Profile Likelihood Estimation of Partially Linear Panel Data Models with Fixed Effects*

LIANGJUN SU

Guanghua School of Management, Peking University, lsu@gsm.pku.edu.cn

AMAN ULLAH

Department of Economics, University of California, Riverside, aman.ullah@ucr.edu

May, 2005

Abstract

We consider consistent estimation of partially linear panel data models with fixed effects. We propose profile-likelihood-based estimators for both the parametric and nonparametric components in the models and establish convergence rates and asymptotic normality for both estimators.

Keywords: Fixed effects, Panel data models, Partially linear models, Profile likelihood.

JEL Classification: C13, C14, C23.

*Address Correspondence to: Liangjun Su, Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, P.R. China; E-mail: lsu@gsm.pku.edu.cn, Phone: (0086)10-62767444. Aman Ullah, Department of Economics, UCR, Riverside, CA 92521-0427 (E-mail: aman.ullah@ucr.edu; Phone: 951-827-1591). The second author gratefully acknowledges the financial support from the Academic Senate, UCR.

1 Introduction

As Baltagi and Li (2002) notice, there is a rich literature on semiparametric estimation of panel data models whereas few studies focus on consistent estimation of semiparametric panel data models with *fixed effects*. By taking the first difference to eliminate the fixed effects and using the series method, they establish asymptotic normality for the finite dimensional parameter of interest in the model and consistency for the nonparametric object. Even though their approach overcomes several drawbacks associated with the kernel approach of Li and Stengos (1996), they don't estimate the slope parameter for the nonparametric component. Recently, Mundra (2005) considers nonparametric estimation of the slope parameter for fixed-effect panel data models, but with her approach we cannot obtain estimators for the nonparametric regression mean.

In this paper, we consider estimation of partially linear panel data models with fixed effects without taking the first difference explicitly. Our approach draws support from the literature on profile likelihood, which is extremely useful for estimating semiparametric models. Given the finite dimensional parameter of interest and the fixed effect parameter, we can estimate the nonparametric object as a function of these parameters. Plugging this nonparametric object into a least squares type of objective function to minimize, we can get consistent estimator of the parameter of interest. Meanwhile, we obtain consistent estimators for both the nonparametric component in the regression mean and its slope parameter.

The paper is structured as follows. In Section 2 we introduce the profile likelihood estimators for the partially linear panel data models with fixed effects. We study the asymptotic properties of the estimators in Section 3. All technical details are relegated to the Appendix. Throughout the paper, let $\|a\| = \left(\sum_{i=1} a_i^2\right)^{1/2}$ denote the usual Euclidean norm of a vector a. More generally, for a matrix $A = (a_{ij})$, we denote its norm by $\|A\| = \{tr\,(A'A)\}^{1/2} = \left(\sum_{j=1}^{n} \sum_{i=1} a_{ij}^2\right)^{1/2}$, where prime means transpose unless otherwise stated. Let I_n denote the $n \times n$ identity matrix.

2 The Model and Estimators

Consider the following partially linear model with fixed effects:

$$y_{it} = \alpha_i + x_{it}'\beta + m\,(z_{it}) + v_{it}, i = 1, ...n, t = 1...., T, \tag{2.1}$$

where x_{it} and z_{it} are of dimensions $p \times 1$ and $q \times 1$, respectively, β is a $p \times 1$ vector of unknown parameters, $m\,(\cdot)$ is an unknown smooth function, $\alpha_i's$ are fixed effects, and $v_{it}'s$ are the random disturbances. For simplicity, we assume that (x_{it}, z_{it}) are strictly exogenous variables. We are interested in consistent estimation of β, $m\,(\cdot)$ and $\dot{m}\,(\cdot)$, where $\dot{m}\,(\cdot)$ is the first derivative of $m\,(\cdot)$. We establish the asymptotic theory by letting n approach infinity and holding T fixed.

Let K denote a kernel function on $\mathbf{R}^q$ and $H = diag\,(h_1, ..., h_q)$, a matrix of bandwidth sequences. Set $K_H\,(z) = |H|^{-1} K\left(H^{-1} z\right)$, where $|H|$ is the determinant of H. Let $Z_{it}\,(z) = [1, \left(H^{-1}\,(z_{it} - z)\right)']'$. Further denote $\mathbf{K}_H\,(z) = diag(K_H\,(z_{11} - z), ..., K_H\,(z_{1T} - z), K_H\,(z_{21} - z), ..., K_H\,(z_{nT} - z))$ and $\overrightarrow{Z}\,(z) = (Z_{11}\,(z), ..., Z_{1T}\,(z), Z_{21}\,(z), ..., Z_{nT}\,(z))'$.

