ANNALS OF ECONOMICS AND FINANCE **3**, 103–116 (2002)

Series Estimation of Partially Linear Panel Data Models with Fixed Effects *

Badi H. Baltagi

Texas A&M University
Department of Economics
College Station, TX 77843-4228
E-mail: Badi@econ.tamu.edu

and

Dong Li

Kansa State University
Department of Economics
Manhattan, KS 66506
E-mail: dongli@eudoramail.com

This paper considers the problem of estimating a partially linear semiparametric fixed effects panel data model with possible endogeneity. Using the series method, we establish the root N normality result for the estimator of the parametric component, and we show that the unknown function can be consistently estimated at the standard nonparametric rate. © 2002 Peking University Press

Key Words: Panel data; Series method; Fixed effects; Additive models.
JEL Classification Numbers: C13, C14.

1. INTRODUCTION

There is a rich literature on semiparametric estimation of panel data models. However, to our knowledge, no one has proposed a consistent estimation method for a dynamic partially linear panel data model with *fixed effects*. In this paper we show that one can use the series method to con-

* We would like to thank Qi Li for stimulating our interest in this topic and for his helpful comments and suggestions. Baltagi thanks the Advanced Research Program, Texas Higher Education Board and the Bush School Program in the Economics of Public Policy for their support.

1529-7373/2002
Copyright © 2002 by Peking University Press
All rights of reproduction in any form reserved.

sistently estimate a semiparametric panel model with fixed effects. We use the standard approach of taking the first difference to eliminate the fixed effects. This results in a semiparametric additive model with the restriction that the two additive functions have the same functional form. Series estimation methods are more convenient than kernel methods under certain type of restrictions (such as additivity or shape-preserving estimation, see Dechevsky and Penez (1997)). It is also computationally convenient because the results can be summarized by relatively few coefficients.

Recently semiparametric estimation of additive models and additive partially linear models have received much attention, see Linton and Nielsen (1995), Newey (1994), Fan and Li (1996), Fan et al. (1998) and Li (2000), to mention a few. The additive regression model partially avoids the 'curse of dimensionality' problem which may circumvent the estimation of a fully nonparametric regression model. Linton and Nielsen (1995), Newey (1994), and Tojstheim and Auestad (1994) propose to estimate additive models using the kernel marginal integration method. Li (2000), on the other hand, uses the series method to estimate an additive partially l inear model.

Li (2000) considered only the independent data case, while this paper considers panel data with correlated observations. The model in Li (2000) does not allow for endogenous variables (or a lagged dependent variable) among the regressors, while this paper allows for endogenous regressors. We establish the root N normality result for the estimator of the parametric component, and show that the unknown function can be consistently estimated at the standard nonparametric rate.

The paper is organized as follows. In Section 2, we consider the estimation of a static partially linear semiparametric panel data model with fixed effects. In Section 3, we consider a dynamic model with a lagged dependent variable. Section 4 concludes the paper. The proofs of the main results are given in the Appendix.

2. STATIC MODEL AND RESULTS

Consider the following partially linear semiparametric panel data model

$$y_{it} = x_{it}'\gamma + g(z_{it}) + u_{it}, \quad i = 1, ..., N; \ t = 1, ..., T \tag{1}$$

with one-way error component disturbances $u_{it} = \mu_i + \nu_{it}$, x_{it} and z_{it} are of dimensions $k_1 \times 1$ and $k_2 \times 1$, respectively, and the μ_i's are fixed effects, ν_{it} are assumed to be i.i.d. $(0, \sigma_\nu^2)$. Throughout the paper, z_{it}'s are strictly exogenous variables, x_{it} is assumed to be exogenous in this section. However, in Section 3 we will allow x_{it} to be correlated with the error term ν_{it}. The asymptotic theory in this paper assumes a finite value of T while letting the number of individuals N approach infinity. This is the typical micro panel data case.

First differencing (1) to eliminate the fixed effects, we get

$$y_{it} - y_{i,t-1} = (x_{it} - x_{i,t-1})\gamma + [g(z_{it}) - g(z_{i,t-1})] + u_{it} - u_{i,t-1} \quad (2)$$

or

$$Y_{it} = X_{it}\gamma + G(z_{it}, z_{i,t-1}) + U_{it} \quad (3)$$

where $Y_{it} = y_{it} - y_{i,t-1}$, $X_{it} = x_{it} - x_{i,t-1}$, $G(z_{it}, z_{i,t-1}) = g(z_{it}) - g(z_{i,t-1})$, and $U_{it} = \nu_{it} - \nu_{i,t-1}$. In matrix-vector form, we have

$$Y = X\gamma + G + U \quad (4)$$

where Y is a $NT \times 1$ vector with typical element Y_{it}, and X, G and U are similarly defined. To keep our notation consistent, let $Z_{it} = z_{it}$ and Z is an $NT \times 1$ vector with typical element Z_{it}.

Li and Stengos (1996) considered the estimation of γ of the above panel data model using the kernel instrumental variable method. There are several drawbacks of the method proposed by Li and Stengos (1996). First, in order to eliminate the unknown function $G(z_{it}, z_{i,t-1}) = g(z_{it}) - g(z_{i,t-1})$, they suggest estimating $E(Y_{it}|Z_{it}, Z_{i,t-1})$ and $E(X_{it}|Z_{it}, Z_{i,t-1})$ by the nonparametric kernel method. This suffers from the 'curse of dimensionality' because $(z_{it}, z_{i,t-1})$ has a higher dimension than that of z_{it}, i.e., it ignores the additive structure of model (2). Secondly, although their proposed method can estimate $G(z_{it}, z_{i,t-1})$, they did not propose a method to estimate the original unknown function $g(z_{it})$. The problem with Li and Stengos' (1996) method is that they ignore the additive structure of $G(z_{it}, z_{i,t-1}) = g(z_{it}) - g(z_{i,t-1})$. In this paper we propose a series method for estimating model (2) that does not suffer the above mentioned drawbacks. Although one can also use the kernel marginal integration method to estimate the additive function $g(z_{it})$, the asymptotic theory involved in using the kernel method to estimate an additive partially linear model is quite involved, see Fan, Härdle and Mammen (1998) and Fan and Li (1996). Neither Fan, Härdle and Mammen (1998) nor Fan and Li (1996) considered the case of different additive functions to contain overlapping variables. Sperlich et al. (1999) considered the problem of using the kernel method to estimate an additive model with second order interaction terms, but their model does not have a parametric linear component. To our knowledge, there is no asymptotic theory developed for estimating an additive partially linear model by *kernel* method that allows interaction terms in the additive functions.

