

Profile likelihood estimation of partially linear panel data models with fixed effects[☆]

Liangjun Su^{a,*}, Aman Ullah^{b,1}

^a *Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, P. R. China*

^b *Department of Economics, University of California, Riverside, CA 92521-0427, USA*

Received 7 June 2005; received in revised form 18 December 2005; accepted 6 January 2006

Available online 2 May 2006

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.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Fixed effects; Panel data; Partially linear models; Profile likelihood

JEL classification: C13; C14; C23

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

[☆] The first author gratefully acknowledges the financial support from the NSFC (Project 705010001). The second author gratefully acknowledges the financial support from the Academic Senate, UCR.

* Corresponding author. Tel.: +86 10 62767444.

E-mail addresses: lsu@gsm.pku.edu.cn (L. Su), aman.ullah@ucr.edu (A. Ullah).

¹ Tel.: +1 951 827 1591.

establish asymptotic normality for the finite dimensional parameter of interest in the model and consistency for the nonparametric object. Although their approach overcomes several drawbacks associated with the kernel approach of Li and Stengos (1996), they do not estimate the slope parameter for the nonparametric component. Recently, Mundra (2005) considered nonparametric estimation of the slope parameter for fixed-effect panel data models, but with her approach, we cannot obtain estimators for the nonparametric regression function.

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 a consistent estimator of the parameter of interest. Meanwhile, we obtain consistent estimators for both the nonparametric regression function 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 their asymptotic properties in Section 3. All technical details are relegated to the Appendix. Throughout the paper, we denote the norm of a matrix A by $\|A\| = \{tr(A'A)\}^{1/2}$, where prime means transpose. Let I_n denote the $n \times n$ identity matrix and i_n denote the $n \times 1$ vector of ones.

2. The model and estimators

Consider the following partially linear model with fixed effects:

$$y_{it} = \alpha_i + x'_{it}\beta + m(z_{it}) + v_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (2.1)$$

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, α_i 's are fixed effects, and v_{it} 's are the random disturbances. For identification purpose, we impose $\sum_{i=1}^n \alpha_i = 0$. 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 \mathbb{R}^q and $H = \text{diag}(h_1, \dots, h_q)$, a matrix of bandwidth sequences. Set $K_H(z) = |H|^{-1} K(H^{-1}z)$, where $|H|$ is the determinant of H . Let $Z_{it}(z) = [1, \{H^{-1}(z_{it} - z)\}']'$. Further denote $\mathbf{K}_H(z) = \text{diag}(K_H(z_{11} - z), \dots, K_H(z_{1T} - z), K_H(z_{21} - z), \dots, K_H(z_{nT} - z))$ and $\vec{Z}(z) = (Z_{11}(z), \dots, Z_{1T}(z), Z_{21}(z), \dots, Z_{nT}(z))'$.

Let $\alpha = (\alpha_2, \dots, \alpha_n)'$ and $\theta = (\alpha', \beta')'$. Given θ , we estimate $M(z) \equiv (m(z), (H\dot{m}(z))')$ by

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

where $Y = (y_{11}, \dots, y_{1T}, y_{21}, \dots, y_{nT})'$, $X = (x_{11}, \dots, x_{1T}, x_{21}, \dots, x_{nT})'$, $D = (I_n \otimes i_T) d_n$, $d_n = [-i_n - 1 I_{n-1}]'$. Define the smoothing operator by $S(z) = [\vec{Z}(z)' \mathbf{K}_H(z) \vec{Z}(z)]^{-1} \vec{Z}(z)' \mathbf{K}_H(z)$.

