possibly divergent sequences of compact sets. A.8(i) imposes moment conditions on Z_{it}, $\bar{W}_{it}$, and p_{it}. A.8(ii)-(iii) are standard (see, e.g., Newey (1997)). A.8(iv) is a high level assumption. A.9 imposes further restrictions on L to control the sieve approximation bias and variance.

We establish the asymptotic normality of $\hat{\theta}_{sieve}$ in the following theorem.

Theorem 3.1 *Suppose Assumptions A.1(i)-(ii) and A.7-A.9 hold. Then* $\sqrt{NT_2}(\hat{\theta}_{sieve} - \theta_0) \xrightarrow{d} N(0, Q_1^{-1}Q_2\Omega_1 Q_2' Q_1^{-1})$.

Remark 3.1. The above theorem gives the asymptotic distribution of $\hat{\theta}_{sieve}$. The sieve IV estimation uses all observations in the estimation procedure and it is easier to implement than the kernel-based semiparametric GMM estimation. Nevertheless, it seems difficult to make a theoretical comparison between the two types of estimates. We compare them only through Monte Carlo simulations.

The following theorem reports the convergence rates and asymptotic normality of $\hat{m}_{sieve}(u)$.

Theorem 3.2 *Suppose Assumptions A.1(i)-(ii) and A.7-A.9 hold. Suppose* $\left\|p^L(u)\right\| \geq c > 0$. *Then*

(i) $\int [\hat{m}_{sieve}(u) - m(u)]^2 \alpha(u)\, du = O_P\left(L/N + L^{-2\gamma/(d_x+1)}\right)$,

(ii) $\frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} [\hat{m}_{sieve}(U_{i,t-1}) - m(U_{i,t-1})]^2 = O_P\left(L/N + L^{-2\gamma/(d_x+1)}\right)$,

(iii) $\sup_{u\in\mathcal{U}_N} |\hat{m}_{sieve}(u) - m(u)| = O_P(\zeta_0(L)(\sqrt{L/N} + L^{-\gamma/(d_x+1)}))$,

(iv) $\sqrt{NT_2} A_L(u)^{-1/2} [\hat{m}_{sieve}(u) - m(u)] \xrightarrow{d} N(0, 1)$,

where $A_L(u) \equiv p^L(u)' Q_3^{-1} Q_4 \Omega_1 Q_4' Q_3^{-1} p^L(u)$.

Remark 3.2. Following Newey (1997), we can also study the consistency and asymptotic normality of the plug-in estimates of linear or nonlinear functionals of m. The argument is standard and we do not repeat it to save space. To improve the finite sample performance of the nonparametric estimate, we can treat $\hat{\theta}_{sieve}$ as the true value θ_0 and then apply the method in Su and Lu (2013) to re-estimate the unknown function $m(u)$. Given $\hat{\theta}_{sieve}$, we can solve the following integral equation

$$m(u) = \mathcal{A}m(u) + \left[r_y(u) - \hat{\theta}_{sieve}' r_z(u)\right]. \tag{3.3}$$

Solving the empirical integral equation gives the new local polynomial estimator $\hat{m}(u)$ of $m(u)$, which can be written as $\hat{m}^{b}_{\hat{\theta}_{sieve}}(u)$ using notation defined in Section 2 when the bandwidth is chosen to be b and the same kernel function is used as before. We denote these estimators as $\hat{m}_{s,iter}$ and $\hat{m}_{s,noniter}$ when iterative and noniterative algorithms are used, respectively. It is trivial to show that both estimators have the same asymptotic distribution as $\hat{m}^{b}_{\hat{\theta}_{gmm}}(u)$ given in Theorem 2.3.

Remark 3.3. Once we obtain the estimators $\hat{m}_{s,iter}$ and $\hat{m}_{s,noniter}$, we can update the estimates of θ by treating $\hat{m}_{s,iter}$ or $\hat{m}_{s,noniter}$ as the true unknown function m by IV or OLS method based on the following first differenced models:

$$\Delta Y_{it} - [\hat{m}_{s,iter}(U_{i,t-1}) - \hat{m}_{s,iter}(U_{i,t-2})] = \Delta Z'_{it}\theta_0 + \Delta\varepsilon_{it} \tag{3.4}$$

or

$$\Delta Y_{it} - [\hat{m}_{s,noniter}(U_{i,t-1}) - \hat{m}_{s,noniter}(U_{i,t-2})] = \Delta Z'_{it}\theta_0 + \Delta\varepsilon_{it}. \tag{3.5}$$

When Z_{it} is endogenous, one could use the same IV as before, i.e., we can use a subvector of $\bar{W}_{it}$ that corresponds to the IV for ΔZ_{it}. If Z_{it} is strictly exogenous, then OLS regression will yield consistent estimates of θ. But because $m(\cdot)$ can only be estimated sufficiently well on compact sets, we only use observations for which both $U_{i,t-1}$ and $U_{i,t-2}$ lie on $\mathcal{U}$ in the above regression. We denote these two estimates of θ as $\hat{\theta}_{s,iter}$ or $\hat{\theta}_{s,noniter}$, respectively. Following the proof of Theorem 2.2, we can readily establish their $\sqrt{N}$-consistency and asymptotic normality under some regularity conditions. We omit the details to save space.

4 Testing for the linearity of the unknown nonparametric component

In this section we maintain the correct specification of the partially linear panel data model and consider testing the linearity of the nonparametric component $m(\cdot)$ in the partially linear model. The null hypothesis is

$$\mathbb{H}_0 : m(U_{i,t-1}) = \upsilon_0 + \gamma_0' U_{i,t-1} \text{ a.s. for some } (\upsilon_0, \gamma_0')' \in \Upsilon \subset \mathbb{R}^{d_x+2}$$

where $i = 1, ..., N$, $t = 2, ..., T$, and Υ is a compact subset of $\mathbb{R}^{d_x+2}$. The alternative hypothesis is given by

$$\mathbb{H}_1 : \Pr[m(U_{i,t-1}) = \upsilon + \gamma' U_{i,t-1}] < 1 \text{ for } \forall\ (\upsilon, \gamma')' \in \Upsilon \text{ for some } t = 2, ..., T.$$

There are several tests for linearity in nonparametric panel data models. Lee (2013) proposes a residual-based test to check the validity of the linear *dynamic* models with both large N and T and her test requires the consistent estimation of the generalized spectral derivatives which is impossible for fixed T. In the spirit of Härdle and Mammen (1993), Su and Lu (2013) introduce a nonparametric test for linearity in nonparametric *dynamic* panel data models by comparing two estimates, i.e., the restricted estimate under $\mathbb{H}_0$ and the unrestricted estimate under $\mathbb{H}_1$ when N is large and T is fixed, and Lin et al. (2014) propose a similar test for a linear functional form in *static* panel data models when both N and T are large. This idea is also used in Su and Zhang (2015) who propose a test for linear functional form in nonparametric *dynamic* panel data model with *interactive* fixed effects when both N and T are large. In addition, Li et al. (2011) propose a test for linearity of nonparametric component in a cross-sectional partially linear model and propose to obtain the p-value by using fiducial method. Nevertheless, to the best of our knowledge, there is no test for linearity of nonparametric component in partially linear dynamic panel data models with fixed effects.

Following Su and Lu (2013), we consider the smooth functional

$$\Gamma \equiv \int_{\mathcal{U}} \left[m(u) - \upsilon_0 - u'\gamma_0 \right]^2 a(u) f(u) \, du,$$

where $a(u)$ is a user-specified nonnegative weighting function with compact support $\mathcal{U}$. Clearly, we have $\Gamma = 0$ under $\mathbb{H}_0$ and generally $\Gamma > 0$ under $\mathbb{H}_1$. This motivates us to propose a test based on Γ. Under $\mathbb{H}_0$, we estimate the following linear panel data model

$$Y_{it} = Z_{it}'\theta_0 + \upsilon_0 + \gamma_0' U_{i,t-1} + \alpha_i + \varepsilon_{it}$$

by applying the usual IV/GMM method to the first-differenced model. For example, if Z_{it} and X_{it} are strictly exogenous, we can follow Anderson and Hsiao (1980) or Arellano and Bond (1991) to obtain the IV/GMM estimates of θ and γ. Let $(\breve{\theta}, \breve{\gamma})$ be the IV/GMM estimate. Then we can estimate υ by $\breve{\upsilon} = \frac{1}{NT_2} \sum_{i=1}^{N} \sum_{t=3}^{T} \left(Y_{it} - Z_{it}'\breve{\theta} - U_{i,t-1}'\breve{\gamma} \right)$ under the identification restriction that $E(\alpha_i) = E(\varepsilon_{it}) = 0$. Then we have two natural test statistics

$$\Gamma_{NT,1} = \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} \left[\hat{m}(U_{i,t-1}) - \breve{\upsilon} - U_{i,t-1}'\breve{\gamma} \right]^2 a(U_{i,t-1}),$$

and

$$\Gamma_{NT,2} = \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} \left[\hat{\theta}' Z_{it} + \hat{m}(U_{i,t-1}) - Z_{it}'\breve{\theta} - \breve{\upsilon} - U_{i,t-1}'\breve{\gamma} \right]^2 a(U_{i,t-1}),$$

where $a(\cdot)$ is a nonnegative weight function, $\hat{\theta}$ is either one of the two estimates ($\hat{\theta}_{gmm}$ and $\hat{\theta}_{\text{sieve}}$) discussed above, $\hat{m}$ is either one of the four estimates: $\hat{m}_{gmm,iter}$, $\hat{m}_{gmm,noniter}$, $\hat{m}_{s,iter}$, and $\hat{m}_{s,noniter}$. We do not recommend the use of $\hat{m}_{sieve}$ because simulations indicate that it tends to be outperformed by the other four estimates.

Note that all parametric estimates have the usual parametric convergence rate under the null hypotheses. Following the asymptotic results in Sections 2 and 3 and the analysis in Su and Lu (2013), one can easily show that $\Gamma_{NT,1}$ and $\Gamma_{NT,2}$ are asymptotically equivalent under the null, and after being suitably normalized, they share the same asymptotic normal null distribution as the corresponding test statistic Γ_{NT} in Su and Lu (2013):

$$\left(NT_1 (b!)^{1/2} \Gamma_{NT,s} - \mathbb{B}_{NT} \right) \xrightarrow{D} N\left(0, \sigma_0^2\right) \text{ for } s = 1, 2 \text{ under } \mathbb{H}_0,$$

where $\mathbb{B}_{NT}$ and σ_0^2 are as defined in Su and Lu (2013, eqns. (3.7) and (3.8)). Under $\mathbb{H}_0$, we can consistently estimate them by $\hat{\mathbb{B}}_{NT}$ and $\hat{\sigma}_{NT}^2$ whose formulae are also given in the latter paper. The feasible test statistics are then given by

$$J_{NT,s} \equiv \left(NT_1 (b!)^{1/2} \Gamma_{NT,s} - \hat{\mathbb{B}}_{NT} \right) / \sqrt{\hat{\sigma}_{NT}^2}, \; s = 1, 2,$$

which are asymptotically distributed as $N(0,1)$ under the null. Following Su and Lu (2013), one can also study the local power properties of $J_{NT,1}$ and $J_{NT,2}$ and demonstrate that they can detect local alternatives converging to the null at the usual nonparametric rate $(NT_1)^{-1/2}(b!)^{-1/4}$. The global consistency of the tests can also be established, following almost identical arguments as used in Su and Lu (2013). We omit the details to save space.

Remark 4.1. (Specification test for the partially linear model) One might also be interested in testing for the correct specification of the partially linear model. For simplicity, assume that Z_{it} is also sequentially exogenous and $E(\varepsilon_{it}|Z_{it}, U_{i,t-1}, \alpha_i) = 0$. In this case, the null and alternative hypotheses are

$$\mathbb{H}_0' : E(Y_{it}|Z_{it}, U_{i,t-1}, \alpha_i) = Z_{it}'\theta_0 + m_0(U_{i,t-1}) + \alpha_i \text{ a.s. for some } \theta_0 \in \Theta \in \mathbb{R}^{d_z} \text{ and } m_0 \in \mathcal{M}$$

and

$$\mathbb{H}_1' : \Pr[E(Y_{it}|Z_{it}, U_{i,t-1}, \alpha_i) = Z_{it}'\theta + m(U_{i,t-1}) + \alpha_i] < 1 \text{ for all } \theta \in \Theta \in \mathbb{R}^{d_z} \text{ and } m \in \mathcal{M},$$

where $\mathcal{M}$ is certain class of smooth functions. There are several possible ways to conduct the tests. For example, one may follow Fan and Li (1996) and construct a residual-based kernel-smooth test. For another example, one may follow Delgado and González-Manteiga (2001) and Li et al. (2003) and consider a residual-based non-smooth test via the use of empirical process theory. Alternatively, one may consider a weighted square-distance-based test by comparing the semiparametric estimate of the regression function under the null with the purely nonparametric estimate under the alterative. Such ideas have been frequently applied in the nonparametric literature to construct specification tests for parametric or semiparametric models in either cross section or time series studies. A major problem that pertains to our panel data model is that the fixed effects α_i's cannot be consistently estimated in the case of fixed T and thus it is difficult to construct a test based on the estimated residuals. That motivates Su and Lu (2013) to take the third approach mentioned above. Apparently, this approach is also problematic if $d_z + d_x + 1$, the dimension of $(Z_{it}, U_{i,t-1})$, is high because it is then hard to estimate the nonparametric function under the alternative. Of course, other approaches to specification tests are also typically subject to the "curse of dimensionality". It remains an open question how to derive a practical test in this framework and we leave it for future research.

Remark 4.2. (A bootstrap version of the test) It is well known that nonparametric tests based on their asymptotic normal null distributions may perform poorly in finite samples. As an alternative, people frequently rely on bootstrap p-values to make inference. Below we propose a recursive bootstrap procedure to obtain the bootstrap p-values for our test. Let J_{NT} be one of $J_{NT,1}$ and $J_{NT,2}$ defined above. The procedure goes as follows:

1. Estimate the restricted model under $\mathbb{H}_0$ and obtain the residuals $\breve{\varepsilon}_{it} = Y_{it} - Z_{it}'\breve{\theta} - \breve{\upsilon} - (Y_{i,t-1}, X_{it}')\breve{\gamma}$, where $\breve{\theta}$, $\breve{\upsilon}$, and $\breve{\gamma}$ are the IV or GMM estimates of θ, υ, and θ, under the null. Calculate the test statistic J_{NT} based on the original sample $\{Y_{it}, X_{it}, Z_{it}, V_{it}\}$. Let $\breve{\alpha}_i \equiv \overline{\breve{\varepsilon}}_i \equiv T^{-1}\sum_{t=1}^{T} \breve{\varepsilon}_{it}$.

2. Obtain the bootstrap error $\varepsilon_{it}^* = \left(\breve{\varepsilon}_{it} - \overline{\breve{\varepsilon}}_{\cdot t}\right)\epsilon_{it}$ for $i = 1, 2, \ldots, N$ and $t = 2, \ldots, T$, where $\overline{\breve{\varepsilon}}_{\cdot t} \equiv N^{-1}\sum_{i=1}^{N} \breve{\varepsilon}_{it}$ and ϵ_{it}'s are IID across both i and t and follow a two-point distribution: $\epsilon_{it} =$

$(1-\sqrt{5})/2$ with probability $(1+\sqrt{5})/2\sqrt{5}$ and $\epsilon_{it} = (\sqrt{5}+1)/2$ with probability $(\sqrt{5}-1)/2\sqrt{5}$. Generate the bootstrap analogue Y_{it}^* of Y_{it} as

$$Y_{it}^* = Z_{it}'\breve{\theta} + \breve{\upsilon} + \left(Y_{i,t-1}^*, X_{it}'\right)\breve{\gamma} + \breve{\alpha}_i + \varepsilon_{it}^* \text{ for } i = 1, 2, \ldots, N \text{ and } t = 2, \ldots, T,$$

where $Y_{i1}^* = Y_{i1}$.

3. Given the bootstrap resample $\{Y_{it}^*, X_{it}, Z_{it}, V_{it}\}$, estimate both the restricted (linear) and unrestricted (semi-parametric) first-differenced model and calculate the bootstrap test statistic J_{NT}^* analogously to J_{NT}.

4. Repeat steps 2 and 3 for B times and index the bootstrap test statistics as $\{J_{NT,l}^*\}_{l=1}^B$. The bootstrap p-value is calculated by $p^* \equiv B^{-1}\sum_{l=1}^B \mathbf{1}(J_{NT,l}^* > J_{NT})$.

Note that we impose the null hypothesis of linear dynamic panel data models in step 2. Conditional on the data, $(Y_{it}^*, \varepsilon_{it}^*)$ are independently but not identically distributed (INID) across i, and ε_{it}^* are also independently distributed across t. So we need to resort to the CLT for second order U-statistics with INID data (e.g., de Jong (1987)) to justify the asymptotic validity of the above bootstrap procedure. See Su and Lu (2013) for more discussions.

5 Simulations

In this section, we conduct a small set of Monte Carlo simulations to examine the finite sample performance of our proposed estimators and test statistics.

5.1 Data generating processes

We first consider the following three data generating processes (DGPs) where both Z_{it} and X_{it} are strictly exogenous:

DGP 1: $Y_{it} = 0.5Z_{it} + 0.25Y_{i,t-1} + X_{it} + \alpha_i + \varepsilon_{it}$;
DGP 2: $Y_{it} = 0.5Z_{it} + \phi\left(Y_{i,t-1}\right) + X_{it}^2 + \alpha_i + \varepsilon_{it}$;
DGP 3: $Y_{it} = 0.5Z_{it} + \phi(Y_{i,t-1} - Y_{i,t-1}^2)\left[1.5 + \phi(X_{it})\right] + \alpha_i + \varepsilon_{it}$;

where $\phi\left(\cdot\right)$ is the standard normal PDF, ε_{it} are IID $N\left(0,1\right)$ across both i and t, α_i are IID $U\left(-1/2, 1/2\right)$, $X_{it} = 0.25\alpha_i + \varepsilon_{x,it}$ with $\{\varepsilon_{x,it}\}$ being IID $N\left(0,1\right)$ across both i and t and independent of $\{\alpha_i\}$ and $\{\varepsilon_{it}\}$, $Z_{it} = 0.25\alpha_i + \varepsilon_{z,it}$ with $\{\varepsilon_{z,it}\}$ being IID $N\left(0,1\right)$ and independent of $\{\alpha_i, \varepsilon_{x,it}, \varepsilon_{it}\}$. DGP 1 is a linear dynamic panel data model whereas DGPs 2-3 specify partially linear dynamic panel data models. The lagged dependent variable $Y_{i,t-1}$ and regressor X_{it} enter the model additively in DGP 2 and multiplicatively in DGP 3.