Below we introduce some definitions and assumptions.

Definition 2.1. A function $\xi(z_{it}, z_{i,t-1})$ is said to belong to an additive class of functions $\mathcal{G}$ ($\xi \in \mathcal{G}$) if $\xi(z_{it}, z_{i,t-1}) = g(z_{it}) - g(z_{i,t-1})$, $g(\cdot)$ is twice differentiable in the interior of its support $\mathcal{S}$, which is a compact subset of $\mathbb{R}^{k_2}$, and $E[g^2(z)] < \infty$.

We use series $p^K(z)$ of dimension $K \times 1$ to approximate $g(z)$. The approximation function $p^K(z)$ has the following properties: (i) $p^K(z) \in \mathcal{G}$; (ii) as K grows, there is a linear combination of $p^K(z)$ that can approximate any $g \in \mathcal{G}$ arbitrarily well in mean square error. Therefore, $p^K(z)$ approximates $g(z)$ and $p^K(z_{it}, z_{i,t-1}) \equiv (p^K(z_{it}) - p^K(z_{i,t-1}))$ approximates $G(z_{it}, z_{i,t-1}) = g(z_{it}) - g(z_{i,t-1})$:

$$p^K(z_{it}, z_{i,t-1}) = \begin{pmatrix} p_1(z_{it}) - p_1(z_{i,t-1}) \\ p_2(z_{it}) - p_2(z_{i,t-1}) \\ \cdots \\ p_K(z_{it}) - p_K(z_{i,t-1}) \end{pmatrix}. \tag{5}$$

Define $p_{it}^K = p^K(z_{it}, z_{i,t-1})$, and $P = (p_{11}^K, p_{12}^K, ..., p_{1T}^K, p_{21}^K, ..., p_{NT}^K)'$. P is of dimension $NT \times K$.

For any scalar or vector function $W(z)$, we use the notation of $E_{\mathcal{G}}(W(z))$ to denote the projection of $W(z)$ onto the additive functional space $\mathcal{G}$ (under the L_2-norm). That is, $E_{\mathcal{G}}(W(z))$ is an element that belongs to $\mathcal{G}$ (has an additive structure) and it is the closest function to $W(z)$ among all the functions in $\mathcal{G}$. More specifically, we have

$$\begin{aligned} & E\{[W(z_{it}) - E_{\mathcal{G}}(W(z_{it}))][W(z_{it}) - E_{\mathcal{G}}(W(z_{it}))]'\} \\ = & \inf_{\xi \in \mathcal{G}} E\{[W(z_{it}) - \xi(z_{it})][W(z_{it}) - \xi(z_{it})]'\}, \end{aligned} \tag{6}$$

where the infimum is in the sense that

$$\begin{aligned} & E\{[W(z_{it}) - E_{\mathcal{G}}(W(z_{it}))][W(z_{it}) - E_{\mathcal{G}}(W(z_{it}))]'\} \\ \leq & E\{[W(z_{it}) - \xi(z_{it})][W(z_{it}) - \xi(z_{it})]'\} \end{aligned} \tag{7}$$

for all $\xi \in \mathcal{G}$, where for square matrices A and B, $A \leq B$ means that $A - B$ is negative semidefinite.

Define $\theta(z) = E(X|Z = z)$. We will use $h(z)$ to denote the projection of $\theta(z)$ onto $\mathcal{G}$, i.e., $h(z) = E_{\mathcal{G}}(\theta(z))$. Note that the series method proposed in this paper can deal with the case where z_{it} and $z_{i,t-1}$ have overlapping variables. For example, z_{it} can have current and lagged exogenous variables.

Assumption 2.1. *(i) We assume that (Y_{it}, X_{it}, Z_{it})'s are independent across individuals, i.e., (Y_i, X_i, Z_i) are independent and identically dis-*

tributed as (Y_1, X_1, Z_1)*, where* $Y_i = (Y_{i1}, ..., Y_{iT})'$ *and* X_i *and* Z_i *are similarly defined; (ii) The support of* (X_1, Z_1) *is a compact subset of* $\mathbb{R}^{k_1+k_2}$*; (iii)* $\theta(z)$ *and* $var(Y_1|X_1 = x, Z_1 = z)$ *are both bounded functions on the support of* (X_1, Z_1)*.*

Assumption 2.2. *(i) For every* K *there is a nonsingular matrix* B *such that for* $P^K(z) = B\,p^k(z)$*: the smallest eigenvalue of* $E[P^K(Z_{it})P^K(Z_{it})']$ *is bounded away from zero uniformly in* K*; (ii) There is a sequence of constants* $\zeta_0(K)$ *satisfying* $\sup_{z\in\mathbb{S}} \left\|P^K(z)\right\| \leq \zeta_0(K)$ *and* $K = K(N)$ *such that* $(\zeta_0(K))^2K/N \to 0$ *as* $N \to \infty$*.*

Assumption 2.3. *(i) For* $f = g$ *or* $f = h_{(s)}$ $(s = 1, ..., k_1)$*, there exist some* $\delta(> 0)$*,* $\beta_f = \beta_f(K)$*,* $\sup_{z\in\mathcal{Z}} |f(z)-P^K(z)\beta_f| = O(K^{-\delta})$ *as* $K \to \infty$*; (ii)* $\sqrt{N}K^{-\delta} \to 0$ *and* $N \to \infty$*.*

Assumption 2.1 is quite standard in the additive models. Assumption 2.2 ensures that $P'P$ is asymptotically nonsingular. Newey (1997) gives some primitive conditions for power series and splines such that assumptions 2.2 and 2.3 hold.