Then

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

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

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

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

The parameter θ is then estimated by the *profile likelihood* method (more precisely, it is a *profile least squares* method in the current context):

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

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

$$\hat{\beta} = [X^{*'}M^{*}X^{*}]^{-1}X^{*'}M^{*}Y^{*}, \quad (2.6)$$

$$\hat{\alpha} = (\hat{\alpha}_2, \dots, \hat{\alpha}_n)' = [D^{*'}D^{*}]^{-1}D^{*'}(Y^{*} - X^{*}\hat{\beta}), \quad (2.7)$$

where $D^{*} = (I_{nT} - S)D$, $Y^{*} = (I_{nT} - S)Y$, $X^{*} = (I_{nT} - S)X$, $M^{*} = I_{nT} - D^{*}[D^{*'}D^{*}]^{-1}D^{*}$, $S = (s_{11}, \dots, s_{1T}, s_{21}, \dots, s_{nT})'$, and $s_{it} = s(z_{it})$. α_1 is estimated by $\hat{\alpha}_1 = -\sum_{i=2}^n \hat{\alpha}_i$.

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

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

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

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

We can also use local constant estimator for $m(z)$. But it is well known that the local constant estimator is subject to the boundary bias problem.

3. Asymptotic properties for the estimators

In this section we first state assumptions that are used to establish asymptotic properties of the proposed estimators. We then study the asymptotic normality of the proposed estimators.

3.1. Assumptions

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

A1. $(\alpha_i, v_i, x_i, z_i)$, $i=1, \dots, n$, are i.i.d., where $v_i = (v_{i1}, \dots, v_{iT})'$ 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. z_{it} has a continuous density function $f_i(\cdot)$ with compact support C_f on \mathbb{R}^q . $f_i(\cdot)$ is bounded away from zero and infinity on C_f for each $t=1, \dots, T$.

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 \mathbb{R}^q . All odd order moments of K vanish.

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

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 polynomial estimation (Fan and Gijbels (1996)). A5 rules out time-invariant terms in x_{it} . The requirement that K is compactly supported in A6 can be removed at the cost of lengthier arguments used in the proofs. A7 is easily satisfied by considering $H = \text{diag}(h_1, \dots, h_q)$ 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 do not expect large q in practice.

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

Theorem 3.1.

(i) Under Assumptions A1–A7

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

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

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

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

Theorem 3.2. Under Assumptions A1–A7,

$$\sqrt{n|H|} \left(\hat{M}(z) - M(z) - Q^{-1} \left(\frac{\tilde{f}(z)}{0} \text{tr} \left(\int_{\mathbb{R}^q} u u' K(u) du H \ddot{m}(z) H \right) \right) \right) \xrightarrow{d} N(0, Q^{-1} \Gamma Q^{-1}), \quad (3.2)$$

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

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

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 $\dot{m}(z)$, and they have different rates of convergence (see the definition of $M(z)$). Furthermore, the asymptotic normal distribution given by Theorem 3.2 can be used to calculate pointwise confidence intervals for $M(z)$.

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 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: $E(v_{it}^2 | x_{it}, z_{it}) = \sigma^2$, our estimator $\hat{\beta}$ for β and theirs share the same asymptotic variance $(\sigma^2 / (T-1)) [E(\tilde{x}_{it} \tilde{x}_{it}')^{-1}]$. Yet we also prove the asymptotic normality for both estimators of $m(z)$ and $\dot{m}(z)$.

Appendix A

We first provide some lemmas that are used in the proof of the main theorems in the text. Note that $S = (s_{11}, \dots, s_{1T}, s_{21}, \dots, s_{nT})'$, where $s_{it} = s(z_{it})$. Denote a typical entry of $s(z)$ by $s(z_{it}, z) = e' [\bar{Z}(z)' K_H(z) \bar{Z}(z)]^{-1} \bar{Z}_{it}(z) K_H(z_{it} - z)$, where $\bar{Z}_{it}(z)$ is a typical column of $\bar{Z}(z)'$: $\bar{Z}(z)' = (\bar{Z}_{11}(z), \dots, \bar{Z}_{1T}(z), \bar{Z}_{21}(z), \dots, \bar{Z}_{nT}(z))$. Let $P = (I_{nT} - S)'(I_{nT} - S)$.

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 positive constant whose exact value may vary from case to case. We state some lemmas, the proof of which is available upon request.

Lemma 4.1.

