Semiparametric Estimation of Fixed Effects Panel Data Varying Coefficient Models

Yiguo Sun
Department of Economics, University of Guelph
Guelph, ON, Canada N1G2W1

Raymond J. Carroll
Department of Statistics, Texas A&M University
College Station, TX 77843-3134, USA

Dingding Li
Department of Economics, University of Windsor
Windsor, ON, Canada N9B3P4

April 2, 2009

Abstract

We consider the problem of estimating a varying coefficient panel data model with fixed effects using a local linear regression approach. Unlike first-differenced estimator, our proposed estimator removes fixed effects using kernel-based weights. This results a one-step estimator without using back-fitting technique. The computed estimator is shown to be asymptotically normally distributed. A modified least-squared cross-validatory method is used to select the optimal bandwidth automatically. Moreover, we propose a test statistic for testing the null hypothesis of a random effects against a fixed effects varying coefficient panel data model. Monte Carlo simulations show that our proposed estimator and test statistic have satisfactory finite sample performance.

Key words: Consistent test; Fixed effects; Panel data; Varying coefficients model.

1 INTRODUCTION

Panel data traces information on each individual unit across time. Such two-dimensional information set enables researchers to estimate complex models and extract information and inferences which may not be possible using pure time-series data or cross-section data. With the increased availability of panel data, both theoretical and applied work in panel data analysis have become more popular in the recent years.

Arellano (2003), Baltigi (2005), and Hsiao (2003) provide excellent overview of parametric panel data model analysis. However, it is well known that a misspecified parametric panel data model may give misleading inferences. To avoid imposing the strong restrictions assumed in the parametric panel data models, econometricians and statisticians have worked on theories of nonparametric and semiparametric panel data regression models. For example, Henderson, Carroll, and Li (2008) considered the fixed-effects nonparametric panel data model. Henderson and Ullah (2005), Lin and Carroll (2000, 2001, 2006), Lin, Wang, Welsh and Carroll (2004), Lin and Ying (2001), Ruckstuhl, Welsh and Carroll (1999), Wang (2003), and Wu and Zhang (2002) considered the random-effects nonparametric panel data models. Li and Stengos (1996) considered a partially linear panel data model with some regressors being endogenous via IV approach, and Su and Ullah (2006) investigated a fixed-effects partially linear panel data model with exogenous regressors.

A purely nonparametric model suffers from the 'course of dimensionality' problem; while a partially linear semiparametric model may be too restrictive as it only allows for some additive nonlinearities. The varying coefficient model considered in this paper includes both pure nonparametric model and partially linear regression model as special cases. Moreover, we assume a fixed-effects panel data model. By fixed effects we mean that the individual effects are correlated with the regressors in an unknown way. Consistent with the well-known results in parametric panel data model estimation, we show that random effects estimators are inconsistent if the true model is one with fixed effects and that fixed effects estimators are consistent under both random and fixed effects panel data model, although the random effects estimator is more efficient than the fixed effects estimator when the random effects model holds true. Therefore, estimation of random effects models is appropriate only when individual effects are uncorrelated with regressors. As in practice, economists often view the assumptions required for the random effects model as being unsupported by the data, this paper emphasizes more on estimating a fixed effects panel data varying coefficient model, and we propose to use the local linear method to estimate unknown smooth coefficient functions. We also propose a test statistic for testing a random effects against a fixed effects varying coefficient panel data model. Simulation results show that our proposed estimator and test statistic have satisfactory finite sample performances.

Recently, Cai and Li (2008) studied a dynamic nonparametric panel data model with unknown varying coefficients. As Cai and Li (2008) allow the regressors not appearing in the varying coefficient curves to be endogenous, the GMM-based IV estimation method plus local linear regression approach is used to deliver consistent estimator of the unknown smooth coefficient curves. In this paper, all the regressors are assumed to be exogenous. Therefore, the least-squared method combining with local linear regression approach can be used to produce consistent estimator of the unknown smoothing coefficient curves. In addition, the asymptotic results are given when the time length is finite.

The rest of the paper is organized as follows. In section 2 we set up the model and discuss transformation methods that are used to remove fixed effects. Section 3 proposes a nonparametric fixed effects estimator and studies its asymptotic properties. In section 4 we suggest a statistic for testing the null hypothesis of a random effects against a fixed effects varying coefficient model. Section 5 reports simulation results that examine the finite sample performance of our semiparametric estimator and the test statistic. Finally we concludes the paper in section 6. The proofs of the main results are collected in an Appendix.

2 FIXED-EFFECTS VARYING COEFFICIENT PANEL DATA MODELS

We consider the following fixed-effects varying coefficient panel data regression model

$$Y_{it} = X_{it}^{\top}\theta(Z_{it}) + \mu_i + \nu_{it}, \qquad (i = 1, ..., n;\ t = 1, ..., m) \tag{1}$$

where the covariate $Z_{it} = (Z_{it,1}, ..., Z_{it,q})^{\top}$ is of dimension q, $X_{it} = (X_{it,1}, ..., X_{it,p})^{\top}$ is of dimension p, $\theta(\cdot) = \{\theta_1(\cdot), \cdots, \theta_p(\cdot)\}^{\top}$ contains p unknown functions, and all other variables are scalars. None of the variables in X_{it} can be obtained from Z_{it} and vice versa. The random errors v_{it} are assumed to be i.i.d. with a zero mean, finite variance $\sigma_v^2 > 0$ and independent of μ_j, Z_{js}, and X_{js} for all i, j,

s and t. The unobserved individual effects μ_i are assumed to be i.i.d. with a zero mean and a finite variance $\sigma_\mu^2 > 0$. We allow for μ_i to be correlated with Z_{it} and/or X_{it} with an unknown correlation structure. Hence, model (1) is a fixed-effects model. Alternatively, when μ_i is uncorrelated with Z_{it} and X_{it}, model (1) becomes a random-effects model.

A somewhat simplistic explanation for consideration of fixed effects models and the need for estimation of the function $\theta(\cdot)$ arises from considerations such as the following. Suppose that Y_{it} is the (logarithm) income of individual i at time period t, and X_{it} is education of individual i at time period t , e.g., number of years of schooling; and Z_{it} is the age of individual i at time t. The fixed effects term μ_i in (1) includes the individual's unobservable characteristics such as ability, e.g., IQ level, characteristics which are not observable for the data at hand. In this problem, economists are interested in the marginal effects of education on income, after controlling for the unobservable individual ability factors. Hence, they are interested in the marginal effects in the income change for an additional year of education regardless of whether the person has high or low ability. In this simple example, it is reasonable to believe that ability and education are positively correlated. If one does not control for the unobserved individual effects, then one would over-estimate the true marginal effects of education on income (i.e., with an upwards bias).

When $X_{it} \equiv 1$ for all i and t and $p = 1$, model (1) reduces to Henderson, Carroll, and Li's (2008) nonparametric panel data model with fixed effects as a special case. One may also interpret $X_{it}^\top \theta(Z_{it})$ as an interactive term between X_{it} and Z_{it} where we allow $\theta(Z_{it})$ to have a flexible format since the popularly used parametric setup such as Z_{it} and/or Z_{it}^2 may be misspecified.

For a given fixed effects model, there are many ways of removing the unknown fixed effects from the model.

The usual first-differenced (FD) estimation method deducts one equation from another to remove the time-invariant fixed effects. For example, deducting equation for time t from that for time $t-1$, we have for $t = 2, \cdots, m$

$$\widetilde{Y}_{it} = Y_{it} - Y_{it-1} = X_{it}^\top \theta(Z_{it}) - X_{it-1}^\top \theta(Z_{it-1}) + \widetilde{v}_{it}, \qquad \text{with } \widetilde{v}_{it} = v_{it} - v_{it-1}; \tag{2}$$

or deducting equation for time t from that for time 1, we obtain for $t = 2, \cdots, m$

$$\widetilde{Y}_{it} = Y_{it} - Y_{i1} = X_{it}^\top \theta(Z_{it}) - X_{i1}^\top \theta(Z_{i1}) + \widetilde{v}_{it}, \qquad \text{with } \widetilde{v}_{it} = v_{it} - v_{i1}. \tag{3}$$

The conventional fixed-effects (FE) estimation method, on the other hand, removes the fixed effects by deducting each equation from the cross-time average of the system, and it gives for $t = 2, \cdots, m$

$$\begin{aligned}\widetilde{Y}_{it} &= Y_{it} - \frac{1}{m}\sum_{s=1}^{m} Y_{is} = X_{it}^{\top}\theta(Z_{it}) - \frac{1}{m}\sum_{s=1}^{m} X_{is}^{\top}\theta(Z_{is}) + \widetilde{v}_{it} \\ &= \sum_{s=1}^{m} q_{ts} X_{is}^{\top}\theta(Z_{is}) + \widetilde{v}_{it} \qquad \text{with } \tilde{v}_{it} = v_{it} - \tfrac{1}{m}\textstyle\sum_{s=1}^{m} v_{is} \end{aligned} \tag{4}$$

where $q_{ts} = -1/m$ if $s \neq t$ and $1 - 1/m$ otherwise, and $\sum_{s=1}^{m} q_{ts} = 0$ for all t.

Many nonparametric local smoothing methods can be used to estimate the unknown function $\theta(\cdot)$. However, for each i, the right-hand sides of equations (2)-(4) contain linear combination of $X_{it}^{\top}\theta(Z_{it})$ for different time t. If X_{it} contains a time-invariant term, say the first component of X_{it}, and let $\theta_1(Z_{it})$ denote the first component of $\theta(Z_{it})$, then a first difference of $X_{it,1}\theta_1(Z_{it}) \equiv X_{i,1}\theta_1(Z_{it})$ gives $X_{i,1}\left(\theta_1(Z_{it}) - \theta_1(Z_{i,t-1})\right)$, which is an additive function with the same function form for the two functions but evaluated at different observation points. Kernel based estimator usually requires some backfitting algorithms to recover the unknown function, which will suffer the common problems as indicated in estimating nonparametric additive model. Moreover, if $\theta_1(Z_{it})$ contains an additive constant term, say $\theta(Z_{it}) = c + g_1(Z_{it})$, where c is a constant, then the first difference will wipe out the additive constant c. As a consequence, one cannot consistently estimate $\theta_1(\cdot)$ one were to estimate a first-differenced model in general (if $X_{i,1} \equiv 1$, one can recover c by averaging $Y_{it} - X_{it}^{\top}\hat{\theta}(Z_{it})$ for all cross sections and across time).

Therefore, in this paper we consider an alternative way of removing the unknown fixed effects, motivated by a least squares dummy variable (LSDV) model in parametric panel data analysis. We will describe how the proposed method removes fixed effects by deducting a smoothed version of cross-time average from each individual unit. As we will show later, this transformation method will not wipe the additive constant c in $\theta_1(Z_{it}) = c + g_1(Z_{it})$. Therefore, we can consistently estimate $\theta_1(\cdot)$ as well as other components of $\theta(\cdot)$ when at most one of the variables in X_{it} is time invariant.