Let $M = P(P'P)^-P'$, where $(.)^-$ denotes any symmetric generalized inverse. Define $\tilde{A} = MA = P\beta_A$ where $\beta_A = (P'P)^-P'A$. Then premultiplying (4) by M leads to

$$\tilde{Y} = \tilde{X}\gamma + \tilde{G} + \tilde{U}. \tag{8}$$

Subtracting (8) from (4) gives

$$Y - \tilde{Y} = (X - \tilde{X})\gamma + G - \tilde{G} + U - \tilde{U}. \tag{9}$$

We estimate γ by least squares regression of $Y - \tilde{Y}$ on $X - \tilde{X}$:

$$\hat{\gamma} = [(X - \tilde{X})'(X - \tilde{X})]^-(X - \tilde{X})'(Y - \tilde{Y}). \tag{10}$$

Plug (9) into (10), we have

$$\begin{aligned} \hat{\gamma} &= [(X - \tilde{X})'(X - \tilde{X})]^-(X - \tilde{X})'(Y - \tilde{Y}) \\ &= \gamma + [(X - \tilde{X})'(X - \tilde{X})]^-(X - \tilde{X})'(G - \tilde{G} + U - \tilde{U}). \end{aligned} \tag{11}$$

$g(z)$ is estimated by $\hat{g}(z) = p^K(z)'\hat{\beta}$ where $\hat{\beta}$ is given by

$$\hat{\beta} = (P'P)^-P'(Y - X\hat{\gamma}). \tag{12}$$

THEOREM 2.1. *Define* $\epsilon_{it} = X_{it} - h(Z_{it})$, *where* $h(Z_{it}) = E_{\mathcal{G}}(\theta(Z_{it}))$ *is defined above. Assume that* $\Phi \stackrel{def}{=} \frac{1}{T}\sum_{t=1}^{T} E(\epsilon_{it}\epsilon_{it}')$ *is positive definite, then we have (i)* $\sqrt{N}(\hat{\gamma} - \gamma) \to N(0, \Sigma)$ *in distribution, where* $\Sigma = \Phi^{-1}\Omega\Phi^{-1}$, $\Omega = \frac{1}{T}\sum_t E[\sigma_u^2(X_{it}, Z_{it})\epsilon_{it}\epsilon_{it}']$ *and* $\sigma_u^2(X_{it}, Z_{it}) = E(U_{it}^2 | X_{it} = x, Z_{it} = z)$; *(ii) A consistent estimator of* Σ *is given by* $\hat{\Sigma} = \hat{\Phi}^{-1}\hat{\Omega}\hat{\Phi}^{-1}$, *where* $\hat{\Phi} = (NT)^{-1}\sum_i\sum_t (X_{it} - \tilde{X}_{it})(X_{it} - \tilde{X}_{it})'$, $\hat{\Omega} = (NT)^{-1}\sum_i\sum_t \hat{U}_{it}^2(X_{it} - \tilde{X}_{it})(X_{it} - \tilde{X}_{it})'$, $\hat{U}_{it} = Y_{it} - X_{it}'\hat{\gamma} - \hat{G}(Z_{it})$.

The proof is given in the appendix.

Theorem 2.1 establishes the $\sqrt{N}$-consistency and asymptotic normality of $\hat{\gamma}$. The next theorem establishes the consistency for the nonparametric component estimator.

THEOREM 2.2. *Under assumptions 1–3, we have*
(i) $\sup_{z \in \mathcal{S}} |\hat{g}(z) - g(z)| = O_p(\zeta_0(K))(\sqrt{K}/\sqrt{N} + K^{-\delta})$;
(ii) $N^{-1}(\hat{g}(z) - g(z))^2 = O_p(K/N + K^{-2\delta})$;
(iii) $\int (\hat{g}(z) - g(z))^2 dF(z) = O_p(K/N + K^{-2\delta})$, *where* $F(.)$ *is the cumulative distribution function of* Z.

3. INSTRUMENT VARIABLE ESTIMATION OF A PANEL DATA MODEL

In this section we allow the partially linear panel data model in (2) to have variables in the parametric component that may be correlated with the error term. We still assume that the nonparametric component is a function of exogenous variables. A special case of this model is a dynamic panel data model with a lagged dependent variable as one of the regressors in x_{it}. In this case, model (2) reduces to a dynamic panel data model

$$y_{it} = \gamma_0 y_{i,t-1} + \gamma_1' x_{it}^- + g(z_{it}) + u_{it}. \tag{13}$$

where x_{it}^- is x_{it} excluding $y_{i,t-1}$.

