Available online at www.sciencedirect.com

ScienceDirect

Journal of Econometrics 144 (2008) 257–275

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Nonparametric estimation and testing of fixed effects panel data models

Daniel J. Henderson[a,*], Raymond J. Carroll[b], Qi Li[c,d]

[a]Department of Economics, State University of New York at Binghamton, Binghamton, NY 13902-6000, USA
[b]Department of Statistics, Texas A&M University, College Station, TX 77843-3134, USA
[c]Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA
[d]Department of Economics, Tsinghua University, Beijing 100084, PR China

Received 11 May 2006; received in revised form 17 January 2008; accepted 31 January 2008
Available online 11 February 2008

Abstract

In this paper we consider the problem of estimating nonparametric panel data models with fixed effects. We introduce an iterative nonparametric kernel estimator. We also extend the estimation method to the case of a semiparametric partially linear fixed effects model. To determine whether a parametric, semiparametric or nonparametric model is appropriate, we propose test statistics to test between the three alternatives in practice. We further propose a test statistic for testing the null hypothesis of random effects against fixed effects in a nonparametric panel data regression model. Simulations are used to examine the finite sample performance of the proposed estimators and the test statistics.
Published by Elsevier B.V.

Keywords: Fixed effects models; Model specification tests; Nonparametric kernel method; Panel data; Partially linear model; Profile method; Random effects models; Semiparametric efficiency bound

1. Introduction

Nonparametric and semiparametric kernel methods are increasingly popular tools for statisticians/econometricians. Researchers have begun to gravitate towards nonparametric and semiparametric methods when there is little prior knowledge on specific (regression) functional forms or some known parametric specifications are deemed inadequate for the problem at hand. This often occurs when formal rejection of a parametric model yields no clues as to the direction in which to search for an improved parametric model. This growing popularity of nonparametric methods stems from their ability to relax functional form assumptions of an unknown model and let the data determine a function tailored to the data. This capacity to potentially reveal structure in the data that may be missed by common parametric specifications has encouraged growth in a variety of areas of statistics and econometrics.

The estimation of panel data models is no exception. The focus has been on both semiparametric (e.g. see Ke and Wang, 2001; Li and Stengos, 1996; Ullah and Roy, 1998) and nonparametric estimation of random

*Corresponding author. Tel.: +1 607 777 4480; fax: +1 607 777 2681.
E-mail addresses: djhender@binghamton.edu (D.J. Henderson), carroll@stat.tamu.edu (R.J. Carroll), qi@econmail.tamu.edu (Q. Li).

0304-4076/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jeconom.2008.01.005

effects models (e.g. see Henderson and Ullah, 2005; Lin and Carroll, 2000, 2001, 2006; Lin et al., 2004; Lin and Ying, 2001; Ruckstuhl et al., 2000; Wang, 2003; Wu and Zhang, 2002). Estimation of these types of models is appropriate when the individual effect is independent of the regressors. This is common in many applications, where researchers often treat any unobserved individual heterogeneity as being distributed independently of the regressors. However, random effects estimators are inconsistent if the true model is one with fixed effects, i.e., individual effects which are correlated with the regressors (e.g. see Wooldridge, 2002). Indeed, economists often view the assumptions for the random effects model as being unsupported by the data. In light of this we seek to develop both nonparametric and semiparametric fixed effects estimation procedures. These procedures will be consistent under either the random or fixed effects assumptions.

We present both nonparametric and semiparametric models which either take or do not take the correlation structure into account when estimating a fixed effects nonparametric/semiparametric panel data model. Our results show that for the nonparametric model, incorporating or ignoring the within-subject correlation leads to consistent estimation results. This is shown with a sketch of the proof similar to that in Lin and Carroll (2006). However, incorporation of the correlation leads to an improvement in the estimated variance when the number of time periods is greater than two. For the semiparametric partially linear model, we also find that taking into account the correlation structure leads to efficient estimation of the finite dimensional (parametric) parameter.

Given that nonparametric estimators suffer from the curse of dimensionality, it is desirable to apply the consistent estimator with the fastest rate of convergence. Although nonparametric models are consistent under minimal assumptions, their rate of convergence is relatively slow. In contrast, semiparametric models allow for $\sqrt{n}$-convergent estimation of the parametric components, and parametric models allow all parameters to be estimated at that rate when their respective functional form restrictions are appropriate. To choose between parametric, semiparametric and nonparametric alternatives we propose in Section 4.1 tests between these three models, using a simple and practical bootstrap testing approach.

The question of whether to use random or fixed effects naturally arises with panel data. We know that when the individual effect is correlated with any of the regressors, the random effects estimator becomes biased and inconsistent. The fixed effect estimator wipes out these individual effects and leads to consistent estimates. On the other hand, if the individual effects are independent of the regressors, both estimators are consistent. In this case the random effects estimator is more efficient. This trade-off is common in econometrics and is often resolved using a testing procedure. In Section 4.2 we develop a Hausman style test for the presence of fixed versus random effects. We suggest a separate bootstrap procedure for the implementation of this test in practice.

The remainder of the paper is organized as follows: Section 2 gives the nonparametric estimation procedures when we both account for and ignore the correlation structure. Section 3 generalizes the results to the case of a semiparametric partially linear model. In Section 4 we propose test statistics for testing between parametric, semiparametric and nonparametric alternatives as well as a test statistic for testing random effects against fixed effects in nonparametric panel data regression models. Section 5 examines the finite sample properties with a small Monte Carlo study. Finally, Section 6 gives concluding remarks.

2. Fixed effects nonparametric panel data models

Consider the following nonparametric panel data regression model with fixed effects:

$$Y_{it} = \theta(Z_{it}) + \mu_i + v_{it} \quad (i = 1, \ldots, n; t = 1, \ldots, m), \tag{1}$$

where the functional form of $\theta(\cdot)$ is not specified. The covariate $Z_{it} = (Z_{it,1}, \ldots, Z_{it,q})$ is of dimension q, and all other variables are scalars. The random errors v_{it} are assumed to be i.i.d. with a zero mean, finite variance and independent of Z_{it} for all i and t.[1] Further, μ_i has a zero mean and finite variance. We allow μ_i to be correlated with Z_{it} with an unknown correlation structure. Hence, (1) is a fixed effects model. Alternatively, when μ_i is assumed to be uncorrelated with Z_{it}, model (1) is a random effects model. Note that we only consider the balanced data case in this paper for notational simplicity. The results of this paper can be generalized to the unbalanced data case.

[1]The independence between v_{it} and Z_{it} can be relaxed so that v_{it} and Z_{it} are mean independent ($\mathrm{E}(v_{it}|Z_{it}) = 0$). Simulations not reported in the paper were performed to confirm this. These results are available upon request.

We consider the usual case of large n and fixed m, and assume that the data are independent across the i index. We take a first difference to remove the fixed effects

$$\widetilde{Y}_{it} \equiv Y_{it} - Y_{i1} = \theta(Z_{it}) - \theta(Z_{i1}) + v_{it} - v_{i1}. \tag{2}$$

Note that the above difference is to subtract observation $t = 1$ from t. One can also use the alternative transformation of $Y_{it} - Y_{i,t-1} = \theta(Z_{it}) - \theta(Z_{i,t-1}) + v_{it} - v_{i,t-1}$ to remove the fixed effects. The asymptotic analysis is similar under various transformations. In this paper we will focus on the transformation given in (2). From (1) we know that $\mathrm{E}\{\theta(Z_{it})\} = \mathrm{E}(Y_{it})$. Under this condition, $\theta(\cdot)$ defined in (2) is identified. We discuss two nonparametric estimators for $\theta(\cdot)$, one which utilizes the variance structure of the terms $\widetilde{Y}_{it}$, and the other which ignores the structure. We will show that the former estimator is asymptotically more efficient when $m > 2$.

2.1. An estimator using the variance structure

Our goal in this section is to derive an estimator which exploits the variance structure. We start by defining $\widetilde{v}_{it} = v_{it} - v_{i1}$ and $\widetilde{v}_i = (\widetilde{v}_{i2}, \ldots, \widetilde{v}_{im})^{\mathrm{T}}$, where the superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose of a matrix $(\cdot)$. The variance–covariance matrix of $\widetilde{v}_i$, defined as $\Sigma = \mathrm{cov}(\widetilde{v}_i | Z_{i1}, \ldots, Z_{im}) = \mathrm{cov}(\widetilde{v}_i)$, is given by

$$\Sigma = \sigma_v^2 (I_{m-1} + e_{m-1} e_{m-1}^{\mathrm{T}}),$$

where I_{m-1} is an identity matrix of dimension $(m-1) \times (m-1)$, and e_{m-1} is a $(m-1) \times 1$ vector of ones. It is easy to check that $\Sigma^{-1} = \sigma_v^{-2}(I_{m-1} - e_{m-1} e_{m-1}^{\mathrm{T}}/m)$. Following Wang (2003) and Lin and Carroll (2006) we use a profile likelihood approach to estimate $\theta(\cdot)$. The criterion function for individual i is given by ($Y_i = (Y_{i1}, \ldots, Y_{im})$)

$$\mathscr{L}_i(\cdot) = \mathscr{L}(Y_i, \theta_i) = -\tfrac{1}{2}(\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1})^{\mathrm{T}} \Sigma^{-1} (\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1}), \tag{3}$$

where $\widetilde{Y}_i = (\widetilde{Y}_{i2}, \ldots, \widetilde{Y}_{im})^{\mathrm{T}}$, $\theta_{it} = \theta(Z_{it})$ and $\theta_i = (\theta_{i2}, \ldots, \theta_{im})^{\mathrm{T}}$.

Defining $\mathscr{L}_{i,t\theta} = \partial \mathscr{L}_i(\cdot)/\partial\theta_{it}$, and $\mathscr{L}_{i,ts\theta} = \partial^2 \mathscr{L}_i(\cdot)/(\partial\theta_{it}\partial\theta_{is})$, from (3) we obtain

$$\mathscr{L}_{i,1\theta} = -e_{m-1}^{\mathrm{T}} \Sigma^{-1} (\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1}),$$

$$\mathscr{L}_{i,t\theta} = c_{t-1}^{\mathrm{T}} \Sigma^{-1} (\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1}) \quad \text{for } t \geqslant 2,$$

where c_{t-1} is a vector of dimension $(m-1) \times 1$ with the $(t-1)$ element being 1 and all other elements being 0.