Let $\alpha = (\alpha_1, ..., \alpha_n)'$ and $\theta = \left(\alpha', \beta'\right)'$. Given θ, we estimate $M\,(z) \equiv \left(m\,(z), H\dot{m}\,(z)\right)'$ by

$$M_\theta\,(z) = \underset{M \in \mathbf{R}^{q+1}}{\arg\min} \left(Y - D\alpha - X\beta - \overrightarrow{Z}\,(z)\,M\right)' \mathbf{K}_H\,(z) \left(Y - D\alpha - X\beta - \overrightarrow{Z}\,(z)\,M\right) \tag{2.2}$$

where $Y = (y_{11}, ..., y_{1T}, y_{21}, ..., y_{nT})'$, $X = (x_{11}, ..., x_{1T}, x_{21}, ..., x_{nT})'$, $D = [d_1, ..., d_n]$, d_i is a dummy variable indicating the ith unit. Define the smoothing operator by $S\,(z) = [\overrightarrow{Z}\,(z)'\,\mathbf{K}_H\,(z)$

$\overrightarrow{Z}(z)]^{-1}\overrightarrow{Z}(z)' \mathsf{K}_H(z)$. Then

$$M_\theta(z) = S(z)(Y - D\alpha - X\beta). \tag{2.3}$$

In particular, the estimator for $m(z)$ is given by

$$m_\theta(z) = s(z)(Y - D\alpha - X\beta), \tag{2.4}$$

where $s(z)' = e'S(z)$, and $e = (1, 0, ..., 0)'$ is a $(q+1) \times 1$ vector.

The parameter θ is then estimated by a profile-likelihood-based least squares regression:

$$\hat{\theta} = \arg\min_\theta (Y - D\alpha - X\beta - m_\theta(Z))'(Y - D\alpha - X\beta - m_\theta(Z)), \tag{2.5}$$

where $m_\theta(Z) = (m_\theta(z_{11}), ..., m_\theta(z_{1T}), ..., m_\theta(z_{nT}))'$. Plugging (2.4) into (2.5) and using the formula for partitioned regression, we obtain

$$\hat{\beta} = [X^{*\prime}M^*X^*]^{-1} X^{*\prime}M^*Y^*, \tag{2.6}$$

$$\hat{\alpha} = [D^{*\prime}D^*]^{-1} D^{*\prime} \left(Y^* - X^*\hat{\beta}\right), \tag{2.7}$$

where $D^* = (I_{nT} - S)D, Y^* = (I_{nT} - S)Y, X^* = (I_{nT} - S)X$, and $M^* = I_{nT} - D^*[D^{*\prime}D^*]^{-1}D^{*\prime}$.

The profile likelihood estimatos for $M(z)$ is given by

$$\hat{M}(z) = M_{\hat{\theta}}(z) = S(z)\left(Y - D\hat{\alpha} - X\hat{\beta}\right). \tag{2.8}$$

In particular, the profile likelihood estimatos for $m(z)$ is

$$\hat{m}(z) = m_{\hat{\theta}}(z) = s(z)'\left(Y - D\hat{\alpha} - X\hat{\beta}\right). \tag{2.9}$$

To proceed, let us make two remarks. (a) We can also use local constant estimator for $m(z)$. It is well known that the bias of the local constant estimator is higher for z near the boundary than for z away from the boundary. One needs to use either a "trimming" scheme (e.g., Robinson (1988)), or some weight functions (Li and Stengos (1996)) to avoid the random denominator problem in the local constant estimation. (b) From the perspective of efficient estimation, one can minimize a weighted least squares function in (2.5) by taking into account the dependence structure in the error term v_{it}. But early studies in the panel data models with random effects (e.g., Lin and Carroll (2001)) suggest that one cannot gain by this approach.

3 Asymptotic Properties for the Estimators

In this section we first state a set of assumptions that are used to establish asymptotic properties of the estimators proposed in the previous section. We then study the consistency and asymptotic normality of the proposed estimators in the partially linear panel data models.

3.1 Assumptions

To provide a rigorous analysis, we make the following assumptions:

A1. $(\alpha_i, v_i, x_i, z_i)$, $i = 1, ..., n$, are i.i.d., where $v_i = (v_{i1}, ..., v_{iT})'$ and x_i and z_i are similarly defined. $E\|x_{it}\|^{2+\delta} < \infty$ and $E|v_{it}|^{2+\delta} < \infty$ for some $\delta > 0$. Let $\sigma^2(x, z) = \text{var}(y_{it}|x_{it} = x, z_{it} = z)$

and $\sigma^2(z) = \text{var}(x_{it}|z_{it} = z)$. $\sigma^2(x, z)$ and $\sigma^2(z)$ are uniformly bounded from above from infinity and below from 0.

A2. $E(y_{it}|x_i, z_i, \alpha_i) = E(y_{it}|x_{it}, z_{it}, \alpha_i) = \alpha_i + x'_{it}\beta + m(z_{it})$.

A3. The random vector z_{it} has a continuous density function $f_t(\cdot)$ with compact support C_f on $\mathbf{R}^q$. $f_t(\cdot)$ is bounded away from zero and infinity on C_f for each $t = 1, ..., T$.

A4. Let $p(z) = E(x_{it}|z_{it} = z)$. The functions $m(\cdot)$ and $p(\cdot)$ have bounded second partial derivatives on C_f.

A5. Let $\tilde{x}_{it} = x_{it} - E(x_{it}|z_{it})$. $\Phi = \sum_t E\{\tilde{x}_{it}[\tilde{x}_{it} - \sum_s \tilde{x}_{is}/T]'\}$ is positive definite.