As in Section 2, taking first difference of equation (2) leads to equation (10). Now instead of using OLS-type semiparametric estimator as given in (11), we use an instrumental variable (IV) semiparametric estimation method. We assume that there exists a set of instrument variables $W_{it} \in R^{k_3}$ with $k_3 \geq k_1$, such that $E(U_{it}|W_{it}) = 0$ and $Cov(W_{it}, X_{it}) \neq 0$. Hence we can estimate γ by the following IV method (recall that $\tilde{A} \stackrel{def}{=} P(P'P)^- P'A$):

$$\hat{\gamma}_{IV} = [(W - \tilde{W})'(X - \tilde{X})]^{-1}(W - \tilde{W})'(Y - \tilde{Y})$$

$$= \gamma + [(W - \tilde{W})'(X - \tilde{X})]^{-1}(W - \tilde{W})'[(G - \tilde{G}) + (U - \tilde{U})]. \quad (14)$$

The nonparametric part $g(z)$ can be estimated by $\hat{g}(z) = p^K(z)'\hat{\beta}_{IV}$ where $\hat{\beta}_{IV}$ is given by

$$\hat{\beta}_{IV} = (P'P)^{-}P'(Y - X\hat{\gamma}_{IV}). \quad (15)$$

We have the following theorem for the IV estimators:

THEOREM 3.1. *Define* $\epsilon_{it} = W_{it} - E_{\mathcal{G}}(W_{it})$, $\tau_{it} = X_{it} - E_{\mathcal{G}}(X_{it})$, *and assume that* $\Psi \stackrel{def}{=} \frac{1}{T}\sum_{t=1}^{T} E(\epsilon_{it}\tau_{it}')$ *is positive definite, we have*

(i) $\sqrt{N}(\hat{\gamma}_{IV} - \gamma) \to N(0, \Upsilon)$ *in distribution, where* $\Upsilon = \Psi^{-1}\Lambda\Psi^{-1}$, $\Lambda = \frac{1}{T}\sum_t E[\sigma_u^2(W_{it}, Z_{it})\epsilon_{it}\epsilon_{it}']$ *and* $\sigma_u^2(W_{it}, Z_{it}) = E(U_{it}^2|W_{it} = w, Z_{it} = z)$;

(ii) A consistent estimator of Υ *is given by* $\hat{\Upsilon} = \hat{\Psi}^{-1}\hat{\Lambda}\hat{\Psi}^{-1}$, *where* $\hat{\Psi} = (NT)^{-1}\sum_i\sum_t(W_{it} - \tilde{W}_{it})(X_{it} - \tilde{X}_{it})'$, $\hat{\Lambda} = (NT)^{-1}\sum_i\sum_t \hat{U}_{it}^2(W_{it} - \tilde{W}_{it})(W_{it} - \tilde{W}_{it})'$, *and* $\hat{U}_{it} = Y_{it} - X_{it}\hat{\gamma}_{IV} - \hat{G}(Z_{it})$.

The proof of Theorem 3.1 is similar to the proof of Theorem 2.1, a sketchy proof is given in the Appendix.

In this section we have established the asymptotic results for a partially linear panel data model with fixed effects and with possible endogeneity in the regressors. These results are an extension of those in Section 2. If all the regressors in (1) are exogenous, we can simply let the instrument variable W be X itself, and the results are the same as that in Section 2. The model discussed in this Section encompasses an important model - the dynamic panel data model with lagged dependent variables among the regressors, see Baltagi (1995) for a discussion of parametric estimation of this model.

4. CONCLUSION

In this paper we consider the problem of estimating a partially linear semiparametric (dynamic) panel data model with fixed effects. We extend the results in Li (2000) to a dynamic panel data model which allows for dependent observations and the presence of lagged dependent variables among the regressors. Using the series method, we establish the root N normality result for the estimator of the parametric component, and we show that the unknown function can be consistently estimated at the standard nonparametric rate. These results are of special interest to applications in micro level panel data.

APPENDIX A

Proof of Lemmas:

First we prove some lemmas which will be utilized in the proofs of theorems.

We use $\mathbf{1}$ to denote an indicator function that takes value one if $(P'P)$ is invertible and zero otherwise. Whenever we have $(P'P)^{-1}$, it should be understood as $\mathbf{1}(P'P)^{-1}$ and since $Prob(\mathbf{1}=1)\to 1$ almost surely, we will often omit the indicator function $\mathbf{1}$.

We define $S_{A,B}=n^{-1}A'B=n^{-1}\sum_i A_iB_i'$. Also, we define $S_A=S_{A,A}$.

LEMMA A.1. $\hat{Q}-I=O_p(\zeta_0(K)\sqrt{K}/\sqrt{N})$, *where* $\hat{Q}=P'P/(NT)$.

Proof. See the proof of Theorem 1 in Newey (1997, pp. 161–162). ∎

LEMMA A.2. $\left\|\tilde{\beta}_f-\beta_f\right\|=O_p(K^{-\delta})$, *where* $\tilde{\beta}_f=(P'P)^{-}P'f$, β_f *satisfies assumption (2.3), for* $f=g$ *or* $f=h_{(s)}$ $(s=1,...,r)$.

Proof. We have

$$\begin{aligned}
&\mathbf{1}\left\|\tilde{\beta}_f-\beta_f\right\| \overset{def}{=} \mathbf{1}\left\|(P'P)^{-}P'(f-P\beta_f)\right\| \\
&= \mathbf{1}\{(f-P\beta_f)'P(P'P)^{-1}(P'P/(NT))^{-1}P'(f-P\beta_f)/(NT)\}^{1/2} \\
&= \mathbf{1}O_p(1)\{(f-P\beta_f)'P(P'P)^{-1}P'(f-P\beta_f)/(NT)\}^{1/2} \\
&\le O_p(1)\{(f-P\beta_f)'(f-P\beta_f)/(NT)\}^{1/2}=O_p(K^{-\delta}) \qquad \text{(A.1)}
\end{aligned}$$

by lemma A.1, assumption 2.3 and the fact that $P(P'P)^{-1}P'$ is idempotent. So $\left\|\tilde{\beta}_f-\beta_f\right\|=O_p(K^{-\delta})$ since $Prob(\mathbf{1}=1)\to 1$. ∎

LEMMA A.3. $P'\eta/(NT)=O_p(\zeta_0(K)/\sqrt{N})=o_p(1)$.