Next, we consider another three DGPs with exogenous X_{it} but endogenous Z_{it} :

DGP 4: $Y_{it} = 0.5Z_{it} + 0.25Y_{i,t-1} + X_{it} + \alpha_i + \varepsilon_{it}$;
DGP 5: $Y_{it} = 0.5Z_{it} + \phi\left(Y_{i,t-1}\right) + X_{it}^2 + \alpha_i + \varepsilon_{it}$;
DGP 6: $Y_{it} = 0.5Z_{it} + \phi(Y_{i,t-1} - Y_{i,t-1}^2)\left[1.5 + \phi(X_{it})\right] + \alpha_i + \varepsilon_{it}$;

where the regression functions parallel those in DGPs 1-3 in order, $\{\alpha_i, X_{it}\}$ are generated in the same way as in DGPs 1-3, the endogenous regressor Z_{it} is generated according to

$$Z_{it} = 0.25\alpha_i + V_{it} + \varepsilon_{z,it},$$

$(\varepsilon_{it}, \varepsilon_{z,it})$ are IID draws from $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}\right)$, and V_{it} are IID $N(0,1)$ and independent of $\{\varepsilon_{it}, \varepsilon_{z,it}\}$.

Obviously, m is defined as follows: $m(y,x) = 0.25y + x$ in DGPs 1 and 4, $m(y,x) = \phi(y) + x^2$ in DGPs 2 and 5, and $m(y,x) = \phi(y - y^2)[1.5 + \phi(x)]$ in DGPs 3 and 6. All the six DGPs are used to evaluate the performance of our estimates and tests for the linearity of the function $m(\cdot)$. In the case of testing for the nonlinearity of $m(\cdot)$, DGPs 1 and 4 are used for the size study and all the other four DGPs are for power comparisons.

5.2 Implementation

In the estimation, we consider the following estimates for θ and $m(\cdot)$:

1. $\hat{\theta}_{gmm,iter}$ and $\hat{m}_{gmm,iter}$: semiparametric GMM estimate of θ and local quadratic estimate of m, respectively; both are based on the local quadratic estimates of $m_y(u)$ and $m_z(u)$ via the iterative algorithm. We choose $W_{it} = (\Delta Z_{it}, U'_{i,t-2})'$ in DGPs 1-3 and $W_{it} = (V_{i,t-1}, U'_{i,t-2})'$ in DGPs 4-6.

2. $\hat{\theta}_{gmm,noniter}$ and $\hat{m}_{gmm,noniter}$: semiparametric GMM estimate of θ and local quadratic estimate of m, respectively; both are based on the local quadratic estimates of $m_y(u)$ and $m_z(u)$ via the non-iterative algorithm. We choose $W_{it} = (\Delta Z_{it}, U'_{i,t-2})'$ in DGPs 1-3 and $W_{it} = (V_{i,t-1}, U'_{i,t-2})'$ in DGPs 4-6.

3. $\hat{\theta}_{sieve}$ and $\hat{m}_{sieve}$: sieve IV estimates of θ and m, respectively. We choose $\bar{W}_{it} = (\Delta Z_{it}, p'_{i,t-2})'$ in DGPs 1-3 and $\bar{W}_{it} = (V_{i,t-1}, p'_{i,t-2})'$ in DGPs 4-6.

4. $\hat{m}_{s,iter}$ and $\hat{m}_{s,noniter}$: local quadratic estimates of m by finding the empirical solution to eqn. (3.3) through the iterative and non-iterative algorithms, respectively. See remark 7 in Section 3.

5. $\hat{\theta}_{s,iter}$ and $\hat{\theta}_{s,noniter}$: IV or OLS estimates of θ by running the regression in (3.4) and (3.5), respectively. See remark 8 in Section 3 for more details.

To obtain $\hat{\theta}_{gmm,iter}$ or $\hat{\theta}_{gmm,noniter}$, we consider the semiparametric 2SLS estimates of θ by setting $A_{NT} = (\frac{1}{\tilde{n}}\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}$. Note that when a local polynomial regression is called upon, we always apply the local quadratic regression so that $q = 2$. For all iterative local quadratic estimates of the nonparametric components, we follow Su and Lu (2013) and choose the sieve-IV estimates as the initial estimates. For the sieve estimates, we choose the cubic B-spline as the sieve basis and include the tensor product terms to approximate the bivariate function $m(y,x)$. Along each dimension of the covariate in $m(\cdot)$, we let $L_0 = \lfloor (NT_2)^{1/4} \rfloor + 1$ and choose L_0 sieve approximating terms, where $\lfloor a \rfloor$ denotes the integer part of a.[5]

[5] We also apply $L_0 = \lfloor c(NT_2)^{1/4} \rfloor + 1$ for different values of c (0.5, 1, 2) and find our iterative estimates are quite robust to the choice of c.

For the convergence criterion in the iterative algorithms, we terminate the procedure if

$$\frac{\sum_{j=1}^{J}\left[m_a^{(l+1)}(u_j)-m_a^{(l)}(u_j)\right]^2}{\sum_{j=1}^{J}\left[m_a^{(l)}(u_j)\right]^2+0.0001}<0.001,\ a=y,\ z_1,...,z_{d_z},$$

where u_j, $j=1,...,J$, are the J evaluation points. In practice, researchers can choose the evaluation points on their own. Here we let the number of evaluation points be 625, with 25 grid points along each dimension. For each DGP, the evaluation points are fixed across replications and approximately evenly distributed between 0.2 quantile and 0.8 quantile of the data points. A similar convergence criterion is used in the literature; see Nielsen and Sperlich (2005), Henderson et al. (2008), Mammen et al. (2009), and Su and Lu (2013), to name just a few. In the specification test, we choose the data points on $\mathcal{U}$ as the evaluation points.

For both estimation and testing, we need to choose the kernel function and the bandwidth sequence. We use the Epanechnikov kernel $k(z)=0.75\left(1-z^2\right)\mathbf{1}(|z|\leq 1)$ and choose the bandwidths by Silverman's "rule of thumb": $h=2.35S_U(NT_2)^{-1/[2(q+2)-0.5]}$ and $b=2.35S_U(NT_2)^{-1/[2(q+2)+d_x+1]}$ where $d_x=1$, $q=2$, $S_U=(S_Y,S_X)$, and S_Y and S_X denote the sample standard deviation of $\{Y_{i,t-1}\}$ and $\{X_{it}\}$, respectively. Note that h satisfies the undersmoothing requirement, and b has the optimal rate for local quadratic regressions. Admittedly, these choices of bandwidth sequences are usually not optimal for either the estimation or the testing issue. There is a room to improve the performance of our estimates and tests by developing a data-driven rule for the selection of "optimal" bandwidth sequences. It is well known that a bandwidth that is optimal for the estimation problem is typically not optimal for the testing problems. One has to separately consider these two issues. Given the complication of our kernel estimates, we do not address these issues in this paper and leave them for future research.

In the sieve IV estimates discussed in Section 3, we also use cubic B-spline, choose $L_0=\lfloor (NT_2)^{1/4}\rfloor+1$ sieve approximating terms along each dimension of the covariates in $m(\cdot)$, and include the tensor product terms in the approximation.

Also, we need to choose the compact set $\mathcal{U}$. For this, we trim out the data on the two-sided 5% tails along each dimension in $U_{i,t-1}$ or $U_{i,t-2}$. This trimming is also used for implementing our tests, in which case the weight function $a(\cdot)$ is an indicator function: it takes value 1 if $U_{i,t-1}$ lies in $\mathcal{U}$ and 0 otherwise. We only consider the testing results for $J_{NT,1}$ defined in Section 4 by comparing the semiparametric estimates of $m(\cdot)$ with its Anderson-Hsiao-type IV estimate under the linear null hypothesis. Given the poor performance of $\hat{m}_{sieve}$ in simulations, we only consider the other four estimates of $m(\cdot)$: $\hat{m}_{gmm,iter}$, $\hat{m}_{gmm,noniter}$, $\hat{m}_{s,iter}$, and $\hat{m}_{s,noniter}$, and denote the corresponding J-test statistics as $J_{gmm,iter}$, $J_{gmm,noniter}$, $J_{s,iter}$, and $J_{s,noniter}$, respectively.

For the (N,T) pair, we consider $N=25$, 50, 100 and $T=4$, 6. For each scenario, the number of replications is 1000 for the estimation. For the test, we use 500 replications and 200 bootstrap resamples for the size study and 250 replications and 200 bootstrap resamples for the power study.

5.3 Estimation and testing results

Table 1 reports the bias (Bias) and root mean squared error (RMSE) of various estimates of θ. We summarize the findings from Table 1. First, for all DGPs under investigation, the RMSEs decrease as

Table 1. Estimation results for θ

			$\hat{\theta}_{gmm,iter}$		$\hat{\theta}_{gmm,noniter}$		$\hat{\theta}_{sieve}$		$\hat{\theta}_{s,iter}$		$\hat{\theta}_{s,noniter}$	
DGP	T	N	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	4	25	-0.035	0.195	-0.033	0.195	-0.089	0.236	-0.043	0.185	-0.041	0.184
		50	-0.027	0.136	-0.026	0.137	-0.049	0.150	-0.028	0.129	-0.028	0.129
		100	-0.035	0.097	-0.035	0.097	-0.024	0.094	-0.034	0.095	-0.095	0.095
	6	25	-0.026	0.134	-0.025	0.134	-0.043	0.147	-0.029	0.128	-0.028	0.128
		50	-0.027	0.094	-0.027	0.094	-0.019	0.095	-0.027	0.090	-0.027	0.090
		100	-0.029	0.070	-0.028	0.069	-0.015	0.066	-0.027	0.067	-0.027	0.067
2	4	25	0.009	0.285	0.006	0.266	-0.071	0.239	-0.000	0.260	-0.002	0.242
		50	0.007	0.201	0.007	0.189	-0.035	0.155	0.002	0.187	0.004	0.176
		100	0.003	0.134	0.001	0.122	-0.014	0.097	0.001	0.129	0.001	0.118
	6	25	0.005	0.203	0.002	0.190	-0.027	0.154	0.003	0.188	0.000	0.177
		50	0.009	0.140	0.004	0.129	-0.001	0.098	0.007	0.132	0.005	0.123
		100	0.002	0.095	-0.003	0.088	-0.004	0.065	0.001	0.091	0.002	0.086
3	4	25	0.000	0.181	0.001	0.184	-0.099	0.234	-0.015	0.174	-0.016	0.174
		50	0.006	0.127	0.006	0.130	-0.061	0.158	0.000	0.123	-0.001	0.124
		100	0.000	0.086	0.000	0.086	-0.030	0.100	-0.001	0.085	-0.002	0.085
	6	25	0.004	0.132	0.004	0.132	-0.047	0.155	-0.001	0.127	-0.001	0.128
		50	0.003	0.088	0.002	0.088	-0.022	0.096	0.000	0.085	0.001	0.085
		100	0.000	0.061	0.001	0.061	-0.013	0.067	0.000	0.060	0.000	0.060
4	4	25	-0.075	0.297	-0.070	0.298	0.059	0.180	0.095	0.159	0.096	0.161
		50	-0.065	0.197	-0.062	0.194	0.051	0.137	0.081	0.129	0.082	0.129
		100	-0.068	0.145	-0.067	0.146	0.030	0.104	0.036	0.095	0.036	0.095
	6	25	-0.060	0.182	-0.058	0.182	0.072	0.147	0.086	0.132	0.087	0.132
		50	-0.051	0.133	-0.050	0.134	0.035	0.104	0.046	0.099	0.045	0.099
		100	-0.059	0.104	-0.058	0.103	0.019	0.073	0.017	0.071	0.018	0.072
5	4	25	0.014	0.371	0.011	0.367	0.080	0.183	0.145	0.237	0.143	0.228
		50	-0.002	0.246	0.002	0.230	0.078	0.149	0.122	0.195	0.129	0.193
		100	0.003	0.172	-0.004	0.162	0.045	0.110	0.082	0.157	0.092	0.163
	6	25	0.011	0.243	0.011	0.229	0.091	0.156	0.123	0.201	0.134	0.202
		50	0.011	0.171	0.009	0.160	0.059	0.115	0.102	0.170	0.111	0.174
		100	0.004	0.115	0.000	0.106	0.033	0.078	0.064	0.124	0.073	0.130
6	4	25	0.009	0.236	0.007	0.244	0.056	0.171	0.135	0.185	0.134	0.185
		50	0.000	0.161	0.001	0.159	0.047	0.137	0.120	0.154	0.120	0.154
		100	0.001	0.111	0.000	0.112	0.029	0.105	0.083	0.117	0.081	0.116
	6	25	0.014	0.151	0.014	0.152	0.075	0.148	0.128	0.163	0.127	0.163
		50	0.007	0.108	0.006	0.108	0.042	0.105	0.092	0.121	0.091	0.120
		100	-0.002	0.073	-0.002	0.073	0.020	0.074	0.063	0.087	0.062	0.086

Table 2. Estimation results for m (RMSE)

			$\hat{m}_{gmm,iter}$		$\hat{m}_{gmm,noniter}$		$\hat{m}_{sieve}$		$\hat{m}_{s,iter}$		$\hat{m}_{s,noniter}$	
DGP	T	N	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean
1	4	25	0.410	0.390	0.392	0.369	0.751	0.732	0.399	0.381	0.408	0.384
		50	0.320	0.304	0.310	0.295	0.625	0.614	0.311	0.295	0.320	0.307
		100	0.250	0.241	0.246	0.236	0.444	0.437	0.245	0.235	0.253	0.241
	6	25	0.306	0.289	0.294	0.278	0.633	0.618	0.297	0.284	0.298	0.285
		50	0.243	0.233	0.235	0.224	0.441	0.439	0.236	0.227	0.239	0.229
		100	0.190	0.184	0.186	0.179	0.362	0.354	0.186	0.179	0.189	0.184
2	4	25	0.707	0.680	0.438	0.412	0.712	0.690	0.459	0.432	0.457	0.430
		50	0.621	0.602	0.324	0.311	0.611	0.590	0.353	0.336	0.341	0.320
		100	0.543	0.540	0.244	0.231	0.437	0.429	0.278	0.264	0.254	0.239
	6	25	0.594	0.581	0.320	0.305	0.608	0.597	0.344	0.329	0.323	0.305
		50	0.515	0.504	0.240	0.228	0.434	0.423	0.271	0.258	0.247	0.234
		100	0.454	0.449	0.179	0.173	0.359	0.356	0.211	0.203	0.184	0.178
3	4	25	0.381	0.364	0.377	0.351	0.703	0.679	0.378	0.356	0.380	0.358
		50	0.302	0.291	0.301	0.288	0.597	0.578	0.304	0.291	0.306	0.296
		100	0.236	0.227	0.234	0.227	0.440	0.437	0.235	0.230	0.236	0.229
	6	25	0.293	0.280	0.292	0.276	0.604	0.589	0.295	0.282	0.292	0.281
		50	0.231	0.220	0.230	0.221	0.432	0.423	0.232	0.225	0.231	0.223
		100	0.181	0.174	0.182	0.175	0.360	0.353	0.183	0.175	0.184	0.177
4	4	25	0.397	0.375	0.404	0.380	0.727	0.698	0.388	0.370	0.395	0.375
		50	0.313	0.296	0.312	0.296	0.625	0.613	0.304	0.292	0.312	0.301
		100	0.243	0.234	0.245	0.238	0.445	0.439	0.238	0.233	0.245	0.238
	6	25	0.302	0.290	0.300	0.286	0.623	0.611	0.291	0.277	0.292	0.277
		50	0.239	0.232	0.236	0.230	0.431	0.423	0.232	0.224	0.236	0.236
		100	0.190	0.183	0.189	0.181	0.365	0.360	0.185	0.177	0.188	0.181
5	4	25	0.717	0.690	0.455	0.425	0.704	0.672	0.450	0.431	0.453	0.425
		50	0.633	0.617	0.338	0.318	0.613	0.602	0.348	0.335	0.342	0.325
		100	0.544	0.536	0.254	0.242	0.441	0.427	0.278	0.268	0.259	0.246
	6	25	0.601	0.592	0.334	0.315	0.603	0.589	0.341	0.325	0.328	0.312
		50	0.523	0.518	0.248	0.238	0.424	0.415	0.268	0.256	0.250	0.239
		100	0.461	0.454	0.184	0.178	0.357	0.354	0.212	0.205	0.186	0.179
6	4	25	0.367	0.352	0.381	0.370	0.710	0.687	0.369	0.353	0.372	0.363
		50	0.288	0.276	0.295	0.290	0.622	0.595	0.292	0.280	0.294	0.281
		100	0.223	0.216	0.232	0.223	0.428	0.421	0.230	0.221	0.231	0.224
	6	25	0.287	0.270	0.296	0.281	0.616	0.606	0.292	0.272	0.288	0.270
		50	0.228	0.219	0.233	0.222	0.433	0.426	0.232	0.220	0.232	0.217
		100	0.179	0.174	0.181	0.177	0.359	0.354	0.181	0.176	0.181	0.177

either N or T increases and are roughly halved as N is quadrupled. Second, for the linear DGPs (DGPs 1 and 4), $\hat{\theta}_{s,iter}$ and $\hat{\theta}_{s,noniter}$ tend to have smaller RMSEs than $\hat{\theta}_{sieve}$, $\hat{\theta}_{gmm,iter}$, and $\hat{\theta}_{gmm,noniter}$. Third, for nonlinear DGPs, $\hat{\theta}_{sieve}$ usually has the smallest RMSE but largest bias among all estimates under study, $\hat{\theta}_{s,iter}$ and $\hat{\theta}_{s,noniter}$ tend to have smaller RMSE than $\hat{\theta}_{gmm,iter}$ and $\hat{\theta}_{gmm,noniter}$ when N is small.

Table 2 presents the median and mean RMSEs for the estimates of the nonparametric component m. Clearly, we can see that both median and mean RMSEs of $\hat{m}_{sieve}$ are much larger than those of other estimates. $\hat{m}_{s,iter}$ and $\hat{m}_{s,noniter}$ can improve the original sieve IV estimate $\hat{m}_{sieve}$ significantly. As expected, $\hat{m}_{s,iter}$ and $\hat{m}_{s,noniter}$ have similar performance as $\hat{m}_{gmm,iter}$ and $\hat{m}_{gmm,noniter}$ for most DGPs. Note that in DGPs 2 and 5, both $\hat{m}_{gmm,iter}$ and $\hat{m}_{s,iter}$ have much larger median and mean RMSEs than their non-iterative analogues, which may be caused by the bad convergence of the iterative algorithm. In empirical applications, both estimates based on iterative and noniterative algorithms can be reported. When there is a large difference between the two estimates, we recommend the use of the noniterative algorithm when NT_2 is not excessively large (e.g., $NT_2 < 1000$) and that of the iterative algorithm otherwise.

