

# Nonparametric and Semiparametric Panel Econometric Models: Estimation and Testing\*

Liangjun Su<sup>a</sup>, Aman Ullah<sup>b</sup>

<sup>a</sup>*School of Economics, Singapore Management University, ljsu@smu.edu.sg*

<sup>b</sup>*Department of Economics, University of California, Riverside, aman.ullah@ucr.edu*

April 6, 2010

## ABSTRACT

This paper gives a selective review on the recent developments of nonparametric and semi-parametric panel data models. We focus on the conventional panel data models with one-way error component structure, partially linear panel data models, varying coefficient panel data models, nonparametric panel data models with multi-factor error structure, and nonseparable nonparametric panel data models. For each area, we discuss the basic models and ideas of estimation, and comment on the asymptotic properties of different estimators and specification tests. Much theoretical and empirical research is needed in this emerging area.

**KEY WORDS:** Cross section dependence; fixed effects; hypothesis testing; nonadditive model; nonparametric GMM; nonseparable model; partially linear panel data model; random effects; varying coefficient model;

---

\*The first author gratefully acknowledges the financial support from the NSFC under the grant numbers 70501001 and 70601001. The second author acknowledges the financial support from the academic senate, UCR.

# 1 Introduction

There exists enormous literature on the development of panel data models in the last five decades or so. The readers are referred to Arellano (2003), Hsiao (2003), and Baltagi (2008) for an overview of this literature. Nevertheless, these books only focus on the study of parametric panel data models which can be misspecified. Estimators from misspecified models are often inconsistent, invalidating the subsequent statistical inference. For this reason, we also observe a rapid growth of the literature on nonparametric (NP) and semiparametric (SP) panel data models in the last fifteen years. For an early review on this latter literature, the readers are referred to Ullah and Roy (1998). See also Ai and Li (2008) whose survey focuses on partially linear and limited dependent NP and SP panel data models.

In this paper, we review the recent literature on nonparametric and semiparametric panel data models. Given the space limitation, it is impossible to survey all the important developments in this literature. We choose to focus on the following areas:

- nonparametric panel data models with random effects
- nonparametric panel data models with fixed effects
- partially linear panel data models
- varying coefficient panel data models
- nonparametric panel data models with cross section dependence
- nonseparable nonparametric panel data models

The first two areas are limited to the conventional nonparametric panel data models with one-way error component structure:

$$y_{it} = m(x_{it}) + \varepsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (1.1)$$

where  $x_{it}$  is a  $p \times 1$  random vector,  $m(\cdot)$  is unknown smooth function,  $\varepsilon_{it}$  is the disturbance term that exists the one-way error component structure:

$$\varepsilon_{it} = \alpha_i + u_{it}. \quad (1.2)$$

Here,  $\alpha_i$  represents the cross sectional heterogeneity parameters, and  $u_{it}$  is the idiosyncratic error term. As in the parametric framework,  $\alpha_i$  can be treated as either random or fixed so that we will have random effects or fixed effects nonparametric panel data models.

Given the notorious “curse of dimensionality” problem in the nonparametric literature, applications of (1.1) may be limited in practice. This motivates the fast developments of two classes of semiparametric panel data models, namely, partially linear panel data models and varying coefficient panel data models. In Section 4, we study the estimation of the following partially linear panel data models

$$y_{it} = x_{it}'\beta_0 + m(z_{it}) + \alpha_i + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (1.3)$$

where  $x_{it}$  and  $z_{it}$  are of dimensions  $p \times 1$  and  $q \times 1$ , respectively,  $\beta_0$  is a  $p \times 1$  vector of unknown parameters,  $m(\cdot)$  is an unknown smooth function,  $\alpha_i$  and  $u_{it}$  are as defined above. In Section 5, we study the estimation of the following varying coefficient panel data models

$$y_{it} = x_{it}'m(z_{it}) + \alpha_i + u_{it} = \sum_{d=1}^p x_{it,d}m_d(z_{it}) + \alpha_i + u_{it} \quad (1.4)$$

where the covariate  $z_{it}$  is a  $q \times 1$  vector,  $x_{it} = (x_{it,1}, \dots, x_{it,p})'$ , and  $m(\cdot) = (m_1(\cdot), \dots, m_p(\cdot))'$  has  $p$  unknown smooth functions.

The literature on the estimation of parametric panel data models with cross section dependence has been growing rapidly in the last decade. See Pesaran (2006) and Bai (2009) and the references therein. In Section 6 we consider the estimation of  $m_i$  in

$$y_{it} = m_i(x_{it}) + \gamma_{1i}'f_{1t} + \gamma_{2i}'f_{2t} + \varepsilon_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (1.5)$$

where  $m_i(\cdot)$  is an unknown smooth function from,  $f_{1t}$  is a  $q_1 \times 1$  vector of observed common factors,  $f_{2t}$  is a  $q_2 \times 1$  vector of unobserved common factors,  $\gamma_{1i}$  and  $\gamma_{2i}$  are factor loadings,  $\varepsilon_{it}$  is the usual idiosyncratic disturbance. Since  $\gamma_{2i}'f_{2t} + \varepsilon_{it}$  is treated as the error term, we say it exhibits multi-factor error structure. Specification tests can be conducted to test the homogeneous relationship ( $m_i$  does not depend on  $i$ ) and the existence of cross section dependence.

All previous works assume that the unobserved heterogeneity and idiosyncratic error term enter the nonparametric panel data model additively. In Section 7, we focus on the estimation of the following two models

$$y_{it} = m(x_{it}, \alpha_i) + u_{it} \quad (1.6)$$

and

$$y_{it} = m(x_{it}, \alpha_i, u_{it}) \quad (1.7)$$

where both  $m(\cdot, \cdot)$  and  $m(\cdot, \cdot, \cdot)$  are unknown functions, and  $\alpha_i$  and  $u_{it}$  are as defined above. Clearly, (1.6) is a partially separable model because the idiosyncratic disturbance enters the model additively; (1.7) is fully nonseparable. We also remark that specification testing can

be developed to test the monotonicity of the response variable in the individual heterogeneity parameter.

Throughout the paper, we restrict our attention to the balanced panel. We use  $i = 1, \dots, n$  to denote an individual and  $t = 1, \dots, T$  to denote time, but keep in mind that in some applications, the index  $t$  may not really mean time. For example,  $i$  may denote a family and  $t$  a specific child in the family. Unless otherwise stated, all asymptotic theories are established by passing  $n$  to infinity.  $T$  may also pass to infinity in some scenarios, say, in some dynamic panel data models or the panel data models with cross section dependence. For a natural number  $a$ , we use  $I_a$  to denote an  $a \times a$  identity matrix and  $l_a$  an  $a \times 1$  vector of ones.  $\otimes$  and  $\odot$  denote the Kronecker and Hadarmard products, respectively.

## 2 Nonparametric Panel Data Models with Random Effects

In this section, we consider nonparametric panel data models with random effects:

$$y_{it} = m(x_{it}) + \alpha_i + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (2.1)$$

where  $x_{it}$  is  $p \times 1$  vector of exogenous variables,  $\alpha_i$  is independently and identically distributed (i.i.d.)  $(0, \sigma_\alpha^2)$ ,  $u_{jt}$  is i.i.d.  $(0, \sigma_u^2)$ , and  $\alpha_i$  and  $u_{jt}$  are uncorrelated for all  $i, j = 1, \dots, n$  and  $t = 1, \dots, T$ . We remark that some of these assumptions can be relaxed and specification testing is also possible.

Let  $\varepsilon_{it} = \alpha_i + u_{it}$ ,  $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})'$  and  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ . Then  $\Sigma \equiv E(\varepsilon_i \varepsilon_i') = \sigma_\alpha^2 I_T + \sigma_u^2 l_T l_T'$  and  $\Omega \equiv E(\varepsilon \varepsilon') = I_n \otimes \Sigma$ . We first discuss local linear least squares (LLLS) estimator of  $m$  and its first order derivatives by ignoring the information contained in the variance-covariance matrix  $\Omega$  and then proceed to the more efficient estimation of  $m$  and its derivatives by exploring the information on  $\Omega$ .

### 2.1 Local linear least squares estimator

A local linear approximation of the model (2.1) can be written as

$$\begin{aligned} y_{it} &\approx m(x) + (x_{it} - x)' \beta(x) + \alpha_i + u_{it} \\ &= x_{it}(x) \delta(x) + \alpha_i + u_{it} \end{aligned}$$

where  $x_{it}$  is ‘‘close’’ to  $x$ ,  $x_{it}(x) = (1 \ (x_{it} - x)')'$ ,  $\beta(x) = \partial m(x) / \partial x$ , and  $\delta(x) = (m(x) \ \beta(x)')'$ . In a vector form, we can write

$$Y \approx \mathbf{X}(x) \delta(x) + \varepsilon \quad (2.2)$$

where  $Y = (y_{11}, \dots, y_{1T}, \dots, y_{n1}, \dots, y_{nT})'$ , and  $\mathbf{X}(x) = ((x_{11}(x), \dots, x_{1T}(x), \dots, x_{n1}(x), \dots, x_{nT}(x))'$ .

Let  $K_h(x) = h^{-p}K(x/h)$ , where  $K$  is a kernel function and  $h \equiv h(n)$  is a bandwidth parameter. Then the LLLS estimator of  $\delta(x)$  is obtained by choosing  $\delta$  to minimize

$$(Y - \mathbf{X}(x)\delta)' \mathbf{K}(x) (Y - \mathbf{X}(x)\delta). \quad (2.3)$$

where  $\mathbf{K}(x) = \text{diag}(K_h(x_{11} - x), \dots, K_h(x_{1T} - x), \dots, K_h(x_{n1} - x), \dots, K_h(x_{nT} - x))$  is an  $nT \times nT$  diagonal matrix. The solution to this minimization problem is given by

$$\hat{\delta}(x) = [\mathbf{X}(x)' \mathbf{K}(x) \mathbf{X}(x)]^{-1} \mathbf{X}(x)' \mathbf{K}(x) Y. \quad (2.4)$$

Denote the first component of  $\hat{\delta}(x)$  as  $\hat{m}(x)$  which estimates  $m(x)$ . It is straightforward to study the asymptotic properties of  $\hat{\delta}(x)$  and  $\hat{m}(x)$ ; see, e.g., see Li and Racine (2007).

## 2.2 More efficient estimation

Clearly, the estimator in (2.4) ignores the information on  $\Omega$ . To incorporate this, we can define a weighted LLLS estimator of  $\delta(x)$  by choosing  $\delta$  to minimize

$$[Y - \mathbf{X}(x)\delta]' \mathbf{W}(x) [Y - \mathbf{X}(x)\delta]$$

which gives

$$\hat{\delta}_{\mathbf{W}}(x) = [\mathbf{X}(x)' \mathbf{W}(x) \mathbf{X}(x)]^{-1} \mathbf{X}(x)' \mathbf{W}(x) Y \quad (2.5)$$

where  $\mathbf{W}(x)$  is a kernel-based weight matrix, see Henderson and Ullah (2005). Lin and Carroll (2000) have considered  $\mathbf{W}(x) = \mathbf{K}(x)^{1/2} \Omega^{-1} \mathbf{K}(x)^{1/2}$  and  $\mathbf{W}(x) = \Omega^{-1} \mathbf{K}(x)$ , and Ullah and Roy (1998) have suggested  $\mathbf{W}(x) = \Omega^{-\frac{1}{2}} \mathbf{K}(x) \Omega^{-\frac{1}{2}}$ . When  $\Omega$  is a diagonal matrix, these choices of  $\mathbf{W}(x)$  are the same.

For an operational estimate, we need to estimate  $\Omega$ . For this purpose, define

$$\hat{\sigma}_1^2 = \frac{T}{n} \sum_{i=1}^n \bar{\hat{\varepsilon}}_i^2, \quad \hat{\sigma}_u^2 = \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T (\hat{\varepsilon}_{it} - \bar{\hat{\varepsilon}}_i)^2 \quad (2.6)$$

where  $\bar{\hat{\varepsilon}}_i = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_{it}$  and  $\hat{\varepsilon}_{it} = y_{it} - \hat{m}(x_{it})$  is the LLLS residual. Noting that  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_u^2$  estimate  $\sigma_1^2 = T\sigma_\alpha^2 + \sigma_u^2$  and  $\sigma_u^2$ , respectively, we can estimate  $\sigma_\alpha^2$  by  $\hat{\sigma}_\alpha^2 = \frac{1}{T} (\hat{\sigma}_1^2 - \hat{\sigma}_u^2)$ . With these estimates, one can obtain an estimate  $\hat{\Omega}$  of  $\Omega$  with  $\sigma_\alpha^2$  and  $\sigma_u^2$  replaced by  $\hat{\sigma}_\alpha^2$  and  $\hat{\sigma}_u^2$ , respectively. The operational estimator of  $\delta(x)$  is given by

$$\hat{\delta}_{\hat{\mathbf{W}}}(x) = [\mathbf{X}(x)' \hat{\mathbf{W}}(x) \mathbf{X}(x)]^{-1} \mathbf{X}(x)' \hat{\mathbf{W}}(x) Y \quad (2.7)$$

where  $\widehat{\mathbf{W}}(x)$  is  $\mathbf{W}(x)$  with  $\Omega$  replaced by  $\hat{\Omega}$ . However, Lin and Carroll (2000) demonstrate that one can not achieve asymptotic improvement over the LLLS estimator by such weighted LLLS estimation. Henderson and Ullah (2008) also find similar observations in their Monte Carlo study by comparing these weighted estimators. They also show that the following two step estimator of Reckstuhl, Welsh, and Carroll (2000) is more efficient than the above weighted estimators as well as the conventional LLLS estimator.

This two step estimator of Ruckstuhl, Welsh, and Carroll (2000) is developed as follows. Let us write (2.1) in vector form:

$$Y = \mathbf{m}(X) + \varepsilon, \quad (2.8)$$

where  $X = (x_{11}, \dots, x_{1T}, \dots, x_{n1}, \dots, x_{nT})'$ ,  $\mathbf{m}(X) = (m(x_{11}), \dots, m(x_{1T}), \dots, m(x_{n1}), \dots, m(x_{nT}))'$ ,  $\varepsilon = \alpha \otimes l_T + U$ ,  $U = (u_{11}, \dots, u_{1T}, \dots, u_{n1}, \dots, u_{nT})'$ . Multiplying both sides of (2.8) by  $\Omega^{-\frac{1}{2}}$  yields

$$\begin{aligned} \Omega^{-\frac{1}{2}}Y &= \Omega^{-\frac{1}{2}}\mathbf{m}(X) + \Omega^{-\frac{1}{2}}\varepsilon \\ &= \Omega^{-\frac{1}{2}}\mathbf{m}(X) - \mathbf{m}(X) + \mathbf{m}(X) + \Omega^{-\frac{1}{2}}\varepsilon \end{aligned}$$

or

$$Y^* = \mathbf{m}(X) + \Omega^{-\frac{1}{2}}\varepsilon \quad (2.9)$$

where  $Y^* = \Omega^{-\frac{1}{2}}Y + (I - \Omega^{-\frac{1}{2}})\mathbf{m}(X)$  is the transformed variable and  $\Omega^{-\frac{1}{2}}\varepsilon$  has an identity variance-covariance matrix. However,  $Y^*$  is not observed. So, a feasible estimator based on this transformed model can be obtained via a two-step procedure. In the first step we can run the LLLS regression  $Y$  on  $X$  to obtain the estimate  $\hat{m}(x)$  of  $m(x)$  at each data point and the residuals, based on which we can obtain consistent estimate  $\hat{\Omega}$  of  $\Omega$  as discussed above. This gives  $\hat{Y}^* = \hat{\Omega}^{-\frac{1}{2}}Y + (I - \hat{\Omega}^{-\frac{1}{2}})\hat{\mathbf{m}}(X)$ , where  $\hat{\mathbf{m}}(X) = (\hat{m}(x_{11}), \dots, \hat{m}(x_{1T}), \dots, \hat{m}(x_{n1}), \dots, \hat{m}(x_{nT}))'$ . In the second step, we run the LLLS regression of  $\hat{Y}^*$  on  $X$ . Such two-step estimation performs better than the weighted LLLS estimator (Henderson and Ullah (2008)). The asymptotic property of this type of two-step estimators is established in Su and Ullah (2007). See also Martins-Filho and Yao (2009) and Su, Ullah and Wang (2010) for related research along this line.

### 3 Nonparametric Panel Data Model with Fixed Effects

In this section, we consider the following nonparametric panel data model with fixed effects

$$y_{it} = m(x_{it}) + \alpha_i + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (3.1)$$

where the covariate (regressor)  $x_{it}$  is of dimension  $p \times 1$ ,  $m(\cdot)$  is an unknown smooth function,  $\alpha_i$ 's are fixed effects heterogeneity parameters, and  $u_{it}$  is i.i.d. with zero mean, finite variance  $\sigma_u^2$  and independent of  $x_{jt}$  for all  $i, j$  and  $t$ . We assume  $\sum_{i=1}^n \alpha_i = 0$  (so that  $\alpha_1 = -\sum_{i=2}^n \alpha_i$ ) for the purpose of identification. Also, for the sake of simplicity,  $x_{it}$  is strictly exogenous. We are interested in consistent estimation of  $m(\cdot)$  and its first order derivative.