Proof. Note that $E(P_{it}\eta_{it})=0$ since $p^K(.)\in\mathcal{G}$ and $\eta(.)\perp\mathcal{G}$. Define $P_i=(P_{i1}',...,P_{iT}')'$ and $\eta_i=(\eta_{i1},...,\eta_{iT})'$. In this case

$$\begin{aligned}
E\left\|P'\eta/(NT)\right\|^2 &= (NT)^{-2}E(\eta'PP'\eta)=(NT)^{-2}E\{(\sum_i\eta_i'P_i)(\sum_j P_j'\eta_j)\} \\
&= (NT)^{-2}E(\sum_i\sum_j\eta_i'P_iP_j'\eta_j) \\
&= (NT)^{-2}\{E(\sum_i\eta_i'P_iP_i'\eta_i)+E(\sum_i\sum_{j\neq i}\eta_i'P_iP_j'\eta_j) \\
&= (NT)^{-2}\sum_i E(\eta_i'P_iP_i'\eta_i)+0=N^{-1}T^{-2}E(tr(P_iP_i'\eta_i\eta_i'))
\end{aligned}$$

$$
\begin{aligned}
&= N^{-1}T^{-2}tr(E(P_iP_i'\eta_i\eta_i')) \\
&\leq N^{-1}T^{-2}C\,tr(E(P_iP_i')) = N^{-1}T^{-2}C\sum_{t=1}^{T}\sum_{l=1}^{K}E(P_{it,l}^2) \\
&\leq N^{-1}T^{-1}C(\zeta_0(K))^2 = O((\zeta_0^2(K))^2/N)
\end{aligned}
$$

by assumption 2.2. Hence $P'\eta/(NT) = O_p(\zeta_0(K)/\sqrt{N}) = o_p(1)$. ▮

LEMMA A.4. *$S_{f-\tilde{f}} = O_p(K^{-2\delta}) = o_p(N^{-1/2})$, where $f = G$ or $f = h$.*

Proof. Note that $\tilde{f} \equiv P\tilde{\beta}_f$, so that

$$
\begin{aligned}
S_{f-\tilde{f}} &= N^{-1}\left\|f-\tilde{f}\right\|^2 \leq N^{-1}\{\|f-P\beta_f\|^2 + \left\|P(\beta_f-\tilde{\beta}_f)\right\|^2 \\
&= N^{-1}\|f-P\beta_f\|^2 + T(\beta_f-\tilde{\beta}_f)'\{P'P/(NT)\}(\beta_f-\tilde{\beta}_f) \\
&= O(K^{-2\delta}) + O_p(K^{-2\delta}) \\
&= O_p(K^{-2\delta})
\end{aligned}
$$

by assumption 2.3, lemmas A.1 and A.2. ▮

LEMMA A.5. *(i) $S_{\tilde{a}} = O_p(K/N)$, for $a = U$ or v; (ii) $S_{\tilde{\eta}} = o_p(1)$.*

Proof. (i) Similar to the proof of Theorem 1 in Newey (1997), for $a = v$, we have

$$
\begin{aligned}
E(S_{\tilde{v}}|Z) &= (NT)^{-1}E\{v'P(P'P)^{-1}P'v|Z\} = (NT)^{-1}tr[P(P'P)^{-1}P'E(vv'|Z)] \\
&\leq (NT)^{-1}C\,tr[P(P'P)^{-1}P'] \\
&= O(K/N),
\end{aligned}
$$

which implies $S_{\tilde{v}} = O_p(K/N)$. It is similar for $a = u$.

(ii) $S_{\tilde{\eta}} = (NT)^{-1}\tilde{\eta}'\tilde{\eta} = (\eta'P/(NT))(P'P/(NT))^{-1}(P'\eta/(NT)) = O_p(\zeta_0^2(K)K/N)$ $= o_p(1)$ by lemmas A.1 and A.3. ▮

LEMMA A.6. *For any $T \times T$ matrix $A = (a_{ts})$ in which elements are bounded $|a_{ts}| \leq M < \infty$, there is a constant C such that $CI_T - A$ is positive definite.*

Proof. Let $C = \bar{M}T$ for some $\bar{M} > M$ (recall T is a fixed value). Then for any vector $x \neq 0$,

$$
x'(CI_T - A)x = Cx'x - x'Ax = \bar{M}T\sum_{t=1}^{T}x_t^2 - \sum_{t=1}^{T}\sum_{s=1}^{T}a_{ts}x_tx_s
$$

$$\geq \bar{M}T\sum_{t=1}^{T}x_t^2 - \frac{M}{2}\sum_{t=1}^{T}\sum_{s=1}^{T}(x_t^2 + x_s^2) > 0.$$

▮

Proof of Theorem 2.1:

Recall that $\theta(Z_{it}) = E(X_{it}|Z_{it})$, $v_{it} = X_{it} - \theta(Z_{it})$, $\epsilon_{it} = X_{it} - h(Z_{it})$, and $\eta_i = \theta(Z_{it}) - h(Z_{it})$ where $h(z) = E_{\mathcal{G}}(\theta(z))$.

We will use the following short-hand notations: $\theta_{it} = \theta(Z_{it})$, $g_{it} = g(Z_{it})$, $h_{it} = h(Z_{it})$. Hence, $v_{it} = X_{it} - \theta_{it}$, $\epsilon_{it} = X_{it} - h_{it}$, and $\eta_{it} = \theta_{it} - h_{it}$. θ is the $NT \times k_1$ matrix with typical element θ_{it}. h, G, η, ϵ, v, and U are similarly defined.

From the definitions, we have $X = \eta + v + h$ and $\tilde{X} = \tilde{\eta} + \tilde{v} + \tilde{h}$. Subtracting these two equalities yields

$$X - \tilde{X} = \eta - \tilde{\eta} + v - \tilde{v} + h - \tilde{h}. \tag{A.2}$$

Note that if $S_{X-\tilde{X}}^{-1}$ exists, then from (9) and (10), we have

$$\sqrt{N}(\hat{\gamma} - \gamma) = S_{X-\tilde{X}}^{-1}\sqrt{N}S_{X-\tilde{X},G-\tilde{G}+U-\tilde{U}}. \tag{A.3}$$

We need to show the followings: (i) $S_{X-\tilde{X}} = \Phi + o_p(1)$, (ii) $S_{X-\tilde{X},G-\tilde{G}} = o_p(N^{-1/2})$, (iii) $S_{X-\tilde{X},\tilde{U}} = o_p(N^{-1/2})$, and (iv) $\sqrt{N}S_{X-\tilde{X},U} \to N(0,\Omega)$ in distribution.