Table 3 gives the empirical rejection frequency for our proposed tests. From this table, we can see that the levels behave reasonably well for DGPs 1 and 4. All the four tests are slightly undersized for DGP 1 and oversized for DGP 4. DGPs 2-3 and 5-6 examine the empirical power of our tests. The powers of all four tests are reasonably good.

6 An empirical application: the impact of IPR protection on economic growth

6.1 Motivation

In this section, we apply the partially linear dynamic panel data model to study the classical question of how intellectual property right (IPR) protection affects economic growth. IPRs, as the rights to use and sell knowledge and inventions, with the aim of guaranteeing adequate returns for innovators and creators, has played a central role in the long-standing debates concerning economic policy. In forming a decision on how to protect IPR, there is a typical trade-off: if IPR protection is stronger, only the owner of knowledge design will use it and the impact on economic growth will be smaller; if IPR protection is weaker, the diffusion or transfer of knowledge and technology will be accelerated and the adopters will benefit without paying adequate costs. This could cause higher economic growth, but at the same time weaken the incentive of innovation and then reduce the growth enhancement. Many theoretic growth models have discussed this topic, but the conclusions are ambiguous. Some advocate stronger IPR protection reform and the others oppose to this. For instance, Dinopoulos and Segerstrom (2010) develop an endogenous economic growth model to evaluate the effect of stronger IPR protection in developing countries, and conclude that stronger IPR protection in the South promotes innovation in the global economy and explains faster growth rate of some developing countries. In some North-South trade models, Branstetter et al. (2006) and Glass and Wu (2007) support a similar view that patent reform increases the economic growth rate permanently. However, Furukawa (2007) proves that IPR cannot increase economic growth in an endogenous growth model with costless imitation, whereas Eicher

Table 3: Empirical rejection frequency

			$J_{s,iter}$			$J_{s,noniter}$			$J_{gmm,iter}$			$J_{gmm,noniter}$		
DGP	T	N	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	4	25	0.008	0.042	0.074	0.012	0.036	0.066	0.008	0.034	0.056	0.010	0.036	0.064
		50	0.012	0.032	0.060	0.008	0.030	0.056	0.008	0.026	0.038	0.006	0.032	0.070
		100	0.008	0.040	0.058	0.002	0.024	0.060	0.008	0.028	0.060	0.002	0.020	0.062
	6	25	0.012	0.040	0.060	0.010	0.040	0.080	0.016	0.036	0.062	0.008	0.042	0.086
		50	0.006	0.042	0.084	0.010	0.044	0.084	0.008	0.050	0.092	0.012	0.036	0.080
		100	0.010	0.058	0.088	0.014	0.046	0.086	0.014	0.036	0.086	0.018	0.036	0.078
2	4	25	0.020	0.116	0.192	0.100	0.268	0.392	0.004	0.024	0.060	0.076	0.236	0.376
		50	0.080	0.276	0.472	0.364	0.592	0.696	0.024	0.056	0.132	0.316	0.556	0.696
		100	0.380	0.676	0.792	0.804	0.932	0.980	0.064	0.192	0.312	0.804	0.952	0.980
	6	25	0.168	0.396	0.584	0.464	0.708	0.840	0.112	0.260	0.400	0.416	0.716	0.836
		50	0.376	0.712	0.848	0.932	0.992	1.000	0.308	0.542	0.692	0.900	0.992	1.000
		100	0.572	0.840	0.996	0.992	0.996	1.000	0.760	0.920	0.956	1.000	1.000	1.000
3	4	25	0.024	0.100	0.184	0.016	0.056	0.128	0.032	0.100	0.204	0.020	0.052	0.120
		50	0.104	0.212	0.300	0.052	0.160	0.268	0.104	0.232	0.316	0.048	0.164	0.244
		100	0.152	0.288	0.392	0.116	0.224	0.312	0.136	0.276	0.416	0.124	0.228	0.292
	6	25	0.080	0.200	0.296	0.052	0.168	0.288	0.088	0.220	0.336	0.028	0.144	0.264
		50	0.144	0.288	0.420	0.112	0.248	0.360	0.136	0.320	0.420	0.096	0.232	0.316
		100	0.360	0.584	0.688	0.284	0.528	0.612	0.416	0.572	0.708	0.244	0.420	0.624
4	4	25	0.032	0.070	0.130	0.026	0.066	0.116	0.034	0.070	0.136	0.028	0.078	0.120
		50	0.018	0.046	0.088	0.020	0.052	0.102	0.020	0.068	0.114	0.018	0.062	0.110
		100	0.018	0.066	0.100	0.014	0.054	0.104	0.014	0.048	0.110	0.028	0.074	0.126
	6	25	0.024	0.060	0.114	0.024	0.074	0.102	0.032	0.078	0.116	0.024	0.060	0.122
		50	0.016	0.050	0.122	0.014	0.058	0.098	0.020	0.060	0.104	0.020	0.062	0.112
		100	0.012	0.060	0.112	0.020	0.072	0.108	0.022	0.078	0.116	0.016	0.076	0.124
5	4	25	0.020	0.108	0.236	0.076	0.240	0.364	0.012	0.024	0.056	0.084	0.232	0.312
		50	0.100	0.288	0.432	0.280	0.512	0.688	0.020	0.072	0.124	0.244	0.484	0.624
		100	0.276	0.576	0.728	0.700	0.868	0.928	0.048	0.128	0.248	0.656	0.860	0.932
	6	25	0.040	0.204	0.396	0.396	0.628	0.752	0.012	0.084	0.184	0.324	0.588	0.732
		50	0.188	0.424	0.600	0.824	0.972	0.984	0.088	0.240	0.328	0.792	0.932	0.972
		100	0.544	0.856	0.904	0.980	1.000	1.000	0.268	0.532	0.724	0.984	1.000	1.000
6	4	25	0.052	0.108	0.176	0.036	0.076	0.104	0.044	0.092	0.144	0.020	0.068	0.108
		50	0.064	0.180	0.284	0.056	0.144	0.236	0.104	0.208	0.316	0.044	0.120	0.224
		100	0.148	0.292	0.472	0.096	0.220	0.328	0156	0.304	0.424	0.112	0.216	0.340
	6	25	0.048	0.192	0.288	0.056	0.140	0.244	0.088	0.208	0.276	0.036	0.152	0.248
		50	0.196	0.380	0.496	0.132	0.344	0.448	0.196	0.400	0.536	0.116	0.260	0.420
		100	0.388	0.708	0.776	0.336	0.592	0.700	0.384	0.636	0.776	0.312	0.544	0.672

and Garcia-Peñalosa (2008) show that the relationship between IPR and economic growth is ambiguous. From another side, there is also no conclusive result from a large body of empirical studies. For example, Chen and Puttitanun (2005) find a positive effect for developing countries, Park and Ginarte (1997) and Kanwar and Evenson (2003) discover a general positive effect; Groizard (2009) and Falvey et al. (2009) find an ambiguous relationship. In addition, there are also some studies on the nonlinear relationship between IPR protection and economic growth; see Furukawa (2007) and Panagopoulos (2009) for an inverted-U relationship, Chen and Puttitanun (2005) for a U-shape relationship between optimal IPR and economic growth for developing countries, and Falvey et al. (2009) for a nonlinear relationship which depends on other variables such as the level of development, the imitative ability and the market size of the importing country. In summary, from the theoretical or empirical point of view, there is no conclusive relationship between IPR and economic growth. Note that most of the empirical studies on this question use linear models and exclude dynamic nature. In this section, we reinvestigate this topic using our proposed partially linear dynamic panel data models that allow general nonlinearity of unknown form for the lagged dependent variable and IPR and the usual linearity of control variables.

6.2 Data and variables

The data set includes 93 countries or regions for the years 1975-2005. We examine the five-year economic growth rate. Let $\Delta \ln GDP_{it} = \ln GDP_{i,s_t} - \ln GDP_{i,s_t-5}$ denote the growth rate of country i over the t-th five-year period, where GDP_{i,s_t} and GDP_{i,s_t-5} denote the real GDP per capita for country i in the end year s_t and the starting year $s_t - 5$, respectively. For example, $s_1 = 1980$ and $\Delta \ln GDP_{i1} = \ln GDP_{i,1980} - \ln GDP_{i,1975}$. We include two more sets of regressors other than the lagged growth rate ($\Delta \ln GDP_{i,t-1}$). The first set includes IPR (IPR_{it}), the variable of our main interest, measured by the updated Ginarte-Park index of patent rights (Park, 2008), which enters the model together with $\Delta \ln GDP_{i,t-1}$ nonparametrically. The second set includes linear control variables $(Z_{1,it}, \cdots, Z_{9,it}) = (FDI_{it}, FDI_{it}^2, Schooling_{it}, Invest_{it}, GonvC_{it}, Institutions_{it}, Openness_{it}, POP_{it}, INF_{it})$, where FDI_{it} denotes the inward FDI inflows as a share of GDP and FDI_{it}^2 is its squared term, $Schooling_{it}$ denotes the average of human capital measured as percentage of secondary school enrollment in total population, $Invest_{it}$ denotes the domestic investment measured by gross capital formation as a share of GDP, $GonvC_{it}$ denotes the general government final consumption expenditure as a share of GDP, $Institutions_{it}$ is a measure of market distortions, as proxied by the Fraser Institute' Index of Legal Structure and Security of Property Rights, $Openness_{it}$ is measured by imports plus exports as a share of GDP, POP_{it} is the population growth rate, and INF_{it} denotes the inflation measured by percentage change in the GDP deflator. Following the literature on economic growth, we take five-year averages of annual values for IPR_{it} and all the control variables. For the list of countries/regions and the sources of all variables, see the data appendix. Table 4 provides summary statistics on the data set. From Table 4 we see that the five-year GDP growth rates (non-annualized) range from -53.44% to 55.93%. There is also considerable variation in the five-year average IPR with the smallest value 0 and the largest value 4.88. The summary statistics for control variables are also listed in the table.

Table 4. Descriptive statistics (93 countries, 1975-2005)

Variable	Mean	Median	Std deviation	Max	Min
GDP growth	0.0846	0.0884	0.1430	0.5593	-0.5344
IPR	2.5683	2.3850	1.1016	4.8800	0.0000
FDI	0.0183	0.0101	0.0253	0.2079	-0.0335
Domestic investment	0.2116	0.2097	0.0625	0.5182	0.0248
Schooling	0.2766	0.2545	0.1689	0.7523	0.0053
Government consumption	0.1553	0.1494	0.0572	0.4036	0.0409
Population growth rate	0.0177	0.0187	0.0111	0.0631	-0.0464
Institutions	5.4397	5.3800	2.0275	9.6200	1.1400
Openness	5.9992	6.1350	2.2893	9.7200	0.0000
Inflation	0.3942	0.0799	2.6733	48.2871	-0.0673

6.3 Estimation results

We consider the following partially linear dynamic panel data model with fixed effects

$$\begin{aligned}\Delta \ln GDP_{it} &= m\left(\Delta \ln GDP_{i,t-1}, IPR_{it}\right) + \theta_1 FDI_{it} + \theta_2 FDI_{it}^2 + \theta_3 Schooling_{it} + \theta_4 Invest_{it} \\ &\quad + \theta_5 GonvC_{it} + \theta_6 Institutions_{it} + \theta_7 Openness_{it} + \theta_8 POP_{it} + \theta_9 INF_{it} \\ &\quad + \alpha_i + \varepsilon_{it},\end{aligned}$$

$i = 1, ..., 93$, $t = 2, ..., 6$. For both the estimation and testing, we follow the details in Section 5.2 and consider five estimates for the parametric component, five estimates for the nonparametric component, and four test statistics to test the linearity or quadraticity of the nonparametric component. To make inferences on the estimated parameters, we also report the t-values based on bootstrap standard errors (s.e.'s) where we resample the N individual countries with replacement to form 400 bootstrap resamples. The details of implementation of our estimation and tests are the same as in the simulations.

Table 5 presents the semiparametric estimation results for the parametric part and the parametric estimation results based on Anderson-Hsiao estimator. We report five semiparametric estimates, $\hat{\theta}_{sieve}$, $\hat{\theta}_{s,iter}$ $\hat{\theta}_{s,noniter}$, $\hat{\theta}_{gmm,iter}$, and $\hat{\theta}_{gmm,noniter}$ as explained in Section 5.2. The numbers in parentheses are the t-values based on the bootstrap s.e.'s when 400 bootstrap resamples are used. We also report the linear and quadratic estimates $\hat{\theta}_{linear}$ and $\hat{\theta}_{quadratic}$ in the last two columns. Their bootstrap t-values are given in the parentheses based on 400 wild bootstrap resamples with the estimated parametric functional forms being imposed in the bootstrap world. We summarize the main findings as follows. First, we can see that the estimated results are quite robust among all estimating procedures but $\hat{\theta}_{sieve}$. Second, for the semiparametric estimates, only the estimated coefficients for *Investment* and *GonvC* are *significant* at the 5% significance level. Third, most control variables are not significant at the 5% level possibly due to the inclusion of the lagged dependent variable in the model. Forth, there are large differences between parametric and semiparametric estimates for the control variables *FDI*, *FDI*^{2}, *Schooling*, *Institutions*, *Openness* and *POP.* Sixth, we find negative but insignificant effects of IPR protection on economic growth rate in both the linear and quadratic models at the 5% level. Last, we see an insignificant U-shape relationship between IPR and economic growth rate in the quadratic model, which is similar to the findings in Chen and Puttitanun (2005).

In Figure 1, we plot the estimated surface of $m(\cdot,\cdot)$ to present the relationship between the economic growth rate and IPR and the lagged five-year growth rate as well. Apparently, as in the simulation the sieve estimate in Figure 1(a) is quite different from the other four estimates. Subplots (b)-(e) in Figure 1 clearly demonstrate the existence of a nonlinear relationship between economic growth rate and IPR protection and the lagged growth rate. In sharp contrast to the parametric estimation results, it seems that there exists a (basically) positive association between the economic growth rate and the IPR protection, which is conformable with many empirical studies and implies that the positive effects of IPR dominate its negative effects.

Figure 2 present the estimated two-dimensional relationships between economic growth rate and its lag and IPR. Since the sieve estimate is less accurate as shown in the simulation, we only report the other four estimates for the nonparametric component. Figures 2(a), 2(c), and 2(e) present the relationships between economic growth rate and its lag when IPR is fixed at its 0.25, 0.5 and 0.75 sample quantiles, respectively. Specifically, they show the estimates of $m(\cdot;IPR_{0,25})$, $m(\cdot;IPR_{0,5})$, and $m(\cdot;IPR_{0,75})$, where IPR_{α} is the α-th empirical quantile of data $\{IPR_{it}\}$. We summarize the main findings as follows. First, we can see clearly a significant nonlinear relationships between economic growth rate and its lag, and the shapes of $\hat{m}(\cdot;IPR_{\alpha})$, $\alpha=0.25, 0.5$ and 0.75 are drastically different. It seems that the patterns of dynamics in economic growth rate vary with the change of the level of IPR protection. Figures 2(b), 2(d), and 2(f) give the estimated relationships between economic growth rate and IPR when the lagged growth rate is fixed at its 0.25, 0.5 and 0.75 sample quantiles, respectively. From this column, we first find slight nonlinear relationships between economic growth rate and IPR. Second, it seems that IPR protection has a global positive impact on economic growth rate, which contradicts the findings in the parametric models. Third, for all the three estimated curves, their first order derivatives neither monotonically decrease nor increase as IPR increases, which implies the relationships between IPR and economic growth rate is neither inverted-U-shaped nor U-shaped. Figures 3 and 4 report the 90% pointwise confidence bands for the semiparametric GMM and sieve estimates of the nonparametric component, respectively, based on 400 bootstraps. As usual, these bands seem to be wide given the fact the effective number of observations used in our estimation is given by $NT_2 = 93\times(6-2) = 372$, which is not very large.

6.4 Specification test results

Based on our proposed testing procedure, we can test two commonly used linear and quadratic functional form specifications. The null hypotheses for linear and quadratic functional forms can be stated as follows:

$$
\begin{aligned}
H_0 &: m\left(Y_{i,t-1}, IPR_{it}\right) = \rho Y_{i,t-1} + \theta IPR_{it} \text{ for some } (\rho,\theta)\in\mathbb{R}^2, \\
H_0' &: m\left(Y_{i,t-1}, IPR_{it}\right) = \rho Y_{i,t-1} + \theta_1 IPR_{it} + \theta_2 IPR_{it}^2 \text{ for some } (\rho,\theta_1,\theta_2)\in\mathbb{R}^3.
\end{aligned}
$$

Table 6 reports the bootstrap p-values for our test $J_{NT,1}$ for different estimation procedures based on 2000 bootstrap resamples. We can reject either H_0 or H_0' at the 5% significance level for the $J_{s,iter}$, $J_{s,noniter}$ and $J_{noniter}$ statistics and at the 10% significance level for the J_{iter} statistic. In general, we can conclude that the nonparametric component $m(\cdot,\cdot)$ is neither linear nor quadratic in IPR at least at

Table 5. Estimation results of the parametric part

Dependent variable: five-year per capita GDP growth rate							
Var.\Coeff. est.	$\hat{\theta}_{sieve}$	$\hat{\theta}_{s,iter}$	$\hat{\theta}_{s,noniter}$	$\hat{\theta}_{gmm,iter}$	$\hat{\theta}_{gmm,noniter}$	$\hat{\theta}_{linear}$	$\hat{\theta}_{quadratic}$
FDI	0.3882 (0.8692)	0.6254 (1.6878)	0.6417 (1.7245)	0.5176 (1.2643)	0.5574 (1.3236)	2.3772 (4.9943)	2.3875 (4.9363)
FDI^2	−0.5619 (−2.1829)	−0.3876 (−0.9849)	−0.4907 (−1.2886)	−0.4074 (−1.2252)	−0.4876 (−1.5242)	−7.1854 (−2.8423)	−7.3474 (−2.8347)
Schooling	0.0050 (0.0274)	−0.1367 (−0.6889)	−0.0747 (−0.4031)	−0.1245 (−0.6149)	−0.0827 (−0.4157)	0.3041 (2.2252)	0.2987 (2.2043)
Investment	0.9915 (3.3873)	0.8973 (3.5642)	0.9018 (3.4605)	0.8746 (3.1676)	0.9144 (3.3107)	1.0011 (4.3078)	1.0017 (4.2820)
GonvC	−1.0128 (−3.0822)	−1.0427 (−3.4668)	−1.0181 (−3.3812)	−1.0843 (−3.4157)	−0.9291 (−2.9157)	−0.8670 (−2.9374)	−0.8808 (−2.9478)
Institutions	0.0000 (0.0014)	0.0035 (0.4465)	0.0026 (0.3172)	0.0035 (0.4060)	0.0030 (0.3396)	0.0073 (1.1756)	0.0073 (1.1654)
Openness	0.0003 (0.0393)	−0.0044 (−0.6566)	−0.0030 (−0.4359)	−0.0060 (−0.8376)	−0.0048 (−0.6764)	0.0021 (0.3807)	0.0023 (0.4151)
Inflation	−0.0022 (−0.2752)	−0.0023 (−0.2709)	−0.0026 (−0.3078)	−0.0027 (−0.2562)	−0.0028 (−0.2811)	−0.0045 (−3.2430)	−0.0045 (−3.2569)
Population	−0.8022 (−1.5938)	−1.0487 (−1.4287)	−0.9330 (−0.9484)	−1.0934 (−1.6077)	−0.9765 (−1.4023)	−0.4786 (−0.4481)	−0.4951 (−0.4590)
Lag of growth rate	–	–	–	–	–	0.0155 (0.2411)	0.0193 (0.3000)
IPR	–	–	–	–	–	−0.0250 (−1.9342)	−0.0402 (−0.9043)
IPR^2	–	–	–	–	–	–	0.0029 (0.3793)

Note. The numbers in parentheses are the t-values based on bootstrap standard errors.