Following the notation in the previous section, we can approximate the model in (3.1) as follows

$$Y \approx \mathbf{X}(x) \delta(x) + D\alpha + U \quad (3.2)$$

where  $\alpha = (\alpha_2, \dots, \alpha_n)'$ ,  $D = (I_n \otimes I_T) d_n$ ,  $d_n = [-I_{n-1} \ I_n]'$ , and other notations are as defined above. Note that  $\alpha$  contains heterogeneity parameters that may be correlated with the idiosyncratic error term  $u_{it}$  and the regressor  $x_{it}$  as well. So the LLS estimator is generally inconsistent in this case.

### 3.1 Profile least squares estimators

We argue that  $\delta(x)$  in (3.2) can be estimated by using the idea of profile least squares. There are two alternative approaches here. In the first approach, one can profile out the individual effects parameter  $\alpha$  and consider the concentrated least squares for  $\delta(x)$ . In the second approach, one profiles out the nonparametric component  $\delta(x)$  and consider the concentrated least squares for  $\alpha$ . We discuss the first approach, followed by the second approach.

For the moment, we pretend  $\alpha$  is known and then we can estimate  $\delta(x)$  in (3.2) by choosing  $\delta$  to minimize the following criterion function

$$[Y - \mathbf{X}(x) \delta - D\alpha]' \mathbf{K}(x) [Y - \mathbf{X}(x) \delta - D\alpha]. \quad (3.3)$$

We denote the solution to the above minimization problem as  $\delta_\alpha(x)$ , which is the LLS estimator of  $\delta(x)$  by regressing  $y_{it} - \alpha_i$  on  $x_{it}$ . It is easy to verify that

$$\delta_\alpha(x) = S(x)(Y - D\alpha) \quad (3.4)$$

where

$$S(x) = [\mathbf{X}(x)' \mathbf{K}(x) \mathbf{X}(x)]^{-1} \mathbf{X}(x)' \mathbf{K}(x) \quad (3.5)$$

is a  $(p+1) \times nT$  matrix. In particular, the LLS estimator of  $m(x)$  is given by

$$m_\alpha(x) = e_1' \delta_\alpha(x) = e_1' S(x)(Y - D\alpha) = s(x)'(Y - D\alpha) \quad (3.6)$$

where  $e_1 = (1, 0, \dots, 0)'$  is a  $(p+1) \times 1$  vector, and  $s(x)' = e_1' S(x)$ .

However,  $\delta_\alpha(x)$  is not operational since it depends on the unknown parameter  $\alpha$ . This motivates us to profile out the nonparametric component  $m(x)$  in (3.1). Note that (3.1) can be written as

$$Y = \mathbf{m}(X) + D\alpha + U \quad (3.7)$$

To profile out  $\mathbf{m}(X)$  in the above regression, we consider choosing  $\alpha$  to minimize the following criterion function

$$[Y - D\alpha - \mathbf{m}_\alpha(X)]' [Y - D\alpha - \mathbf{m}_\alpha(X)] = (Y^* - D^*\alpha)'(Y^* - D^*\alpha), \quad (3.8)$$

where

$$\begin{aligned} \mathbf{m}_\alpha(X) &= [m_\alpha(x_{11}) \cdots m_\alpha(x_{1T}), \cdots, m_\alpha(x_{n1}) \cdots m_\alpha(x_{nT})] = \mathbf{S}(Y - D\alpha), \\ Y^* &= (I_{nT} - \mathbf{S})Y, \\ D^* &= (I_{nT} - \mathbf{S})D, \end{aligned}$$

$\mathbf{S} = (s_{11}, \cdots, s_{1T}, \cdots, s_{n1}, \cdots, s_{nT})'$  is an  $nT \times nT$  matrix, and  $s_{it} = s(x_{it})$ . Then the solution to the above minimization problem is given by

$$\hat{\alpha} = (D^*D^*)^{-1}D^*Y^* = (D'QD)^{-1}D'QY, \quad (3.9)$$

where  $Q = (I_{nT} - \mathbf{S})'(I_{nT} - \mathbf{S})$ . The estimator for  $\alpha_1$  is  $\hat{\alpha}_1 = -\sum_{i=2}^n \hat{\alpha}_i$ , where  $\hat{\alpha} = (\hat{\alpha}_2, \cdots, \hat{\alpha}_n)'$ .

The profile least squares estimator for  $\delta(x)$  and  $m(x)$  are given respectively by

$$\hat{\delta}(x) = \delta_{\hat{\alpha}}(x) = S(x)(Y - D\hat{\alpha}) = S(x)MY \quad (3.10)$$

and

$$\hat{m}(x) = m_{\hat{\alpha}}(x) = s(x)(Y - D\hat{\alpha}) = s(x)MY \quad (3.11)$$

where  $M = I_{nT} - D(D'QD)^{-1}D'Q$  is an  $nT \times nT$  matrix such that  $MD = 0$ . The asymptotic properties of  $\hat{\delta}(x)$  have been studied in Su and Ullah (2006) in the framework of partially linear panel data models.

An alternative way to obtain the estimates of  $\alpha$  and  $\delta(x)$  is to profile out  $\alpha$  first by choosing  $\alpha$  to minimize the following criterion function:

$$[Y - \mathbf{X}(x)\delta(x) - D\alpha]' \mathbf{K}(x) [Y - \mathbf{X}(x)\delta(x) - D\alpha]. \quad (3.12)$$

The solution to this minimization problem is given by

$$\tilde{\alpha}(x) = [D'\mathbf{K}(x)D]^{-1} D'\mathbf{K}(x) [Y - \mathbf{X}(x)\delta(x)]. \quad (3.13)$$



In the second stage, we substitute  $\tilde{\alpha}(x)$  in (3.12) to obtain the following concentrated weighted least squares objective function

$$[Y - \mathbf{X}(x)\delta(x)]' \mathbf{K}^*(x) [Y - \mathbf{X}(x)\delta(x)] \quad (3.14)$$

where  $\mathbf{K}^*(x) = M(x)\mathbf{K}(x)M(x)$  and  $M(x) = I_{nT} - D(D'\mathbf{K}(x)D)^{-1}D'\mathbf{K}(x)$  is such that  $M(x)D = 0$ . Choosing  $\delta(x)$  to minimize (3.14) yields the solution

$$\tilde{\delta}(x) = [\mathbf{X}(x)'\mathbf{K}^*(x)\mathbf{X}(x)]^{-1} \mathbf{X}(x)'\mathbf{K}^*(x)Y.$$

See Sun, Carroll, and Li (2009) for this estimator in a more general framework and its asymptotic properties. An operational estimator of  $\alpha(x)$  is obtained by substituting  $\delta(x)$  with  $\hat{\delta}(x)$  in (3.13). This approach, however, does not provide an estimator of  $\alpha$ .

### 3.2 Measure of goodness-of-fit

Now we present the measure of goodness-of-fit in the fixed effects model which can be similarly defined in other types of models. Let  $\hat{\mathbf{m}}(X) = (\hat{m}(x_{11}), \dots, \hat{m}(x_{1T}), \dots, \hat{m}(x_{n1}), \dots, \hat{m}(x_{nT}))'$ , and  $\hat{U} = Y - \hat{\mathbf{m}}(X) - D\hat{\alpha}$ . Noting that  $\hat{\mathbf{m}}(X) = \mathbf{S}MY$  and  $\hat{\alpha} = (D'QD)^{-1}D'QY$ , we have

$$\begin{aligned} Y &= \hat{\mathbf{m}}(X) + D\hat{\alpha} + \hat{U} \\ &= \mathbf{S}MY + D(D'QD)^{-1}D'QY + \hat{U} \\ &= \mathbf{S}MY + (I_{nT} - M)Y + \hat{U} = \hat{Y} + \hat{U}, \end{aligned}$$

where  $\hat{Y} = [I_{nT} + (\mathbf{S} - I_{nT})M]Y$  is the stack of the fitted values, and thus  $\hat{U} = (I_{nT} - \mathbf{S})MY$ . Under the assumption that  $u_{it}$  is i.i.d. across both  $i$  and  $t$ , we can estimate its variance  $\sigma_u^2$  by

$$\hat{\sigma}_u^2 = \frac{\hat{U}'\hat{U}}{\text{tr}(N)} = \frac{Y'NY}{\text{tr}(N)}$$

where  $N = M'QM$ . Conditional on  $X$ , we have

$$E(\hat{\sigma}_u^2|X) = \sigma_u^2 + \frac{1}{\text{tr}(N)}\mathbf{m}(X)'N\mathbf{m}(X). \quad (3.15)$$

Thus,  $\hat{\sigma}_u^2$  is unbiased only if  $N\mathbf{m}(X) = 0$ . In general, we can establish only the consistency of  $\hat{\sigma}_u^2$  for  $\sigma_u^2$ .

A global goodness of fit measure can be defined as

$$R^2 = \frac{\hat{Y}'\hat{Y}}{Y'Y}, \quad (3.16)$$

or obtained by calculating the square of correlation between  $Y$  and  $\hat{Y}$ . However, this may not have the same interpretation as in the case of linear regression model because  $Y'Y = \hat{Y}'\hat{Y} + \hat{U}'\hat{U} + 2\hat{U}'\hat{Y}$  but  $\hat{U}'\hat{Y}$  is not guaranteed to be zero.

In view of the above problem, we propose an alternative way to construct a goodness-of-fit measure as follows. First, we define a local  $R^2$  and then the global  $R^2$ . We write from (3.2)

$$\begin{aligned} Y &= \mathbf{X}(x)\hat{\delta}(x) + D\hat{\alpha} + \hat{U}_x \\ &= \mathbf{Z}(x)\hat{\gamma}(x) + \hat{U}_x, \end{aligned} \quad (3.17)$$

where  $\mathbf{Z}(x) = [\mathbf{X}(x) \ D]$ ,  $\hat{\gamma}(x) = [\hat{\delta}'(x) \ \hat{\alpha}']'$ , and  $\hat{U}_x \equiv Y - \mathbf{Z}(x)\hat{\gamma}(x)$ . Then

$$(Y - LY)' \vec{\mathbf{K}}(x)(Y - LY) = [Z(x)\hat{\gamma}(x) - LY]' \mathbf{K}(x) [Z(x)\hat{\gamma}(x) - LY] + \hat{U}_x' \mathbf{K}(x) \hat{U}_x \quad (3.18)$$

where  $L = l_{nT} l_{nT}' / (nT)$ ,  $\vec{\mathbf{K}}(x)$  is a diagonal matrix with typical elements  $K_h(x_{it} - x) / (nT \hat{f}(x))$  for  $i = 1, \dots, n$ , and  $t = 1, \dots, T$ ,  $\hat{f}(x) = (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T K_h(x_{it} - x)$ . Observe that  $\hat{\gamma}(x)$  can be written as

$$\hat{\gamma}(x) = A(x)Y, \quad A(x) = \begin{pmatrix} S(x)M \\ (DQD)^{-1}D'Q \end{pmatrix} \quad (3.19)$$

Thus we can write (3.18) as

$$Y'N_1(x)Y = Y'N_2(x)Y + Y'N_3(x)Y \quad (3.20)$$

where  $N_1(x) = (I_{nT} - L)'K(x)(I_{nT} - L)$ ,  $N_2(x) = [I_{nT} - \mathbf{Z}(x)A(x) - L]' \vec{\mathbf{K}}(x) [I_{nT} - \mathbf{Z}(x)A(x) - L]$ ,  $N_3 = [I_{nT} - \mathbf{Z}(x)A(x)]' \vec{\mathbf{K}}(x) [I_{nT} - \mathbf{Z}(x)A(x)]$ , and  $N_2(x)N_3(x) = 0$ . It follows that

$$TSS(x) = SSR(x) + RSS(x) \quad (3.21)$$

where  $TSS(x) = Y'N_1(x)Y$ ,  $SSR(x) = Y'N_2(x)Y$ , and  $RSS(x) = Y'N_3(x)Y$ .

Thus (3.21) represents a local analysis of variance (ANOVA) so that we can define a local  $R^2$  as

$$R^2(x) = \frac{SSR(x)}{TSS(x)} = 1 - \frac{RSS(x)}{TSS(x)} \quad (3.22)$$

where  $0 \leq R^2 \leq 1$  by construction. Further, a global  $R^2$  can be defined as

$$R^2 = \frac{SSR}{TSS} = 1 - \frac{RSS}{TSS} \quad (3.23)$$

where  $SSR = \int_x SSR(x)\hat{f}(x)dx$ ,  $TSS = \int_x TSS(x)\hat{f}(x)dx$  and  $RSS = \int_x RSS(x)\hat{f}(x)dx$ . It is worth pointing out that  $TSS = \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \bar{y})^2$  where  $\bar{y} = (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T y_{it}$ .

### 3.3 Differencing method

Let  $\Delta y_{it} = y_{it} - y_{i,t-1}$ .  $\Delta u_{it}$  is similarly defined. As in the usual differencing method, we can consider subtracting the model in (3.1) for time  $t$  from that for time  $t - 1$  so that

$$\Delta y_{it} = m(x_{it}) - m(x_{i,t-1}) + \Delta u_{it} \quad (3.24)$$

or subtracting the equation for time  $t$  from that for time 1 so that

$$y_{it} - y_{i1} = m(x_{it}) - m(x_{i1}) + u_{it} - u_{i1}. \quad (3.25)$$

Another method, which is conventional, removes the fixed effects by deducting each equation from the cross-time average. This gives

$$y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it} = m(x_{it}) - \frac{1}{T} \sum_{s=1}^T m(x_{is}) + u_{it} - \frac{1}{T} \sum_{s=1}^T u_{is} \quad (3.26)$$

or

$$y_{it}^* = \sum_{s=1}^T d_{ts} m(x_{is}) + u_{it}^* \quad (3.27)$$

where  $d_{ts} = -\frac{1}{T}$  if  $s \neq t$  and  $1 - \frac{1}{T}$  otherwise, and  $\sum_{s=1}^T d_{ts} = 0$  for all  $t$ ,  $y_{it}^* = y_{it} - T^{-1} \sum_{t=1}^T y_{it}$ , and  $u_{it}^* = u_{it} - T^{-1} \sum_{t=1}^T u_{it}$ .

For each  $i$ , the right hand sides of equations (3.24)-(3.26) contain linear combination of  $m(x_{is})$ ,  $s = 1, \dots, T$ . We discuss the estimation corresponding to each of these differencing methods. To proceed, it is worth mentioning that some components of the function  $m(\cdot)$  may not be fully identified via differencing methods. For example, if  $m(x_{it}) = a + m_1(x_{it})$ , then the difference will wipe out  $a$  and hence we can only estimate  $m(x_{it})$  under some identification restriction. Similar issues arise when we consider the case of varying functional coefficient models later on if differencing methods are called upon.

For the first differencing (FD) model in (3.24), Li and Stengos (1996) suggest estimation of  $m(x_{it}, x_{i,t-1}) = m(x_{it}) - m(x_{i,t-1})$  by doing a local linear regression of  $\Delta y_{it}$  on  $x_{it}$  and  $x_{i,t-1}$ . Then we can obtain estimates of  $m(x)$  by the method of estimating nonparametric additive models, e.g., by the marginal integration method of Linton and Nielson (1995) or by the backfitting method. For example, after we obtain estimates  $\hat{m}(x, x_{i,t-1})$  of  $m(x, x_{i,t-1})$  for  $i = 1, \dots, n$ , and  $t = 2, \dots, T$ , we can estimate  $m(x)$  by  $\hat{m}(x) = (n(T-1))^{-1} \sum_{i=1}^n \sum_{t=2}^T \hat{m}(x, x_{i,t-1})$ , apart from the concerns discussed above for the differencing method. We also note that this method suffers from the ‘‘curse of dimensionality’’ problem in calculating  $\hat{m}(x, x_{i,t-1})$  because it involves estimating a  $2p$ -dimensional nonparametric object. In view of this, Baltagi and Li

(2002) obtain consistent estimators of  $m(x)$  by considering the first differencing method and using series approximation for the nonparametric component.

Based on the differencing model in (3.25), Henderson, Carroll, and Li (2008) propose an iterative kernel estimator of  $m(x)$  and establish the asymptotic normality for their estimator. But this estimator is also subject to the comments on differencing given above. Since this method is elaborated in detail in Li and Racine (2007), we skip it for brevity.