- (a) $s(z_{it}, z) = n^{-1} K_H(z_{it} - z) \bar{f}^{-1}(z) \{1 + o_p(1)\}$, where $\bar{f}(z) = \sum_{t=1}^T f_t(z)$.
- (b) $\lim P_n \{n^{-1} [Z(z)' K_H(z) Z(z)]_{ij} \leq C \text{ for } z \in C_f, i, j = 1, \dots, q+1\} = 1 \text{ for some } C$.
- (c) $\lim P_n \{ \sup_z \max_{1 \leq i \leq n, 1 \leq t \leq T} |s(z_{it}, z)| \leq C n^{-1} |H|^{-1} \} = 1 \text{ for some } C$.

Lemma 4.2.

$$(D'PD)^{-1} = (D'D)^{-1} + O_p(\varsigma_n) = T^{-1} I_n + O_p(n) \text{ for sufficiently large } n, \text{ where } n \\ = i_n i_n' n^{-1} |H|^{-1} \sqrt{\ln n}.$$

Lemma 4.3.

- (a) $n^{-1} X' P X \xrightarrow{p} \sum_{t=1}^T E[(x_{it} - p(z_{it}))(x_{it} - p(z_{it}))']$.
- (b) $n^{-1} X' P D (D'D)^{-1} D' P X \xrightarrow{p} T^{-1} \sum_{t=1}^T \sum_{s=1}^T E[(x_{it} - p(z_{it}))(x_{is} - p(z_{is}))']$.

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

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

Lemma 4.6.

$$(a) \quad n^{-1/2}X'PV=n^{-1/2}\sum_{i=1}^n\sum_{t=1}^T(x_{it}-p(z_{it}))v_{it}+o_p(1). \\ (b) \quad n^{-1/2}X'PD(D'D)^{-1}D'PV=n^{-1/2}T^{-1}\sum_{i=1}^n\sum_{t=1}^T\sum_{s=1}^T(x_{it}-p(z_{it}))v_{is}+o_p(1).$$

Lemma 4.7. $n^{-1/2}X^{*'}M^*(I_{nT}-S)V^dN(0,\Omega)$.

Proof of Theorem 3.1. (i) Noting that $M^*D^*=0$, $\hat{\beta}-\beta=(X^{*'}M^*X^*)^{-1}X^{*'}M^*Y^*-\beta=(X^{*'}M^*X^*)^{-1}X^{*'}M^*[D^*\alpha+X^*\beta+(I_{nT}-S)(m(Z)+V)]-\beta=(X^{*'}M^*X^*)^{-1}X^{*'}M^*(I_{nT}-S)(m(Z)+V)$. Thus by (Lemmas 4.4, 4.5, 4.7), and the CLT, $\sqrt{n}(\hat{\beta}-\beta)=[n^{-1}X^{*'}M^*X^*]^{-1}n^{-1/2}X^{*'}M^*(I_{nT}-S)V+[n^{-1}X^{*'}M^*X^*]^{-1}n^{-1/2}X^{*'}M^*(I_{nT}-S)m(Z)=\Phi^{-1}n^{-1/2}\sum_{i=1}^n\sum_{t=1}^T\tilde{x}_{it}(v_{it}-T^{-1}\sum_{s=1}^Tv_{is})+o_p(1)^dN(0,\Phi^{-1}\Omega\Phi^{-1})$.

(ii) It suffices to show $\Phi=\Phi+o_p(1)$ and $\Omega=\Omega+o_p(1)$. The first part follows from the arguments of Lemma 4.3. Let $\tau_n=n^{-1/2}|H|^{-1/2}\sqrt{\ln(n)}+||H||^2$. By Theorem 3.2 and standard arguments for uniform convergence (e.g., Masry (1996)), $m(z)-\hat{m}(z)=O_p(\tau_n)$ uniformly in z . By (2.6), (2.7) and (3.1) it is easy to show $v_{it}=\hat{v}_{it}+o_p(1)$ uniformly. Also, uniformly in i and t , $\tilde{x}_{it}'=x_{it}'-s(z_{it})'X=\tilde{x}_{it}'+O_p(\tau_n)$. Consequently, by the law of large numbers $\hat{\Omega}=n^{-1}\sum_i\sum_t\sum_s\tilde{x}_{it}(\tilde{x}_{is}-\sum_l\tilde{x}_{il}/T)'\tilde{v}_{it}\tilde{v}_{is}+o_p(1)^p\Omega$. \square