Table 6. Bootstrap p-values based on 2000 bootstrap resamples

Test statistics	$J_{s,iter}$	$J_{s,noniter}$	J_{iter}	$J_{noniter}$
Linear function (H_0 vs H_1)	0.0055	0.0075	0.0575	0.0110
Quadratic function (H_0' vs H_1')	0.0075	0.0070	0.0560	0.0105

the 10% significance level.

7 Conclusion

This paper provides two types of estimates of partially linear dynamic panel models where the lagged dependent variable enters the nonparametric component. One is based on the solutions to a class of Fredholm integral equations of the second kind and the other is based on the sieve approximation. We prove the asymptotic normality for the estimates of the parametric component and uniform consistency and asymptotic normality for the estimates of the nonparametric component. We also consider a specification test for the linearity of the nonparametric component based on the weighted squared distance between the semiparametric and parametric estimates. Monte Carlo simulations show that our estimators and tests perform reasonably well in finite samples. We illustrate our methods with an empirical application

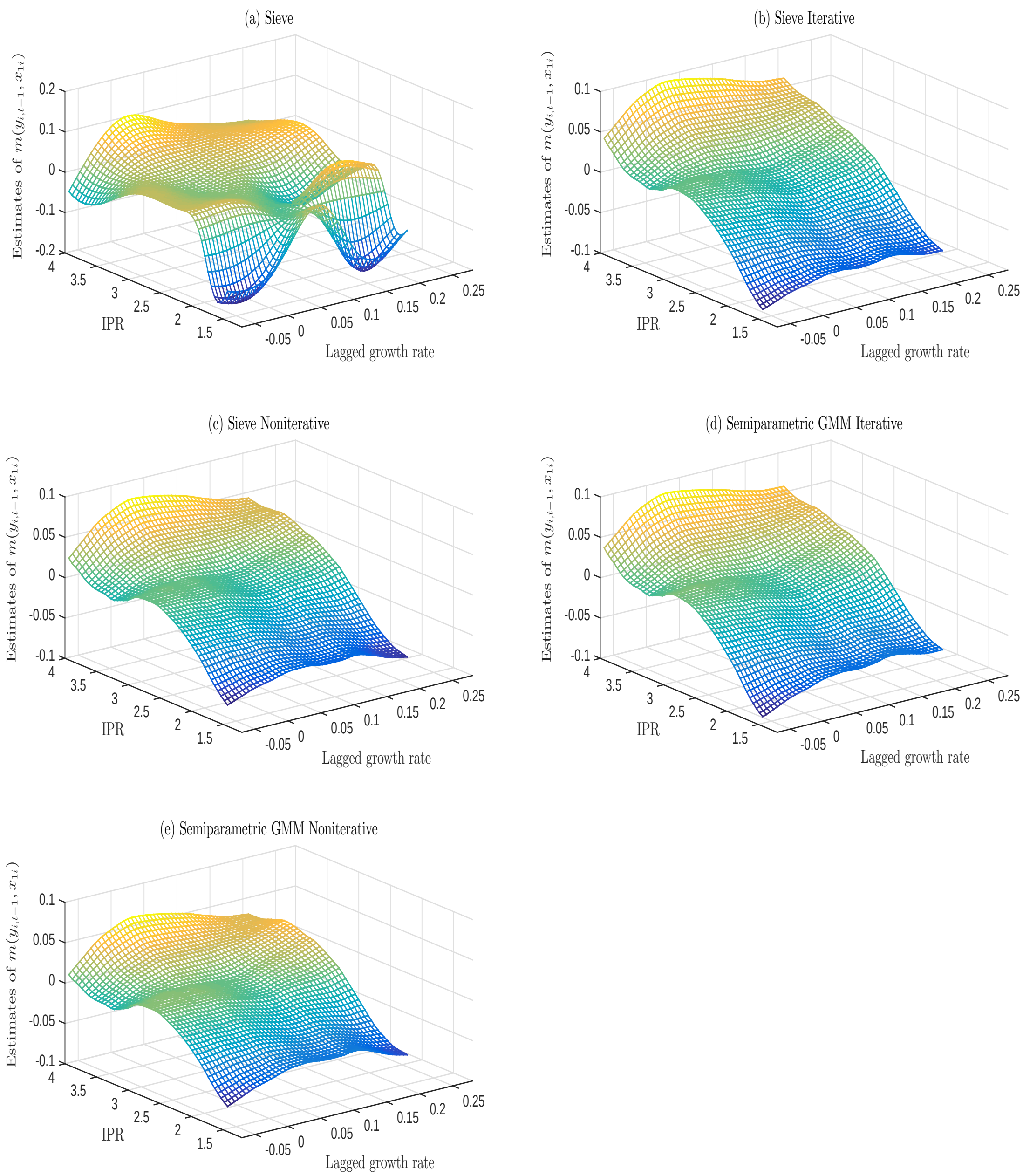

Figure 1: Relationship between the economic growth rate and the lagged growth rate ($y_{i,t-1}$) and IPR (x_{it})

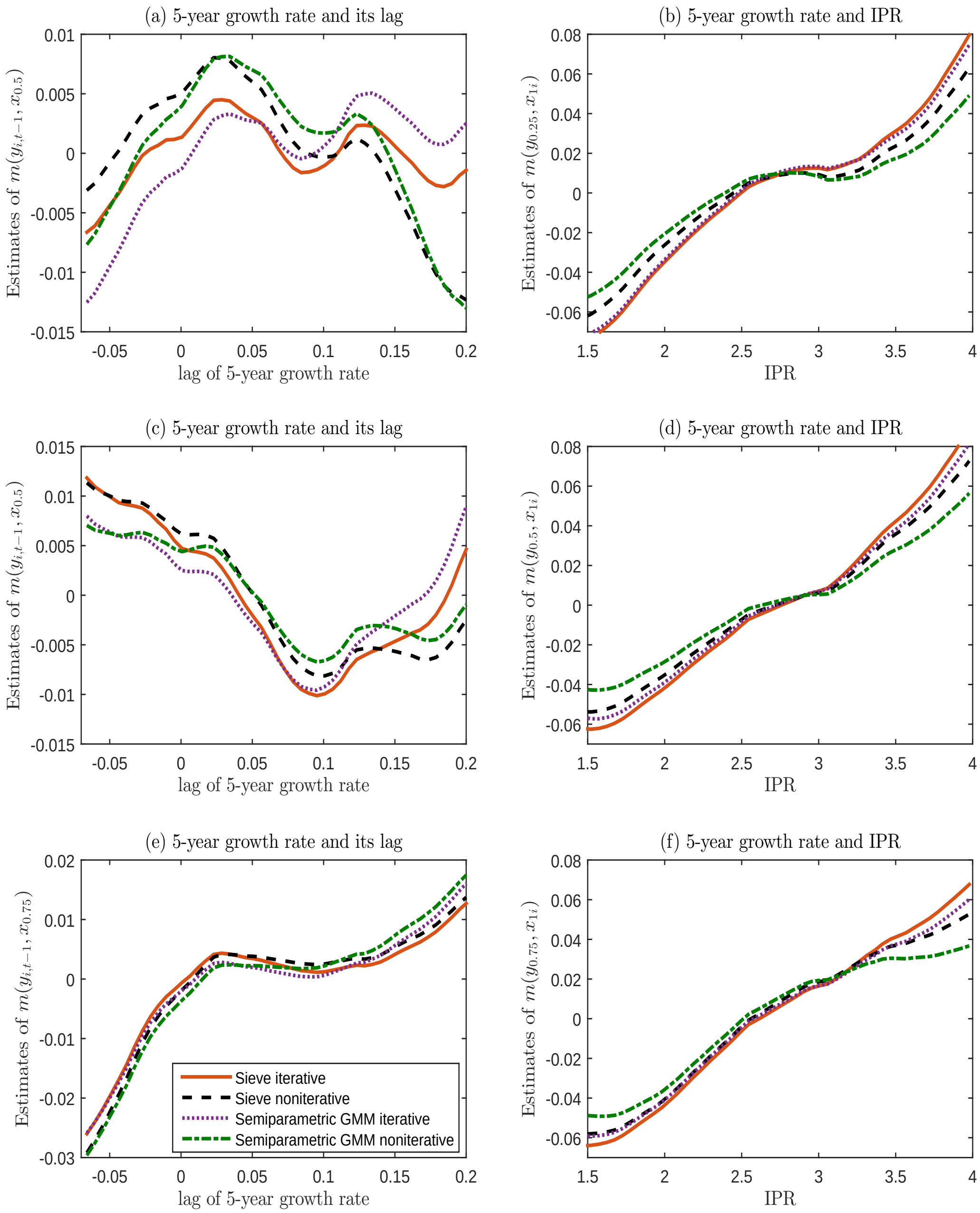

Figure 2: Relationship between the economic growth rate and the lagged growth rate ($y_{i,t-1}$) and IPR (x_{it})

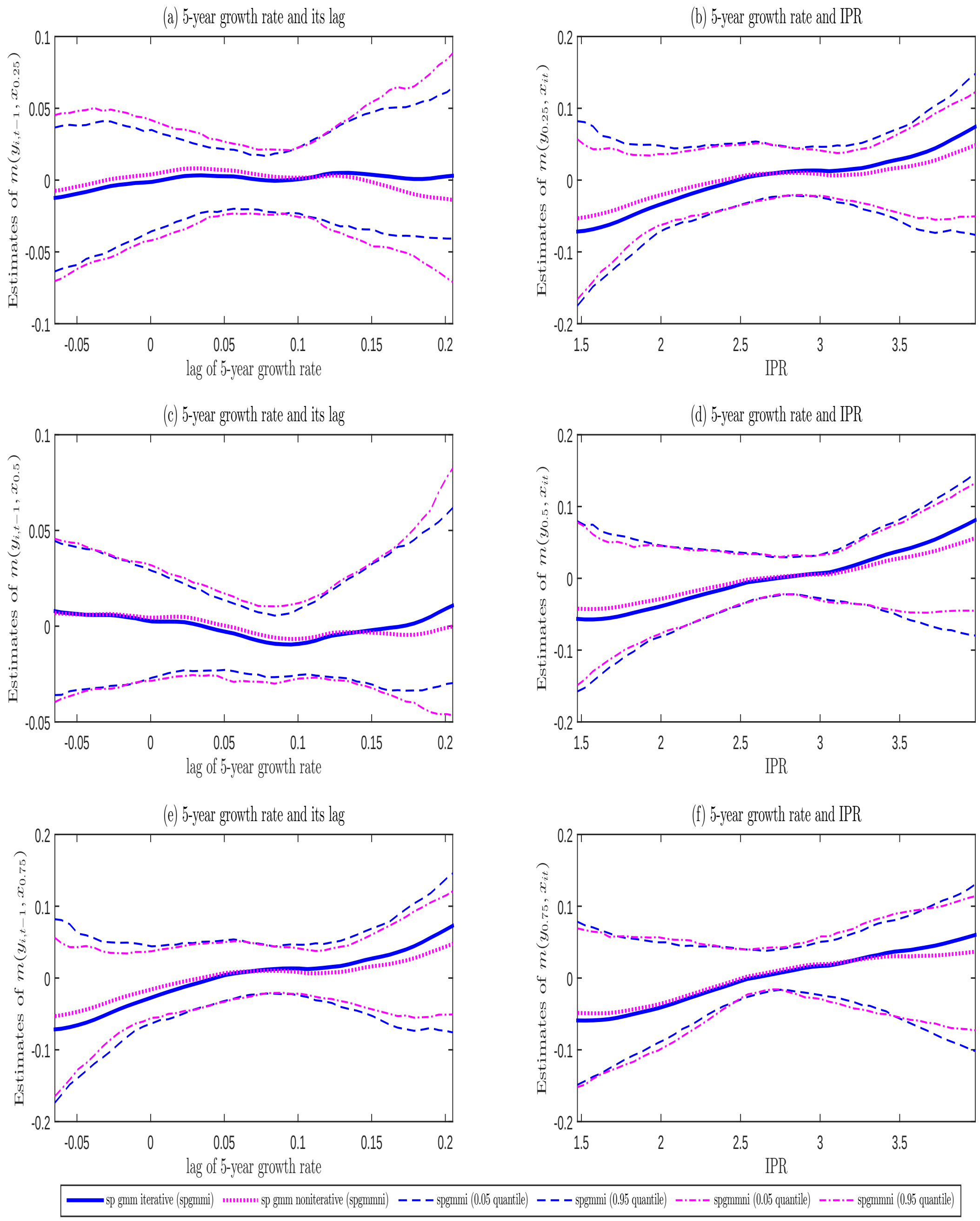

Figure 3: 90% confidence bands for the semiparametric GMM estimates of the nonparametric components at variance quantiles

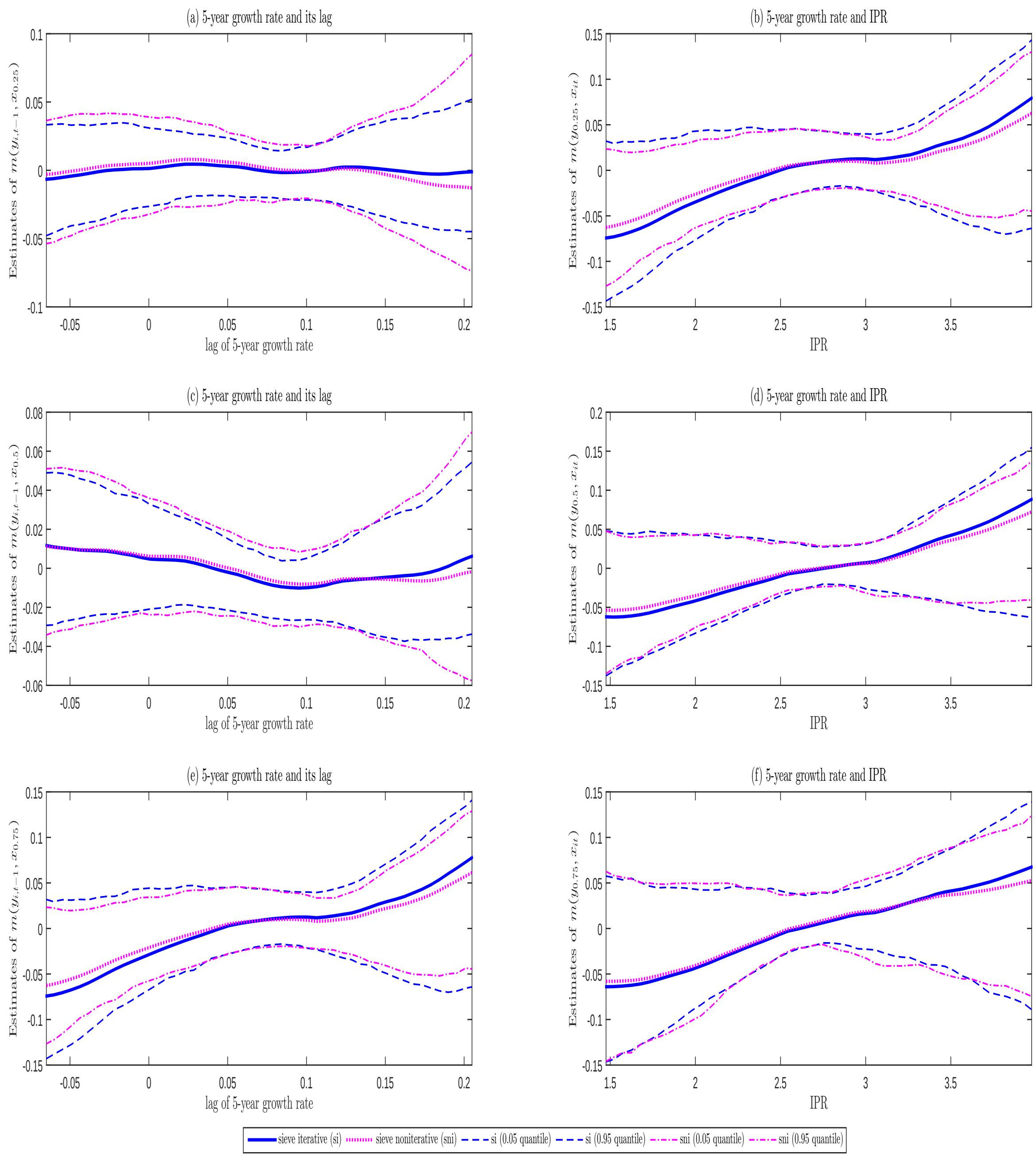

Figure 4: 90% confidence bands for the sieve estimates of the nonparametric components at variance quantiles

on economic growth and find a positive nonlinear relationship between the IPR protection and economic growth rate.

There are several interesting topics for further research. First, we do not address the choice of optimal IVs in this paper. Since we need IVs for the estimation of both the parametric and nonparametric components, we need a separate consideration of the choice of IVs for the estimation of the nonparametric component and the parametric component. Due to the nature of our first approach, we choose the IV for the nonparametric component as $U_{i,t-2}$ in order to derive a Fredholm integral equation of the second kind. More researches are needed to consider the optimal choice of IV for nonparametric estimation despite the fact that it is only relatively well studied how to choose optimal IVs for the parametric estimation. Second, we do not allow the endogenous regressors to enter the nonparametric component to avoid the ill-posed inverse problem. It is of great interest to allow the endogenous regressors to enter the nonparametric component in the panel data framework. We leave these for future research.