(i) Proof of $S_{X-\tilde{X}} = \Phi + o_p(1)$.

From (A.2) we have $S_{X-\tilde{X}} = S_{(\eta+v)+(h-\tilde{h})-\tilde{\eta}-\tilde{v}} = S_{\eta+v} + S_{(h-\tilde{h})-\tilde{\eta}-\tilde{v}} + 2S_{(\eta+v),(h-\tilde{h})-\tilde{\eta}-\tilde{v}}$.

First, we have

$$\begin{aligned} S_{\eta+v} &= (NT)^{-1}\sum_{i,t}(\eta_{it} + v_{it})(\eta_{it} + v_{it})' = (NT)^{-1}\sum_i\sum_t \epsilon_{it}\epsilon_{it}' \\ &= N^{-1}\sum_i(T^{-1}\sum_t \epsilon_{it}\epsilon_{it}') = \Phi + o_p(1) \end{aligned}$$

by virtue of a law of large numbers.

Second, $S_{(h-\tilde{h})-\tilde{\eta}-\tilde{v}} \leq 2S_{(h-\tilde{h})} + 4S_{\tilde{\eta}} + 4S_{\tilde{v}} = o_p(1)$ by lemmas A.4 and A.5.

Finally, $S_{(\eta+v),(h-\tilde{h})-\tilde{\eta}-\tilde{v}} \leq (S_{\eta+v}S_{(h-\tilde{h})-\tilde{\eta}-\tilde{v}})^{1/2} = [O_p(1)o_p(1)]^{1/2} = o_p(1)$ by the above results.

(ii) Proof of $S_{X-\tilde{X},G-\tilde{G}} = o_p(N^{-1/2})$.

By equation (A.2), we have

$$S_{X-\tilde{X},G-\tilde{G}} = S_{\eta-\tilde{\eta}+v-\tilde{v}+h-\tilde{h},G-\tilde{G}}$$

$$= S_{\eta+v,G-\tilde{G}} + S_{h-\tilde{h},G-\tilde{G}} - S_{\tilde{v},G-\tilde{G}} - S_{\tilde{\eta},G-\tilde{G}}. \quad \text{(A.4)}$$

For these four terms, we have

(1) $S_{\eta+v,G-\tilde{G}} \leq \{S_{\eta+v}S_{G-\tilde{G}}\}^{1/2} = O_p(K^{-\delta})$ by lemma A.4 and $S_{\eta+v} = O_p(1)$.

(2) $S_{h-\tilde{h},G-\tilde{G}} \leq \{S_{h-\tilde{h}}S_{G-\tilde{G}}\}^{1/2} = O_p(K^{-2\delta})$ by lemma A.4.

(3) $S_{\tilde{v},G-\tilde{G}} \leq \{S_{\tilde{v}}S_{G-\tilde{G}}\}^{1/2} = o_p(1)O_p(K^{-\delta})$ by lemmas A.4 and A.5.

(4) $S_{\tilde{\eta},G-\tilde{G}} \leq \{S_{\tilde{\eta}}S_{G-\tilde{G}}\}^{1/2} = o_p(1)O_p(K^{-\delta})$ by lemmas A.4 and A.5.

(iii) Proof of $S_{X-\tilde{X},\tilde{U}} = o_p(N^{-1/2})$.

By equation (A.2), $S_{X-\tilde{X},\tilde{U}} = S_{\eta-\tilde{\eta}+v-\tilde{v}+h-\tilde{h},\tilde{U}} = S_{\eta,\tilde{U}}+S_{v,\tilde{U}}+S_{h-\tilde{h},\tilde{U}} - S_{\tilde{v},\tilde{U}} - S_{\tilde{\eta},\tilde{U}}$. We consider these five terms separately.

(1) We have

$$\begin{aligned} E[\left\|S_{\eta,\tilde{U}}\right\|^2 |Z] &= (NT)^{-2}E[U'P(P'P)^{-1}P'\eta\eta'P(P'P)^{-1}P'U|Z] \\ &= (NT)^{-2}tr[P(P'P)^{-1}P'\eta\eta'P(P'P)^{-1}P'E(UU'|Z)] \\ &\leq C(NT)^{-2}tr(\tilde{\eta}\tilde{\eta}') = C(NT)^{-1}tr(S_{\tilde{\eta}}) = o_p(N^{-1}) \end{aligned}$$

by lemma A.5. Hence $S_{\eta,\tilde{U}} = O_p(N^{-1/2})$.

(2) We have

$$\begin{aligned} E[\left\|S_{v,\tilde{U}}\right\|^2 |X,Z] &= (NT)^{-2}tr[P(P'P)^{-1}P'vv'P(P'P)^{-1}P'E(UU'|X,Z)] \\ &\leq C(NT)^{-2}tr(\tilde{v}\tilde{v}') = C(NT)^{-1}tr(S_{\tilde{v}}) = O(K/N^2) \end{aligned}$$

by lemma A.5.

(3) $S_{h-\tilde{h},\tilde{U}} \leq \{S_{h-\tilde{h}}S_{\tilde{U}}\}^{1/2} = O_p(K^{-\delta})O_p(\sqrt{K}/\sqrt{N})$ by lemmas A.4 and A.5.

(4) $S_{\tilde{v},\tilde{U}} \leq \{S_{\tilde{v}}S_{\tilde{U}}\}^{1/2} = O_p(K/N)$ by lemma A.5.

