1-2013

Nonparametric Dynamic Panel Data Models: Kernel Estimation and Specification Testing

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

Xun LU
Hong Kong University of Science and Technology

Nonparametric Dynamic Panel Data Models: Kernel Estimation and Specification Testing *

Liangjun Su[a] and Xun Lu[b]

[a] School of Economics, Singapore Management University

[b] Department of Economics, Hong Kong University of Science and Technology

January 4, 2012

Abstract

Motivated by the first differencing method for linear panel data models, we propose a class of iterative local polynomial estimators for nonparametric dynamic panel data models with or without exogeous regressors. The estimators utilize the additive structure of the first-differenced model, the fact that the two additive components have the same functional form, and the unknown function of interest is implicitly defined as a solution of a Fredholm integral equation of the second kind. We establish the uniform consistency and asymptotic normality of the estimators. We also propose a consistent test for the correct specification of linearity in typical dynamic panel data models based on the L_2 distance of our nonparametric estimates and the parametric estimates under the linear restriction. We derive the asymptotic distributions of the test statistic under the null hypothesis and a sequence of Pitman local alternatives, and prove its consistency against global alternatives. Simulations suggest that the proposed estimators and tests perform well in finite samples. We apply our new methods to study the relation between economic growth, initial economic condition and capital accumulation and find the nonlinear relation between economic growth and initial economic condition.

Key words: Additive models, Dynamic panel data models, Fredholm integral equation, Iterative estimator, Linearity, Local polynomial regression, Specification test

JEL Classification: C14, C33, C36.

*Address correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903; Phone: +65 6828 0386; e-mail: ljsu@smu.edu.sg. We would like to express our appreciation to Qi Li, Zudi Lu, Hashem Pesaran, and Yiguo Sun for discussions on the subject matter. We also thank the participants at the seminars at National Technological University and National University of Singapore who provided valuable suggestions and discussions. The first author gratefully acknowledges financial support from the SMU research grant (Grant number: 11-C244-SMU-002).

1 Introduction

There exists an enormous literature on parametric (often linear) panel data models; see the books by Arellano (2003), Hsiao (2003), and Baltagi (2008) for an excellent overview. Nevertheless, the parametric functional forms may be misspecified and estimators based on misspecified models are often inconsistent and thus invalidate subsequent statistical inferences. For this reason, we also observe a rapid growth of the literature on nonparametric (NP) and semiparametric (SP) panel data models in the last two decades. See Su and Ullah (2011) for a recent survey on this topic.

To the best of our knowledge, there lacks satisfactory development in NP or SP dynamic panel data models where a lagged dependent variable enters the nonparametric component instead of a parametric (usually linear) component of the models. For NP panel data models with fixed effects, the main focus has been on the model

$$Y_{it} = m\left(X_{it}\right) + \alpha_i + \varepsilon_{it}, \tag{1.1}$$

$i = 1, ..., N$, $t = 1, ..., T$, where the functional form of $m(\cdot)$ is not specified, the covariate X_{it} is of $d \times 1$ dimension, and α_i is a fixed effect that can be correlated with X_{it}, and ε_{it}'s are idiosyncratic error terms. Motivated by the first differencing method for linear panel data models, one can consider the following first-differenced model

$$\Delta Y_{it} = m\left(X_{it}\right) - m\left(X_{i,t-1}\right) + \Delta\varepsilon_{it} \tag{1.2}$$

where $\Delta Y_{it} \equiv Y_{it} - Y_{i,t-1}$ and $\Delta\varepsilon_{it} \equiv \varepsilon_{it} - \varepsilon_{i,t-1}$. Li and Stengos (1996) suggest estimating $m(X_{it}, X_{i,t-1}) \equiv m(X_{it}) - m(X_{i,t-1})$ by first running a local linear regression of ΔY_{it} on X_{it} and $X_{i,t-1}$, and then obtaining estimates of $m(\cdot)$ by standard methods of estimating nonparametric additive models, e.g., by the marginal integration method of Linton and Nielson (1995) or by the backfitting method (e.g., Opsomer and Ruppert (1997), and Mammen, Linton and Nielsen (1999)). Apparently, this method suffers from the "curse of dimensionality" problem because the first step local linear regression involves estimating a $2d$-dimensional nonparametric object, and it does not utilize the fact that the two additive components share the same functional form. In view of this, Baltagi and D. Li (2002) obtain consistent estimators of $m(\cdot)$ by considering the first differencing method and using series approximation for the nonparametric component.[1] Also based on the difference model in (1.2), Henderson, Carroll and Li (2008) introduce an iterative nonparametric kernel estimator of $m(\cdot)$ and conjecture its asymptotic bias and variance and asymptotic normality. The crucial assumptions in this latter paper include: 1) ε_{it}'s are independent and identically distributed (IID) across both i and t, and are independent of X_{it}, and 2) there exists a consistent initial estimator. More recently, Lee (2010) considers the sieve estimation of (1.1) when $X_{it} = Y_{i,t-1}$ via within-group transformation. That is, he considers the following transformed model

$$Y_{it} - T^{-1}\sum_{s=1}^{T} Y_{is} = m\left(Y_{i,t-1}\right) - T^{-1}\sum_{s=1}^{T} m\left(Y_{i,s-1}\right) + \varepsilon_{it} - T^{-1}\sum_{s=1}^{T}\varepsilon_{is} \tag{1.3}$$

[1] Both Li and Stengos (1996) and Baltagi and D. Li (2002) consider a more general model, namely, a partially linear model.

and approximate the unknown smooth function $m(\cdot)$ by some basis functions. Under the assumption that $\lim_{N,T\to\infty} N/T = \kappa \in (0,\infty)$, he finds that the series estimator is asymptotically biased. So he proposes a bias-corrected series estimator and establishes its asymptotic normality.

Another method that is adopted to estimate the model in (1.1) is based on the profile likelihood or least squares method in the statistical literature. For example, Su and Ullah (2006a) propose to estimate the unknown function by profile least squares under the identification condition $\sum_{i=1}^{N} \alpha_i = 0$, which boils down to a local linear analogue of the least squares dummy variable (LSDV) estimator for typical linear panel data models. Under the weaker identification condition that $E(\alpha_i) = 0$, Li and Sun (2011) propose to mimic the parametric LSDV estimation method by removing the fixed effects nonparametrically and establish the asymptotic normality of their estimator under the assumption that ε_{it} are independent of α_j and $E(\varepsilon_{it}|\mathcal{X}) = 0$ for all i, j and t where $\mathcal{X} = \{X_{js}, j = 1, ..., N, s = 1, ..., T\}$, and that $T \to \infty$ sufficiently fast as $N \to \infty$.

In addition, it is worth mentioning that there are also some studies on semiparametric panel data models that include the model in (1.1) as a special case. One example is the paper by Sun, Carroll and Li (2009) who consider the local linear estimation of the varying coefficient panel data models with fixed effects

$$Y_{it} = Z_{it}'\theta(X_{it}) + \alpha_i + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, ..., T, \tag{1.4}$$

where the idiosyncratic error terms ε_{it} are independent of X_{js}, Z_{js} and α_j for all i,j,t, and s. Note that this model reduces to (1.1) when $Z_{it} \equiv 1$. Obviously, they do not allow $Y_{i,t-1}$ to enter X_{it}. Another example is the partially linear panel data models with fixed effects: Baltagi and Q. Li (2002) propose a semiparametric instrumental variable estimator for estimating a partially linear dynamic panel data models; Qian and Wang (2011) consider the marginal integration estimator of the nonparametric additive component resulting from the first differencing of a partially linear panel data model. Unfortunately, none of these models allow the lagged dependent variable to enter the nonparametric component. In addition, Mammen, Støve and Tjøstheim (2009) consider the consistent estimation of nonparametric additive panel data models with time effects or with both time and individual effects via backfitting. But they only establish the asymptotic normality of the resulting estimator in the presence of time effects only.

In this paper, we propose an iterative kernel estimation of nonparametric dynamic models of the form

$$Y_{it} = m(Y_{i,t-1}, X_{it}) + \alpha_i + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, ..., T, \tag{1.5}$$

where X_{it} is a $d \times 1$ vector of regressors, $m(\cdot,\cdot)$ is an unknown but smooth function defined on $\mathbb{R}^{d+1}$, α_i's are the individual-specific fixed effects, and ε_{it}'s are idiosyncratic error terms. Let $\underline{X}_{it} \equiv (X_{it}', ..., X_{i1}')'$ and $\underline{Y}_{i,t-1} \equiv \left(Y_{i,t-1}', ..., Y_{i1}'\right)'$. We assume that $E\left(\varepsilon_{it}|\underline{Y}_{i,t-1}, \underline{X}_{it}\right) = 0$ and consider the first-differenced model

$$\Delta Y_{it} = m(Y_{i,t-1}, X_{it}) - m(Y_{i,t-2}, X_{i,t-1}) + \Delta\varepsilon_{it}. \tag{1.6}$$

Apparently, (1.6) has a simple additive structure. But it is different from the typical additive models in two aspects. First, the error term $\Delta\varepsilon_{it}$ forms a moving average process of order 1 [MA(1)] and is correlated with the regressor $Y_{i,t-1}$ in general. Second, the two additive components in (1.6) share the same functional form. The first observation indicates that the traditional kernel estimation based on either marginal integration or backfitting method does not yield a consistent estimator of $m(\cdot, \cdot)$. The second observation, in conjunction with the fact that $E[\Delta\varepsilon_{it}|Y_{i,t-2}, X_{i,t-1}] = 0$, implies that $m(\cdot, \cdot)$ implicitly solves a Fredholm integral equation of the second kind [see (2.7) in Section 2.1], so that we can propose a simple local polynomial regression-based iterative estimator for it.

Under fairly general conditions which allow nonstationarity of $(Y_{i,t-1}, X_{it}, \varepsilon_{it})$ along the time dimension and conditional heteroskedasticity among ε_{it}, we establish the uniform consistency of the proposed estimator over a compact set and study its asymptotic normality by passing the cross sectional unit N to ∞ and holding the time dimension T as a fixed constant as in typical micro panel data models. We also remark that under some suitable conditions, one can plug our estimate of $m(Y_{i,t-1}, X_{it})$ into (1.6) to obtain a new estimate of $m(\cdot)$ to achieve certain "oracle" properties.

Based on our kernel estimator, we also propose a test for the correct specification of linear dynamic panel data models. There have been various specification tests for parametric panel data models in the literature; see Hausman (1978), Hausman and Taylor (1981), Arellano (1990), Arellano and Bond (1991), Li and Stengos (1992), Metcalf (1996), Baltagi (1999), Fu, Li and Fung (2002), Inoue and Solon (2006), Okui (2009), among others. Nevertheless, none of these tests are designed to check the correct specification of linearity in the panel data models. Recently, Lee (2011) proposes a class of residual-based specification tests for linearity in dynamic panel data models by characterizing the correct specification of linearity as the martingale difference property of the error terms in the model and extending the generalized spectral analysis of Hong (1999) to dynamic panel data models. To eliminate the problem of incidental parameters, she focuses on dynamic panel data models with both large N and large T. So her test can not be applied to typical micro panel data where T is usually small.

In this paper, we consider a specification test for the linearity of a dynamic panel data model when N is large and T is small/fixed. Under the null hypothesis of correct specification of linear dynamic panel data models, various methods can be called upon to estimate the unknown parameters in the linear regression model. Under the alternative, the functional form of the regression model is left unspecified as in (1.5) and one can estimate the unknown function by using the method proposed in this paper. We base our test statistic on the L_2 distance between the two functional estimates in the spirit of Härdle and Mammen (1993), and study its asymptotic properties under the null hypothesis, a sequence of Pitman local alternatives, and global alternatives.

We use Monte Carlo simulations to examine the finite sample performance of our estimators and tests. Our iterative estimators can reduce the root mean square error (RMSE) of the initial estimators (the nonparametric sieve estimators) by 30-40%. Both the levels and powers of our test perform well in finite samples. We apply our new method to two empirical studies. In the first application, we study the relationship between economic growth, initial economic condition and capital accumulation. We find substantial nonlinearity in the relation between a country's economic growth and its initial economic condition. We find that the very poor and very rich countries tend to have relatively low economic growth rates, while medium initial income countries tend to enjoy fast economic growth. In the second application, we examine the relation between a firm's labor inputs and its sales. We do not find nonlinear relationship between them, thus we validate the use of linear models in this context.

The rest of the paper is organized as follows. In Section 2 we introduce the iterative kernel estimator for nonparametric dynamic panel data models and study its asymptotic properties. In Section 3, we propose a consistent test for the correct specification of linear panel data models that are a routine in empirical studies. The test statistic is based on the L_2 distance between the kernel estimate of the nonparametric regression function under the alternative and the parametric estimate of the linear dynamic panel data models under the null. In Section 4, we conduct a small number of Monte Carlo simulations to evaluate the finite sample performance of our estimators and tests. We apply our method to study (i) the relation between economic growth, initial economic condition and capital accumulations and (ii) the relation between a firm's sales and its labor inputs in Section 5. Final remarks are contained in Section 6. All technical details are relegated to the Appendix.

Throughout the paper, we restrict our attention to the balanced panel. We use $i = 1, \cdots, N$ to denote an individual and $t = 1, \cdots, T$ 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 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 $l = 1, 2$. We use $\otimes$ and $\odot$ denote the Kronecker and Hadarmard products, respectively. For conformable vectors u and v, we use u/v to denote elementwise division. $\overset{P}{\to}$ and $\overset{D}{\to}$ signify convergence in probability and distribution, respectively.

2 Kernel estimation of nonparametric dynamic panel data models

In this section, we first propose a kernel estimator for nonparametric dynamic panel data models based on a Fredholm integral equation of the second kind. Then we study the uniform consistency and asymptotic normality of the proposed estimator.

2.1 Kernel estimation based on Fredholm integral equation of the second kind

We consider the following nonparametric dynamic panel data models

$$Y_{it} = m\left(Y_{i,t-1}, X_{it}\right) + \alpha_i + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, ..., T, \tag{2.1}$$

where X_{it} is a $d \times 1$ vector of regressors, $m(\cdot,\cdot)$ is an unknown but smooth function defined on $\mathbb{R}^{d+1}$, α_i's are the individual-specific fixed effects, and ε_{it}'s are idiosyncratic error terms. We assume that $(Y_{it}, X_{it}, \varepsilon_{it})$ are independently and identically distributed (IID) along the individual dimension but may not be strictly stationary along the time dimension. In addition, we assume that $E\left(\varepsilon_{it}|\underline{Y}_{i,t-1}, \underline{X}_{it}\right) = 0$, where $\underline{Y}_{i,t-1} \equiv \left(Y'_{i,t-1}, ..., Y'_{i1}\right)'$ and $\underline{X}_{it} \equiv \left(X'_{it}, ..., X'_{i1}\right)'$.

To proceed, we make several remarks on the model in (2.1). First, even though we signify the dynamic nature of our nonparametric model, the asymptotic theory developed in this paper goes through if $Y_{i,t-1}$ is missing from $m(\cdot,\cdot)$, or if X_{it} contains higher order lagged dependent variables such as $Y_{i,t-2}$ and $Y_{i,t-3}$. Second, X_{it} may contain some time-invariant regressors and our estimation strategy can recover their effects on Y_{it}. This is in sharp contrast with the typical dynamic panel data models because time-invariant regressors will be wiped out after first-differencing. This observation allows us to re-investigate the long-run relationship between economic growth, initial economic condition (which is time-invariant), and capital accumulation through our nonparametric dynamic panel data models. Third, to stay focused, we do not allow time effects in (2.1). That is, one cannot have the one-way component structure for ε_{it}: $\varepsilon_{it} = \gamma_t + v_{it}$. The inclusion of time effects will significantly complicate the analysis and be left for future research.

Motivated by the first-differencing method for linear dynamic panel data models, we consider the following first-differenced model

$$\Delta Y_{it} = m(Y_{i,t-1}, X_{it}) - m(Y_{i,t-2}, X_{i,t-1}) + \Delta\varepsilon_{it} \tag{2.2}$$

where $\Delta Y_{it} \equiv Y_{it} - Y_{i,t-1}$ and $\Delta\varepsilon_{it} \equiv \varepsilon_{it} - \varepsilon_{i,t-1}$. Apparently, the above model is an additive model where the two additive components share the same functional form. In addition, as in the linear dynamic panel data models, $\Delta\varepsilon_{it}$ is correlated with the regressor $Y_{i,t-1}$ on the right hand side of (2.2) so that one has to take into account the endogeneity issue in order to estimate the model. These observations indicate that the two standard techniques in the kernel literature to handle additive models, namely, marginal integration and backfitting, are not appropriate without proper modifications. In principle, one can modify either technique to take into account the two additional features of the above model. But we are not sure whether the modification is straightforward. One thing that seems transparent to us is that the marginal integration method is involved with a higher dimension estimation of the two additive components in the first step and may not utilize the above two specific features effectively.

In this paper, we consider a kernel estimate of m by taking into account both features mentioned above. In view of the fact that $\Delta\varepsilon_{it}$ is (conditionally) mean-independent of $\underline{U}_{i,t-2} \equiv (\underline{Y}'_{i,t-2}, \underline{X}'_{i,t-1})'$, we obtain the following conditional moment conditions

$$E\left[\Delta Y_{it} - m(Y_{i,t-1}, X_{it}) + m(Y_{i,t-2}, X_{i,t-1}) \,|\underline{U}_{i,t-2}\right] = 0. \tag{2.3}$$

Clearly, for large t the conditioning information set $\underline{U}_{i,t-2}$ contains a large number of valid instrument variables (IVs) for the local nonparametric identification of m. But for technical reasons, it is unrealistic to use all variables in $\underline{U}_{i,t-2}$ for our nonparametric regression. So we consider only a small number of IVs that are measurable with respect to $\underline{U}_{i,t-2}$. In fact, as the following estimation strategy suggests, we only consider $U_{i,t-2} \equiv (Y_{i,t-2}, X'_{i,t-1})'$ and leave the efficient choice of IVs for future research.

To proceed, we define some notation. Let $\mathcal{U}$ denote a compact set on $\mathbb{R}^{d+1}$.[2] We assume that $U_{i,t-2}$ has a positive density on $\mathcal{U}$ and denote the conditional probability density function (PDF) of $U_{i,t-2}$

[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 $U_{i,t-2}$.

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

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

where $1(\cdot)$ is the usual indicator function. By the weak law of large numbers (WLLN) for IID data, $n_{t-2}/N \xrightarrow{P} p_{t-2} \equiv P\left(U_{i,t-2} \in \mathcal{U}\right)$ and $n/N \xrightarrow{P} p$ where $p \equiv \sum_{t=3}^{T} p_{t-2}$. Let $r_{t|t-2}(u) \equiv -E\left(\Delta Y_{it}|U_{i,t-2} = u\right)$. Put

$$f(u) \equiv \sum_{t=3}^{T} \frac{p_{t-2}}{p} f_{t-2}(u), \; f(\bar{u}|u) \equiv \sum_{t=3}^{T} \frac{p_{t-2}}{p} f_{t-1|t-2}(\bar{u}|u) \text{ and } r(u) \equiv \sum_{t=3}^{T} \frac{p_{t-2}}{p} r_{t|t-2}(u), \tag{2.4}$$

where we suppress the dependence of p, f and r on T. Note that both $f(\cdot)$ and $f(\cdot|\cdot)$ are mixture densities and in the stationary case $f(\cdot)$ and $f(\cdot|\cdot)$ respectively denote the marginal and transitional densities, conditional on $U_{i,t-2} \in \mathcal{U}$. We will assume that $f(\cdot)$ is uniformly bounded and bounded below from 0 on $\mathcal{U}$.

By the law of iterated expectations, (2.3) implies that

$$\begin{aligned} m(u) &= -E\left(\Delta Y_{it}|U_{i,t-2} = u\right) + E\left[m\left(U_{i,t-1}\right)|U_{i,t-2} = u\right], \\ &= r_{t|t-2}(u) + \int m(\bar{u}) f_{t-1|t-2}(\bar{u}|u)\, d\bar{u} \quad \text{for } t = 3, ..., T. \end{aligned} \tag{2.5}$$

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

$$m(u) = r(u) + \int m(\bar{u}) f(\bar{u}|u)\, d\bar{u} \tag{2.6}$$

where we have used the fact that $\sum_{t=3}^{T} p_{t-2}/p = 1$. (2.6) suggests that the parameter of interest, m, is implicitly defined as a solution to a Fredholm integral equation of the second kind in an infinite dimensional Hilbert space $\mathcal{L}_2(f)$ under certain regularity conditions:

$$m = r + \mathcal{A}m \tag{2.7}$$

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

$$\mathcal{A}m(u) = \int m(\bar{u}) f(\bar{u}|u)\, d\bar{u} \tag{2.8}$$

and $\mathcal{L}_2(f)$ is a Hilbert space with norm

$$\|m\|_2 \equiv \{< m, m >\}^{1/2} \equiv \left\{\int_{\mathcal{U}} m(u)^2 f(u)\, du\right\}^{1/2}.$$

In general, $< m_1, m_2 > \equiv \int_{\mathcal{U}} m_1(u) m_2(u) f(u)\, du$ for any $m_1,\, m_2 \in \mathcal{L}_2(f)$.