ACKNOWLEDGMENT

The authors thank the editors, three referees, Aman Ullah, and the participants at the AIE Conference in Honor of Aman Ullah for valuable comments. Su gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. Zhang gratefully acknowledges the Renmin University of China for Academic Research Fund under grant number 14XNF045.

Appendix

A Proof of the results in Sections 2 and 3

Proof of Theorem 2.1. The proof follows from that of Theorem 2.1 in Su and Lu (2013, **SL** hereafter) closely. The major difference is that we also maintain the second order variance term that plays an important role in the determination of the asymptotic variance of our parametric estimator but not in that of our nonparametric estimator. Let $|m|_\infty \equiv \sup_{u\in\mathcal{U}} |m(u)|$. Let $\Delta_n \equiv (nh!)^{-1/2}$ and $\nu_n \equiv \Delta_n(\log n)^{1/2} + \|h\|^{\bar{q}+1}$. We prove the first part of the theorem by verifying the conditions in Mammen and Yu (2009):

(B1) $\sup_{\|m\|_2\le 1} |\mathcal{A}m|_\infty < \infty$;

(B2) $\sup_{\|m\|_2\le 1} \left\|(\mathcal{I}-\mathcal{A})^{-1} m\right\|_2 < \infty$;

(B3) $\sup_{\|m\|_2\le 1} \left|(\widehat{\mathcal{A}}-\mathcal{A})m\right|_\infty = O_P(\nu_n)$; and

(B4) For each $a \in \{y, z_1, ..., z_{d_z}\}$, there exists a decomposition $\hat{r}_a - r_a + (\widehat{\mathcal{A}}-\mathcal{A})m_a = V_{a,NT} + B_{a,NT} + R_{a,NT}$ with random functions $V_{a,NT}$, $B_{a,NT}$ and $R_{a,NT}$ such that: (B4a) $||V_{a,NT}||_2 = O_P(\Delta_n)$, (B4b) $|\mathcal{A}(\mathcal{I}-\mathcal{A})^{-1} V_{a,NT}|_\infty = O_P(\sqrt{\log n/n})$, (B4c) $||B_{a,NT}||_2 = O_P(\|h\|^{\bar{q}+1})$, and (B4d) $|R_{a,NT}|_\infty = O_P[\Delta_n(\log n)^{1/2}\nu_n]$.

To see this, we focus on the case $a = y$ in the theorem and in (B4) since all the other cases can be proved similarly. Using $\mathcal{A}^{-1} - \mathcal{C}^{-1} = \mathcal{A}^{-1}(\mathcal{C}-\mathcal{A})\mathcal{C}^{-1}$ and $(\mathcal{I}-\mathcal{A})^{-1} = \mathcal{I} + \mathcal{A}(\mathcal{I}-\mathcal{A})^{-1}$, we have

$$
\begin{aligned}
\hat{m}_y - m_y &= (\mathcal{I}-\widehat{\mathcal{A}})^{-1}\hat{r}_y - (\mathcal{I}-\mathcal{A})^{-1}r_y \\
&= (\mathcal{I}-\widehat{\mathcal{A}})^{-1}(\hat{r}_y - r_y) + [(\mathcal{I}-\widehat{\mathcal{A}})^{-1} - (\mathcal{I}-\mathcal{A})^{-1}]r_y \\
&= (\mathcal{I}-\widehat{\mathcal{A}})^{-1}\left[(\hat{r}_y - r_y) + (\widehat{\mathcal{A}}-\mathcal{A})(\mathcal{I}-\mathcal{A})^{-1}r_y\right] \\
&= (\mathcal{I}-\widehat{\mathcal{A}})^{-1}\left[(\hat{r}_y - r_y) + (\widehat{\mathcal{A}}-\mathcal{A})m_y\right] \\
&= (\mathcal{I}-\widehat{\mathcal{A}})^{-1}\left[V_{y,NT} + B_{y,NT} + R_{y,NT}\right] \\
&= V_{y,NT} + \mathcal{A}(\mathcal{I}-\mathcal{A})^{-1}V_{y,NT} + (\mathcal{I}-\mathcal{A})^{-1}B_{y,NT} \\
&\quad + \mathcal{D}_1 V_{y,NT} + \mathcal{D}_2 B_{y,NT} + (\mathcal{I}-\widehat{\mathcal{A}})^{-1}R_{y,NT},
\end{aligned} \tag{A.1}
$$

where $\mathcal{D}_1 \equiv \widehat{\mathcal{A}}(\mathcal{I}-\widehat{\mathcal{A}})^{-1} - \mathcal{A}(\mathcal{I}-\mathcal{A})^{-1}$, $\mathcal{D}_2 \equiv (\mathcal{I}-\widehat{\mathcal{A}})^{-1} - (\mathcal{I}-\mathcal{A})^{-1}$, and the fifth equality follows from (A.2) below. Following the proof of Theorem 5 in Mammen et al. (2009), we can show that

$$
\begin{aligned}
|\mathcal{D}_1 V_{y,NT}|_\infty &= O_P(\nu_n\Delta_n) \text{ by (B2), (B3), (B4a), and (B4b)}, \\
|\mathcal{D}_2 B_{y,NT}|_\infty &= O_P(\nu_n \|h\|^{\bar{q}+1}) \text{ by (B2), (B3), and (B4c)}, \\
|(\mathcal{I}-\widehat{\mathcal{A}})^{-1}R_{y,NT}|_\infty &= O_P[\Delta_n(\log n)^{1/2}\nu_n] \text{ by (B2), (B3) and (B4d)}.
\end{aligned}
$$

With these results, by the fact that $(\mathcal{I}-\mathcal{A})^{-1} = \mathcal{I} + \mathcal{A}(\mathcal{I}-\mathcal{A})^{-1}$ and Minkowski inequality, we have

$$
\begin{aligned}
&\left|\hat{m}_y - m_y - (\mathcal{I}-\mathcal{A})^{-1}V_{y,NT} - (\mathcal{I}-\mathcal{A})^{-1}B_{y,NT}\right|_\infty \\
\le\ & |\mathcal{D}_1 V_{y,NT}|_\infty + |\mathcal{D}_2 B_{y,NT}|_\infty + \left|(\mathcal{I}-\widehat{\mathcal{A}})^{-1}R_{y,NT}\right|_\infty \\
=\ & O_P\left(\nu_n\Delta_n + \nu_n\|h\|^{\bar{q}+1} + \Delta_n(\log n)^{1/2}\nu_n\right) = O_P\left(\nu_n^2\right).
\end{aligned}
$$

That is, the first part of Theorem 2.1 holds for $a = y$. Analogously, it holds for $a = z_1, ..., z_{d_z}$. Noting that $\hat{m}_\theta(u) - m_\theta(u) = \hat{m}_y(u) - m_y(u) - \theta' [\hat{m}_z(u) - m_z(u)]$, $V_{\theta,NT}(u) = V_{y,NT}(u) - \theta' V_{z,NT}(u)$, and $B_{\theta,NT}(u) = B_{y,NT}(u) - \theta' B_{z,NT}(u)$, we have by Minkowski and Cauchy-Schwarz inequalities and the fact that Θ is compact

$$\begin{aligned}
& \sup_{\theta\in\Theta}\sup_{u\in\mathcal{U}} \left| \hat{m}_\theta(u) - m_\theta(u) - (\mathcal{I}-\mathcal{A})^{-1} V_{\theta,NT}(u) - (\mathcal{I}-\mathcal{A})^{-1} B_{\theta,NT}(u) \right| \\
\leq & \sup_{u\in\mathcal{U}} \left| \hat{m}_y(u) - m_y(u) - (\mathcal{I}-\mathcal{A})^{-1} V_{y,NT}(u) - (\mathcal{I}-\mathcal{A})^{-1} B_{y,NT}(u) \right| \\
& + \sup_{\theta\in\Theta} \|\theta\| \sup_{u\in\mathcal{U}} \left\| \hat{m}_z(u) - m_z(u) - (\mathcal{I}-\mathcal{A})^{-1} V_{z,NT}(u) - (\mathcal{I}-\mathcal{A})^{-1} B_{z,NT}(u) \right\| \\
= & O_P\left(\nu_n^2\right) + O_P\left(\nu_n^2\right) = O_P\left(\nu_n^2\right).
\end{aligned}$$

Thus, the second part of Theorem 2.1 also follows.

Now, we verify (B1)-(B4). Apparently, Assumption A.1(vii) ensures (B1). By the discussion of Assumption A.1(v) in **SL**, Assumption A.1(v) implies (B2). To verify (B3), letting $\bar{m}_a(u) \equiv \mathcal{A} m_a(u)$, we have the following bias-variance decomposition

$$\begin{aligned}
& (\widehat{\mathcal{A}} - \mathcal{A}) m_a(u) \\
= & \frac{1}{n} \sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) m_a(U_{i,t-1}) - \bar{m}_a(u) \\
= & \frac{1}{n} \sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) [m_a(U_{i,t-1}) - \bar{m}_a(U_{i,t-2})] + \frac{1}{n} \sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) [\bar{m}_a(U_{i,t-2}) - \bar{m}_a(u)] \\
\equiv & A_{a,1NT}(u) + A_{a,2NT}(u), \text{ say,}
\end{aligned}$$

where in the second equation we use the fact that $\frac{1}{n}\sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) = 1$ because $S_{NT}(u) S_{NT}(u)^{-1} = \mathbf{I}_G$. By the same fact, $A_{a,2NT}(u) = \frac{1}{n}\sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) D\bar{m}_{a,it}(u)$, where $D\bar{m}_{a,it}(u) \equiv \bar{m}_a(U_{i,t-2}) - \bar{m}_a(u) - \sum_{1\leq|\mathbf{j}|\leq q} \frac{1}{\mathbf{j}!} \bar{m}_a^{(\mathbf{j})}(u) (U_{i,t-2} - u)^{\mathbf{j}}$. Using the arguments as used in Masry (1996) and Hansen (2008), we can readily show that $\sup_{u\in\mathcal{U}} |A_{a,1NT}(u)| = O_P[\Delta_n (\log n)^{1/2}]$ and $\sup_{u\in\mathcal{U}} |A_{a,2NT}(u)| = O_P(\|h\|^{\bar{q}+1})$. Consequently (B3) follows. Noting that $\Delta Y_{it} = m_y(U_{i,t-1}) - m_y(U_{i,t-2}) + \eta_{y,it}$ and $r_y(u) + \bar{m}_y(u) = m_y(u)$ (see eqns. (2.10) and (2.11)), we have

$$\begin{aligned}
& \hat{r}_y(u) - r_y(u) + \widehat{\mathcal{A}} m_y(u) - \mathcal{A} m_y(u) \\
= & \frac{1}{n} \sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) [-\Delta Y_{it} + m_y(U_{i,t-1})] - [r_y(u) + \bar{m}_y(u)] \\
= & \frac{1}{n} \sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) \left\{ \left[m_y(U_{i,t-2}) - \eta_{y,it} \right] - m_y(u) \right\} \\
= & \frac{-1}{n} \sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) \eta_{y,it} + \frac{1}{n} \sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) [m_y(U_{i,t-2}) - m_y(u)] \\
= & \frac{-1}{n} \sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) \eta_{y,it} + \frac{1}{n} \sum_{i=1}^N \sum_{t=3}^T \mathcal{K}_{it}(u) \mathcal{D} m_{y,it}(u) \\
= & V_{y,NT}(u) + B_{y,NT}(u) + R_{y,NT}(u), \qquad \text{(A.2)}
\end{aligned}$$

where

$$
\begin{aligned}
B_{y,NT}(u) &\equiv \frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\bar{\mathcal{K}}_{it}(u)\,Dm_{y,it}(u),\\
V_{y,NT}(u) &\equiv \frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\bar{\mathcal{K}}_{it}(u)\,\eta_{y,it},\\
R_{y,NT}(u) &= e_1'\{[S_{NT}(u)]^{-1}-\left[\bar{S}_{NT}(u)\right]^{-1}\}\left[R_{y,1NT}(u)+R_{y,2NT}(u)\right], \qquad (A.3)
\end{aligned}
$$

$Dm_{y,it}(u)\equiv m_y(U_{i,t-2})-m_y(u)-\sum_{1\le|\mathbf{j}|\le q}\frac{1}{\mathbf{j}!}m_y^{(\mathbf{j})}(u)(U_{i,t-2}-u)^{\mathbf{j}}$, $R_{y,1NT}(u)=\frac{-1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathbf{1}_{it}K_h(U_{i,t-2}-u)\,\mu_h(U_{i,t-2}-u)\,\eta_{y,it}$, and $R_{y,2NT}(u)=\frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathbf{1}_{it}K_h(U_{i,t-2}-u)\,\mu_h(U_{i,t-2}-u)\,Dm_{y,it}(u)$. Using the above decomposition and following **SL**, one can readily verify conditions (B4a)-(B4d). ■

Proof of Theorem 2.2. We make the following decomposition:

$$
\begin{aligned}
\hat{\theta}_{gmm}-\theta_0 &= (\tilde{\theta}_{gmm}-\theta_0)\\
&\quad+\left(\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_z\right)^{-}\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_y-\left(\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_z\right)^{-}\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_y\\
&= (\tilde{\theta}_{gmm}-\theta_0)+\left(\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_z\right)^{-}\left(\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_y-\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_y\right)\\
&\quad+\left[\left(\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_z\right)^{-}-\left(\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_z\right)^{-}\right]\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_y\\
&\quad+\left[\left(\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_z\right)^{-}-\left(\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_z\right)^{-}\right]\left(\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_y-\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_y\right)\\
&\equiv (\tilde{\theta}_{gmm}-\theta_0)+B_1+B_2+B_3,\ \text{say.}
\end{aligned}
$$

Let $\Delta\boldsymbol{\varepsilon}\equiv(\Delta\varepsilon_1,...,\Delta\varepsilon_N)'$ where $\Delta\varepsilon_i\equiv(\Delta\varepsilon_{i3},...,\Delta\varepsilon_{iT})'$. It suffices to prove the theorem by showing that

(i) $\frac{1}{\tilde{n}^2}\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_z=Q_{wz}'AQ_{wz}+o_P(1)$,

(ii) $\sqrt{\tilde{n}}(\tilde{\theta}_{gmm}-\theta_0)=(Q_{wz}'AQ_{wz})^{-1}Q_{wz}'A\frac{1}{\sqrt{\tilde{n}}}\tilde{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}+o_P(1)$,

(iii) $\sqrt{\tilde{n}}B_1=(Q_{wz}'AQ_{wz})^{-1}\frac{\sqrt{\tilde{n}}}{N}\sum_{i=1}^{N}[Q_{wz}'A\varphi_1(\chi_{y,i})+\varphi_{1,z}(\chi_{z,i})'AQ_{wz}\theta_0]+o_P(1)$,

(iv) $\sqrt{\tilde{n}}B_2=-(Q_{wz}'AQ_{wz})^{-1}\frac{\sqrt{\tilde{n}}}{N}\sum_{i=1}^{N}[Q_{wz}'A\varphi_{1,z}(\chi_{z,i})+\varphi_{1,z}(\chi_{z,i})'AQ_{wz}]\theta_0+o_P(1)$, and

(v) $\sqrt{\tilde{n}}B_3=o_P(1)$,

because these results, in conjunction with Slutsky lemma, the continuous mapping theorem, and Assumption A.5(iii), imply that

$$
\begin{aligned}
\sqrt{\tilde{n}}\left(\hat{\theta}_{gmm}-\theta_0\right) &= (Q_{wz}'AQ_{wz})^{-1}Q_{wz}'A\left\{\frac{1}{\sqrt{\tilde{n}}}\tilde{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}+\frac{\sqrt{\tilde{n}}}{N}\sum_{i=1}^{N}[\varphi_1(\chi_{y,i})-\varphi_{1,z}(\chi_{z,i})\theta_0]\right\}+o_P(1)\\
&\xrightarrow{D} N\left(0,(Q_{wz}'AQ_{wz})^{-1}Q_{wz}'A\Omega_0AQ_{wz}(Q_{wz}'AQ_{wz})^{-1}\right).
\end{aligned}
$$

We first show (i). By Theorem 2.1, Remark 1, and Assumption A.6, we have that uniformly in $u\in\mathcal{U}$,

$$
\begin{aligned}
\hat{m}_a(u)-m_a(u) &= (\mathcal{I}-\mathcal{A})^{-1}V_{a,NT}(u)+o_P(n^{-1/2})\ \text{for}\ a=y,\ z,\ \text{and} \qquad (A.4)\\
\hat{m}_a(u)-m_a(u) &= O_P(\nu_n)\ \text{for}\ a=y,\ z. \qquad (A.5)
\end{aligned}
$$

It follows that $\max_{i,t}\tilde{\mathbf{1}}_{it}\left\|\Delta m_{a,it}-\Delta\hat{m}_{a,it}\right\| = O_P(\nu_n)$ for $a=y$ and z. Noting that $\hat{\eta}_{z,it}-\eta_{z,it}=\Delta m_{z,it}-\Delta\hat{m}_{z,it}$, we have by Assumption A.5(ii) and Markov inequality

$$\begin{aligned}\left\|\frac{1}{\tilde{n}}\left(\hat{\boldsymbol{\eta}}_z-\boldsymbol{\eta}_z\right)'\tilde{\mathbf{W}}\right\| &\leq \frac{1}{\tilde{n}}\sum_{i=1}^{N}\sum_{t=3}^{T}\tilde{\mathbf{1}}_{it}\left\|\Delta m_{z,it}-\Delta\hat{m}_{z,it}\right\|\left\|W_{it}\right\| \\ &\leq \max_{i,t}\tilde{\mathbf{1}}_{it}\left\|\Delta m_{a,it}-\Delta\hat{m}_{a,it}\right\|\frac{1}{\tilde{n}}\sum_{i=1}^{N}\sum_{t=3}^{T}\left\|W_{it}\right\|\tilde{\mathbf{1}}_{it}=O_P(\nu_n). \qquad (A.6)\end{aligned}$$

With this result, one can readily show that $\frac{1}{\tilde{n}^2}\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_z=\frac{1}{\tilde{n}^2}\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_z+o_P(1)$. Then (i) follows by noticing that

$$\frac{1}{\tilde{n}^2}\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_z \xrightarrow{P} Q_{wz}'AQ_{wz}>0 \qquad (A.7)$$

by Assumption A.5(i) and Slutsky lemma.