A6. The kernel function $K(\cdot)$ is a continuous density with compact support on $\mathbf{R}^q$. All odd order moments of K vanish.

A7. As $n \to \infty, \|H\| \to 0, n|H|^2 \to \infty$, $\|H\|^4 |H|^{-1} \to 0$ and $n|H| \, ||H||^4 \to c \in [0, \infty)$.

Assumption A1 is standard in the literature. A2 is assumed to simplify the proof and it is also assumed in Lin and Carroll (2001) and Hu et al. (2004). It implies $E(v_{it}|x_i, z_i, \alpha_i) = E(v_{it}|x_{it}, z_{it}) = 0$. A3 and A4 are standard in the literature on local linear/polynomial estimation. See Fan and Gijbels (1996). A5 rules out time-invariant terms in x_{it}. As Hall et al. (1999) remark, the requirement that K is compactly supported in A6 can be removed at the cost of lengthier arguments used in the proofs, and in particular, Gaussian kernel is allowed. A7 is easily satisfied by considering $H = diag(h_1, ..., h_q)$ with $h_i \propto n^{-1/(4+q)}$ for $q < 4$. When $q \geq 4$, higher order local polynomial can be used to achieve bias reduction. Nevertheless, due to the "curse of dimensionality", we don't expect large q in practice.

3.2 Asymptotic properties of $\hat{\beta}$ and $\hat{M}(z)$

Theorem 3.1 *(i) Under Assumptions A1-A7,*

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N(0, \Sigma), \tag{3.1}$$

where $\Sigma = \Phi^{-1}\Omega\Phi^{-1}$, *and* $\Omega = \sum_s \sum_t E\{\tilde{x}_{it}[\tilde{x}_{is} - \sum_l \tilde{x}_{il}/T]' v_{it} v_{is}\}$.

(ii) A consistent estimator of Σ *is given by* $\hat{\Sigma} = \hat{\Phi}^{-1}\hat{\Omega}\hat{\Phi}^{-1}$, *where* $\hat{\Phi} = (nT)^{-1} \sum_i \sum_t \overrightarrow{x}_{it}(\overrightarrow{x}_{it} - \sum_l \overrightarrow{x}_{il}/T)'$, $\hat{\Omega} = n^{-1} \sum_i \sum_t \sum_s \overrightarrow{x}_{it}(\overrightarrow{x}_{is} - \sum_l \overrightarrow{x}_{il}/T)' \hat{b}_{it}\hat{b}_{is}$, $\overrightarrow{x}_{it} = x_{it} - s^{(1)}(z_{it})' X$, *and* $\hat{b}_{it} = y_{it} - x'_{it}\hat{\beta} - \hat{m}(z_{it}) - \hat{b}_i$.

The proof is given in the Appendix. To study the asymptotic property of $\hat{M}(z)$, let $\overline{f}(z) = \sum_{t=1}^{T} f_t(z)$, $\tilde{v}_{it} = v_{it} - T^{-1} \sum_{s=1}^{T} v_{is}$, $\sigma_t^2(z) = E\left(\tilde{v}_{it}^2 | z_{it} = z\right)$, and $\overline{\sigma}^2(z) = \sum_{t=1}^{T} \sigma_t^2(z) f_t(z)$.

Theorem 3.2 *Under Assumptions A1-A7,*

$$\sqrt{n|H|}\left(\begin{pmatrix} \hat{M}(z) - M(z) \\ \dot{\hat{m}}(z) - \dot{m}(z) \end{pmatrix} - Q^{-1}\begin{pmatrix} \frac{1}{2} tr \int_{\mathbf{R}^q} uu'K(u)\,du H \ddot{m}(z) H \\ 0 \end{pmatrix}\right) \xrightarrow{d} N\left(0, Q^{-1}\Gamma Q^{-1}\right), \tag{3.2}$$

where $\ddot{m}(z)$ *is the second order derivative matrix of* $m(\cdot)$ *at* z,

$$Q = \overline{f}(z)\begin{pmatrix} 1 & 0' \\ 0 & \int_{\mathbf{R}^q} uu'K(u)\,du \end{pmatrix}, \text{ and } \Gamma = \overline{\sigma}^2(z)\begin{pmatrix} \int_{\mathbf{R}^q} K(u)^2\,du & 0' \\ 0 & \int_{\mathbf{R}^q} uu'K(u)^2\,du \end{pmatrix}. \tag{3.3}$$

Remark 1. Theorem 3.2 tells us that $\hat{M}(z)$ is asymptotically distributed as if the finite dimensional parameters $\theta = (\alpha', \beta')'$ is known. In particular the estimator for $m(z)$ are asymptotically independent of the estimator for $\dot{m}(z)$. When K is product kernel of k that is a symmetric density on $\mathbf{R}$, the bias and variance formulae of $\hat{M}(z)$ can be simplified. Furthermore, the asymptotic

normal distribution given by Theorem 3.2 can be used to calculate pointwise confidence intervals for $M(z)$. The procedure is standard and we omit it for brevity.