Assume that nonparametric estimators of r and $\mathcal{A}$ are given by $\hat{r}$ and $\widehat{\mathcal{A}}$. The plug-in estimator $\hat{m}$ is then given by the solution of

$$\hat{m} = \hat{r} + \widehat{\mathcal{A}}\hat{m}. \tag{2.9}$$

In this paper we consider the local polynomial estimates of r and $\mathcal{A}m$. Let $u \equiv (y, x')' \equiv (u_0, ..., u_d)'$ be a $(d+1) \times 1$ vector, where x is $d \times 1$ and y is a scalar. Let $\boldsymbol{j} \equiv (j_0, j_1, ..., j_d)$ be a $(d+1)$-vector of non-negative integers. Following Masry (1996), we adopt the notation

$$u^{\boldsymbol{j}} \equiv \boldsymbol{\Pi}_{i=0}^{d} u_i^{j_i}, \;\; \boldsymbol{j}! \equiv \boldsymbol{\Pi}_{i=1}^{d} j_i!, \;\; |\boldsymbol{j}| \equiv \sum_{i=0}^{d} j_i, \quad \sum_{0 \leq |\boldsymbol{j}| \leq q} \equiv \sum_{k=0}^{q} \sum_{\substack{j_0=0 \\ j_0+j_1+\cdots+j_d=k}}^{k} \cdots \sum_{j_d=0}^{k} .$$

From the definition of $u^{\boldsymbol{j}}$, we see that the j_i's represent powers applied to the elements of u when constructing polynomials.

We first describe the q-th order local polynomial estimators $\hat{r}$ of r and $\widehat{\mathcal{A}m}$ of $\mathcal{A}m$. Given observations $\{(Y_{i,t}, X_{it})\}$, we estimate $r(u)$ by $\hat{r}(u)$ as the minimizing constant in the following weighted least squares problem:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \sum_{t=3}^{T} \left[-\Delta Y_{it} - \sum_{0 \leq |\boldsymbol{j}| \leq q} \beta_{\boldsymbol{j}}' \left((U_{i,t-2} - u)/h \right)^{\boldsymbol{j}} \right]^2 K_h (U_{i,t-2} - u) \, 1 (U_{i,t-2} \in \mathcal{U}), \tag{2.10}$$

where $\boldsymbol{\beta}$ stacks the $\beta_{\boldsymbol{j}}$'s ($0 \leq |\boldsymbol{j}| \leq q$) in lexicographic order (with $\beta_{\mathbf{0}}$, indexed by $\mathbf{0} \equiv (0, ..., 0)$, in the first position, the element with index $(0, 0, ..., 1)$ next, etc.), $K_h(u) = h_0^{-1} k(y/h_0) \Pi_{j=1}^{d} h_j^{-1} k(x_j/h_j)$ for $u \equiv (y, x')'$, k is a univariate PDF, $h = (h_0, h_1, ..., h_d)'$ is a bandwidth sequence that shrinks to zero as $N \to \infty$. Note that in (2.10) we have used an indicator function $1(\cdot)$ to handle the non-compact support of $U_{i,t-2}$.

Let $Q_l \equiv (l+d)!/(l!d!)$ be the number of distinct $(d+1)$-tuples $\boldsymbol{j}$ with $|\boldsymbol{j}| = l$. In the above estimation problem, this denotes the number of distinct lth order partial derivatives of $r(u)$ with respect to u. Let $Q \equiv \sum_{l=0}^{q} Q_l$. Let $\mu_h(\cdot) = \mu(\cdot/h)$, where μ is a stacking function such that $\mu_h(U_{i,t-2} - u)$ denotes a $Q \times 1$ 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 $q = 1$). Then it is easy to verify that

$$\begin{aligned} \hat{r}(u) &= e_1' [S_{NT}(u)]^{-1} \frac{1}{n} \sum_{i=1}^{N} \sum_{t=3}^{T} 1_{it} K_h (U_{i,t-2} - u) \, \mu_h (U_{i,t-2} - u) \, \Delta Y_{it} \\ &= \frac{1}{n} \sum_{i=1}^{N} \sum_{t=3}^{T} \mathcal{K}_{it}(u) \, \Delta Y_{it} \end{aligned} \tag{2.11}$$

where $e_1 \equiv (1, 0, ..., 0)'$ is a $Q \times 1$ vector with 1 in the first position and zeros elsewhere, $1_{it} \equiv 1(U_{i,t-2} \in \mathcal{U})$,

$$S_{NT}(u) \equiv \frac{1}{n} \sum_{i=1}^{N} \sum_{t=3}^{T} 1_{it} K_h (U_{i,t-2} - u) \, \mu_h (U_{i,t-2} - u) \, \mu_h (U_{i,t-2} - u)', \tag{2.12}$$

and

$$\mathcal{K}_{it}(u) \equiv 1_{it} e_1' S_{NT}(u)^{-1} \mu_h (U_{i,t-2} - u) \, K_h (U_{i,t-2} - u) \tag{2.13}$$

is analogous to the "equivalent kernel" for the above local polynomial regression. $\widehat{\mathcal{A}m}(u)$ is analogously defined as the estimator of $\mathcal{A}m(u)$ with $m(U_{i,t-1})$ in place of $-\Delta Y_{it}$ in the latter's definition. Given these estimators, we can study the asymptotic properties of the resulting plug-in estimator $\hat{m}$ below.

Like Mammen, Linton and Nielsen (1999) and Mammen, Støve and Tjøstheim (2009, MST hereafter), our estimators only use observations in the smoothing if the corresponding covariates $U_{i,t-2}$ lie in a compact set $\mathcal{U}$ on $\mathbb{R}^{d+1}$. All other observations are thrown away and not used in the construction of the estimator. This device will greatly facilitate our asymptotic study by allowing $U_{i,t-2}$ to have non-compact support. If $U_{i,t-2}$ is compactly supported, then one can always chooses $\mathcal{U}$ to be its support and then the indicator function becomes redundant in the above definitions.

In terms of numerical algorithm, if $\mathcal{A}$ is well behaved in the sense to be clear later, it is well known that (2.7) implies that $m = (I - \mathcal{A})^{-1} r = \sum_{j=0}^{\infty} \mathcal{A}^j r$ where I is an identity operator. In this case, the sequence of approximations

$$m^{(l)} = \mathcal{A} m^{(l-1)} + r, \ \ l = 1, 2, 3, ...$$

converges to the truth from any starting point $m^{(0)}$. If in addition $\widehat{\mathcal{A}}$ and $\hat{r}$ are sufficiently close to $\mathcal{A}$ and r respectively, then

$$\hat{m}^{(l)} = \widehat{\mathcal{A}} \hat{m}^{(l-1)} + \hat{r}, \ \ l = 1, 2, 3, ...$$

converges to $\hat{m}$.

Note that $m(\cdot)$ is identified only upon to a location shift in (2.2). Under our model assumptions (A.1(i)-(ii)) in the next section, we have $E[m(U_{i,t-1})] = E[Y_{it}]$. This motivates us to recenter $\hat{m}^{(l)}(u)$ in each iteration to obtain

$$\hat{m}^{(l)}(u) + \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} [Y_{it} - \hat{m}^{(l)}(U_{i,t-1})],$$

which is then used in the next iteration.

2.1.1 Using the sieve estimator as an initial estimator

The above iterative procedure requires an initial estimator which may be consistent or not. If we use a consistent estimator as the initial estimator, then we expect that the procedure converges soon after a few iterations. So in practice it is desirable to start with a consistent initial estimator. Noting that (2.2) is an additive model, we propose to estimate m by the sieve method. For an excellent review on the sieve method, see Chen (2007).

To proceed, let $\{q_l(u), \ l = 1, 2, \cdots\}$ 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

$$\begin{aligned} q^L(u) &\equiv (q_1(u), q_2(u), \cdots, q_L(u))', \ q_{i,t-1} \equiv q^L(U_{i,t-1}), \ \Delta q_{i,t-1} \equiv q_{i,t-1} - q_{i,t-2}, \\ \Delta q_i &\equiv (\Delta q_{i2}, \cdots, q_{iT-1})', \text{ and } \mathbf{\Delta Q} \equiv (\Delta q_1', \Delta q_2', \cdots, \Delta q_N')'. \end{aligned}$$

Obviously we have suppressed the dependence of q_{it}, Δq_{it}, Δq_i and $\mathbf{\Delta Q}$ on L, N, or T. In particular, Δq_i and $\mathbf{\Delta Q}$ 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 (2.2) by $\beta_m' \Delta q_{i,t-1}$ for some $L \times 1$ vector β_m. This motivates us to consider the following model

$$\Delta Y_{it} = \beta_m' \Delta q_{i,t-1} + \Delta\varepsilon_{it} + R_{it},$$

where $R_{it} \equiv m(Y_{i,t-1}, X_{it}) - m(Y_{i,t-2}, X_{i,t-1}) - \beta_m' \Delta q_{i,t-1}$ signifies the approximation error. To estimate β_m in the above model, we run the regression of ΔY_{it} on $\Delta q_{i,t-1}$ by using an $\bar{L} \times 1$ valid instrument vector which we denote as Z_{it}, where $\bar{L} \geq L$. Noting that any measurable function of $\underline{U}_{i,t-2}$ can be a valid instrument, we can choose elements of Z_{it} simply as a sequence of measurable functions of $\underline{U}_{i,t-2}$. Following Anderson and Hsiao (1981), the simplest choice of Z_{it} is

$$Z_{it} = q_{i,t-2} = q^L(U_{i,t-2})$$

in which case $\bar{L} = L$. Let

$$\begin{aligned} Z_i &\equiv (Z_{i3}, \cdots, Z_{iT})', \ \mathbf{Z} \equiv (Z_1', Z_2', \cdots, Z_N')', \ \mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^- \mathbf{Z}, \\ \Delta Y_i &\equiv (\Delta Y_{i3}, \cdots, \Delta Y_{iT})', \ \mathbf{\Delta Y} \equiv (\Delta Y_1', \Delta Y_2', \cdots, \Delta Y_N')', \end{aligned}$$

where $(\cdot)^-$ denotes any symmetric generalized inverse. The two-stage least squares (2SLS) estimate of β_m is given by

$$\tilde{\beta}_m = \left(\mathbf{\Delta Q}'\mathbf{P_Z}\mathbf{\Delta Q}\right)^- \mathbf{\Delta Q}'\mathbf{P_Z}\mathbf{\Delta Y}.$$

Then we can obtain a preliminary estimate of $m(u)$ by

$$\tilde{m}_0(u) \equiv \tilde{\beta}_m' q^L(u).$$

As before, we recenter $\tilde{m}_0(u)$ to estimate $m(u)$ by

$$\tilde{m}(u) = \tilde{m}_0(u) + \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} \left[Y_{it} - \tilde{m}_0(U_{i,t-1})\right].$$

Under certain regularity conditions, we can show that $\tilde{m}(u)$ is a consistent estimator of $m(u)$ and establish its asymptotic normality. To conserve the space, we do not report the formal proofs in this paper.

2.2 Asymptotic properties of $\hat{m}(u)$

Let $Y_i = (Y_{i1}, ..., Y_{iT})'$, $X_i = (X_{i1}, ..., X_{iT})'$, $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$. We make the following assumptions on $\{Y_{it}, X_{it}, \alpha_i, \varepsilon_{it}\}$, the function of interest m, the kernel function k and the bandwidth h.

Assumptions

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

(ii) $E\left(\varepsilon_{it}|\underline{U}_{i,t-1}\right) = 0$ a.s., $E\left(\varepsilon_{it}^2|U_{i,t-2}\right) = \sigma_{t|t-2}^2(U_{i,t-2})$ a.s., and $E\left(\varepsilon_{i,t-1}^2|U_{i,t-2}\right) = \sigma_{t-1|t-2}^2(U_{i,t-2})$ a.s. Let $\sigma_{t-2}^2(\cdot) \equiv \sigma_{t|t-2}^2(\cdot) + \sigma_{t-1|t-2}^2(\cdot)$.

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

(iv) $\|m\|_2 < C$ for some $C < \infty$.

(v) $\int_{\mathcal{U}} \int [m(\bar{u}) - m(u)]^2 f(u) f(\bar{u}|u) d\bar{u}du > 0$ 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) $\sup_{u \in \mathcal{U}} \int |m(\bar{u})| f(\bar{u}|u) d\bar{u} < \infty$.

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

(ii) $m(\cdot)$ has all $(q+1)$th partial derivatives that are uniformly continuous on $\mathcal{U}$.

(iii) For $t = 3, ..., T$, $\sigma^2_{t-2}(\cdot)$ have all second order partial derivatives that 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 compact support.

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

A.1 (i) is standard in the panel data literature. For simplicity, we do no allow cross sectional dependence among $\{X_i, \alpha_i, \varepsilon_i\}$. But we allow *nonstationarity* in the time series $\{Y_{it}, X_{it}, \varepsilon_{it}, t = 1, ..., T\}$. In sharp contrast to Lee (2010), we do not need any mixing condition along the time dimension because we assume T is fixed in this paper. A.1(ii) indicates the process $\{\varepsilon_{it}, t \geq 0\}$ is a martingale difference sequence (m.d.s.) with respect to the filter generated by $\underline{U}_{i,t-1}$. It allows *conditional heteroskedasticity* of unknown form. A.1(iii) restricts that $f(\cdot)$ is well behaved on $\mathcal{U}$. A.1(iv) requires the finite second moment of $m(U_{i,t-1})$ in order for $\mathcal{L}_2(f)$ to be well defined.

A.1 (v) imposes assumptions on the functional forms of the regression function $m(\cdot)$ and the mixture densities $f(\cdot)$ and $f(\cdot|\cdot)$. If $\{U_{it}, t \geq 1\}$ is strictly stationary with PDF $f(\cdot)$ and transition density $f(\cdot|\cdot)$ and the support of f is $\mathcal{U}$, then A.1(v) requires that $0 < E[m(U_{i,t-1}) - m(U_{i,t-2})]^2/2 =< m, m > - < m, \mathcal{A}m >$. Therefore

$$< m, m > - < m, \mathcal{A}m > \text{ is positive for all } m \neq 0. \tag{2.14}$$

In other words, there is no nonzero m such that $\mathcal{A}m = m$ and hence the operator $(I - \mathcal{A})$ is one-to-one. In addition, (2.14) implies that $(I - \mathcal{A})$ has eigenvalues (in absolute values) bounded from below by a positive number $1 - \gamma$, say, for some $\gamma \in (0, 1)$. It follows that $(I - \mathcal{A})$ is invertible and $(I - \mathcal{A})^{-1}$ has eigenvalues that are bounded by $(1 - \gamma)^{-1}$. This implies that

$$\sup_{\|m\|_2 \leq 1} \left\| (I - \mathcal{A})^{-1} m \right\|_2 < \infty. \tag{2.15}$$

A.1(vi) further ensures that the operator $\mathcal{A}$ is Hilbert-Schmidt and a fortiori compact. It amounts to saying that there is no much dependence between $U_{i,t-1}$ and $U_{i,t-2}$ under the mixture transition density $f(\cdot|\cdot)$, and is tightly related to typical mixing conditions on time series. See Carrasco, Florens, and Renault (2007) for more discussions. A.1(vii) is an assumption on the operator $\mathcal{A}$, which can be easily satisfied.

A.2 specifies mainly the smooth conditions on f_{t-2}, m, and σ^2_{t-2}. A.3 mainly 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.

Let $\bar{S}_{NT}(u) \equiv E[S_{NT}(u)]$. Define

$$B_{NT}(u) \equiv \frac{1}{n} \sum_{i=1}^{N} \sum_{t=3}^{T} \bar{\mathcal{K}}_{it}(u) Dm_{it}(u) \text{ and } V_{NT}(u) \equiv \frac{-1}{n} \sum_{i=1}^{N} \sum_{t=3}^{T} \bar{\mathcal{K}}_{it}(u) \Delta\varepsilon_{it}, \tag{2.16}$$

where $Dm_{it}(u) \equiv m(U_{i,t-2}) - m(u) - \sum_{1\leq|\mathbf{j}|\leq q} \frac{1}{\mathbf{j}!} m^{(\mathbf{j})}(u)(U_{i,t-2}-u)^{\mathbf{j}}$, and

$$\bar{\mathcal{K}}_{it}(u) \equiv e_1' \left[\bar{S}_{NT}(u)\right]^{-1} 1_{it} K_h(U_{i,t-2}-u)\mu_h(U_{i,t-2}-u). \tag{2.17}$$

Note that we use the non-stochastic term $\bar{S}_{NT}(u)$ in the definition of $V_{NT}(u)$ and $B_{NT}(u)$ to facilitate the analysis in the next section. By the standard local polynomial regression theory [e.g., Masry (1996), Hansen (2008)], $\sup_{u\in\mathcal{U}}|V_{NT}(u)| = (nh!)^{-1/2}[\log(n)]^{1/2}$ and $\sup_{u\in\mathcal{U}}|B_{NT}(u)| = \|h\|^{q+1}$.

The following theorem states the uniform consistency of the estimator $\hat{m}(u)$.

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

$$\sup_{u\in\mathcal{U}}\left|\hat{m}(u) - m(u) - V_{NT}(u) - (I-\mathcal{A})^{-1}B_{NT}(u)\right| = O_P\left[n^{-1/2}(\log n)^{1/2} + v_n^2\right], \tag{2.18}$$

where $\nu_n \equiv (nh!)^{-1/2}(\log n)^{1/2} + \|h\|^{q+1}$.

Remark 1. The result in Theorem 2.1 is stronger than that in Theorem 1 of Mammen and Yu (2009) because we specify the exact probability order on the right hand side of (2.18) and $n^{-1/2}(\log n)^{1/2}+v_n^2 = o[(nh!)^{-1/2}]$ under Assumption A.4. In the above theorem $V_{NT}(u)$ and $(I-\mathcal{A})^{-1}B_{NT}(u)$ signifies the asymptotic variance and bias of $\hat{m}(u)$, respectively. For simplicity, in the proof we restrict our attention to the case where $\mathcal{U}$ is a compact set which does not expand to $\mathbb{R}^{d+1}$ as $N\to\infty$. Under some extra regularity conditions on f and m, we can follow Hansen (2008) and Li, Lu and Linton (2011) to allow $\mathcal{U}(\equiv\mathcal{U}_{NT})$ to expand to $\mathbb{R}^{d+1}$ slowly as $N\to\infty$. In this latter case, we have to adjust the uniform convergence rate accordingly.

The next theorem establishes the asymptotic normality of $\hat{m}(u)$.

Theorem 2.2 *Suppose Assumptions A.1-A.4 hold. Then for any* $u\in$*interior*$(\mathcal{U})$,

$$\sqrt{nh!}\left[\hat{m}(u) - m(u) - (I-\mathcal{A})^{-1}B_0(u)\right] \xrightarrow{D} N\left(0, \frac{\sigma^2(u)}{f(u)} e_1'\mathbb{S}^{-1}\mathbb{K}\mathbb{S}^{-1}e_1\right),$$

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

Remark 2. The asymptotic bias and variance formulae appear a little bit complicated in the above theorem. This is because we allow for general order of local polynomial regressions and distinct bandwidths for different covariates. In the special case where $q=1$, one can easily verify that

$$\mathbb{S} = \begin{pmatrix} 1 & \mathbf{0}_{1\times(d+1)} \\ \mathbf{0}_{(d+1)\times 1} & I_{d+1}\int u^2k(u)du \end{pmatrix},\ \mathbb{K} = \begin{pmatrix} \left[\int k(u)^2du\right]^{d+1} & \mathbf{0}_{1\times(d+1)} \\ \mathbf{0}_{(d+1)\times 1} & I_{d+1}\left[\int u^2k(u)^2du\right]^{d+1} \end{pmatrix},$$

the asymptotic variance reduces to $\sigma^2(u)[\int k(u)^2du]^{d+1}/f(u)$, and $B_0(u) = \frac{1}{2}\sum_{l=0}^{d}h_l^2\partial^2 m(u)/u_l^2$.

Remark 3. Theorem 2.2 indicates that the asymptotic variance of the estimator $\hat{m}(u)$ shares the same structure as that of a typical local polynomial estimator of either m in the model

$$\Delta Y_{it} = m(U_{i,t-1}) - m(U_{i,t-2}) + \Delta\varepsilon_{it} \tag{2.19}$$

by pretending the other one is known. Nevertheless, the asymptotic bias of $\hat{m}(u)$ is different from the case where one of the two m's is known in (2.19). The operator $(I-\mathcal{A})^{-1}$ signifies the accumulated bias during the iterative procedure. Noting that the error term $\Delta\varepsilon_{it}$ in (2.19) has the structure of moving averaging of order 1 (MA(1)), one may be tempted to explore such an MA(1) structure to obtain a more efficient estimate of $m(u)$, say, by following the lead of Xiao, Linton, Carroll and Mammen (2003) and Su and Ullah (2006b). Unfortunately, the latter papers require that the error process is invertible, which is not the case here. So we cannot follow either paper to obtain a more efficient estimate of $m(u)$.