We will use I_n to denote an identity matrix of dimension n, and e_m to denote an $m \times 1$ vector with all elements being ones. Rewriting model (1) in a matrix format yields

$$Y = B\{X, \theta(Z)\} + D_0\mu_0 + V, \tag{5}$$

where $Y = (Y_1^\top, \cdots, Y_n^\top)^\top$ and $V = (v_1^\top, \cdots, v_n^\top)^\top$ are $(nm) \times 1$ vectors; $Y_i^\top = (Y_{i1}, ..., Y_{in})$ and $v_i^\top = (v_{i1}, ..., v_{in})$. $B\{X, \theta(Z)\}$ stacks all $X_{it}^\top \theta(Z_{it})$ into an $(nm) \times 1$ vector with the (i, t) subscript matching that of the $(nm) \times 1$ vector of Y; $\mu_0 = (\mu_1, \cdots, \mu_n)^\top$ is an $n \times 1$ vector, and $D_0 = I_n \otimes e_m$ is an $(nm) \times n$ matrix with main diagonal blocks being e_m, where $\otimes$ refers to Kronecker product operation. However, we can not estimate model (5) directly due to the existence of the fixed effects term. Therefore, we need some identification conditions. Su and Ullah (2006) assume $\sum_{i=1}^n \mu_i = 0$. We show that assuming an i.i.d sequence of unknown fixed effects, μ_i, with zero mean and a finite variance is enough to identify the unknown coefficient curves asymptotically. We therefore impose this weaker version of identification condition in this paper.

To introduce our estimator, we first assume that model (1) holds with the restriction $\sum_{i=1}^n \mu_i = 0$ (note that we do not impose this restriction for our estimator, and this restriction is added here for motivating our estimator). Define $\mu = (\mu_2, \cdots, \mu_n)^\top$. We then rewrite (5) as

$$Y = B\{X, \theta(Z)\} + D\mu + V, \tag{6}$$

where $D = [-e_{n-1} \ \ I_{n-1}]^\top \otimes e_m$ is an $(nm) \times (n-1)$ matrix. Note that $D\mu = \mu_0 \otimes e_m$ with $\mu_0 = (-\sum_{i=2}^n \mu_i, \mu_2, ..., \mu_n)^\top$ so that the restriction $\sum_{i=1}^n \mu_i = 0$ is imposed in (6).

Define an $m \times m$ diagonal matrix $K_H(Z_i, z) = \text{diag}\{K_H(Z_{i1}, z), \cdots, K_H(Z_{im}, z)\}$ for each i, and a $(nm) \times (nm)$ diagonal matrix $W_H(z) = \text{diag}\{K_H(Z_1, z), \cdots, K_H(Z_n, z)\}$, where $K_H(Z_{it}, z) = K\{H^{-1}(Z_{it} - z)\}$ for all i and t, and $H = \text{diag}(h_1, \cdots, h_q)$ is a $q \times q$ diagonal bandwidth matrix. We then solve the following optimization problem

$$\min_{\theta(Z), \mu} [Y - B\{X, \theta(Z)\} - D\mu]^\top W_H(z) [Y - B\{X, \theta(Z)\} - D\mu], \tag{7}$$

where we use the local weight matrix $W_H(z)$ to ensure locality of our nonparametric fitting, and place no weight matrix for data variation since the $\{v_{it}\}$ are i.i.d. across equations. Taking first-order condition with respect to μ gives

$$D^\top W_H(z)[Y - B\{X, \theta(Z)\} - D\widehat{\mu}(z)] = 0, \tag{8}$$

which yields

$$\widehat{\mu}(z) = \{D^\top W_H(z) D\}^{-1} D^\top W_H(z) [Y - B\{X, \theta(Z)\}]. \tag{9}$$

Define $S_H(z) = M_H(z)^\top W_H(z) M_H(z)$ and $M_H(z) = I_{n\times m} - D\{D^\top W_H(z)D\}^{-1}D^\top W_H(z)$, where $I_{n\times m}$ denotes an identity matrix of dimension nm by nm. Replacing μ in (7) by $\widehat{\mu}(z)$, we obtain the concentrated weighted least squares

$$\min_{\theta(Z)}[Y - B\{X, \theta(Z)\}]^\top S_H(z)[Y - B\{X, \theta(Z)\}], \tag{10}$$

Note that $M_H(z)D\mu \equiv 0_{(nm)\times 1}$ for all z. Hence, the fixed effects term μ is removed in model (10).

To see how $M_H(z)$ transforms the data, simple calculations give

$$M_H(z) = I_{n\times m} - D\{A^{-1} - A^{-1}e_{n-1}e_{n-1}^\top A^{-1}/\sum_{i=1}^{n} c_H(Z_i, z)\}D^\top W_H(z),$$

where $c_H(Z_i, z)^{-1} = \sum_{t=1}^{m} K_H(Z_{it}, z)$ for $i = 1, \cdots, n$ and $A = \text{diag}\{c_H(Z_2, z)^{-1}, \ldots\cdots, c_H(Z_n, z)^{-1}\}$. We use the formula $(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}$ to derive the inverse matrix, see Appendix B in Poirier (1995).

3 NONPARAMETRIC ESTIMATOR AND ASYMPTOTIC THEORY

A local linear regression approach is commonly used to estimate non-/semi-parametric models. The basic idea of this method is to apply Taylor expansion up to the second-order derivative. Throughout the paper we will use the notation that $A_n \approx B_n$ to denote that B_n is the leading term of A_n, i.e., $A_n = B_n + (s.o.)$, where $(s.o.)$ denotes terms having probability order smaller than that of B_n. For each $l = 1, \cdots, p$, we have the following Taylor expansion around z:

$$\theta_l(z_{it}) \approx \theta_l(z) + \{H\theta_l'(z)\}^\top [H^{-1}(z_{it} - z)] + \frac{1}{2} r_{H,l}(z_{it}, z), \tag{11}$$

where $\theta_l'(z) = \partial\theta_l(z)/\partial z$ is the $q\times 1$ vector of the first order derive function, $r_{H,l}(z_{it}, z) = \{H^{-1}(z_{it} - z)\}^\top\{H \frac{\partial^2\theta_l(z)}{\partial z \partial z^\top} H\}\{H^{-1}(z_{it} - z)\}$. Of course, $\theta_l(z)$ approximates $\theta_l(z_{it})$ and $\theta_l'(z)$ approximates $\theta_l'(z_{it})$ when z_{it} is close to z. Define $\beta_l(z) = \{\theta_l(z), [H\theta_l'(z)]^\top\}^\top$, a $(q + 1) \times 1$ column vector for $l = 1, 2, \cdots, p$, and $\beta(z) = \{\beta_1(z), \cdots, \beta_p(z)\}^\top$, a $p \times (q + 1)$ parameter matrix. The first column of $\beta(z)$ is $\theta(z)$. Therefore, we will replace $\theta(Z_{it})$ in (1) by $\beta(z)G_{it}(z, H)$ for each i and t, where $G_{it}(z, H) = [1, \{H^{-1}(Z_{it} - z)\}^\top]^\top$ is a $(q + 1) \times 1$ vector.

To make matrix operations simpler, we stack the matrix $\beta(z)$ into a $p(q+1)\times 1$ column vector and denote it by $\text{vec}\{\beta(z)\}$. Since $\text{vec}(ABC) = (C^\top \otimes A)\text{vec}(B)$ and $(A\otimes B)^\top = A^\top \otimes B^\top$, where

$\otimes$ refers to Kronecker product, we have $X_{it}^{\top}\beta(z)G_{it}(z,H) = \{G_{it}(z,H)\otimes X_{it}\}^{\top}\text{vec}\{\beta(z)\}$ for all i and t. Thus, we consider the following minimization problem

$$\min_{\beta(z)}[Y - R(z,H)\text{vec}\{\beta(z)\}]^{\top}S_H(z)[Y - R(z,H)\text{vec}\{\beta(z)\}] \tag{12}$$

where

$$\begin{aligned} R_i(z,H) &= \begin{bmatrix} (G_{i,1}(z,H)\otimes X_{i1})^{\top} \\ \vdots \\ (G_{i,m}(z,H)\otimes X_{im})^{\top} \end{bmatrix} \text{ is an } m\times[p(q+1)] \text{ matrix,} \\ R(z,H) &= [R_1(z,H)^{\top},\cdots,R_n(z,H)^{\top}]^{\top} \text{ is an } (nm)\times[p(q+1)] \text{ matrix.} \end{aligned}$$

Simple calculations give

$$\begin{aligned} \text{vec}\{\widehat{\beta}(z)\} &= \{R(z,H)^{\top}S_H(z)R(z,H)\}^{-1}R(z,H)^{\top}S_H(z)Y \\ &= \text{vec}\{\beta(z)\} + \{R(z,H)^{\top}S_H(z)R(z,H)\}^{-1}(A_n/2 + B_n + C_n), \end{aligned} \tag{13}$$

where $A_n = R(z,H)^{\top}S_H(z)\Pi(z,H)$, $B_n = R(z,H)^{\top}S_H(z)D_0\mu_0$, and $C_n = R(z,H)^{\top}S_H(z)V$. The $\{t+(i-1)m\}^{\text{th}}$ element of the column vector $\Pi(z,H)$ is $X_{it}^{\top}r_H(\widetilde{Z}_{it},z)$, where $r_H(\cdot,\cdot) = \{r_{H,1}(\cdot,\cdot),\cdots,r_{H,p}(\cdot,\cdot)\}^{\top}$ and $r_{H,l}(\widetilde{Z}_{it},z) = \{H^{-1}(Z_{it}-z)\}^{\top}\{H\frac{\partial^2\theta_l(\widetilde{Z}_{it})}{\partial z\partial z^{\top}}H\}\{H^{-1}(Z_{it}-z)\}$ with $\widetilde{Z}_{it}$ lying between Z_{it} and z for each i and t. Both A_n and B_n contribute to the bias term of the estimator. Also, if $\sum_{i=1}^{n}\mu_i = 0$ holds true, $B_n = 0$; if we only assume μ_i being iid with zero mean and finite variance, the bias due to the existence of unknown fixed effects can be asymptotically ignored.

To derive the asymptotic distribution of $\text{vec}\{\widehat{\beta}(z)\}$, we first give some regularity conditions. Throughout this paper, we use $M > 0$ to denote a finite constant, which may take a different value at different places.

Assumption 1: The random variables (X_{it}, Z_{it}) are independently and identically distributed (i.i.d.) across the i index, and

(a) $E\|X_{it}\|^{2(1+\delta)} \le M < \infty$ and $E\|Z_{it}\|^{2(1+\delta)} \le M < \infty$ hold for some $\delta > 0$ and for all i and t.

(b) The Z_{it} are continuous random variables with a p.d.f. $f_t(z)$. Also, for each $z \in R^q$, $f(z) = \sum_{t=1}^{m} f_t(z) > 0$.

(c) Denote $\lambda_{it} = K_H(Z_{it},z)$ and $\varpi_{it} = \lambda_{it}/\sum_{t=1}^{m}\lambda_{it} \in (0,1)$ for all i and t. $\Psi(z) = |H|^{-1}\sum_{t=1}^{m} E\left[(1-\varpi_{it})\lambda_{it}X_{it}X_{it}^{T}\right]$ is a nonsingular matrix.

(d) Let $f_t(z|X_{it})$ be the conditional pdf of Z_{it} at $Z_{it} = z$ conditional on X_{it} and $f_{t,s}(z_1, z_2|X_{it}, X_{js})$ be the joint conditional pdf of (Z_{it}, Z_{js}) at $(Z_{it}, Z_{js}) = (z_1, z_2)$ conditional on (X_{it}, X_{js}) for $t \neq s$ and any i and j. Also, $\theta(z)$, $f_t(z)$, $f_t(\cdot|X_{it})$, $f_{t,s}(\cdot, \cdot|X_{it}, X_{js})$ are uniformly bounded in the domain of Z and are all twice continuously differentiable at $z \in R^q$ for all $t \neq s$, i and j.

Assumption 2: Both X and Z have full column rank; $\{X_{it,1}, ..., X_{it,p}, \{X_{it,l}Z_{it,j} : l = 1, ..., p, j = 1, ..., q\}\}$ are linearly independent. If $X_{it,l} \equiv X_{i,l}$ for at most one $l \in \{1, \cdots, p\}$, i.e., $X_{i,l}$ does not depend on t, we assume $E(X_{i,l}) \neq 0$. The unobserved fixed effects μ_i are i.i.d. with zero mean and a finite variance $\sigma_\mu^2 > 0$. The random errors v_{it} are assumed to be i.i.d. with a zero mean, finite variance σ_v^2 and independent of Z_{it} and X_{it} for all i and t. Y_{it} is generated by equation (1).