For (ii), by (2.23) and the fact that $\eta_{y,it}=\theta_0'\eta_{z,it}+\Delta\varepsilon_{it}$ we have

$$\sqrt{\tilde{n}}\left(\tilde{\theta}_{gmm}-\theta_0\right)=\left(\frac{1}{\tilde{n}^2}\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_z\right)^{-1}\frac{1}{\tilde{n}}\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\frac{1}{\sqrt{\tilde{n}}}\tilde{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}.$$

Then (ii) follows by (A.7), Assumption A.5(i) and the fact that $\frac{1}{\sqrt{\tilde{n}}}\tilde{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}=O_P(1)$.

To show (iii), let the operator $\mathcal{L}(\bar{u},u)$ be as defined in (2.28). As in Linton and Mammen (2005, p.821), we can show that $\int_{\mathcal{U}}\int_{\mathcal{U}}\mathcal{L}(u,\bar{u})^2 m(\bar{u})f(\bar{u})f(u)\,d\bar{u}du<\infty$ under Assumptions A.1(iv)-(vi) and

$$\begin{aligned}(\mathcal{I}-\mathcal{A})^{-1}V_{y,NT}(u) &= \int_{\mathcal{U}}\mathcal{L}(u,\bar{u})\frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathbf{1}_{it}e_1'\bar{S}_{NT}(\bar{u})^{-1}\mu_h(U_{i,t-2}-\bar{u})K_h(U_{i,t-2}-\bar{u})\,d\bar{u}\eta_{y,it} \\ &= \frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\bar{\mathcal{L}}(U_{i,t-2},u)\eta_{y,it}, \qquad (A.8)\end{aligned}$$

where $\bar{\mathcal{L}}(v,u)\equiv\int_{\mathcal{U}}\mathcal{L}(u,\bar{u})\mathbf{1}(v\in\mathcal{U})e_1'\bar{S}_{NT}(\bar{u})^{-1}\mu_h(v-\bar{u})K_h(v-\bar{u})\,d\bar{u}$.[6] Next, we make the following decomposition

$$\begin{aligned}&\frac{1}{\tilde{n}^{3/2}}\left(\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_y-\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_y\right) \\ =\;&\frac{1}{\tilde{n}^{3/2}}\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'(\hat{\boldsymbol{\eta}}_y-\boldsymbol{\eta}_y)+\frac{1}{\tilde{n}^{3/2}}(\hat{\boldsymbol{\eta}}_z-\boldsymbol{\eta}_z)'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_y+\frac{1}{\tilde{n}^{3/2}}(\hat{\boldsymbol{\eta}}_z-\boldsymbol{\eta}_z)'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'(\hat{\boldsymbol{\eta}}_y-\boldsymbol{\eta}_y) \\ \equiv\;& I_1+I_2+I_3,\text{ say.}\end{aligned}$$

Apparently, by (A.6) and Assumptions A.5(i) and A.6, $I_3=\sqrt{\tilde{n}}O_P(\nu_n^2)=o_P(1)$. To study I_1 and I_2, we resort to the results in (A.4) and (A.8) and the theory for second order U-statistics. For $a=y,z_1,...,z_{d_z}$, define

$$\begin{aligned}\varphi^0\left(\chi_{a,i},\chi_{a,j}\right) &= -\frac{N(N-1)}{n\tilde{n}}\sum_{s=3}^{T}\sum_{t=3}^{T}\left[\bar{\mathcal{L}}(U_{j,s-2},U_{i,t-1})-\bar{\mathcal{L}}(U_{j,s-2},U_{i,t-2})\right]\tilde{\mathbf{1}}_{it}W_{it}\eta_{a,js}, \\ \varphi\left(\chi_{a,i},\chi_{a,j}\right) &= [\varphi^0\left(\chi_{a,i},\chi_{a,j}\right)+\varphi^0\left(\chi_{a,j},\chi_{a,i}\right)]/2,\end{aligned}$$

[6] Define the operator $\mathcal{L}^{\dagger}$ by $\mathcal{A}(\mathcal{I}-\mathcal{A})^{-1}m(u)=\int_{\mathcal{U}}\mathcal{L}^{\dagger}(u,\bar{u})m(\bar{u})f(\bar{u})\,d\bar{u}$. Strictly speaking, Linton and Mammen (2005, p.281) show that $\mathcal{L}^{\dagger}(u,\bar{u})$ satisfies the square integrability condition. The result for $\mathcal{L}$ follows from the fact that $(\mathcal{I}-\mathcal{A})^{-1}=\mathcal{I}+\mathcal{A}(\mathcal{I}-\mathcal{A})^{-1}$ and that we restrict the integration to be done over the compact set $\mathcal{U}$.

where $\chi_{a,i} = (U_{i1}',...,U_{i,T-1}', W_{i3}',...,W_{iT}',\ \eta_{a,i3},...,\eta_{a,iT})'$. Let $\varphi_1(\chi_{a,i}) = E[\varphi(\chi_{a,i},\chi_{a,j})|\chi_{a,i}]$ and $\varphi_2(\chi_{a,i},\chi_{a,j}) = \varphi(\chi_{a,i},\chi_{a,j}) - \varphi_1(\chi_{a,i}) - \varphi_1(\chi_{a,j}) + E[\varphi_1(\chi_{a,i})]$ for any $i \neq j$. It is easy to verify that φ_1 is as given in (2.29) and $E[\varphi_1(\chi_{a,i})] = 0$. By (A.4) and (A.8),

$$\begin{aligned}
\frac{1}{\sqrt{\tilde{n}}}\tilde{\mathbf{W}}'(\hat{\boldsymbol{\eta}}_y - \boldsymbol{\eta}_y) &= -\frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{N}\sum_{t=3}^{T}(\Delta\hat{m}_{y,it} - \Delta m_{y,it})\tilde{\mathbf{1}}_{it}W_{it} \\
&= -\frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{N}\sum_{t=3}^{T}[(\hat{m}_{y,it} - m_{y,it}) - (\hat{m}_{y,i,t-1} - m_{y,i,t-1})]\tilde{\mathbf{1}}_{it}W_{it} \\
&= -\frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{N}\sum_{t=3}^{T}\left[\frac{1}{n}\sum_{j=1}^{N}\sum_{s=3}^{T}[\bar{\mathcal{L}}(U_{j,s-2},U_{i,t-1}) - \bar{\mathcal{L}}(U_{j,s-2},U_{i,t-2})]\eta_{y,js}\right]\tilde{\mathbf{1}}_{it}W_{it} + o_P(1) \\
&= \sqrt{\tilde{n}}\mathcal{V}_{y,NT} + o_P(1)
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{V}_{y,NT} &= -\frac{1}{n\tilde{n}}\sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{s=3}^{T}\sum_{t=3}^{T}[\bar{\mathcal{L}}(U_{j,s-2},U_{i,t-1}) - \bar{\mathcal{L}}(U_{j,s-2},U_{i,t-2})]\tilde{\mathbf{1}}_{it}W_{it}\eta_{y,js} \\
&= \frac{2}{N(N-1)}\sum_{1\leq i<j\leq N}\varphi(\chi_{y,i},\chi_{y,j}) + \frac{1}{n\tilde{n}}\sum_{i=1}^{N}\varphi(\chi_{y,i},\chi_{y,i}) \equiv \mathcal{V}_{y,NT1} + \mathcal{V}_{y,NT2},\ \text{say,}
\end{aligned}$$

By Markov inequality, we can readily show that $\mathcal{V}_{y,NT2} = O_P(N^{-1})$. Let $\mathcal{H}_{y,NT}^{(1)} = \frac{1}{N}\sum_{i=1}^{N}\varphi_1(\chi_{y,i})$ and $\mathcal{H}_{y,NT}^{(2)} = \frac{1}{N(N-1)}\sum_{1\leq i<j\leq N}\varphi_2(\chi_{y,i},\chi_{y,j})$. By Hoeffding decomposition (e.g., Lee (1990, p.26)), $\mathcal{V}_{y,NT1} = 2\mathcal{H}_{y,NT}^{(1)} + \mathcal{H}_{y,NT}^{(2)}$. By straightforward moment calculations, $E\left\|\mathcal{H}_{y,NT}^{(2)}\right\|^2 = O(N^{-2})$. So $\mathcal{H}_{y,NT}^{(2)} = O_P(N^{-1}) = o_P(\tilde{n}^{-1/2})$. Then we have $\frac{1}{\sqrt{\tilde{n}}}\tilde{\mathbf{W}}(\hat{\boldsymbol{\eta}}_y - \boldsymbol{\eta}_y) = \sqrt{\tilde{n}}\mathcal{V}_{y,NT1} + o_P(1) = \frac{\sqrt{\tilde{n}}}{N}\sum_{i=1}^{N}\varphi_1(\chi_{y,i})$ $+o_P(1)$ and

$$I_1 = \frac{1}{\tilde{n}}\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\frac{1}{\tilde{n}^{1/2}}\tilde{\mathbf{W}}'(\hat{\boldsymbol{\eta}}_y - \boldsymbol{\eta}_y) = Q_{wz}'A\frac{\sqrt{\tilde{n}}}{N}\sum_{i=1}^{N}\varphi_1(\chi_{y,i}) + o_P(1).$$

By the same token we can show that $\frac{1}{\tilde{n}^{1/2}}(\hat{\boldsymbol{\eta}}_z - \boldsymbol{\eta}_z)'\tilde{\mathbf{W}} = \frac{\sqrt{\tilde{n}}}{N}\sum_{i=1}^{N}\varphi_{1,z}(\chi_{z,i}) + o_P(1)$, where $\varphi_{1,z}(\chi_{z,i}) = (\varphi_1(\chi_{z_1,i}),...,\varphi_1(\chi_{z_{d_z},i}))$ is a $d_w \times d_z$ matrix. Using $\eta_{y,it} = \theta_0'\eta_{z,it} + \Delta\varepsilon_{it}$ again, we can readily show that $\frac{1}{\tilde{n}}\tilde{\mathbf{W}}'\boldsymbol{\eta}_y = \frac{1}{\tilde{n}}\tilde{\mathbf{W}}'\boldsymbol{\eta}_z\theta_0 + \frac{1}{\tilde{n}}\tilde{\mathbf{W}}'\Delta\boldsymbol{\varepsilon} = Q_{wz}\theta_0 + o_P(1)$. It follows that $I_2 = \frac{1}{\tilde{n}^{1/2}}(\hat{\boldsymbol{\eta}}_z - \boldsymbol{\eta}_z)'\tilde{\mathbf{W}}A_{NT}\frac{1}{\tilde{n}}\tilde{\mathbf{W}}'\boldsymbol{\eta}_y = \frac{\sqrt{\tilde{n}}}{N}\sum_{i=1}^{N}\varphi_{1,z}(\chi_{z,i})'AQ_{wz}\theta_0 + o_P(1)$. Consequently,

$$\frac{1}{\tilde{n}^{3/2}}\left(\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_y - \boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_y\right) = \frac{\sqrt{\tilde{n}}}{N}\sum_{i=1}^{N}\left[Q_{wz}'A\varphi_1(\chi_{y,i}) + \varphi_{1,z}(\chi_{z,i})'AQ_{wz}\theta_0\right] + o_P(1),$$

and (iii) follows.

Now, we show (iv). By the fact that $A_1^{-1} - A_2^{-1} = A_1^{-1}(A_2 - A_1)A_2^{-1}$ for any two conformable nonsingular matrices A_1 and A_2 and that $\eta_{y,it} = \theta_0'\eta_{z,it} + \Delta\varepsilon_{it}$, we have

$$\sqrt{\tilde{n}}B_2 = \hat{\Phi}_{NT}^{-1}\sqrt{\tilde{n}}\left(\Phi_{NT} - \hat{\Phi}_{NT}\right)\theta_0 + \hat{\Phi}_{NT}^{-1}\sqrt{\tilde{n}}\left(\Phi_{NT} - \hat{\Phi}_{NT}\right)\Phi_{NT}^{-1}\frac{1}{\tilde{n}^2}\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\Delta\boldsymbol{\varepsilon} \equiv I_4 + I_5,\ \text{say,}$$

where $\hat{\Phi}_{NT}=\frac{1}{\tilde{n}^2}\hat{\boldsymbol{\eta}}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_z$ and $\Phi_{NT}=\frac{1}{\tilde{n}^2}\boldsymbol{\eta}_z'\tilde{\mathbf{W}}A_{NT}\tilde{\mathbf{W}}'\boldsymbol{\eta}_z$. By arguments as used in the derivation of (i) and (iii), we can readily show that

$$
\begin{aligned}
I_4 &= -\left(Q_{wz}'AQ_{wz}\right)^{-1}\frac{\sqrt{\tilde{n}}}{N}\sum_{i=1}^{N}\left[Q_{wz}'A\varphi_{1,z}\left(\chi_{z,i}\right)+\varphi_{1,z}\left(\chi_{z,i}\right)'AQ_{wz}\right]\theta_0+o_P(1) \text{ and} \\
I_5 &= O_P(1)\,O_P(1)\,O_P(1)\,O_P(\tilde{n}^{-1/2})=o_P(1).
\end{aligned}
$$

Then (iv) follows.

Lastly, noting that $\frac{1}{\tilde{n}}\tilde{\mathbf{W}}'\hat{\boldsymbol{\eta}}_a-\frac{1}{\tilde{n}}\tilde{\mathbf{W}}'\boldsymbol{\eta}_a=O_P(\nu_n)$ for $a=y$ and z, we can readily show that $\sqrt{\tilde{n}}B_3=\sqrt{\tilde{n}}O_P(\nu_n^2)=o_P(1)$ by Slutsky lemma and Assumptions A.5(i) and A.6. This completes the proof of the theorem. ■

Proof of Theorem 2.3. Let $\hat{m}_{\theta_0}^b(u)\equiv\hat{m}_y^b(u)-\theta_0'\hat{m}_z^b(u)$. Let $B_y(u)$, $B_z(u)$, and $B_\theta(u)$ be analogously defined as $B_0(u)$ in Theorem 2.3 with m being replaced by $m_y(u)$, $m_z(u)$, and $m_\theta(u)$, respectively. In view of the fact that $m_{\theta_0}(u)=m(u)$ and $B_{\theta_0}(u)=B_0(u)$, we have

$$
\begin{aligned}
\sqrt{nb!}[\hat{m}^b_{\hat{\theta}_{gmm}}(u)-m(u)-(\mathcal{I}-\mathcal{A})^{-1}B_0(u)] &= \sqrt{nb!}[\hat{m}^b_{\theta_0}(u)-m_{\theta_0}(u)-(\mathcal{I}-\mathcal{A})^{-1}B_{\theta_0}(u)] \\
&\quad+\sqrt{nb!}[\hat{m}^b_{\hat{\theta}_{gmm}}(u)-\hat{m}_{\theta_0}(u)].
\end{aligned}
$$

It suffices to prove the theorem by showing that (i) $\sqrt{nb!}[\hat{m}^b_{\theta_0}(u)-m_{\theta_0}(u)-(\mathcal{I}-\mathcal{A})^{-1}B_{\theta_0}(u)]\xrightarrow{d} N\left(0,\frac{\sigma_0^2(u)}{f(u)}e_1'\mathbb{S}^{-1}\mathbb{K}\mathbb{S}^{-1}e_1\right)$, (ii) $\sup_{u\in\mathcal{U}}|\hat{m}^b_{\theta_0}(u)-m_{\theta_0}(u)|=O_P\left((nb!)^{-1/2}(\log n)^{1/2}+\|b\|^{\bar{q}+1}\right)$, (iii) $\sup_{u\in\mathcal{U}}\left\|(\mathcal{I}-\mathcal{A})^{-1}B_0(u)\right\|=O_P\left(\|b\|^{\bar{q}+1}\right)$, and (iv) $\sqrt{nb!}\sup_{u\in\mathcal{U}}|\hat{m}^b_{\hat{\theta}_{gmm}}(u)-\hat{m}^b_{\theta_0}(u)|=o_P(1)$.

Noting that $\hat{m}^b_{\theta_0}(u)-m_{\theta_0}(u)=[\hat{m}^b_y(u)-m_y(u)]-\theta_0'\left[\hat{m}^b_z(u)-m_z(u)\right]$, $B_{\theta_0}(u)=B_y(u)-\theta_0'B_z(u)$, and $\Delta\varepsilon_{it}(\theta_0)=\eta_{y,it}-\theta_0'\eta_{z,it}=\Delta\varepsilon_{it}$, the proof of (i)-(iii) follows straightforward from that of Theorems 2.1 and 2.2 in **SL**. Noting that $\hat{m}^b_\theta(u)=\hat{m}^b_y(u)-\theta'\hat{m}^b_z(u)$ for $\theta=\hat{\theta}_{gmm}$ and θ_0, we have by Theorems 2.1 and 2.2

$$
\begin{aligned}
\sqrt{nb!}\sup_{u\in\mathcal{U}}|\hat{m}^b_{\hat{\theta}_{gmm}}(u)-\hat{m}^b_{\theta_0}(u)| &\leq \sqrt{nb!}\left\|\hat{\theta}_{gmm}-\theta_0\right\|\sup_{u\in\mathcal{U}}\left|\hat{m}^b_z(u)\right| \\
&= -\sqrt{nb!}\left\|\hat{\theta}_{gmm}-\theta_0\right\|\left(\sup_{u\in\mathcal{U}}|m_z(u)|+o_P(1)\right) \\
&= \sqrt{nb!}O_P(\tilde{n}^{-1/2})O_P(1)=o_P(1).
\end{aligned}
$$

Thus (iv) follows. ■

Proof of Theorem 3.1. Noting that $\mathbf{Y}_w=\mathbf{Z}_w\theta_0+\mathbf{p}_w\beta+\mathbf{P}_{\bar{\mathbf{W}}}\Delta\boldsymbol{\varepsilon}+\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R}$ and $\mathbf{M}_{\mathbf{p}_w}\mathbf{p}_w=0$, we have

$$
\begin{aligned}
\hat{\theta}_{sieve}-\theta_0 &= \left(\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{Z}_w\right)^{-}\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{Y}_w-\theta_0 \\
&= \left(\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{Z}_w\right)^{-}\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{P}_{\bar{\mathbf{W}}}\Delta\boldsymbol{\varepsilon}+\left(\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{Z}_w\right)^{-}\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R}.
\end{aligned}
$$

It suffices to prove the theorem by showing that (i) $Q_{1,NT}\equiv\frac{1}{NT_2}\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{Z}_w\xrightarrow{P}Q_1>0$, (ii) $R_{NT}\equiv\frac{1}{NT_2}\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R}=o_P(N^{-1/2})$, and (iii) $\frac{1}{\sqrt{NT_2}}\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{P}_{\bar{\mathbf{W}}}\Delta\boldsymbol{\varepsilon}\xrightarrow{D}N\left(0,Q_2\Omega_1Q_2'\right)$.