Remark 4. By pretending $m(U_{i,t-1})$ is known in (2.19), we can estimate the second additive component $m(\cdot)$ in (2.19) by the local polynomial estimator of m as $\hat{m}^{(oracle)}$. It is interesting to know whether we can propose an estimator that can be as asymptotically efficient as this "oracle" estimator. The answer is yes if $U_{i,t-1}$, $t=2,...,T$, are compactly supported and $f(\cdot)$ is bounded away from zero on the union $\mathcal{U}$ of their supports. For this, we use the above kernel estimate $\hat{m}(\cdot)$ with well-chosen undersmoothing bandwidth as an estimate for the first additive component[3] and then run a local polynomial regression to obtain an estimator for the second additive component. Specifically, we propose the following procedure:

1. For $u=U_{i,t-1}$, $i=1,...,N$, and $t=2,...,T$, obtain the estimate $\hat{m}_{h^0}(u)$ as $\hat{m}(u)$ defined above by using an undersmoothing bandwidth sequence $h^0=(h_0^0,h_1^0,...,h_d^0)$ and ignoring the trimming function $1(\cdot)$.
2. Run the qth order local polynomial regression of $\hat{m}_{h^0}(U_{i,t-1})-\Delta Y_{it}$ on $U_{i,t-2}$ to obtain the estimate $\hat{m}_h^*(u)$ of $m(u)$ with the typical optimal rate of bandwidth h.

Under some suitable conditions, we can prove that the resulting estimator, $\hat{m}_h^*(u)$, say, is asymptotically equivalent to the oracle one.[4]

In the case where $U_{i,t-1}$ is not compactly supported for all $t=2,...,T$, we may not obtain uniformly consistent estimates of $m(U_{i,t-1})$ at proper rates so that we have to trim out certain observations in the second stage. The rate of convergence for $\hat{m}_h^*(u)$ will be affected proportionally by the amount of trimming. For example, let $1_{it}^*=1(U_{i,t-1}\in\mathcal{U})$ and $n^*=\sum_{i=1}^N\sum_{t=2}^T 1_{it}^*$. If we only use observations with $1_{it}^*=1$ in the second stage regression, then the rate of convergence for $\hat{m}_h^*(u)$ would become $(n^*h!)^{1/2}$ instead of $(NTh!)^{1/2}$ for the oracle estimate. To conserve space, we do not report the details here.

Remark 5. Here we propose an iterative estimation method. In fact, one can also use non-iterative method to solve a Fredholm integral equation of the second kind; see, e.g., Linton and Mammen (2005), Darolles, Fan, Florens, and Renault (2011). This involves solving a linear system of equations. Specifically, (2.3) implies

$$m(u)-E[m(Y_{i,t-1},X_{it})\mid U_{i,t-2}=u]=-E[\Delta Y_{it}\mid U_{i,t-2}=u]\quad\text{for all } u. \tag{2.20}$$

[3] In this case we can simply ignore the trimming function used in the above estimation procedure.

[4] In addition, it is also interesting to estimate the first order partial derivatives $Dm(u)$ of $m(\cdot)$ in applications, which typically requires the choice of $q\geq 2$. Note that we can obtain an estimate of $Dm(u)$ in the second step above, say, by $\hat{D}_h^*m(u)$. We can show that $\hat{D}_h^*m(u)$ is also asymptotically oracle.

Replacing the unknown conditional expectations by their local polynomial estimates yields

$$m(u) - n^{-1} \sum_{j=1}^{N} \sum_{s=3}^{T} \mathcal{K}_{js}(u)\, m(U_{j,s-1}) = -n^{-1} \sum_{j=1}^{N} \sum_{s=3}^{T} \mathcal{K}_{js}(u)\, \Delta Y_{js}. \tag{2.21}$$

Evaluating (2.21) at $u = (Y_{i,t-1}, X'_{i,t})'$ for $i = 1, ..., N$ and $t = 3, ..., T$ yields the following linear system of equations with NT_2 equations and NT_2 unknowns ($m(U_{i,t-1})$, $i = 1, ..., N$ and $t = 3, ..., T$):

$$\mathcal{M} - \mathcal{K}\mathcal{M} = -\mathcal{K}\mathcal{Y}, \tag{2.22}$$

where

$$\mathcal{M} \equiv \begin{bmatrix} m(U_{1,2}) \\ \vdots \\ m(U_{N,T-1}) \end{bmatrix}, \ \mathcal{K} \equiv n^{-1} \begin{bmatrix} \mathcal{K}_{13}(U_{1,2}) & \dots & \mathcal{K}_{NT}(U_{1,2}) \\ \vdots & \vdots & \vdots \\ \mathcal{K}_{13}(U_{N,T-1}) & \dots & \mathcal{K}_{NT}(U_{N,T-1}) \end{bmatrix}, \text{ and } \mathcal{Y} \equiv \begin{bmatrix} \Delta Y_{13} \\ \dots \\ \Delta Y_{NT} \end{bmatrix}.$$

The solution to the above linear system of equations is given by

$$\hat{\mathcal{M}} = -\left(I_{NT_2} - \mathcal{K}\right)^{-1} \mathcal{K}\mathcal{Y}.$$

For any evaluation point $u \in \mathcal{U}$, the non-iterative estimator of $m(u)$ is given by

$$\breve{m}(u) = n^{-1} \left[\mathcal{K}_{13}(u), \cdots, \mathcal{K}_{NT}(u)\right] \left(\hat{\mathcal{M}} - \mathcal{Y}\right).$$

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. Therefore, in practice, especially when the sample size is large, the iterative method may be preferred to.

3 A specification test for linear dynamic panel data models

In this section we consider the functional form specification test for dynamic panel data models. We focus on testing the correct specification of the most widely used linear dynamic panel data models versus the nonparametric dynamic panel data models considered above.

3.1 Hypotheses

To be concrete, we consider the model in (2.1). The null hypothesis of interest is

$$\mathbb{H}_0 : m(U_{i,t-1}) = \beta_0' U_{i,t-1} \text{ a.s. for some } \beta_0 \in \mathcal{B} \subset \mathbb{R}^{d+1} \tag{3.1}$$

where $i = 1, ..., N$, $t = 2, ..., T$, and $\mathcal{B}$ is compact subset of $\mathbb{R}^{d+1}$. The alternative hypothesis is

$$\mathbb{H}_1 : \Pr\left[m(U_{i,t-1}) = \beta' U_{i,t-1}\right] < 1 \ \forall\ \beta \in \mathcal{B} \text{ for some } t = 2, ..., T. \tag{3.2}$$

Recently, Lee (2011) proposes a residual-based test to check the validity of the linear dynamic models with both large N and T. Her test requires the consistent estimation of the generalized spectral derivatives which is impossible for fixed T.

In this paper we propose an alternative test for $\mathbb{H}_0$ versus $\mathbb{H}_1$ which is applicable for fixed T.[5] The proposed test is based on the comparison of the restricted estimate under $\mathbb{H}_0$ and the unrestricted estimate under $\mathbb{H}_1$, say in the spirit of Härdle and Mammen (1993). We consider the following smooth functional

$$\Gamma \equiv \int \left[m(u) - \beta_0' u\right]^2 a(u) f(u) du, \tag{3.3}$$

where $a(u)$ is a user-specified nonnegative weighting function with compact support $\mathcal{U}$. Clearly $\Gamma = 0$ under $\mathbb{H}_0$ and is generally nonzero under $\mathbb{H}_1$. Hence we can consider a test based on Γ.

The previous section gives a consistent estimate of $m(u)$ under $\mathbb{H}_1$, and β_0 can be estimated by various ways consistently under $\mathbb{H}_0$. Let $\hat{\beta}$ denote a $\sqrt{N}$-consistent estimator of β_0. Then a natural feasible test statistic could be

$$\Gamma_{NT} \equiv \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} \left[\hat{m}(U_{i,t-1}) - \hat{\beta}' U_{i,t-1}\right]^2 a(U_{i,t-1}). \tag{3.4}$$

We will show that after being appropriately centered and scaled, Γ_{NT} is asymptotically normally distributed under some suitable assumptions.

3.2 Asymptotic null distribution

Define the asymptotic bias and variance for our test statistic respectively:

$$\mathbb{B}_{NT} \equiv (h!)^{-1/2} T_1 p^{-1} \int \int e_1' \bar{S}_{NT}(u)^{-1} \mu(v) \mu(v)' \bar{S}_{NT}(u)^{-1} e_1 K(v)^2 \sigma^2(u+hv) a(u) \bar{f}(u) dvdu, \tag{3.5}$$

and

$$\begin{aligned} \sigma_0^2 \equiv\; & 2T_1^2 p^{-2} \int \int \int \int e_1' \bar{S}_{NT}(v)^{-1} \mu(u) \mu(u+w)' \bar{S}_{NT}(v)^{-1} e_1 e_1' \bar{S}_{NT}(v)^{-1} \mu(\tilde{u}) \mu(\tilde{u}+w)' \\ & \times \bar{S}_{NT}(v)^{-1} e_1 K(u) K(\tilde{u}) K(u+w) K(\tilde{u}+w) \sigma^4(v) a(v)^2 \bar{f}(v)^2 d\tilde{u}dudvdw \Big], \end{aligned} \tag{3.6}$$

where $\bar{f}(u) \equiv T_1^{-1} \sum_{t=2}^{T} f_{t-1}^{(uc)}(u)$ and $f_{t-1}^{(uc)}(\cdot)$ is the unconditional PDF of $U_{i,t-1}$ (without conditioning on that $U_{i,t-2}$ lies in $\mathcal{U}$).

If $d < 3$, it suffices to base our test on the local linear regression. In this case, we can use the simplified version of $\mathbb{B}_{NT}$ and σ_0^2 as

$$\mathbb{B}_{NT}^{(ll)} \equiv (h!)^{-1/2} T_1 p^{-1} C_1^{d+1} \int \sigma^2(u) a(u) \bar{f}(u) f(u)^{-2} du, \tag{3.7}$$

and

$$\sigma_0^{2(ll)} \equiv 2T_1^2 p^{-2} C_2^{d+1} \int \sigma^4(u) a(u)^2 \bar{f}(u)^2 f(u)^{-4} du, \tag{3.8}$$

where $C_1 \equiv \int_{\mathbb{R}} k(z)^2 dz$, and $C_2 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} k(z) k(z+\bar{z}) dz\right)^2 d\bar{z}$. For any k, we can calculate C_1 and C_2 explicitly. If we use the Gaussian kernel,[6] i.e., $k(z) = \left(1/\sqrt{2\pi}\right) e^{-z^2/2}$, then $C_1 = 1/(2\sqrt{\pi})$ and

[5] We conjecture that our asymptotic theory can also be extended to the case of large T.

[6] While the Gaussian kernel does not have compact support, it can be approximated arbitrarily well by compactly supported kernels. See Ahn (1997, p.13).

$C_2 = 1/(2\sqrt{2\pi})$. If we use the Epanechnikov kernel instead, i.e., $k(z) = 0.75(1 - z^2)1(|z| \leq 1)$, then $C_1 = 0.6$, and $C_2 = 0.4338$. One can readily show that $\mathbb{B}_{NT} = \mathbb{B}_{NT}^{(ll)} + o_P(1)$ and $\sigma_0^2 = \sigma_0^{2(ll)}$ if $d < 3$.

We add the following assumptions.

A.5 (i) $\sqrt{N}(\hat{\beta} - \beta_0) = O_P(1)$ under $\mathbb{H}_0$.

(ii) $\max_{2 \leq t \leq T} E[\|U_{i,t-1}\|^2] < \infty$.

(iii) The weight function $a(\cdot)$ is a nonnegative function that is uniformly continuous and bounded on its compact support $\mathcal{U}$.

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

A.5(i) is weak and can be met for various estimates of β_0 in correctly specified linear dynamic panel data models. A.5(ii) specifies the weak conditions on $a(\cdot)$. The simple indicator function $1\{\cdot \in \mathcal{U}\}$ suffices. A.6 requires the bandwidth be undersmoothing in comparison with the optimal rate of bandwidth in estimating m.

Theorem 3.1 *Under Assumptions A.1-A.3 and A.5-A.6,* $NT_1(h!)^{1/2}\Gamma_{NT} - \mathbb{B}_{NT} \xrightarrow{D} N(0, \sigma_0^2)$ *under* $\mathbb{H}_0$.

Remark 6. The proof of the above theorem is quite involved. Because we do not have a closed form estimate for $m(\cdot)$ under the alternative, we can only rely on the consistent estimate $\hat{m}$ studied previously. By the stochastic expression reported in Theorem 2.1, we can demonstrate that $\hat{m}(u) - m(u) - V_{NT}(u) - (I - \mathcal{A})^{-1} B_{NT}(u)$ has asymptotic negligible effect on the asymptotic distribution of our test statistic. In addition, $B_{NT}(u) = 0$ under $\mathbb{H}_0$, which implies that $V_{NT}(u)$ alone contributes to both the asymptotic bias and variance of our test statistic. Then we can write the leading term of our test statistic as a well-behaved third order V-statistic. This V-statistic can be further decomposed as a second U-statistic, plus a bias term ($\mathbb{B}_{NT}$) and some asymptotically negligible terms. See the proof of Theorem 3.1 in the appendix.

Remark 7. To implement, we need consistent estimates of the asymptotic bias and variance. To achieve this goal, we propose to estimate the error term $\Delta\varepsilon_{it}$ in the first-differenced model consistently. There is a temptation to use the nonparametric residuals

$$\tilde{\Delta}\varepsilon_{it} = \Delta Y_{it} - \hat{m}(U_{i,t-1}) + \hat{m}(U_{i,t-2}).$$

Unfortunately, this estimate does not serve our purpose unless $\{U_{i,t-1}\}$ has a compact support. For the infinite support case, our device to use observations with $U_{i,t-2}$ lying on $\mathcal{U}$ can ensure the consistency of $\hat{m}(U_{i,t-2})$ with $m(U_{i,t-2})$ at a proper rate but not that of $\hat{m}(U_{i,t-1})$ with $m(U_{i,t-1})$ whenever $U_{i,t-1}$ lies in the tail of the distributions. Below we propose consistent estimates for $\mathbb{B}_{NT}^{(ll)}$ and $\sigma_0^{2(ll)}$ (or $\mathbb{B}_{NT}$ and σ_0^2), based on the parametric residuals

$$\hat{\Delta}\varepsilon_{it} \equiv \Delta Y_{it} - \hat{\beta}' U_{i,t-1} + \hat{\beta}' U_{i,t-2}.$$

Note that $\hat{\Delta}\varepsilon_{it} - \Delta\varepsilon_{it}$ is $O_P(N^{-1/2})$ under $\mathbb{H}_0$ by Assumption A.5(i), and is $O_P(\lambda_{NT})$ under the local alternative $\mathbb{H}_1(\lambda_{NT})$ defined in (3.9) below.

Define[7]

$$\hat{\sigma}^2(u) \equiv \frac{1}{n}\sum_{j=1}^N\sum_{s=3}^T 1_{js}L_h(U_{j,s-2}-u)\left(\hat{\Delta}\varepsilon_{js}\right)^2, \ \hat{f}(u) \equiv \frac{1}{n}\sum_{j=1}^N\sum_{s=3}^T 1_{js}L_h(U_{j,s-2}-u),$$

$$\hat{\bar{f}}(u) \equiv \frac{1}{NT_1}\sum_{j=1}^N\sum_{s=2}^T L_h(U_{j,s-1}-u),$$

where $L_h(u) = h_0^{-1}l(y/h_0)\Pi_{j=1}^d h_j^{-1}l(x_j/h_j)$ for $u \equiv (y,x')'$, and l is a univariate kernel function with compact support in $\mathbb{R}$. The condition on l is specified in the following assumption.

A.7 Let $\gamma \geq 2$. The kernel function $l : \mathbb{R} \to \mathbb{R}$ is symmetric, continuous and compactly supported such that $l(\cdot)$ is a γth order kernel: $\int s^j l(s)\,du = \delta_{j0}$ for $j = 1, \ldots, \gamma-1$, $\int s^\gamma l(s)\,ds = \kappa_\gamma < \infty$, where δ_{ij} is the Kronecker's delta. The γth order derivatives of $f_{t-2}(\cdot)$, $t = 3, ..., T$, exist and are continuous.

When $d < 3$ and local linear regressions are applied, we propose to estimate the asymptotic bias $\mathbb{B}_{NT}^{(ll)}$ by

$$\hat{\mathbb{B}}_{NT}^{(ll)} \equiv (h!)^{-1/2}T_1Nn^{-1}C_1^{d+1}\frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \hat{\sigma}^2(U_{i,t-1})\,a(U_{i,t-1})\,\hat{f}(U_{i,t-1})^{-2}$$

and the asymptotic variance $\sigma_0^{2(ll)}$ by

$$\hat{\sigma}_{NT}^{2(ll)} \equiv 2T_1^2N^2n^{-2}C_2^{d+1}\frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \left[\hat{\sigma}^2(U_{i,t-1})\right]^2 a(U_{i,t-1})^2\,\hat{\bar{f}}(U_{i,t-1})\,\hat{f}(U_{i,t-1})^{-4}.$$

For the general case when $d \geq 3$ or higher order local polynomial is used, we propose to apply the residuals $\hat{\Delta}\varepsilon_{it}$ to estimate $\mathbb{B}_{NT}$ in (3.5) and σ_0^2 in (3.6) directly as in Hoderlein, Su and White (2011). Let $\hat{W}_{i,t-2} \equiv (U'_{i,t-2}, \hat{\Delta}\varepsilon_{it})'$, and $\hat{\zeta}(\hat{W}_{i,t-1}, \hat{W}_{j,s-2}) \equiv \mathcal{K}_{js}(U_{i,t-1})\,\hat{\Delta}\varepsilon_{js}$. Then we can estimate the asymptotic bias, $\mathbb{B}_{NT}$, and variance, σ_0^2, by

$$\hat{\mathbb{B}}_{NT} = (h!)^{1/2}n^{-2}\sum_{i=1}^N\sum_{t=2}^T\sum_{j=1}^N\sum_{s=3}^T\left[\hat{\zeta}\left(\hat{W}_{i,t-1},\hat{W}_{j,s-2}\right)\right]^2 a(U_{i,t-1}), \text{ and}$$

$$\hat{\sigma}_{NT}^2 = 2h!T_1^2N^2n^{-4}\sum_{i=1}^N\sum_{t=3}^T\sum_{j=1}^N\sum_{s=3}^T\left[\frac{1}{NT_1}\sum_{l=1}^N\sum_{r=2}^T\hat{\zeta}\left(\hat{W}_{l,r-1},\hat{W}_{i,t-2}\right)\hat{\zeta}\left(\hat{W}_{l,r-1},\hat{W}_{j,s-2}\right)a(U_{l,r-1})\right]^2.$$

Clearly, the computation now becomes quite involved.

It is tedious to show the consistency of either type of estimates under either $\mathbb{H}_0$ or $\mathbb{H}_1(\lambda_{NT})$ with $\lambda_{NT} = (NT_1)^{-1/2}(h!)^{-1/4}$ [e.g., $\hat{\mathbb{B}}_{NT} - \mathbb{B}_{NT} = o_P(1)$ and $\hat{\sigma}_{NT}^2 - \sigma_0^2 = o_P(1)$]. Then the following feasible test statistics

$$J_{NT} \equiv \left[NT_1(h!)^{1/2}\Gamma_{NT} - \hat{\mathbb{B}}_{NT}\right]/\sqrt{\hat{\sigma}_{NT}^2}$$

and

$$J_{NT}^{(ll)} \equiv \left[NT_1(h!)^{1/2}\Gamma_{NT} - \hat{\mathbb{B}}_{NT}^{(ll)}\right]/\sqrt{\hat{\sigma}_{NT}^{2(ll)}}$$

are asymptotically distributed as $N(0,1)$ under suitable conditions. The result is summarized in the following corollary.

[7] Note that $\hat{\sigma}^2(u)$ estimates $\sigma^2(u) \equiv \sum_{s=3}^T(p_{s-2}/p)\,\sigma_{s-2}^2(u)\,f_{s-2}(u)$.

Corollary 3.2 *Under Assumptions A.1-A.3 and A.5-A.7,* $J_{NT} \stackrel{D}{\to} N(0,1)$ *under* $\mathbb{H}_0$*. If* $d < 3$ *and* $q = 1$*,* $J_{NT}^{(ll)} \stackrel{D}{\to} N(0,1)$ *under* $\mathbb{H}_0$*.*

Noting that the J_{NT} or $J_{NT}^{(ll)}$ test is one-sided, we reject the null for large values of J_{NT} or $J_{NT}^{(ll)}$.