Here we will maximize a kernel-weighted objective function. We start by defining the product kernel

$$K_h(v) = \prod_{j=1}^{q} h_j^{-1} k(v_j/h_j),$$

where $k(\cdot)$ is a univariate kernel function. Further, define $(Z_{it} - z)/h = \{(Z_{it,1} - z_1)/h_1, \ldots, (Z_{it,q} - z_q)/h_q\}^{\mathrm{T}}$ and $G_{it}(z, h) = [1, \{(Z_{it} - z)/h\}^{\mathrm{T}}]^{\mathrm{T}}$, where G_{it} is of dimension $(q+1) \times 1$. Finally, define $\theta^{(1)}(z) = \partial\theta(z)/\partial z$ as the first order derivative of $\theta(\cdot)$ with respect to z. We estimate the unknown function $\theta(z)$ by solving the first order condition

$$0 = \sum_{i=1}^{n} \sum_{t=1}^{m} K_h(Z_{it} - z) G_{it}(z, h) \mathscr{L}_{i,t\theta}[Y_i, \widehat{\theta}(Z_{i1}), \ldots, \widehat{\theta}(z) + \{(Z_{it} - z)/h\} \widehat{\theta}^{(1)}(z), \ldots, \widehat{\theta}(Z_{im})], \tag{4}$$

where the argument $\mathscr{L}_{i,t\theta}$ is $\widehat{\theta}(Z_{is})$ for $s \neq t$ and $\widehat{\theta}(z) + \{(Z_{it} - z)/h\} \widehat{\theta}^{(1)}(z)$ when $s = t$.

2.1.1. An iterative procedure for nonparametric estimation

Eq. (4) suggests the following iterative procedure. Suppose the current estimate of $\theta(z)$ at the $[\ell - 1]$th step is $\widehat{\theta}_{[\ell-1]}(z)$. Then the next step estimate of $\theta(z)$ is $\widehat{\theta}_{[\ell]}(z) = \widehat{\alpha}_0(z)$, where $(\widehat{\alpha}_0, \widehat{\alpha}_1)$ solve the following equation:

$$0 = \sum_{i=1}^{n} \sum_{t=1}^{m} K_h(Z_{it} - z) G_{it}(z, h) \mathscr{L}_{i,t\theta}[Y_i, \widehat{\theta}_{[\ell-1]}(Z_{i1}), \ldots, \widehat{\alpha}_0 + \{(Z_{it} - z)/h\}^{\mathrm{T}} \widehat{\alpha}_1, \ldots, \widehat{\theta}_{[\ell-1]}(Z_{im})]. \tag{5}$$

Below we give an algorithm for estimating $\theta(\cdot)$. We note here that we need to use the restriction that

$$\sum_{i=1}^{n}\sum_{t=1}^{m}\{Y_{it}-\widehat{\theta}(Z_{it})\}=0$$

in order for $\theta(\cdot)$ to be uniquely defined based on (2), since $\mathrm{E}(Y_{it})=\mathrm{E}\{\theta(Z_{it})\}$. The algorithm is linear in the Y_{it}'s. By defining

$$H_{i,[\ell-1]}=\begin{pmatrix} Y_{i2}-\widehat{\theta}_{[\ell-1]}(Z_{i2})\\ \vdots \\ Y_{im}-\widehat{\theta}_{[\ell-1]}(Z_{im})\end{pmatrix}-\{Y_{i1}-\widehat{\theta}_{[\ell-1]}(Z_{i1})\}e_{m-1},$$

we get

$$\begin{aligned}0=&\sum_{i=1}^{n}K_h(Z_{i1}-z)G_{i1}[-e_{m-1}^{\mathrm{T}}\Sigma^{-1}H_{i,[\ell-1]}+e_{m-1}^{\mathrm{T}}\Sigma^{-1}e_{m-1}\{\widehat{\theta}_{[\ell-1]}(Z_{i1})-G_{i1}^{\mathrm{T}}(\alpha_0,\alpha_1)^{\mathrm{T}}\}]\\&+\sum_{i=1}^{n}\sum_{t=2}^{m}K_h(Z_{it}-z)G_{it}[c_{t-1}^{\mathrm{T}}\Sigma^{-1}H_{i,[\ell-1]}+c_{t-1}^{\mathrm{T}}\Sigma^{-1}c_{t-1}\{\widehat{\theta}_{[\ell-1]}(Z_{it})-G_{it}^{\mathrm{T}}(\alpha_0,\alpha_1)^{\mathrm{T}}\}],\end{aligned}$$

where $G_{it}=G_{it}(z,h)$. Further, by defining

$$D_1=n^{-1}\sum_{i=1}^{n}\left\{e_{m-1}^{\mathrm{T}}\Sigma^{-1}e_{m-1}K_h(Z_{i1}-z)G_{i1}G_{i1}^{\mathrm{T}}+\sum_{t=2}^{m}c_{t-1}^{\mathrm{T}}\Sigma^{-1}c_{t-1}K_h(Z_{it}-z)G_{it}G_{it}^{\mathrm{T}}\right\},$$

$$D_2=n^{-1}\sum_{i=1}^{n}\left\{e_{m-1}^{\mathrm{T}}\Sigma^{-1}e_{m-1}K_h(Z_{i1}-z)G_{i1}\widehat{\theta}_{[\ell-1]}(Z_{i1})+\sum_{t=2}^{m}c_{t-1}^{\mathrm{T}}\Sigma^{-1}c_{t-1}K_h(Z_{it}-z)G_{it}\widehat{\theta}_{[\ell-1]}(Z_{it})\right\},$$

$$D_3=n^{-1}\sum_{i=1}^{n}\left\{\sum_{t=2}^{m}K_h(Z_{it}-z)G_{it}c_{t-1}^{\mathrm{T}}\Sigma^{-1}H_{i,[\ell-1]}-K_h(Z_{i1}-z)G_{i1}e_{m-1}^{\mathrm{T}}\Sigma^{-1}H_{i,[\ell-1]}\right\},$$

and then solving for α_0 and α_1 leads to $\{\widehat{\alpha}_0(z),\widehat{\alpha}_1(z)\}^{\mathrm{T}}=D_1^{-1}(D_2+D_3)$. The next step estimate of $\theta(z)$ is given by $\widehat{\theta}_{[\ell]}(z)=\widehat{\alpha}_0(z)$, while $\widehat{\alpha}_1(z)$ gives the next step derivative estimator of $\theta(z)$.

Wang (2003) considered the random effects case. In her model a consistent initial estimator can be obtained by replacing Σ by an identity matrix. The simulations reported in Wang (2003) show that a one-step iteration is nearly as efficient as the result for full convergence, and that it usually only takes 3–4 iterations to achieve full convergence. In our case, even when one replaces Σ by an identity matrix, (2) is an additive model with the restriction that the two additive functions have the same functional form, and an initial consistent estimator of $\theta(\cdot)$ can be obtained by the standard backfitting method, see for example Opsomer and Ruppert (1997). Alternatively, one can use a nonparametric series method to obtain an initial consistent estimator of $\theta(\cdot)$. The advantage of using the series method to estimate an additive model is that one can easily impose the additive structure. This is because the method involves only a least squares estimation procedure. We suggest to use the series method to obtain an initial estimator for $\theta(\cdot)$.

2.1.2. Asymptotic theory

To study the asymptotic distribution of $\widehat{\theta}(z)$, we first give some regularity conditions and definitions.

Assumption 1. The random variables (Y_{it},Z_{it}) are independent and identically distributed across the i index and Y_{it} has finite fourth moments for all t. Let $f_t(\cdot)$ denote the density function of Z_{it}; then both $f_t(\cdot)$ and $\theta(\cdot)$ are twice continuously differentiable functions. Let $\mathscr{S}_t$ denote the support of Z_{it}; then $f_t(z)$ is bounded from both below and above by some positive constant for all $z\in\mathscr{S}_t$. The initial estimator of $\theta(z)$ that is used to start the iteration is a consistent estimator of $\theta(\cdot)$.

Assumption 2. The univariate kernel function $k(\cdot)$ is a bounded, symmetric probability density function where $\kappa_2 \overset{\text{def}}{=} \int k(v)v^2\,dv$ is finite. As $n \to \infty$, $h_r \to 0$ for all $r = 1, \ldots, q$ and $nh_1 \cdots h_q \to \infty$.

We first make the following general definitions, the calculations of which will follow after the statement of the main result. Define

$$\Omega(z) = -\sum_{t=1}^{m} f_t(z)\mathrm{E}\{\mathscr{L}_{i,tt\theta}|Z_{it} = z\}.$$

Further define $\varepsilon_{it} = \mathscr{L}_{t\theta}\{\widetilde{Y}_i, Z_i, \theta(Z_{i1}), \ldots, \theta(Z_{it}), \ldots, \theta(Z_{im})\}$,

$$\eta_n = \sum_{s=1}^{q} h_s^2 + (nh_1 \cdots h_q)^{-1/2},$$

and define $b_r(z)$ to be a bounded[2] and continuous function that is the solution to

$$b_r(z) = \frac{\kappa_2}{2}\theta_{rr}(z) - \sum_{t=1}^{m}\sum_{s\neq t}^{m} f_t(z)\mathrm{E}\{\mathscr{L}_{i,ts\theta}(\cdot)b_r(Z_{is})|Z_{it} = z\}/\Omega(z), \tag{6}$$

where $\theta_{rr}(z) = \partial^2\theta(z)/\partial z_r^2$. Note that for our model $\mathscr{L}_{ts\theta}$ is non-stochastic. Specifically, $\mathscr{L}_{i,1s} = -1/(m\sigma_v^2)$ for $s\neq 1$ and $\mathscr{L}_{i,ts\theta} = 1/(m\sigma_v^2)$ for $t, s \geqslant 2$ and $t \neq s$. Hence, (6) can be written as

$$b_r(z) = \frac{\kappa_2}{2}\theta_{rr}(z) + \frac{1}{m\sigma_v^2}\sum_{t=1}^{m}\sum_{s\neq t}^{m} d_{ts} f_t(z)\mathrm{E}\{b_r(Z_{is})|Z_{it} = z\}/\Omega(z), \tag{7}$$

where $d_{ts} = 1$ if $t = 1$ or $s = 1$, and $d_{ts} = -1$ otherwise. If Z_{it} and Z_{is} are independent for $t \neq s$, we obtain a closed form solution for $b_r(z)$. In the general dependence case, $b_r(\cdot)$ does not have a closed form expression.