Remark 2. Baltagi and Li (2002) obtain consistent estimators for β and $m(z)$ by taking the first difference to eliminate the fixed effects and using series approximation for the nonparametric component. They establish asymptotic normality for the their estimator of β and consistency for their estimator of $m(z)$. In the simplest case where (x_{it}, z_{it}, v_{it}) are i.i.d. across both i and t and v_{it} are conditional homoskedastic: $E\left(v_{it}^2 | x_{it}, z_{it}\right) = \sigma^2$, our estimator $\hat{\beta}$ for β and theirs share the same asymptotic variance $\sigma^2/(T-1)\left[E\left(\tilde{x}_{it}\tilde{x}_{it}'\right)\right]^{-1}$. Yet we also prove the asymptotic normality for both estimators of $m(z)$ and $\dot{m}(z)$.

4 Appendix

We first provide some lemmas that are used in the proof of the main theorems in the text. Note that $S = (s_{11}, ..., s_{1T}, s_{21}, ..., s_{nT})'$, where $s_{it} = s(z_{it})$. Denote a typical entry of $s(z)$ by $s(z_{it}, z) = e'\left[\overrightarrow{Z}(z)' \mathbf{K}_H(z) \overrightarrow{Z}(z)\right]^{-1} \overrightarrow{Z}_{it}(z) K_H(z_{it} - z)$, where $\overrightarrow{Z}_{it}(z)$ is a typical column of $\overrightarrow{Z}(z)'$: $\overrightarrow{Z}(z)' = \left[\overrightarrow{Z}_{11}(z), ..., \overrightarrow{Z}_{1T}(z), \overrightarrow{Z}_{21}(z), ... \overrightarrow{Z}_{nT}(z)\right]$. Let $\hat{p}(z) = s(z)' X$, $\hat{q}(z) = s(z)' m(Z)$, $\hat{b}(z) = s(z)' V$, and $P = (I_{nT} - S)'(I_{nT} - S)$. Let $A \approx B$ denote $A = B(1 + o_p(1))$ componentwise for any matrices A, B of the same dimension. Let C signify a generic constant whose exact value may vary from case to case.

Lemma 4.1 *(a)* $s(z_{it}, z) = n^{-1} K_H(z_{it} - z) \overline{f}^{-1}(z) \{1 + o_p(1)\}$, *where* $\overline{f}(z) = \sum_{t=1}^{T} f_t(z)$.

(b) There is a positive constant C such that $\lim_n P\left(n^{-1}\left|\overrightarrow{Z}(z)' \mathbf{K}_H(z) \overrightarrow{Z}(z)\right|_{ij} \leq C \text{ for } z \in C_f, i, j = 1, ..., q+1\right\} = 1$.

(c) There is a positive constant C such that $\lim_n P\left(\sup_z \max_{1 \leq i \leq n, 1 \leq t \leq T} |s(z_{it}, z)| \leq C n^{-1} |H|^{-1}\right) = 1$.

Proof. By the law of large numbers and Assumptions A1-A7, $n^{-1}\overrightarrow{Z}(z)' \mathbf{K}_H(z) \overrightarrow{Z}(z) = Q + o_p(1)$, where Q is defined in (3.3). Consequently, $s(z_{it}, z) = n^{-1} e'\left[n^{-1}\overrightarrow{Z}(z)' \mathbf{K}_H(z) \overrightarrow{Z}(z)\right]^{-1} \overrightarrow{Z}_{it}(z) K_H(z_{it} - z) = n^{-1} K_H(z_{it} - z) \overline{f}^{-1}(z) \{1 + o_p(1)\}$. Par (b) follows from (a) and the Hoeffding's inequality. Part (c) follows from (a) and (b). ■

Lemma 4.2 $(D'PD)^{-1} = T^{-1} I_n + O_p\left(i_n i_n' n^{-1} |H|^{-1} \sqrt{\ln n}\right)$ *for sufficiently large n, where i_n is a $n \times 1$ column vector of ones,.*

Proof. Write $D'PD = D'(I_{nT} - S)'(I_{nT} - S) D = DD' - D'SD - D'S'D + D'S'SD \equiv A_{11} - A_{12} - A_{13} + A_{14}$. First, $A_{11} = T I_n$. By Lemma 4.1, the (i,j)th element of A_{12} is $[A_{12}]_{ij} = \sum_{t=1}^{T} \sum_{s=1}^{T} s(z_{is}, z_{jt})' = O_p\left(n^{-1} |H|^{-1}\right)$. Similarly, $[A_{1l}]_{ij} = O_p\left(n^{-1} |H|^{-1}\right)$, $l = 3, 4$. Consequently, $[D'PD]_{ij} = [T I_n]_{ij} + O_p\left(n^{-1} |H|^{-1}\right)$. Following the proof of Lemma 3.1 in Opsomer and Ruppert (1997), we can prove that the above equality holds uniformly for all i, j and with $O_p(\cdot)$ replaced by $O_{as}(\cdot)$.