3.3 Local power property and consistency

To derive the asymptotic power function of J_{NT} or $J_{NT}^{(ll)}$ under a sequence of Pitman local alternatives, we need to consider the multi-array process[8]

$$\{(X_{it}^{[NT]\prime}, Y_{it}^{[NT]})', \ i = 1, ..., N, \ t = 1, ..., T, \ N = 1, 2, ..., \ T = 1, 2, ...\}.$$

Let $f_{t-2}^{[NT]}(\cdot)$ denote the PDF of $U_{i,t-2}^{[NT]} \equiv (Y_{i,t-2}^{[NT]\prime}, X_{i,t-1}^{[NT]\prime})'$ conditional on that $U_{i,t-2}^{[NT]}$ lies in $\mathcal{U}$. Let $f_{t-1|t-2}^{[NT]}(\cdot|\cdot)$ denote the conditional PDF of $U_{i,t-1}^{[NT]}$ given $U_{i,t-2}^{[NT]}$ conditional on that $U_{i,t-2}^{[NT]}$ lies in $\mathcal{U}$. Define $f^{[NT]}(\cdot)$ and $f^{[NT]}(\cdot|\cdot)$ analogously as $f(\cdot)$ and $f(\cdot|\cdot)$ in (2.4). We consider the following sequence of Pitman local alternatives

$$\mathbb{H}_1(\lambda_{NT}) : m(U_{i,t-1}^{[NT]}) = \beta_0' U_{i,t-1}^{[NT]} + \lambda_{NT}\delta_{NT}(U_{i,t-1}^{[NT]}) \text{ a.s.} \tag{3.9}$$

where $\lambda_{NT} \to 0$ as $N \to \infty$ and $\delta_{NT}(\cdot)$ is a measurable function such that $\mu_0 \equiv \lim_{N\to\infty} \int \delta_{NT}(u)^2 a(u) f^{[NT]}(u)\,du$ exists and is finite. Let $\underline{U}_{i,t-1}^{[NT]} \equiv (U_{i,t-1}^{[NT]\prime}, U_{i,t-2}^{[NT]\prime}, ..., U_{i,1}^{[NT]\prime})'$.

The following theorem studies the asymptotic local power property of J_{NT} and $J_{NT}^{(ll)}$ under $\mathbb{H}_1(\lambda_{NT})$.

Theorem 3.3 *Let Assumptions A.1-A.2 hold for the process* $(X_{it}^{[NT]}, Y_{it}^{[NT]}, \alpha_i, \varepsilon_{it})$ *with obvious modifications, e.g., with* $f^{[NT]}(\cdot)$ $f^{[NT]}(\cdot|\cdot)$*, and* $\underline{U}_{i,t-1}^{[NT]}$*, replacing* $f(\cdot)$*,* $F(\cdot|\cdot)$*, and* $\underline{U}_{i,t-1}$*, respectively. Let Assumptions A.3 and A.5-A.6 hold. Suppose that* $\lambda_{NT} = (NT_1)^{-1/2}(h!)^{-1/4}$ *in* $\mathbb{H}_1(\lambda_{NT})$*. Then* $\Pr(J_{NT} \geq z|\mathbb{H}_1(\lambda_{NT})) \to 1 - \Phi(z - \mu_0/\sigma_0)$ *and the same result also holds for* $J_{NT}^{(ll)}$ *if* $d < 3$ *and* $q = 1$*.*

The above theorem says that our test statistic J_{NT} or $J_{NT}^{(ll)}$ has nontrivial power against $\mathbb{H}_1(\lambda_{NT})$ with $\lambda_{NT} = (NT_1)^{-1/2}(h!)^{-1/4}$ whenever $\mu_0 > 0$. The rate $\lambda_{NT} = (NT_1)^{-1/2}(h!)^{-1/4}$ is slower than the parametric rate $(NT_1)^{-1/2}$ as $h! \to 0$.

The following theorem shows that the test is consistent.

Theorem 3.4 *Suppose Assumptions A.1-A.7 hold. Suppose that* $\mu_A \equiv T_1^{-1}\sum_{t=2}^{T} E\{[m(U_{i,t-1}) - \beta_0' U_{i,t-1}]^2 a(U_{i,t-1})\} > 0$*. Then* $P(J_{NT} > \alpha_{NT}) \to 1$ *as* $N \to \infty$ *for any nonstochastic sequence* $\lambda_{NT} = o[NT_1(h!)^{1/2}]$*.*

3.4 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. Therefore it is worthwhile to propose a bootstrap procedure to improve the finite sample performance of our test. Below we propose a recursive bootstrap procedure to obtain the bootstrap p-values for our test. The procedure goes as follows:

[8] It is also fine to allow ε_{it} and α_i to be $[NT]$-dependent.

1. Estimate the restricted model under $\mathbb{H}_0$ and obtain the residuals $\hat{\varepsilon}_{it} = Y_{it} - \hat{\rho} Y_{i,t-1} - \hat{\beta}'_{-1} X_{it}$, where $\hat{\beta} = (\hat{\rho},\, \hat{\beta}'_{-1})'$ is any $\sqrt{N}$-consistent IV or GMM estimate of β. Calculate the test statistic J_{NT} based on $\{Y_{it}, X_{it}\}$. Let $\hat{\alpha}_i \equiv \bar{\hat{\varepsilon}}_i \equiv T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_{it}$.

2. Obtain the bootstrap error $\varepsilon^*_{it} = \left(\hat{\varepsilon}_{it} - \bar{\hat{\varepsilon}}_i\right) \eta_{it}$ for $i = 1, 2, \ldots, N$ and $t = 2, \ldots, T$, where η_{it}'s are IID across both i and t and follow a two point distribution: $\eta_{it} = (1 - \sqrt{5})/2$ with probability $(1 + \sqrt{5})/2\sqrt{5}$ and $\eta_{it} = (\sqrt{5} + 1)/2$ with probability $(\sqrt{5} - 1)/2\sqrt{5}$. Generate the bootstrap analogue Y^*_{it} of Y_{it} as

$$Y^*_{it} = \hat{\rho} Y^*_{i,t-1} + \hat{\beta}'_{-1} X_{it} + \hat{\alpha}_i + \varepsilon^*_{it} \text{ for } i = 1, 2, \ldots, N \text{ and } t = 2, \ldots, T,$$

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

3. Given the bootstrap resample $\{Y^*_{it},\, X_{it}\}$, estimate both the restricted (linear) and unrestricted (nonparametric) first-differenced model and calculate the bootstrap test statistic J^*_{NT} or $J^{(ll)*}_{NT}$ analogously to J_{NT} or $J^{(ll)}_{NT}$.

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 $p^* \equiv B^{-1} \sum_{l=1}^{B} 1\left(J^*_{NT,l} > J_{NT}\right)$.

Remark 8. We make a few remarks on the above bootstrap procedure. First, we impose the null hypothesis of linear dynamic panel data models in step 2. Second, in view of the fact that T is fixed, the process $\{Y^*_{it},\, t = 1, 2, ..., T\}$ cannot have the identical marginal distribution (conditional or unconditional on the data) for any $i = 1, ..., N$. Fortunately, the asymptotic theories we developed so far allows nonstationarity along the time dimension. Third, 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. In particular, we conjecture that one can show that $J^*_{NT} \xrightarrow{D} N(0, 1)$ conditionally on the observed sample. To conserve the space, we omit the details.

Remark 9. We can also construct the specification test using the non-iterative estimators by replacing the iterative estimator $\hat{m}$ with $\check{m}$ in Remark 5.

4 Simulations

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

4.1 Data generating processes

We consider the following six data generating processes (DGPs):

DGP 1: $Y_{it} = 0.25 Y_{i,t-1} + \alpha_i + \varepsilon_{it}$;

DGP 2: $Y_{it} = 0.25Y_{i,t-1} - 0.75X_{it} + \alpha_i + \varepsilon_{it}$;

DGP 3: $Y_{it} = \cos(Y_{i,t-1}) + \alpha_i + \varepsilon_{it}$;

DGP 4: $Y_{it} = 2\Phi(Y_{i,t-1} - Y_{i,t-1}^2) + \alpha_i + \varepsilon_{it}$;

DGP 5: $Y_{it} = 2\cos(Y_{i,t-1}) + \exp(X_{it}) + \alpha_i + \varepsilon_{it}$;

DGP 6: $Y_{it} = 2\Phi(Y_{i,t-1} - Y_{i,t-1}^2)\,[1 + \Phi(X_{it})] + \alpha_i + \varepsilon_{it}$;

where $\Phi(\cdot)$ is the standard normal CDF, α_i are IID $U(-1/2, 1/2)$, ε_{it} are IID $N(0,1)$ across both i and t, $X_{it} = 0.5\alpha_i + \eta_{it}$, η_{it} are IID $U(-1,1)$ across both i and t, $\{\alpha_i\}$, $\{\varepsilon_{it}\}$, and $\{\eta_{it}\}$ are mutually independent.

Apparently, DGPs 1 and 2 are linear models with and without exogenous regressors, respectively. DGP 3 and 4 are nonlinear dynamic panel data models without exogenous regressors, and DGPs 5 and 6 are nonlinear dynamic panel data model with one exogenous regressor X_{it}. The lagged dependent variable $Y_{i,t-1}$ and the exogenous regressor X_{it} enter DGP 5 additively and DGP 6 multiplicatively through some nonlinear transformations. Using the notation in Sections 2-3, $m(\cdot)$ is defined as follows:

DGP 1: $m(y) = 0.25y$;

DGP 2: $m(y,x) = 0.25y - 0.75x$;

DGP 3: $m(y) = \cos(y)$;

DGP 4: $m(y) = 2\Phi\left(y - y^2\right)$;

DGP 5: $m(y,x) = 2\cos(y) + \exp(x)$;

DGP 6: $m(y,x) = 2\Phi\left(y - y^2\right)[1 + \Phi(x)]$.

We assume that these functional forms are completely unknown. Our purpose here is to estimate m and test for the linearity of m.

4.2 Implementation

To implement our estimation and testing procedures, we need to obtain the initial sieve estimator. We choose Hermite polynomials as the sieve base (see Blundell, Chen, and Kristensen, 2007). For DGPs 1, 3 and 4, we have only one endogenous regressor $Y_{i,t-1}$ in the unknown function $m(\cdot)$ so that we can approximate the one-dimensional function $m(y)$ by the Hermite polynomials:

$$q^{L_0}(y) \equiv \left[1, \left(y - \bar{Y}\right), \left(y - \bar{Y}\right)^2, ..., \left(y - \bar{Y}\right)^{L_0 - 1}\right]' \exp\left(-(y - \bar{Y})^2 / \left[2S_Y^2\right]\right) \tag{4.1}$$

where $\bar{Y}$ and S_Y are the sample mean and standard deviation of $\{Y_{i,t-2}\}$, respectively. In DGPs 2, 5 and 6, we have two terms in the unknown function $m(y,x)$. Other than $q^{L_0}(y)$ and $q^{L_0}(x)$,[9] we also use their cross product terms to approximate $m(y,x)$. We simply choose $L_0 = \lfloor (NT_2)^{1/4} \rfloor + 1$, where $\lfloor a \rfloor$ denotes the integer part of a. In this case, the total number of approximating terms in the sieve base is given by L_0 in DGPs 1, 3 and 4 and $L_0(2 + L_0)$ in DGPs 2, 5 and 6.

For the estimation and testing, we need to choose both the kernel function and the bandwidth sequence. We use the Epanechnikov kernel $k(z) = 0.75\left(1 - z^2\right)1(|z| \le 1)$, and choose the bandwidth

[9] $q^{L_0}(x) \equiv \left[1, \left(x - \bar{X}\right), \left(x - \bar{X}\right)^2, ..., \left(x - \bar{X}\right)^{L_0 - 1}\right]' \exp\left(-(x - \bar{X})^2 / \left[2S_X^2\right]\right)$, where $\bar{X}$ and S_X are the sample mean and standard deviation of $\{X_{i,t-1}\}$, respectively.

by the Silverman's "rule of thumb": $h = 2.35 S_U (NT_2)^{-1/(5+d)}$ where $S_U = S_Y$ and $d = 0$ in DGPs 1, 3 and 4, and $S_U = (S_Y, S_X)$ and $d = 1$ in DGPs 2, 5 and 6. This bandwidth is usually not optimal, especially for dependent data. There may exist some other bandwidth that improves the final estimates. But we leave the development of a data-driven rule for the selection of "optimal" bandwidth for the proposed algorithm for future research.

The convergence criterion we use for the estimation is as follows: stop the iteration procedure if

$$\frac{\sum_{j=1}^{J} \left[m^{(l+1)}(u_j) - m^{(l)}(u_j) \right]^2}{\sum_{j=1}^{J} \left[m^{(l)}(u_j) \right]^2 + 0.0001} < 0.001,$$

where u_j, $j = 1, ..., J$, are the J evaluation points. In practice, researchers can choose the evaluation points they are interested in. Here we let the number of evaluation points be 50 for DGPs 1, 3 and 4 and 225 for DGPs 2, 5 and 6. 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. The similar convergence criterion is used in, e.g., Nielsen and Sperlich (2005), Henderson, Carroll and Li (2008), and MST (2009). For the specification test, we let the data points be the evaluation points.

For the (N, T) pair, we consider $N = 50$, 100, 200, and $T = 4$, 6. For each scenario, the number of replications is 1000 for the estimation and 250 for the test. The number of bootstrap resamples for the test is 200. 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 $U_{i,t-2}$.

Further, for both estimation and testing, we use both iterative and non-iterative methods; see Remarks 5 and 9.

4.3 Estimation results

Table 1 reports the estimation results for $T = 4$. For all the DGPs, the median or average RMSEs of both iterative and non-iterative estimators are smaller than those of the initial estimators. Relative to the initial estimators, for most DGPs the RMSEs can be reduced by 30-40% using either iterative or non-iterative estimators. Comparing the performances of iterative estimators and non-iterative estimators, iterative estimators are slightly better. For the iterative estimators, Table 3 presents the median of number of iterations. For all the DGPs, the number of iteration is quite small; after 3-4 iterations, the estimates converge. This means that the iterations do not require much computation time. It also suggests that our initial estimates are well chosen. Figure 1 illustrates the estimation results for DGP 1 when $N = 50$ and 100. Figure 2 shows the estimation results for DGP 3.[10]

Table 2 presents the estimation results for $T = 6$. Again, for all the DGPs, the median or average RMSEs of initial estimators can be reduced by up to 30-40%. The performances of iterative estimators or non-iterative estimators are similar. The median number of iteration for the iterative estimators are quite modest: around 2 to 3 as shown in Table 3.

[10] Figures for other DGPs are available upon request.

4.4 Specification test results

We examine the empirical level and power of our test statistics. Table 4 shows the empirical rejection frequencies at the three conventional nominal levels (1%, 5% and 10%) for $N = 50$, 100, and 200 and $T = 4$. We use DGPs 1 and 2 to examine the level behavior of the test. The levels behave reasonably well for both DGPs. For DGP 1, the levels for both iterative and non-iterative methods are slightly under-sized. For DGP 2, the rejection frequencies for the iterative method are similar to the nominal levels and those for non-iterative methods are slightly smaller than the nominal levels. We use DGPs 3-6 to examine the empirical power of our test. The powers of both iterative and non-iterative methods are good. They increase rapidly as N increases; when N increases to 200, the powers are about 90% or higher even for the 1% test.

Table 5 presents the rejection frequencies for $N = 50$, 100 and $T = 6$. Again, for DGPs 1 and 2, the rejection frequencies are close to the nominal levels. For DGP 3-6, the power of tests increase with both N and T. When $N = 100$ and $T = 6$, the powers are above 90% for most cases.

5 Empirical applications

5.1 Economic growth, initial economic condition, and capital accumulation

In this subsection, we apply our new nonparametric dynamic panel data models to study the important question of economic growth. Specifically, we examine two questions in details. First, what is the relation between a country's economic growth and its initial level of income? Second, what is the relation between a country's economic growth and its capital accumulation? For the first question, Solow's (1956) growth model predicts that economic growth rates are negatively associated with the initial income levels. The endogenous growth models (e.g., Romer, 1986 and Lucas, 1988) argue that the differences in initial income levels are transitory and do not affect the long-run economic growth. Barro (1991) examines the question empirically using a cross-section of countries in the period 1960-1985 and finds that the growth rate of real GDP per capita is negatively related to the initial (1960) level of real GDP per capita. For the second question, different models predict different relationships between economic growth and physical capital investment. Solow's (1956) growth model shows that there is no association between economic growth and investment in the steady state. Endogenous growth models (e.g., Romer, 1986), predict a positive association between economic growth and investment. In a neoclassic growth model, Carroll and Weil (1994) show that "exogenous increase in growth makes subsequent saving fall." Thus, from the theoretical point of view, the relation between economic growth and investment is not conclusive. There are also many empirical studies on the relationship between economic growth and capital accumulation (e.g., Bond, Leblebicioglu, and Schiantarelli, 2010, BLS). The empirical evidence is also mixed. Some suggest little or no association between investment and economic growth rate (e.g., Jones, 1995); others show positive relationship (e.g., BLS). Most of the empirical studies on these two questions use linear models. Nevertheless, as Durlauf (2000) puts it, for many growth theories, "the linear regression is a misspecification of the growth process." In this subsection, we investigate these two questions using our nonparametric models that allow general nonlinearity of unknown form.

We use a panel data of 71 countries over 41 years (1960-2000). The dataset is same as in BLS.[11] We first examine the ten year growth rate. Let Y_{it} denote the economic growth rate of country i over the tth decade. For example, Y_{i1}=ln $(GDP_{i,1970}) - \ln(GDP_{i,1960})$, where $GDP_{i,s}$ is the real GDP per worker for country i in year s as in BLS. We include two more regressors other than the lagged Y_{it}: a country's initial economic condition ($X_{1,i}$) and its investment share ($X_{2,it}$). Specifically, $X_{1,i}$ is the logarithm of country i's real GDP per worker in 1960, which represents its initial economic condition. $X_{2,it}$ is the logarithm of the average share of physical investment of country i over its GDP over the tth decade. So, here $N = 71$ and $T = 4$. We consider the nonparametric fixed effects model:

$$Y_{it} = m\left(Y_{i,t-1}, X_{1,i}, X_{2,it}\right) + \alpha_i + \varepsilon_{it},\ i = 1, ..., 71,\ t = 1, ..., 4. \tag{5.1}$$

We first present the estimation results in Figure 3. Figure 3(a) presents the relation between a country's ten year economic growth rate and its lagged ten year growth rate. Specifically, it shows the estimate of $m\left(\cdot, \bar{x}_1, \bar{x}_2\right)$, where $\bar{x}_1$ and $\bar{x}_2$ are fixed at the medians of $X_{1,i}$ and $X_{2,it}$, respectively. It is clear that this figure suggests a nonlinear relationship between the current growth rate of real DGP per worker and its lagged value. When the lagged growth rate is in the relatively low range (-0.1 to 0.2), the relation between the current growth rate and the lagged growth rate is positive; when the lagged growth rate is in the relatively high range (0.2 to 0.5), the relation between them becomes negative.

In Figure 3(b), we present the relation between a country's ten year economic growth and its initial GDP per worker; i.e., we plot the estimates of $m\left(\bar{y}, \cdot, \bar{x}_2\right)$, where $\bar{y}$ and $\bar{x}_2$ are the medians of $Y_{i,t-1}$ and $X_{2,it}$, respectively. Again, we observe substantial nonlinearity. When initial income levels are high, as found in much of the early literature they tend to be negatively associated with economic growth rates. Nevertheless, when initial income levels are low, they tend to be positively associated with economic growth rates. This suggests that both very poor countries and very rich countries tend to have low economic growth rates and countries with medium initial income levels may enjoy fast economic growth.

Figure 3(c) shows the relation between a country's economic growth and its investment share. We observe a positive relation between them. The figure appears linear, though it becomes flatter when the investment share is large.

It is apparent that there is substantial nonlinear relationship among the three variables. Our formal specification tests of linearity soundly reject the null. When we use 200 bootstrap resamples, the bootstrap p-values of the tests based on iterative and non-iterative estimates are 0 and 0.01, respectively.

As a robust check, we perform the same analysis using five year growth rates over the same period (1960-2000). That is, we let Y_{it} denote the growth rate of real GDP per worker for country i over the tth five-year period; for example, $Y_{i1} = \ln\left(GDP_{i,1965}\right) - \ln\left(GDP_{i,1960}\right)$. $X_{1,i}$ again is the logarithm of real GDP per worker in 1960. $X_{2,it}$ is the logarithm of average investment share for country i over

[11] There are several differences between BLS and our study. For example, BLS study the annual economic growth rates; we study the long-run (5 and 10 year) growth rates. BLS use a linear fixed effect model; we allow nonlinearity. We only include one lag of Y_{it}, and X_{it} as explantory variables; BLS use multiple lags of Y_{it} and X_{it}.

the tth five-year period. So, here we have $N = 71$ and $T = 8$. We find very similar patterns for the relations among the three economic variables. The estimation results are presented in Figures 3(d)-(f). Our specification tests also reject the null of linearity at the 5% significance level. Table 6 reports all the specification test results.

In summary, we find the nonlinear relation between a country's economic growth rate and its lagged value. We also find that the relation between a country's economic growth rate and its initial economic condition is nonlinear. This study shows that using linear models to characterize the relationship among these economic variables may overlook the important nonlinearity.