Now we consider eliminating the fixed effects via the sample average over time. Following (3.26), we write

$$y_{it} - \bar{y}_i = m(x_{it}) - \bar{m}_i + u_{it} - \bar{u}_i$$

where  $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$ ,  $\bar{u}_i = T^{-1} \sum_{t=1}^T u_{it}$ , and  $\bar{m}_i = T^{-1} \sum_{t=1}^T m(x_{it})$ . Then writing  $m(x_{it}) \approx m(x) + (x_{it} - x)' \beta(x)$  with  $\beta(x) = \partial m(x) / \partial x$ , we get

$$y_{it} - \bar{y}_i \approx (x_{it} - \bar{x}_i)' \beta(x) + u_{it} - \bar{u}_i,$$

where  $\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$ . The local linear within-group estimator of  $\beta(x)$  then follows as

$$\hat{\beta}_W(x) = \left( \sum_{i=1}^n \sum_{t=1}^T (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' K_h(x_{it} - x) \right)^{-1} \sum_{i=1}^n \sum_{t=1}^T (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) K_h(x_{it} - x).$$

Similarly, if we use the first differencing method, then the local linear estimator of  $\beta(x)$  for some fixed element  $x$  in  $\{x_{it}, i = 1, \dots, n; t = 1, \dots, T\}$  is given by

$$\hat{\beta}_D(x) = \left( \sum_{i=1}^n \sum_{t=1}^T \Delta x_{it} \Delta x_{it}' K_h(x_{it} - x) \right)^{-1} \sum_{i=1}^n \sum_{t=1}^T \Delta x_{it} \Delta y_{it} K_h(x_{it} - x).$$

Lee and Mukherjee (2008) study the asymptotic properties of the above two estimators. For the case where  $x_{it}$  is a scalar random variable (i.e.,  $p = 1$ ), they show that under some standard assumptions,

$$E \left[ \hat{\beta}_W(x) - \beta(x) | X \right] = \frac{m^{(2)}(x) [\mu_1(x) \mu_2(x) + \mu_3(x)]}{2 [\mu_1^2(x) + \mu_2(x)]} + O_p(h^2)$$

and

$$E \left[ \hat{\beta}_D(x) - \beta(x) | X \right] = \frac{m^{(2)}(x) \mu_3(x)}{2 \mu_2(x)} + O_p(h^2),$$

where  $\mu_j(x) = E(x_{it} - x)^j < \infty$  for  $j = 1, 2, 3$ , and  $m^{(2)}(x) = \partial^2 m(x) / \partial x^2$ .

It is clear from the above expressions that both the conventional within-group estimator and first-difference estimator are inconsistent because as  $n \rightarrow \infty$  and  $h \rightarrow 0$  we have a non-degenerating bias. This bias, however, is zero when the true regression function  $m(x)$  is linear

in  $x$  or  $x_{it}$  is symmetric around the point of evaluation  $x$  such that  $\mu_j(x) = 0$  for  $j = 1$  and  $3$ . As Lee and Mukherjee (2008) observed, the non-vanishing biases arise because the difference equations are not locally weighted by the differenced variables whereas the original model is a local approximation around the point  $x$  of the original variable  $x_{it}$ . In other words, the differenced equations are initially localized around a value of  $x_{it}$  without considering the rest of values  $x_{is}$ ,  $s \neq t$ . But  $|x_{is} - x|$  can not be small enough uniformly over all  $i$  and  $s \neq t$  such that  $\max_{i,s} |x_{is} - x| < Ch$  for some  $C < \infty$ , so that the differenced remainder terms cannot be tending to zero. Here the remainder term is  $\bar{R}_{it} = (T - 1)^{-1} \sum_{s=1, s \neq t}^T R_{is}(x)$  when  $x = x_{it}$ , where  $R_{is}(x) = \frac{1}{2}m^{(2)}(x_{is}^*)(x_{is} - x)^2$  and  $x_{is}^*$  lies between  $x_{is}$  and  $x$ . Obviously, the biases do not vanish even when  $T \rightarrow \infty$ . Again, this is due to the local approximation of  $m(x)$  at given  $x_{it}$  as indicated in the kernel weight function  $K_h(x_{it} - x)$ , but the local estimator involves the average of  $(x_{is} - x)$  for all  $i$  and  $s \neq t$ .

We notice that the estimator  $\hat{\beta}_W(x)$ , based on conventional within average differencing, was introduced in Ullah and Roy (1998), whereas the estimator  $\hat{\beta}_D(x)$  is based on the first differencing method in Li and Stengos (1996) and Mundra (2005). In views of this, Mukherjee (2002) and Mukherjee and Ullah (2003)(also Henderson and Ullah (2005), p.406) proposed elimination of the fixed effects by taking the within differencing in using local weighted average at  $x$ .

Define the locally weighted averages as

$$\bar{x}_i(x) = \frac{\sum_{s=1, s \neq t}^T x_{is} K_h(x_{is} - x)}{\sum_{s=1, s \neq t}^T K_h(x_{is} - x)}, \text{ and } \bar{y}_i(x) = \frac{\sum_{s=1, s \neq t}^T y_{is} K_h(x_{is} - x)}{\sum_{s=1, s \neq t}^T K_h(x_{is} - x)}.$$

The local-within leave-one-out estimator of  $\beta(x)$  for  $x = x_{it}$  is given by

$$\tilde{\beta}(x) = \left[ \sum_{i=1}^n \sum_{s=1, s \neq t}^T x_{is}^*(x) x_{is}^*(x)' K_h(x_{is} - x) \right]^{-1} \sum_{i=1}^n \sum_{s=1, s \neq t}^T x_{is}^*(x) y_{is}^*(x) K_h(x_{is} - x),$$

where  $x_{is}^*(x) = x_{is} - \bar{x}_i(x)$  and  $y_{is}^*(x) = y_{is} - \bar{y}_i(x)$ . Clearly, this estimator is the solution to the problem

$$\min_{\beta} \sum_{i=1}^n \sum_{s=1, s \neq t}^T [y_{is}^*(x) - x_{is}^*(x)' \beta]^2 K_h(x_{is} - x)$$

For  $p = 1$ , Lee and Mukherjee (2008) provide the following results under the standard regularity assumptions: (i)  $u_{it}$  is i.i.d. with mean 0 and variance  $\sigma^2$  and it is independent of  $\alpha_i$  and  $x_{it}$  for all  $i$  and  $t$ , (ii)  $\alpha_i$  is i.i.d., (iii)  $x_{it}$  is i.i.d. with probability density function (p.d.f.)

$f(x)$  whose support is bounded, and for the interior point  $x$ , it is twice differentiable with bounded second-order derivative, (iv)  $m(x)$  is twice differentiable with bounded second-order derivative, (v)  $K$  is compactly supported, bounded, and symmetric second-order kernel, (vi)  $h \rightarrow 0$  as  $nh \rightarrow 0$ ,  $Th \rightarrow 0$  and  $nTh^3 \rightarrow 0$  as  $n, T \rightarrow \infty$ . Under these assumptions,

$$\begin{aligned} E \left[ \tilde{\beta}(x) - \beta(x) | X \right] &= \frac{h^2}{2} \left( \frac{m^{(2)}(x)f^{(1)}(x)}{f(x)} \right) \left( \frac{\kappa_4 - \kappa_2^2}{\kappa_2} \right) + O(h^2) \\ \text{Var}(\tilde{\beta}(x)|X) &= \frac{1}{nTh^3} \left( \frac{\sigma^2}{f(x)} \right) \frac{\omega_2}{\kappa_2^2} + O_p \left( \frac{1}{nTh^3} \right) \end{aligned}$$

where  $f^{(1)}(x) = \partial f(x)/\partial x$ ,  $\omega_2 = \int x^2 K(x)^2 dx$ , and  $\kappa_l = \int z^l K(x) dx$  for  $l = 2, 4$ . Further, using the above results one can show that the optimal bandwidth in minimizing  $MSE(\tilde{\beta}(x))$  is proportional to  $(nT)^{-1/7}$ . If  $m(x)$  is three times differentiable then in the bias of  $\tilde{\beta}(x)$  we add an additional term  $h^2 m^{(3)}(x) \kappa_4 / (6\kappa_2)$  where  $m^{(3)}(x) = \partial^3 m(x) / \partial x^3$ . These results show that for the local weighted average differencing the orders of magnitudes of bias and variance are the same as those of the local linear derivative estimator. See Pagan and Ullah (1999) and Li and Racine (2007). However, the magnitude of bias differs with  $-h^2 m^{(2)}(x) f^{(1)}(x) \kappa_2 / (2f(x))$  which arises due to the local weighted average differencing, but the magnitude of variance remains the same.

A similar idea can be applied to the case of time differenced model. Lee and Mukherjee (2008) suggest estimating  $\beta(x)$  by

$$\hat{\beta}(x) = \min_{\beta} \sum_{i=1}^n \sum_{t=1}^T (\Delta y_{it} - \beta \Delta x_{it})^2 K_h(x_{it} - x, x_{i,t-1} - x).$$

But this method does not go through when the model has time-heterogeneity.

Finally, although the estimator of  $m(x)$  is not directly obtained from the objective function, an estimator of  $m(x)$  could be written as

$$\tilde{m}(x) = \frac{1}{n} \sum_{i=1}^n \tilde{m}_i(x)$$

where  $\tilde{m}_i(x) = \bar{y}_i(x) - \tilde{\beta}(x) \bar{x}_i(x)$ . See Lee and Mukherjee (2008) for an alternative proposal. The properties of  $\tilde{m}_i(x)$  are not yet known, also the asymptotic normality of  $\tilde{\beta}(x)$ .

### 3.4 Series estimation

The above estimation procedures are invalid if  $x_{it}$  contains lagged dependent variables. Lee (2008) considers series estimation of the following nonparametric dynamic panel data model:

$$y_{it} = m(y_{i,t-1}) + \alpha_i + u_{it}, \tag{3.28}$$

where  $\alpha_i$  can be eliminated via first differencing or within-group difference. Let  $m^*(y_{i,t-1}^*) = m^*(y_{i,t-1}) - T^{-1} \sum_{s=1}^T m(y_{i,s-1})$  and similarly define  $y_{it}^*$  and  $u_{it}^*$ . Then we have the within-group transformation of the above model as follows:

$$y_{it}^* = m^*(y_{i,t-1}^*) + u_{it}^*. \quad (3.29)$$

Lee's (2008) series estimator of  $m$  is based on the above within-group transformation. Under the assumption that  $\lim_{n,T \rightarrow \infty} n/T = \kappa \in (0, \infty)$ , he finds that the series estimator is asymptotically biased and proposes a bias-corrected series estimator. Asymptotic normality is also established.

### 3.5 A nonparametric Hausman test

To test the random effects against the fixed effects specification in the model  $y_{it} = m(x_{it}) + \alpha_i + u_{it}$ , we can specify the null and alternative hypotheses as

$$H_0 : E(\alpha_i | x_{i1}, \dots, x_{iT}) = 0 \text{ a.s. versus } H_1 : \text{the negation of } H_0,$$

where a.s. is an abbreviation for almost surely. If we maintain the assumption that  $E(u_{it} | x_{i1}, \dots, x_{iT}) = 0$ , the null hypothesis can also be written as

$$H_0 : E(\varepsilon_{it} | x_{i1}, \dots, x_{iT}) = 0 \text{ a.s.}$$

where  $\varepsilon_{it} = \alpha_i + u_{it}$ . Then one can propose a test based on the sample analogue of

$$J = E\{\varepsilon_{it} E(\varepsilon_{it} | x_{it}) f(x_{it})\}$$

where  $f(\cdot)$  is the p.d.f. of  $x_{it}$  because  $J = 0$  under  $H_0$  and  $J = E\{[E(\varepsilon_{it} | x_{it})]^2 f(x_{it})\} > 0$  under  $H_1$ . A feasible test statistic is given by

$$J_n = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \hat{\varepsilon}_{it} \hat{E}_{-it}(\hat{\varepsilon}_{it} | x_{it}) \hat{f}_{-it}(x_{it})$$

where  $\hat{\varepsilon}_{it}$  is the residual from the random effects regression,  $\hat{f}_{-it}(x_{it})$  and  $\hat{E}_{-it}(\hat{\varepsilon}_{it} | x_{it})$  are leave-one-out kernel estimates of  $f(x_{it})$  and  $E(\varepsilon_{it} | x_{it})$ , respectively by using observations on  $\{x_{it}, \hat{\varepsilon}_{it}\}$ . This test statistic is considered in Henderson, Carroll, and Li (2008). But they do not provide a formal asymptotic distributional analysis. Instead, they propose a bootstrap method to obtain the critical values and demonstrate through simulations that  $J_n$  works reasonably well in finite samples.

## 4 Partially Linear Panel Data Models

In this section we review the literature on partially linear panel data models. We focus on the following model

$$y_{it} = x'_{it}\beta_0 + m(z_{it}) + \alpha_i + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (4.1)$$

where  $x_{it}$  and  $z_{it}$  are of dimensions  $p \times 1$  and  $q \times 1$ , respectively,  $\beta_0$  is a  $p \times 1$  vector of unknown parameters,  $m(\cdot)$  is an unknown smooth function,  $\alpha_i$  is random or fixed effects, and  $u_{it}$  is the idiosyncratic disturbance. We will first discuss the estimation of (4.1) when  $\alpha_i$  represents the random effects and then the fixed effects model. We also comment on extensions and specification tests.

### 4.1 Partially linear panel data models with random effects

Let  $\varepsilon_{it} = \alpha_i + u_{it}$ . We can rewrite (4.1) as

$$y_{it} = x'_{it}\beta_0 + m(z_{it}) + \varepsilon_{it}. \quad (4.2)$$

In the literature, it is frequently assumed that

$$E(\varepsilon_{it}|z_{it}) = 0. \quad (4.3)$$

Note that this assumption does not rule out the dependence between  $x_{it}$  and  $\varepsilon_{it}$ . As a matter of fact, some or all the components of  $x_{it}$  may be correlated with the error  $\varepsilon_{it}$ . Li and Stengos (1996) discuss the estimation of (4.1) for the case of random effects model.

Under the assumption in (4.3), we can take conditional expectation of (4.1) given  $z_{it}$  on both sides to yield

$$E(y_{it}|z_{it}) = E(x_{it}|z_{it})'\beta_0 + m(z_{it}). \quad (4.4)$$

Subtracting (4.4) from (4.1), we have

$$Y_{it} = X'_{it}\beta_0 + \varepsilon_{it}. \quad (4.5)$$

Let  $Y_{it} = y_{it} - E(y_{it}|z_{it})$  and  $X_{it} = x_{it} - E(x_{it}|z_{it})$ . So (4.5) is a linear panel data model with dependent variable  $Y_{it}$  and independent variable  $X_{it}$ . If  $(Y_{it}, X_{it})$  were observable, we can estimate  $\beta_0$  by the parametric methods. For simplicity, we assume that there exists an instrumental variable (IV)  $w_{it} \in \mathbb{R}^p$ , such that

$$E(\varepsilon_{it}|w_{it}, z_{it}) = 0 \text{ and } E(x'_{it}w_{it}) \neq 0. \quad (4.6)$$



We then can estimate  $\beta_0$  by the IV method<sup>1</sup>:

$$\tilde{\beta} = (W'X)^{-1} W'Y = \beta_0 + (W'X)^{-1} W'\varepsilon, \quad (4.7)$$

where  $W_{it} = w_{it} - E(w_{it}|z_{it})$ ,  $Y = (Y_{11}, \dots, Y_{1T}, \dots, Y_{n1}, \dots, Y_{nT})'$ ,  $X$ ,  $W$ , and  $\varepsilon$  are similarly defined. Under (4.6), we have  $E(\varepsilon_{it}|W_{it}) = 0$ , so the IV estimator  $\tilde{\beta}$  is consistent. Nevertheless, it is infeasible since the conditional expectations  $E(y_{it}|z_{it})$ ,  $E(x_{it}|z_{it})$ , and  $E(w_{it}|z_{it})$  are unknown to us. As before, these conditional expectations can be consistently estimated using nonparametric methods. To avoid random denominator problem, we choose to use the marginal p.d.f.  $f(\cdot)$  of  $z_{it}$  as the weighting function as in Li and Stengos (1996).

Multiplying (4.5) by  $f_{it} = f(z_{it})$ , we have

$$Y_{it}f_{it} = (X_{it}f_{it})' \beta_0 + \varepsilon_{it}f_{it}. \quad (4.8)$$

Now one can estimate the unknown finite dimensional parameter  $\beta_0$  by regressing  $Y_{it}f_{it}$  on  $X_{it}f_{it}$  using  $W_{it}f_{it}$  as an IV. The infeasible IV estimator is obtain

$$\tilde{\beta}_f = \left( \sum_{i=1}^n \sum_{t=1}^T W_{it} X'_{it} f_{it}^2 \right)^{-1} \sum_{i=1}^n \sum_{t=1}^T W_{it} Y_{it} f_{it}^2, \quad (4.9)$$

It is easy to show that  $\tilde{\beta}_f$  is asymptotically normally distributed, i.e.,

$$\sqrt{n} (\tilde{\beta}_f - \beta_0) \xrightarrow{d} N \left( 0, \Phi_f^{-1} \Psi_f \Phi_f^{-1} \right), \quad (4.10)$$

where  $\Phi_f = T^{-1} \sum_{t=1}^T E [W_{it} X'_{it} f_{it}^2]$ , and  $\Psi_f = T^{-2} \sum_{t=1}^T \sum_{s=1}^T E (u_{it} u_{is} W_{it} W'_{it} f_{it}^2 f_{is}^2)$ .