If X_{it} contains a time invariant regressor, say the l^{th} component of X_{it} is $X_{it,l} = W_i$. Then the corresponding coefficient $\theta_l(\cdot)$ is estimable if $M_H(z)(W \otimes e_m) \neq 0$ for a given z, where $W = (W_1, ..., W_n)^\top$. Simple calculations give $M_H(z)(W \otimes e_m) = (n^{-1} \sum_{i=1}^n W_i) M_H(z) \times (e_n \otimes e_m)$. The proof of Lemma A.2 in Appendix 7.1 can be used to show that $M_H(z)(e_n \otimes e_m) \neq 0$ for any given z with probability one. Therefore, $\theta_l(\cdot)$ is asymptotically identifiable if $n^{-1} \sum_{i=1}^n X_{it,l} \equiv n^{-1} \sum_{i=1}^n W_i \nrightarrow 0$ while $\bar{\mu} \overset{a.s.}{\to} 0$. For example, if X_{it} contains a constant, say, $X_{it,1} = W_i \equiv 1$, then $\theta_1(\cdot)$ is estimable because $n^{-1} \sum_{i=1}^n W_i = 1 \neq 0$.

Assumption 3: $K(v) = \prod_{s=1}^q k(v_s)$ is a product kernel, and the univariate kernel function $k(\cdot)$ is a uniformly bounded, symmetric (around zero) probability density function with a compact support $[-1, 1]$. In addition, define $|H| = h_1 \cdots h_q$ and $\|H\| = \sqrt{\sum_{j=1}^q h_j^2}$. As $n \to \infty$, $\|H\| \to 0$, $n|H| \to \infty$.

The assumptions listed above are regularity assumptions commonly seen in nonparametric estimation literature. Assumption 1 apparently excludes the case of either X_{it} or Z_{it} being I(1); other than the moment restrictions, we do not impose I(0) structure on X_{it} across time, since this paper considers the case that m is a small finite number. Also, instead of imposing the smoothness assumption on $f_t(\cdot|X_{it})$ and $f_{t,s}(\cdot, \cdot|X_{it}, X_{is})$ as in Assumption 1(d), we can assume that $f_t(z) E\left(X_{it}X_{it}^T|z\right)$ and $f_{t,s}(z_1, z_2) E\left(X_{it}X_{js}^T|z_1, z_2\right)$ are uniformly bounded in the domain of Z and are all twice continuously differentiable at $z \in R^q$ for all $t \neq s$ and i and j. Our version of the smoothness assumption simplifies our notation in the proofs.

Assumption 2 indicates that X_{it} can contain a constant term of ones. The kernel function having a compact support in Assumption 3 is imposed for the sake of brevity of proof and can be removed at the cost of lengthy proofs. Specifically, the Gaussian kernel is allowed.

We use $\widehat{\theta}(z)$ to denote the first column of $\widehat{\beta}(z)$. Then $\hat{\theta}(z)$ estimates $\theta(z)$.

THEOREM 3.1 *Under Assumptions 1-3, we obtain the following bias and variance for $\widehat{\theta}(z)$, given a finite integer $m > 0$:*

$$\begin{aligned} bias(\widehat{\theta}(z)) &= \Psi(z)^{-1}\Lambda(z)/2 + O\left(n^{-1/2}|H|\ln(\ln n)\right) + o(\|H\|^2), \\ var(\widehat{\theta}(z)) &= n^{-1}|H|^{-1}\sigma_v^2\Psi(z)^{-1}\Gamma(z)\Psi(z)^{-1} + o(n^{-1}|H|^{-1}), \end{aligned}$$

where $\Psi(z) = |H|^{-1}\sum_{t=1}^m E\left[(1-\varpi_{it})\lambda_{it}X_{it}X_{it}^T\right]$, $\Lambda(z) = |H|^{-1}\sum_{t=1}^m E\left[(1-\varpi_{it})\lambda_{it}X_{it}X_{it}^T r_H\left(\tilde{Z}_{it}, z\right)\right]$ $= O\left(\|H\|^2\right)$, *and* $\Gamma(z) = |H|^{-1}\sum_{t=1}^m E\left[(1-\varpi_{it})^2\lambda_{it}^2X_{it}X_{it}^T\right]$.

The first term of bias($\widehat{\theta}(z)$) results from the local approximation of $\theta(z)$ by a linear function of z, which is of order $O\left(\|H\|^2\right)$ as usual. The second term of bias($\widehat{\theta}(z)$) results from the unknown fixed effects μ_i: (a) if we assumed $\sum_{i=1}^n \mu_i = 0$, this term is zero exactly; (b) the result indicates that the bias term is dominated by the first term and will vanish as $n \to \infty$.

In Appendix, we show that

$$\begin{aligned} |H|^{-1}\sum_{t=1}^m E\left(\lambda_{it}X_{it}X_{it}^T\right) &= \Phi(z) + o(\|H\|^2), \\ |H|^{-1}\sum_{t=1}^m E\left[\lambda_{it}X_{it}X_{it}^T r_H\left(\tilde{Z}_{it}, z\right)\right] &= \kappa_2\Phi(z)\Theta_H(z) + o(\|H\|^2), \\ |H|^{-1}\sum_{t=1}^m E\left(\lambda_{it}^2X_{it}X_{it}^T\right) &= \left(\int K^2(u)\,du\right)\Phi(z) + o(\|H\|^2), \end{aligned}$$

where $\kappa_2 = \int k(v)v^2dv$, $\Phi(z) = \sum_{t=1}^m f_t(z)E\left(X_{1t}X_{1t}^T|z\right)$, and $\Theta_H(z) = \left[tr\left(H\frac{\partial^2\theta_1(z)}{\partial z\partial z^T}H\right), \cdots, tr\left(H\frac{\partial^2\theta_p(z)}{\partial z\partial z^T}H\right)\right]^T$. Since $\varpi_{it} \in [0,1)$ for all i and t, the results above imply the existence of $\Psi(z)$, $\Lambda(z)$, and $\Gamma(z)$. However, given a finite integer $m > 0$, we can not obtain explicitly the asymptotic bias and variance due to the random denominator appearing in ϖ_{it}.

Further, the following Theorem gives the asymptotic normality results for $\widehat{\theta}(z)$.

THEOREM 3.2 *Under Assumptions 1-3, and assuming in addition that* $E|v_{it}|^{2+\delta} < \infty$ *for some* $\delta > 0$*, and that* $\sqrt{n|H|}\|H\|^2 = O(1)$ *as* $\to \infty$*, we have*

$$\sqrt{n|H|}\{\widehat{\theta}(z) - \theta(z) - \Psi(z)^{-1}\Lambda(z)/2\} \xrightarrow{d} N(0, \Sigma_{\theta(z)}),$$

where $\Sigma_{\theta(z)} = \sigma_v^2 \lim_{n\to\infty} \Psi(z)^{-1}\Gamma(z)\Psi(z)^{-1}$*. Moreover, a consistent estimator for* $\Sigma_{\theta(z)}$ *is given as follows:*

$$\begin{aligned}
\widehat{\Sigma}_{\theta(z)} &= S_p\widehat{\Omega}(z,H)^{-1}\widehat{J}(z,H)\widehat{\Omega}(z,H)^{-1}S_p^\top \xrightarrow{p} \Sigma_{\theta(z)}, \\
\widehat{\Omega}(z,H) &= n^{-1}|H|^{-1}R(z,H)^\top S_H(z)R(z,H) \\
\widehat{J}(z,H) &= n^{-1}|H|^{-1}R(z,H)^\top S_H(z)\widehat{V}\widehat{V}^\top S_H(z)R(z,H)
\end{aligned}$$

where $\widehat{V}$ *is the vector of estimated residuals and* S_p *includes the first* p *rows of the* $p(q+1)$ *identify matrix. Finally, a consistent estimator for the leading bias can be easily obtained based on a nonparametric local quadratic regression result.*

4 TESTING RANDOM EFFECTS VERSUS FIXED EFFECTS

In this section we discuss how to test for the presence of random effects versus fixed effects in a semiparametric varying coefficient panel data model. The model remains as (1). The random effects specification assumes that μ_i is uncorrelated with the regressors X_{it} and Z_{it}, while for the fixed effects case, μ_i is allowed to be correlated with X_{it} and/or Z_{it} in an unknown way.

We are interested in testing the null hypothesis (H_0) that μ_i is a random effect versus the alternative hypothesis (H_1) that μ_i is a fixed effect. The null and alternative hypotheses can be written as

$$H_0 \ : \ \Pr\{E(\mu_i|Z_{i1},...,Z_{im},X_{i1},\cdots,X_{im}) = 0\} = 1 \text{ for all } i, \tag{14}$$

$$H_1 \ : \ \Pr\{E(\mu_i|Z_{i1},...,Z_{im},X_{i1},\cdots,X_{im}) \neq 0\} > 0 \text{ for some } i\ , \tag{15}$$

while we keep the same setup given in model (1) under both H_0 and H_1.

Our test statistic is based on the squared difference between the FE and RE estimators, which is asymptotically zero under H_0 and positive under H_1. To simplify the proofs and save computing time, we use local constant estimator instead of local linear estimator for constructing our test.

Then following the argument in Section 2 and Appendix 7.2, we have

$$\begin{aligned}\widehat{\theta}_{FE}(z) &= \{X^\top S_H(z)X\}^{-1}X^\top S_H(z)Y\\ \widehat{\theta}_{RE}(z) &= \{X^\top W_H(z)X\}^{-1}X^\top W_H(z)Y\end{aligned}$$

where X is a $(nm)\times p$ matrix with $X=(X_1^\top,\cdots,X_n^\top)$, and for each i, $X_i=(X_{i1},\cdots,X_{im})^\top$ is an $m\times p$ matrix with $X_{it}=[X_{it,1},\cdots,X_{it,p}]^\top$. Motivated by Li, et al. (2002), we remove the random denominator of $\widehat{\theta}_{FE}(z)$ by multiplying $X^\top S_H(z)X$ and our test statistic will be based on

$$\begin{aligned}T_n &= \int\{\widehat{\theta}_{FE}(z)-\widehat{\theta}_{RE}(z)\}^\top\{X^\top S_H(z)X\}^\top\{X^\top S_H(z)X\}\{\widehat{\theta}_{FE}(z)-\widehat{\theta}_{RE}(z)\}dz\\ &= \int \widetilde{U}(z)^\top S_H(z)XX^\top S_H(z)\widetilde{U}(z)dz\end{aligned}$$

since $\{X^\top S_H(z)X\}\{\widehat{\theta}_{FE}(z)-\widehat{\theta}_{RE}(z)\}=X^\top S_H(z)\{Y-X\widehat{\theta}_{RE}(z)\}\equiv X^\top S_H(z)\widetilde{U}(z)$. To simplify the statistic, we make several changes in T_n. Firstly, we simplify the integration calculation by replacing $\widetilde{U}(z)$ by $\widehat{U}$, where $\widehat{U}\equiv\widehat{U}(Z)=Y-B\{X,\widehat{\theta}_{RE}(Z)\}$ and $B\{X,\widehat{\theta}_{RE}(Z)\}$ stacks up $X_{it}^\top\widehat{\theta}_{RE}(Z_{it})$ in the increasing order of i first then of t. Secondly, to overcome the complexity caused by the random denominator in $M_H(z)$, we replace $M_H(z)$ by $M_D=I_{n\times m}-m^{-1}I_n\otimes(e_me_m^\top)$ such that the fixed effects can be removed due to the fact that $M_DD_0=0$. With the above modification and also removing the $i=j$ terms in T_n (since T_n contains two summations $\sum_i\sum_j$), our further modified test statistic is given by