Here we state our main result. A sketch of the proof is in the appendix and it follows along the lines of the more detailed arguments in Lin and Carroll (2006).

Main result: The estimator $\widehat{\theta}(z)$ has the asymptotic expansion

$$\widehat{\theta}(z) - \theta(z) = (\kappa_2/2)\sum_{r=1}^{q} h_r^2 b_r(z) - n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it} - z)\varepsilon_{is}/\Omega(z) + \mathrm{o_p}(\eta_n). \tag{8}$$

Thus, the asymptotic bias and variance of $\widehat{\theta}(z)$ are

$$\text{bias} = (\kappa_2/2)\sum_{r=1}^{q} h_r^2 b_r(z) + \mathrm{o}\left(\sum_{r=1}^{q} h_r^2\right),$$

$$\text{variance} = \frac{\kappa^q}{nh_1 \cdots h_q}\frac{1}{\Omega(z)} + \mathrm{o}\{(nh_1 \cdots h_q)^{-1}\}.$$

Remark 1. The particular form of our problem means that many of the terms have simple expressions. In particular,

$$\mathscr{L}_{i,11\theta} = -e_{m-1}^{\mathrm{T}}\Sigma^{-1}e_{m-1} = -(m-1)/(m\sigma_v^2),$$

$$\mathscr{L}_{i,tt\theta} = -c_{t-1}^{\mathrm{T}}\Sigma^{-1}c_{t-1} = -(m-1)/(m\sigma_v^2) \quad \text{for } t \geqslant 2,$$

$$\mathscr{L}_{i,1t\theta} = -c_{t-1}^{\mathrm{T}}\Sigma^{-1}e_{m-1} = -\frac{1}{m\sigma_v^2} \quad \text{for } t \geqslant 2,$$

[2]Note that $b_r(z)$ is bounded because Z is bounded. However, we leave the rigorous proof of the existence and continuity of $b_r(\cdot)$ to future research.

$$\mathscr{L}_{i,ts\theta} = -c_{t-1}^{\mathrm{T}}\Sigma^{-1}c_{s-1} = \frac{1}{m\sigma_v^2} \quad \text{for } t,s\geqslant 2 \text{ and } t\neq s,$$

$$\Omega(z) = -\sum_{t=1}^{m} f_t(z)\mathrm{E}\{\mathscr{L}_{tt\theta}(\cdot)|Z_t = z\} = \frac{m-1}{m\sigma_v^2}\sum_{t=1}^{m} f_t(z).$$

Remark 2. If we further assume that $f_t(z) = f(z)$ for all t, then the asymptotic variance becomes

$$\mathrm{avar}\{\sqrt{nh_1\cdots h_q}\widehat{\theta}(z)\} = \frac{\sigma_v^2\kappa^q}{(m-1)f(z)}. \tag{9}$$

Under the assumption that $h_r \sim n^{-1/(4+q)}$, and by defining $\kappa = \int k^2(v)\,\mathrm{d}v$, we obtain the following asymptotic distribution for $\widehat{\theta}(z)$:

$$(nh_1\cdots h_q)^{1/2}\left\{\widehat{\theta}(z) - \theta(z) - \sum_{r=1}^{q} h_r^2 b_r(z)\right\} \to \mathrm{N}(0, \kappa^q/\Omega(z)) \quad \text{in distribution.} \tag{10}$$

Remark 3. Obviously, $\Omega(z)$ can be consistently estimated by

$$\widehat{\Omega}(z) = (m-1)\sum_{t=1}^{m}\widehat{f}_t(z)/(m\widehat{\sigma}_v^2),$$

where

$$\widehat{f}_t(z) = n^{-1}\sum_{i=1}^{n} K_h(Z_{it} - z)$$

and

$$\widehat{\sigma}_v^2 = \frac{1}{2n(m-1)}\sum_{i=1}^{n}\sum_{t=2}^{m}(Y_{it} - Y_{i1} - \{\widehat{\theta}(Z_{it}) - \widehat{\theta}(Z_{i1})\})^2.$$

If Z_{it} is strictly stationary in t, $\Omega(z) = (m-1)f(z)/\sigma_v^2$, and one can estimate $\Omega(z)$ by $(m-1)\widehat{f}(z)/\widehat{\sigma}_v^2$, where

$$\widehat{f}(z) = (nm)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it} - z).$$

Here we note that $\widehat{\sigma}_v^2$ is only necessary in order to estimate the covariance matrix of the limit distribution. It is not necessary for the estimates of the unknown function and its derivative because given our specification, as discussed in Remark 1, in the estimation of $\{\widehat{\alpha}_0(z), \widehat{\alpha}_1(z)\}^{\mathrm{T}} = D_1^{-1}(D_2 + D_3)$, $\widehat{\sigma}_v^2$ simply drops out.

Remark 4. In the appendix we provide an outline proof for the main result (8). Therefore, one may view (8) as a conjecture, rather than a rigorous proved result. Our proof follows similar arguments as in Wang (2003) and Lin and Carroll (2006) who did an iterative construction of their estimators. We leave the rigorous proof of the main results of the paper to future research. Also, as a referee correctly pointed out, since we did not provide formal proofs, the imposed assumptions presented in the paper may not be sufficient. It is possible that additional and potentially restrictive assumptions may have to be added to make the results presented in the paper rigorously valid.

2.2. An estimator ignoring the correlation structure

In this section we study the asymptotic distribution of a fixed effects estimator that ignores the variance structure Σ. Past research in nonparametric panel data estimation has shown that the ‘working independence’ method has the same rate of convergence as methods which incorporate the correlation structure (e.g. see Lin and Carroll, 2000). Thus, here we examine the estimator ignoring this structure to determine whether our estimator has an asymptotic improvement. In this case the objective function (3) is modified by replacing Σ^{-1}

by I_{m-1}, and thus becomes

$$\mathscr{L}_i(\cdot) = \mathscr{L}(Y_i, \theta_i) = -\tfrac{1}{2}(\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1})^{\mathrm{T}}(\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1}).$$

Then from (3) we obtain $\mathscr{L}_{i,t\theta}$ and $\mathscr{L}_{i,ts\theta}$ as

$$\mathscr{L}_{i,1\theta} = -e_{m-1}^{\mathrm{T}}(\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1}),$$

$$\mathscr{L}_{i,t\theta} = c_{t-1}^{\mathrm{T}}(\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1}) \quad \text{for } t \geqslant 2.$$

The iterative procedure is similar to before. Eq. (5) remains of the same form, and we solve

$$\begin{aligned} 0 = {} & \sum_{i=1}^{n} K_h(Z_{i1} - z) G_{i1}[-e_{m-1}^{\mathrm{T}} H_{i,[\ell-1]} + (m-1)\{\widehat{\theta}_{[\ell-1]}(Z_{i1}) - G_{i1}^{\mathrm{T}}(\alpha_0, \alpha_1)^{\mathrm{T}}\}] \\ & + \sum_{i=1}^{n}\sum_{t=2}^{m} K_h(Z_{it} - z) G_{it}[c_{t-1}^{\mathrm{T}} H_{i,[\ell-1]} + \{\widehat{\theta}_{[\ell-1]}(Z_{it}) - G_{it}^{\mathrm{T}}(\alpha_0, \alpha_1)^{\mathrm{T}}\}]. \end{aligned}$$

By replacing Σ^{-1} with I_{m-1}, and noting that $e_{m-1}^{\mathrm{T}} e_{m-1} = m-1$ and $c_{t-1}^{\mathrm{T}} c_{t-1} = 1$, analogous to the definitions of D_1, D_2 and D_3, we obtain

$$J_1 = n^{-1}\sum_{i=1}^{n}\left\{(m-1)K_h(Z_{i1} - z)G_{i1}G_{i1}^{\mathrm{T}} + \sum_{t=2}^{m} K_h(Z_{it} - z)G_{it}G_{it}^{\mathrm{T}}\right\},$$

$$J_2 = n^{-1}\sum_{i=1}^{n}\left\{(m-1)K_h(Z_{i1} - z)G_{i1}\widehat{\theta}_{[\ell-1]}(Z_{i1}) + \sum_{t=2}^{m} K_h(Z_{it} - z)G_{it}\widehat{\theta}_{[\ell-1]}(Z_{it})\right\},$$

$$J_3 = n^{-1}\sum_{i=1}^{n}\left\{\sum_{t=2}^{m} K_h(Z_{it} - z)G_{it}c_{t-1}^{\mathrm{T}} H_{i,[\ell-1]} - K_h(Z_{i1} - z)G_{i1}e_{m-1}^{\mathrm{T}} H_{i,[\ell-1]}\right\}.$$

Then solving for α_0 and α_1 leads to $\{\widetilde{\alpha}_0(z), \widetilde{\alpha}_1(z)\} = J_1^{-1}(J_2 + J_3)$. One can use the results of Section 2.1 to derive the asymptotic distribution of $\widetilde{\alpha}_0(z)$ by replacing Σ^{-1} with I_{m-1}. However, direct calculation of the asymptotic variance is quite simple. Under the assumption that $f_t(z) = f(z)$ for all $t = 1, \ldots, m$, it is easy to see that $J_1 = 2(m-1)f(z)\mathrm{diag}(1, \kappa_2 I_{m-1}) + o_p(1)$. It can be shown that the asymptotic variance of $J_2 + J_3$ comes from J_3 by replacing $H_{i,[\ell-1]}$ with $\widetilde{v}_i$.