To proceed, we estimate  $f_{it}$  by  $\hat{f}(z_{it}) = (nT)^{-1} \sum_{j=1}^n \sum_{s=1}^T K_{it,js}$  and  $E(y_{it}|z_{it})$  by  $\hat{y}_{it} = (nT)^{-1} \sum_{j=1}^n \sum_{s=1}^T y_{js} K_{it,js} / \hat{f}(z_{it})$ , where  $K_{it,js} = K_h(z_{it} - z_{js})$ . The estimators  $\hat{x}_{it}$  and  $\hat{w}_{it}$  of  $E(x_{it}|z_{it})$  and  $E(w_{it}|z_{it})$  are similarly defined. A feasible estimator of  $\beta_0$  can be obtained by replacing  $Y_{it}$ ,  $X_{it}$ ,  $Z_{it}$ , and  $f_{it}$  with  $y_{it} - \hat{y}_{it}$ ,  $x_{it} - \hat{x}_{it}$ ,  $w_{it} - \hat{w}_{it}$ , and  $\hat{f}(z_{it})$ . This leads to the following feasible density-weighted estimator of  $\beta_0$ :

$$\hat{\beta}_f = \left( \sum_{i=1}^n \sum_{t=1}^T (w_{it} - \hat{w}_{it})(x_{it} - \hat{x}_{it})' \hat{f}(z_{it})^2 \right)^{-1} \sum_{i=1}^n \sum_{t=1}^T (w_{it} - \hat{w}_{it})(y_{it} - \hat{y}_{it}) \hat{f}(z_{it})^2. \quad (4.11)$$

Under some regularity conditions, Li and Stengos (1996) have established the asymptotic normality of  $\hat{\beta}_f$ :

$$\sqrt{n} (\hat{\beta}_f - \beta_0) \xrightarrow{d} N \left( 0, \Phi_f^{-1} \Psi_f \Phi_f^{-1} \right).$$

---

<sup>1</sup>If the dimension of  $w_{it}$  is  $l \geq p$ , the IV estimator of  $\beta_0$  is given by  $\tilde{\beta}_1 = (X'W(W'W)W'X)^{-1} X'W(W'W)W'Y$

For statistical inference on  $\beta_0$ , we need to estimate the asymptotic variance-covariance of  $\hat{\beta}_f$  consistently, which is straightforward.

After obtaining a  $\sqrt{n}$ -consistent estimator  $\hat{\beta}_f$  of  $\beta_0$ , we can estimate  $m(z)$  consistently by  $\hat{m}(z) = (nT)^{-1} \sum_{j=1}^n \sum_{s=1}^T (y_{js} - x'_{js} \hat{\beta}_f) K_{\tilde{h}}(z_{js} - z) / \tilde{f}(z)$ , where the bandwidth  $\tilde{h}$  is typically different from  $h$ , and  $\tilde{f}(z) = (nT)^{-1} \sum_{j=1}^n \sum_{s=1}^T K_{\tilde{h}}(z_{js} - z)$ . Since the nonparametric kernel estimator has a slower convergence rate than the parametric  $\sqrt{n}$ -rate, it is easy to show  $\hat{m}(z)$  has the same asymptotic distribution as  $\tilde{m}(z) = (nT)^{-1} \sum_{j=1}^n \sum_{s=1}^T (y_{js} - x'_{js} \beta_0) K_h(z_{js} - z) / \tilde{f}(z)$ .

It is worth mentioning the above method works in a variety of applications. In particular, it allows  $x_{it}$  to contain lagged dependent variable. Nevertheless, the above IV estimator of  $\beta_0$  is generally inefficient. When the error follows a one-way error component structure in the partially linear panel data model, Li and Ullah (1998) propose a feasible semiparametric generalized least squares (GLS) type estimator for estimating  $\beta_0$  and show that is asymptotically more efficient than the semiparametric ordinary least squares (OLS) type estimator. They also discuss the case for which the regressor of the parametric component is correlated with the error, and propose an IV GLS-type semiparametric estimator. They show that their estimator for the finite dimensional parameter is efficient. For brevity, we refer the reader directly to their paper.

## 4.2 Partially linear panel data models with fixed effects

We now discuss the estimation of (4.1) when  $\alpha_i$  represents the fixed effect. For the identification purpose, we can impose  $\sum_{i=1}^n \alpha_i = 0$ . For simplicity, we assume that  $z_{it}$  is strictly exogenous but allow  $x_{it}$  to be correlated with the error term  $u_{it}$ . We are interested in consistent estimation of  $\beta_0$  and  $m(\cdot)$ . As usual, we focus on the case where  $n$  is approaching infinity and  $T$  is fixed.

In principle, we can apply Li and Stengos (1996) or the method introduced in the previous section to estimate the fixed effect model. From (4.1), we can take the first difference as in the linear panel data model to obtain

$$y_{it} - y_{i,t-1} = (x_{it} - x_{i,t-1})' \beta_0 + [m(z_{it}) - m(z_{i,t-1})] + (u_{it} - u_{i,t-1}), \quad (4.12)$$

or

$$Y_{it} = X'_{it} \beta_0 + M(z_{it}, z_{i,t-1}) + U_{it}, \quad (4.13)$$

where  $Y_{it} = y_{it} - y_{i,t-1}$ ,  $X_{it} = x_{it} - x_{i,t-1}$ ,  $U_{it} = u_{it} - u_{i,t-1}$ , and  $M(z_{it}, z_{i,t-1}) = m(z_{it}) - m(z_{i,t-1})$ . Equation (4.13) is basically the same as (4.2) except that we know that  $U_{it}$  has a moving average structure. Nevertheless, this approach has several drawbacks. First, in order to

eliminate  $M(z_{it}, z_{i,t-1})$ , 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 it ignores the additive structure of (4.12) and requires the kernel function to be defined on  $\mathbb{R}^{2q}$  instead of  $\mathbb{R}^q$ . Secondly, although they propose a method to estimate the finite dimensional parameter  $\beta_0$  and their method can estimate  $M(z_{it}, z_{i,t-1})$ , they did not suggest how to estimate the original unknown function  $m(z_{it})$ . For this reason, Baltagi and Li (2002) consider series estimation of the model that imposes the additive structure of  $M(z_{it}, z_{i,t-1}) = m(z_{it}) - m(z_{i,t-1})$ .

In matrix form, (4.13) can be rewritten as

$$Y = X\beta_0 + M + U \quad (4.14)$$

where  $Y$  is an  $nT \times 1$  vector with typical element  $Y_{it}$ , and  $X$ ,  $M$  and  $U$  are similarly defined. Let  $Z$  denote an  $nT \times q$  matrix with typical row given by  $z_{it}$ .

A function  $\xi(z_{it}, z_{i,t-1})$  is said to be an additive class of functions  $\mathcal{M}$  if  $\xi(z_{it}, z_{i,t-1}) = m(z_{it}) - m(z_{i,t-1})$ ,  $m(\cdot)$  is twice differentiable in the interior of its support  $\mathcal{Z}$ , which is a compact subset of  $\mathbb{R}^q$  and  $E[m^2(z_{it})] < \infty$ . We will use series  $p^L(z)$  of  $L \times 1$  dimension to approximate  $m(z)$ , where  $L = L(n)$ . The approximation function  $p^L(z)$  has the following properties: (a)  $p^L(z, \tilde{z}) \equiv p^L(z) - p^L(\tilde{z}) \in \mathcal{M}$ ; (b) as  $L$  grows, there is a linear combination of  $p^L(z, \tilde{z})$  that can approximate any function in  $\mathcal{M}$  arbitrarily well in the sense of mean squared error. Therefore,  $p^L(z)$  approximates  $m(z)$  and  $p^L(z, \tilde{z}) \equiv p^L(z) - p^L(\tilde{z})$  approximates  $M(z, \tilde{z}) = m(z) - m(\tilde{z})$ :

$$p^L(z_{it}, z_{i,t-1}) = \begin{pmatrix} p_1(z_{it}) - p_1(z_{i,t-1}) \\ p_2(z_{it}) - p_2(z_{i,t-1}) \\ \vdots \\ p_L(z_{it}) - p_L(z_{i,t-1}) \end{pmatrix}. \quad (4.15)$$

For notational simplicity, define  $p_{it}^L = p^L(z_{it}, z_{i,t-1})$  and  $P = (p_{11}^L, p_{12}^L, \dots, p_{1T}^L, \dots, p_{n1}^L, p_{n2}^L, \dots, p_{nT}^L)'$ . Clearly,  $P$  is a  $nT \times L$  matrix.

For any scalar or vector function  $g(z)$ , denote  $E_{\mathcal{M}}(g(z))$  the projection onto the additive function space  $\mathcal{M}$  (under the  $L_2$  norm). That is,  $E_{\mathcal{M}}(g(z))$  is an element that belongs to  $\mathcal{M}$  and it is the closest function to  $g(z)$  in the  $L_2$  norm for all the functions in  $L_2$  in  $\mathcal{M}$ . Define  $\theta(z) = E(X|Z = z)$  and  $h(z) = E_{\mathcal{M}}(\theta(z))$ .

Let  $\bar{P} = P(P'P)^{-}P'$ , where  $(\cdot)^{-}$  denotes any symmetric generalized inverse. Let  $\tilde{A} = \bar{P}A = P\beta_A$ , where  $\beta_A = (P'P)^{-}P'A$ . Premultiplying (4.14) by  $\bar{P}$  yields

$$\tilde{Y} = \tilde{X}\beta_0 + \tilde{M} + \tilde{V}. \quad (4.16)$$

Subtracting (4.16) from (4.14) by  $\bar{P}$  leads to

$$Y - \tilde{Y} = (X - \tilde{X})\beta_0 + (M - \tilde{M}) + (V - \tilde{V}). \quad (4.17)$$

We estimate  $\beta_0$  by the least squares regression of  $Y - \tilde{Y}$  on  $(X - \tilde{X})$ :

$$\hat{\beta} = \left[ (X - \tilde{X})'(X - \tilde{X}) \right]^{-1} (X - \tilde{X})'(Y - \tilde{Y}). \quad (4.18)$$

Upon obtaining  $\hat{\beta}$ , we can estimate  $m(z)$  by

$$\hat{m}(z) = p^L(z)' \hat{\gamma} \quad (4.19)$$

where  $\hat{\gamma} = (P'P)^{-1} P'(Y - X\hat{\beta})$ .

Let  $\epsilon_{it} = X_{it} - h(z_{it})$ , where  $h(z_{it}) = E_{\mathcal{M}}(\theta(z_{it}))$ . Let  $\Phi = T^{-1} \sum_{t=1}^T E(\epsilon_{it}\epsilon'_{it})$  and  $\Psi = T^{-1} \sum_{t=1}^T E(\sigma^2(X_{it}, Z_{it})\epsilon_{it}\epsilon'_{it})$  where  $\sigma^2(X_{it}, Z_{it}) = E[V_{it}^2 | X_{it}, Z_{it}]$ . Baltagi and Li (2002) prove the following asymptotic normality of  $\hat{\beta}$ :

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(\Phi^{-1}\Psi\Phi^{-1})$$

Baltagi and Li (2002) also establish the consistency rate of  $\hat{m}(z)$  but not the asymptotic normality.

If  $x_{it}$  contains the lagged dependent variable, then the above estimation procedure has to be modified. For example, consider the following partially linear dynamic panel data model

$$y_{it} = \beta_{0,1}y_{i,t-1} + \beta'_{0,2}x_{it}^{(2)} + m(z_{it}) + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (4.20)$$

where  $x_{it}^{(2)}$  is  $x_{it}$  excluding its first element  $y_{i,t-1}$ . Assume the existence of an IV  $w_{it} \in \mathbb{R}^l$  with  $l \geq p$  such that

$$E(U_{it}|w_{it}, z_{it}) = 0, \text{ and } \text{Cov}(w_{it}, X_{it}) \neq 0. \quad (4.21)$$

We can estimate  $\beta_0 = (\beta_{0,1}, \beta'_{0,2})'$  by the IV method for the case  $l = p$ :

$$\hat{\beta}_{IV} = \left[ (W - \tilde{W})'(X - \tilde{X}) \right]^{-1} (W - \tilde{W})'(Y - \tilde{Y}), \quad (4.22)$$

and estimate  $m(z)$  by

$$\hat{m}_{IV}(z) = p^L(z)' \hat{\gamma}_{IV} \quad (4.23)$$

where  $\hat{\gamma}_{IV} = (P'P)^{-1} P'(Y - X\hat{\beta}_{IV})$ . The asymptotic normality of  $\hat{\beta}_{IV}$  is established in Baltagi and Li (2002). See also Baltagi and Li (2000). Obviously, in the case where all elements in  $X_{it}$  are exogenous, we can simply set  $w_{it} = X_{it}$ , and the results will be the same as discussed above.

When  $z_{it} = y_{i,t-1}$ , (4.1) becomes the partially linear dynamic model studied by Lee (2008). He establishes the asymptotic normality of a bias-corrected series estimator of  $\beta_0$ .

### 4.3 Extensions

Traditionally, the dependent variable in a partially linear model is a continuous random variable. This may not be the case in applications. Lin and Carroll (2001a, 2001b) consider a generalized partially linear panel data model using generalized estimating equations. Given the covariates  $x_{it}$  and  $z_{it}$ , they assume that the mean  $\mu_{it}$  of the dependent variable  $y_{it}$  satisfies

$$g(\mu_{it}) = x'_{it}\beta_0 + m(z_{it}), \quad (4.24)$$

where  $z_{it}$  may be time-dependent or not, and  $g$  is some known link function. They develop kernel estimating equations for the nonparametric component  $m(\cdot)$  and profile estimating equations for the parametric component  $\beta_0$ .

If the dimension  $q$  of  $z_{it}$  is large, the estimation of the parametric and nonparametric components in (4.1) becomes difficult. In this case, we can consider the following additive partially linear panel data models

$$y_{it} = x'_{it}\beta_0 + m_1(z_{it,1}) + \cdots + m_q(z_{it,q}) + \alpha_i + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (4.25)$$

where  $\beta_0$ ,  $x_{it}$ ,  $\alpha_i$ , and  $u_{it}$  are defined as above,  $m_l(\cdot)$ ,  $l = 1, \dots, q$ , are unknown smooth functions. Obviously the individual functions  $m_l(\cdot)$ ,  $l = 1, \dots, q$ , are not identified without further conditions. In the literature on kernel estimation, one may assume that  $E[m_l(z_{it,l})] = 0$  whereas in the literature on series estimation, it seems convenient to assume that  $g_l(0) = 0$  for  $l = 2, \dots, q$ . Li (2000) consider the series estimation of the above model in the cross section framework. It seems straightforward to extend his method to the panel framework.

### 4.4 Specification tests

Various specification tests can be conducted for partially linear models. These include tests in for correct specification of functional forms, tests for random effects versus fixed effects, tests for individual effects, tests for serial correlation, and tests for heteroskedasticity in the disturbance terms, etc. Despite the importance of specification testing in panel data models, only few papers consider this.

Henderson, Carroll, and Li (2008) considers testing the functional form by considering the following possible specifications

$$y_{it} = x'_{it}\beta_0 + z'_{it}\gamma_0 + \varepsilon_{it}, \quad (4.26)$$

$$y_{it} = x'_{it}\beta_0 + m(z_{it}) + \varepsilon_{it}, \quad (4.27)$$

$$y_{it} = g(x_{it}, z_{it}) + \varepsilon_{it}, \quad (4.28)$$

where the definitions of parameters and functions are self-evident. The three pairs of null and alternative hypotheses are

$$\begin{aligned} H_0^a & : (4.26) \text{ versus } H_1^a : (4.27), \\ H_0^b & : (4.26) \text{ versus } H_1^b : (4.28), \\ H_0^c & : (4.27) \text{ versus } H_1^c : (4.28), \end{aligned}$$

where for example, “ $H_0^a : (4.26)$ ” means the model in (4.26) is the true model under the first null hypothesis  $H_0^a$ . For each case, they estimate the models under the null and alternative and compared the squared distance between the estimated models. For example, to test  $H_0^a : (4.26)$  versus  $H_1^a : (4.27)$ , they estimate both (4.26) and (4.27) and base their test statistic on

$$J_n^a = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left[ x'_{it} \tilde{\beta} + z'_{it} \tilde{\gamma} - x'_{it} \hat{\beta} - \hat{m}(z_{it}) \right]^2$$

where  $\tilde{\beta}$  and  $\tilde{\gamma}$  are estimates of  $\beta_0$  and  $\gamma_0$  under  $H_0^a$ ,  $\hat{\beta}$  and  $\hat{m}(z_{it})$  are estimates of  $\beta_0$  and  $m(z_{it})$  under  $H_1^a$ . Without deriving the asymptotic distribution for such a test statistic, they propose a bootstrap method to obtain the critical values and demonstrate through simulations that the proposed tests work fairly well in finite samples.

Li and Hsiao (1998) consider testing serial correlation in a partially linear panel data models that could allow lagged dependent variables as explanatory variables. They consider the following model

$$y_{it} = x'_{it} \beta_0 + m(z_{it}) + u_{it}. \quad (4.29)$$

where variables are defined as above and  $u_{it}$  satisfies  $E(u_{it} | x_{it}, z_{it}) = 0$  a.s. The null hypothesis is

$$H_0 : u_{it} \text{ is a martingale difference sequence (m.d.s.)}$$

Clearly, under the above null hypothesis,  $u_{it}$  cannot contain the individual effects. Based on the residuals from the above partially linear model, they propose three test statistics that test zero first-order serial correlation, higher-order serial correlations, and individual effects, respectively. These test statistics have either asymptotic normal or chi-square distribution under the null hypothesis of an m.d.s. error process.