(5) We have

$$\begin{aligned} E[\left\|S_{\tilde{\eta},\tilde{U}}\right\|^2 |Z] &= (NT)^{-2}tr[P(P'P)^{-1}P'\eta\eta'P(P'P)^{-1}P'E(UU'|Z)] \\ &\leq C(NT)^{-2}tr(\tilde{\eta}\tilde{\eta}') = C(NT)^{-1}tr(S_{\tilde{\eta}}) = o_p(N^{-1}) \end{aligned}$$

by lemma A.5. Hence $S_{\tilde{\eta},\tilde{U}} = o_p(N^{-1/2})$.

(iv) Proof of $\sqrt{N}S_{X-\tilde{X},U} \to N(0,\Omega)$ in distribution.

$S_{X-\tilde{X},U} = S_{\eta-\tilde{\eta}+v-\tilde{v}+h-\tilde{h},U} = S_{\eta+v,U} + S_{h-\tilde{h},U} - S_{\tilde{v},U} - S_{\tilde{\eta},U}$. We consider them separately.

(1) $\sqrt{N}S_{\eta+v,U} = \sqrt{N}S_{\eta+v,U} = \sqrt{N}\sum_i(\eta_i + v_i)u_i' \to N(0,\Omega)$ by Levi-Lindberg central limit theorem.

(2) We have

$$E[\left\|S_{h-\tilde{h},U}\right\|^2 |Z] = (NT)^{-2}tr[(h-\tilde{h})(h-\tilde{h})'E(UU'|Z)]$$

$$\leq C(NT)^{-1}tr[(h-\tilde{h})(h-\tilde{h})'/NT]$$
$$= C(NT)^{-1}tr(S_{h-\tilde{h}}) = o_p(N^{-1})$$

by lemma A.4 and lemma A.5. Hence, $S_{h-\tilde{h},U} = o_p(N^{-1/2})$.

(3) Similarly, we have

$$E[\|S_{\tilde{v},U}\|^2 |Z] \leq C(NT)^{-1}tr[S_{\tilde{v}}] = o_p(N^{-1})$$

by lemma A.5. Hence, $S_{\tilde{v},U} = o_p(N^{-1/2})$.

(4) We have

$$E[\|S_{\tilde{\eta},U}\|^2 |Z] \leq C(NT)^{-1}tr[S_{\tilde{\eta}}] = o_p(N^{-1})$$

by lemma A.5. Hence, $S_{\tilde{\eta},U} = o_p(N^{-1/2})$.

(i) – (iv) above imply that $\sqrt{N}(\hat{\gamma}-\gamma) = \Phi^{-1}N(0,\Omega) + o_p(1) \rightarrow N(0,\Phi^{-1}\Omega\Phi^{-1})$.

Proof of $\hat{\Sigma} = \Sigma + o_p(1)$:

$\hat{\Phi} \equiv S_{X-\tilde{X}} = \Phi + o_p(1)$ is proved in (i) of Theorem 2.1.

Similarly we can prove that $\hat{\Omega} = \Omega + o_p(1)$. Notice that $\hat{\gamma} - \gamma = O_p(N^{-1/2})$ and $\hat{G} - G = o_p(1)$. It is easy to prove that $\hat{u}_{it} = u_{it} + o_p(1)$. Also, we know that $h_{it} - \tilde{h}_{it} = o_p(1)$, $\tilde{v}_{it} = o_p(1)$ and $\tilde{\eta}_{it} = o_p(1)$. Hence $X_{it} - \tilde{X}_{it} = \epsilon_{it} + o_p(1)$. These results lead to $\hat{\Omega} = \Omega + o_p(1)$ by a law of large numbers.

Proof of Theorem 2.2:

This theorem is almost the same as Theorem 1 in Newey (1997) and Theorem 4.1 in Newey (1995) except that our estimator $\hat{g}$ has an extra term $X(\gamma-\hat{\gamma})$. It suffices to show that the contribution of this extra term is asymptotically negligible. One would expect this to be true because $(\gamma - \hat{\gamma}) = O_p(N^{-1/2})$ which has a smaller order than that of the nonparametric series estimation convergence rate.

Let β_g satisfy assumption 2.3 with $f = g$. We have

$$\begin{aligned}\hat{\beta} &= (P'P)^{-}P'(Y-X\hat{\gamma}) = (P'P)^{-}P'[(Y-X\gamma)-X(\hat{\gamma}-\gamma)]\\ &= \beta_g + (P'P)^{-}P'[(G-P\beta_g)+U] - (P'P)^{-}P'X(\hat{\gamma}-\gamma)\\ &= \beta_g + D_{1N} - D_{2N}(\hat{\gamma}-\gamma), \end{aligned} \tag{A.5}$$

where $D_{1N} = (P'P)^{-}P'[(G-P\beta_g)+U]$ and $D_{2N} = (P'P)^{-}P'X$. $\|D_{1N}\| = O_p(K^{-\delta} + \sqrt{K}/\sqrt{N})$ was proved by Theorem 4.1 in Newey (1995).

Next we will show that $D_{2N}(\hat{\gamma}-\gamma) = O_p(N^{-1/2})$.

$$\|D_{2N}\|^2 = \left\|(P'P)^{-}P'(\eta+v+h)\right\|^2 = \left\|\tilde{\beta}_\eta + \tilde{\beta}_v + \tilde{\beta}_h\right\|^2$$

$$\begin{aligned} &\leq C(\left\|\tilde{\beta}_\eta\right\|^2 + \left\|\tilde{\beta}_v\right\|^2 + \left\|\tilde{\beta}_h\right\|^2) \\ &= O_p(\frac{1}{N}\{\|\tilde{\eta}\|^2 + \|\tilde{v}\|^2 + \left\|\tilde{h}\right\|^2\}) \\ &= N^{-1}\{O_p(\|h\|^2) + o_p(1)\} = O_p(1). \end{aligned}$$

So $D_{2N}(\hat{\gamma}-\gamma) = O_p(\hat{\gamma}-\gamma) = O_p(1/\sqrt{N})$ which has an order smaller than $O_p(\sqrt{K}/\sqrt{N})$. Thus $\left\|\hat{\beta} - \beta_g\right\| = O_p(K^{-\delta} + \sqrt{K}/\sqrt{N})$ as in Newey (1995). The rest of proofs (to prove (i)–(iii) in our Theorem 2.2) follows the same arguments as in the proofs of Theorem 1 in Newey (1997) for (i) and that of Theorem 4.1 in Newey (1995) for (ii) and (iii).