$$\widetilde{T}_n\overset{def}{=}\sum_{i=1}^n\sum_{j\neq i}\widehat{U}_i^\top Q_m\int K_H(Z_i,z)X_i^\top X_jK_H(Z_j,z)dzQ_m\widehat{U}_j,$$

where $Q_m=I_m-m^{-1}e_me_m^\top$. If $|H|\to0$ as $n\to\infty$, we obtain

$$\begin{aligned}&|H|^{-1}\int K_H(Z_i,z)X_i^\top X_jK_H(Z_j,z)dz \qquad (16)\\ =&\begin{bmatrix}\bar{K}_H(Z_{i,1},Z_{j,1})X_{i,1}^\top X_{j,1} & \cdots & \bar{K}_H(Z_{i,1},Z_{j,m})X_{i,1}^\top X_{j,m}\\ \vdots & \ddots & \vdots\\ \bar{K}_H(Z_{i,m},Z_{j,1})X_{i,m}^\top X_{j,1} & \cdots & \bar{K}_H(Z_{i,m},Z_{j,m})X_{i,m}^\top X_{j,m}\end{bmatrix},\end{aligned}$$

where $\bar{K}_H(Z_{it},Z_{js})=\int K\{H^{-1}(Z_{it}-Z_{js})+\omega\}K(\omega)d\omega$. We then replace $\bar{K}_H(Z_{it},Z_{js})$ by $K_H(Z_{it},Z_{js})$; this replacement will not affect the essence of the test statistic since the local weight is untouched. Now, our proposed test statistic is given by

$$\widehat{T}_n=\frac{1}{n^2|H|}\sum_{i=1}^n\sum_{j\neq i}^n\widehat{U}_i^\top Q_mA_{i,j}Q_m\widehat{U}_j \qquad (17)$$

where $A_{i,j}$ equals to the right-hand side of equation (16) after replacing $\bar{K}_H(Z_{it}, Z_{js})$ by $K_H(Z_{it}, Z_{js})$. Finally, to remove the asymptotic bias term of the proposed test statistic, we calculate the leave-one-unit-out random-effects estimator of $\theta(Z_{it})$; that is, for a given pair of (i, j) in the double summation of (17) with $i \neq j$, $\widehat{\theta}_{RE}(Z_{it})$ is calculated without using the observations on the j^{th}-unit, $\{(X_{jt}, Z_{jt}, Y_{jt})\}_{t=1}^{m}$ and $\widehat{\theta}_{RE}(Z_{jt})$ is calculated without using the observations on the i^{th}-unit.

We present the asymptotic properties of this test below and delay the proofs to Appendix 7.3.

THEOREM 4.1 *Under Assumptions 1-3, and $f_t(z)$ has a compact support $\mathcal{S}$ for all t, and $n\sqrt{|H|}\,\|H\|^4 \to 0$ as $n \to \infty$, then we have under H_0 that*

$$J_n = n\sqrt{|H|}\widehat{T}_n/\widehat{\sigma}_0 \xrightarrow{d} N(0,1) \tag{18}$$

where $\widehat{\sigma}_0^2 = \frac{2}{n^2|H|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\widehat{V}_i^{\top} Q_m A_{i,j} Q_m \widehat{V}_j)^2$ is a consistent estimator of

$$\sigma_0^2 = 4(1-1/m)^2\sigma_v^4 \int K^2(u)du \sum_{t=2}^{m}\sum_{s=1}^{t-1} E\left[f_t(Z_{1s})(X_{1s}^{\top}X_{2t})^2\right],$$

where $\widehat{V}_{it} = Y_{it} - X_{it}^{\top}\widehat{\theta}_{FE}(Z_{it}) - \widehat{\mu}_i$ and for each pair of (i, j), $i \neq j$, $\widehat{\theta}_{FE}(Z_{it})$ is a leave-two-unit-out FE estimator without using the observations from the i^{th} and j^{th} units and $\widehat{\mu}_i = \bar{Y}_i - m^{-1}\sum_{t=1}^{m} X_{it}^{\top}\widehat{\theta}_{FE}(Z_{it})$. Under H_1, $\Pr[J_n > B_n] \to 1$ as $n \to \infty$, where B_n is any nonstochastic sequence with $B_n = o(n\sqrt{|H|})$.

Assuming that $f_t(z)$ has a compact support $\mathcal{S}$ for all t is to simplify the proof of $\sup_{z\in\mathcal{S}} ||\widehat{\theta}_{RE}(z) - \theta(z)|| = o_p(1)$ as $n \to \infty$; otherwise, some trimming procedure has to be placed to show the uniform convergence result and the consistency of $\widehat{\sigma}_0^2$ as an estimator of σ_0^2. Theorem 4.1 states that the test statistic $J_n = n\sqrt{|H|}\widehat{T}_n/\widehat{\sigma}_0$ is a consistent test for testing H_0 against H_1. It is a one-sided test. If J_n is greater than the critical values from the standard normal distribution, we reject the null hypothesis at the corresponding significance levels.

5 MONTE CARLO SIMULATIONS

In this section we report some Monte Carlo simulation results to examine the finite sample performance of the proposed estimator. The following data generating process is used:

$$Y_{it} = \theta_1(Z_{it}) + \theta_2(Z_{it})X_{it} + \mu_i + v_{it}, \tag{19}$$

where $\theta_1(z) = 1 + z + z^2$, $\theta_2(z) = \sin(z\pi)$, $Z_{it} = w_{it} + w_{i,t-1}$, w_{it} is i.i.d. uniformly distributed in $[0, \pi/2]$, $X_{it} = 0.5X_{i,t-1} + \xi_{it}$, ξ_{it} is i.i.d. $\mathrm{N}(0,1)$. In addition, $\mu_i = c_0\bar{Z}_{i\cdot} + u_i$ for $i = 2, \cdots, n$ with $c_0 = 0$, 0.5, and 1.0, u_i is i.i.d. $\mathrm{N}(0,1)$. When $c_0 \neq 0$, μ_i and Z_{it} are correlated; we use c_0 to control the correlation between μ_i and $\bar{Z}_{i\cdot} = m^{-1}\sum_{t=1}^{m} Z_{it}$. Moreover, v_{it} is i.i.d. $\mathrm{N}(0,1)$, w_{it}, ξ_{it}, u_i and v_{it} are independent of each other.

We report estimation results for both the proposed fixed-effects (FE) estimator and the random-effects (RE) estimator, see Appendix 7.2 for the asymptotic results of the RE estimator. To learn how the two estimators perform when we have fixed-effects model and when we have random-effects model, we use the integrated squared error as a standard measure of estimation accuracy:

$$ISE(\widehat{\theta}_l) = \int \{\widehat{\theta}_l(z) - \theta_l(z)\}^2 f(z)dz, \tag{20}$$

which can be approximated by the average mean squared error

$$AMSE(\widehat{\theta}_l) = (nm)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}[\widehat{\theta}_l(Z_{it}) - \theta_l(Z_{it})]^2$$

for $l = 1, 2$. In Table 1 we present the average value of $AMSE(\widehat{\theta}_l)$ from 1000 Monte Carlo experiments. We choose $m = 3$ and $n = 50$, 100, and 200.

Since the bias and variance of the proposed FE estimator do not depend on the values of the fixed effects, our estimates are the same for different values of c_0; however, it is not true under the random-effects model. Therefore, the results derived from the FE estimator are only reported once in Table 1 since it is invariant to different values of c_0.

It is well-known that the performance of non/semiparametric models depends on the choice of bandwidth. Therefore, we propose a leave-**one-unit**-out cross validation method to automatically select the optimal bandwidth for estimating both the FE and RE models. Specifically, when estimating $\theta(\cdot)$ at a point Z_{it}, we remove $\{(X_{it}, Y_{it}, Z_{it})\}_{t=1}^{m}$ from the data and only use the rest of $(n-1)m$ observations to calculate $\widehat{\theta}_{(-i)}(Z_{it})$. In computing the RE estimate, the leave-one-unit-out cross validation method is just a trivial extension of the conventional leave-one-out cross validation method. The conventional leave-one-out method fails to provide satisfying result due to the existence of unknown fixed effects. Therefore, when calculating the FE estimator, we use the following modified leave-one-unit-out cross validation method:

$$\widehat{H}_{opt} = \arg\min_{H} [Y - B\{X, \widehat{\theta}_{(-1)}(Z)\}]^\top M_D^\top M_D [Y - B\{X, \widehat{\theta}_{(-1)}(Z)\}], \tag{21}$$

where $M_D = I_{n\times m} - m^{-1} I_n \otimes (e_m e_m^\top)$ satisfies $M_D D_0 = 0$; this is used to remove the unknown fixed effects. In addition, $B\{X, \widehat{\theta}_{(-1)}(Z)\}$ stacks up $X_{it}^\top \widehat{\theta}_{(-i)}(Z_{it})$ in the increasing order of i first then of t. Simple calculations give

$$\begin{aligned} & [Y - B\{X, \widehat{\theta}_{(-1)}(Z)\}]^\top M_D^\top M_D [Y - B\{X, \widehat{\theta}_{(-1)}(Z)\}] \\ = \; & [B\{X, \theta(Z)\} - B\{X, \widehat{\theta}_{(-1)}(Z)\}]^\top M_D^\top M_D [B\{X, \theta(Z)\} - B\{X, \widehat{\theta}_{(-1)}(Z)\}] \\ & + 2[B\{X, \theta(Z)\} - B\{X, \widehat{\theta}_{(-1)}(Z)\}]^\top M_D^\top M_D V + V^\top M_D M_D V, \end{aligned} \tag{22}$$

where the last term does not depend on the bandwidth. If v_{it} is independent of the $\{X_{js}, Z_{js}\}$ for all i, j, s and t, or (X_{it}, Z_{it}) is strictly exogenous variable, then the second term has zero expectation because the linear transformation matrix M_D removes a cross-time **not** cross-sectional average from each variable, e.g. $\widetilde{Y}_{it} = Y_{it} - m^{-1} \sum_{s=1}^{m} Y_{is}$ for all i and t. Therefore, the first term is the dominant term in large samples and (21) is used to find an optimal smoothing matrix minimizing a weighted mean squared error of $\{\widehat{\theta}(Z_{it})\}$. Of course, we could use other weight matrices in (21) instead of M_D as long as the weight matrices can remove the fixed effects and do not trigger a non-zero expectation of the second term in (22).

Table 1 shows that the RE estimator performs better than the FE estimator when the true model is a random effects model. However, the FE estimator performs much better than the RE estimator when the true model is a fixed-effects model. This is expected since the RE estimator is inconsistent when the true model is the fixed effects model. Therefore, our simulation results indicate that a test for random effects against fixed effects will be always in demand when we analyze panel data models. In Table 2 we report simulation results of the proposed nonparametric test of random effects against fixed effects.

For the selection of the bandwidth h, for univariate case, Theorem 4.1 indicates that $h \to 0$, $nh \to \infty$, and $nh^{9/2} \to 0$ as $n \to \infty$; if we take $h \sim n^{-\alpha}$, Theorem 4.1 requires $\alpha \in (\frac{2}{9}, 1)$. To fulfill both conditions $nh \to \infty$ and $nh^{9/2} \to 0$ as $n \to \infty$, we use $\alpha = 2/7$. Therefore, in producing Table 2, we use $h = c(nm)^{-2/7}\widehat{\sigma}_z$ to calculate the RE estimator with c taking a value from .8 , 1.0, and 1.2. Since the computation is very time consuming, we only report results for $n = 50$ and 100.