## 5 Varying Coefficient Panel Data Models

In this section, we review the literature on varying coefficient models. We consider the following model

$$\begin{aligned} y_{it} &= x'_{it}m(z_{it}) + \alpha_i + u_{it} \\ &= \sum_{d=1}^p x_{it,d}m_d(z_{it}) + \alpha_i + u_{it} \end{aligned} \quad (5.1)$$

where the covariate  $z_{it}$  is a  $q \times 1$  vector,  $x_{it} = (x_{it,1}, \dots, x_{it,p})'$ ,  $m(\cdot) = (m_1(\cdot), \dots, m_p(\cdot))'$  has  $p$  unknown smooth functions, and  $u_{it}$  is i.i.d. with zero mean and finite variance  $\sigma_u^2$ . We make explicit assumptions on the dependence of  $\alpha_i$  and  $u_{it}$  on the covariates  $x_{it}$  and  $z_{it}$  only when needed. The model in (5.1) is useful where the response parameter (slope coefficient) depends on the variable  $z_{it}$ . For example, in a wage equation,  $y_{it}$  denotes the logarithm of wage,  $x_{it}$  denotes the years of schooling (education), and the rate of return to education may depend on the individual characteristic  $z_{it}$ . In a special case where  $p = 1$ ,  $x_{it} = 1$  for all  $i$  and  $t$ , and  $\alpha_i$  can be correlated with  $x_{it}$  and  $u_{it}$ , model (5.1) reduces to the conventional fixed effects panel data models considered by Su and Ullah (2006) and Henderson, Carroll and Li (2008).

Note that the model in (5.1) includes the partially linear model as special cases:  $x'_{it}m(z_{it}) = m_1(z_{it}) + \tilde{x}'_{it}\beta_0$  where  $x_{it} = (1, \tilde{x}'_{it})'$ , and  $m(z_{it}) = (m_1(z_{it}), \beta'_0)'$  for some real-valued function  $m_1$  and  $(p-1) \times 1$  vector  $\beta_0$ . The latter model was considered by Li and Hsiao (1998) and Kniesner and Li (2002) who assumes that  $E(u_{it}|z_{it}, x_{it}) = 0$ . Li and Stengos (1996) and Baltagi and Li (2002) considered the same model but allowed  $E(u_{it}|x_{it}) \neq 0$ . See Section 4.

### 5.1 Profile least squares method

We first consider the estimation of  $m$  in (5.1) when  $\alpha_i$  is treated as fixed effects which can be correlated with either  $x_{it}$  or  $u_{it}$ .

For any given  $z$  and  $d \in \{1, 2, \dots, p\}$ , it follows from a first order Taylor expansion that

$$\begin{aligned} m_d(z_{it}) &\approx m_d(z) + (z_{it} - z)' \beta_d(z) \\ &= z_{it}(z)' \delta_d(z) \end{aligned}$$

where  $z_{it}(z) = (1 \ (z_{it} - z)')'$ ,  $\delta_d(z) = (m_d(z) \ \beta_d(z)')'$ , and  $\beta_d(z) = \partial m_d(z) / \partial z$ . Then following the LLS estimation procedure in Section 2, we can write the estimate of  $\delta(z) = (m_1(z), \dots, m_p(z), \beta_1(z)', \dots, \beta_p(z)')'$  as

$$\delta_\alpha(z) = \min_{\delta} (Y^* - \tilde{X}\delta)' \mathbf{K}(z) (Y^* - \tilde{X}\delta)$$

where  $Y^* = Y - D\alpha$ ,  $\tilde{X} = (X_{11}, \dots, X_{1T}, \dots, X_{n1}, \dots, X_{nT})'$  is an  $nT \times p(q+1)$  matrix with  $X_{it} = X_{it}(z) = (X'_{it}, X'_{it} \otimes (z_{it} - z)')$ , and  $\mathbf{K}(z) = \text{diag}(K_h(z_{11} - z), \dots, K_h(z_{1T} - z), \dots, K_h(z_{n1} - z), \dots, K_h(z_{nT} - z))$  is an  $nT \times nT$  diagonal matrix. So, given  $\alpha$ , the LLLS estimator of  $\delta(z)$  is simply

$$\hat{\delta}_\alpha(z) = [\tilde{X}'\mathbf{K}(z)\tilde{X}]^{-1} \tilde{X}'\mathbf{K}(z)Y^*.$$

In particular, the estimator of  $m(z) = (m_1(z), \dots, m_p(z))'$  is given by

$$\hat{m}_\alpha(z) = \mathbf{e}\hat{\delta}_\alpha(z) = \mathbf{s}(z)(Y - D\alpha)$$

where  $\mathbf{e} = (I_p, \mathbf{0}_{p \times pq})$  is a  $p \times p(q+1)$  selection matrix with  $\mathbf{0}_{p \times pq}$  denoting a  $p \times pq$  matrix of zeros, and  $\mathbf{s}(z) = \mathbf{e}\mathbf{S}(z) = \mathbf{e}(\tilde{X}'\mathbf{K}(z)\tilde{X})^{-1}\tilde{X}'\mathbf{K}(z)$  is a  $p \times nT$  matrix.

Let  $Z = (z_{11}, \dots, z_{1T}, \dots, z_{n1}, \dots, z_{nT})'$ . We can write  $m(x_{it}, z_{it}) \equiv x'_{it}m(z_{it})$ ,  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ , in vector form as

$$\mathbf{m}(X, Z) = \sum_{d=1}^p \mathbf{x}_d \odot \mathbf{m}_d(\mathbf{z})$$

where  $\mathbf{x}_d = (x_{11,d}, \dots, x_{1T,d}, \dots, x_{n1,d}, \dots, x_{nT,d})'$ ,  $\mathbf{m}_d(Z) = (m_d(z_{11}), \dots, m_d(z_{1T}), \dots, m_d(z_{n1}), \dots, m_d(z_{nT}))'$  for  $d = 1, \dots, p$ , and  $\odot$  is the Hadamard product. Thus

$$\hat{\mathbf{m}}_\alpha(X, Z) = \sum_{d=1}^p \mathbf{x}_d \odot \hat{\mathbf{m}}_{\alpha,d}(Z) = \sum_{d=1}^p \mathbf{x}_d \odot (\mathbf{S}_d(z)(Y - D\alpha))$$

where  $\hat{\mathbf{m}}_{\alpha,d}(z) = (\hat{m}_{\alpha,d}(z_{11}), \dots, \hat{m}_{\alpha,d}(z_{1T}), \dots, \hat{m}_{\alpha,d}(z_{n1}), \dots, \hat{m}_{\alpha,d}(z_{nT}))'$ ,  $\hat{m}_{\alpha,d}(z)$  is the  $d$ th element of  $\hat{m}_\alpha(z) : \hat{m}_{\alpha,d}(z) = e'_d \hat{m}_\alpha(z) = e'_d \mathbf{s}(z)(Y - D\alpha)$  with  $e_d$  being a  $p \times 1$  vector with 1 in the  $d$ th element and 0 elsewhere, and  $\mathbf{S}_d(Z)' = (\mathbf{s}(z_{11})'e_d, \dots, \mathbf{s}(z_{1T})'e_d, \dots, \mathbf{s}(z_{n1})', \dots, \mathbf{s}(z_{nT})'e_d)$ .

Noting that

$$\hat{\mathbf{m}}_\alpha(X, Z) = \left( \sum_{d=1}^p (\mathbf{x}_d \otimes l'_{nT}) \odot \mathbf{S}_d(Z) \right) (Y - D\alpha),$$

the estimate of  $\alpha$  is given by

$$\hat{\alpha} = (D'Q_1D)^{-1}D'Q_1Y$$

where  $Q_1 = (I_{nT} - \sum_{d=1}^p (\mathbf{x}_d \otimes l'_{nT}) \odot \mathbf{S}_d(Z))'(I_{nT} - \sum_{d=1}^p (\mathbf{x}_d \otimes l'_{nT}) \odot \mathbf{S}_d(Z))$ . Further, the estimator for  $\delta(z)$  and  $m(z)$  follows by  $\hat{\delta}_{\hat{\alpha}}(z)$  and  $\hat{m}_{\hat{\alpha}}(z)$ , respectively.

Sun, Carroll, and Li (2009) suggest an alternative profile least squares estimator for the above model by profiling out the nonparametric component  $m$ . They also propose a test for testing a random effects model against a fixed effects alternative model. Notice that if the vector  $z_{it}$  contains both the discrete and continuous variables, then Su, Chen, Ullah (2009) can be extended to this panel framework.



## 5.2 Differencing method

As in Section 3, we can consider subtracting the model (5.1) for time  $t$  from that of time  $t - 1$  so that

$$\Delta y_{it} = x'_{it}m(z_{it}) - X'_{i,t-1}m(z_{i,t-1}) + \Delta u_{it} \quad (5.2)$$

or subtracting the equation from time  $t$  from that for time 1 so that

$$y_{it} - y_{i1} = x'_{it}m(z_{it}) - x'_{i1}m(z_{i1}) + u_{it} - u_{i1}. \quad (5.3)$$

Alternatively, the within-group differencing method yields

$$y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it} = x'_{it}m(z_{it}) - \frac{1}{T} \sum_{t=1}^T x'_{it}m(z_{it}) + u_{it} - \frac{1}{T} \sum_{t=1}^T u_{it} \quad (5.4)$$

or

$$y_{it}^* = \sum_{s=1}^T d_{ts}x'_{is}m(x_{is}) + u_{it}^* \quad (5.5)$$

where  $d_{ts} = -\frac{1}{T}$  if  $s \neq t$  and  $1 - \frac{1}{T}$  otherwise, and  $\sum_{s=1}^T d_{ts} = 0$  for all  $t$ ,  $y_{it}^* = y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it}$ , and  $u_{it}^* = u_{it} - \frac{1}{T} \sum_{t=1}^T u_{it}$ .

For each  $i$ , the right hand side of (5.2)-(5.4) contains linear combination of  $X'_{it}m(z_{it})$  for different  $t$ . If there is an intercept term in  $x_{it}$  and  $m_1(z_{it})$  is the first element of  $m(z_{it})$ , then the difference of the first element of  $x'_{it}m(z_{it}) = \sum_{d=1}^p x_{it,d}m_d(z_{it})$  gives  $m_1(z_{it}) - m_1(z_{i,t-1})$ . This is an additive function with the same functional form (strong assumption) at different times. The kernel estimation requires some backfitting algorithms or marginal integration method to recover the unknown function, which causes computation burden as well as complications in asymptotic analyses.

For this reason, Sun, Carroll, and Li (2009) focus on the profile estimation of  $m$ . But we believe that the asymptotic analyses based on differencing methods in the conventional fixed effects panel data models can be extended to this model.

## 5.3 Nonparametric GMM estimation

In the above model  $E(u_{it}|z_{it}) = 0$  and  $E(u_{it}|x_{it}) = 0$ . However, in various economic models  $E(u_{it}|x_{it}) \neq 0$ , for example, when  $x_{it}$  is correlated with  $u_{it}$  (endogeneity),  $x_{it}$  has measurement errors, and  $x_{it}$  has lagged dependent variable. The result for the case of  $E(u_{it}|z_{it}) \neq 0$  has not been developed yet.

The IV estimation of the general model  $m(x_{it}, z_{it}) = x'_{it}m(z_{it})$  has been considered by Das (2005), Cai et al. (2006), and Cai and Xiong (2006) for discrete and continuous variables in

the cross-sectional setup. In addition, with no endogeneity, this model is covered in González, Teräsvirta, and van Dijk (2005) and the threshold non-dynamic model in Hansen (1999). Here we present the nonparametric GMM estimation of Cai and Li (2008) for this model.

Cai and Li (2008) consider the model

$$y_{it} = x'_{it}m(z_{it}) + \varepsilon_{it} \quad (5.6)$$

where  $\varepsilon_{it}$  plays the role of  $\alpha_i + u_{it}$  in (5.1),  $E(\varepsilon_{it}|z_{it}) = 0$ , and  $E(\varepsilon_{it}|x_{it}) \neq 0$ . Note that  $E(y_{it}|x_{it}, z_{it}) \neq x'_{it}m(z_{it})$  because  $E(\varepsilon_{it}|x_{it}) \neq 0$ . Let  $w_{it}$  be the  $l \times 1$  instrument variables such that

$$E(\varepsilon_{it}|v_{it}) = 0, \quad (5.7)$$

where  $v_{it} = (w'_{it}, z'_{it})'$ . Multiplying both sides of (5.7) by  $\pi(v_{it}) = E(x_{it}|v_{it})$  and taking expectations, conditional on  $z_{it} = z$ , we have

$$E[\pi(v_{it})y_{it}|z_{it}] = E[\pi(v_{it})x'_{it}|z_{it} = z]m(z) = E[\pi(v_{it})\pi(v_{it})'|z_{it} = z]m(z).$$

This gives  $m(z) = \{E[\pi(v_{it})\pi(v_{it})'|z_{it} = z]\}^{-1} E[\pi(v_{it})y_{it}|z_{it} = z]$  under the assumption of positive definiteness of  $E[\pi(v_{it})\pi(v_{it})'|z_{it} = z]$ . This assumption guarantees that  $m(\cdot)$  is identified locally. To obtain the estimator of  $m(\cdot)$ , one can consider a two stage nonparametric procedure. At the first stage  $\hat{\pi}(v_{it})$  is obtained by a nonparametric estimation of  $x_{it}$  on  $v_{it}$ . Then at the second stage, one estimate  $m(\cdot)$  based on the varying coefficient model:  $y_{it} \approx \hat{\pi}(v_{it})'m(z_{it}) + \varepsilon_{it}$ . The asymptotic property of such a two-stage nonparametric estimator is however quite complicated.

In viewing this, Cai and Li (2008) propose a one step nonparametric GMM (NPGMM) estimation of  $m(z)$ . According to this, an  $m_1 \times 1$  vector function  $g(v_{it})$  is chosen such that

$$E[g(v_{it})\varepsilon_{it}|v_{it}] = E[g(v_{it})\{y_{it} - x'_{it}m(z_{it})\}|v_{it}] = 0. \quad (5.8)$$

Let us write the sample GMM orthogonality conditions based on the local linear approximation of  $m(z_{it})$  in a neighborhood of  $z$  as

$$\sum_{i=1}^n \sum_{t=1}^T g(v_{it})(y_{it} - V'_{it}\delta)K_h(z_{it} - z) = 0, \quad (5.9)$$

where

$$V_{it} = \begin{pmatrix} x_{it} \\ x_{it} \otimes (z_{it} - z) \end{pmatrix}$$

is an  $m_2 \times 1$  vector with  $m_2 = p(q+1)$ ,  $\delta = \delta(z)$  is an  $m_2 \times 1$  vector of parameters whose true value corresponds to  $(m(z)', \partial m_1(z)/\partial z', \dots, \partial m_p(z)/\partial z')'$ . When  $m_1 \geq m_2$ , the solution to  $\delta$  is given by

$$\hat{\delta}(z) = (P'P)^{-1} P'Q$$

where

$$P = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T g(v_{it}) V_{it}' K_h(z_{it} - z), \text{ and } Q = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T g(v_{it}) K_h(z_{it} - z) y_{it}.$$

This gives the NPGMM estimate of  $m(z)$  and its first-order derivatives  $\partial m_j(z)/\partial z$  for  $j = 1, \dots, p$ . This is one stage estimator which is simpler compared to the two-stage NP estimator described above and studied in Cai et al. (2006) in the cross-sectional setup. Note that the two-stage estimation involves a NP regression of higher dimensions, and requires two smoothing parameters compared to one step NP estimation which only needs one smoothing component. If the dimension of  $w_{it}$  is higher than that of  $z_{it}$ , one expects that the one-step estimator has much better finite sample performance than that for the two-stage estimator. When there is no endogenous variables ( $w_{it} = x_{it}$ ) then one can choose  $g(v_{it}) = V_{it}$ . In this case the GMM conditions become

$$\sum_{i=1}^n \sum_{t=1}^T V_{it} K_h(z_{it} - z) (y_{it} - V_{it}' \delta) = 0,$$

which is the normal equation of the following LLS problem of the varying coefficient model:

$$\min_{\delta} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - V_{it}' \delta)^2 K_h(z_{it} - z)$$

and it gives the ordinary LLS estimator studied above.

For the choice of  $g(v_{it})$  one solution is to consider the  $p(q+1) \times 1$  vector

$$g(v_{it}) = \begin{pmatrix} w_{it} \\ w_{it} \otimes (z_{it} - z)/h \end{pmatrix}$$

In this case  $m_1 = l(q+1) \geq m_2$  implies  $l \geq p$ . Although it is simple, it may not be optimal. The optimality could be developed by using results analogous to those in Newey (1999) and Ai and Chen (2003).