Let $A = (A_{12} + A_{13} - A_{14})/T$. Write $\rho(A)$ for the spectral radius of A. Then for large enough n, $\rho(A) < 1$ so that $I_n - A$ is invertible a.s.. Assume now that $(I_n - A)^{-1}$ exists for $n \geq N$. We can write $(I_n - A)^{-1} = I_n + \sum_{l=1}^{\infty} A^l$. It is easy to show by induction that for all $l = 1, 2, ..., \max_{i,j} \left|A^l\right|_{ij} \leq (1 - \epsilon)^l n^{-1} |H|^{-1} \sqrt{\ln n}$ for some $\epsilon > 0$. So $\max_{i,j} \left|(I_n - A)^{-1} - I_n\right|_{i,j} \leq C n^{-1} |H|^{-1} \sqrt{\ln n}$ for $n > N$. The desired result follows. ■

Lemma 4.3 *(a)* $n^{-1}X'PX \xrightarrow{p} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[(x_{it} - p(z_{it}))(x_{it} - p(z_{it}))'\right]$.
(b) $n^{-1}X'PDD'PX \xrightarrow{p} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[(x_{it} - p(z_{it}))(x_{is} - p(z_{is}))'\right]$.

Proof. To prove (a), write

$$\begin{aligned}
& n^{-1}X'PX \\
= \; & n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \hat{p}(z_{it}))(x_{it} - \hat{p}(z_{it}))' \\
= \; & n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - p(z_{it}))(x_{it} - p(z_{it}))' + n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (p(z_{it}) - \hat{p}(z_{it}))(p(z_{it}) - \hat{p}(z_{it}))' \\
& + n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - p(z_{it}))(p(z_{it}) - \hat{p}(z_{it}))' + n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (p(z_{it}) - \hat{p}(z_{it}))(x_{it} - p(z_{it}))' \\
\equiv \; & A_{21} + A_{22} + A_{23} + A_{23}'.
\end{aligned}$$

By the law of large numbers, $A_{21} \xrightarrow{p} \sum_{t=1}^{T} E\left[(x_{it} - p(z_{it}))(x_{it} - p(z_{it}))'\right]$. By Masry (1996) or Hamilton and Truong (1997, Lemma 4) and Assumptions A1-A7, $A_{22} = O_p\left(n^{-1}|H|^{-1} + \|H\|^4\right) = o_p(1)$. Then by the Cauchy-Schwartz inequality, $A_{23} = o_p(1)$. The result follows.

To prove (b), write

$$\begin{aligned}
n^{-1}X'PDD'PX &= n^{-1}X'(I_{nT} - S)'(I_{nT} - S)DD'(I_{nT} - S)'(I_{nT} - S)X \\
&= n^{-1}X'(I_{nT} - S)'DD'(I_{nT} - S)X - n^{-1}X'(I_{nT} - S)'DD'S'(I_{nT} - S)X \\
&\quad - n^{-1}X'(I_{nT} - S)'SDD'(I_{nT} - S)X + n^{-1}X'(I_{nT} - S)'SDD'S'(I_{nT} - S)X \\
&\equiv A_{31} - A_{32} - A_{33} + A_{34}.
\end{aligned}$$

Using an argument similar to the above, $A_{31} = n^{-1}X'(I_{nT} - S)'DD'(I_{nT} - S)X = n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} (x_{it} - \hat{p}(z_{it}))(x_{is} - \hat{p}(z_{is}))' \xrightarrow{p} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[(x_{it} - p(z_{it}))(x_{is} - p(z_{is}))'\right]$. To show $A_{32} = o_p(1)$, we only consider the case $p = 1$ because the vector x_{it} case follows from the same proof as the scalar x_{it} case and the Cauchy inequality. $|A_{32}| = \left|n^{-1}X'(I_{nT} - S)'DD'S'(I_{nT} - S)X\right| \leq \left|n^{-1/2}D'(I_{nT} - S)X\right| \left|n^{-1/2}D'S'(I_{nT} - S)X\right|$. To show $A_{32} = o_p(1)$, by (a) it suffices to show $||n^{-1/2}D'S'(I_{nT} - S)X|| = n^{-1} \sum_{j=1}^{n} \left[\sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} (x_{it} - \hat{p}(z_{it}))\, s(z_{js}, z_{it})\right]^2 = o_p(1)$. By Lemma 4.1, $\sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \hat{p}(z_{it}))\, s(z, z_{it}) \approx n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \hat{p}(z_{it}))\, K_H(z - z_{it})\, \overline{f}^{-1}(z_{it})$ $= n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - p(z_{it}))\, K_H(z - z_{it})\, \overline{f}^{-1}(z_{it}) + n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (p(z_{it}) - \hat{p}(z_{it}))\, K_H(z - z_{it})\, \overline{f}^{-1}(z_{it}) \equiv A_{41}(z) + A_{42}(z)$. By the argument of Masry (1996), $\sup |A_{41}(z)| = O_p\left(n^{-1/2}|H|^{-1/2}\sqrt{\ln n}\right) = o_p(1)$, and $\sup_z |A_{42}(z)| \leq \sup_z |p(z) - \hat{p}(z)| \sup_z \left|n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} K_H(z - z_{it})\, \overline{f}^{-1}(z_{it})\right| = O_p\left(n^{-1/2}|H|^{-1/2}\sqrt{\ln n} + \|H\|^2\right) = o_p(1)$. So $\sup_z \left|\sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \hat{p}(z_{it}))\, s(z, z_{it})\right| = o_p(1)$ uniformly in z and $A_{32} = o_p(1)$. Similarly, one can show $A_{3j} = o_p(1)$, $j = 3, 4$. This concludes the proof of (b). ■

Lemma 4.4 $n^{-1}X^{*\prime}M^*X^* \xrightarrow{p} \Phi$.