We decompose J_3 into $J_3 = J_{3,1} - J_{3,2}$,

$$J_{3,1} = n^{-1}\sum_{i=1}^{n}\sum_{t=2}^{m} K_h(Z_{it} - z)G_{it}c_{t-1}^{\mathrm{T}} H_{i,[\ell-1]}$$

and

$$J_{3,2} = n^{-1}\sum_{i=1}^{n} K_h(Z_{i1} - z)G_{i1}e_{m-1}^{\mathrm{T}} H_{i,[\ell-1]}.$$

It is also easy to show that

$$\mathrm{avar}(J_{3,1}) = \frac{1}{nh_1 \cdots h_q}\{2(m-1)\sigma_v^2 \kappa^{q-1} f(z)\}\mathrm{diag}(\kappa, \kappa_{22} I_{m-1}),$$

where $\kappa_{22} = \int v^2 k(v)^2 \, \mathrm{d}v$, and

$$\mathrm{avar}(J_{3,2}) = \frac{1}{nh_1 \cdots h_q}\{(m-1)m\sigma_v^2 \kappa^{q-1} f(z)\}\mathrm{diag}(\kappa, \kappa_{22} I_{m-1}),$$

and that $\mathrm{cov}(J_{3,1}, J_{3,2})$ has an order smaller than $O\{(nh_1 \cdots h_q)^{-1}\}$. Hence, we have that

$$\mathrm{avar}(J_3) = \frac{1}{nh_1 \cdots h_q}\{4(m-1)\sigma_v^2 \kappa^{q-1} f(z)\}\left\{\frac{(2+m)}{4}\right\}\mathrm{diag}(\kappa, \kappa_{22} I_{m-1}).$$

Thus, we immediately obtain the asymptotic variance of $\{\widetilde{\alpha}_0(z), \widetilde{\alpha}_1(z)\}^{\mathrm{T}}$ which is given by

$$\frac{1}{4(m-1)^2 f(z)^2} \frac{1}{nh_1 \cdots h_q} \{4(m-1)\sigma_v^2 \kappa^{q-1} f(z)\} \left\{\frac{(2+m)}{4}\right\} \operatorname{diag}\{\kappa, (\kappa_{22}/\kappa_2^2) I_{m-1}\}$$
$$= \frac{1}{nh_1 \cdots h_q} \frac{\kappa^{q-1}\sigma_v^2}{(m-1)f(z)} \frac{(2+m)}{4} \operatorname{diag}\{\kappa, (\kappa_{22}/\kappa_2^2) I_{m-1}\}. \tag{11}$$

Comparing (11) with (9), we see that the relative asymptotic variance of $\widetilde{\alpha}_0(z)$ and $\widehat{\alpha}_0(z)$ is

$$\frac{\operatorname{avar}(\widetilde{\alpha}_0(z))}{\operatorname{avar}(\widehat{\alpha}_0(z))} = \frac{2+m}{4},$$

which equals one if $m = 2$ (as expected) and is greater than one when $m > 2$.

Even though the estimator that ignores the variance structure Σ has a large asymptotic variance (when $m > 2$), it has the advantage that it is robust to possible misspecification in Σ. While the asymptotic distribution of the estimator that uses Σ^{-1} requires that the variance structure Σ is correctly specified. For example, if v_{it} is serially correlated but one ignores the serial correlation, then form of Σ will be misspecified. Finally, the estimator that ignores Σ is computationally simple compared with the estimator that uses Σ^{-1}.

Here we note that when we replace Σ^{-1} with I_{m-1}, $\mathscr{L}_{i,11\theta} = -(m-1)$ and $\mathscr{L}_{i,tt\theta} = -1$ for $t \geqslant 2$, and $\mathscr{L}_{i,ts\theta} = -1$ for $t, s = 1, \ldots, m-1$ when t is different from s. Thus, $\Omega(z) = 2(m-1)f(z)$ and the leading bias term becomes

$$b_r(z) = \frac{\kappa_2}{2}\theta_{rr}(z) - \sum_{t=1}^{m}\sum_{s \neq t}^{m} f_t(z) \mathrm{E}\{\mathscr{L}_{i,ts\theta}(\cdot) b_r(Z_{is}) | Z_{it} = z\} / \Omega(z)$$
$$= \frac{\kappa_2}{2}\theta_{rr}(z) + \frac{1}{2(m-1)} \sum_{t=1}^{m} \sum_{s=1, s\neq t}^{q} \mathrm{E}\{b_r(Z_{is}) | Z_{it} = z\}.$$

3. A partially linear model with fixed effects

Nonparametric regression suffers from the curse of dimensionality problem when the dimension of the regressors is high. In this section we consider a semiparametric partially linear model where only a subset of the regressors enter the regression model nonparametrically. A partially linear panel data regression model with fixed effects is given by

$$Y_{it} = X_{it}^{\mathrm{T}}\beta + \theta(Z_{it}) + \mu_i + v_{it} \quad (i = 1, \ldots, n; t = 1, \ldots, m),$$

where X_{it} is of dimension $d \times 1$, and the other variables are as defined in Section 2.

Again we take the first difference to eliminate the fixed effects:

$$\widetilde{Y}_{it} = \widetilde{X}_{it}^{\mathrm{T}}\beta + \theta(Z_{it}) - \theta(Z_{i1}) + \widetilde{v}_{it} \quad (i = 1, \ldots, n; t = 2, \ldots, m), \tag{12}$$

where $\widetilde{X}_{it} \equiv X_{it} - X_{i1}$. The criterion function for individual i is modified to

$$\mathscr{L}_i(\cdot) = \mathscr{L}(Y_i, X_i, \beta, \theta_i)$$
$$= -\tfrac{1}{2}(\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta - \theta_i + \theta_{i1}e_{m-1})^{\mathrm{T}}\Sigma^{-1}(\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta - \theta_i + \theta_{i1}e_{m-1}),$$

where $\widetilde{X}_i = (\widetilde{X}_{i2}, \ldots, \widetilde{X}_{im})^{\mathrm{T}}$. The derivative functions become

$$\mathscr{L}_{i,1\theta} = -e_{m-1}\Sigma^{-1}(\widetilde{Y}_i - \widetilde{X}_i\beta - \theta_i + \theta_{i1}e_{m-1}),$$

$$\mathscr{L}_{i,t\theta} = c_{t-1}\Sigma^{-1}(\widetilde{Y}_i - \widetilde{X}_i\beta - \theta_i - \theta_{i1}e_{m-1}) \quad \text{for } t \geqslant 2,$$

and the second derivatives of $\mathscr{L}_{i,ts\theta}(\cdot)$ are analogous to those given in Section 2.

Following Wang et al. (2005), we estimate $\theta(\cdot)$ and β by a profile-kernel approach. For given values of β and a current stage estimator $\widehat{\theta}(\cdot)$, we estimate $\theta(z)$ by $\widehat{\alpha}_0$, where $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$ satisfy the following first order condition:

$$0=\sum_{i=1}^{n}\sum_{t=1}^{m}K_h(Z_{it}-z)G_{it}(z,h)\mathscr{L}_{t\theta}[Y_i,X_i,\beta,\widehat{\theta}(Z_{i1},\beta),\ldots,\widehat{\alpha}_0+\{(Z_{it}-z)/h\}^{\mathrm{T}}\widehat{\alpha}_1,\ldots,\widehat{\theta}(Z_{im},\beta)].$$

Compare (12) with (2). Let $\widehat{\theta}_y(\cdot)$ be the nonparametric estimator in model (2) and let $\widehat{\theta}_{x,r}(\cdot)$ be the nonparametric estimator in model (2) if Y_{it} is replaced by the rth component of X_{it}. Further, let $\widehat{\theta}_x(z)=\{\widehat{\theta}_{x,1}(z),\ldots,\widehat{\theta}_{x,d}(z)\}^{\mathrm{T}}$. It is obvious by the linearity of the smoother and from (12) that

$$\widehat{\theta}(z,\beta)=\widehat{\theta}_y(z)-\widehat{\theta}_x(z)^{\mathrm{T}}\beta. \tag{13}$$

This means that $(\partial/\partial\beta)\widehat{\theta}(z,\beta)=-\widehat{\theta}_x(z)$. Therefore, we estimate β by the minimization of

$$\sum_{i=1}^{n}\begin{bmatrix}\widetilde{Y}_{i2}-\widetilde{X}_{i2}^{\mathrm{T}}\beta-\{\widehat{\theta}(Z_{i2},\beta)-\widehat{\theta}(Z_{i1},\beta)\}\\ \vdots\\ \widetilde{Y}_{im}-\widetilde{X}_{im}^{\mathrm{T}}\beta-\{\widehat{\theta}(Z_{im},\beta)-\widehat{\theta}(Z_{i1},\beta)\}\end{bmatrix}^{\mathrm{T}}\Sigma^{-1}\begin{bmatrix}\widetilde{Y}_{i2}-\widetilde{X}_{i2}^{\mathrm{T}}\beta-\{\widehat{\theta}(Z_{i2},\beta)-\widehat{\theta}(Z_{i1},\beta)\}\\ \vdots\\ \widetilde{Y}_{im}-\widetilde{X}_{im}^{\mathrm{T}}\beta-\{\widehat{\theta}(Z_{im},\beta)-\widehat{\theta}(Z_{i1},\beta)\}\end{bmatrix}.$$

We can now invoke (13) to get an explicit solution for $\widehat{\beta}$ and an explicit covariance matrix for the asymptotic distribution of $n^{1/2}(\widehat{\beta}-\beta)$. By defining $\widetilde{Y}_{it*}=\widetilde{Y}_{it}-\{\widehat{\theta}_y(Z_{it})-\widehat{\theta}_y(Z_{i1})\}$, $\widetilde{Y}_{i*}=(\widetilde{Y}_{i2*},\ldots,\widetilde{Y}_{im*})^{\mathrm{T}}$, $\widetilde{X}_{it*}=\widetilde{X}_{it}-\{\widehat{\theta}_x(Z_{it})-\widehat{\theta}_x(Z_{i1})\}$ and $\widetilde{X}_{i*}=(\widetilde{X}_{i2*},\ldots,\widetilde{X}_{im*})$, the estimate of β is given by

$$\widehat{\beta}=\left(\sum_{i=1}^{n}\widetilde{X}_{i*}^{\mathrm{T}}\Sigma^{-1}\widetilde{X}_{i*}\right)^{-1}\left(\sum_{i=1}^{n}\widetilde{X}_{i*}^{\mathrm{T}}\Sigma^{-1}\widetilde{Y}_{i*}\right).$$

Note that since we have a closed form solution for $\widehat{\beta}$, no iteration is needed. We estimate $\theta(\cdot)$ by the same method as discussed in Section 2 except now that $\widetilde{Y}_{it}$ is replaced by $\widetilde{Y}_{it}-\widetilde{X}_{it}^{\mathrm{T}}\widehat{\beta}$ whenever it occurs. At convergence, the resulting $\widehat{\theta}(z)$ has the same asymptotic distribution as described in Section 2. Then of course, we have a nonparametric regression model as covered in Section 2. Next, notice that $\widehat{\beta}-\beta=O_p(n^{-1/2})$ converges to zero faster than the nonparametric estimator $\widehat{\theta}(z)-\theta(z)$. Therefore, replacing β by $\widehat{\beta}$ will not affect the asymptotic distribution of $\widehat{\theta}(z)$.