Under the usual assumptions such as that  $h \rightarrow 0$  and  $nh^q \rightarrow 0$  as  $n \rightarrow \infty$ ,  $K$  is a symmetric, nonnegative, and bounded second-order kernel, and that  $E[g(v_{it})g(v_{it})'|z_{it} = z]$  is positive definite, Cai and Li (2008) showed that, for fixed  $T$ ,

$$\sqrt{nTh^q} \left[ H(\hat{\delta} - \delta) - \frac{h^2}{2} \begin{pmatrix} B_m(z) \\ \mathbf{0}_{pq \times 1} \end{pmatrix} + o_p(h^2) \right] \xrightarrow{d} N \left( \mathbf{0}_{p(q+1) \times 1}, \frac{\Delta}{f(z)} \right),$$

where  $H = \text{diag}(I_p, hI_{qp})$  is an  $m_2 \times m_2$  matrix,  $B_m(z) = \int A(u, z)K(u)du$  is a  $p \times 1$  vector,  $A(u, z) = (u'm_1^{(2)}(z)u, \dots, u'm_p^{(2)}(z)u)'$ ,  $m_j^{(2)}(z) = d^2m_j(z)/dzdz'$ ,  $\Omega = \Omega(z) = E(w_{it}x'_{it}|z_{it} = z)$ , and  $\Delta = \text{diag}\{v_0\Omega_m, \Omega_m \otimes [\mu_2^{-1}(K) \mu_2(K^2)\mu_2^{-1}(K)]\}$  with  $\Omega_m = (\Omega'\Omega)^{-1} \Omega'\Omega_1\Omega(\Omega'\Omega)^{-1}$ ,  $\Omega_1 = \Omega_1(z) = \text{Var}(w_{it}\varepsilon_{it}|z_{it} = z)$ ,  $\mu_2(K) = \int uu'K(u)du$ , and  $v_0 = \int K^2(u)du$ . If  $T \rightarrow \infty$ , then

$$\sqrt{nTh^q} \left[ \hat{m}(z) - m(z) - \frac{h^2}{2}B_m(z) + o_p(h^2) \right] \xrightarrow{d} N \left( \mathbf{0}_{p \times 1}, \frac{v_0\Omega_m}{f(z)} \right),$$

where  $\hat{m}(z)$  as the first  $p$  elements  $\hat{\delta}(z)$  is the estimator of  $m(z)$ . For details in proofs and the assumptions, see Cai and Li (2008). It is clear from the above results that  $\hat{m}(z)$  has the same leading bias and variance for both finite and large  $T$  cases. Therefore the asymptotic *MSE* ( $T$  is fixed or large) is the same and the optimal  $h$  is proportional to  $(nT)^{-1/(p+4)}$ . However, when  $T$  is large and  $n$  is small, some modification in the results may be needed.

Finally, when  $w_{it} = x_{it}$ , we have

$$\sqrt{nTh^q} \left[ \hat{m}(z) - m(z) - \frac{h^2}{2}B_m(z) + O_p(h^2) \right] \xrightarrow{d} N \left( \mathbf{0}_{p \times 1}, v_0f^{-1}(z)\Omega_m^*(z) \right)$$

where  $\Omega_m^*(z) = [E(x_{it}x'_{it}|z_{it} = z)]^{-1} E[\sigma^2(v_{it})x_{it}x'_{it}|z_{it} = z] [E(x_{it}x'_{it}|z_{it} = z)]^{-1}$  and  $\sigma^2(v_{it}) = \text{Var}(\varepsilon_{it}|v_{it})$ .

The efficiency property of Cai et al.'s (2006) two-stage estimator compared to the single stage estimator is not fully known. For special cases of asymptotic efficiency, see Cai and Li (2008, p.1333).

The above estimation procedure is valid when  $\varepsilon_{it}$  is serially correlated and/or  $x_{it}$  contains lagged dependent variables. But as remarked earlier, it is unclear how to estimate the model if  $z_{it}$  contains the lagged dependent variable. We conjecture that it may be easier to establish the asymptotic theory for estimators based on series method rather than the kernel method.

In a recent paper Tran and Tsionas (2010) considered a two step NP GMM estimation with a general weighting matrix, and where  $n$  is large but  $T$  is fixed. They claim that their two step estimation may lead to potential gain in asymptotic efficiency. They also analyze the finite sample efficiency of their estimator and provide an empirical application.

In addition, Cai and Xiong (2006) have considered the following varying coefficient IV model

$$\begin{aligned} Y &= m(x, z_1) + u \\ &= m_1(z_{11})'z_{12} + m_2(z_{11})'x_1 + \beta_1'z_{13} + \beta_2'x_2 + u \end{aligned}$$

where  $x = (x'_1, x'_2)'$  is a vector of endogenous variables,  $z_1 = (z'_{11}, z'_{12}, z'_{13})'$  is a vector of exogenous variables,  $z = (z'_1, z'_2)$  with  $z_2$  being a vector of IVs, and  $E(u|z) = 0$ . If there is

no endogenous variable, this model becomes the partially varying coefficient model studied by Ahmad, Leelahanon, and Li (2005) and that of the model in Cai et al. (2006) if the parametric part is absent. And if  $x$  is a discrete endogenous variable, then the model is as studied by Das (2005), as a special case. The estimation of the above model and its asymptotic properties are developed in Cai and Xiong (2006) in the cross-sectional setup, which can be potentially extended to the panel data framework.

#### 5.4 Testing random effects versus fixed effects

Based on their profile least squares estimates, Sun, Carroll, and Li (2009) propose a test of random effects against fixed effects in model (5.1). The null hypothesis is

$$H_0 : E(\alpha_i | x_{i1}, \dots, x_{iT}, z_{i1}, \dots, z_{iT}) = 0 \text{ a.s.}$$

and the alternative hypothesis  $H_1$  is the negation of  $H_0$ . Their test statistic is based on the weighted squared difference between the random effects and fixed effects estimators, where the weights are used to get around the random denominator issue in the kernel literature. They show that their test statistic is asymptotically normally distributed under the null and diverges to infinity under the fixed alternative.

## 6 Nonparametric Panel Data Models with Cross Section Dependence

In this section, we consider a semiparametric panel data model with cross section dependence:

$$y_{it} = m_i(x_{it}) + \gamma'_{1i} f_{1t} + e_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T \quad (6.1)$$

where  $x_{it}$  is a  $p \times 1$  vector of observed individual-specific regressors,  $m_i(\cdot)$  is an unknown smooth function from,  $f_{1t}$  is a  $q_1 \times 1$  vector of observed common factors, and  $\gamma_{1i}$ ,  $i = 1, \dots, n$ , are factor loadings. Here we assume that  $f_{1t}$  includes the intercept term and impose the condition  $E[m_i(x_{it})] = 0$  in order to identify  $m_i(\cdot)$ . The error term  $e_{it}$  in (6.1) follows the multi-factor structure

$$e_{it} = \gamma'_{2i} f_{2t} + \varepsilon_{it}, \quad (6.2)$$

where  $f_{2t}$  is a  $q_2 \times 1$  vector of unobserved common factors,  $\varepsilon_{it}$  is the idiosyncratic error assumed to be independently distributed of  $(x_{it}, f_{1t}, f_{2t})$ , and  $\gamma_{2i}$ ,  $i = 1, \dots, n$ , are factor loadings. We are interested in the estimation of  $g_i(\cdot)$  in the presence of multi-factor error structure.

Like Pesaran (2006), the unobserved factor  $f_{2t}$  could be correlated with  $(x_{it}, f_{1t})$ . To allow for such a possibility, we adopt the following fairly general model for the individual-specific regressors,

$$x_{it} = \Gamma'_{1i}f_{1t} + \Gamma'_{2i}f_{2t} + v_{it}, \quad (6.3)$$

where  $\Gamma_{1i}$  and  $\Gamma_{2i}$  are  $q_1 \times p$  and  $q_2 \times p$  factor loading matrices, and  $v_{it}$  is a  $p \times 1$  vector of individual-specific components of  $x_{it}$ .

The model specified in (6.1)-(6.3) is fairly general and includes a variety of panel data models as special cases. First, Pesaran's (2006) model corresponds to the case where  $m_i(x) = \beta'_i x$  for some  $d \times 1$  vector  $\beta_i$  so that model (6.1) becomes  $y_{it} = \beta'_i x_{it} + \gamma'_{1i} f_{1t} + e_{it}$ . Second, it includes the conventional fixed or random effects models and the models of Bai (2009) in particular. Third, it includes the usual nonparametric panel data model  $y_{it} = m(x_{it}) + \alpha_i + v_t + \varepsilon_{it}$ , where the individual effects  $\alpha_i$  and the time effects  $v_t$  enter the model additively. Huang (2006) studies the kernel estimation of (6.1) when the unobserved factor  $f_{2t}$  in (6.2) is a scalar random variable.

## 6.1 Common correlated effect (CCE) estimator

Let  $\bar{x}_t \equiv n^{-1} \sum_{i=1}^n x_{it}$  and  $\bar{y}_t \equiv n^{-1} \sum_{i=1}^n y_{it}$ . Then (6.1)-(6.3) implies that

$$\begin{pmatrix} \bar{x}_t \\ \bar{y}_t \end{pmatrix} = \begin{pmatrix} \bar{\Gamma}'_1 \\ \bar{\gamma}'_1 \end{pmatrix} f_{1t} + \begin{pmatrix} \bar{\Gamma}'_2 \\ \bar{\gamma}'_2 \end{pmatrix} f_{2t} + \begin{pmatrix} \bar{v}_t \\ \bar{m}_t + \bar{\varepsilon}_t \end{pmatrix}, \quad (6.4)$$

where  $\bar{\Gamma}_1, \bar{\Gamma}_2, \bar{\gamma}_1, \bar{\gamma}_2, \bar{v}_t$ , and  $\bar{\varepsilon}_t$  are sample averages of  $\Gamma_{1i}, \Gamma_{2i}, \gamma_{1i}, \gamma_{2i}, v_{it}$ , and  $\varepsilon_{it}$  over  $i$ , respectively, and  $\bar{m}_t = n^{-1} \sum_{i=1}^n m_i(x_{it})$ . Let  $\bar{\Gamma}_2^* \equiv (\bar{\Gamma}_2, \bar{\gamma}_2)$ . Premultiplying both sides of (6.4) by  $\bar{\Gamma}_2^*$  and solving for  $f_{2t}$  yields

$$f_{2t} = \left( \bar{\Gamma}_2^* \bar{\Gamma}_2^{*'} \right)^{-1} \bar{\Gamma}_2 \left( \begin{pmatrix} \bar{x}_t \\ \bar{y}_t \end{pmatrix} - \begin{pmatrix} \bar{\Gamma}'_1 \\ \bar{\gamma}'_1 \end{pmatrix} f_{1t} - \begin{pmatrix} \bar{v}_t \\ \bar{g}_t + \bar{\varepsilon}_t \end{pmatrix} \right) \quad (6.5)$$

provided that

$$\text{rank}(\bar{\Gamma}_2^*) = q_2 \leq p + 1 \text{ for sufficiently large } n. \quad (6.6)$$

As  $n \rightarrow \infty$ ,  $\bar{v}_t \xrightarrow{p} 0$ ,  $\bar{\varepsilon}_t \xrightarrow{p} 0$  and  $\bar{m}_t \xrightarrow{p} 0$  for each  $t$  under weak conditions. It follows

$$f_{2t} - \left( \bar{\Gamma}_2^* \bar{\Gamma}_2^{*'} \right)^{-1} \bar{\Gamma}_2 \left( \begin{pmatrix} \bar{x}_t \\ \bar{y}_t \end{pmatrix} - \begin{pmatrix} \bar{\Gamma}'_1 \\ \bar{\gamma}'_1 \end{pmatrix} f_{1t} \right) \xrightarrow{p} 0 \text{ as } n \rightarrow \infty. \quad (6.7)$$

The last line suggests that we can use  $h_t \equiv (f'_{1t}, \bar{x}'_t, \bar{y}_t)'$  as observable proxies for  $f_{2t}$  and consider the following semiparametric regression:

$$y_{it} \approx m_i(x_{it}) + \vartheta'_i h_t + e_{it}. \quad (6.8)$$

Clearly, (6.8) is an additive semiparametric model and series method has its advantage over the kernel method. For this reason, Su and Jin (2010) propose to estimate  $m_i(\cdot)$  by sieve method.

To proceed, let  $\{p_l(x), l = 1, 2, \dots\}$  denote a sequence of known basis functions that can approximate any square-integrable function of  $x$  very well. Let  $L \equiv L(T)$  be some integer such that  $L \rightarrow \infty$  as  $T \rightarrow \infty$ . Let  $p^L(x) = (p_1(x), p_2(x), \dots, p_L(x))'$ ,  $p_{it} = p^L(x_{it})$ , and  $p_i = (p_{i1}, p_{i2}, \dots, p_{iT})'$ . Under fairly weak conditions, we can approximate  $m_i(x)$  in (6.8) very well by  $\alpha'_{m_i} p^L(x)$  for some  $L \times 1$  vector  $\alpha_{m_i}$ .

To estimate  $\alpha_{m_i}$ , we run the regression of  $y_{it}$  on  $p^L(x_{it})$  and  $h_t \equiv (f'_{1t}, \bar{x}'_t, \bar{y}_t)'$

$$y_{it} = \alpha'_{m_i} p^L(x_{it}) + \vartheta'_i h_t + u_{it} \quad (6.9)$$

where  $u_{it}$  is the new error term. Let  $y_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$ ,  $h = (h_1, h_2, \dots, h_T)'$ , and  $u_i = (u_{i1}, u_{i2}, \dots, u_{iT})'$ . We can rewrite (6.9) in vector form

$$y_i = p_i \alpha_{m_i} + h \vartheta_i + u_i \quad (6.10)$$

By the formula for partitioned regression, the estimator of  $\alpha_{m_i}$  in (6.9) or (6.10) is given by

$$\hat{\alpha}_{m_i} = (p'_i b_h p_i)^{-} p'_i b_h y_i, \quad (6.11)$$

where  $b_h \equiv I_T - h(h'h)^{-} h$ , and  $(\cdot)^{-}$  denotes any symmetric generalized inverse. The estimator of  $m_i(x)$  is then given by

$$\hat{m}_i(x) = p^L(x)' \hat{\alpha}_{m_i}. \quad (6.12)$$

Su and Jin (2010) establish the consistency and asymptotic normality of  $\hat{m}_i(x)$  by passing  $T \rightarrow \infty$ .

## 6.2 Estimating the homogenous relationship

In practice, one may also be interested in estimating a restricted submodel of (6.1):

$$y_{it} = m(x_{it}) + \gamma'_{1i} f_{1t} + e_{it}. \quad (6.13)$$

That is,  $m_i(x) = m(x)$  for all  $i$  in model (6.1). In the case where  $\gamma_{1i} = 0$ , (6.13) can be regarded as a nonparametric extension of Bai's (2009) linear panel data model with multi-factor error structure or a simple extension of Huang's (2006) nonparametric panel data from his single-factor error structure to multiple-factor error structure.

If model (6.13) is assumed to be correctly specified in conjunction with (6.2)- (6.3), we can estimate  $m(\cdot)$  by

$$\hat{m}(x) = p^L(x)' \tilde{\alpha}_m. \quad (6.14)$$

where  $\tilde{\alpha}_m = (\sum_{i=1}^n p_i' b_h p_i)^{-1} \sum_{i=1}^n p_i' b_h y_i$  and  $L$  is now allowed to depend on both  $n$  and  $T$ . The asymptotic normality of  $\hat{m}(x)$  is also studied in Su and Jin (2010) by passing both  $n$  and  $T$  to infinity.

Clearly, besides the multi-factor error structure, the key assumption that underlines the asymptotic analysis of Su and Jin (2010) is (6.3) that specifies the relationship between the individual-specific regressor and the factors. The violation of such an assumption may invalidate their analysis. Therefore it is desirable to propose an alternative estimator without imposing such an assumption. By combining the series method with the principal component analysis, Su and Zhang (2010a) consider the estimation of homogenous relationship ( $m$ ) in a simpler model

$$y_{it} = m(x_{it}) + \gamma_i' f_t + \varepsilon_{it}$$

where  $f_t$  is a  $q \times 1$  vector of unobservable factors and  $\gamma_i$ 's are factor loadings,  $m$ ,  $\varepsilon_{it}$ , and  $x_{it}$  are as defined above. If  $m(x_{it}) = x_{it}' \beta_0$  for a  $p \times 1$  vector  $\beta_0$ , the model reduces to that of Bai (2009).

### 6.3 Specification tests

Various specification tests can be conducted for the model in (6.1). This includes tests for homogenous relationship ( $m_i = m$  for all  $i$ ) and tests for cross section independence or uncorrelatedness.

Jin and Su (2010) propose a nonparametric test for poolability in (6.1). The null hypothesis is

$$H_0 : m_i(x) = m_j(x) \text{ a.e. on the joint support of } m_i \text{ and } m_j \text{ and for all } i, j = 1, \dots, n, \quad (6.15)$$

where a.e. is the abbreviation for almost everywhere. They propose a test statistic based on series estimation and the measure

$$\Gamma = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \int (m_i(x) - m_j(x))^2 w(x) dx, \quad (6.16)$$

where  $w(x)$  is a nonnegative weight function, and establish the asymptotic normality of their test under the null and a sequence of local alternatives. This extends and complements the work of Baltagi, Hidalgo, and Li (1996) who propose a kernel-based test for poolability in conventional panel data models.

Chen, Gao, and Li (2009) propose a kernel-based test for cross section *uncorrelatedness* in

$$y_{it} = m_i(x_{it}) + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (6.17)$$



where the error term satisfies  $E(u_{it}) = 0$ . They test whether  $E(u_{it}u_{jt}) = 0$  for all  $t \geq 1$  and  $i \neq j$  by allowing both  $n$  and  $T$  to pass to the infinity. If  $n$  is fixed, then their test complements the test of conditional uncorrelatedness in Su and Ullah (2009). Su and Zhang (2000b) propose a test of cross section *independence* for the model in (6.17). It is based on the comparison of the joint densities and the product of marginal densities and thus has extra power in detecting deviations from cross section independence when compared with the test of Chen, Gao, and Li (2009).