Proof. By Lemmas 4.2 and 4.3, $n^{-1}X^{*\prime}M^*X^* = n^{-1}X^{*\prime}\left[I_{nT} - D^*(D^{*\prime}D^*)^{-1}D^{*\prime}\right]X^* = n^{-1}X'PX - n^{-1}X'PD(D'PD)^{-1}D'PX \approx n^{-1}X'PX - n^{-1}T^{-1}X'PDD'PX \xrightarrow{p} \Phi$. ■

Lemma 4.5 $n^{-1/2}X^{*\prime}M^*(I_{nT} - S)\,m(Z) = o_p(1)$.

Proof. By Lemma 4.2, $n^{-1/2}X^{*\prime}M^*\left(I_{nT}-S\right)m\left(Z\right)=n^{-1/2}X^{*\prime}(I_{nT}-D^*(D^{*\prime}D^*)^{-1}D^{*\prime})(I_{nT}-S)m\left(Z\right)=n^{-1/2}X'Pm\left(Z\right)-n^{-1/2}X'PD\left(D'PD\right)^{-1}D'Pm\left(Z\right)\approx n^{-1/2}X'Pm\left(Z\right)-n^{-1/2}T^{-1}X'PDD'Pm\left(Z\right)=A_{51}-A_{52}$. So it suffices to show $A_{5i}=o_p\left(1\right),\ i=1,2$. $A_{51}=n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[x_{it}-p\left(z_{it}\right)\right]\left[m\left(z_{it}\right)-\hat{\phi}\left(z_{it}\right)\right]+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[p\left(z_{it}\right)-\hat{\phi}\left(z_{it}\right)\right]\left[m\left(z_{i}\right)-\hat{\phi}\left(z_{it}\right)\right]\equiv A_{61}+A_{62}$. By Masry (1996) and Assumptions A1-A7, $\left\|A_{62}\right\|\leq n^{1/2}O_p\left(n^{-1}\left|H\right|^{-1}+\left\|H\right\|^4\right)=o_p\left(1\right)$. Next, by Assumptions A1-A2 and the law of iterated expectation, $E\left[A_{61}\right]=0$ and for $p=1$,

$$\begin{aligned}
E\left[A_{61}\right]^2 &= n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\{\left[x_{it}-p\left(z_{it}\right)\right]\left[x_{is}-p\left(z_{is}\right)\right]\left[m\left(z_{it}\right)-\hat{\phi}\left(z_{it}\right)\right]\left[m\left(z_{is}\right)-\hat{\phi}\left(z_{is}\right)\right]\right\}\\
&\quad+n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\{\left[x_{it}-p\left(z_{it}\right)\right]\left[x_{js}-p\left(z_{js}\right)\right]\left[m\left(z_{it}\right)-\hat{\phi}\left(z_{it}\right)\right]\left[m\left(z_{is}\right)-\hat{\phi}\left(z_{is}\right)\right]\right\}\\
&= O\left(n^{-1}\left|H\right|^{-1}\ln n+\left\|H\right\|^4\right)+0=o(1).
\end{aligned}$$

It follows by the Chebyshev inequality that $A_{61}=o_P\left(1\right)$ and so $A_{51}=o_P\left(1\right)$. The result for $p>1$ case then follows by an application of the Cauchy inequality. Similarly, $A_{52}=o_P\left(1\right)$. Consequently, $n^{-1/2}X^{*\prime}M^*\left(I_{nT}-S\right)m\left(Z\right)=o_p\left(1\right)$. ■

Lemma 4.6 *(a)* $n^{-1/2}X'PV=n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-p\left(z_{it}\right)\right)v_{it}+o_p\left(1\right)$.
(b) $n^{-1/2}X'PDD'PV=n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(x_{it}-p\left(z_{it}\right)\right)v_{is}+o_p\left(1\right)$.