Proof of Theorem 3.2. Denote $S(z_{it}, z)$ as a typical column of $S(z)$, i.e., $S(z)=(S(z_{11}, z), \dots, S(z_{iT}, z), \dots, S(z_{nT}, z))$. By (2.7), (2.9) and Lemma 4.2, $\hat{M}(z)=S(z)(Y-D\alpha-X\beta)=S(z)(I_{nT}-D(D'PD)^{-1}D'P)(Y-X\beta)=s(z)(I_{nT}-D(D'PD)^{-1}D'P)\{m(Z)+V-X(\hat{\beta}-\beta)\}$. By the Taylor expression, $m(z_{it})=\bar{Z}_{it}'M(z)+\frac{1}{2}(z_{it}-z)'\ddot{m}(z)(z_{it}-z)+o_p(||H||^2)$. So $\sqrt{n|H|}(\hat{M}(z)-M(z))\frac{\sqrt{n|H|}}{2}\sum_{i=1}^n\sum_{t=1}^T S(z_{it}, z)(z_{it}-z)'\ddot{m}(z)(z_{it}-z)+\sqrt{n|H|}S(z)(I_{nT}-D(D'PD)^{-1}D'P)V-\sqrt{n|H|}S(z)D(D'DP)^{-1}D'Pm(Z)-\sqrt{n|H|}S(z)(I_{nT}-D(D'PD)^{-1}D'P)X(\hat{\beta}-\beta)\equiv B_{11}+B_{12}-B_{13}-B_{14}$. It is easy to show $B_{13}=o_p(1)$ $B_{14}=o_p(1)$ by using results in Lemmas 4.1–4.3, Lemma 4.5 and Theorem 3.1. Noting $[n^{-1}\bar{Z}'(z)'K_H(z)\bar{Z}(z)]^{-1}=Q^{-1}+o_p(1)$, $B_{11}=\frac{\sqrt{n|H|}}{2}\sum_{i=1}^n\sum_{t=1}^T S(z_{it}, z)(z_{it}-z)'\ddot{m}(z)(z_{it}-z)\frac{\sqrt{n|H|}}{2}Q^{-1}n^{-1}\sum_{i=1}^n\sum_{t=1}^T \bar{Z}_{it}'(z)K_H(z_{it}-z)(z_{it}-z)'\ddot{m}(z)(z_{it}-z)=\frac{\sqrt{n|H|V(z)}}{2}Q^{-1}(tr(\int_{\mathbb{R}^q}uu'K(u)duH\ddot{m}(z)H^0)+o_p(1))$. $B_{12}=\sqrt{n|H|}S(z)(I_{nT}-D(D'PD)^{-1}D'P)V\sqrt{n|H|}Q^{-1}n^{-1}\sum_{i=1}^n\sum_{t=1}^T \bar{Z}_{it}'(z)K_H(z_{it}-z)\tilde{v}_{it}^dN(0, Q^{-1}\Gamma Q^{-1})$, where $\tilde{v}_{it}=v_{it}-T^{-1}\sum_{s=1}^Tv_{is}$. This completes the proof. \square

References

- Baltagi, B.H., Li, D., 2002. Series estimation of partially linear panel data models with fixed effects. *Annals of Economic and Finance* 3, 103–116.
- Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and its Applications*. Chapman and Hall.
- Hu, Z., Wang, N., Carroll, R.J., 2004. Profile-kernel versus backfitting in the partially linear models for longitudinal/clustering data. *Biometrika* 91, 251–262.

- Li, Q., Stengos, T., 1996. Semiparametric estimation of partially linear panel data models. *Journal of Econometrics* 71, 289–397.
- Lin, X., Carroll, R.J., 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.