To derive the asymptotic distribution of $\widehat{\beta}$, we first give some definitions. Let $\mathscr{G}$ denote the space of bounded, twice continuously differentiable functions. Define $g(\cdot)=(g_1(\cdot),g_2(\cdot),\ldots,g_d(\cdot))$ to be a $1\times d$ vector function, with $g_j\in\mathscr{G}$. Define a $d\times 1$ vector function $\phi(\cdot)$ as the function that minimizes the following objective function:

$$\phi(\cdot)=\underset{\phi(\cdot)=g(\cdot)\in\mathscr{G}}{\operatorname{argmin}}\,\mathrm{E}\left[\begin{pmatrix}\widetilde{X}_{i2}-\{g(Z_{i2})-g(Z_{i1})\}\\ \vdots\\ \widetilde{X}_{im}-\{g(Z_{im})-g(Z_{i1})\}\end{pmatrix}^{\mathrm{T}}\Sigma^{-1}\begin{pmatrix}\widetilde{X}_{i2}-\{g(Z_{i2})-g(Z_{i1})\}\\ \vdots\\ \widetilde{X}_{im}-\{g(Z_{im})-g(Z_{i1})\}\end{pmatrix}\right].$$

Here we modify Assumption 1 as Assumption 3.

Assumption 3. The random variables (Y_{it},X_{it},Z_{it}) are independent and identically distributed across the i index and Y_{it} and each component of X_{it} has finite fourth moments for all t. Let $f_t(\cdot)$ denote the density function of Z_{it}; then both $f_t(\cdot)$ and $\theta(\cdot)$ are twice continuously differentiable functions. Let $\mathscr{S}_t$ denote the support of Z_{it}; then $f_t(z)$ is bounded from both below and above by some positive constant for all $z\in\mathscr{S}_t$. The initial estimator of $\theta(z)$ that is used to start the iteration is a consistent estimator of $\theta(\cdot)$.

Then, under Assumptions 2 and 3, we obtain the asymptotic distribution for the convergent $\widehat{\beta}$ as given by

$$n^{1/2}(\widehat{\beta}-\beta)\to \text{Normal }(0,V^{-1})\text{ in distribution,} \tag{14}$$

where

$$V = \mathrm{E}\left[\begin{pmatrix} \widetilde{X}_{i2} - \{\phi(Z_{i2}) - \phi(Z_{i1})\} \\ \vdots \\ \widetilde{X}_{im} - \{\phi(Z_{im}) - \phi(Z_{i1})\} \end{pmatrix}^{\mathrm{T}} \Sigma^{-1} \begin{pmatrix} \widetilde{X}_{i2} - \{\phi(Z_{i2}) - \phi(Z_{i1})\} \\ \vdots \\ \widetilde{X}_{im} - \{\phi(Z_{im}) - \phi(Z_{i1})\} \end{pmatrix}\right].$$

Moreover, V can be consistently estimated by

$$\widehat{V} = \frac{1}{n}\sum_{i=1}^{n} \widetilde{X}_{i*}^{\mathrm{T}} \widehat{\Sigma}^{-1} \widetilde{X}_{i*},$$

where $\widetilde{X}_{i*}$ is defined in constructing our estimator $\widehat{\beta}$, $\widehat{\Sigma}^{-1}$ is a consistent estimator of Σ, which relies on a consistent estimator of σ_v^2. It is easy to show that

$$\widehat{\sigma}_v^2 = [2n(m-1)]^{-1} \sum_{i=1}^{n}\sum_{t=2}^{m} [\widetilde{Y}_{it}^{*} - \widetilde{X}_{it}^{*}\widehat{\beta}]^2$$

is a consistent estimator of σ_v^2.

Bickel et al. (1993), Bickel and Kwon (2002), and Chamberlain (1992) provide general treatment on inferences and efficient bounds analysis for semiparametric models. By following the same arguments as in Lin and Carroll (2006), one can show that V^{-1} is the semiparametric efficient lower bound for the asymptotic variance, among all estimators of β based upon the differences $Y_{it} - Y_{i1}$, when the regression errors v_{it} in (1) have a Gaussian distribution. Su and Ullah (2006) consider an alternative of β in a fixed effects partially linear panel data model. However, they did not address the problem of semiparametric efficient estimation of β.

4. Specification testing

In this section we consider two types of specification tests. The first type is to test the functional form assumptions of a regression. Specifically, we present tests to test a parametric model versus a semiparametric model, a parametric model versus a nonparametric model, and finally, a semiparametric model versus a nonparametric model. The other type is to test a random effects against a fixed effects specification.

4.1. Regression functional form specification tests

Our tests consider the following possible specifications:

$$Y_{it} = X_{it}^{\mathrm{T}}\beta + Z_{it}^{\mathrm{T}}\gamma + u_{it}, \tag{15}$$

$$Y_{it} = X_{it}^{\mathrm{T}}\beta + \theta(Z_{it}) + u_{it}, \tag{16}$$

$$Y_{it} = g(X_{it}, Z_{it}) + u_{it}, \tag{17}$$

and the functional forms of $\theta : \mathscr{R}^q \to \mathscr{R}$; and $g : \mathscr{R}^{d+q} \to \mathscr{R}$ are not specified. We assume that $u_{it} = \mu_i + v_{it}$, and we allow for the possibility that μ_i is correlated with X_{it} and/or Z_{it} in an unspecified manner.

We let $\widetilde{\beta}$ and $\widetilde{\gamma}$ denote consistent estimators of β and γ based on model (15), and let $\widehat{\beta}$ and $\widehat{\theta}(\cdot)$ denote the consistent estimators of β and $\theta(\cdot)$ based on model (16). Finally, $\widehat{g}(\cdot)$ denotes the consistent estimator of $g(\cdot)$ based on (17).

In our first test we use H_0^a to denote the null hypothesis of the linear regression model (15), against H_1^a: the corresponding alternative is the partially linear model (16).

Our test statistic for testing H_0^a is

$$I_n^a = \frac{1}{nm}\sum_{i=1}^{n}\sum_{t=1}^{m} [X_{it}^{\mathrm{T}}\widetilde{\beta} + Z_{it}^{\mathrm{T}}\widetilde{\gamma} - X_{it}^{\mathrm{T}}\widehat{\beta} - \widehat{\theta}(Z_{it})]^2.$$

Under H_0^a, I_n^a converges to 0 in probability, and I_n^a converges to a positive constant under H_1^a. Therefore, I_n^a can be used to detect whether H_0^a is true or not. We conjecture that I_n^a, after proper normalization and centering, is asymptotically normally distributed. However, the derivation of such a result is quite complicated due to the iterative procedure involved in computing $\widehat{\theta}(\cdot)$. We leave the study of its asymptotic distribution to future research. Even if one derives the asymptotic distribution of I_n^a, it is well known that asymptotic theory does not provide good approximations for nonparametric kernel-based tests in finite sample applications (e.g. see Härdle and Mammen, 1993; Lee and Ullah, 2000; Li and Wang, 1998; Whang and Andrews, 1993). Therefore, we propose the following bootstrap procedure to approximate the finite sample null distribution of I_n^a.

Let $\widehat{u}_{i,a} = (\widehat{u}_{i2,a}, \ldots, \widehat{u}_{im,a})^{\mathrm{T}}$, where

$$\widehat{u}_{it,a} = \widetilde{u}_{it,a} - (nm)^{-1}\sum_{j=1}^{n}\sum_{s=2}^{m}\widetilde{u}_{js,a}$$

is the re-centered parametric fixed effects residual, $\widetilde{u}_{it,a} = Y_{it} - Y_{i1} - (X_{it} - X_{i1})^{\mathrm{T}}\widetilde{\beta} - (Z_{it} - Z_{i1})^{\mathrm{T}}\widetilde{\gamma}$. We obtain $u_{i,a}^*$ by random draw from $\{\widehat{u}_{j,a}\}_{j=1}^n$ with replacement. Note here that we are resampling the entire set of fixed effect vector residuals for a particular cross-sectional unit ($t = 2, \ldots, m$). Then generate $Y_{it}^* - Y_{i1}^* = (X_{it} - X_{i1})^{\mathrm{T}}\widetilde{\beta} + (Z_{it} - Z_{i1})^{\mathrm{T}}\widetilde{\gamma} + u_{it,a}^*$. Call $\{X_{it}, Z_{it}, Y_{it,a}^*\}$ the bootstrap sample. Use the bootstrap sample to estimate β and γ based on model (15), say $\widetilde{\beta}^*$ and $\widetilde{\gamma}^*$, and estimate β and $\theta(\cdot)$ based on model (16), say $\widehat{\beta}^*$ and $\widehat{\theta}^*(\cdot)$. Then compute I_n^{a*}, where I_n^{a*} is obtained from I_n^a with $\widetilde{\beta}$, $\widetilde{\gamma}$, $\widehat{\beta}$ and $\widehat{\theta}(\cdot)$ replaced by $\widetilde{\beta}^*$, $\widetilde{\gamma}^*$, $\widehat{\beta}^*$ and $\widehat{\theta}^*(\cdot)$, respectively. We use the empirical distribution of I_n^{a*} to approximate the null distribution of I_n^a. We expect that this bootstrap procedure works because Y_{it}^* is generated according to the null fixed effects (linear) model. Hence, I_n^{a*} mimics the null behavior of I_n^a, whether the null hypothesis H_0^a holds true or not.[3]

We use H_0^b to denote the null hypothesis that model (15) is the correct specification and we use H_1^b to denote that (17) is the correct model. For testing H_0^b we use the test statistic

$$I_n^b = \frac{1}{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}[X_{it}^{\mathrm{T}}\widetilde{\beta} + Z_{it}^{\mathrm{T}}\widetilde{\gamma} - \widehat{g}(X_{it}, Z_{it})]^2.$$

Since H_0^a and H_0^b are identical (both testing a linear null model), we use the same bootstrap sample $\{X_{it}, Z_{it}, Y_{it,a}^*\}$ to compute I_n^b to obtain I_n^{b*}, where I_n^{b*} is obtained from I_n^b with $\widetilde{\beta}$, $\widetilde{\gamma}$, and $\widehat{g}(\cdot)$ being replaced by $\widetilde{\beta}^*$, $\widetilde{\gamma}^*$, and $\widehat{g}^*(\cdot)$, respectively. We use the empirical distribution of I_n^{b*} to approximate the null distribution of I_n^b.