With $m = 3$, the effective sample size is 150 and 300, which is small but moderate sample size. Although the bandwidth chosen this way may not be optimal, the results in Tables 2, 3, and 4 show that the proposed test statistic is not very sensitive to the choice of h when c changes and that a moderate size distortion and decent power are consistent with the findings in the nonparametric tests literature. We conjecture that some bootstrap procedures can be used to reduce the size distortion in finite samples. We will leave this as a future research topic.

6 CONCLUSION

In this paper we proposed a local linear least squares method to estimate a fixed effects varying coefficient panel data model when the number of observations across time is finite; a data-driven method was introduced to automatically find the optimal bandwidth for the proposed FE estimator. In addition, we introduced a new test statistic to test for a random effects model against a fixed effects model. Monte Carlo simulations indicate that the proposed estimator and test statistic have good finite sample performance.

7 APPENDIX

7.1 Proof of Theorem 3.1

To make our mathematical formula short, we introduce some simplified notations first: for each i and t, $\lambda_{it} = K_H(Z_{it}, z)$ and $c_H(Z_i, z)^{-1} = \sum_{t=1}^{m} \lambda_{it}$, and for any positive integers i, j, t, s

$$\begin{aligned} [\cdot]_{it,js} &= G_{it}(z,H)\, G_{js}^T(z,H) = \begin{bmatrix} 1 & G_{js1} & \cdots & G_{jsq} \\ G_{it1} & G_{it1}G_{js1} & \cdots & G_{it1}G_{jsq} \\ \vdots & \vdots & \ddots & \vdots \\ G_{itq} & G_{itq}G_{js1} & \cdots & G_{itq}G_{jsq} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \left(H^{-1}(Z_{js}-z)\right)^T \\ H^{-1}(Z_{it}-z) & H^{-1}(Z_{it}-z)\left(H^{-1}(Z_{js}-z)\right)^T \end{bmatrix} \end{aligned} \tag{A.1}$$

where the $(l+1)^{th}$ element of $G_{js}(z,H)$ is $G_{jsl} = (Z_{jsl} - z_l)/h_l$, $l = 1, \cdots, q$. Simple calculations show that

$$\begin{aligned} [\cdot]_{i_1t_1,i_2t_2}\,[\cdot]_{j_1s_1,j_2s_2} &= \left(1 + \sum_{j=1}^{q} G_{j_1s_1j}G_{i_2t_2j}\right)[\cdot]_{i_1t_1,j_2s_2}, \\ R_i(z,H)^T K_H(Z_i,z)\, e_m e_m^T K_H(Z_j,z)\, R_j(z,H) &= \sum_{t=1}^{m}\sum_{s=1}^{m} \lambda_{it}\lambda_{js}\,[\cdot]_{it,js} \otimes \left(X_{it}X_{js}^T\right) \end{aligned}$$

In addition, we obtain for a finite positive integer j

$$|H|^{-1}\sum_{t=1}^{m}E\left[\lambda_{it}^{j}\left[\cdot\right]_{it,it}|X_{it}\right] = \sum_{t=1}^{m}E\left(S_{t,j,1}|X_{it}\right)+O_{p}\left(\|H\|^{2}\right), \tag{A.2}$$

$$|H|^{-1}\sum_{t=1}^{m}E\left[\lambda_{it}^{2j}\sum_{j'=1}^{q}G_{itj'}^{2}\left[\cdot\right]_{it,it}|X_{it}\right] = \sum_{t=1}^{m}E\left(S_{t,j,2}|X_{it}\right)+O_{p}\left(\|H\|^{2}\right), \tag{A.3}$$

where

$$S_{t,j,1} = \begin{bmatrix} f_t\left(z|X_{it}\right)\int K^{j}\left(u\right)du & \frac{\partial f_t(z|X_{it})}{\partial z^T}HR_{K,j} \\ R_{K,j}H\frac{\partial f_t(z|X_{it})}{\partial z} & f_t\left(z|X_{it}\right)R_{K,j} \end{bmatrix} \tag{A.4}$$

$$S_{t,j,2} = \begin{bmatrix} f_t\left(z|X_{it}\right)\int K^{2j}\left(u\right)u^T u du & \frac{\partial f_t(z|X_{1t})}{\partial z^T}H\Gamma_{K,2j} \\ \Gamma_{K,2j}H\frac{\partial f_t(z|X_{it})}{\partial z} & f_t\left(z|X_{it}\right)\Gamma_{K,2j} \end{bmatrix} \tag{A.5}$$

where $R_{K,j}=\int K^{j}\left(u\right)uu^{T}du$ and $\Gamma_{K,2j}=\int K^{2j}\left(u\right)\left(u^{T}u\right)\left(uu^{T}\right)du$.

Moreover, for any finite positive integer j_1 and j_2, we have

$$|H|^{-2}\sum_{t=1}^{m}\sum_{s\neq t}^{m}E\left[\lambda_{it}^{j_1}\lambda_{is}^{j_2}\left[\cdot\right]_{it,is}|X_{it},X_{is}\right] \tag{A.6}$$

$$= \sum_{t=1}^{m}\sum_{s\neq t}^{m}E\left(T_{j_1,j_2,1}^{(t,s)}|X_{it},X_{is}\right)+O_{p}\left(\|H\|^{2}\right)$$

$$|H|^{-2}\sum_{t=1}^{m}\sum_{s\neq t}^{m}E\left[\lambda_{it}^{j_1}\lambda_{is}^{j_2}\left(\sum_{j'=1}^{q}G_{itj'}G_{isj'}\right)\left[\cdot\right]_{it,is}|X_{it},X_{is}\right] \tag{A.7}$$

$$= \sum_{t=1}^{m}\sum_{s\neq t}^{m}E\left(T_{j_1,j_2,2}^{(t,s)}|X_{it},X_{is}\right)+O_{p}\left(\|H\|^{2}\right)$$

where we define $b_{j1,j_2,i_1,i_2=}\int K^{j_1}\left(u\right)u_1^{2i_1}du\int K^{j_2}\left(u\right)u_1^{2i_2}du$

$$T_{j_1,j_2,1}^{(t,s)}=\begin{bmatrix} f_{t,s}\left(z,z|X_{it},X_{is}\right)b_{j_1,j_2,0,0} & \bigtriangledown_s^T f_{t,s}\left(z,z|X_{it},X_{is}\right)Hb_{j_1,j_2,0,1} \\ H\bigtriangledown_t f_{t,s}\left(z,z|X_{it},X_{is}\right)b_{j_1,j_2,1,0} & H\bigtriangledown_{t,s}^2 f_{t,s}\left(z,z|X_{it},X_{is}\right)Hb_{j_1,j_2,1,1} \end{bmatrix}$$

and

$$T_{j_1,j_2,2}^{(t,s)}=\begin{bmatrix} tr\left(H\bigtriangledown_{t,s}^2 f_{t,s}\left(z,z|X_{it},X_{is}\right)H\right) & \bigtriangledown_t^T f_{t,s}\left(z,z|X_{it},X_{is}\right)H \\ H\bigtriangledown_s f_{t,s}\left(z,z|X_{it},X_{is}\right) & f_{t,s}\left(z,z|X_{it},X_{is}\right)I_{q\times q} \end{bmatrix}b_{j_1,j_2,1,1},$$

with $\bigtriangledown_s f_{t,s}\left(z,z|X_{it},X_{is}\right)=\partial f_{t,s}\left(z,z|X_{it},X_{is}\right)/\partial z_s$ and $\bigtriangledown_{t,s}^2 f_{t,s}\left(z,z|X_{it},X_{is}\right)=\partial^2 f_{t,s}\left(z,z|X_{it},X_{is}\right)/\left(\partial z_t\partial z_s^T\right)$.

The conditional bias and variance of $vec\left(\widehat{\beta}(z)\right)$ are given as follows:

$$Bias\left[vec\left(\widehat{\beta}(z)\right)|\left\{X_{it},Z_{it}\right\}\right]=\left[R\left(z,H\right)^{T}S_{H}\left(z\right)R\left(z,H\right)\right]^{-1}R\left(z,H\right)^{T}S_{H}\left(z\right)\left[\Pi\left(z,H\right)/2+D_0\mu_0\right];$$

$$
\begin{aligned}
Var\left[vec\left(\widehat{\beta}(z)\right) \mid \{X_{it}, Z_{it}\}\right] &= \sigma_v^2 \left[R(z,H)^T S_H(z) R(z,H)\right]^{-1} \left[R(z,H)^T S_H^2(z) R(z,H)\right] \\
&\times \left[R(z,H)^T S_H(z) R(z,H)\right]^{-1}.
\end{aligned}
$$

Lemma A.1 *If Assumption A3 holds, we have*

$$
\left[\sum_{i=1}^{n} c_H(Z_i, z)\right]^{-1} = O_p\left(n^{-1}|H|\ln(\ln n)\right). \tag{A.8}
$$

Proof: Simple calculations give $E\left(\sum_{t=1}^m K_H(Z_{it}, z)\right) = |H| f(z) + O\left(|H| \|H\|^2\right)$ and $K_H(Z_{it}, z) = |H| f_t(z) + O\left(|H| \|H\|^2\right)$, where $f(z) = \sum_{t=1}^m f_t(z)$. Next, we obtain for any small $\varepsilon > 0$

$$
\begin{aligned}
&\Pr\left\{\max_{1\le i\le n} \sum_{t=1}^{m} \lambda_{it} > \varepsilon^{-1} f(z)|H|\ln(\ln n)\right\} = 1 - \Pr\left\{\max_{1\le i\le n} \sum_{t=1}^{m} \lambda_{it} \le \varepsilon^{-1} f(z)|H|\ln(\ln n)\right\} \\
= \;& 1 - \left\{1 - \Pr\left\{\sum_{t=1}^{m} \lambda_{it} > \varepsilon^{-1} f(z)|H|\ln(\ln n)\right\}\right\}^n \le 1 - \left\{1 - \frac{\varepsilon E\left(\sum_{t=1}^m \lambda_{it}\right)}{f(z)|H|\ln(\ln n)}\right\}^n \\
\le \;& 1 - \left\{1 - \varepsilon\left(1 + M\|H\|^2\right)/\ln(\ln n)\right\}^n \to 0 \text{ as } n \to \infty
\end{aligned}
$$

where the first inequality uses the the generalized Chebyshev inequality and the limit is derived using the l'Hôpital's rule. This will complete the proof of this lemma.

Lemma A.2 *Under Assumptions 1-3, we have*

$$
n^{-1}|H|^{-1} R(z,H)^T S_H(z) R(z,H) \approx |H|^{-1} \sum_{t=1}^{m} E\left(\varpi_{it}\lambda_{it}[\cdot]_{it,it} \otimes \left(X_{it}X_{it}^T\right)\right)
$$

where $\varpi_{it} = \lambda_{it}/\sum_{t=1}^m \lambda_{it} \in (0,1)$ *for all* i *and* t.