Proof of Theorem 3.1:

Note that

$$\hat{\gamma}_{IV} = \gamma + [(W-\tilde{W})'(X-\tilde{X})]^{-1}(W-\tilde{W})'[(G-\tilde{G}) + (U-\tilde{U})] \quad (A.6)$$

So we have

$$\sqrt{N}(\hat{\gamma}_{IV} - \gamma) = S^{-1}_{(W-\tilde{W}),(Y_{-1}-\tilde{Y}_{-1})}\sqrt{N}S_{(W-\tilde{W}),(G-\tilde{G})+(U-\tilde{U})}. \quad (A.7)$$

We will show the followings:
(i) $S_{(W-\tilde{W}),(X-\tilde{X})} = \Psi + o_p(1)$;
(ii) $S_{(W-\tilde{W}),(G-\tilde{G})} = o_p(N^{-1/2})$;
(iii) $S_{(W-\tilde{W}),\tilde{U}} = o_p(N^{-1/2})$; and
(iv) $\sqrt{N}S_{(W-\tilde{W}),U} \to N(0,\Lambda)$ in distribution.
(i) Proof of $S_{(W-\tilde{W}),(X-\tilde{X})} = \Psi + o_p(1)$.

Let $W = [W - E_{\mathcal{G}}(W)] + E_{\mathcal{G}}(W) = \epsilon + h$. Then $\tilde{W} = \tilde{\epsilon} + \tilde{h}$. Let $X = [X - E_{\mathcal{G}}(X)] + E_{\mathcal{G}}(X) = \tau + \varkappa$. Then $\tilde{X} = \tilde{\tau} + \tilde{\varkappa}$. Follow the proof in Theorem 2.1, we have

$$S_{(W-\tilde{W}),(X-\tilde{X})} = S_{\epsilon+(h-\tilde{h}-\tilde{\epsilon}),\tau+(\varkappa-\tilde{\varkappa}-\tilde{\tau})} = \frac{1}{T}\sum_{t=1}^{T} E(\epsilon_{it}\tau'_{it}) + o_p(1) = \Psi + o_p(1). \quad (A.8)$$

(ii) Proof of $S_{(W-\tilde{W}),(G-\tilde{G})} = o_p(N^{-1/2})$.

The proof is similar to that in Theorem 2.1 and thus omitted.

(iii) Proof of $S_{(W-\tilde{W}),\tilde{U}} = o_p(N^{-1/2})$.

The proof is similar to that in Theorem 2.1 and thus omitted.

(iv) Proof of $\sqrt{N}S_{(W-\tilde{W}),U} \to N(0,\Lambda)$ in distribution.

$S_{(W-\tilde{W}),U} = S_{\epsilon+(h-\tilde{h}-\tilde{\epsilon}),U} = S_{\epsilon,U} + S_{(h-\tilde{h}),U} - S_{\tilde{\epsilon},U}$. We consider them separately.

(1) We have $\sqrt{N} S_{\epsilon,U} = \sqrt{N} \sum_i \epsilon_i u_i' \to N(0, \Lambda)$ by Levi-Lindberg central limit theorem.

(2) We have

$$E[\left\| S_{(h-\tilde{h}),U} \right\|^2 W, Z] = (NT)^{-2} tr[(h-\tilde{h})(h-\tilde{h})' E(UU'|Z)]$$
$$\leq C(NT)^{-1} tr(S_{h-\tilde{h}}) = o_p(N^{-1})$$

by lemma A.4 and lemma A.5. Hence, $S_{(h-\tilde{h}),U} = o_p(N^{-1/2})$.

(3) Similar to that in the proof of Theorem 2.1, $S_{\tilde{\epsilon},U} = o_p(N^{-1/2})$.

So we have $\sqrt{N} S_{(W-\tilde{W}),U} \to N(0, \Lambda)$ in distribution.

From (i) - (iv), we have $\sqrt{N}(\hat{\gamma}_{IV} - \gamma) \to N(0, \Psi^{-1}\Lambda\Psi^{-1})$ in distribution.

REFERENCES

Baltagi, B. H., 1995, *Econometric Analysis of Panel Data*. Wiley, Chichester.

Dechevsky, L. and S. Penez, 1997, On shape-preserving probabilistic wavelet approximators. *Stochastic Analysis and Applications* **15**, 187–215

Fan, J., W. Härdle, and E. Mammen, 1998, Direct estimation of low dimensional components in additive models. *Annals of Statistics* **26**, 943–971.

Fan, Y. and Q. Li, 1996, On estimating additive partially linear models, unpublished manuscript.

Li, Q., 2000, Efficient estimation of additive partially linear models. *International Economic Review* **41**, 1073–1092.

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

Linton, O. B. and J. P. Nielson, 1995, A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 91–100

Newey, W. K., 1994, Kernel estimation of partial means in a general variance estimator. *Econometric Theory* **10**, 233-253

Newey, W. K., 1997, Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* **79**, 147–168

Sperlich, S., D. Tjstheim, and L. Yang, 1999, Nonparametric estimation and testing of interaction in additive models, forthcoming, *Econometric Theory*

Tjstheim, D. and B. H. Auestad, 1994, Nonparametric identification of nonlinear time series: Prejections. *Journal of American Statistical Association* **89**, 1398–1409.