Proof.

$$\begin{aligned}
n^{-1/2}X'PV &= n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-\hat{\phi}\left(z_{it}\right)\right)\left(v_{it}-\hat{\phi}\left(z_{it}\right)\right)\\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-p\left(z_{it}\right)\right)v_{it}-n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-p\left(z_{it}\right)\right)\hat{\phi}\left(z_{it}\right)\\
&\quad+n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(p\left(z_{it}\right)-\hat{\phi}\left(z_{it}\right)\right)v_{it}+n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(p\left(z_{it}\right)-\hat{\phi}\left(z_{it}\right)\right)\hat{\phi}\left(z_{it}\right).
\end{aligned}$$

It is straightforward to show that the last three terms is $o_p\left(1\right)$ by using arguments analogous to those in the proof of Lemma 4.5. Similarly, we can show $n^{-1/2}X'PDD'PV=n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t=1}^{T}\left(x_{it}-\hat{\phi}\left(z_{it}\right)\right)\left(v_{is}-\hat{\phi}\left(z_{is}\right)\right)=n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(x_{it}-p\left(z_{it}\right)\right)v_{is}+o_p\left(1\right)$. ■

Lemma 4.7 $n^{-1/2}X^{*\prime}M^*\left(I_{nT}-S\right)V\xrightarrow{d}N\left(0,\Omega\right)$.

Proof. By Lemma 4.2, $n^{-1/2}X^{*\prime}M^*\left(I_{nT}-S\right)V=n^{-1/2}X^{*\prime}(I_{nT}-D^*(D^{*\prime}D^*)^{-1}D^{*\prime})\left(I_{nT}-S\right)V=n^{-1/2}X'P\left(I_{nT}-D\left(D'PD\right)^{-1}D'P\right)V\approx n^{-1/2}X'P\left(I_{nT}-T^{-1}DD'P\right)V$. Let $\tilde{e}_{it}=v_{it}-T^{-1}\sum_{s=1}^{T}v_{is}$. By Lemma 4.6 and the central limit theorem $n^{-1/2}X'(PI_{nT}-T^{-1}DD'P)V=n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}x_{it}\tilde{e}_{it}+o_p\left(1\right)\xrightarrow{d}N\left(0,\Omega\right)$. ■

Proof of Theorem 3.1

(i) Noting that $M^*D^*=0$, $\hat{\beta}-\beta=\left(X^{*\prime}M^*X^*\right)^{-1}X^{*\prime}M^*Y^*-\beta=\left(X^{*\prime}M^*X^*\right)^{-1}X^{*\prime}M^*\left[D^*\alpha+X^*\beta+\left(I_{nT}-S\right)\left(m\left(Z\right)+V\right)\right]-\beta=\left(X^{*\prime}M^*X^*\right)^{-1}X^{*\prime}M^*\left(I_{nT}-S\right)\left(m\left(Z\right)+V\right)$. Thus by Lemmas 4.4, 4.5, 4.7, and the CLT, $\sqrt{n}\left(\hat{\beta}-\beta\right)=\left(n^{-1}X^{*\prime}M^*X^*\right)^{-1}n^{-1/2}X^{*\prime}M^*\left(I_{nT}-S\right)V$

$$+\left(n^{-1}X^{*\prime}M^{*}X^{*}\right)^{-1}n^{-1/2}X^{*\prime}M^{*}\left(I_{nT}-S\right)m\left(Z\right)=\Phi^{-1}n^{-1/2}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\varepsilon_{it}v_{it}-T^{-1}\sum_{s=1}^{T}v_{is}\right)\right)$$
$$+o_{p}\left(1\right)\xrightarrow{d}N\left(0,\Phi^{-1}\Omega\Phi^{-1}\right).$$

(ii) It suffices to show $\hat{\Phi}=\Phi+o_{p}\left(1\right)$ and $\hat{\Omega}=\Omega+o_{p}\left(1\right)$. The first part follows from the arguments of Lemma 4.3. Let $\tau_{n}=n^{-1/2}\left|H\right|^{-1/2}\sqrt{\ln\left(n\right)+\left\|H\right\|^{2}}$. By Theorem 3.2 and standard arguments for uniform convergence (e.g., Masry (1996)), $\hat{m}\left(z\right)-m\left(z\right)=O_{p}\left(\tau_{n}\right)$ uniformly in z. By (2.6), (2.7) and (3.1) it is easy to show $\hat{\varepsilon}_{it}=v_{it}+o_{p}\left(1\right)$ uniformly. Also, uniformly in i and t, $\overrightarrow{\tilde{x}}_{it}=\tilde{x}_{it}-s\left(z_{it}\right)'\tilde{X}=\varepsilon_{it}+O_{p}\left(\tau_{n}\right)$. Consequently, by the law of large numbers $\hat{\Omega}=n^{-1}\sum_{i}\sum_{t}\sum_{s}\varepsilon_{it}\left(\varepsilon_{is}-\sum_{l}\varepsilon_{il}/T\right)'v_{it}v_{is}+o_{p}\left(1\right)\xrightarrow{p}\Omega.$ ¥