5.2 Firm labor inputs and sales

In this subsection, we illustrate our methods by studying the relation between a firm's sales and its labor inputs. Let $Y_{it} = \ln(S_{it}/K_{it})$ and $X_{it} = \ln(L_{it}/K_{it})$, where S_{it}, K_{it}, and L_{it} are firm i's sales, capital inputs, and labor inputs in year t, respectively. Again, we consider the nonparametric dynamic panel data model:

$$Y_{it} = m(Y_{i,t-1}, X_{it}) + \alpha_i + \varepsilon_{it}.$$

The lagged sales $Y_{i,t-1}$ can potentially affect the current sales Y_{it}, for example, through the change of inventories. Labor inputs X_{it} also affect Y_{it} through the change of production. We expect positive relationship between Y_{it} and X_{it}. However, without looking into the data, we have no good reason to believe that the relation between them is linear.

We use the same dataset as in Bond (2002) to investigate the question.[12] The number of firms is $N = 509$ and the number of time periods is $T = 8$. Figure 4 presents the estimation results. In Figures 4(a)-(c), we set $\bar{x}$ to be fixed at 0.25, 0.5 and 0.75 quantiles of X_{it}, respectively, and plot the estimates of $m(\cdot, \bar{x})$. In all the three figures, $Y_{i,t-1}$ is positively associated with Y_{it} and the relation appears linear. In Figures 4(d)-(f), we set $\bar{y}$ to be the 0.25, 0.5 and 0.75 quantiles of Y_{it}, respectively, and draw the estimates of $m(\bar{y}, \cdot)$. Without any surprise, X_{it} is positively associated with Y_{it}, as high labor inputs lead to high sales. However, it is interesting to observe that the relation between labor inputs and sales appears linear in both Figures 4(e) and (f). In Figure 4(d) where $\bar{y}$ is the 0.25 quantile of Y_{it}, we observe some nonlinearity: when X_{it} increases to the value around -3.5, Y_{it} declines with X_{it}. Nevertheless, our formal specification tests do not reject the linearity of $m(\cdot, \cdot)$. The p-values of our specification tests based on the iterative and non-iterative estimates are 0.35 and 0.41, respectively. In conclusion, our study validates the use of linear model for this context.

[12] There are some differences between our study and Bond (2002). Bond (2002) uses a linear static model with correlated errors (ε_{it}). He uses sales as the dependent variable and labor and capital inputs as the covariates. We study a dynamic model, but assume errors (ε_{it}) are uncorrelated. Also, to reduce the number of covariates, we use the ratio of sales over capitals as the dependent variable and the ratio of labors over capitals as the covariates.

6 Concluding remarks

This paper provides a new iterative estimation method for nonparametric dynamic panel models. We consider a short panel where the number of time periods T is fixed. The new estimator utilizes the additive structure of the first differenced model and is defined as a solution to a Fredholm integral equation of the second kind. We prove its uniform consistency and asymptotic normality. This paper also provides specification tests for the linearity of dynamic panel models. The tests are based on the L_2 distance between parametric and nonparametric estimators. Monte Carlo simulations show that our estimators and tests perform well in finite samples. We illustrate our methods with two empirical applications on economic growth and on firm sales. In the economic growth application, we find that the relationship between economic growth rates and initial income levels are nonlinear. However, we do not find nonlinearity between firm sales and labor inputs in the second application.

There are many interesting topics for further research. First, we only consider the test for neglected nonlinearity in linear panel data models. But the tools developed in this paper can be used to test for the correct specification of many parametric or semiparametric panel data models, including the widely used partially linear models where the lagged dependent variables enter the model linearly and the other regressors (usually exogenous) enter the model nonparametrically. Second, in terms of estimation, we did not explore all valid instruments in the information set. (2.3) suggests that all variables contained in $\underline{U}_{i,t-2}$ can be utilized for the estimation purpose but our iterative estimation strategy only requires the use of $U_{i,t-2}$, a subset of $\underline{U}_{i,t-2}$, in the spirit of Anderson and Hsiao (1981). It is not clear whether one can follow Arellano and Bond (1991) in the parametric framework and use other lagged variables in $\underline{U}_{i,t-2}$ to improve the efficiency of our estimate. It seems desirable to study the optimal choice of instruments. Third, our nonparametric panel data models can be extended along several dimensions. For example, we can also allow time effects in our model so that the mode in (2.1) becomes

$$Y_{it} = m\left(Y_{i,t-1}, X_{it}\right) + \alpha_i + \gamma_t + \varepsilon_{it},$$

where the extra term γ_t signifies the time effects. The appearance of the time effects will significantly complicate the analysis. For another example, one can also extend our estimation method to partially linear models

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

or functional coefficient models

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

where Z_{it} is a $p \times 1$ vector of regressors. We leave these for the future research.

Table 1: Estimation results ($T = 4$)

DGP	N	Median RMSE			Mean RMSE		
		Initial Estimator	Iterative Estimator	Non-iterative Estimator	Initial Estimator	Iterative Estimator	Non-iterative Estimator
1	50	0.270	0.186 (68.89%)	0.199 (73.70%)	0.302	0.196 (64.90%)	0.215 (71.19%)
	100	0.203	0.138 (67.98%)	0.153 (75.37%)	0.227	0.148 (65.20%)	0.162 (71.37%)
	200	0.128	0.100 (78.13%)	0.109 (85.16%)	0.138	0.108 (78.26%)	0.119 (86.23%)
2	50	0.504	0.292 (57.94%)	0.277 (54.96%)	0.524	0.306 (58.40%)	0.294 (56.11%)
	100	0.344	0.221 (64.24%)	0.216 (62.79%)	0.358	0.231 (64.53%)	0.227 (63.41%)
	200	0.266	0.170 (63.91%)	0.169 (63.53%)	0.273	0.176 (64.47%)	0.174 (63.74%)
3	50	0.286	0.182 (63.64%)	0.202 (70.63%)	0.312	0.202 (64.74%)	0.229 (73.40%)
	100	0.216	0.146 (67.59%)	0.159 (73.61%)	0.241	0.156 (64.73%)	0.173 (71.78%)
	200	0.134	0.115 (85.82%)	0.115 (85.82%)	0.144	0.122 (84.72%)	0.127 (88.19%)
4	50	0.273	0.189 (69.23%)	0.214 (78.39%)	0.297	0.205 (69.02%)	0.242 (81.48%)
	100	0.204	0.145 (71.08%)	0.166 (81.37%)	0.224	0.157 (70.09%)	0.183 (81.70%)
	200	0.133	0.113 (84.96%)	0.129 (96.99%)	0.144	0.118 (81.94%)	0.138 (95.83%)
5	50	0.601	0.435 (72.38%)	0.472 (78.54%)	0.644	0.452 (70.19%)	0.506 (78.57%)
	100	0.418	0.343 (82.06%)	0.369 (88.28%)	0.442	0.353 (79.86%)	0.392 (88.69%)
	200	0.333	0.266 (79.88%)	0.282 (84.68%)	0.353	0.277 (78.47%)	0.296 (83.85%)
6	50	0.539	0.374 (69.39%)	0.385 (71.43%)	0.570	0.386 (67.72%)	0.405 (71.05%)
	100	0.377	0.299 (79.31%)	0.309 (81.96%)	0.390	0.306 (78.46%)	0.320 (82.05%)
	200	0.285	0.231 (81.05%)	0.241 (84.56%)	0.291	0.237 (81.44%)	0.249 (85.57%)

Note: The numbers in brackets are the ratios of the iterative or non-iterative estimator's RMSE over that of the initial estimator.

Table 2: Estimation results ($T = 6$)

DGP	N	Median RMSE			Mean RMSE		
		Initial Estimator	Iterative Estimator	Non-iterative Estimator	Initial Estimator	Iterative Estimator	Non-iterative Estimator
1	50	0.194	0.141 (72.68%)	0.146 (75.26%)	0.206	0.151 (73.30%)	0.152 (73.79%)
	100	0.131	0.105 (80.15%)	0.110 (83.97%)	0.135	0.111 (82.22%)	0.116 (85.93%)
	200	0.092	0.080 (86.96%)	0.080 (86.96%)	0.099	0.084 (84.85%)	0.086 (86.87%)
2	50	0.345	0.217 (62.90%)	0.210 (60.87%)	0.353	0.229 (64.87%)	0.218 (61.76%)
	100	0.266	0.170 (63.91%)	0.166 (62.41%)	0.270	0.175 (64.81%)	0.173 (64.07%)
	200	0.198	0.132 (66.67%)	0.129 (65.15%)	0.200	0.135 (67.50%)	0.132 (66.00%)
3	50	0.196	0.142 (72.45%)	0.149 (76.02%)	0.215	0.153 (71.16%)	0.157 (73.02%)
	100	0.126	0.115 (91.27%)	0.111 (88.10%)	0.135	0.121 (89.63%)	0.120 (88.89%)
	200	0.097	0.091 (93.81%)	0.083 (85.57%)	0.102	0.096 (94.12%)	0.088 (86.27%)
4	50	0.183	0.140 (76.50%)	0.151 (82.51%)	0.195	0.153 (78.46%)	0.168 (86.15%)
	100	0.126	0.112 (88.89%)	0.120 (95.24%)	0.133	0.119 (89.47%)	0.132 (99.25%)
	200	0.093	0.083 (89.25%)	0.093 (100.00%)	0.100	0.088 (88.00%)	0.100 (100.00%)
5	50	0.393	0.331 (84.22%)	0.357 (90.84%)	0.418	0.340 (81.34%)	0.375 (89.71%)
	100	0.321	0.268 (83.49%)	0.279 (86.92%)	0.336	0.272 (80.95%)	0.291 (86.61%)
	200	0.220	0.210 (95.45%)	0.205 (93.18%)	0.228	0.213 (93.42%)	0.213 (93.42%)
6	50	0.359	0.293 (81.62%)	0.300 (83.57%)	0.371	0.300 (80.86%)	0.309 (83.29%)
	100	0.281	0.236 (83.99%)	0.240 (85.41%)	0.285	0.240 (84.21%)	0.249 (87.37%)
	200	0.208	0.187 (89.90%)	0.188 (90.38%)	0.210	0.189 (90.00%)	0.191 (90.95%)

Note: The numbers in brackets are the ratios of the iterative or non-iterative estimator's RMSE over that of the initial estimator.

Table 3: Median number of iterations

DGP \ (N, T)	$(50, 4)$	$(100, 4)$	$(200, 4)$	$(50, 6)$	$(100, 6)$	$(200, 6)$
1	4	3	3	3	3	3
2	5	4	4	4	4	3
3	3	3	2	3	2	2
4	3	2	2	2	2	2
5	4	3	3	3	3	3
6	4	3	3	3	3	3

Table 4: Empirical rejection frequency ($T = 4$)

DGP	N	T	Iterative Method			Non-iterative Method		
			1%	5%	10%	1%	5%	10%
1	50	4	0.012	0.040	0.060	0.000	0.008	0.036
	100	4	0.000	0.052	0.080	0.004	0.032	0.040
	200	4	0.008	0.044	0.064	0.004	0.024	0.056
2	50	4	0.024	0.056	0.096	0.004	0.024	0.048
	100	4	0.000	0.040	0.100	0.008	0.036	0.068
	200	4	0.028	0.052	0.104	0.016	0.036	0.060
3	50	4	0.232	0.488	0.664	0.208	0.404	0.536
	100	4	0.576	0.828	0.892	0.392	0.692	0.824
	200	4	0.892	0.988	0.992	0.844	0.972	0.988
4	50	4	0.336	0.648	0.764	0.164	0.432	0.568
	100	4	0.692	0.896	0.952	0.452	0.704	0.824
	200	4	0.976	1.000	1.000	0.944	0.992	1.000
5	50	4	0.088	0.284	0.524	0.108	0.392	0.612
	100	4	0.420	0.752	0.916	0.500	0.900	0.984
	200	4	0.912	1.000	1.000	0.992	1.000	1.000
6	50	4	0.148	0.372	0.524	0.152	0.372	0.480
	100	4	0.472	0.760	0.864	0.468	0.728	0.844
	200	4	0.900	0.980	0.992	0.896	0.988	0.988

Table 5: Empirical rejection frequency ($T = 6$)

DGP	N	T	Iterative Method			Non-iterative Method		
			1%	5%	10%	1%	5%	10%
1	50	6	0.000	0.032	0.064	0.008	0.028	0.060
	100	6	0.008	0.032	0.084	0.016	0.024	0.072
2	50	6	0.032	0.064	0.096	0.028	0.068	0.092
	100	6	0.012	0.076	0.148	0.016	0.068	0.136
3	50	6	0.556	0.828	0.900	0.648	0.852	0.884
	100	6	0.912	0.980	0.996	0.952	0.992	1.000
4	50	6	0.740	0.904	0.944	0.708	0.880	0.932
	100	6	0.992	1.000	1.000	0.984	1.000	1.000
5	50	6	0.328	0.708	0.900	0.768	0.968	0.992
	100	6	0.888	0.992	1.000	1.000	1.000	1.000
6	50	4	0.484	0.784	0.880	0.576	0.796	0.884
	100	4	0.880	0.976	0.992	0.936	0.980	0.996

Table 6: Specification test results for economic growth application (1960-2000)

	10 year growth rates	5 year growth rates
Iterative Method	0	0.04
Non-iterative Method	0.01	0.05

Note: The numbers in the main entries are the p-values based on 200 bootstraps.

