Nonparametric Panel Data Regression Models

Yiguo Sun
Department of Economics
University of Guelph
Guelph, Ontario N1G 2W1, Canada

Yu Yvette Zhang
Department of Agriculture Economics
Texas A&M University
College Station, TX 77843

Qi Li*
Department of Economics
Texas A&M University
College Station, TX 77832-4228

This version: June 28, 2013

This chapter selectively reviews some recent developments on nonparametric panel data regression models, including different estimation methods developed for nonparametric panel data mean regression models, some introduction on nonparametric panel data quantile regression models, nonseparable nonparametric panel data models and nonparametric poolability tests and cross-sectional independence tests.

Key Words: Panel data models; Fixed effects; Kernel estimator; Penalized spline smoothing estimator; Quantile regressions; Nonparametric poolability tests; Cross-sectional dependence.

1 Introduction

The increasing availability of panel data has nourished fast growing development in panel data econometrics analysis. Textbooks and survey articles have been published to help readers get a quick jump-start into this exciting field. For parametric panel data analysis, the most popularly cited econometrics textbooks include Arellano (2003), Baltagi (2005) and Hsiao (2003). For non-/semi-parametric panel data analysis, recent survey articles include Arellano and Honore (2001), Ai and Li (2008) and Su and Ullah (2011). Arellano and Honore (2001) and Ai and Li (2008) reviewed parametric and semiparametric panel data censored, discrete choice, and sample selective models with fixed effects, while Su and Ullah (2011) focused on nonparametric panel data models, partially

*The corresponding author: Qi Li, email: qi@econmail.tamu.edu
linear and varying coefficient panel data models and nonparametric panel data models with cross-sectional dependence and under nonseparability. This chapter joins the others to provide readers with a selective survey on nonparametric panel data analysis in the framework of conditional mean or conditional quantile regressions. Unavoidably, our survey has some overlaps with Su and Ullah’s (2011) survey on nonparametric panel data estimation.

Given a panel data set \((X_{i,t}, Y_{i,t}) : i = 1, ..., n; t = 1, ..., T\), we consider a nonparametric panel data model

\[ Y_{i,t} = m(X_{i,t}) + u_{i,t}, \quad i = 1, 2, ..., n; t = 1, ..., T, \quad (1.1) \]

where \(Y_{i,t}\) and \(X_{i,t}\) are read-valued random variables, \(m(\cdot)\) is an unknown smooth measurable function, and \(u_{i,t}\) is the error term. This chapter mainly focuses on estimation methods and test statistics developed for nonparametric panel data models, with the exception that semiparametric models as well as nonparametric models are included in the presence of cross-sectional dependent errors in Section 2.3.

Note that model (1.1) does allow \(X_{i,t}\) to be a random vector of discrete and/or continuous components. However, we choose to consider the case that \(X_{i,t}\) is a continuously distributed scalar random variable out of two concerns. First, the estimation methods and test statistics included in this chapter can be naturally extended to the case that \(X_{i,t}\) is a random vector. However, this extra generality increases notation complexity in matrix representation which may hurt readers’ focus on grasping the essence of the methodology explained in the chapter. Second, the curse-of-dimensionality becomes evident in finite samples if \(X_{i,t}\) contains more than two continuous variables, which makes semiparametric models strong competitors to pure nonparametric regression models. We refer the readers to Li and Racine’s (2007) textbook for popularly studied semiparametric models for cross-sectional and time series data.

This rest of this chapter is organized as follows. Section 2 focuses on nonparametric estimation of conditional mean regression models in panel data setup. Section 3 is devoted to nonparametric conditional quantile regression models. Section 4 discusses nonseparable panel data models. Section 5 contains nonparametric tests on poolability and cross-sectional independence/uncorrelationness. The last section concludes.