Proof: First, simple calculation gives

$$\begin{aligned}
A_n &= R(z,H)^T S_H(z) R(z,H) = R(z,H)^T W_H(z) M_H(z) R(z,H) \\
&= \sum_{i=1}^n R_i(z,H)^T K_H(Z_i,z) R_i(z,H) \\
&\quad - \sum_{j=1}^n \sum_{i=1}^n q_{ij} R_i(z,H)^T K_H(Z_i,z) e_m e_m^T K_H(Z_j,z) R_j(z,H) \\
&= \sum_{i=1}^n \sum_{t=1}^m \lambda_{it} [\cdot]_{it,it} \otimes \left(X_{it} X_{it}^T\right) - \sum_{i=1}^n q_{ii} \sum_{t=1}^m \sum_{s=1}^m \lambda_{it}\lambda_{is} [\cdot]_{it,is} \otimes \left(X_{it} X_{is}^T\right) \\
&\quad - \sum_{j=1}^n \sum_{i\neq j}^n q_{ij} \sum_{t=1}^m \sum_{s=1}^m \lambda_{it}\lambda_{js} [\cdot]_{it,js} \otimes \left(X_{it} X_{js}^T\right) \\
&= A_{n1} - A_{n2} - A_{n3},
\end{aligned}$$

where $M_H(z) = I_{n\times m} - \left[Q \otimes \left(e_m e_m^T\right)\right] W_H(z)$, and the typical elements of Q are $q_{ii} = c_H(Z_i,z) - c_H(Z_i,z)^2 / \sum_{i=1}^n c_H(Z_i,z)$ and $q_{ij} = -c_H(Z_i,z) c_H(Z_j,z) / \sum_{i=1}^n c_H(Z_i,z)$ for $i \neq j$. Here, $c_H(Z_i,z) = \left(\sum_{t=1}^m \lambda_{it}\right)^{-1}$ for all i.

Applying (A.2), (A.3), (A.6), and (A.7) to A_{n1}, we have $n^{-1}|H|^{-1} A_{n1} \approx \sum_{t=1}^m E\left[S_{t,1,1} \otimes \left(X_{it} X_{it}^T\right)\right]$ $+O_p\left(\|H\|^2\right)$ $+O_p\left(n^{-\frac{1}{2}} |H|^{-\frac{1}{2}}\right)$ if $\|H\| \to 0$ and $n|H| \to \infty$ as $n \to \infty$.

Apparently, $\sum_{t=1}^m \varpi_{it} = 1$ for all i. In addition, since the kernel function $K(\cdot)$ is zero outside the unit circle by Assumption 3, the summations in A_{n2} are taken over units such that $\left\|H^{-1}(Z_{it} - z)\right\| \leq 1$. By Lemma A.1 and by the LLN given Assumption 1 (a), we obtain

$$\left\| \frac{1}{n|H| \sum_{i=1}^n c_H(Z_i,z)} \sum_{i=1}^n \sum_{t=1}^m \sum_{s=1}^m \varpi_{it}\varpi_{is} [\cdot]_{it,is} \otimes \left(X_{it} X_{is}^T\right) \right\| = O_p\left(n^{-1} \ln(\ln n)\right)$$

and $\left\| \frac{1}{n|H|} \sum_{i=1}^n \sum_{t=1}^m \sum_{s\neq t}^m \frac{\lambda_{it}\lambda_{is}}{\sum_{t=1}^m \lambda_{it}} [\cdot]_{it,is} \otimes \left(X_{it} X_{is}^T\right) \right\| \leq \frac{1}{2n|H|} \sum_{i=1}^n \sum_{t=1}^m \sum_{s\neq t}^m \sqrt{\lambda_{it}\lambda_{is}} \left\| [\cdot]_{it,is} \otimes \left(X_{it} X_{is}^T\right) \right\|$ $= O_p(|H|)$, where we use $\sum_{t=1}^m \lambda_{it} \geq \lambda_{it} + \lambda_{is} \geq 2\sqrt{\lambda_{it}\lambda_{is}}$ for any $t \neq s$.

Hence, we have $n^{-1}|H|^{-1} A_{n2} = n^{-1}|H|^{-1} \sum_{i=1}^n \sum_{t=1}^m \varpi_{it}\lambda_{it} [\cdot]_{it,it} \otimes \left(X_{it} X_{it}^T\right) + O_p(|H|)$. Denote $d_{it} = \varpi_{it}\lambda_{it} [\cdot]_{it,it} \otimes \left(X_{it} X_{it}^T\right)$ and $\Delta_n = n^{-1}|H|^{-1} \sum_{i=1}^n \sum_{t=1}^m (d_{it} - Ed_{it})$. It is easy to show that $n^{-1}|H|^{-1} \Delta_n = O_p\left(n^{-1/2} |H|^{-1/2}\right)$. Since $E(\|d_{it}\|) \leq E\left[\lambda_{it} \left\| [\cdot]_{it,it} \otimes \left(X_{it} X_{it}^T\right) \right\| \right] \leq M|H|$ holds for all i and t, $n^{-1}|H|^{-1} A_{n2} = |H|^{-1} \sum_{t=1}^m E\left[\varpi_{it}\lambda_{it} [\cdot]_{it,it} \otimes \left(X_{it} X_{it}^T\right)\right] + o_p(1)$ exists, but we can not calculate the exact expectation due to the random denominator.

Consider A_{n3}. We have $n^{-1}|H|^{-1} \|A_{n3}\| = O_p\left(|H|^2 \ln(\ln n)\right)$ by Lemma A.1, Assumption 1, and the fact that $n^{-1}|H|^{-1} \sum_{i=1}^n \sum_{t=1}^m I\left(\left\|H^{-1}(Z_{it} - z)\right\| \leq 1\right) = 2f(z) + O_p\left(\|H\|^2\right) +$

$O_p\left(n^{-1/2}\left|H\right|^{-1/2}\right)$.

Hence, we obtain

$$\begin{aligned}
n^{-1}\left|H\right|^{-1}A_n &\approx n^{-1}\left|H\right|^{-1}A_{n1}-n^{-1}\left|H\right|^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\varpi_{it}\lambda_{it}\left[\cdot\right]_{it,it}\otimes\left(X_{it}X_{it}^T\right)\\
&= n^{-1}\left|H\right|^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(1-\varpi_{it}\right)\lambda_{it}\left[\cdot\right]_{it,it}\otimes\left(X_{it}X_{it}^T\right)\\
&= \left|H\right|^{-1}\sum_{t=1}^{m}E\left[\left(1-\varpi_{it}\right)\lambda_{it}\left[\cdot\right]_{it,it}\otimes\left(X_{it}X_{it}^T\right)\right]+o_p\left(1\right).
\end{aligned}$$

This will complete the proof of this Lemma.

Lemma A.3 *Under Assumptions 1-3, we have*

$$n^{-1}\left|H\right|^{-1}R\left(z,H\right)^T S_H\left(z\right)\Pi\left(z,H\right)\approx\left|H\right|^{-1}\sum_{t=1}^{m}E\left[\left(1-\varpi_{it}\right)\lambda_{it}\left(G_{it}\otimes X_{it}\right)X_{it}^T r_H\left(\tilde{Z}_{it},z\right)\right].$$

Proof: Simple calculations give

$$\begin{aligned}
B_n &= R\left(z,H\right)^T S_H\left(z\right)\Pi\left(z,H\right)\\
&= \sum_{i=1}^{n}\sum_{t=1}^{m}\lambda_{it}\left(G_{it}\otimes X_{it}\right)X_{it}^T r_H\left(\tilde{Z}_{it},z\right)-\sum_{j=1}^{n}\sum_{i=1}^{n}q_{ij}\sum_{s=1}^{m}\sum_{t=1}^{m}\lambda_{js}\lambda_{it}\left(G_{it}\otimes X_{it}\right)X_{js}^T r_H\left(\tilde{Z}_{js},z\right)\\
&= \sum_{i=1}^{n}\sum_{t=1}^{m}\lambda_{it}\left(G_{it}\otimes X_{it}\right)X_{it}^T r_H\left(\tilde{Z}_{it},z\right)-\sum_{i=1}^{n}q_{ii}\sum_{t=1}^{m}\lambda_{it}^2\left(G_{it}\otimes X_{it}\right)X_{it}^T r_H\left(\tilde{Z}_{it},z\right)\\
&\quad -\sum_{i=1}^{n}q_{ii}\sum_{t=1}^{m}\sum_{s\neq t}^{m}\lambda_{is}\lambda_{it}\left(G_{it}\otimes X_{it}\right)X_{is}^T r_H\left(\tilde{Z}_{is},z\right)\\
&\quad -\sum_{j=1}^{n}\sum_{i\neq j}^{n}q_{ij}\sum_{t=1}^{m}\sum_{s=1}^{m}\lambda_{js}\lambda_{it}\left(G_{it}\otimes X_{it}\right)X_{js}^T r_H\left(\tilde{Z}_{js},z\right)\\
&= B_{n1}-B_{n2}-B_{n3}-B_{n4},
\end{aligned}$$

where $\Pi\left(z,H\right)$ is defined in Section 3. Using the same method in the proof of Lemma A.2, we show $n^{-1}\left|H\right|^{-1}B_n\approx n^{-1}\left|H\right|^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(1-\varpi_{it}\right)\lambda_{it}\left(G_{it}\otimes X_{it}\right)X_{it}^T r_H\left(\tilde{Z}_{it},z\right)$.

For $l=1,\cdots,k$ we have

$$\begin{aligned}
&\left|H\right|^{-1}E\left[\lambda_{it}r_{H,l}\left(Z_{it},z\right)|X_{it}\right]=\kappa_2 f_t\left(z|X_{it}\right)\Theta_H\left(z\right)+O_p\left(\left\|H\right\|^4\right)\\
&\left|H\right|^{-1}E\left[\lambda_{it}r_{H,l}\left(Z_{it},z\right)H^{-1}\left(Z_{it}-z\right)|X_{it}\right]=O_p\left(\left\|H\right\|^3\right),
\end{aligned}$$

and $E\left(n^{-1}|H|^{-1}B_{n1}\right)\approx\left\{\kappa_2\left[\Phi(z)\Theta_H(z)\right]^T,O\left(\|H\|^3\right)\right\}^T$, where

$$\Theta_H(z)=\left[tr\left(H\frac{\partial^2\theta_1(z)}{\partial z\partial z^T}H\right),\cdots,tr\left(H\frac{\partial^2\theta_k(z)}{\partial z\partial z^T}H\right)\right]^T.$$

Similarly we can show that $Var\left(n^{-1}|H|^{-1}B_{n1}\right)=O\left(n^{-1}|H|^{-1}\|H\|^4\right)$ if $E\left(\left\|X_{it}X_{is}^TX_{it}X_{is}^T\right\|\right)<M<\infty$ for all t and s.

In addition, it is easy to show that $n^{-1}|H|^{-1}\sum_{i=1}^n\sum_{t=1}^m\varpi_{it}\lambda_{it}\left(G_{it}\otimes X_{it}\right)X_{it}^Tr_H\left(\tilde{Z}_{it},z\right)=n^{-1}|H|^{-1}\sum_{i=1}^n\sum_{t=1}^mE\left[\varpi_{it}\lambda_{it}\left(G_{it}\otimes X_{it}\right)X_{it}^Tr_H\left(\tilde{Z}_{it},z\right)\right]+O_p\left(n^{-1/2}|H|^{-1/2}\|H\|^2\right)$, where $|H|^{-1}\sum_{t=1}^mE\left[\varpi_{it}\lambda_{it}\left(G_{it}\otimes X_{it}\right)X_{it}^Tr_H\left(\tilde{Z}_{it},z\right)\right]\leq|H|^{-1}\sum_{t=1}^mE\left[\lambda_{it}\left\|\left(G_{it}\otimes X_{it}\right)X_{it}^Tr_H\left(\tilde{Z}_{it},z\right)\right\|\right]$ $\leq M\|H\|^2<\infty$ for all i and t.

This will complete the proof of this lemma.