Finally, we use H_0^c to denote the null hypothesis that model (16) is the correct specification and we use H_1^c to denote that (17) is the correct model. For testing H_0^c we use the test statistic

$$I_n^c = \frac{1}{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}[X_{it}^{\mathrm{T}}\widehat{\beta} + \widehat{\theta}(Z_{it}) - \widehat{g}(X_{it}, Z_{it})]^2.$$

Let $\widehat{u}_{i,c} = (\widehat{u}_{i2,c}, \ldots, \widehat{u}_{im,c})^{\mathrm{T}}$, where

$$\widehat{u}_{it,c} = \widetilde{u}_{it,c} - (nm)^{-1}\sum_{j=1}^{n}\sum_{s=2}^{m}\widetilde{u}_{js,c}$$

and $\widetilde{u}_{it,c} = Y_{it} - Y_{i1} - (X_{it} - X_{i1})^{\mathrm{T}}\widehat{\beta} - \{\widehat{\theta}(Z_{it}) - \widehat{\theta}(Z_{i1})\}$. We obtain $u_{i,c}^*$ from $\{\widehat{u}_{j,c}\}_{j=1}^n$ with replacement. Again note that we are resampling the entire set of vector residuals for a particular cross-sectional unit ($t = 2, \ldots, m$). Next generate $Y_{it}^* - Y_{i1}^* = (X_{it} - X_{i1})^{\mathrm{T}}\widehat{\beta} + \widehat{\theta}(Z_{it}) - \widehat{\theta}(Z_{i1}) + u_{it,c}^*$. Call $\{X_{it}, Z_{it}, Y_{it,c}^*\}$ the bootstrap sample. Use the bootstrap sample to estimate β and $\theta(\cdot)$ based on model (16), say $\widehat{\beta}^*$and $\widehat{\theta}^*(\cdot)$, and estimate $g(\cdot)$ based on model (16), say $\widehat{g}^*(\cdot)$. Then compute I_n^{c*}, where I_n^{c*} is obtained from I_n^c with $\widehat{\beta}$, $\widehat{\theta}(\cdot)$ and $\widehat{g}(\cdot)$ replaced by $\widehat{\beta}^*$, $\widehat{\theta}^*(\cdot)$ and $\widehat{g}^*(\cdot)$, respectively. As before, we use the empirical distribution of the bootstrapped test statistic I_n^{c*} to approximate the null distribution of I_n^c.

The finite sample performance of the above bootstrap procedures is examined via simulations in Section 5.

[3] For a similar bootstrap procedure in the context of additive nonparametric models see Fan and Jiang (2005).

4.2. Testing random versus fixed effects: a nonparametric Hausman test

The efficiency/consistency trade off between random and fixed effects models is well known in econometrics. In this section we discuss how to test for the presence of random effects versus fixed effects in a nonparametric panel data model. The model remains as (1) with $u_{it} = \mu_i + v_{it}$. The random effects specification assumes that μ_i is uncorrelated with the regressor Z_{it}, while for the fixed effects case, μ_i is allowed to be correlated with Z_{it} in an unknown way.

We are interested in testing the null hypothesis that μ_i is a random effect versus the alternative hypothesis that μ_i is a fixed effect. The null hypothesis can be written as

$$\mathrm{H}_0 : \mathrm{E}(\mu_i|Z_{i1}, \ldots, Z_{im}) = 0 \quad \text{almost everywhere.}$$

The alternative hypothesis is the negation of the null, i.e., $\mathrm{H}_1 : \mathrm{E}(\mu_i|Z_{i1}, \ldots, Z_{im}) \neq 0$ on a set with positive measure. We maintain the assumption that $\mathrm{E}(v_{it}|Z_{i1}, \ldots, Z_{im}) = 0$ under either H_0 or H_1. The null and the alternative hypotheses can then be equivalently written as

$$\mathrm{H}_0 : \mathrm{E}(u_{it}|Z_{i1}, \ldots, Z_{im}) = 0 \quad \text{almost everywhere,}$$

where $u_{it} = \mu_i + v_{it}$, and

$$\mathrm{H}_1 : \mathrm{E}(u_{it}|Z_{i1}, \ldots, Z_{im}) \neq 0 \quad \text{on a set with positive measure.}$$

Our proposed test is based on the sample analogue of $J = \mathrm{E}\{u_{it}\mathrm{E}(u_{it}|Z_{it})f(Z_{it})\}$. Note that $J = 0$ under H_0 and $J = \mathrm{E}[\{\mathrm{E}(u_{it}|Z_{it})\}^2 f(Z_{it})] > 0$ if the null is false. Hence, J serves as a proper candidate for testing H_0.

For notational simplicity, we impose an additional assumption that $f_t(\cdot) = f(\cdot)$ for all $t = 1, \ldots, m$. Let $\widehat{\theta}(z)$ denote a consistent estimator of $\theta(z)$ under the fixed effects assumption. Then a consistent estimator of u_{it} is given by $\widehat{u}_{it} = y_{it} - \widehat{\theta}(z_{it})$. Our feasible test statistic is given by

$$\begin{aligned}\widehat{J} &= (nm)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\widehat{u}_{it}\widehat{\mathrm{E}}_{-it}(\widehat{u}_{it}|Z_{it})\widehat{f}_{-it}(Z_{it})\\ &= \{nm(nm-1)\}^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\sum_{j=1}^{n}\sum_{s=1,\{j,s\}\neq\{i,t\}}^{m}\widehat{u}_{it}\widehat{u}_{js}K_{h,it,js},\end{aligned}$$

where $K_{h,it,js} = K_h(Z_{it} - Z_{js})$,

$$K_h(v) = \prod_{\ell=1}^{q} h_l^{-1} k(v_\ell/h_\ell),$$

$k(\cdot)$ is a univariate kernel function, and

$$\mathrm{E}_{-it}(\widehat{u}_{it}|Z_{it}) = \{n(m-1)\}^{-1}\sum_{j=1}^{n}\sum_{s=1,\{js\}\neq\{it\}}^{m}\widehat{u}_{js}K_{h,it,js}/\widehat{f}_{-it}(Z_{it})$$

and

$$\widehat{f}_{-it}(Z_{it}) = \{n(m-1)\}^{-1}\sum_{j=1}^{n}\sum_{s=1,js,\{js\}\neq\{it\}}^{m}K_{h.it,js}$$

are the leave-one-out estimators of $\mathrm{E}(u_{it}|Z_{it})$ and $f(Z_{it})$, respectively. Li and Wang (1998) consider a similar test statistic with cross-sectional data. As in Li and Wang (1998), we use the leave-one-out kernel estimator in order to remove an asymptotic non-negligible center term.

It can be shown that $\widehat{J}$ is a consistent estimator of J. Hence, $\widehat{J} \xrightarrow{\mathrm{p}} 0$ under the null, and $\widehat{J} \xrightarrow{\mathrm{p}} C$ if H_0 is false, where $C > 0$ is a positive constant. Therefore, one rejects H_0 when $\widehat{J}$ takes large positive values.

Again, we conjecture that $\widehat{J}$, after proper normalization and centering, is asymptotically normally distributed. However, the derivation of this result is also complicated due to the iterative procedure involved in computing $\widehat{\theta}(\cdot)$. We leave the study of its asymptotic distribution to future research. Therefore, we propose the following bootstrap procedure to approximate the finite sample null distribution of $\widehat{J}$.

The bootstrap procedure below will be different from those of Section 4.1 because the null hypothesis is a random effects model now, and we cannot use the fixed effects residual as the basis for the bootstrap. Instead we must base it on the random effects residual to carry out the bootstrap procedure so that the null model is imposed on the bootstrap sample. Let $\widehat{u}_i = (\widehat{u}_{i1}, \ldots, \widehat{u}_{im})^{\mathrm{T}}$, where $\widehat{u}_{it} = Y_{it} - \widetilde{\theta}(Z_{it})$ is the residual from the random effects model, and $\widetilde{\theta}(z)$ is the random effects estimator of $\theta(z)$. Compute the two-point wild bootstrap errors by $u_i^* = \{(1-\sqrt{5})/2\}\widehat{u}_i$ with probability $r = (1+\sqrt{5})/(2\sqrt{5})$ and $u_i^* = \{(1+\sqrt{5})/2\}\widehat{u}_i$ with probability $1-r$. Then generate Y_{it}^* via $Y_{it}^* = \widetilde{\theta}(Z_{it}) + u_{it}^*$. Call $\{Z_{it}, Y_{it}^*\}_{i=1,t=1}^{n,m}$ the bootstrap sample. Using the bootstrap sample to estimate $\theta(z)$ via the random effects method, denote the estimate by $\widetilde{\theta}^*(z)$, and then obtain the bootstrap residual by $\widehat{u}_{it}^* = Y_{it}^* - \widetilde{\theta}^*(Z_{it})$. The bootstrap test statistic $\widehat{J}^*$ is obtained as in $\widehat{J}$ except that $\widehat{u}_{it}$ ($\widehat{u}_{js}$) is replaced by $\widehat{u}_{it}^*$ ($\widehat{u}_{js}^*$) wherever it occurs. This process is repeated a large number (B) of times. The empirical distribution of the B bootstrap statistics is then used to approximate the null distribution of the test statistic $\widehat{J}$. The above bootstrap-based testing procedure can be generalized to test random effects against fixed effects in a semiparametric partially linear model where the modifications are that one replaces the nonparametric residual by the residual obtained from estimating the partially linear model, and also generates Y_{it}^* based on a partially linear model.

The finite sample performance of this bootstrap procedure is also examined via simulations in Section 5.