## 7 Nonseparable Nonparametric Panel Data Models

In this section we review papers on nonseparable nonparametric panel data models. We focus on two types of models. The first type is the partially separable nonparametric panel data model

$$y_{it} = m(x_{it}, \alpha_i) + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (7.1)$$

where  $x_{it}$  is a  $p \times 1$  vector of explanatory variables, the scalar  $\alpha_i$  is a parameter that represents unobserved individual heterogeneity,  $u_{it}$  is a scalar idiosyncratic error term, and  $m$  is an unknown smooth function. The second type is the fully nonseparable nonparametric panel data models

$$y_{it} = m(x_{it}, \alpha_i, u_{it}), \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (7.2)$$

where  $x_{it}$ ,  $\alpha_i$ , and  $u_{it}$  are defined as above, and the structural function  $m(x, \alpha, u)$  is unknown.

### 7.1 Partially separable nonparametric panel data models

Evdokimov (2009) studies the identification and estimation of the structural function  $m(x, \alpha)$  in (7.1). For simplicity, we focus on the case where  $T = 2$ . Let  $f_{A|B}(\cdot|b)$  and  $\phi_{A|B}(\cdot|b)$  denote the conditional p.d.f. and conditional characteristic function (c.h.f.) of  $A$  given  $B = b$ , respectively. Let  $x_{i,(-t)} = x_i \setminus x_{i,t}$  and  $u_{i,(-t)} = u_i \setminus u_{i,t}$ , where  $x_i = (x_{i1}, x_{i2})$  and  $u_i = (u_{i1}, u_{i2})$ .

We first assume that (i)  $\{x_i, \alpha_i, u_i\}$  is an i.i.d. random sample; (ii)  $f_{u_{it}|x_{it}, \alpha_i, x_{i,(-t)}, u_{i,(-t)}}(u_t|x, \alpha, x_{(-t)}, u_{(-t)}) = f_{u_{it}|x_{it}}(u_t|x)$ ; (iii)  $E(u_{it}|x_{it}) = 0$  a.s.; (iv)  $\phi_{u_{it}|x_{it}}(u|x)$  does not vanish for all  $u \in \mathbb{R}$ ,  $x$  on the support  $\mathcal{X}$  of  $x_{it}$ , and  $t \in \{1, 2\}$ ; (v)  $E[|m(x_{it}, \alpha_i)|x_i]$  and  $E[|u_{it}|x_{it}]$  are bounded a.s. for each  $t$ ; (vi) the joint p.d.f.  $f_{x_{i1}, x_{i2}}(\cdot, \cdot)$  of  $x_{i1}$  and  $x_{i2}$  satisfies  $f_{x_{i1}, x_{i2}}(x, x) > 0$  for all  $x \in \mathcal{X}$ ; (vii)  $m(x, \alpha)$  is increasing in  $\alpha$  for all  $x$ ; (viii)  $\alpha_i$  and  $x_i = (x_{i1}, x_{i2})$  are independent, (ix)  $\alpha_i$  has a uniform distribution on  $[0, 1]$ . Under these conditions, Evdokimov (2009) shows that

1. Under Assumptions (i)-(iv), the conditional distributions of  $m(x, \alpha_i)$ ,  $u_{i1}$  and  $u_{i2}$  given  $x_{i1} = x_{i2} = x$  is identified for all  $x \in \mathcal{X}$ .
2. Under (ii) and (iv), the c.h.f. of  $m(x, \alpha_i)$  given  $x_{it} = x$  is identified as  $\phi_{m(x, \alpha_i)|x_{it}}(s|x) = \phi_{y_{it}|x_{it}}(s|x) / \phi_{u_{it}|x_{it}}(s|x)$ .
3. By the equivalence of c.h.f. and conditional cumulative distribution function (c.d.f.), this implies that the conditional c.d.f.  $F_{m(x, \alpha_i)|x_{it}}(\cdot|x)$  of  $m(x, \alpha_i)$  given  $x_{it} = x$  is identified by

$$F_{m(x, \alpha_i)|x_{it}}(w|x) = \frac{1}{2} - \lim_{A \rightarrow \infty} \int_{-A}^A \frac{e^{-\mathbf{i}sw}}{2\pi \mathbf{i}s} \phi_{m(x, \alpha_i)|x_{it}}(s|x) ds \text{ for all } (w, x) \in \mathbb{R} \times \mathcal{X}$$

at the continuity of the c.d.f. in  $w$ , where  $\mathbf{i} = \sqrt{-1}$ .

4.  $m(x, \alpha)$  is then identified by  $m(x, \alpha) = F_{m(x, \alpha_i)|x_{it}}^{-1}(\alpha|x)$  for all  $x \in \mathcal{X}$  and  $\alpha \in (0, 1)$ , where  $F_{m(x, \alpha_i)|x_{it}}^{-1}(\cdot|x)$  is the inverse function of  $F_{m(x, \alpha_i)|x_{it}}(\cdot|x)$ .

The key in the proof of the above identification results lie in the first step. Consider the special case when  $u_{i1}$  and  $u_{i2}$  are identically and symmetrically distributed, conditional on  $x_{i1} = x_{i2} = x$ . Then the c.h.f. of  $y_{i2} - y_{i1}$  given  $x_{i1} = x_{i2} = x$  equals

$$\begin{aligned} \phi_{y_{i2}-y_{i1}}(s|x_{i1} = x_{i2} = x) &= E[\exp(\mathbf{i}s(y_{i2} - y_{i1})) | x_{i1} = x_{i2} = x] \\ &= E[\exp(\mathbf{i}s(u_{i2} - u_{i1})) | x_{i1} = x_{i2} = x] \\ &= \phi_U(s|x) \phi_U(-s|x) = \phi_U(s|x)^2 \end{aligned}$$

where  $\phi_U(s|x)$  denotes the c.h.f. of  $U_{it}$  given  $x_{it} = x$ . Consequently,  $\phi_U(s|x)$  is identified because  $\phi_{y_{i2}-y_{i1}}(s|x_{i1} = x_{i2} = x)$  can be identified from the observed data.

When Assumption (viii) is violated, Evdokimov (2009) shows that the structural equation in a correlated random effects model can also be identified. The key assumption in this case is the normalization condition: there exists  $\bar{x} \in \mathcal{X}$  such that  $m(\bar{x}, \alpha) = \alpha$  for all  $\alpha$ . Similar conditions are imposed in early literature on nonseparable nonparametric models, see, Matzkin (2003) and Altonji and Matzkin (2005).

Based on the identification results, Evdokimov (2009) considers consistent estimation of the structural function  $m(x, \alpha)$  which boils down to the estimation of c.d.f. and conditional quantile functions. Nevertheless, he needs to estimate  $\phi_{m(x, \alpha_i)|x_{it}}(s|x)$  by a conditional deconvolution approach which yields extremely slow convergence rates. In particular, if the idiosyncratic error term is normally distributed, the conditional deconvolution estimator converges to its truth at

the logarithm rate. Besides, no distributional theory has yet been established so far for such an estimator, and no dynamic lagged dependent variable is allowed to be a regressor in the structural equation.

## 7.2 Fully nonseparable nonparametric panel data models

For the fully nonseparable nonparametric panel data model in (7.2), Altonji and Matzkin (2005), Bester and Hansen (2007), and Hoderlein and White (2009) study conditions for identification and estimation of the structural functional itself or the local average derivatives. Here we focus on the two estimators of Altonji and Matzkin (2005) and remark on other estimators.

Both estimators of Altonji and Matzkin (2005) involve nonseparable unobservable terms and endogenous regressors, and both are based on a conditional density restriction

$$f(\alpha, u|x', z') = f(\alpha, u|x'', z'') \quad (7.3)$$

for specific values  $(x', z')$  and  $(x'', z'')$  of the vector of conditioning variables  $(x_{it}, z_{it})$ . Here  $f(\cdot, \cdot|x, z)$  denotes the conditional p.d.f. of  $(\alpha_i, u_{it})$  given  $(x_{it}, z_{it}) = (x, z)$ . Similarly,  $f(\cdot, \cdot|x)$  denotes the conditional p.d.f. of  $(\alpha_i, u_{it})$  given  $x_{it} = x$ .

### 7.2.1 Local average response (LAR) estimator

The local average response (LAR) estimator is based on the identification of average marginal effects by assuming the existence of a control variable (CV)  $z_{it}$  that is sufficient for  $x_{it}$  in the distribution of unobservables.

Let  $\varepsilon_{it} = (\alpha_i, u_{it})$ . Then (7.2) can be written as  $y_{it} = m(x_{it}, \varepsilon_{it})$ . When  $m(x, \varepsilon)$  is differentiable in  $x$ ,<sup>2</sup> we can define the local average response (LAR)  $\beta(x)$  as

$$\beta(x) = \int m_x(x, \varepsilon) f(\varepsilon|x) d\varepsilon \quad (7.4)$$

where here and below the use of function arguments as subscripts to functions denotes partial derivatives. Under the conditional independence assumption that

$$f(\varepsilon|x, z) = f(\varepsilon|z), \quad (7.5)$$

---

<sup>2</sup>The LAR  $\beta(x)$  can also be defined if  $m$  is not differentiable as in the binary response case. See Altonji and Matzkin (2005).

$\beta(x)$  can be identified as follows

$$\begin{aligned}
\beta(x) &= \int m_x(x, \varepsilon) f(\varepsilon|x, z) f(z|x) dz d\varepsilon \\
&= \int E[m_x(x, \varepsilon_{it})|x, z] f(z|x) dz \\
&= \int E_x[y_{it}|x, z] f(z|x) dz,
\end{aligned} \tag{7.6}$$

where  $f(z|x)$  denotes the conditional p.d.f. of  $z_{it}$  given  $x_{it}$ . (7.6) forms the basis of Altonji and Matzkin's LAR estimator.

Let  $\hat{E}_x[y_{it}|x, z]$  and  $\hat{f}(z|x)$  denote kernel estimators of  $E_x[y_{it}|x, z]$  and  $f(z|x)$ , respectively. In principle, one could estimate  $\beta(x)$  by

$$\hat{\beta}(x) = \int \hat{E}_x[y_{it}|x, z] \hat{f}(z|x) dz.$$

But this estimator is not easy to analyze because it involves a random denominator problem. Noting that

$$\begin{aligned}
\beta(x) &= \int E_x[y_{it}|x, z] f(z|x) dz \\
&= \frac{\int y f_x(y, x) dy}{f(x)} - \int \frac{f_x(x, z) \int y f(y, x, z) dy}{f(x, z) f(x)} dz,
\end{aligned}$$

where  $f(x)$ ,  $f(x, z)$ ,  $f(y, x, z)$  denote the p.d.f.'s of  $x_{it}$ ,  $(x_{it}, z_{it})$ , and  $(y_{it}, x_{it}, z_{it})$ , respectively, we can estimate  $\beta(x)$  by

$$\hat{\beta}(x) = \frac{\int y \hat{f}_x(y, x) dy}{\hat{f}(x)} - \int \frac{\tau(\hat{f}(x, z), b) \hat{f}_x(x, z) \int y \hat{f}(y, x, z) dy}{\hat{f}(x, z) \hat{f}(x)} dz$$

where  $\hat{f}(x)$ ,  $\hat{f}(x, z)$ ,  $\hat{f}_x(x, z)$ ,  $\hat{f}(y, x, z)$  are kernel estimates of  $f(x)$ ,  $f(x, z)$ ,  $f_x(x, z)$ , and  $f(y, x, z)$ , respectively, and  $\tau$  is a trimming function defined by

$$\tau(s, b) = \begin{cases} 1/s & \text{if } s \geq 2b \\ \left[ \frac{49(s-b)^3}{b^4} - \frac{76(s-b)^4}{b^5} + \frac{31(s-b)^5}{b^6} \right] / 8 & \text{if } b \leq s < 2b \\ 0 & \text{if } s < b \end{cases} .$$

Altonji and Matzkin (2005) establish the asymptotic normality of  $\hat{\beta}(x)$  first for the case of  $T = 1$ . If  $T > 1$ , then one can proceed by first observing an estimator of  $\beta(x)$  for each  $t = 1, \dots, T$ , and averaging the  $T$  estimators to obtain the final estimator of  $\beta(x)$ . It is well known from standard asymptotic analysis in the kernel literature, these  $T$  estimators are asymptotically independent because the covariance between each two of them is of smaller magnitude than the individual variances.

When the dimensional of  $x_{it}$  is high, the rate of convergence of  $\hat{\beta}(x)$  can be undesirably slow. As an alternative, one can consider some weighted average measure of the nonparametric LAR estimator to increase the precision of the estimator. For example, one can consider estimating

$$\bar{\beta} = \int \beta(x) w(x) dx \quad (7.7)$$

for some prescribed weight function  $w$ . As usual, the estimator of  $\bar{\beta}(x)$  will have the regular  $\sqrt{n}$ -rate of convergence.

Clearly, the key assumption underlying the above analysis is the conditional independence assumption in (7.5). This requires (7.3) holds for all values of  $(\alpha, u)$ ,  $(x, x')$  and  $(z', z'')$  such that  $z' = z''$ . The LAR estimator is based upon the (derivative of) conditional expectation  $E(y|x, z)$ . Because of (7.5), holding  $z$  constant also holds the distribution of the unobservable term ( $\varepsilon$ ) constant. Then one can undo the effect of conditioning on  $z$  by integrating  $E_x(y|x, z)$  over an estimate of the distribution of  $z$  given  $x$ .

Bester and Hansen (2007) consider identification and estimation of average marginal effects in a correlated random coefficients models. Instead of assuming the existence of the known CV vector  $z_{it}$ , they assume the existence of a set of sufficient statistics for  $x_{it}$  in the distribution of individual heterogeneity, which is not known but takes on some index form. To be concrete, Bester and Hansen assume that

$$F(\alpha_i|\mathbf{x}_i) = F(\alpha_i|h_1(\mathbf{x}_{i,1}), \dots, h_p(\mathbf{x}_{i,p})) \quad (7.8)$$

for some unknown real-valued functions  $h_s(\mathbf{x}_{i,s})$ ,  $s = 1, \dots, p$ , where for example,  $F(\alpha_i|\mathbf{x}_i)$  denotes the conditional c.d.f. of  $\alpha_i$  given  $\mathbf{x}_i = (x_{i1}, \dots, x_{iT})'$ , a  $T \times p$  matrix, and  $\mathbf{x}_{i,s}$  denotes the  $s$ th column in  $\mathbf{x}_i$ . This assumption is neither more or less general than the conditional independence assumption (7.5) of Altonji and Matzkin (2005): the set of sufficient statistics is unknown but restricted so that there is one sufficient statistic for each covariate in  $x_{it}$  for the restriction in (7.8); the CV  $z_{it}$  has to be unknown but may include interactions of covariates in (7.5). In addition, neither Bester and Hansen's (2007) nor Altonji and Matzkin's (2005) LAR approach identifies the structural function itself.

### 7.2.2 Structural function and distribution (SFD) estimator

To define the structural function and distribution (SFD) estimator, we impose the following assumptions: (i) There exists a real valued function  $g(\varepsilon)$  such that  $y_{it} = m(x_{it}, e_{it})^3$  for

---

<sup>3</sup>We keep using the notation  $m$  in  $m(x, e)$  but keep in mind that this is different from the original structural function  $m(x, \alpha, u)$ .

$e_{it} = g(\varepsilon_{it})$ ; (ii)  $m(x, e)$  is strictly increasing in  $e$  for all  $x$ ; (iii) there exists some value  $\bar{x}$  of  $x$ ,  $m(\bar{x}, e) = e$  for all  $e$ ; (iv) for any value  $\tilde{x}$  of  $x$  there exist values  $\tilde{z}$  and  $\tilde{z}'$  of  $z$  such that  $f(e|\tilde{x}, \tilde{z}) = f(e|\bar{x}, \tilde{z}')$  where  $f(e|x, z)$  denotes the conditional p.d.f. of  $e_{it}$  given  $(x_{it}, z_{it})$ ; (v) for all  $(x, z)$ ,  $f(e|x, z)$  is strictly positive everywhere.

Clearly, (i) indicates that the effect of the vector  $\varepsilon_{it}$  can be aggregated by a scalar-valued unobservable random term  $e_{it} = g(\varepsilon_{it})$ . (ii) assumes monotonicity in the unobservable and (iii) is a normalization restriction. (iv) can be satisfied under some exchangeability conditions. (v) and (ii) guarantee that the conditional c.d.f.  $F(\cdot|x, z)$  of  $y_{it}$  given  $(x_{it}, z_{it}) = (x, z)$  is strictly increasing so that  $m(x, e)$  can be identified via

$$m(x, e) = F^{-1}(F(e|\bar{x}, z')|x, z). \quad (7.9)$$

To see this, noticing that for any value  $x$  there exist values  $z$  and  $z'$  such that for any value  $e$ , we have

$$\begin{aligned} P(e_{it} \leq e|x, z) &= P(e_{it} \leq e|\bar{x}, z') \quad (\text{by (iv)}) \\ &\implies \\ P(m(x, e_{it}) \leq m(x, e)|x, z) &= P(m(\bar{x}, e_{it}) \leq m(\bar{x}, e)|\bar{x}, z') \quad (\text{by (ii)}) \text{ or} \\ P(y_{it} \leq m(x, e)|x, z) &= P(y_{it} \leq m(\bar{x}, e)|\bar{x}, z') \text{ or} \\ F(m(x, e)|x, z) &= F(m(\bar{x}, e)|\bar{x}, z'). \end{aligned}$$

The last line implies (7.9) by (ii) and (v). Let  $F_{e_{it}|x_{it}}(\cdot|x)$  and  $F_{y_{it}|x_{it}}(\cdot|x)$  denote the conditional c.d.f. of  $e_{it}$  and  $y_{it}$  given  $x_{it} = x$ , respectively. Then under (ii),  $F_{e_{it}|x_{it}}$  is identified via

$$F_{e_{it}|x_{it}}(e|x) = F_{y_{it}|x_{it}}(m(x, e)|x). \quad (7.10)$$

Given the above identification results, we can obtain estimators of  $m(x, e)$  and  $F_{e_{it}|x_{it}}(e|x)$  straightforwardly via the kernel method. Both estimators involve the kernel estimates of  $F(\cdot|x, z)$  and its inverse function (conditional quantile function)  $F^{-1}(\cdot|x, z)$ . The latter also involves the estimation of  $F_{y_{it}|x_{it}}(\cdot|x)$ . Altonji and Matzkin (2005) formally establish the asymptotic normality of either estimator.