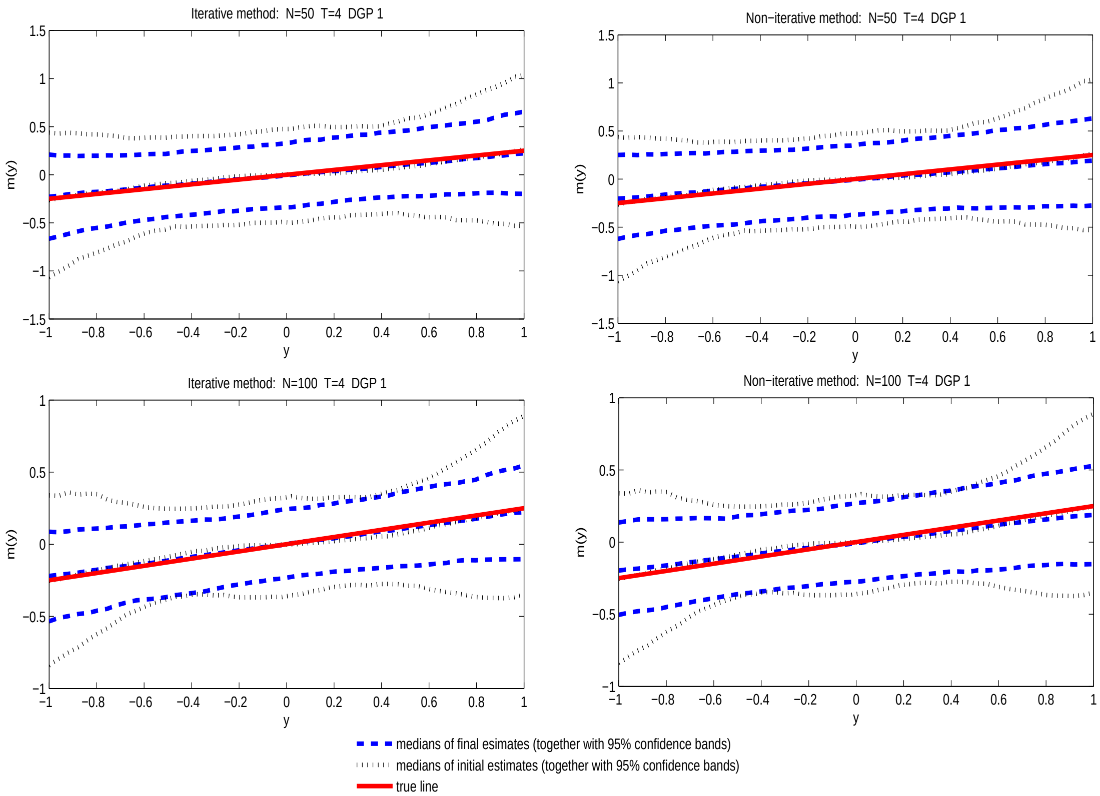

Figure 1: Estimation results for DGP 1

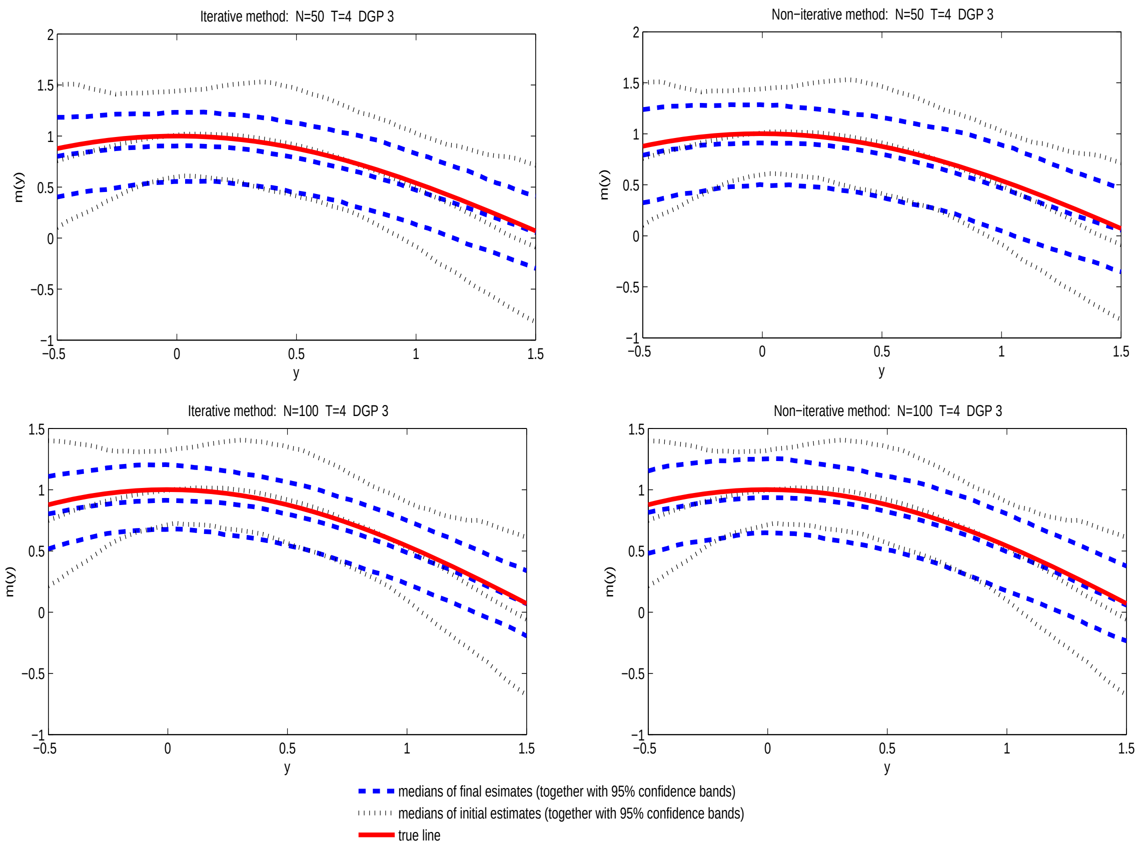

Figure 2: Estimation results for DGP 3

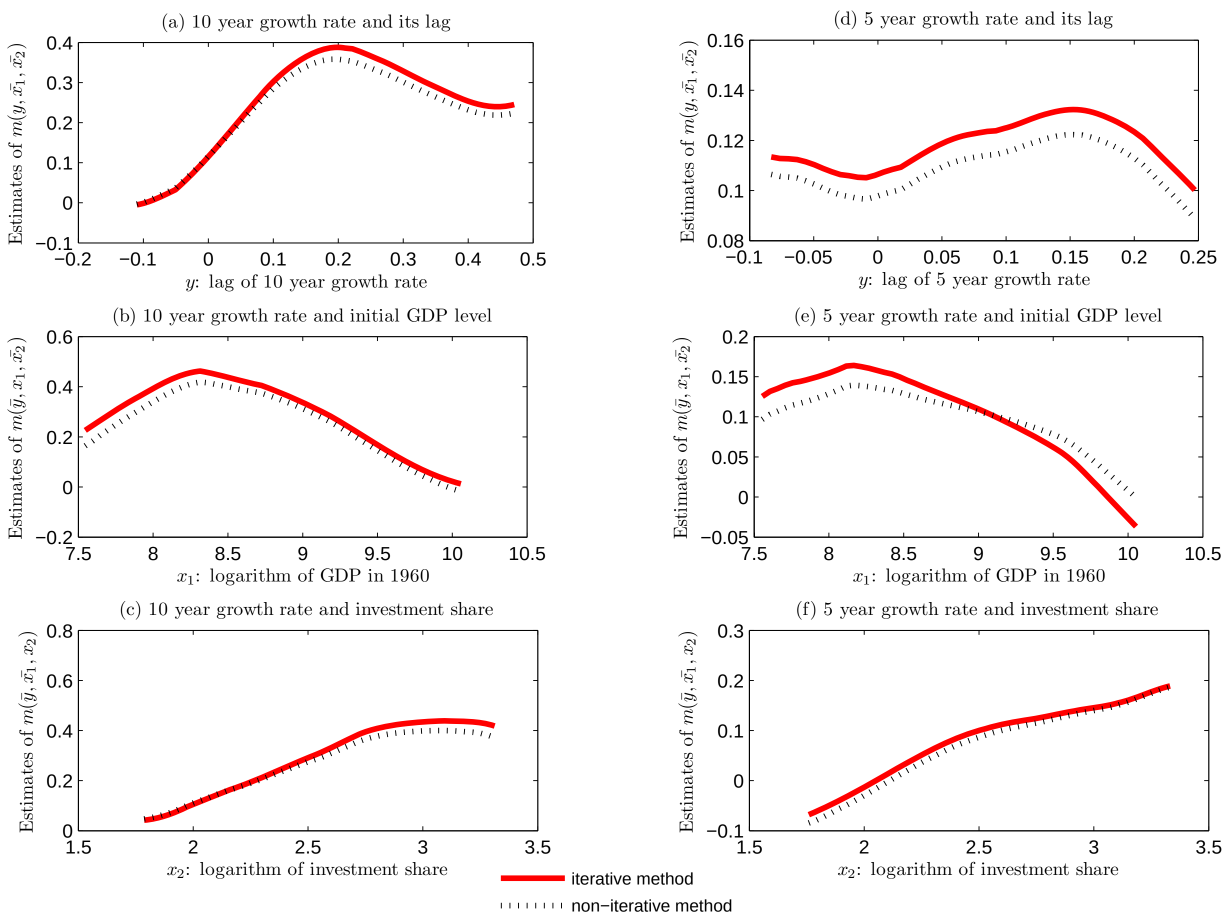

Figure 3: Economic growth, initial economic condition, and capital accumulation

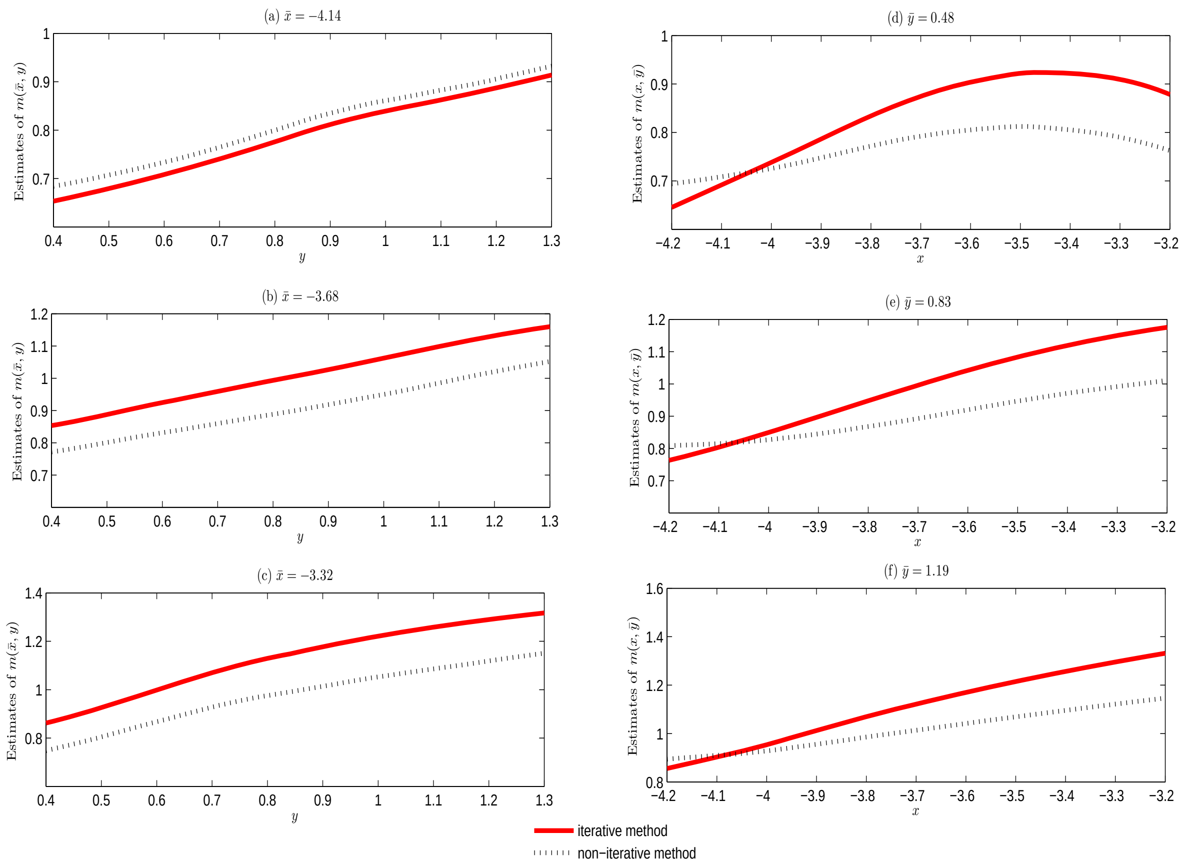

Figure 4: Firm labor inputs and sales

Appendix

A Proof of the results in Sections 2 and 3

Proof of Theorem 2.1

Let $|m|_\infty \equiv \sup_{u\in\mathcal{U}} |m(u)|$. Let $\Delta_n \equiv (nh!)^{-1/2}$ and $\nu_n \equiv \Delta_n (\log n)^{1/2} + \|h\|^{q+1}$. Following the proof of Theorem 1 in Mammen and Yu (2009, MY), we can prove the theorem by verifying the following conditions:

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

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

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

(B4) There exists a decomposition $\hat{r} - r + (\widehat{\mathcal{A}}-\mathcal{A})m = V_{NT} + B_{NT} + R_{NT}$ with random functions V_{NT}, B_{NT} and R_{NT} such that: a) $||V_{NT}||_2 = O_P(\Delta_n)$, b) $|\mathcal{A}(I-\mathcal{A})^{-1} V_{NT}|_\infty = O_P(\sqrt{\log n/n}\ )$, c) $||B_{NT}||_2 = O_P(\|h\|^{q+1})$, and d) $|R_{NT}|_\infty = O_P[\Delta_n(\log n)^{1/2}\nu_n]$. To see this, noting that $\mathcal{A}^{-1} - \mathcal{C}^{-1} = \mathcal{A}^{-1}(\mathcal{C}-\mathcal{A})\mathcal{C}^{-1}$ and $(I-\mathcal{A})^{-1} = I + \mathcal{A}(I-\mathcal{A})^{-1}$, we have

$$
\begin{aligned}
\hat{m} - m &= (I-\widehat{\mathcal{A}})^{-1}\hat{r} - (I-\mathcal{A})^{-1} r \\
&= (I-\widehat{\mathcal{A}})^{-1}(\hat{r}-r) + [(I-\widehat{\mathcal{A}})^{-1} - (I-\mathcal{A})^{-1}]r \\
&= (I-\widehat{\mathcal{A}})^{-1}\left[(\hat{r}-r) + (\widehat{\mathcal{A}}-\mathcal{A})(I-\mathcal{A})^{-1}r\right] \\
&= (I-\widehat{\mathcal{A}})^{-1}\left[(\hat{r}-r) + (\widehat{\mathcal{A}}-\mathcal{A})m\right] = (I-\widehat{\mathcal{A}})^{-1}[V_{NT} + B_{NT} + R_{NT}] \\
&= V_{NT} + \widehat{\mathcal{A}}(I-\widehat{\mathcal{A}})^{-1}V_{NT} + (I-\widehat{\mathcal{A}})^{-1}B_{NT} + (I-\widehat{\mathcal{A}})^{-1}R_{NT} \\
&= V_{NT} + (I-\mathcal{A})^{-1}B_{NT} + \left[\widehat{\mathcal{A}}(I-\widehat{\mathcal{A}})^{-1}V_{NT} + \mathcal{D}B_{NT} + (I-\widehat{\mathcal{A}})^{-1}R_{NT}\right], \qquad (A.1)
\end{aligned}
$$

where $\mathcal{D} \equiv (I-\widehat{\mathcal{A}})^{-1} - (I-\mathcal{A})^{-1}$. It follows that

$$
\left|\hat{m} - m - V_{NT} - (I-\mathcal{A})^{-1}B_{NT}\right|_\infty \leq \left|\widehat{\mathcal{A}}(I-\widehat{\mathcal{A}})^{-1}V_{NT}\right|_\infty + |\mathcal{D}B_{NT}|_\infty + \left|(I-\widehat{\mathcal{A}})^{-1}R_{NT}\right|_\infty. \qquad (A.2)
$$

Following the proof of Theorem 5 in MST, we can prove that under conditions (B1)-(B4),

$$
\begin{aligned}
|\widehat{\mathcal{A}}(I-\widehat{\mathcal{A}})^{-1}V_{NT}|_\infty &= O_P(\sqrt{\log n/n}),\ \ |\mathcal{D}B_{NT}|_\infty = O_P(\nu_n\,\|h\|^{q+1}),\ \text{and} \\
|(I-\widehat{\mathcal{A}})^{-1}R_{NT}|_\infty &= O_P[\Delta_n(\log n)^{1/2}\nu_n]. \qquad (A.3)
\end{aligned}
$$

Then the result in Theorem 2.1 follows.

First, Assumption A.1(vii) ensures (B1) and Assumption A.1(v) ensures (B2) as remarked in Section 2.2. Next, we verify (B3). Let $\bar{m}(u) \equiv \mathcal{A}m(u)$. Then we have the following bias-variance decomposition

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

where we have used the result that $\frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathcal{K}_{it}(u)=1$ in view of the fact $S_{NT}(u)^{-1}S_{NT}(u)=I_Q$. By the latter fact again, it is well known that we can write $A_{2NT}(u)$ as $A_{2NT}(u)=\frac{1}{n}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathcal{K}_{it}(u)$ $\times D\bar{m}_{it}(u)$, where $D\bar{m}_{it}(u)\equiv\bar{m}(U_{i,t-2})-\bar{m}(u)-\sum_{1\leq|\mathbf{j}|\leq q}\frac{1}{\mathbf{j}!}\bar{m}^{(\mathbf{j})}(u)(U_{i,t-2}-u)^{\mathbf{j}}$. By the standard arguments for local polynomial regressions [e.g., Masry (1996), Hansen (2008)],

$$\sup_{u\in\mathcal{U}}|A_{1NT}(u)|=O_P[\Delta_n(\log n)^{1/2}] \text{ and } \sup_{u\in\mathcal{U}}|A_{2NT}(u)|=O_P(\|h\|^{q+1}).$$

Then (B3) follows.

Now, we verify condition (B4) with V_{NT} and B_{NT} defined in (2.16) and R_{NT} given by

$$R_{NT}(u)=e_1'\{[S_{NT}(u)]^{-1}-\left[\bar{S}_{NT}(u)\right]^{-1}\}\left[R_{1NT}(u)+R_{2NT}(u)\right], \tag{A.4}$$

where

$$\begin{aligned} R_{1NT}(u) &= \frac{-1}{NT_2}\sum_{i=1}^{N}\sum_{t=3}^{T}1_{it}K_h(U_{i,t-2}-u)\,\mu_h(U_{i,t-2}-u)\,\Delta\varepsilon_{it}, \\ R_{2NT}(u) &= \frac{1}{NT_2}\sum_{i=1}^{N}\sum_{t=3}^{T}1_{it}K_h(U_{i,t-2}-u)\,\mu_h(U_{i,t-2}-u)\,Dm_{it}(u), \end{aligned}$$

and $Dm_{it}(u)\equiv m(U_{i,t-2})-m(u)-\sum_{1\leq|\mathbf{j}|\leq q}\frac{1}{\mathbf{j}!}m^{(\mathbf{j})}(u)(U_{i,t-2}-u)^{\mathbf{j}}$. Noting that $-\Delta Y_{it}+m(U_{i,t-1})=m(U_{i,t-2})-\Delta\varepsilon_{it}$ and $r(u)+\bar{m}(u)=m(u)$ by (2.6), we have

$$\begin{aligned} \hat{r}(u)-r(u)+\widehat{\mathcal{A}}m(u)-\mathcal{A}m(u) &= \frac{1}{NT_2}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathcal{K}_{it}(u)\left[-\Delta Y_{it}+m(U_{i,t-1})\right]-\left[r(u)+\bar{m}(u)\right] \\ &= \frac{1}{NT_2}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathcal{K}_{it}(u)\{-\Delta\varepsilon_{it}+\left[m(U_{i,t-2})-m(u)\right]\} \\ &= \frac{1}{NT_2}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathcal{K}_{it}(u)\{-\Delta\varepsilon_{it}+Dm_{it}(u)\} \\ &= V_{NT}(u)+B_{NT}(u)+R_{NT}(u). \end{aligned}$$

As in Masry (1996) and Hansen (2008), we can show that $\sup_{u\in\mathcal{U}}|R_{1NT}(u)|=O_P[\Delta_n(\log n)^{1/2}]$, $\sup_{u\in\mathcal{U}}|R_{2NT}(u)|=O_P(\|h\|^{q+1})$, and $\sup_{u\in\mathcal{U}}|S_{NT}(u)-\bar{S}_{NT}(u)|=O_P[\Delta_n(\log n)^{1/2}]$. It follows that $R_{NT}(u)=O_P[\Delta_n(\log n)^{1/2}\nu_n]$. This verifies (B4d).

By the Fubini theorem, Assumptions A.1(i), (ii), and (iii), it is easy to show that

$$\begin{aligned} E\|V_{NT}\|_2^2 &= \frac{1}{N^2T_2^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=3}^{T}\int_{\mathcal{U}}e_1'\left[\bar{S}_{NT}(u)\right]^{-1}E\left[1_{it}K_h(U_{i,t-2}-u)\,\mu_h(U_{i,t-2}-u)\,\Delta\varepsilon_{it}\right. \\ &\quad \left.\times\,1_{is}K_h(U_{i,s-2}-u)\,\mu_h(U_{i,s-2}-u)\,\Delta\varepsilon_{is}\right]\left[\bar{S}_{NT}(u)\right]^{-1}e_1f(u)\,du=O\left(\Delta_n^2\right). \end{aligned}$$

It follows that $\|V_{NT}\|_2=O_P(\Delta_n)$ by the Chebyshev inequality. This verifies (B4a). Define the operator $\mathcal{L}(\bar{u},u)$ by $\mathcal{A}(I-\mathcal{A})^{-1}m(u)=\int_{\mathcal{U}}\mathcal{L}(u,\bar{u})\,m(\bar{u})\,f(\bar{u})\,d\bar{u}$. Following Linton and Mammen (2005, p.821), we can show that $\int_{\mathcal{U}}\int_{\mathcal{U}}\mathcal{L}(u,\bar{u})^2 m(\bar{u}) f(\bar{u}) f(u)\, d\bar{u}du < \infty$ and

$$\begin{aligned}\mathcal{A}(I-\mathcal{A})^{-1} V_{NT}(u) &= \int_{\mathcal{U}} \mathcal{L}(u,\bar{u}) \frac{1}{NT_2}\sum_{i=1}^{N}\sum_{t=3}^{T} 1_{it} e_1' \bar{S}_{NT}(\bar{u})^{-1} \mu_h(U_{i,t-2}-\bar{u}) K_h(U_{i,t-2}-\bar{u})\, d\bar{u}\Delta\varepsilon_{it} \\ &= \frac{1}{NT_2}\sum_{i=1}^{N}\sum_{t=3}^{T} \xi(U_{i,t-2},u)\Delta\varepsilon_{it},\end{aligned}$$

where $\xi(v,u) = \int_{\mathcal{U}} \mathcal{L}(u,\bar{u}) 1(v\in\mathcal{U}) e_1'\bar{S}_{NT}(\bar{u})^{-1}\mu_h(v-\bar{u}) K_h(v-\bar{u})\, d\bar{u}$. Then we can apply the exponential inequality for IID data (as in Masry's proof for strong mixing data) to show $\sup_{u\in\mathcal{U}} |\mathcal{A}(I-\mathcal{A})^{-1}$ $V_{NT}(u)| = O_P(\sqrt{\log n/n})$, i.e., (B4b) holds. Finally, $|B_{NT}|_\infty = O_P(\|h\|^{q+1})$, implying that $\|B_{NT}\|_2 = O_P(\|h\|^{q+1})$, i.e., (B4c) holds. ■

Proof of Theorem 2.2

Theorem 2.1 implies that $\sqrt{nh!}[\hat{m}(u) - m(u) - (I-\mathcal{A})^{-1} B_{NT}(u)] = \sqrt{nh!}V_{NT}(u) + o_P(1)$. We prove the theorem by showing that

$$\begin{aligned}\sqrt{nh!}V_{NT}(u) &= \frac{-\sqrt{h!}}{\sqrt{n}}\sum_{i=1}^{N}\sum_{t=3}^{T} 1_{it} e_1' \left[\bar{S}_{NT}(u)\right]^{-1}\mu_h(U_{i,t-2}-u) K_h(U_{i,t-2}-u)\Delta\varepsilon_{it} \\ &\xrightarrow{D} N\left(0, \frac{\sigma^2(u)}{f(u)} e_1'\mathbb{S}^{-1}\int K(\bar{u})^2 \mu(\bar{u})\mu(\bar{u})'\, d\bar{u}\mathbb{S}^{-1}e_1\right) \qquad \text{(A.5)}\end{aligned}$$

and

$$\sqrt{nh!}\left[B_{NT}(u) - B_0(u)\right] = o_P(1). \qquad \text{(A.6)}$$

(A.5) can be proved by the Liapounov central limit theorem (CLT). To prove (A.6), we first calculate the bias

$$\begin{aligned}E\left[B_{NT}(u)\right] &= e_1'\left[\bar{S}_{NT}(u)\right]^{-1}\sum_{t=3}^{T}\frac{n_{t-2}}{n}\frac{1}{n_{t-2}}\sum_{i=1}^{N} E\left[1_{it}K_h(U_{i,t-2}-u)\mu_h(U_{i,t-2}-u) D_{it} m(u)\right] \\ &= e_1'\left[\bar{S}_{NT}(u)\right]^{-1}\sum_{t=3}^{T}\frac{p_{t-2}}{p}\int 1(u+h\odot w\in\mathcal{U}) K(w)\mu_h(w) \\ &\quad \times \sum_{|\mathbf{j}|=q+1}\frac{1}{\mathbf{j}!} m^{(\mathbf{j})}(u)(w\odot h)^{\mathbf{j}} f_{t-2}(u+h\odot w)\, dw + o(\|h\|^{q+1}) \\ &= e_1'\left[\bar{S}_{NT}(u)\right]^{-1}\sum_{|\mathbf{j}|=q+1}\frac{1}{\mathbf{j}!} m^{(\mathbf{j})}(u)\int K(w)\mu_h(w)(w\odot h)^{\mathbf{j}}\, dw f(u) + o(\|h\|^{q+1}) \\ &= e_1'\mathbb{S}^{-1}\sum_{|\mathbf{j}|=q+1}\frac{1}{\mathbf{j}!} m^{(\mathbf{j})}(u)\int K(w)\mu_h(w)(w\odot h)^{\mathbf{j}}\, dw + o(\|h\|^{q+1}),\end{aligned}$$

where recall $f(u) = \sum_{t=3}^{T}(p_{t-2}/p) f_{t-2}(u)$, and the last line follows from the fact that $\bar{S}_{NT}(u) \to f(u)\mathbb{S}$ for all $u \in$interior$(\mathcal{U})$ as $N\to\infty$. In addition, by the straightforward variance calculations, we have $\mathrm{Var}(B_{NT}(u)) = O((nh!)^{-1}\|h\|^2)$. Hence (A.6) follows by the fact that $\sqrt{nh!}\,\|h\|^{q+1} = O(1)$ by Assumption A.4. ■

Proof of Theorem 3.1

By (A.1), $\hat{m}(u) = m(u) + V_{NT}(u) + \hat{R}_{NT}(u)$, where

$$\hat{R}_{NT}(u) = (I - \mathcal{A})^{-1} B_{NT}(u) + \widehat{\mathcal{A}}(I - \widehat{\mathcal{A}})^{-1} V_{NT}(u) + \mathcal{D}B_{NT}(u) + (I - \widehat{\mathcal{A}})^{-1} R_{NT}(u). \qquad \text{(A.7)}$$

Let $b_{NT} \equiv NT_1 (h!)^{1/2}$ and $a_{it} \equiv a(U_{i,t-1})$. It follows that

$$\begin{aligned}
b_{NT}\Gamma_{NT} &= (h!)^{1/2} \sum_{i=1}^{N}\sum_{t=2}^{T} \left[m(U_{i,t-1}) + V_{NT}(U_{i,t-1}) - \hat{\beta}' U_{i,t-1} + \hat{R}_{NT}(U_{i,t-1})\right]^2 a_{it} \\
&= (h!)^{1/2} \sum_{i=1}^{N}\sum_{t=2}^{T} \left[m(U_{i,t-1}) + V_{NT}(U_{i,t-1}) - \hat{\beta}' U_{i,t-1}\right]^2 a_{it} \\
&\quad +2 (h!)^{1/2} \sum_{i=1}^{N}\sum_{t=2}^{T} \left[m(U_{i,t-1}) + V_{NT}(U_{i,t-1}) - \hat{\beta}' U_{i,t-1}\right] \hat{R}_{NT}(U_{i,t-1}) a_{it} \\
&\quad + (h!)^{1/2} \sum_{i=1}^{N}\sum_{t=2}^{T} \hat{R}_{NT}(U_{i,t-1})^2 a_{it} \\
&\equiv \Gamma_{NT1} + 2\Gamma_{NT2} + \Gamma_{NT3}, \text{ say.}
\end{aligned} \qquad \text{(A.8)}$$

First, we dispense with the term Γ_{NT3} that is easiest to analyze. Noting that under $\mathbb{H}_0$, $B_{NT}(u) = 0$ for all $u \in \mathbb{R}^{d+1}$ as $Dm_{it}(u) = 0$, we have

$$\hat{R}_{NT}(u) = \widehat{\mathcal{A}}(I - \widehat{\mathcal{A}})^{-1} V_{NT}(u) + (I - \widehat{\mathcal{A}})^{-1} R_{NT}(u) \equiv \tilde{R}_{NT}(u), \qquad \text{(A.9)}$$

and the result in (A.3) in the proof of Theorem 2.1 can be strengthened to

$$\left|(I - \widehat{\mathcal{A}})^{-1} R_{NT}\right|_{\infty} = O_p\left(\Delta_n^2 \log n\right). \qquad \text{(A.10)}$$

This, together with (A.3), implies that

$$\left|\tilde{R}_{NT}\right|_{\infty} = O_P[(\log n/n)^{1/2} + \Delta_n^2 \log n]. \qquad \text{(A.11)}$$

By (A.9), (A.11) and Assumption A.6, $\Gamma_{NT3} = NT (h!)^{1/2} O_P\left(\log n/n + \Delta_n^4 (\log n)^2\right) = o_P(1)$.