Lemma A.4 *Under Assumptions 1-3, we have*

$$n^{-1}|H|^{-1}R(z,H)^TS_H(z)D_0\mu_0=O_p\left(n^{-1/2}|H|\ln(\ln n)\right).$$

Proof: Simple calculations give $M_H(z)D_0\mu_0=\bar{\mu}M_H(z)\left(e_n\otimes e_m\right)$, where $\bar{\mu}=n^{-1}\sum_{i=1}^n\mu_i$. It follows that

$$\begin{aligned}C_n&=R(z,H)^TS_H(z)D_0\mu_0=\bar{\mu}R(z,H)^TS_H(z)\left(e_n\otimes e_m\right)\\&=\bar{\mu}\sum_{i=1}^n\sum_{t=1}^mR_i^TK_ie_m-\bar{\mu}\sum_{j=1}^n\left(\sum_{t=1}^m\lambda_{jt}\right)\sum_{i=1}^nq_{ij}R_i^TK_ie_m\\&=\bar{\mu}\sum_{i=1}^n\sum_{t=1}^m\lambda_{it}\left(G_{it}\otimes X_{it}\right)-\bar{\mu}\sum_{j=1}^n\left(\sum_{t=1}^m\lambda_{jt}\right)\sum_{i=1}^nq_{ij}\sum_{t=1}^m\lambda_{it}\left(G_{it}\otimes X_{it}\right)\\&=n\bar{\mu}\left[\sum_{i=1}^n\left(\sum_{t=1}^m\lambda_{it}\right)^{-1}\right]^{-1}\sum_{i=1}^n\sum_{t=1}^m\varpi_{it}\left(G_{it}\otimes X_{it}\right)\end{aligned}$$

and we obtain $n^{-1}|H|^{-1}C_n=\bar{\mu}O_p\left(|H|\ln(\ln n)\right)$ by (a) Lemma A.1, (b) for all $l=1,\cdots,q$, $k\left(\left(Z_{it,l}-z_l\right)/h\right)=0$ if $\left|Z_{it,l}-z_l\right|>h$ by Assumption 3, (c) $\varpi_{it}\leq1$, and (d) $E\|X_{it}\|^{1+\delta}<M<\infty$ for some $\delta>0$ by Assumption 1. Since $\mu_i\sim iid\left(0,\sigma_\mu^2\right)$, we have $\bar{\mu}=O_p\left(n^{-1/2}\right)$. It follows that $n^{-1}|H|^{-1}C_n=O_p\left(n^{-1/2}|H|\ln(\ln n)\right)$.

Lemma A.5 *Under Assumptions 1-3, we have*

$$n^{-1}|H|^{-1}R(z,H)^TS_H^2(z)R(z,H)\approx|H|^{-1}\sum_{t=1}^mE\left[\left(1-\varpi_{it}\right)^2\lambda_{it}^2\left[\cdot\right]_{it}\otimes\left(X_{it}X_{it}^T\right)\right].$$

Proof: Simple calculations give

$$\begin{aligned} D_n &= R(z,H)^T S_H^2(z) R(z,H) = R(z,H)^T W_H(z) M_H(z) M_H(z)^T W_H(z) R(z,H) \\ &= \sum_{i=1}^n R_i(z,H)^T K_H^2(Z_i,z) R_i(z,H) - 2\sum_{j=1}^n \sum_{i=1}^n q_{ji} R_j(z,H)^T K_H^2(Z_j,z) e_m e_m^T K_H(Z_i,z) R_i(z,H) \\ &\quad + \sum_{j=1}^n \sum_{i=1}^n \sum_{i'=1}^n q_{ij} q_{ji'} R_i(z,H)^T K_H(Z_i,z) e_m e_m^T K_H^2(Z_j,z) e_m e_m^T K_H(Z_{i'},z) R_{i'}(z,H) \\ &= D_{n1} - 2D_{n2} + D_{n3}. \end{aligned}$$

Using the same method in the proof of Lemma A.2, we show $D_n \approx \sum_{i=1}^n \sum_{t=1}^m (1-\varpi_{it})^2 \lambda_{it}^2 [\cdot]_{it,it} \otimes \left(X_{it}X_{it}^T\right)$. It is easy to show that $n^{-1}|H|^{-1} D_{n1} = n^{-1}|H|^{-1} \sum_{i=1}^n \sum_{t=1}^m \lambda_{it}^2 [\cdot]_{it,it} \otimes \left(X_{it}X_{it}^T\right) = \sum_{t=1}^m E\left[S_{t,2,1} \otimes \left(X_{it}X_{it}^T\right)\right] + O_p\left(\|H\|^2\right) + O_p\left(n^{-1/2}|H|^{-1/2}\right)$.

Also, we obtain $n^{-1}|H|^{-1} \sum_{i=1}^n \sum_{t=1}^m (1-\varpi_{it})^2 \lambda_{it}^2 [\cdot]_{it,it} \otimes \left(X_{it}X_{it}^T\right) = \varkappa(z) + O_p\left(n^{-1/2}|H|^{-1/2}\right)$, where $\varkappa(z) = |H|^{-1} \sum_{t=1}^m E\left[(1-\varpi_{it})^2 \lambda_{it}^2 [\cdot]_{it,it} \otimes \left(X_{it}X_{it}^T\right)\right] \le |H|^{-1} \sum_{t=1}^m E\left[\lambda_{it}^2 \left\|[\cdot]_{it,it} \otimes \left(X_{it}X_{it}^T\right)\right\|\right]$ $\le M < \infty$ for all i and t.

The four lemmas above are enough to give the result of Theorem 3.1. Moreover, applying Liaponuov's CLT will give the result of Theorem 3.2. Since the proof is a rather standard procedure, we drop the details for compactness of the paper.

7.2 Technical Sketch–Random Effects Estimator

The RE estimator, $\hat{\theta}_{RE}(\cdot)$, is the solution to the following optimization problem:

$$\min_{\beta(z)} [Y - R(z,H) vec(\beta(z))]^T W_H(z) [Y - R(z,H) vec(\beta(z))];$$

that is, we have

$$\begin{aligned} & vec\left(\hat{\beta}_{RE}(z)\right) \\ =\ & \left[R(z,H)^T W_H(z) R(z,H)\right]^{-1} R(z,H)^T W_H(z) Y \\ =\ & vec(\beta(z)) + \left[R(z,H)^T W_H(z) R(z,H)\right]^{-1} \left(\tilde{A}_n/2 + \tilde{B}_n + \tilde{C}_n\right) \end{aligned}$$

where $\tilde{A}_n = R(z,H)^T W_H(z) \Pi(z,H)$, $\tilde{B}_n = R(z,H)^T W_H(z) D_0\mu_0$, and $\tilde{C}_n = R(z,H)^T W_H(z) V$. Its asymptotic properties are as follows.

Lemma A.6 *Under Assumptions 1-3, and* $E\left(X_{it}X_{it}^T|z\right)$ *and* $E\left(\mu_i X_{it}|z\right)$ *have continuous second-order derivative at* $z \in R^q$. *Also,* $\sqrt{n|H|}\,\|H\|^2 = O(1)$ *as* $n \to \infty$, *and* $E\left(|v_{it}|^{2+\delta}\right) < \infty$ *and* $E\left(|\mu_i|^{2+\delta}\right) < M < \infty$ *for all* i *and* t *and for some* $\delta > 0$, *we have under* H_0

$$\sqrt{n|H|}\left(\widehat{\theta}_{RE}(z) - \theta(z) - \kappa_2\Theta_H(z)/2\right) \xrightarrow{d} N\left(0, \Sigma_{\theta(z),RE}\right), \tag{A.9}$$

where $\kappa_2 = \int k(v)\,v^2 dv$, $\Sigma_{\theta(z),RE} = \left(\sigma_\mu^2 + \sigma_v^2\right)\Phi(z)^{-1}\int K^2(u)\,du$ *and* $\Phi(z) = \sum_{t=1}^m f_t(z)\,E\left(X_{1t}X_{1t}^T|z\right)$. *Under* H_1, *we have*

$$\begin{aligned} Bias\left(\widehat{\theta}_{RE}(z)\right) &= \Phi(z)^{-1}\sum_{t=1}^m f_t(z)\,E\left(\mu_1 X_{1t}|z\right) + o(1) \\ Var\left(\hat{\theta}_{RE}(z)\right) &= n^{-1}|H|^{-1}\sigma_v^2\Phi(z)^{-1}\int K^2(u)\,du \end{aligned} \tag{A.10}$$

where $\Theta_H(z)$ *is given in the proof of Lemma A.3.*

Proof of Lemma A.6: First, we have the following decomposition

$$\sqrt{n|H|}\left[\hat{\theta}_{RE}(z) - \theta(z)\right] = \sqrt{n|H|}\left[\hat{\theta}_{RE}(z) - E\left(\hat{\theta}_{RE}(z)\right)\right] + \sqrt{n|H|}\left[E\left(\hat{\theta}_{RE}(z)\right) - \theta(z)\right],$$

where we can show that the first term converges to a normal distribution with mean zero by Liaponuov's CLT (the details are dropped since it is a rather standard proof), and the second term contributes to the asymptotic bias. Since it will cause no notational confusion, we drop the subscription 'RE'. Below, we use $Bias_i\left\{\hat{\theta}(z)\right\}$ and $Var_i\left\{\hat{\theta}(z)\right\}$ to denote the respective bias and variance of $\hat{\theta}_{RE}(z)$ under H_0 if $i = 0$ and under H_1 if $i = 1$.

First, under H_0, the bias and variance of $\hat{\theta}(z)$ are as follows: $Bias_0\left\{\hat{\theta}(z)\,|\,\{(X_{it}, Z_{it})\}\right\} = S_p\left[R(z,H)^T W_H(z)\,R(z,H)\right]^{-1} R(z,H)^T W_H(z)\,\Pi(z,H)/2$ and

$$\begin{aligned} & Var_0\left\{\hat{\theta}(z)\,|\,\{(X_{it}, Z_{it})\}\right\} \\ = \; & S_p\left[R(z,H)^T W_H(z)\,R(z,H)\right]^{-1}\left[R(z,H)^T W_H(z)\,Var(UU^T)W_H(z)\,R(z,H)\right] \\ & \times\left[R(z,H)^T W_H(z)\,R(z,H)\right]^{-1} S_p^T. \end{aligned}$$

It is simple to show that $Var(UU^T) = \sigma_\mu^2 I_n \otimes \left(e_m e_m^T\right) + \sigma_v^2 I_{n\times m}$.

Next, under H_1, we notice that $Bias_1 \left\{ \hat{\theta}(z) \,|\, \{(X_{it}, Z_{it})\} \right\}$ is the sum of $Bias_0 \left\{ \hat{\theta}(z) \,|\, \{(X_{it}, Z_{it})\} \right\}$ plus an additional term $S_p \left[R(z,H)^T W_H(z) R(z,H) \right]^{-1} R(z,H)^T W_H(z) D_0 \mu_0$, and that

$$\begin{aligned} Var_1 \left\{ \hat{\theta}(z) \,|\, \{(X_{it}, Z_{it})\} \right\} &= \sigma_v^2 S_p \left[R(z,H)^T W_H(z) R(z,H) \right]^{-1} \left[R(z,H)^T W_H(z)^2 R(z,H) \right] \\ &\quad \times \left[R(z,H)^T W_H(z) R(z,H) \right]^{-1} S_p^T. \end{aligned}$$

Noting that $R(z,H)^T W_H(z) R(z,H)$ is A_{n1} in Lemma A.2 and that $R(z,H)^T W_H(z) \Pi(z,H)$ is B_{n1} in Lemma A.3, we have

$$Bias_0 \left\{ \hat{\theta}(z) \right\} = \kappa_2 \Theta_H(z)/2 + o\left(\|H\|^2 \right). \tag{A.11}$$

In addition, under Assumptions 1-3, and $E\left(|\mu_i|^{2+\delta} \right) < M < \infty$ and $E\left(\|X_{it}\|^{2+\delta} \right) < M < \infty$ for all i and t and for some $\delta > 0$, we show that

$$\begin{aligned} n^{-1} |H|^{-1} S_p R(z,H)^T W_H(z) D_0 \mu_0 &= n^{-1} |H|^{-1} S_p \sum_{i=1}^{n} \mu_i \sum_{t=1}^{m} \lambda_{it} (G_{it} \otimes X_{it}) \\ = \sum_{t=1}^{m} f_t(z) E(\mu_1 X_{1t}|z) &+ O_p\left(\|H\|^2 \right) + O_p\left((n|H|)^{-1/2} \right), \end{aligned} \tag{A.12}$$

which is a non-zero constant plus a term of $o_p(1)$ under H_1. Combining (A.11) and (A.12), we obtain (A.10). Hence, under H_1, the bias of the RE estimator will not vanish as $n \to \infty$, and this leads to the inconsistency of the RE estimator under H_1.