It is worth mentioning that neither the LAR nor the SFD estimator deals with dynamics in the model. The LAR estimator can be used to estimate the marginal effects of  $x_{it}$  on  $y_{it}$  in a censored regression model but neither can be used to study the effects on a latent dependent variable. The SFD estimator estimates some structural function but it is different from the original one.

### 7.2.3 Nonparametric identification and estimation without monotonicity

Hoderlein and White (2009) consider the general class of nonseparable panel models of the form

$$y_{it} = m(x_{it}, z_{it}, \alpha_i, u_{it}), \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (7.11)$$

where  $z_{it}$  is a  $q \times 1$  vector of observed variables,  $x_{it}, \alpha_i$ , and  $u_{it}$  are defined as before. Their interest centers on the effect of  $x_{it}$  on  $y_{it}$  by controlling the influence of all other variables, whether observed like  $z_{it}$  or unobserved like  $\alpha_i$  and  $u_{it}$ .

Without assuming that  $m(x, z, \alpha, u)$  is monotonic in  $\alpha$  or  $u$ , the structural function  $m$  itself and its derivatives are not identified, but certain of its conditional expectations and their derivatives are. Like early estimators in the nonseparable panel literature, Hoderlein and White's estimator does not allow for lagged dependent variables either. In addition, they can only identify effects for the subpopulation for which  $x_{i1} - x_{i2} = 0$  and  $z_{i1} - z_{i2} = 0$  in the case of  $T = 2$ .

### 7.3 Testing of monotonicity in nonseparable nonparametric panel data models

Despite the wide use of monotonicity of the structural function in individual heterogeneity (e.g., Matzkin (1999), Altonji and Matzkin (2005), Imbens and Newey (2009), Evdokimov (2009), among others), Hoderlein and Mammen (2007, 2009) argue that such an assumption may not be fully justified in economics, say, when the individual effects represent the unobserved heterogeneity in preferences or technologies. Moreover, as Hoderlein, Su, and White (2010) demonstrate, some key identification results fail when monotonicity is violated. This motivates them to consider tests of monotonicity in nonseparable nonparametric panel data models. Under some strict exogeneity conditions, they propose two tests for monotonicity of unobservables in panel nonseparable nonparametric panel data models. The first works under some ideal situation where the unobservables vary across  $i$  but not  $t$  dimension ( $t$  may not be time index). The second works in large dimensional panel where both  $n$  and  $T$  approach  $\infty$  and both time-invariant and time-varying unobservables are present.

Consider first the case where the unobservables vary across individuals but not "time":

$$y_{it} = m(x_{it}, \alpha_i), \quad i = 1, \dots, n, \quad t = 1, \dots, T,$$

where  $\alpha_i$  is i.i.d. uniformly distributed on  $[0, 1]$ , and  $(x_{it}, \alpha_i)$  is identically distributed across

*i.* Under the null hypothesis that  $m(x, \cdot)$  is strictly increasing for any  $x$ , we have

$$\alpha_i = F_t(y_{it}|x_{it}) \text{ a.s. for all } (i, t)$$

where  $F_t(\cdot|x)$  is the conditional c.d.f. of  $y_{it}$  given  $x_{it} = x$ . If we further assume that  $\alpha_i$  is independent of  $x_{it}$  ( $x_{it}$  is exogenous), then we can show that this conditional c.d.f. is time-invariant, that is,  $F_t$  should not depend on  $t$  and can be abbreviated as  $F$ . Thus, we can write the null hypothesis as

$$H_0 : F(y_{is} | x_{is}) = F(y_{is} | x_{is}) \text{ a.s. for all } (t, s) \quad (7.12)$$

Significantly, exogeneity and the time-invariance of  $\alpha_i$  jointly ensure that  $F_t$  is time invariant. When exogeneity or monotonicity fails, we generally have the alternative

$$H_1 : P[F_t(y_{is} | x_{is}) = F_s(y_{is} | x_{is})] < 1 \text{ for some } t \neq s.$$

Let  $\hat{F}_t$  be suitable estimator of  $F_t$ . We can consider the following test statistic

$$D_n \equiv \sum_{t=1}^{T-1} \sum_{s=t+1}^T \sum_{i=1}^n (\hat{F}_t(y_{it} | x_{it}) - \hat{F}_s(y_{is} | x_{is}))^2.$$

Hoderlein, Su, and White (2010) obtain the estimate  $\hat{F}_t(y|x)$  by the local polynomial method and demonstrate after correct centering,  $h^{p/2}D_n$  is asymptotically normality distributed under the null and diverges to infinity under the alternative, where  $h$  is the bandwidth parameter used in the local polynomial estimation.

Now consider the nonseparable structure of the form

$$y_{it} = m(x_{it}, \alpha_i, u_{it}), \quad i = 1, \dots, n, \quad t = 1, \dots, T,$$

$\alpha_i$  is i.i.d. uniformly distributed on  $[0, 1]$ , and  $(x_{it}, \alpha_i, u_{it})$  is i.i.d. across  $i$ , and identically distributed across  $t$ . Define some nonnegative weight functions  $w_\tau(x)$  on the support of  $x_{it}$ ,  $\tau = 1, \dots, T$ . Assuming that  $(x_{it}, u_{it}) \perp \alpha_i$ , we have

$$\tilde{Y}_{\tau,i} = E[y_{it}w_\tau(x_{it})|\alpha_i] = \int m(x, \alpha_i, u)w_\tau(x)dF(x, u) \equiv \bar{m}_\tau(\alpha_i),$$

where  $F(x, u)$  denotes the c.d.f. of  $(x_{it}, u_{it})$ . Clearly,  $\bar{m}_\tau(\cdot)$  is also monotonic under the null hypothesis that  $m(x, \cdot, u)$  is monotone for all  $(x, u)$ . Furthermore,  $\alpha_i$  can be identified as

$$\alpha_i = \bar{m}_\tau^{-1}(\tilde{Y}_{\tau,i}) = \tilde{F}_\tau(\tilde{Y}_{\tau,i})$$



where  $\tilde{F}_\tau$  is the c.d.f. of  $\tilde{Y}_{\tau,i}$ . As a result, we can test the monotonicity by testing the following null hypothesis

$$\tilde{H}_0 : \tilde{F}_\tau(\tilde{Y}_{\tau,i}) = \tilde{F}_\varsigma(\tilde{Y}_{\varsigma,i}) \text{ a.s. for all } (\tau, \varsigma). \quad (7.13)$$

The test statistic is

$$\hat{D}_{nT} \equiv \sum_{\tau=1}^{T-1} \sum_{\varsigma=\tau+1}^T \sum_{i=1}^n (\hat{F}_{n,T,\tau}(\bar{Y}_{T,\tau,i}) - \hat{F}_{n,T,\varsigma}(\bar{Y}_{T,\varsigma,i}))^2.$$

where for  $\tau = 1, \dots, T$ ,  $\hat{F}_{n,T,\tau}(y) = n^{-1} \sum_{i=1}^n \mathbf{1}\{\bar{Y}_{T,\tau,i} \leq y\}$ ,  $\bar{Y}_{T,\tau,i} = T^{-1} \sum_{t=1}^T y_{it} w_\tau(x_{it})$  is a consistent estimate of  $\tilde{Y}_{\tau,i}$  under weak conditions, and  $\mathbf{1}\{\cdot\}$  is the usual indicator function. Under some regularity conditions, Hoderlein, Su, and White (2010) show that limit distribution of  $\hat{D}_{nT}$  is given by weighted chi-squares under the null.

## 8 Concluding Remarks

In this paper we survey some of the recent developments on NP and SP panel data models. Due to space limitation, we omit some of the important areas in this literature. This includes NP and SP limited dependent variable models (see Ai and Li (2008)), and NP and SP panel models with spatial dependence. It is worth mentioning that the latter area is under-developed in econometrics. Other areas that seem promising to us include NP or SP panel data models that impose some curvature restrictions (e.g., monotonicity, concavity, homogeneity) or require less restrictions (e.g., exogeneity, separability, monotonicity). In the nonseparable nonparametric models, no estimator has been proposed to deal with dynamic panel data models. Obviously, this is an interesting yet challenging research topic.

## References

- Ahmad, I., S. Leelahanon and Q. Li. 2005. Efficient estimation of semiparametric varying coefficient models. *Annals of Statistics* 33: 258-283.
- Ai, C. and X. Chen. 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71: 1795–1843.
- Ai, C. and Q. Li. 2008. Semi-parametric and non-parametric methods in panel data models. In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. L. Mátyás and P. Sevestre (eds.), pp. 451-478. 3rd ed. Springer.
- Altonji, J. G. and R. L. Matzkin. 2005. Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica* 73: 1053-1102.
- Arellano, M. 2003. *Panel Data Econometrics*. Oxford: Oxford University Press.

- Bai, J. 2009. Panel data models with interactive fixed effects. *Econometrica* 77, 1229-1279.
- Baltagi, B. H. 2008. *Econometric Analysis of Panel Data*. 4th ed. West Sussex: John Wiley & Sons.
- 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.
- Baltagi, B.H., J. Hidalgo, and Q. Li. 1996. A nonparametric test for poolability using panel data. *Journal of Econometrics* 75: 345-367.
- Baltagi, B.H. and Q. Li. 1996. On instrumental variable estimation of semiparametric dynamic panel data models. *Economics Letters* 76: 1-9.
- Baltagi, B.H. and Q. Li. 2000. Efficient instrumental variable estimation of semiparametric dynamic panel data models. *Working paper*, Dept. of Economics, Texas A&M University.
- Bester, A. and C. Hansen. 2007. Identification of marginal effects in a nonparametric correlated random effects model. *Working paper*, Dept. of Economics, University of Chicago.
- Cai, Z. and Q. Li. 2008. Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory* 24: 1321-1342.
- Cai, Z., M. Das, H. Xiong and Z. Wu. 2006. Functional coefficient instrumental variables models. *Journal of Econometrics* 133, 207-241.
- Cai, Z., and H. Xiong. 2006. Efficient estimation of partially varying-coefficient instrumental variables models: *Working paper*, WISE, Xiamen University.
- Das, M. 2005. Instrumental variables estimation of nonparametric models with discrete endogenous variables. *Journal of Econometrics* 124: 335-361.
- Evdokimov, K. 2009. Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. *Working Paper*, Dept. of Economics, Yale University.
- González, A., T. Teräsvirta and D. van Dijk. 2005. Panel smooth transition regression models. *Working paper*, Stockholm School of Economics.
- Hansen, B. E. 1999. Threshold effects in non-dynamic panels: estimation, testing and inference. *Journal of Econometrics* 93: 345-368.
- Henderson, D. J., R. J. Carroll, and Q. Li. 2008. Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics* 144: 257-275.
- Henderson, D. J. and A. Ullah. 2005. A nonparametric random effects estimator. *Economics Letters* 88: 403-407.
- Henderson, D. J. and A. Ullah. 2008. Nonparametric estimation in a one-way error component model: a Monte Carlo analysis. *Working paper*, Dept. of Economics, University of California, Riverside.

- Hoderlein, S. and E. Mammen 2007. Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* 75: 1513-1518.
- Hoderlein, S. and E. Mammen 2009. Identification and estimation of local average derivatives in non-separable models without monotonicity. *Econometrics Journal* 12: 1-25.
- Hoderlein, S., L. Su, and H. White. 2010. Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Working Paper*, Dept. of Economics, Brown University.
- Hoderlein, S. and H. White. 2009. Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Working Paper*, Dept. of Economics, Brown University.
- Hsiao, C. 2003. *Analysis of Panel Data*. 2nd ed. Cambridge: Cambridge University Press.
- 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.
- Huang, X., 2006. Nonparametric estimation in large panel with cross-section dependence. *Working paper*, Department of Economics & Finance, Kennesaw State University.
- Imbens, G. W. and W. K. Newey. 2009. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77: 1481-1512.
- Jin, S. and L. Su. 2010. A nonparametric poolability test for panel data models with cross section dependence. *Working paper*, School of Economics, Singapore Management University.
- Kniesner, T. and Q. Li. 2002. Semiparametric panel data models with heterogeneous dynamic adjustment: theoretical consideration and an application to labor supply. *Empirical Economics* 27: 131-148.
- Lee, Y. 2008. Nonparametric estimation of dynamic panel models with fixed effects. *Working paper*, Dept. of Economics, Michigan University.
- Lee, Y. and D. Mukherjee. 2008. New nonparametric estimation of the marginal effects in fixed-effects panel models: an application on the environmental Kuznets curve. *Working paper*, Dept. of Economics, Michigan University.
- Li, Q. 2000. Efficient estimation of additive partially linear models. *International Economic Review* 41: 1073-1092.
- Li, Q. and C. Hsiao. 1998. Testing Serial Correlation in Semiparametric Panel Data Models. *Journal of Econometrics* 87: 207-237.
- Li, Q. and J. S. Racine. 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton: Princeton University Press.
- Li, Q. and T. Stengos. 1996. Semiparametric estimation of partially linear panel data models. *Journal of Econometrics* 71: 289-397.

- Li, Q. and A. Ullah. 1998. Estimating partially linear panel data models with one-way error components. *Econometric Reviews* 17: 145-166.
- Lin, X. and R. J. Carroll. 2000. Nonparametric function estimation for clustered data when the predictor is measured without/with error, *Journal of the American Statistical Association* 95: 520-534.
- Lin, X. and R. J. Carroll. 2001a. Semiparametric regression for clustered data. *Biometrika* 88: 1179-1185.
- Lin, X. and R. J. Carroll. 2001b. Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association* 96: 1045-1055.
- Linton, O. and J. P. Nielsen. 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82: 93-101.
- Martins-Filho, C. and F. Yao. 2009. Nonparametric regression estimation with general parametric error covariance, *Journal of Multivariate Analysis* 100: 309-333.
- Matzkin, R. L. 2003. Nonparametric estimation of nonadditive random functions. *Econometrica* 71: 1339-1375.
- Mundra, K. 2005. Nonparametric slope estimation for fixed-effect panel data. *Working Paper*, Department of Economics, San Diego State University.
- Mukherjee, D. 2002. Nonparametric and semiparametric generalized panel data analysis of convergence and growth, *Dissertation*, University of California, Riverside
- Mukherjee, D. and A. Ullah. 2003. Semiparametric analysis of generalized panel data : An application unpublished manuscript, University of California, Riverside.
- Newey, W. K. 1990. Efficient instrumental variables estimation of nonlinear models. *Econometrica* 58: 809-837.
- Pagan, A. and A. Ullah. 1999. *Nonparametric Econometrics*. Cambridge: Cambridge University Press.
- Pesaran, M. H. 2006. Estimation and inference in large heterogeneous panels with multifactor error. *Econometrica* 74: 967-1012.
- Ruckstuhl, A. F., A. H. Welsh, and R. J. Carroll. 2000. Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statistica Sinica* 10: 51-71.
- Su, L., Y. Chen, and A. Ullah. 2009. Functional coefficient estimation with both categorical and continuous data. *Advances in Econometrics* 25: 131-167.
- Su, L. and S. Jin. 2010. Sieve estimation of panel data models with cross section dependence. *Working paper*. School of Economics, Singapore Management University.
- Su, L., and A. Ullah. 2006. Profile likelihood estimation of partially linear panel data models with fixed effects, *Economics Letters* 92: 75-81

- Su, L., and A. Ullah. 2007. More efficient estimation of nonparametric panel data models with random effects. *Economics Letters* 96: 375-380.
- Su, L., and A. Ullah. 2009. Testing conditional uncorrelatedness. *Journal of Business and Economic Statistics* 27: 18-29.
- Su, L., A. Ullah, and Y. Wang. 2010. A note on nonparametric regression estimation with general parametric error covariance. *Working paper*, School of Economics, Singapore Management Univ.
- Su, L. and Y. Zhang. 2010a. Semiparametric panel data model with interactive effects. *Working paper*, School of Economics, Singapore Management Univ.
- Su, L. and Y. Zhang. 2010b. Testing for cross section dependence in nonparametric panel data models. *Working paper*, School of Economics, Singapore Management Univ.
- Sun, Y., R. J. Carroll, and D. Li. 2009. Semiparametric estimation of fixed effects panel data varying coefficient models. *Advances in Econometrics* 25: 101-130.
- Tran, K. C. and E. G. Tsionas. 2010. Local GMM estimation of semiparametric panel data with smooth coefficient models. *Econometric Reviews* 29: 39-61.
- Ullah, A. and N. Roy. 1998. Nonparametric and semiparametric econometrics of panel data. In *Handbook of Applied Economic Statistics*, ed. A. Ullah and D. E. A. Giles, 579-604. New York: Marcel Dekker.