Next, we study Γ_{NT1}. We make the following decomposition:

$$\begin{aligned}
\Gamma_{NT1} &= (h!)^{1/2} \sum_{i=1}^{N}\sum_{t=2}^{T} \left\{V_{NT}(U_{i,t-1}) + \left[m(U_{i,t-1}) - \beta_0' U_{i,t-1}\right] + (\beta_0 - \hat{\beta})' U_{i,t-1}\right\}^2 a(U_{i,t-1}) \\
&= \sum_{l=1}^{6} D_{lNT},
\end{aligned} \qquad \text{(A.12)}$$

where

$$\begin{aligned}
D_{1NT} &= (h!)^{1/2} \sum_{i=1}^{N}\sum_{t=2}^{T} \left[V_{NT}(U_{i,t-1})\right]^2 a_{it}, \\
D_{2NT} &= (h!)^{1/2} \sum_{i=1}^{N}\sum_{t=2}^{T} \left[m(U_{i,t-1}) - \beta_0' U_{i,t-1}\right]^2 a_{it}, \\
D_{3NT} &= (h!)^{1/2} (\beta_0 - \hat{\beta})' \sum_{i=1}^{N}\sum_{t=2}^{T} U_{i,t-1} U_{i,t-1}' a_{it} (\beta_0 - \hat{\beta}),
\end{aligned}$$

$$
\begin{aligned}
D_{4NT} &= 2\left(h!\right)^{1/2}\sum_{i=1}^{N}\sum_{t=2}^{T}V_{NT}\left(U_{i,t-1}\right)\left[m(U_{i,t-1})-\beta_0'U_{i,t-1}\right]a_{it},\\
D_{5NT} &= 2\left(h!\right)^{1/2}(\beta_0-\hat{\beta})'\sum_{i=1}^{N}\sum_{t=2}^{T}U_{i,t-1}V_{NT}\left(U_{i,t-1}\right)a_{it},\\
D_{6NT} &= 2\left(h!\right)^{1/2}(\beta_0-\hat{\beta})'\sum_{i=1}^{N}\sum_{t=2}^{T}U_{i,t-1}\left[m(U_{i,t-1})-\beta_0'U_{i,t-1}\right]a_{it}.
\end{aligned}
$$

Under $\mathbb{H}_0$, $D_{lNT}=0$ for $l=2,\,4,\,6$. By Lemma A.1 below, $D_{1NT}-\mathbb{B}_{NT}\xrightarrow{D}N\left(0,\sigma_0^2\right)$ under $\mathbb{H}_0$. By Assumption A.5 and the Markov inequality, we can readily show that $D_{3NT}=O_P((h!)^{1/2})=o_P\left(1\right).$ Now, write $D_{5NT}=2(\beta_0-\hat{\beta})'\bar{D}_{5NT}$ where

$$
\begin{aligned}
\bar{D}_{5NT} &= (h!)^{1/2}\sum_{i=1}^{N}\sum_{t=2}^{T}U_{i,t-1}V_{NT}\left(U_{i,t-1}\right)a_{it}=\frac{(h!)^{1/2}}{NT_2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{N}\sum_{s=3}^{T}U_{i,t-1}\bar{\mathcal{K}}_{js,it}\Delta\varepsilon_{js}a_{it}\\
&= \frac{(h!)^{1/2}}{NT_2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=3}^{T}U_{i,t-1}\bar{\mathcal{K}}_{is,it}\Delta\varepsilon_{is}a_{it}+\frac{(h!)^{1/2}}{NT_2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j\neq i,j=1}^{N}\sum_{s=3}^{T}U_{i,t-1}\bar{\mathcal{K}}_{js,it}\Delta\varepsilon_{js}a_{it}\\
&\equiv \bar{D}_{5NT,1}+\bar{D}_{5NT,2},
\end{aligned}
$$

$\bar{\mathcal{K}}_{js,it}\equiv\bar{\mathcal{K}}_{js}\left(U_{i,t-1}\right),$ and $\bar{\mathcal{K}}_{it}\left(u\right)$ is defined in (2.17). It is easy to show that $\bar{D}_{5NT,1}=O_P[(h!)^{-1/2}]$. Noting that $E(\bar{D}_{5NT,2})=0$ and $E(\bar{D}_{5NT,2})^2=O(Nh!)$, we have $\bar{D}_{5NT,2}=O_P[(Nh!)^{1/2}]$ by the Chebyshev inequality. Hence $D_{5NT}=O_P[(Nh!)^{-1/2}+(h!)^{1/2}]=o_P\left(1\right).$

Now, we study Γ_{NT2}. We first decompose Γ_{NT2} as follows.

$$
\begin{aligned}
\Gamma_{NT2} &= (h!)^{1/2}\sum_{i=1}^{N}\sum_{t=2}^{T}V_{NT}\left(U_{i,t-1}\right)\hat{R}_{NT}\left(U_{i,t-1}\right)a_{it}\\
&\quad+(h!)^{1/2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[m(U_{i,t-1})-\beta_0'U_{i,t-1}\right]\hat{R}_{NT}\left(U_{i,t-1}\right)a_{it}\\
&\quad+(h!)^{1/2}(\beta_0-\hat{\beta})'\sum_{i=1}^{N}\sum_{t=2}^{T}U_{i,t-1}\hat{R}_{NT}\left(U_{i,t-1}\right)a_{it}\equiv\Gamma_{NT2,1}+\Gamma_{NT2,2}+\Gamma_{NT2,3}. \quad (A.13)
\end{aligned}
$$

Note that $\Gamma_{NT2,2}=0$ under $\mathbb{H}_0$. By (A.9), (A.11) and Assumptions A.5-A.6, we have

$$
\begin{aligned}
\left|\Gamma_{NT2,3}\right| &\leq (h!)^{1/2}\left\|\beta_0-\hat{\beta}\right\|\left|\tilde{R}_{NT}\right|_{\infty}\sum_{i=1}^{N}\sum_{t=2}^{T}\left|U_{i,t-1}a_{it}\right|\\
&= (h!)^{1/2}O_P(N^{-1/2})O_P[(\log n/n)^{1/2}+\Delta_n^2\log n]O_P\left(N\right)=o_P\left(1\right).
\end{aligned}
$$

Under $\mathbb{H}_0$, by (A.9) we can further decompose $\Gamma_{NT2,1}$ as follows

$$
\begin{aligned}
\Gamma_{NT2,1} &= (h!)^{1/2}\sum_{i=1}^{N}\sum_{t=2}^{T}V_{NT}\left(U_{i,t-1}\right)\widehat{\mathcal{A}}(I-\widehat{\mathcal{A}})^{-1}V_{NT}\left(U_{i,t-1}\right)a_{it}\\
&\quad+(h!)^{1/2}\sum_{i=1}^{N}\sum_{t=2}^{T}V_{NT}\left(U_{i,t-1}\right)(I-\widehat{\mathcal{A}})^{-1}R_{NT}\left(u\right)a_{it}\\
&\equiv \Gamma_{NT2,11}+\Gamma_{NT2,12},\ \text{say.}
\end{aligned}
$$

To study $\Gamma_{NT2,11}$, we can first show that $|\widehat{\mathcal{A}}(I-\widehat{\mathcal{A}})^{-1}V_{NT}-\mathcal{A}(I-\mathcal{A})^{-1}V_{NT}|_{\infty}=O_P[\Delta_n^2(\log n)^{1/2}]$. Using this, the uniform bound for $V_{NT}(u)$ and Assumption A.6, we have $\Gamma_{NT2,11}=\bar{\Gamma}_{NT2,11}+N(h!)^{1/2}$ $O_P[\Delta_n^2(\log n)^{1/2}]O_P[\Delta_n(\log n)^{1/2}]=\bar{\Gamma}_{NT2,11}+o_P(1)$, where

$$\bar{\Gamma}_{NT2,11}=(h!)^{1/2}\sum_{i=1}^{N}\sum_{t=2}^{T}V_{NT}(U_{i,t-1})\mathcal{A}(I-\mathcal{A})^{-1}V_{NT}(U_{i,t-1})a_{it}.$$

By straightfoward moment calculations and the Chebyshev inequality, we can show that $\bar{\Gamma}_{NT2,11}=o_P(1)$. Hence $\Gamma_{NT2,11}=o_P(1)$. For $\Gamma_{NT2,12}$, by the Jensen inequality, (A.10), the study of D_{1NT} and Assumption A.6, we have

$$\begin{aligned}|\Gamma_{NT2,12}| &\leq (h!)^{1/2}\left|(I-\widehat{\mathcal{A}})^{-1}R_{NT}\right|_{\infty}\sum_{i=1}^{N}\sum_{t=2}^{T}|V_{NT}(U_{i,t-1})|a_{it}\\ &\leq \sqrt{NT_1}(h!)^{1/2}\left|(I-\widehat{\mathcal{A}})^{-1}R_{NT}\right|_{\infty}\left\{\sum_{i=1}^{N}\sum_{t=2}^{T}[V_{NT}(U_{i,t-1})]^2a_{it}\right\}^{1/2}\\ &= \sqrt{NT_1}(h!)^{1/2}O_p\left(\Delta_n^2\log n\right)O_P[(h!)^{-1/2}]=o_P(1).\end{aligned}$$

Hence $\Gamma_{NT2,1}=o_P(1)$ and $\Gamma_{NT2}=o_P(1)$. This completes the proof of Theorem 3.1. ■

Lemma A.1 $D_{1NT}-\mathbb{B}_{NT}\overset{D}{\to}N(0,\sigma_0^2)$.

Proof. Let $W_{it}\equiv(U_{i,t-2}',\Delta\varepsilon_{it})'$ and $W_i\equiv(W_{i3}',...,W_{iT}')'$. Then we can write D_{1NT} as a third order V-statistic:

$$\begin{aligned}D_{1NT} &= \frac{(h!)^{1/2}}{n^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{N}\sum_{s=3}^{T}\sum_{l=1}^{N}\sum_{r=3}^{T}\bar{\mathcal{K}}_{js,it}\bar{\mathcal{K}}_{lr,it}\Delta\varepsilon_{js}\Delta\varepsilon_{lr}a_{it}\\ &= \frac{(h!)^{1/2}}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\zeta(W_{i_1},W_{i_2},W_{i_3}),\end{aligned}$$

where $\zeta(W_i,W_j,W_l)=N^2n^{-2}\sum_{t=2}^{T}\sum_{s=3}^{T}\sum_{r=3}^{T}\bar{\mathcal{K}}_{js,it}\bar{\mathcal{K}}_{lr,it}\Delta\varepsilon_{js}\Delta\varepsilon_{lr}a_{it}$. To study the asymptotic distribution of D_{1NT}, we need to use the U-statistic theory (e.g., Lee (1990)). Let $\varphi(w_{i_1},w_{i_2})\equiv E[\zeta(W_1,w_{i_1},$ $w_{i_2})]$, and $\bar{\zeta}(w_{i_1},w_{i_2},w_{i_3})\equiv\zeta(w_{i_1},w_{i_2},w_{i_3})-\varphi(w_{i_2},w_{i_3})$. Then we can decompose D_{1NT} as follows

$$\begin{aligned}D_{1NT} &= \frac{(h!)^{1/2}}{N^2}\sum_{i=1}^{N}\sum_{i_2=1}^{N}\varphi(W_{i_1},W_{i_2})+\frac{(h!)^{1/2}}{N^2}\sum_{i=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\bar{\zeta}(W_{i_1},W_{i_2},W_{i_3})\\ &\equiv D_{1NT,1}+D_{1NT,2},\ \text{say.}\end{aligned}$$

First, we consider $D_{1NT,2}$. Write $E(D_{1NT,2})^2=N^{-4}h!\sum_{i_1,...,i_6=1}^{N}E\left[\bar{\zeta}(W_{i_1},W_{i_2},W_{i_3})\bar{\zeta}(W_{i_4},W_{i_5},W_{i_6})\right]$. Observing that $E\left[\bar{\zeta}(W_{i_1},w_{i_2},w_{i_3})\right]=E\left[\bar{\zeta}(w_{i_1},W_{i_2},w_{i_3})\right]=E[\bar{\zeta}(w_{i_1},w_{i_2},W_{i_3})]=0$, $E[\bar{\zeta}(W_{i_1},W_{i_2},$ $W_{i_3})\bar{\zeta}(W_{i_4},W_{i_5},W_{i_6})]=0$ if there are more than three distinct elements in $\{i_1,\ldots,i_6\}$. In view of this, we can show that

$$E(D_{1NT,2})^2=O(N^{-1}(h!)^{-1}+N^{-2}(h!)^{-2}+N^{-3}(h!)^{-3})=o(1).$$

Then $D_{1NT,2} = o_P(1)$ by the Chebyshev inequality.

Now, we consider $D_{1NT,1}$. Note that

$$\begin{aligned}\varphi(W_i, W_j) &= N^2 n^{-2} \sum_{t=2}^{T} \sum_{s=3}^{T} \sum_{r=3}^{T} \int \bar{\mathcal{K}}_{is}(u) \bar{\mathcal{K}}_{jr}(u) \Delta\varepsilon_{js} \Delta\varepsilon_{lr} a(u) f_{t-1}^{(uc)}(u)\, du \\ &= T_1 N^2 n^{-2} \sum_{s=3}^{T} \sum_{r=3}^{T} \int \bar{\mathcal{K}}_{is}(u) \bar{\mathcal{K}}_{jr}(u) \Delta\varepsilon_{js} \Delta\varepsilon_{lr} a(u) \bar{f}(u)\, du,\end{aligned}$$

where recall $\bar{f}(u) \equiv T_1^{-1} \sum_{t=2}^{T} f_{t-1}^{(uc)}(u)$. Then $D_{1NT,1} = \mathbb{B}_{1NT} + \mathbb{V}_{1NT}$, where $\mathbb{B}_{1NT} = N^{-1}(h!)^{1/2} \sum_{i=1}^{N} \varphi(W_i, W_i)$ and $\mathbb{V}_{1NT} = 2N^{-1}(h!)^{1/2} \sum_{1 \le i < j \le N} \varphi(W_i, W_j)$ contribute to the asymptotic bias and variance of our test statistic, respectively. Note that

$$\begin{aligned}\mathbb{B}_{1NT} &= N^{-1}(h!)^{1/2} T_1 N^2 n^{-2} \sum_{j=1}^{N} \sum_{s=3}^{T} \sum_{r=3}^{T} \int e_1' \bar{S}_{NT}(u)^{-1} \mu_h(U_{j,s-2} - u) \mu_h(U_{j,r-2} - u)' \bar{S}_{NT}(u)^{-1} e_1 \\ &\quad \times 1_{js} 1_{jr} K_h(U_{j,s-2} - u) K_h(U_{j,r-2} - u) \Delta\varepsilon_{js} \Delta\varepsilon_{jr} a(u) \bar{f}(u)\, du \\ &= (h!)^{1/2} T_1 N n^{-2} \sum_{j=1}^{N} \sum_{s=3}^{T} \int e_1' \bar{S}_{NT}(u)^{-1} \mu_h(U_{j,s-2} - u) \mu_h(U_{j,s-2} - u)' \bar{S}_{NT}(u)^{-1} e_1 \\ &\quad \times 1_{js} K_h(U_{j,s-2} - u)^2 [\Delta\varepsilon_{js}]^2 a(u) \bar{f}(u)\, du + O_P((h!)^{1/2}) \\ &= (h!)^{1/2} T_1 N n^{-2} \sum_{j=1}^{N} \sum_{s=3}^{T} \int E[e_1' \bar{S}_{NT}(u)^{-1} \mu_h(U_{j,s-2} - u) \mu_h(U_{j,s-2} - u)' \bar{S}_{NT}(u)^{-1} e_1 \\ &\quad \times 1_{js} K_h(U_{j,s-2} - u)^2 \sigma^2(U_{j,s-2})] a(u) \bar{f}(u)\, du + O_P(Nh!)^{-1/2} + (h!)^{1/2}) \\ &= (h!)^{-1/2} T_1 d^{-1} \int \int e_1' \bar{S}_{NT}(u)^{-1} \mu(v) \mu(v)' \bar{S}_{NT}(u)^{-1} e_1 K(v)^2 \sigma^2(u + hv)\, dv a(u) \bar{f}(u)\, du \\ &\quad + O_P(Nh!)^{-1/2} + (h!)^{1/2}) \\ &\equiv \mathbb{B}_{NT} + O_P[(Nh!)^{-1/2} + (h!)^{1/2}], \text{ say,}\end{aligned}$$

where the third equality follows from the straightforward moment calculations and the Chebyshev inequality. Note that $\mathbb{B}_{NT} = O[(h!)^{-1/2}]$. If $d < 3$, noting that $(h!)^{-1/2} \|h\|^2 = o(1)$, we have

$$\begin{aligned}\mathbb{B}_{NT} &= (h!)^{1/2} T_1 N n^{-1} \int e_1' \bar{S}_{NT}(u)^{-1} \int K(v)^2 \mu(v) \mu(v)'\, dv \bar{S}_{NT}(u)^{-1} e_1 \sigma^2(u) a(u) \bar{f}(u) f(u)\, du \\ &\quad + o(1) \\ &= (h!)^{1/2} T_1 N n^{-1} \bar{C}_1 \int \sigma^2(u) a(u) \bar{f}(u) f(u)^{-1}\, du + o(1),\end{aligned}$$

where $\bar{C}_1 = e_1' \mathbb{S}^{-1} \int K(v)^2 \mu(v) \mu(v)'\, dv \mathbb{S}^{-1} e_1$. If $q = 1$, $\bar{C}_1 = \int K(v)^2\, dv = C_1^{d+1}$.

In view of the fact that $\mathbb{V}_{1NT}$ is a second-order degenerate U-statistic and W_i are IID across i, we can easily verify that all the conditions of Theorem 1 of Hall (1984) are satisfied and a central limit theorem applies to it: $\mathbb{V}_{1NT} \xrightarrow{D} N(0, \sigma_0^2)$, where $\sigma_0^2 = \lim_{n\to\infty} \sigma_{NT}^2$ and $\sigma_{NT}^2 = 2h! E[\varphi(W_1, W_2)]^2$.