5. Monte Carlo simulations

5.1. Nonparametric regression

This section uses Monte Carlo simulations to examine the finite sample performance of the nonparametric panel data estimators. Following a methodology similar to Wang (2003), the following data generating process is used: $Y_{it} = \sin(2Z_{it}) + \mu_i + v_{it}$, where Z_{it} is *i.i.d.* uniform$[-1, 1]$, and v_{it} is *i.i.d.* Normal$(0, 1)$. Let v_i denote an *i.i.d.* uniform$[-1, 1]$ sequence of random variables. We generate $\mu_i = v_i + c_0 Z_{i\cdot}$, where

$$Z_{i\cdot} = m^{-1} \sum_{t=1}^{m} Z_{it}.$$

$c_0 = 0$ gives the random effects model, and $c_0 \neq 0$ leads to the fixed effects model. Note that Z_{it} and μ_i are correlated for the fixed effects model. The variances of v_{it} and v_i are both fixed at unity. We use the Gaussian kernel function and the bandwidth is selected as $h = \widehat{\sigma}_z (nm)^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of $\{Z_{it}\}_{i=1,t=1}^{n,m}$.[4]

In the simulations reported below, for the fixed effects estimator we give results both when the variance structure Σ^{-1} is ignored or is used. For the random effects estimator, we ignore the variance structure. Since $m = 3$ in this example, we expect slightly better mean squared error efficiency for the estimator that accounts for the correlation. In this case, the random effects estimator is a simple local constant estimator (no iteration is needed). However, as noted previously, the fixed effects estimation procedure is iterative, even when one ignores the correlation structure of Σ^{-1}, and thus requires information on the previous iteration of $\theta(\cdot)$. Thus we set the initial value of $\theta(\cdot)$ as follows: first, we use OLS to estimate the model using a fourth order polynomial. Next, we calculate the expected value of Y given Z for each observation using the OLS estimates and use these as our starting values for $\theta(\cdot)$. Finally, we use the iterative method discussed in Section 2 to obtain estimates of the fixed effects estimator using the initial estimate of $\theta(\cdot)$ as described above. The convergence criterion is set to be

$$\sum_{i=1}^{n} \sum_{t=1}^{m} \{\widehat{\theta}_{[\ell]}(Z_{it}) - \widehat{\theta}_{[\ell-1]}(Z_{it})\}^2 \Big/ \sum_{i=1}^{n} \sum_{t=1}^{m} \widehat{\theta}_{[\ell-1]}(Z_{it})^2 < 0.001.$$

Unlike the random effects iterative procedure in Wang (2003) which performs well with a one-step iteration, the fixed effects estimation generally needs an average of five to six iterations to obtain convergence.

[4]We use this simple bandwidth selector for computational ease. The relative performance of the estimators are not significantly affected by slight changes in the bandwidth values.

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			Fixed effects without Σ			Fixed effects with Σ		
	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$
Random	0.0477	0.0274	0.0154	0.0538	0.0302	0.0175	0.0495	0.0283	0.0158
Fixed	0.4377	0.4269	0.4266	0.1290	0.0816	0.0475	0.1197	0.0739	0.0436

Fixed effects without Σ means that the covariance matrix Σ is assumed to be the identity, while *Fixed effects with* Σ uses Σ in order to improve efficiency. The number of time periods (m) is set equal to three. The number of Monte Carlo replications (M) is set equal to 1000.

We use both fixed effects and random effects methods to estimate $\theta(\cdot)$, and compute the average mean squared error (AMSE) by

$$\text{AMSE} = M^{-1}\sum_{l=1}^{M}(nm)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\{\widehat{\theta}(Z_{it,l}) - \theta(Z_{it,l})\}^2,$$

where the subscript l denotes the lth replication. In each experiment we use $M = 1000$ replications. The number of time periods (m) is fixed at three, while the number of cross-sections (n) is varied to be 50, 100 and 200. The estimation results are given in Table 1. In summary, we find the following:

- When the data generating process is that of a random effects model ($c_0 = 0$), we see that the random effects estimator has a smaller AMSE than the fixed effects estimator. This result is expected because the fixed effects estimator is not efficient. Also as expected, for both estimators, the AMSE decreases quickly as n gets larger.
- Next, when the data are generated via a fixed effects model ($c_0 = 0.5$), the regressor Z_{it} and the fixed effects μ_i are correlated. In this case the random effects estimator is inconsistent. Indeed, Table 1 shows that the random effects AMSE does not decrease as the sample size increases. In contrast, the fixed effects estimator that removes the fixed effects leads to consistent estimation results. Its AMSE decreases rapidly as n increases.
- Finally, the fixed effects estimator that accounts for Σ has an efficiency gain over the fixed effects estimator that ignores Σ.

5.2. Functional form tests

Next, we present the finite sample performance of the functional form tests. In construction of the tests we consider three simple data generating processes[5]

$$Y_{it} = X_{it}\beta + Z_{it}\gamma + u_{it}, \tag{18}$$

$$Y_{it} = X_{it}\beta + Z_{it}^2\gamma + u_{it}, \tag{19}$$

$$Y_{it} = X_{it}^2\beta + Z_{it}^2\gamma + u_{it}. \tag{20}$$

In each model X_{it} and Z_{it} are scalars which are generated as *i.i.d.* uniform$[-1, 1]$ and uniform$[2, 4]$, respectively. $\beta = 5$ and $\gamma = 2$, and u_{it} is generated as before where $c_0 = 0.5$. We use the Gaussian kernel function and the bandwidth $h_z = \widehat{\sigma}_z(nm)^{-1/(4+q)}$, where $\widehat{\sigma}_z$ is the sample standard deviation of $\{Z_{it}\}_{i=1,t=1}^{n,m}$. h_x is defined similarly.

In the simulations reported in Table 2, we use the variance structure and account for Σ. Further, the initial values of $\theta(\cdot)$ and $g(\cdot)$ are given by the OLS estimates from fourth order polynomial models. The convergence

[5] More complicated data generating processes give qualitatively the same results.

Table 2
Estimated size and power for the $\widehat{I}$ tests for the null hypotheses of a parametric model versus a semiparametric model (a), a parametric model versus a nonparametric model (b), and a semiparametric model versus a nonparametric model (c), respectively

Size and power of the $\widehat{I}$ tests

	$n=50$			$n=100$			$n=200$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Size									
$\widehat{I}_n^a$	0.006	0.031	0.065	0.008	0.041	0.087	0.009	0.046	0.089
$\widehat{I}_n^b$	0.009	0.037	0.092	0.011	0.054	0.108	0.010	0.052	0.093
$\widehat{I}_n^c$	0.009	0.041	0.085	0.011	0.044	0.090	0.011	0.045	0.090
Power									
$\widehat{I}_n^a$	0.855	0.916	0.947	0.929	0.962	1.000	0.981	1.000	1.000
$\widehat{I}_n^b$	0.972	0.986	1.000	0.985	1.000	1.000	1.000	1.000	1.000
$\widehat{I}_n^c$	0.989	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

The number of time periods (m) is set equal to three. The number of Monte Carlo replications (M) is set equal to 1000 and the number of bootstrap replications (B) is set equal to 400.

Table 3
Estimated size and power for the $\widehat{J}$ test for the null hypothesis that the random effects model is true

Size of the $\widehat{J}$ test

c_0	$n=50$			$n=100$			$n=200$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
0	0.012	0.056	0.112	0.015	0.057	0.109	0.011	0.054	0.108
0.25	0.176	0.404	0.518	0.376	0.626	0.734	0.672	0.872	0.922
0.5	0.578	0.834	0.910	0.935	0.990	1.000	1.000	1.000	1.000

The number of time periods (m) is set equal to three. The number of Monte Carlo replications (M) is set equal to 1000 and the number of bootstrap replications (B) is set equal to 400.

criterion is the same as before where $\widehat{\theta}(\cdot)$ is replaced by $\widehat{g}(\cdot)$ when necessary. The number of iterations needed for the semiparametric model to converge is typically between 5 and 7 in the second stage. The number of time periods, cross-sections and replications are the same as before and the number of bootstraps within each replication is set equal to 400. In Table 2 we can see that the estimated sizes of each of the three tests are close to the nominal sizes for relatively small samples. The power of the three tests is also impressive and we can see that they tend towards one as the sample size grows. Again, the limited results show that the tests perform well for the typical panel data situation of large n and small m.

5.3. Nonparametric Hausman test

Finally, we examine the finite sample performance of the nonparametric test for detecting a fixed effects model against a random effects model. The data generating process is the same as in Section 5.1. Again, $c_0 = 0$ gives the random effects model, and $c_0 \neq 0$ leads to the fixed effects model. We consider $c_0 = 0, 0.25, 0.5$. The number of replications (M) here is set equal to 1000 in each setting, and the number of bootstraps within each replication is set at 400. From Table 3 we observe that the estimated sizes of the $\widehat{J}$ test (the case of $c_0 = 0$) are close to the nominal sizes. The last two rows ($c_0 = 0.25, 0.5$) give the estimated power of the $\widehat{J}$ test. We observe that the power increases rapidly as either the sample size increases or as the correlation between the individual effects and the regressor increases (i.e., as c_0 increases). The limited simulation results seem to suggest that the bootstrap-based $\widehat{J}$ test performs well for the typical panel data situation of large n and small m.

6. Conclusion

In this paper we proposed using a kernel-based methodology to estimate a nonparametric panel data model with fixed effects. We extended the estimation method to the case of a partially linear fixed effects model. To determine whether a parametric, semiparametric or nonparametric model was appropriate, we proposed bootstrap procedures to test between the three alternatives in practice. We also suggested using a bootstrap procedure to test for the presence of random effects versus fixed effects in a nonparametric panel data set. Monte Carlo simulations were used to examine the finite sample performance of the proposed estimators and test statistics.

Acknowledgments

The authors would like to thank three anonymous referees and the editor for useful comments which led to an improved version of the paper. The paper also benefitted from comments of participants of the Southern Methodist University Economics Seminar Series, the Virginia Tech University Economics Seminar Series, and the Binghamton University Statistics Seminar Series, as well as participants of the North American Summer Meetings of the Econometric Society at Duke University (June, 2007). Carroll's research was supported by a grant from the National Cancer Institute (CA-57030, CA104620), 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). Li's research was partially supported by the Private Enterprise Research Center, Texas A&M University.

Appendix A. Sketch of technical arguments

In this appendix we provide a sketch of our main result. We conjecture that a rigorous proof may follow similar arguments as given in Mammen et al. (1999), although the model we consider is more complex than the additive model considered by them. We leave this as a future research topic.