2 Conditional Mean Regression Models

Consider a decomposition of the error term \(u_{i,t}\) into the following three cases: (a) \(u_{i,t} = \mu_i + \epsilon_{i,t}\) with \(\mu_i \sim i.i.d. (0, \sigma^2_{\mu})\); (b) \(u_{i,t} = \lambda_{i} + \epsilon_{i,t}\) with \(\lambda_{i} \sim i.i.d. (0, \sigma^2_{\lambda})\); (c) \(u_{i,t} = \mu_i + \delta_i^T g_t + \epsilon_{i,t}\), where \(\epsilon_{i,t} \sim i.i.d. (0, \sigma^2_{\epsilon})\) and \(\delta_i\) and \(g_t\) are unobservable \(d \times 1\) vectors independent of the idiosyncratic errors \(\{\epsilon_{i,t}\}\). Case (a) and case (b) assume that the error term contains a cross-sectional and time fixed effect, respectively. In case (c), \(g_t\) are unobserved time varying common factors and \(\delta_i\) are the factor loadings. As case (b) can be studied in the same spirit as case (a), we choose not to consider
case (b) in this survey.

Section 2.1 discusses random effects panel data models. Section 2.2 deals with fixed effects panel data models, and Section 2.3 allows for cross sectional dependence.

### 2.1 Random effects panel data models

This subsection focuses on random-effects panel data models where \( E(u_{i,t}|X_{i,t}) = 0 \) for all \( i \) and \( t \) for large \( n \) and small \( T \). \( \{(X_{i,t}Y_{i,t})\} \) are assumed to be i.i.d. across index \( i \) with possible within-group correlations, so that \( V = E(uu^T|X) = \text{diag}(V_1, \ldots, V_n) \) with \( V_i = E(u_iu_i^T|X_i) \), where \( w_i = (w_{i,1}, \ldots, w_{i,T})^T \) and \( w = (w_1^T, \ldots, w_n^T)^T \) with \( w = u \) or \( X \). In the presence of within-group correlation, the \((nT) \times (nT)\) conditional covariance matrix \( V \) is not diagonal. We review the current literature on whether and how to take into consideration the within-group correlation, while estimating \( m(\cdot) \) nonparametrically.

Assuming that the unknown function \( m(\cdot) \) is continuously twice-differentiable around an interior point of interest, \( x \), and applying Taylor expansion give

\[
m(X_{i,t}) = m(x) + m'(x)(X_{i,t} - x) + r_{i,t},
\]

where \( r_{i,t} = O\left((X_{i,t} - x)^2\right) \) is negligible if \( X_{i,t} \) is sufficiently close to \( x \). Plugging this approximation into model (1.1) yields

\[
Y_{i,t} \approx m(x) + m'(x)(X_{i,t} - x) + u_{i,t}. \tag{2.1}
\]

Denoting \( w = (w_{1,1}, \ldots, w_{1,T}, w_{2,1}, \ldots, w_{2,T}, \ldots, w_{n,1}, \ldots, w_{n,T})^T \) for \( w = Y, X \) or \( u \), and a 2 by 1 vector \( \beta(x) = \left[m(x), m'(x)\right]^T \), we calculate the local linear weighted least square (LLWLS) estimator of \( \beta(x) \) as follows

\[
\hat{\beta}(x) = \arg\min_{\beta(x)} \|Y - Z(x)\beta(x)\|_W(x) [Y - Z(x)\beta(x)] \tag{2.2}
\]

where the typical row vector of the \((nT)\) by 2 matrix \( Z(x) \) is \((1, X_{i,t} - x)\) and \( W(x) \) is a \((nT)\) by \((nT)\) weighting matrix depending on \( x \). Taking the first order derivative with respect to \( \beta(x) \) gives a nonparametric estimating equation

\[
Z(x)^T W(x) \left[Y - Z(x) \hat{\beta}(x)\right] = 0, \tag{2.3}
\]

which yields

\[
\hat{\beta}(x) = \left[Z(x)^T W(x) Z(x)\right]^{-1} Z(x)^T W(x) Y. \tag{2.4}
\]