We now calculate σ_{NT}^2.

$$\begin{aligned}
\sigma_{NT}^2 &= 2h!T_1^2N^4n^{-4}E\left[\sum_{s=3}^{T}\sum_{r=3}^{T}\int\bar{\mathcal{K}}_{1s}(u)\,\bar{\mathcal{K}}_{2r}(u)\,\Delta\varepsilon_{1s}\Delta\varepsilon_{2r}a(u)\,\bar{f}(u)\,du\right]^2 \\
&= 2h!T_1^2N^4n^{-4}E\left[\sum_{s=3}^{T}\sum_{r=3}^{T}\int e_1'\bar{S}_{NT}(u)^{-1}\mu_h(U_{1,s-2}-u)\,\mu_h(U_{2,r-2}-u)'\,\bar{S}_{NT}(u)^{-1}e_1\right. \\
&\quad \left.\times K_h(U_{1,s-2}-u)\,K_h(U_{2,r-2}-u)\,1_{1s}1_{2r}\Delta\varepsilon_{1s}\Delta\varepsilon_{2r}a(u)\,\bar{f}(u)\,du\right]^2 \\
&= 2(h!)^{-1}T_1^2N^4n^{-4} \\
&\quad \times E\left[\sum_{s=3}^{T}\sum_{r=3}^{T}\int e_1'\bar{S}_{NT}(U_{1,s-2}+h\odot\tilde{u})^{-1}\mu(\tilde{u})\,\mu\left(\tilde{u}+\frac{U_{1,s-2}\text{-}U_{2,r-2}}{h}\right)'\bar{S}_{NT}(U_{1,s-2}+h\odot\tilde{u})^{-1}e_1\right. \\
&\quad \left.\times K(\tilde{u})\,K\left(\tilde{u}+\frac{U_{1,s-2}\text{-}U_{2,r-2}}{h}\right)1_{1s}1_{2r}\Delta\varepsilon_{1s}\Delta\varepsilon_{2r}a(U_{1,s-2}+h\odot\tilde{u})\,\bar{f}(U_{1,s-2}+h\odot\tilde{u})\,d\tilde{u}\right]^2 \\
&= 2(h!)^{-1}T_1^2N^4n^{-4}E\left[\sum_{s=3}^{T}\sum_{r=3}^{T}\int\int e_1'\bar{S}_{NT}(U_{1,s-2})^{-1}\mu(u)\,\mu\left(u+\frac{U_{1,s-2}\text{-}U_{2,r-2}}{h}\right)'\bar{S}_{NT}(U_{1,s-2})^{-1}e_1\right. \\
&\quad \times e_1'\bar{S}_{NT}(U_{1,s-2})^{-1}\mu(\tilde{u})\,\mu\left(\tilde{u}+\frac{U_{1,s-2}\text{-}U_{2,r-2}}{h}\right)'\bar{S}_{NT}(U_{1,s-2})^{-1}e_1 \\
&\quad \times K(u)\,K(\tilde{u})\,K\left(u+\frac{U_{1,s-2}-U_{2,r-2}}{h}\right)K\left(\tilde{u}+\frac{U_{1,s-2}-U_{2,r-2}}{h}\right) \\
&\quad \left.\times 1_{1s}1_{2r}\sigma_{s-2}^2(U_{1,s-2})\,\sigma_{r-2}^2(U_{2,r-2})\,a(U_{1,s-2})^2\,\bar{f}(U_{1,s-2})^2\,d\tilde{u}du\right]+O(\|h\|^2+h!) \\
&= 2T_1^2p^{-2}\int\int\int\int e_1'\bar{S}_{NT}(v)^{-1}\mu(u)\,\mu(u+w)'\,\bar{S}_{NT}(v)^{-1}e_1e_1'\bar{S}_{NT}(v)^{-1}\mu(\tilde{u})\,\mu(\tilde{u}+w)'\,\bar{S}_{NT}(v)^{-1}e_1 \\
&\quad \left.\times K(u)\,K(\tilde{u})\,K(u+w)\,K(\tilde{u}+w)\,\sigma^4(v)\,a(v)^2\,\bar{f}(v)^2\,d\tilde{u}dudvdw\right]+O(\|h\|^2+h!) \\
&= \sigma_0^2+o(1),
\end{aligned}$$

where $T_1^2N^2n^{-2}\to T_1^2/p^2$ as $n/N\overset{P}{\to}p$. In the case where $q=1$, we have $\sigma_0^2=\sigma_0^{2(ll)}=2T_1^2p^{-2}\int[\int K(u)$ $K(u+v)\,du]^2dv\int\sigma^4(\bar{u})\,a(\bar{u})^2\,\bar{f}(\bar{u})^2\,f(\bar{u})^{-4}\,d\bar{u}$. ■

Proof of Corollary 3.2

By Theorem 3.1, it suffices to prove (i) $\hat{B}_{NT}^{(ll)}=B_{NT}^{(ll)}+o_P(1)$ and (ii) $\hat{\sigma}_{NT}^{2(ll)}=\sigma_0^{2(ll)}+o_P(1)$ for the local linear case with $d<3$,[13] and (iii) $\hat{B}_{NT}=B_{NT}+o_P(1)$ and (iv) $\hat{\sigma}_{NT}^2=\sigma_0^2+o_P(1)$ for the general case. We only prove (i) and (ii) because the proofs of (iii) and (iv) are analogous but tedious. In fact, we prove a stronger claim: (i) and (ii) hold under $\mathbb{H}_1(\lambda_{NT})$ with $\lambda_{NT}=(NT_1)^{-1/2}(h!)^{-1/4}$.

We first show (i). In view of the fact that $m(u)=\beta_0'u+\lambda_{NT}\delta_{NT}(u)$ under $\mathbb{H}_1(\lambda_{NT})$, we have

$$\begin{aligned}
\hat{\Delta}\varepsilon_{it} &= \Delta\varepsilon_{it}+(\beta_0-\hat{\beta})'(U_{i,t-1}-U_{i.t-2})+\lambda_{NT}[\delta_{NT}(U_{i,t-1})-\delta_{NT}(U_{i,t-2})] \\
&\equiv \Delta\varepsilon_{it}+A_{1it}+A_{2it},\ \text{say.}
\end{aligned}$$

[13] Strictly speaking, the proof of (i) requires both $q=1$ and $d<3$, whereas that of (ii) requires only $q=1$.

Using the fact that $\hat{\beta}-\beta_0 = O_P(\lambda_{NT})$ under $\mathbb{H}_1(\lambda_{NT})$, and Assumptions A2-A4, we can readily show that uniformly in $u \in \mathcal{U}$,

$$\begin{aligned}
\hat{\sigma}^2(u) &= \frac{1}{n}\sum_{j=1}^{N}\sum_{s=3}^{T} 1_{js} L_h(U_{j,s-2}-u)(\Delta\varepsilon_{js}+A_{1js}+A_{2js})^2 \\
&= \frac{1}{n}\sum_{j=1}^{N}\sum_{s=3}^{T} 1_{js} L_h(U_{j,s-2}-u)\Delta\varepsilon_{js}^2 + O_P(\lambda_{NT}) \\
&= \sum_{s=3}^{T}\frac{n_{s-2}}{n}\frac{1}{n_{s-2}}\sum_{j=1}^{N} E\left[1_{js}L_h(U_{j,s-2}-u)\Delta\varepsilon_{js}^2\right] + O_P[\lambda_{NT}+(Nh!)^{-1/2}] \\
&= \sum_{s=3}^{T}\frac{p_{s-2}}{p}\sigma_{s-2}^2(u) f_{s-2}(u) + O_P[\lambda_{NT}+(Nh!)^{-1/2}+\|h\|^{\gamma}] \\
&= \sigma^2(u) + O_P[(Nh!)^{-1/2}+\|h\|^{\gamma}].
\end{aligned}$$

In addition, $|\hat{f}-f|_\infty = O_P[(Nh!/\log N)^{-1/2}+\|h\|^{\gamma}]$ by standard theory for kernel density estimation. It follows that

$$\begin{aligned}
\hat{\mathbb{B}}_{NT}^{(ll)} &= (h!)^{-1/2} T_1 N n^{-1} C_1^{d+1}\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left\{\sigma^2(U_{i,t-1}) + O_P[(Nh!)^{-1/2}+\|h\|^{\gamma}]\right\} a(U_{i,t-1}) \\
&\quad \times \left\{ f(U_{i,t-1})^{-2} + O_P[(Nh!/\log N)^{-1/2}+\|h\|^{\gamma}]\right\} \\
&= (h!)^{-1/2} T_1 N n^{-1} C_1^{d+1}\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\sigma^2(U_{i,t-1}) a(U_{i,t-1}) f(U_{i,t-1})^{-2} \\
&\quad + (h!)^{-1/2} O_P((Nh!/\log N)^{-1/2}+\|h\|^{\gamma}) \\
&= \mathbb{B}_{NT}^{(ll)} + o_P(1).
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
\hat{\sigma}_{NT}^{2(ll)} &= 2T_1^2 N^2 n^{-2} C_2^{d+1}\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\sigma^4(U_{i,t-1}) a(U_{i,t-1}) \bar{f}(U_{i,t-1}) f(U_{i,t-1})^{-4} + o_P(1) \\
&= \sigma_{NT}^{2(ll)} + o_P(1). \blacksquare
\end{aligned}$$

Proof of Theorem 3.3

The proof follows closely from that of Theorem 3.1, now keeping the additional terms that do not vanish under $\mathbb{H}_1(\lambda_{NT})$ with $\lambda_{NT} = (NT_1)^{-1/2}(h!)^{-1/4}$. In view of the fact $\hat{B}_{NT} = B_{NT} + o_P(1)$ and $\hat{\sigma}_{NT} = \sigma_0^2 + o_P(1)$ [or $\hat{B}_{NT}^{(ll)} = B_{NT}^{(ll)} + o_P(1)$ and $\hat{\sigma}_{NT}^{2(ll)} = \sigma_0^{2(ll)} + o_P(1)$ if $d<3$ and $q=1$] under $\mathbb{H}_1(\lambda_{NT})$ and the results for D_{lNT}, $l=1,3,5$, continue to hold $\mathbb{H}_1(\lambda_{NT})$, it suffices to show that under $\mathbb{H}_1(\lambda_{NT})$, (i) $\Gamma_{NT3} = o_P(1)$, (ii) $\Gamma_{NT2} = o_P(1)$, (iii) $D_{2NT} \xrightarrow{P} \mu_0$ and (iv) $D_{lNT} = o_P(1)$ for $s=4$, 6, where Γ_{NT3}, Γ_{NT2}, and the D's are defined in the proof of Theorem 3.1.

We first show (i). Decompose

$$\begin{aligned}
\hat{R}_{NT}(u) &= \left[(I-\mathcal{A})^{-1} B_{NT}(u) + \mathcal{D}B_{NT}(u)\right] + \left[\widehat{\mathcal{A}}(I-\widehat{\mathcal{A}})^{-1}V_{NT}(u) + (I-\widehat{\mathcal{A}})^{-1}R_{NT}(u)\right] \\
&\equiv \tilde{B}_{NT}(u) + \tilde{R}_{NT}(u).
\end{aligned}$$

Noting that $Dm_{it}(u) = O_P(\lambda_{NT} \|h\|^{q+1})$ and $|B_{NT}|_\infty = O_P(\lambda_{NT} \|h\|^{q+1})$ under $\mathbb{H}_1(\lambda_{NT})$, we can show that $\left|\tilde{B}_{NT}\right|_\infty = O_P(\lambda_{NT} \|h\|^{q+1})$ and $\left|\tilde{R}_{NT}\right|_\infty = O_P[(\log n/n)^{1/2} + \Delta_n^2 \log n + \lambda_{NT} \|h\|^{q+1}]$. It follows that

$$\left|\hat{R}_{NT}\right|_\infty = O_P[(\log n/n)^{1/2} + \Delta_n^2 \log n + \lambda_{NT} \|h\|^{q+1}] \text{ under } \mathbb{H}_1(\lambda_{NT}), \tag{A.14}$$

and

$$\begin{aligned} \Gamma_{NT3} &\leq 2\left|\hat{R}_{NT}\right|_\infty^2 (h!)^{1/2} \sum_{i=1}^{N}\sum_{t=2}^{T} a(U_{i,t-1}^{[NT]}) \\ &= O_P[\log n/n + \Delta_n^4 (\log n)^2 + \lambda_{NT}^2 \|h\|^{2(q+1)}] (h!)^{1/2} O(N) = o_P(1). \end{aligned}$$

Similarly, using (A.14) and the decomposition in (A.13), we can show that $\Gamma_{NT2} = o_P(1)$ under $\mathbb{H}_1(\lambda_{NT})$.

To show (iii), using the WLLN for IID data (along the individual dimension) yields

$$\begin{aligned} D_{2NT} &= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[\delta_{NT}(U_{i,t-1}^{[NT]})\right]^2 a(U_{i,t-1}^{[NT]}) \\ &= \int \delta_{NT}(u)^2 a(u) f^{[NT]}(u)\, du + o_P(1) = \mu_0 + o_P(1) \text{ under } \mathbb{H}_1(\lambda_{NT}). \end{aligned}$$

Let $\bar{\mathcal{K}}_{js,it}^{[NT]}$ be as defined as $\bar{\mathcal{K}}_{js,it}$ with $(U_{i,t-1}^{[NT]}, U_{j,s-2}^{[NT]})$ in place of $(U_{i,t-1}, U_{j,s-2})$. Then under $\mathbb{H}_1(\lambda_{NT})$

$$\begin{aligned} D_{4NT} &= \frac{-2\gamma_{NT}(h!)^{1/2}}{NT_2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{N}\sum_{s=3}^{T}\bar{\mathcal{K}}_{js,it}^{[NT]}\Delta\varepsilon_{js}\delta_{NT}(U_{i,t-1}^{[NT]})a(U_{i,t-1}^{[NT]}) \\ &= -\frac{2\gamma_{NT}(h!)^{1/2}}{NT_2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=3}^{T}\bar{\mathcal{K}}_{is,it}^{[NT]}\Delta\varepsilon_{is}\delta_{NT}(U_{i,t-1}^{[NT]})a(U_{i,t-1}^{[NT]}) \\ &\quad -\frac{2\gamma_{NT}(h!)^{1/2}}{NT_2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j\neq i,j=1}^{N}\sum_{s=3}^{T}\bar{\mathcal{K}}_{js,it}^{[NT]}\Delta\varepsilon_{js}\delta_{NT}(U_{i,t-1}^{[NT]})a(U_{i,t-1}^{[NT]}) \\ &= -2D_{4NT,1} - 2D_{4NT,2}. \end{aligned}$$

It is easy to show that $D_{4NT,1} = O_P[\gamma_{NT}(h!)^{-1/2}] = O_P[N^{-1/2}(h!)^{-3/4}] = o_P(1)$. In view of that $E(D_{4NT,2}) = 0$, and $E(D_{4NT,2})^2 = O_P[(h!)^{1/2}]$, we have $D_{4NT,2} = O_P[(h!)^{1/4}]$ by the Chebyshev inequality. It follows that $D_{4NT} = O_P[N^{-1/2}(h!)^{-3/4} + (h!)^{1/4}] = o_P(1)$. Analogously, we can show that $D_{6NT} = O_P[N^{-1/2}(h!)^{-1/2} + (h!)^{1/2}] = o_P(1)$. This completes the proof of the theorem. ■

Proof of Theorem 3.4

The proof follows closely from that of Theorems 3.1 and 3.3. By (A.8), (A.12), and the fact that $\left|\hat{R}_{NT}\right|_\infty = o_P(1)$ under $\mathbb{H}_1$, we can readily show that

$$\Gamma_{NT} = b_{NT}^{-1}\Gamma_{NT1} + \left(2b_{NT}^{-1}\Gamma_{NT2} + b_{NT}^{-1}\Gamma_{NT3}\right) = b_{NT}^{-1}\sum_{l=1}^{6} D_{lNT} + o_P(1),$$

where recall $b_{NT} \equiv NT_1 (h!)^{1/2}$. It is easy to show that $b_{NT}^{-1} D_{lNT} = o_P(1)$ under $\mathbb{H}_1$ for $l = 1, 3, 4, 5, 6$. Under $\mathbb{H}_1$, by the WLLN for IID data, we have

$$b_{NT}^{-1} D_{2NT} = \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} \left[m(U_{i,t-1}) - \beta_0' U_{i,t-1} \right]^2 a_{it} = \mu_A + o_P(1).$$

In addition, under $\mathbb{H}_1$, we have $b_{NT}^{-1} \hat{B}_{NT} = o_P(1)$ and $\hat{\sigma}_{NT}^2 \xrightarrow{P} \bar{\sigma}^2 < \infty$. It follows that $b_{NT}^{-1} J_{NT} = [\Gamma_{NT} - b_{NT}^{-1} \hat{B}_{NT}] / \sqrt{\hat{\sigma}_{NT}^2} = \mu_A / \sqrt{\bar{\sigma}^2} + o_P(1)$, and the conclusion follows. ■

References

Ahn, H., 1997. Semiparametric estimation of a single-index model with nonparametrically generated regressors. Econometric Theory 13: 3-31.

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

Arellano, M., 1990. Testing for autocorrelation in dynamic random effects models. Reviews of Economic Studies 57: 124-134.

Arellano, M., 2003. Panel Data Econometrics. Oxford: Oxford University Press.

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.

Baltagi, B. H., 1999. Specification tests in panel data model using artificial regressions. Annales D'Economie et de Statistique 55-56: 277-297.

Baltagi, B. H., 2008. Econometric Analysis of Panel Data. 4th ed. West Sussex: John Wiley & Sons.

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., 1991. Economic growth in a cross section of countries. Quarterly Journal of Economics 106: 407-443.

Blundell, R., Chen, X., Kristensen, D., 2007. Semi-nonparametric IV estimation of shape-invariant Engel curves. Econometrica 75: 1613-1669.

Bond, S., 2002. Dynamic panel data models: a guide to micro data methods and practice. Working paper, Dept. of Economics, Institute for Fiscal Studies, London.

Bond, S., Leblebicioglu, A., Schiantarelli, F., 2010. Capital accumulation and growth: a new look at the empirical evidence. Journal of Applied Econometrics 25: 1073–1099.

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.

Carroll, C, Weil D., 1994. Saving and growth: a reinterpretation. Carnegie–Rochester Conference Series on Public Policy 40: 133–192.

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.

de Jong, P., 1987. A central limit theorem for generalized quadratic forms. Probability Theory and Related Fields 75: 261–277.

Darolles, S., Fan, Y., Florens, J., Renault, E., 2011. Nonparametric instrumental regression. Econometrica 79: 1541-1565.

Durlauf, S., 2000. Econometric analysis and the study of economic growth: a skeptical perspective. In R. Backhouse and A. Salanti (eds), Macroeconomics and the Real World, pp. 249-262, Oxford University Press, Oxford.

Fu, B., Li, W. K., Fung, W. K., 2002. Testing model adequacy for dynamic panel data with intercorrelation. Biometrika 89: 591-601.

Hall, P., 1984. Central limit theorem for integrated square error properties of multivariate nonparametric density estimators. Journal of Multivariate Analysis 14: 1-16.

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.

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46: 1251-1271.

Hausman, J., Taylor, W. E., 1981. Panel data and unobservable individual effects. Econometrica 49: 1377-1398.

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.

Hoderlein, S., Su, L., White, H., 2011. Specification testing for nonparametric structural models with monotonicity in unobservables. Working paper, Dept. of Economics, Boston College.

Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. Journal of the American Statistical Association 84: 1201-1220.

Hsiao, C. 2003. Analysis of Panel Data. 2nd ed. Cambridge: Cambridge University Press.

Inoue, A., Solon, G., 2006. A portmanteau test for serially correlated errors in fixed effects models. Econometric Theory 22, 835-851.

Jones, C., 1995. Time series tests of endogenous growth models. Quarterly Journal of Economics 110: 495–525.

Lee, A. J., 1990. U-statistics: theory and practice. Marcel Dekker, New York.

Lee, Y., 2010. Nonparametric estimation of dynamic panel models with fixed effects. Working paper, Dept. of Economics, Michigan University.

Lee, Y-J., 2011. Testing a linear dynamic panel data model against nonlinear alternatives. Working paper, Dept. of Economics, Indiana University.

Li, D., Lu, Z., Linton, O., 2011. Local linear fitting under near epoch dependence: uniform consistency with convergence rates. Forthcoming in Econometric Theory.

Li, Q., Stengos, T., 1992, A Hausman specification test based on root n consistent semiparametric estimators. Economics Letters 40: 141–146.

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

Li, Q., Sun, Y., 2011. A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. Working paper, Dept. of Economics, Texas A&M University.

Linton, O., Nielsen, J. P., 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82: 93-101.

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

Lucas, R., 1988. On the mechanics of economic development. Journal of Monetary Economics 22: 3-42.

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

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.

Metcalf, G. E., 1996. Specification testing in panel data with instrumental variables. Journal of Econometrics 71: 291-307.

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

Okui, R., 2009. Testing serial correlation in fixed effects regression models based on asymptotically unbiased autocorrelation estimators. Mathematics and Computers in Simulation 79: 2897-2909.

Opsomer, J. D., Ruppert, D., 1997. Fitting a bivariate additive model by local polynomial regression. Annals of Statistics 25: 186-211.

Qian, J., Wang, L., 2011. Estimating semiparametric panel data models by marginal integration. Forthcoming in Journal of Econometrics.

Romer, P., 1986. Increasing returns and long-run growth. Journal of Political Economy 94: 1002-1037.

Solow, R., 1956. A contribution to the theory of economic growth. Quarterly Journal of Economics 70: 65–94.

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.

Sun, Y., Carroll, R. J., Li, D., 2009. Semiparametric estimation of fixed effects panel data varying coefficient models. Advances in Econometrics 25: 101-130.

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.