To show (i), note that $Q_{1,NT} = \frac{1}{NT_2}\mathbf{Z}_w'\mathbf{Z}_w - \frac{1}{NT_2}\mathbf{Z}_w'\mathbf{P}_{\mathbf{p}_w}\mathbf{Z}_w = \frac{1}{NT_2}\Delta\mathbf{Z}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{Z} - \frac{1}{NT_2}\Delta\mathbf{Z}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p}$ $(\Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p})^{-}\Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{Z}$. By Assumptions A.1(i) and A.8(i) and Chebyshev inequality, one can readily show that

$$\begin{aligned}\left\|\frac{1}{NT_2}\bar{\mathbf{W}}'\Delta\mathbf{Z} - Q_{\bar{w}z}\right\| &= O_P\left(\sqrt{L/N}\right), \ \left\|\frac{1}{NT_2}\bar{\mathbf{W}}'\Delta\mathbf{p} - Q_{\bar{w}p}\right\| = O_P\left(L/\sqrt{N}\right), \\ \left\|\frac{1}{NT_2}\bar{\mathbf{W}}'\bar{\mathbf{W}} - Q_{\bar{w}}\right\| &= O_P\left(L/\sqrt{N}\right). \end{aligned} \tag{A.9}$$

By (A.9), the triangle inequality, the submultiplicative property of Frobenius norm and Assumption A.8, we can readily show that

$$\frac{1}{NT_2}\Delta\mathbf{Z}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{Z} = \frac{\Delta\mathbf{Z}'\bar{\mathbf{W}}}{NT_2}\left(\frac{\bar{\mathbf{W}}'\bar{\mathbf{W}}}{NT_2}\right)^{-}\frac{\bar{\mathbf{W}}'\Delta\mathbf{Z}}{NT_2} = Q_{\bar{w}z}'Q_{\bar{w}}^{-1}Q_{\bar{w}z} + O_P\left(\frac{L}{\sqrt{N}}\right).$$

Similarly, $\frac{1}{NT_2}\Delta\mathbf{Z}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p} = Q_{\bar{w}z}'Q_{\bar{w}}^{-1}Q_{\bar{w}p} + O_P(L/\sqrt{N})$ and $\frac{1}{NT_2}\Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p} = Q_{\bar{w}p}'Q_{\bar{w}}^{-1}Q_{\bar{w}p} + O_P(L/\sqrt{N})$. It follows that $Q_{1,NT} \xrightarrow{P} Q_{\bar{w}z}'Q_{\bar{w}}^{-1}Q_{\bar{w}z} - Q_{\bar{w}z}'Q_{\bar{w}}^{-1}Q_{\bar{w}p}\left(Q_{\bar{w}p}'Q_{\bar{w}}^{-1}Q_{\bar{w}p}\right)^{-1}Q_{\bar{w}p}'Q_{\bar{w}}^{-1}Q_{\bar{w}z} = Q_1 > 0$.

Next, we show (ii). By the submultiplicative property of the spectral norm, the fact that $\|\mathbf{P}_{\bar{\mathbf{W}}}\|_{\text{sp}} = 1$ and that $\|\mathbf{M}_{\mathbf{p}_w}\|_{\text{sp}} = 1$, we have

$$\begin{aligned}\|R_{NT}\|_{\text{sp}}^2 &= \frac{1}{(NT_2)^2}\|\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R}\|_{\text{sp}}^2 \le \frac{1}{(NT_2)^2}\|\mathbf{Z}_w\|_{\text{sp}}^2\|\mathbf{R}\|^2 \\ &= \lambda_{\max}\left(\frac{1}{NT_2}\mathbf{Z}_w'\mathbf{Z}_w\right)\frac{1}{NT_2}\|\mathbf{R}\|^2 \le \lambda_{\max}\left(\frac{1}{NT_2}\Delta\mathbf{Z}'\Delta\mathbf{Z}\right)\frac{1}{NT_2}\|\mathbf{R}\|^2 \\ &= O_P(1)\,O_P\left(L^{-2\gamma/(1+d_x)}\right)\end{aligned}$$

where we use the fact that $\frac{1}{NT_2}\|\mathbf{R}\|^2 = O_P\left(L^{-2\gamma/(1+d_x)}\right)$ under Assumptions A.1(i) and A.7 by arguments as used in the proof of Lemma A.2 in Su and Jin (2012). Then by Chebyshev inequality and Assumptions A.8 and A.9, $\sqrt{N}\|R_{NT}\| = O_P(\sqrt{N}L^{-\gamma/(1+d_x)}) = o_P(1)$.

Now, we show (iii). Let $Q_{2,NT} \equiv \mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\bar{\mathbf{W}}(\bar{\mathbf{W}}'\bar{\mathbf{W}})^{-}$. Using $\mathbf{M}_{\mathbf{p}_w} = \mathbf{I}_{NT_2} - \mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p}\,(\Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p})^{-}$ $\Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}$, $\mathbf{Z}_w = \mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{Z}$, and $\mathbf{P}_{\bar{\mathbf{W}}}\bar{\mathbf{W}} = \bar{\mathbf{W}}$, we have $Q_{2,NT} = [\Delta\mathbf{Z}'\bar{\mathbf{W}} - \mathbf{\Delta Z}'\mathbf{P}_{\bar{W}}\Delta\mathbf{p}\,(\Delta\mathbf{p}'\mathbf{P}_{\bar{W}}\Delta\mathbf{p})^{-}\Delta\mathbf{p}'\bar{\mathbf{W}}]$ $(\bar{\mathbf{W}}'\bar{\mathbf{W}})^{-}$. Using (A.9), we can readily show that

$$\|Q_{2,NT} - Q_2\| = \left\|\left[\frac{\Delta\mathbf{Z}'\bar{\mathbf{W}}}{NT_2} - \frac{\mathbf{\Delta Z}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p}}{NT_2}\left(\frac{\Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p}}{NT_2}\right)^{-}\frac{\Delta\mathbf{p}'\bar{\mathbf{W}}}{NT_2}\right]\left(\frac{\bar{\mathbf{W}}'\bar{\mathbf{W}}}{NT_2}\right)^{-} - Q_2\right\| = O_P\left(L/\sqrt{N}\right)$$

where $Q_2 = [Q_{\bar{w}z}' - Q_{\bar{w}z}'Q_{\bar{w}}^{-1}Q_{\bar{w}p}(Q_{\bar{w}p}'Q_{\bar{w}}^{-1}Q_{\bar{w}p})^{-1}Q_{\bar{w}p}']Q_{\bar{w}}^{-1}$. It follows that $||\,(Q_{2,NT} - Q_2)\,\frac{1}{\sqrt{NT_2}}\bar{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}||$ $\le ||Q_{2,NT} - Q_2||\frac{1}{\sqrt{NT_2}}\left\|\bar{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}\right\| = O_P(L/\sqrt{N})O_P(\sqrt{L}) = o_P(1)$ as $\frac{1}{\sqrt{NT_2}}\left\|\bar{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}\right\| = O_P(\sqrt{L})$ by Chebyshev inequality. Consequently, we can apply the Liapounov CLT to obtain

$$\begin{aligned}\frac{1}{\sqrt{NT_2}}\mathbf{Z}_w'\mathbf{M}_{\mathbf{p}_w}\mathbf{P}_{\bar{\mathbf{W}}}\Delta\boldsymbol{\varepsilon} &= \frac{1}{\sqrt{NT_2}}Q_2\bar{\mathbf{W}}\Delta\boldsymbol{\varepsilon} + (Q_{2,NT} - Q_2)\frac{1}{\sqrt{NT_2}}\bar{\mathbf{W}}\Delta\boldsymbol{\varepsilon} \\ &= \frac{1}{\sqrt{NT_2}}Q_2\bar{\mathbf{W}}\Delta\boldsymbol{\varepsilon} + o_P(1) \xrightarrow{D} N(0, Q_2\Omega_1Q_2')\end{aligned}$$

where we use the fact that $\lambda_{\max}(Q_2\Omega_1Q_2') < \infty$ by Assumption A.8. ■

Proof of Theorem 3.2. (i) Using $\hat{\beta}_m = (\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)^- \mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{Y}_w$ and $\mathbf{M}_{\mathbf{Z}_w}\mathbf{Y}_w = \mathbf{M}_{\mathbf{Z}_w}(\mathbf{p}_w\beta_m + \mathbf{P}_{\bar{\mathbf{W}}}\Delta\boldsymbol{\varepsilon} + \mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R})$, we have with probability approaching 1 (w.p.a.1)

$$\begin{aligned} & \hat{m}_{sieve}(u) - m(u) \\ = \;& p^L(u)'\hat{\beta}_m - m(u) = p^L(u)'(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)^- \mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{Y}_w - m(u) \\ = \;& p^L(u)'(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)^- \mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{P}_{\bar{\mathbf{W}}}\Delta\boldsymbol{\varepsilon} + p^L(u)'(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)^- \mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R} + \left[p^L(u)'\beta_m - m(u)\right] \\ \equiv \;& D_{NT,1}(u) + D_{NT,2}(u) + D_{NT,3}(u), \text{ say,} \end{aligned} \tag{A.10}$$

where we use the fact that $\frac{1}{NT_2}\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w$ is asymptotically nonsingular (see (A.14) below). Then by Cauchy-Schwarz inequality, $\int[\hat{m}_{sieve}(u) - m(u)]^2\alpha(u)\,du \leq 3\sum_{j=1}^3\int D_{NT,j}(u)^2\alpha(u)\,du$. By Assumptions A.1 and A.8 and (A.9), one can readily show that

$$\begin{aligned} \left\|\frac{1}{NT_2}\Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p} - Q_{\bar{w}p}'Q_{\bar{w}}^{-1}Q_{\bar{w}p}\right\| &= O_P\left(L/\sqrt{N}\right), \\ \left\|\frac{1}{NT_2}\Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{Z} - Q_{\bar{w}p}'Q_{\bar{w}}^{-1}Q_{\bar{w}z}\right\| &= O_P\left(L/\sqrt{N}\right), \\ \left\|\frac{1}{NT_2}\Delta\mathbf{Z}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{Z} - Q_{\bar{w}z}'Q_{\bar{w}}^{-1}Q_{\bar{w}z}\right\| &= O_P\left(L/\sqrt{N}\right). \end{aligned} \tag{A.11}$$

Let $Q_{3,NT} \equiv \frac{1}{NT_2}\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w$. Then we can apply the triangle inequality and the fact that $\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w = \mathbf{p}_w'\mathbf{p}_w - \mathbf{p}_w'\mathbf{Z}_w(\mathbf{Z}_w'\mathbf{Z}_w)^-\mathbf{Z}_w'\mathbf{p}_w = \Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p} - \Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{Z}(\Delta\mathbf{Z}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{Z})^-\Delta\mathbf{Z}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{p}$ to obtain

$$\|Q_{3,NT} - Q_3\| = O_P(L/\sqrt{N}) \tag{A.12}$$

where $Q_3 \equiv Q_{\bar{w}p}'Q_{\bar{w}}^{-1}Q_{\bar{w}p} - Q_{\bar{w}p}'Q_{\bar{w}}^{-1}Q_{\bar{w}z}\left(Q_{\bar{w}z}'Q_{\bar{w}}^{-1}Q_{\bar{w}z}\right)^{-1}Q_{\bar{w}z}'Q_{\bar{w}}^{-1}Q_{\bar{w}p} > 0$ under Assumption A.8. This implies that w.p.a.1

$$\lambda_{\max}\left(\frac{1}{NT_2}\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\right) \leq \lambda_{\max}(Q_3) + O_P(L/\sqrt{N}) \leq 2\lambda_{\max}(Q_3) \tag{A.13}$$

and

$$\lambda_{\min}\left(\frac{1}{NT_2}\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\right) \geq \lambda_{\min}(Q_3) - O_P(L/\sqrt{N}) \geq \frac{1}{2}\lambda_{\min}(Q_3). \tag{A.14}$$

By the rotational property of the trace operator and the fact that $\text{tr}(AB) \leq \text{tr}(A)\,\lambda_{\max}(B)$ for any symmetric matrix B and p.s.d. matrix A (e.g., Bernstein (2005, p.275)),

$$\begin{aligned} & \int D_{NT,1}(u)^2\alpha(u)\,du \\ = \;& \text{tr}\left[(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)^- \mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{P}_{\bar{\mathbf{W}}}\Delta\boldsymbol{\varepsilon}\Delta\boldsymbol{\varepsilon}'\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)^- Q_{pp,\alpha}\right] \\ \leq \;& \lambda_{\max}(Q_{pp,\alpha})\,\text{tr}\left[\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{P}_{\bar{\mathbf{W}}}\Delta\boldsymbol{\varepsilon}\Delta\boldsymbol{\varepsilon}'\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\;(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)^-(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)^-\right] \\ \leq \;& \lambda_{\max}(Q_{pp,\alpha})\,c_{1NT}^{-2}\frac{1}{(NT_2)^2}\text{tr}\left(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{P}_{\bar{\mathbf{W}}}\Delta\boldsymbol{\varepsilon}\Delta\boldsymbol{\varepsilon}'\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\right) \\ = \;& \lambda_{\max}(Q_{pp,\alpha})\,c_{1NT}^{-2}\frac{1}{(NT_2)^2}\text{tr}\left(\bar{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}\Delta\boldsymbol{\varepsilon}'\bar{\mathbf{W}}(\bar{\mathbf{W}}'\bar{\mathbf{W}})^-\bar{\mathbf{W}}'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\bar{\mathbf{W}}(\bar{\mathbf{W}}'\bar{\mathbf{W}})^-\right) \\ \leq \;& \lambda_{\max}(Q_{pp,\alpha})\,c_{1NT}^{-2}c_{2NT}\frac{1}{(NT_2)^2}\text{tr}\left(\bar{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}\Delta\boldsymbol{\varepsilon}'\bar{\mathbf{W}}\right). \end{aligned}$$

where $Q_{pp,\alpha} = \int p^L(u) p^L(u)' \alpha(u) du$, $c_{1NT} = \lambda_{\min}(\frac{1}{NT_2}\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)$, and $c_{2NT} = \lambda_{\max}((\bar{\mathbf{W}}'\bar{\mathbf{W}})^{-}\bar{\mathbf{W}}'\mathbf{M}_{\mathbf{Z}_w}$ $\mathbf{p}_w\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\bar{\mathbf{W}}(\bar{\mathbf{W}}'\bar{\mathbf{W}})^{-}) = O_P(1)$ by using results in (A.9) and (A.11). In addition, $\operatorname{tr}(\bar{\mathbf{W}}'\Delta\varepsilon\Delta\varepsilon'\bar{\mathbf{W}}) = O_P(L/N)$ by straightforward expectation calculations and Markov inequality. It follows that $\int D_{NT,1}(u)^2$ $\alpha(u) du = O_P(L/N)$. Analogously, noting that $\lambda_{\max}(\mathbf{P}_{\bar{\mathbf{W}}}) = 1$ and $\frac{1}{NT_2}\|\mathbf{R}\|^2 = O_P\left(L^{-2\gamma/(d_x+1)}\right)$ by Assumptions A.1 and A.7 and Lemma A.2 of Su and Jin (2012), we have

$$
\begin{aligned}
\int D_{NT,2}(u)^2 \alpha(u) du &\leq \lambda_{\max}(Q_{pp,\alpha}) c_{1NT}^{-2} \frac{1}{(NT_2)^2} \operatorname{tr}\left(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R}\mathbf{R}'\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\right) \\
&= \lambda_{\max}(Q_{pp,\alpha}) c_{1NT}^{-2} \frac{1}{(NT_2)^2} \operatorname{tr}\left(\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R}\mathbf{R}'\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\right) \\
&\leq \lambda_{\max}(Q_{pp,\alpha}) c_{1NT}^{-2} c_{3NT} \frac{1}{NT_2} \operatorname{tr}\left(\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R}\mathbf{R}'\mathbf{P}_{\bar{\mathbf{W}}}\right) \\
&\leq \lambda_{\max}(Q_{pp,\alpha}) c_{1NT}^{-2} c_{3NT} \frac{1}{NT_2} \operatorname{tr}\left(\mathbf{R}\mathbf{R}'\right) \\
&= O(1) O_P(1) O_P(1) O_P\left(L^{-2\gamma/(d_x+1)}\right) = O_P\left(L^{-2\gamma/(d_x+1)}\right),
\end{aligned}
$$

where $c_{3NT} = \lambda_{\max}\left(\frac{1}{NT_2}\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\right) = \lambda_{\max}\left(\frac{1}{NT_2}\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\right) = O_P(1)$ by (A.13) and the fact that $\lambda_{\max}(B'B) = \lambda_{\max}(BB')$ for any real matrix B. In addition, $\int D_{NT,3}(u)^2 \alpha(u) du = O_P\left(L^{-2\gamma/(d_x+1)}\right)$ by Assumption A.7 and standard arguments (see, e.g., Chen et al. (2005) and Su and Jin (2012)). It follows that $\int [\hat{m}_{sieve}(u) - m(u)]^2 \alpha(u) du = O_P\left(L/N + L^{-2\gamma/(d_x+1)}\right)$.

(ii) The proof is analogous to that in part (i) and thus omitted.