Proof of Theorem 3.2

Denote $S\left(z_{it},z\right)$ as a typical column of $S\left(z\right)$, i.e., $S\left(z\right)=\left(S\left(z_{11},z\right),...,S\left(z_{1T},z\right),...,S\left(z_{nT},z\right)\right)$. By (2.7) (2.9) and Lemma 4.2,

$$\begin{aligned}
\hat{M}\left(z\right) &= S\left(z\right)\left(Y-D\hat{\alpha}-X\hat{\beta}\right)\\
&= S\left(z\right)\left(I_{nT}-D\left(D'PD\right)^{-1}D'P\right)\left(Y-X\hat{\beta}\right)\\
&= S\left(z\right)\left(I_{nT}-D\left(D'PD\right)^{-1}D'P\right)\left\{\left(Y-X\beta\right)-X\left(\hat{\beta}-\beta\right)\right\}\\
&= S\left(z\right)\left(I_{nT}-D\left(D'PD\right)^{-1}D'P\right)\left\{\left(D\alpha+m\left(Z\right)+V\right)-X\left(\hat{\beta}-\beta\right)\right\}\\
&\approx S\left(z\right)\left(I_{nT}-T^{-1}DD'P\right)\left(m\left(Z\right)+V-X\left(\hat{\beta}-\beta\right)\right)
\end{aligned}$$

Recall $M\left(z\right)\equiv\left(m\left(z\right),\left(H\dot{m}\left(z\right)\right)'\right)'$. By the second order Taylor expression, $m\left(z_{it}\right)=\overrightarrow{Z}_{it}\left(z\right)'M\left(z\right)+\frac{1}{2}\left(z_{it}-z\right)'\ddot{m}\left(z\right)\left(z_{it}-z\right)+o_{P}\left(\left\|H\right\|^{2}\right)$. So $\sqrt{n\left|H\right|}\left(\hat{M}\left(z\right)-M\left(z\right)\right)\approx\frac{\sqrt{n\left|H\right|}}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}S\left(z_{it},z\right)\left(z_{it}-z\right)'\ddot{m}\left(z\right)\left(z_{it}-z\right)+\sqrt{n\left|H\right|}S\left(z\right)\left(I_{nT}-T^{-1}DD'P\right)V-\sqrt{n\left|H\right|}T^{-1}S\left(z\right)DD'Pm\left(Z\right)-\sqrt{n\left|H\right|}S\left(z\right)\left(I_{nT}-T^{-1}DD'P\right)X\left(\hat{\beta}-\beta\right)\equiv B_{11}+B_{12}-B_{13}-B_{14}$. It is easy to show $B_{13}=o_{p}\left(1\right)$ and $B_{14}=o_{p}\left(1\right)$ by using results in Lemma 4.1, Lemma 4.3, Lemma 4.5 and Theorem 3.1. Recall $\left[n^{-1}\overrightarrow{Z}\left(z\right)'\mathbf{K}_{H}\left(z\right)\overrightarrow{Z}\left(z\right)\right]^{-1}=Q^{-1}+o_{P}\left(1\right)$, $B_{11}=\frac{\sqrt{n\left|H\right|}}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}S\left(z_{it},z\right)\left(z_{it}-z\right)'\ddot{m}\left(z\right)\left(z_{it}-z\right)\approx\frac{\sqrt{n\left|H\right|}}{2}Q^{-1}n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\overrightarrow{Z}_{it}\left(z\right)K_{H}\left(z_{it}-z\right)\left(z_{it}-z\right)'\ddot{m}\left(z\right)\left(z_{it}-z\right)=\frac{\sqrt{n\left|H\right|}}{2}Q^{-1}\left(tr\int_{\mathbf{R}^{q}}uu'K\left(u\right)du\,H\ddot{m}\left(z\right)H\right)+o_{p}\left(1\right)$. $B_{12}=\sqrt{n\left|H\right|}S\left(z\right)\left(I_{nT}-T^{-1}DD'P\right)V\approx\sqrt{n\left|H\right|}Q^{-1}n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\overrightarrow{Z}_{it}\left(z\right)K_{H}\left(z_{it}-z\right)\varepsilon_{it}\xrightarrow{d}N\left(0,Q^{-1}\Gamma Q^{-1}\right)$. This completes the proof. ¥

REFERENCES

Baltagi, B. H. and D. Li (2002) Series Estimation of Partially Linear Panel Data Models with Fixed Effects. *Annals of Economic and Finance* 3, 103-116.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications.* Champman and Hall.

Hall, P., R. C. L. Wolf, and Q. Yao (1999), Methods of estimating a conditional distribution function, *Journal of the American Statistical Association, 94*, 154-163.

Hamilton, S. A. and Y. K. Truong (1997) Local Linear Estimation in Partly Linear Models. *Journal of Multivariate Analysis* 60, 1-19.

Hu, Z., N. Wang and R. J. Carroll (2004) Profile-Kernel versus Backfitting in the Partially Linear Models for Longitudinal/Clustered Data. *Biometrika* 91, 251-262.

Li, Q. and T. Stengos (1996) Semiparametric Estimation of Partially Linear Panel Data Models. *Journal of Econometrics* 71, 289-397.

Lin, X. and R. J. Carroll (2001) Semiparametric Regression for Clustered Data Using Generalized Estimating Equations. *Journal of the American Statistical Association* 96, 1045-1055.

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

Mundra, K. (2005) Nonparametric Slope Estimation for Fixed-Effect Panel Data. *Working Paper*, Department of Economics, San Diego State University.

Opsomer J. D. and D. Ruppert (1997) Fitting a Bivariate Additive Model by Local Polynomial Regression. *Annals of Statistics* 25, 186-211.

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