Lemma 1. *Let $\widehat{\theta}_{[\ell]}(z)$ and $\widehat{\theta}^{(1)}_{[\ell]}(z)$ be the lth step iteration estimators of $\theta(z)$ and $\theta^{(1)}(z)$, respectively, then it satisfies the following equation*:

$$\widehat{\theta}_{[\ell]}(z)-\theta(z)=\frac{\kappa_2}{2}\sum_{r=1}^{q}h_r^2\theta_{rr}(z)-\frac{1}{n\Omega(z)}\sum_{i=1}^{n}\sum_{t=1}^{m}K_h(Z_{it}-z)\varepsilon_{it}$$
$$-\frac{1}{n\Omega(z)}\sum_{i=1}^{n}\sum_{t=1}^{m}\sum_{s\neq t}^{m}K_h(Z_{it}-z)\mathscr{L}_{i,ts\theta}(\cdot)\{\widehat{\theta}_{[\ell-1]}(z)-\theta(z)\}$$
$$+\,\mathrm{o_p}\left\{\sum_{r=1}^{q}h_r^2+(nh_1\cdots h_q)^{-1/2}\right\}, \tag{A.1}$$

where $\varepsilon_{it}=\mathscr{L}_{i,t\theta}\{\widetilde{Y}_i,\theta(Z_{i1}),\ldots,\theta(Z_{it}),\ldots,\theta(Z_{im})\}$.

Proof. By Taylor expansion of the first order equation (5) with respect to $\{\widehat{\theta}_{[\ell]}(z),\widehat{\theta}^{(1)}_{[\ell]}(z)\}$ at $\{\theta(z),\theta^{(1)}(z)\}$, we have that

$$0=n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}K_h(Z_{it}-z)G_{it}(z,h)\mathscr{L}_{i,t\theta}(\cdot)$$

$$+ n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}-z)G_{it}(z,h)G_{it}(z,h)^{\mathrm{T}}\mathscr{L}_{i,tt\theta}(\cdot)\begin{pmatrix}\widehat{\theta}_{[\ell]}(z)-\theta(z)\\ \widehat{\theta}^{(1)}_{[\ell]}(z)-\theta^{(1)}(z)\end{pmatrix}$$

$$+ \mathrm{o_p}\left\{\sum_{r=1}^{q} h_r^2 + (nh_1\cdots h_q)^{-1/2}\right\}, \tag{A.2}$$

where the argument $(\cdot)$ is at $[\widetilde{Y}_i, \widehat{\theta}_{[\ell-1]}(Z_{i1}),\ldots,\theta(z)+\{(Z_{it}-z)/h\}^{\mathrm{T}}\theta^{(1)}(z),\ldots,\widehat{\theta}_{[\ell-1]}(Z_{im})]$. Note that (A.2) is a $(q+1)\times 1$ vector equation. By noting that

$$n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}-z)G_{it}G_{it}^{\mathrm{T}}\mathscr{L}_{i,tt\theta}(\cdot)\sim\sum_{t=1}^{m}\mathrm{E}[K_h(Z_{it}-z)G_{it}G_{it}^{\mathrm{T}}\mathscr{L}_{i,tt\theta}]$$

which converges to $-\Omega(z)\mathrm{diag}(1,\kappa_2\mathrm{I}_q)$, where $\kappa_2=\int k(v)v^2\,\mathrm{d}v$, hence, the first component of the vector equation (A.2) leads to

$$\Omega(z)\{\widehat{\theta}(z)-\theta(z)\} = n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}-z)\mathscr{L}_{i,t\theta}(\cdot)+\mathrm{o_p}(\eta_n)\equiv A_n+\mathrm{o_p}(\eta_n), \tag{A.3}$$

where

$$\eta_n=\sum_{r=1}^{q} h_r^2 + (nh_1\cdots h_q)^{-1/2}.$$

We decompose A_n into $A_n=A_{1n}+A_{2n}$, where A_{1n} is obtained from A_n with $\widehat{\theta}(\cdot)$ replaced by $\theta(\cdot)$, and $A_{2n}=A_n-A_{1n}$. Thus,

$$A_{1n}=n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}-z)\mathscr{L}_{i,t\theta}[\widetilde{Y}_i,\theta(Z_{i1}),\ldots,\theta(z)+\{(Z_{it}-z)/h\}^{\mathrm{T}}\theta^{(1)}(z),\ldots,\theta(Z_{im})]. \tag{A.4}$$

By adding and subtracting terms in A_{1n} we can further write $A_{1n}=A_{1n1}-A_{1n2}$, where

$$A_{1n1}=n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}-z)\varepsilon_{it}, \tag{A.5}$$

$$\begin{aligned}A_{1n2} &= n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}-z)(\mathscr{L}_{i,t\theta}\{\widetilde{Y}_i,\theta(Z_{i1}),\ldots,\theta(Z_{it}),\ldots,\theta(Z_{im})\}\\ &\quad -\mathscr{L}_{i,t\theta}[\widetilde{Y}_i,\theta(Z_{i1}),\ldots,\theta(z)+\{(Z_{it}-z)/h\}^{\mathrm{T}}\theta^{(1)}(z),\ldots,\theta(Z_{im})])\\ &= \frac{1}{2n}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}-z)(Z_{it}-z)\theta^{(2)}(z)(Z_{it}-z)^{\mathrm{T}}\mathscr{L}_{i,tt\theta}(\cdot)+\mathrm{o_p}(\eta_n)\\ &= \frac{\kappa_2}{2}\sum_{r=1}^{q} h_r^2\theta_{rr}(z)\Omega(z)+\mathrm{o_p}(\eta_n),\end{aligned} \tag{A.6}$$

where

$$\Omega(z)=-\sum_{t=1}^{m} f_t(z)\mathrm{E}(\mathscr{L}_{i,tt}|Z_{it}=z)=\{(m-1)/m\}/\sigma_v^2,\theta_{rr}(z)=\partial^2\theta(z)/\partial z_r^2,$$

$\theta^{(2)}(z)$ is the $q\times q$ second order derivative matrix of $\theta(z)$, and

$$\eta_n=\sum_{r=1}^{q} h_r^2 + (nh_1\cdots h_q)^{-1/2}.$$

In addition,

$$
\begin{aligned}
A_{2n} = A_n - A_{1n} &= n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}-z) \\
&\quad \times \mathscr{L}_{i,j\theta}[\widetilde{Y}_i, \widehat{\theta}_{[\ell-1]}(Z_{i1}), \ldots, \theta(z) + \{(Z_{it}-z)/h\}^{\mathrm{T}}\theta^{(1)}(z), \ldots, \widehat{\theta}_{[\ell-1]}(Z_{im})] - A_{1n} \\
&= n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\sum_{s\neq t}^{m} K_h(Z_{it}-z)\mathscr{L}_{i,ts\theta}(\cdot)\{\widehat{\theta}_{[\ell-1]}(Z_{is}) - \theta(Z_{is})\} \\
&\quad + \mathrm{o_p}\left(\sum_{r=1}^{q} h_r^2 + (nh_1\cdots h_q)^{-1/2}\right). \qquad \text{(A.7)}
\end{aligned}
$$

Accumulating these results yields (A.1). □

Here we sketch how one can use (A.1) to show (8). The basic argument is the same as that in Wang (2003), namely repeated application of (A.1). The idea is to start with a consistent estimator and then apply (A.1) to get an expansion for the first step in the iteration. This new expansion is then substituted into (A.1) to get an expansion for the second step in the iteration, etc. Under the assumption that the algorithm converges, the effect of the initial estimator disappears, and (8) emerges. The calculations are merely extremely detailed rather than difficult, and in the interest of space we do not provide them.

It is interesting to note that conversely, (8) is consistent with (A.1). If an estimator with the expansion (8) is used, and then updated, using (A.1) results in (8) once again.

References

Bickel, P.J., Kwon, J., 2002. Inference for semiparametric models: some current frontiers (with discussions). Statistica Sinica 11, 863–960.

Bickel, P.J., Klaassen, A.J., Ritov, Y., Wellner, J.A., 1993. Efficient and Adaptive Inference in Semiparametric Models. Johns Hopkins University Press, Baltimore, MD.

Chamberlain, G., 1992. Efficiency bounds for semiparametric regression. Econometrica 60, 567–596.

Fan, J., Jiang, J., 2005. Nonparametric inferences for additive models. Journal of the American Statistical Association 100, 890–907.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926–1947.

Henderson, D.J., Ullah, A., 2005. A nonparametric random effects estimator. Economics Letters 88, 403–407.

Ke, C., Wang, Y., 2001. Semiparametric nonlinear mixed-effects models and their applications. Journal of the American Statistical Association 96, 1272–1281.

Lee, T.-H., Ullah, A., 2000. Nonparametric bootstrap tests for neglected nonlinearity in time series regression models. Journal of Nonparametric Statistics 13, 425–451.

Li, Q., Stengos, T., 1996. Semiparametric estimation of partially linear regression models. Journal of Econometrics 71, 389–397.

Li, Q., Wang, S., 1998. A simple consistent bootstrap test for a parametric regression function. Journal of Econometrics 87, 145–165.

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

Lin, X., 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., Carroll, R.J., 2001. Semiparametric regression for clustered data using generalised estimation equations. Journal of the American Statistical Association 96, 1045–1056.

Lin, X., 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., Carroll, R.J., 2004. Equivalent kernels of smoothing splines in nonparametric regression for longitudinal/clustered data. Biometrika 91, 177–194.

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

Opsomer, J.D., Ruppert, D., 1997. Fitting a bivariate additive model by local polynomial regression. Annals of Statistics 25, 186–212.

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

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

Ullah, A., Roy, N., 1998. Nonparametric and semiparametric econometrics of panel data. In: Ullah, A., Giles, D.E.A. (Eds.), Handbook of Applied Economics Statistics. Marcel Dekker, New York, pp. 579–604.

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

Wang, N., Carroll, R.J., Lin, X., 2005. Efficient semiparametric marginal estimation for longitudinal/clustered data. Journal of the American Statistical Association 100, 147–157.

Whang, Y., Andrews, D.W.K., 1993. Tests of specification of parametric and semiparametric models. Journal of Econometrics 57, 277–318.

Wooldridge, J., 2002. Econometric Analysis of Cross Section and Panel data. MIT Press, Cambridge, MA.

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