(iii) The proof parallels that of part (i). The major difference is that now we also need to use Assumption A.7(iv) and restrict our attention to the sequence of compact subsets $\{\mathcal{U}_N\}$ of $\mathcal{X}$ that is expanding at controllable rate. To appreciate this, we focus on the analysis of $D_{NT,2}(u)$ and $D_{NT,3}(u)$. As in the proof of (i), we can show that $\left\|(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)^{-}\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R}\right\| = O_P(L^{-\gamma/(d_x+1)})$. Then by Cauchy-Schwarz inequality and Assumption A.7(iv), $\sup_{u\in\mathcal{U}_N}|D_{NT,2}(u)| \leq \|p^L(u)\| \left\|(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w)^{-}\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R}\right\| = \zeta_0(L) O_P(L^{-\gamma/(d_x+1)})$. By Assumptions A.7(ii) and (iv),

$$
\begin{aligned}
\sup_{u\in\mathcal{U}_N}|D_{NT,3}(u)| &= \sup_{u\in\mathcal{U}_N}\left|\left[p^L(u)'\beta_m - m(u)\right]\left(1+\|u\|^2\right)^{-\varpi/2}\right|\left(1+\|u\|^2\right)^{\varpi/2} \\
&\leq \|m - \Pi_{\infty,L}m\|_{\infty,\varpi} \sup_{u\in\mathcal{U}_N}\left(1+\|u\|^2\right)^{\varpi/2} = O(L^{-\gamma/(d_x+1)})\zeta_0(L).
\end{aligned}
$$

(iv) By the decomposition in (A.10), we can prove the theorem by showing that (iv1) $\mathcal{D}_{NT,1} \equiv \sqrt{NT_2}A_L^{-1/2}(u) D_{NT,1}(u) \xrightarrow{D} N(0,1)$, (iv2) $\mathcal{D}_{NT,2} \equiv \sqrt{NT_2}A_L^{-1/2}(u) D_{NT,2}(u) = o_P(1)$, and (iv3) $\mathcal{D}_{NT,3} \equiv \sqrt{NT_2}A_L^{-1/2}(u) D_{NT,3}(u) = o_P(1)$. By the fact that $\|p^L(u)\| \geq c > 0$ and Assumption A.8,

$$
\begin{aligned}
A_L(u) &= p^L(u)' Q_3^{-1} Q_4 \Omega_1 Q_4' Q_3^{-1} p^L(u) \geq \lambda_{\min}(Q_4\Omega_1 Q_4') \lambda_{\min}\left(Q_3^{-1}Q_3^{-1}\right)\|p^L(u)\| \\
&\geq \lambda_{\min}(\Omega_1)[\lambda_{\max}(Q_3)]^{-2}\|p^L(u)\|^2 > 0. \qquad \text{(A.15)}
\end{aligned}
$$

Let $Q_{4,NT} \equiv \mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\bar{\mathbf{W}}(\bar{\mathbf{W}}'\bar{\mathbf{W}})^{-}$. Noting that $\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\bar{\mathbf{W}} = \mathbf{p}_w'\bar{\mathbf{W}} - \mathbf{p}_w'\mathbf{Z}_w(\mathbf{Z}_w'\mathbf{Z}_w)^{-}\mathbf{Z}_w'\bar{\mathbf{W}} = \Delta\mathbf{p}'\bar{\mathbf{W}}$ $-\Delta\mathbf{p}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{Z}(\Delta\mathbf{Z}'\mathbf{P}_{\bar{\mathbf{W}}}\Delta\mathbf{Z})^{-}\Delta\mathbf{Z}'\bar{\mathbf{W}}$, we can readily apply (A.9) and (A.11) to show that

$$
\|Q_{4,NT} - Q_4\| = O_P\left(L/\sqrt{N}\right) \qquad \text{(A.16)}
$$

where $Q_4 \equiv Q'_{\bar{w}p}Q_{\bar{w}}^{-1} - Q'_{\bar{w}p}Q_{\bar{w}}^{-1}Q_{\bar{w}z}\left(Q'_{\bar{w}z}Q_{\bar{w}}^{-1}Q_{\bar{w}z}\right)^{-1}Q'_{\bar{w}z}Q_{\bar{w}}^{-1}$. Note that

$$
\begin{aligned}
\mathcal{D}_{NT,1} &= A_L^{-1/2}(u)\, p^L(u)'\, Q_{3,NT}^{-} Q_{4,NT} \frac{\bar{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}}{\sqrt{NT_2}} \\
&= A_L^{-1/2}(u)'\, p^L(u)'\, Q_3^{-1}Q_4 \frac{\bar{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}}{\sqrt{NT_2}} + A_L^{-1/2}(u)\, p^L(u)'\,(Q_{3,NT}^{-}Q_{4,NT} - Q_3^{-1}Q_4)\frac{\bar{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}}{\sqrt{NT_2}} \\
&\equiv \mathcal{D}_{NT,11} + \mathcal{D}_{NT,12}, \text{ say.}
\end{aligned}
$$

It is standard to verify the Liapounov conditions and show that $\mathcal{D}_{NT,11} \xrightarrow{D} N(0,1)$. By (A.12), (A.16), (A.15), and the fact that $\frac{1}{\sqrt{NT_2}}||\bar{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}|| = O_P(\sqrt{L})$, $|\mathcal{D}_{NT,12}| \leq A_L^{-1/2}(u)\left\|p^L(u)\right\|' ||Q_{3,NT}^{-}Q_{4,NT}$ $-Q_3^{-1}Q_4||\ \frac{1}{\sqrt{NT_2}}\left\|\bar{\mathbf{W}}'\Delta\boldsymbol{\varepsilon}\right\| \leq [\lambda_{\min}(\Omega_1)]^{-1/2}\lambda_{\max}(Q_3)\ O_P(L/\sqrt{N})O_P(\sqrt{L}) = O_P(\sqrt{L^3/N}) = o_P(1)$. Then (iv1) follows.

Next, by the Cauchy-Schwarz inequality, the fact that $a'Ba \leq a'a\lambda_{\max}(B)$ for any vector a and conformable p.s.d. symmetric matrix B, that $\lambda_{\max}(C'C) = \lambda_{\max}(CC')$ for any real matrix C, and that $\lambda_{\max}(\mathbf{P}_{\bar{\mathbf{W}}}) = 1$, we have

$$
\begin{aligned}
\mathcal{D}_{NT,2}^2 &\leq NT_2 c_{NT}\mathbf{R}'\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\left(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\right)^{-}\left(\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\right)^{-}\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R} \\
&\leq c_{NT}\left[\lambda_{\min}(Q_{3,NT})\right]^{-2}\frac{1}{NT_2}\mathbf{R}'\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R} \\
&\leq c_{NT}\left[\lambda_{\min}(Q_{3,NT})\right]^{-2}\lambda_{\max}\left(\frac{1}{NT_2}\mathbf{M}_{\mathbf{Z}_w}\mathbf{p}_w\mathbf{p}_w'\mathbf{M}_{\mathbf{Z}_w}\right)\mathbf{R}'\mathbf{P}_{\bar{\mathbf{W}}}\mathbf{R} \\
&\leq c_{NT}\left[\lambda_{\min}(Q_{3,NT})\right]^{-2}\lambda_{\max}(Q_{3,NT})\left\|\mathbf{R}\right\|^2 \\
&= O(1)\,O_P(1)\,O_P(1)\,O_P\left(NL^{-2\gamma/(d_x+1)}\right) = o_P(1)
\end{aligned}
$$

where $c_{NT} \equiv A_L^{-1}(u)\,p^L(u)'\,p^L(u) \leq [\lambda_{\min}(\Omega_1)]^{-1}[\lambda_{\max}(Q_3)]^2 = O(1)$ by (A.15) and the next to last equality follows because $N^{-1}\left\|\mathbf{R}\right\|^2 = O_P\left(L^{-2\gamma/(d_x+1)}\right)$. Then (iv2) follows. By (A.15) and Assumptions A.8(iii) and A.9,

$$
\begin{aligned}
|\mathcal{D}_{NT,3}| &= \sqrt{NT_2}A_L^{-1/2}(u)\left|\left[p^L(u)'\beta_m - m(u)\right]\left(1+\|u\|^2\right)^{-\varpi/2}\right|\left(1+\|u\|^2\right)^{\varpi/2} \\
&\leq \sqrt{NT_2}A_L^{-1/2}(u)\left\|m - \Pi_{\infty,L}m\right\|_{\infty,\varpi}\left(1+\|u\|^2\right)^{\varpi/2} = O(\sqrt{N}L^{-\gamma/(d_x+1)}) = o(1).
\end{aligned}
$$

Thus (iv3) follows. ■

B Data

List of countries/regions (93)

Algeria, Argentina, Australia, Austria, Bangladesh, Belgium, Benin, Bolivia, Botswana, Brazil, Burundi, Cameroon, Canada, Cyprus, Central African Rep., Chile, China, Colombia, Congo, Costa Rica, Denmark, Dominican Rep., Ecuador, Egypt, El Salvador, Fiji, Finland, France, Gabon, Ghana, Greece, Guatemala, Guyana, Haiti, Honduras, Hong Kong, India, Iceland, Indonesia, Iran, Ireland, Ivory Coast, Israel, Italy, Jamaica, Japan, Jordan, Kenya, Korea (South), Malawi, Malaysia, Mali, Malta, Mauritius, Mexico, Morocco, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Pakistan, Norway, Papua New

Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Rwanda, Senegal, Sierra Leone, Singapore, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Syria, Tanzania, Thailand, Togo, Trinidad & Tobago, Tunisia, Turkey, Uganda, United Kingdom, United States, Uruguay, Venezuela, Zambia, Zimbabwe.

Lists of variables

Variables	Definition	Source
GDP growth rate	Growth rate of real GDP per capita (constant 2005 prices)	UNCTAD
IPR	updated GP index of patent rights	Park (2008)
Foreign Direct Investment	Inward FDI flows (US Dollars at current prices and current exchange rates in millions)	UNCTAD
Government consumption	General Government final consumption expenditure as a share of real GDP	UNCTAD
Schooling	Percentage of secondary schooling attained in population	Barro and Lee (2013)
Domestic Investment	Gross capital formation (US Dollars at current prices and current exchange rates in millions)	UNCTAD
Openness	Fraser Institute's Index of Freedom to trade internationally	Fraser Institute (Gwartney et al., 2010)
Population growth rate	five-year average annual growth rate	UNCTAD
Inflation	Percentage change in the GDP deflator	World Bank

References

Ai, C., & Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica,* 71, 1795-1843.

Anderson, T.W., & Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association,* 76, 598-606.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Reviews of Economic Studies,* 58, 277-297.

Baglan, D. (2010). Efficient estimation of a partially linear dynamic panel data model with fixed Effects: application to unemployment dynamics in the U.S.. *Working Paper*, Dept. of Economics, Howard University.

Baltagi, B.H., & Li, D. (2002). Series estimation of partially linear panel data models with fixed effects. *Annals of Economic and Finance,* 3, 103-116.

Baltagi, B.H., & Li, Q. (2002). On instrumental variable estimation of semiparametric dynamic panel data models. *Economics Letters,* 76, 1-9.

Barro, R.J., & Lee, J.W. (2013). A new data set of educational attainment in the world, 1950-2010. *Journal of Development Economics*,104, 184-198.

Bernstein, D. S. (2005). *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory.* Princeton University Press, Princeton.

Branstetter, L.G., Fisman, R., & Foley, F.C. (2006). Do stronger intellectual property rights increase international technology transfer? Empirical evidence from US firm-level panel data? *The Quarterly Journal of Economics,* 121, 321-49.

Carrasco, M., Florens, J.P., & Renault, E. (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In J.J. Heckman and E. Leamer (eds), *Handbook of Econometrics*, Vol. 6, pp. 5633-5751, North Holland, Amsterdam.

Chen, J., Gao, J., & Li, D. (2012). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics,* 171, 71-85.

Chen, J., Gao, J., & Li, D.(2013a). Estimation in partially linear single-index panel data models with fixed effects. *Journal of Business & Economic Statistics,* 31, 315-330.

Chen, J., Gao, J., & Li, D. (2013b). Estimation in a single-index panel data model with heterogeneous link functions. *Econometric Reviews,* 33, 928-955.

Chen, J., Li, D., & Gao, J. (2013). Non- and semi-parametric panel data models: a selective review. *Working Paper*, University of York.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In J. J. Heckman and E. Leamer (eds), *Handbook of Econometrics*, Vol. 6, pp. 5549-5632, North Holland, Amsterdam.

Chen, X., Hong, H., & Tamer E. (2005). Measurement error models with auxiliary data. *Review of Economic Studies* 72, 343-366.

Chen, X., Linton, O., & Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica,* 71, 1591-1608.

Chen, X., & Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica,* 80, 277-321.

Chen, Y., & Puttitanun, T. (2005). Intellectual property rights and innovation in developing countries. *Journal of Development Economics,* 78, 474-493.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields,* 75, 261-277.

de Jong, R.M. (2002). A note on "Convergence rates and asymptotic normality for series estimators": uniform convergence rates. *Journal of Econometrics,* 111, 1-9.

Darolles, S., Fan, Y., Florens, J., & Renault, E. (2011). Nonparametric instrumental regression. *Econometrica,* 79, 1541-1565.

Delgado, M.A., & González-Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics,* 29, 1469-1507.

Dinopoulos, E., & Segerstrom, P. (2010). Intellectual property rights, multinational firms and economic growth. *Journal of Development Economics,* 92, 13-27.

Dong, C., Gao, J., & B. Peng, B. (2014). Semiparametric single-index panel data regression. *Working Paper*, Monash University.

Dong, C., Gao, J., & B. Peng, B. (2015). Estimation in a semiparametric panel data model under cross-sectional dependence and nonstationarity. *Working Paper*, Monash University.

Eicher, T., & Garcia-Peñalosa, C. (2008). Endogenous strength of intellectual property rights: implications for economic development and growth. *European Economic Review,* 52, 237-258.

Engle, R.F., Granger, C., Rice, J., & Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association,* 81, 310-320.

Escanciano, J., Jacho-Chávez, D., & Lewbel, A. (2014). Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of Econometrics,* 178, 426-473.

Falvey, R., Foster, N., & Greenaway, D. (2009). Trade, imitative ability and intellectual property rights. *Review World Economics,* 145, 373-404.

Fan, Y., & Li, Q. (1996) Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica,* 64, 865-890.

Feng, G., Gao, J., Peng, B., & Zheng, X. (2015). A varying coefficient panel data model with fixed effects: theory and an application to the US commercial banks. *Working Paper,* Monash University.

Florens, J.P., Johannes, J., & Van Bellegem, S. (2012). Instrumental regression in partially linear models. *The Econometrics Journal,* 15, 304-324.

Furukawa, Y. (2007). The protection of intellectual property rights and endogenous growth: Is stronger always better? *Journal of Economic Dynamic & Control,* 31, 3644-3670.

Gwartney, J., Hall, D., Joshua, C., & Lawson, R. (2010). 2010 Economic Freedom Dataset. *Economic Freedom Network*.

Glass, A.J., & Wu, X. (2007). Intellectual property rights and quality improvement. *Journal of Development Economics,* 82, 393-415.

Groizard, J.L. (2009). Technology trade. *Journal of Development Studies,* 45, 1526-1544.

Hahn, J., & Ridder, G. (2013). Asymptotic variance for semiparametric estimators with generated regressors. *Econometrica,* 81, 315-340.

Hansen, B.E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory,* 24, 726-748.

Härdle, W., & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics,* 21, 1926-1947.

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

Ichimura, H., & Lee, S. (2010). Characterization of the asymptotic distribution of semiparametric M-estimators. *Journal of Econometrics,* 159, 252-266.

Kanwar, S., & Evenson, R. (2003). Does intellectual property protection spur technological change? *Oxford Economic Papers,* 55, 235-85.

Kim, N.H., & Saart, P. (2013). Estimation in partially linear semiparametric models with parametric and/or nonparametric endogeneity. *Working Paper*, University of Adelaide.

Kress, R. (1997). *Linear Integral Equation*. Springer, New York.

Kong, E., Linton, O., & Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory* 26, 1529-1564.

Lee, A.J. (1990). *U-statistics: Theory and Practice*. Marcel Dekker, New York.

Lee, Y. (2014). Nonparametric estimation of dynamic panel models with fixed effects. *Econometric Theory,* 30, 1315-1347.

Lee, Y.-J. (2013). Testing a linear dynamic panel data model against nonlinear alternatives. *Journal of Econometrics,* 178, 146-166.

Li, Q., Hsiao, C., & Zinn, J. (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics,* 112, 295-325.

Li, Q., & Racine, J. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Li, Q., & Stengos, T. (1996). Semiparametric estimation of partially linear panel data models. *Journal of Econometrics,* 71, 289-397.

Li, Q., & Ullah, A. (1998). Estimating partially panel data models with one-way error components. *Econometric Reviews,* 17, 145-166.

Li, N., Xu, X., & Jin, P. (2011). Testing the linearity in partially linear models. *Journal of Nonparametric Statistics,* 23, 99-114.

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.

Linton O.B., & Mammen, E. (2005). Estimating semiparametric ARCH(∞) model by kernel smoothing methods. *Econometrica,* 73, 771-836.

Mammen, E., Linton, O.B., & Nielsen, J.P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics,* 27, 1443-1490.

Mammen, E., Rothe, C., & Schienle, M. (2015). Semiparametric estimation with generated covariates. Forthcoming in *Econometric Theory.*

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

Mammen, E., & Yu, K. (2009). Nonparametric estimation of noisy integral equations of the second kind. *Journal of the Korean Statistical Society,* 38, 99-110.

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

Newey, W.K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica,* 62, 1349-1382.

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

Nielsen, J.P., & Sperlich, S. (2005). Smooth backfitting in practice. *Journal of the Royal Statistical Society, Series B,* 67, 43-61.

Panagopoulos, A. (2009). Revisiting the link between knowledge spillovers and growth: an intellectual property perspective. *Economics of Innovation and New Technology,* 18, 533-546.

Park, W.G. (2008). International patent protection: 1965-2005. *Research Policy* 37, 761-766.

Park, W.G., & Ginarte, J.C. (1997). Intellectual property rights and economic growth. *Contemporary Economic Policy,* Vol. XV, 9-1–9-51.

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

Robinson, P.M. (1988). Root-n-consistent semiparametric regression. *Econometrica,* 56, 931-954.

Rust, J. (2000). Nested fixed point algorithm documentation manual. Version 6, Yale University.

Su, L., & Jin, S. (2012). Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics,* 169, 34-47.

Su, L., & Lu, X. (2013). Nonparametric dynamic panel data models: kernel estimation and specification testing. *Journal of Econometrics,* 176, 112-133.

Su, L., & Ullah, A. (2006a). Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters,* 92, 75-81.

Su, L., & Ullah, A. (2006b). More efficient estimation in nonparametric regression with nonparametric autocorrelated errors. *Econometric Theory,* 22, 98-126.

Su, L., & Ullah, A. (2011). Nonparametric and semiparametric panel econometric models: estimation and testing, in *Handbook of Empirical Economics and Finance.* A. Ullah and D.E.A. Giles (eds)., pp. 455-497. Taylor & Francis Group, New York.

Su, L., & Zhang, Y. (2015). Nonparametric dynamic panel data models with interactive fixed effects: sieve estimation and specification testing. *Working Paper*, School of Economics, Singapore Management University.

Sun, Y., Zhang, Y.Y. & Li, Q. (2015). Nonparametric panel data regression models. In B.H. Baltagi (eds), *The Oxford Handbook of Panel Data Econometrics*, pp. 285-324. Oxford University Press, Oxford.

Van der Vaart, A., & Wellner, J.A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York.

Xiao, Z., Linton, O.B., Carroll, R.J., & Mammen, E. (2003). More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *Journal of American Statistical Association,* 98, 980-992.

Yao, F., & Zhang, J. (2015). Efficient kernel-based semiparametric IV estimation with an application to resolving a puzzle on the estimation of the return to schooling. *Empirical Economics,* 48, 253-281.