As for the asymptotic variance, we can easily show that under H_0

$$Var_0 \left\{ \hat{\theta}(z) \right\} = n^{-1} |H|^{-1} \left(\sigma_\mu^2 + \sigma_v^2 \right) \Phi(z)^{-1} \int K^2(u)\, du, \tag{A.13}$$

and under H_1, $Var_1 \left\{ \hat{\theta}(z) \right\} = n^{-1} |H|^{-1} \sigma_v^2 \Phi(z)^{-1} \int K^2(u)\, du$, where we have recognized that $R(z,H)^T W_H(z)^2 R(z,H)$ is D_{n1} in Lemma A.5, and $\left(\sigma_\mu^2 + \sigma_v^2 \right) R(z,H)^T W_H(z)^2 R(z,H)$ is the leading term of $R(z,H)^T W_H(z) Var(UU^T) W_H(z) R(z,H)$.

7.3 Proof of Theorem 4.1

Define $\triangle_i = (\triangle_{i1}, \cdots, \triangle_{im})^T$ with $\triangle_{it} = X_{it}^T \left(\theta(Z_{it}) - \hat{\theta}_{RE}(Z_{it})\right)$. Since $M_D D_0 = 0$, we can decompose the proposed statistic into three terms

$$\begin{aligned}
\hat{T}_n &= \frac{1}{n^2|H|}\sum_{i=1}^{n}\sum_{j\neq i}\hat{U}_i^T Q_m A_{i,j} Q_m \hat{U}_j \\
&= \frac{1}{n^2|H|}\sum_{i=1}^{n}\sum_{j\neq i}\triangle_i^T Q_m A_{i,j} Q_m \triangle_j + \frac{2}{n^2|H|}\sum_{i=1}^{n}\sum_{j\neq i}\triangle_i^T Q_m A_{i,j} Q_m V_j \\
&\quad + \frac{1}{n^2|H|}\sum_{i=1}^{n}\sum_{j\neq i} V_i^T Q_m A_{i,j} Q_m V_j \\
&= T_{n1} + 2T_{n2} + T_{n3}
\end{aligned}$$

where $V_i = (v_{i1}, \cdots, v_{im})^T$ is the $m \times 1$ error vector. Since $\hat{\theta}_{RE}(Z_{it})$ does not depend on the jth unit observations and $\hat{\theta}_{RE}(Z_{jt})$ does not depend on the ith unit observations for a pair of (i,j), it is easy to see that $E(T_{n2}) = 0$. The proofs fall into the standard procedures seen in the literature of nonparametric tests. We therefore give a very brief proof below.

Firstly, applying Hall's (1984) CLT, we can show that under both H_0 and H_1

$$n\sqrt{|H|}T_{n3} \xrightarrow{d} N\left(0, \sigma_0^2\right) \tag{A.14}$$

by defining $H_n(\chi_i, \chi_j) = V_i^T Q_m A_{i,j} Q_m V_j$ with $\chi_i = (X_i, Z_i, V_i)$, which is a symmetric, centred and degenerate variable. We are able to show that

$$\frac{E\left[G_n^2(\chi_1,\chi_2)\right] + n^{-1}E\left[H_n^4((\chi_1,\chi_2))\right]}{\left\{E\left[H_n^2((\chi_1,\chi_2))\right]\right\}^2} = \frac{O\left(|H|^3\right) + O\left(n^{-1}|H|\right)}{O\left(|H|^2\right)} \to 0$$

if $|H| \to 0$ and $n|H| \to \infty$ as $n \to \infty$, where $G_n(\chi_1,\chi_2) = E_{\chi_i}\left[H_n((\chi_1,\chi_i))H_n((\chi_2,\chi_i))\right]$. In addition,

$$\begin{aligned}
var\left(n\sqrt{|H|}T_{n3}\right) &= 2|H|^{-1}E\left(H_n^2(\chi_1,\chi_2)\right) \\
&\approx 2\left(1-m^{-1}\right)^2\sigma_v^4\sum_{t=1}^{m}\sum_{s=1}^{m}|H|^{-1}E\left[K_H^2(Z_{1s},Z_{2t})\left(X_{1s}^T X_{2t}\right)^2\right] = \sigma_0^2 + o(1).
\end{aligned}$$

Secondly, we can show that $n\sqrt{|H|}T_{n2} = O_p\left(\|H\|^2\right) + O_p\left(n^{-1/2}|H|^{-1/2}\right)$ under H_0 and $n\sqrt{|H|}T_{n2} = O_p(1)$ under H_1. Moreover, we have, under H_0, $n\sqrt{|H|}T_{n1} = O_p\left(n\sqrt{|H|}\,\|H\|^4\right)$; under H_1, $n\sqrt{|H|}T_{n1} = O_p\left(n\sqrt{|H|}\right)$.

Finally, to estimate σ_0^2 consistently under both H_0 and H_1, we replace the unknown V_i and V_j in T_{n3} by the estimated residual vectors from the FE estimator. Simple calculations show that the typical element of $\hat{V}_i Q_m$ is $\widetilde{\tilde{v}}_{it} = y_{it} - X_{it}^T \hat{\theta}_{FE}(Z_{it}) - v_{it} - \left(\bar{y}_i - m^{-1}\sum_{t=1}^m X_{it}^T \hat{\theta}_{FE}(Z_{it}) - \bar{v}_i\right)$ $= \tilde{\triangle}_{it} - (v_{it} - \bar{v}_i)$, where $\tilde{\triangle}_{it} = X_{it}^T\left(\theta(Z_{it}) - \hat{\theta}_{FE}(Z_{it})\right) - m^{-1}\sum_{t=1}^m X_{it}^T\left(\theta(Z_{it}) - \hat{\theta}_{FE}(Z_{it})\right) = \sum_{l=1}^m q_{lt} X_{il}^T\left(\theta(Z_{il}) - \hat{\theta}_{FE}(Z_{il})\right)$ with $q_{tt} = 1 - 1/m$ and $q_{lt} = -1/m$ for $l \neq t$. The leave-two-unit-out FE estimator does not use the observations from the ith and jth units for a pair (i,j), and this leads to $E\left(\hat{V}_i^T Q_m A_{i,j} Q_m \hat{V}_j\right)^2 \approx \sum_{t=1}^m \sum_{s=1}^m E\left[K_H^2(Z_{it}, Z_{js})\left(X_{it}^T X_{js}\right)^2\left(\tilde{\triangle}_{it}^2 \tilde{\triangle}_{js}^2 + \tilde{\triangle}_{it}^2 \tilde{v}_{js}^2 + \tilde{\triangle}_{js}^2 \tilde{v}_{it}^2 + \tilde{v}_{it}^2 \tilde{v}_{js}^2\right)\right]$ $\approx \sum_{t=1}^m \sum_{s=1}^m E\left[K_H^2(Z_{it}, Z_{js})\left(X_{it}^T X_{js}\right)^2 \tilde{v}_{it}^2 \tilde{v}_{js}^2\right]$, where $\tilde{v}_{it} = v_{it} - \bar{v}_i$ and $\bar{v}_i = m^{-1}\sum_{t=1}^m v_{it}$.

ACKNOWLEDGEMENTS

Sun's research was supported from the Social Sciences and Humanities Research Council of Canada (SSHRC). Carroll's research was supported by a grant from the National Cancer Institute (CA-57030), and by the Texas A&M Center for Environmental and Rural Health via a grant from the National Institute of Environmental Health Sciences (P30-ES09106). The corresponding author: Yiguo Sun. Email address: yisun@uoguelph.ca.

REFERENCES

Arellano, M. (2003). *Panel Data Econometrics.* Oxford University Press.

Baltagi, B. (2005). *Econometrics Analysis of Panel Data (*2nd edition). Wiley, New York.

Cai, Z. and Li, Q. (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory, 24, 1321-1342.*

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

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

Henderson, D. J. and Ullah, A. (2005). A nonparametric random effects estimator. *Economics Letters*, 88, 403-407.

Hsiao, C. (2003). *Analysis of Panel Data (*2nd edition). Cambridge University Press.

Li, Q. and Huang, C.J., Li, D. and Fu, T. (2002). Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics*, 20, 412-422.

Li, Q. and Stengos, T. (1996). Semiparametric estimation of partially linear panel data models. *Journal of Econometrics*, 71, 389-397.

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *Journal of the American Statistical Association*, 96, 103-126.

Lin, X., and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with Error. *Journal of the American Statistical Association*, 95, 520-534.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimation equations. *Journal of the American Statistical Association*, 96, 1045-1056.

Lin, X. and Carroll, R. J. (2006). Semiparametric estimation in general repeated measures problems. *Journal of the Royal Statistical Society, Series B*, 68, 68-88.

Lin, X., Wang, N., Welsh, A. H. and Carroll, R. J. (2004). Equivalent kernels of smoothing splines in nonparametric regression for longitudinal/clustered data. *Biometrika*, 91, 177-194.

Poirier, D.J. (1995). Intermediate Statistics and Econometrics: a Comparative Approach. The MIT Press.

Ruckstuhl, A. F., Welsh, A. H. and Carroll, R. J. (2000). Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statistica Sinica*, 10, 51-71.

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

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, 90, 43-52.

Wu, H. and Zhang, J. Y. (2002). Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association*, 97, 883-897.

Table 1: Average mean squared errors (AMSE) of the fixed and random effects estimators when the data generation process is a random effects model and when it is a fixed effects model.

Data Process	Random Effects Estimator			Fixed Effects Estimator		
	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
Estimating $\theta_1(\cdot)$:						
$c_0 = 0$	.0951	.0533	.0277			
$c_0 = 0.5$	.6552	.5830	.5544	.1381	.1163	.1021
$c_0 = 1.0$	2.2010	2.1239	2.2310			
Estimating $\theta_2(\cdot)$:						
$c_0 = 0$	.1562	.0753	.0409			
$c_0 = 0.5$	.8629	.7511	.7200	.1984	.1379	.0967
$c_0 = 1.0$	2.8707	2.4302	2.5538			

Table 2: Percentage Rejection Rate When c0=0

c	$n = 50$			$n = 100$		
	1%	5%	10%	1%	5%	10%
0.8	.007	.015	.024	.021	.035	.046
1.0	.011	.023	.041	.025	.040	.062
1.2	.019	.043	.075	.025	.054	.097

Table 3: Percentage Rejection Rate When c0=0.5

c	$n = 50$			$n = 100$		
	1%	5%	10%	1%	5%	10%
0.8	.626	.719	.764	.913	.929	.933
1.0	.682	.780	.819	.935	.943	.951
1.2	.719	.811	.854	.943	.962	.969

Table 4: Percentage Rejection Rate When c0=1.0

c	$n = 50$			$n = 100$		
	1%	5%	10%	1%	5%	10%
0.8	.873	.883	.888	.943	.944	.946
1.0	.908	.913	.921	.962	.966	.967
1.2	.931	.938	.944	.980	.981	.982