Taking \( W(x) \equiv K_h(x) \) yields the local linear least squares (LLLS) estimator of \( \beta(x) \), where the typical element of the \((nT)\) by \((nT)\) diagonal matrix \( K_h(x) \) is \( K((X_{i,t} - x)/h) \), and \( K(\cdot) \) is a second-order kernel function and \( h \) is the bandwidth. **Suppose that** \( h \to 0 \) **as** \( n \to \infty \) **such that** \( K((X_{i,t} - x)/h) \) **assigns a positive weight to** \( X_{i,t} \), **only if** \( |X_{i,t} - x| \leq ch \) **for** some positive constant \( c \). **Do we need this sentence?** The LLLS estimator evidently does not take into account possible correlation between \( u_{i,t} \) and \( u_{i,s} \) for \( t \neq s \). Ruckstuhl, Welsh, and Carroll (2000, Th.1) obtained the asymptotic bias and variance of the LLLS estimator of
where \( f \) and \( m \) were known, transformation through and Carroll’s (2000) estimators by points closer to estimator. These three estimators are equivalent when the covariance matrix the asymptotic bias of the LLLS estimator equals matrix. Under some regularity conditions, Ruckstuhl, Welsh, and Carroll (2000, Th.1) showed that Gaussian errors. Under assumption (i.e., assuming assumption (i.e., assuming assumption (i.e., assuming assumption (i.e., assuming assumption (i.e., assuming assumption (i.e., assuming assumption (i.e., assuming that (i) is a diagonal matrix and with small \( T \) for the local linear regression approach and that the same result holds for \( \hat{\beta}_{LCH}(x) \) with homogeneous Gaussian errors. Under working independence assumption (i.e., assuming \( V \) is a diagonal matrix or ignoring within-group correlation in errors), \( \hat{\beta}(x) \) is the LLLS estimator and is shown to be asymptotically more efficient than the LLWLS estimators by Lin and Carroll (2000) under some regularity conditions. Hence, these results indicate that ignoring within-group correlation in \( \{u_{i,t}\} \) can be beneficial if one estimates \( m(\cdot) \) by the kernel-based local linear regression method.

To facilitate further exposition on whether and how to explore within-group correlation (i.e., \( E(u_{i,t}u_{s,t}|X_{i,t},X_{i,s}) \neq 0 \) for some \( t \neq s \)), we introduce two useful concepts below. Let us write the \( i \)th unit’s covariance matrix \( V_i \) in a variance-correlation form,

\[ V_i = A_i^{1/2} RA_i^{1/2}, \tag{2.7} \]
where $A_i = \text{diag} (\sigma^2_{i1}, \ldots, \sigma^2_{iT})$ with $\sigma^2_{it} = \text{Var} (u_{it} | X_{it})$, and $R$ is a correlation matrix assumed to be common to all the cross-sectional units. This construction allows possible heteroskedastic $u_i$ with the same correlation structure across units. A working independence matrix is a correlation matrix assuming an i.i.d. sequence of errors (i.e., $R = I_{nT}$), while a working correlation matrix is a correlation matrix subjectively assumed to capture the within-group correlation structure of error terms and may or may not equal the true correlation matrix (i.e., $R$ can be any correlation matrix).

Wang (2003) indicated that the reason that the LIWLS estimators perform no better than the LLLS estimator in the presence of within-group correlation is that the kernel estimators asymptotically use at most one observation from each unit to calculate $\hat{\beta} (x)$ with a sufficiently large $n$ and finite $T$, so that the working independence assumption is reasonable with the local linear least squares estimation method. Localizing data first then weighting the localized data by $V^{-1/2}$, Welsh, Lin and Carroll (2002) showed that, with a working independence matrix, the smoothing spline method is equivalent to the kernel method with a specific higher-order kernel. With a working correlation matrix, however, the estimators based on smoothing spline and penalized regression splines behave differently than the kernel-based LIWLS estimators. Specifically, Welsh et al. (2002, p.486) showed that the LIWLS estimator, $\hat{\beta}_{LC} (x)$, is asymptotically local and most efficient under working independence, while the smoothing spline and penalized regression spline estimators generally not asymptotically local and are most efficient when the working covariance matrix equals the true within-group covariance matrix. When $V$ is not a diagonal matrix, their theoretical results and simulation results indicate that the smoothing spline and penalized regression spline methods perform better than the usual kernel estimation method. Lin and Carroll (2006) proposed a profile likelihood approach to estimate $m (x)$ with $W (x) \equiv V^{-1} K_h (x)$, where all the observations from chosen unit $i$ are used to calculate the nonparametric estimator.

Given a known non-diagonal error covariance matrix, $V$, Ruckstuhl et al. (2000) proposed to first transform the error term to an i.i.d. sequence by setting $Y^*_i = \tau V_i^{-1/2} Y_i + (I - \tau V_i^{-1/2}) m (X)$, which gives

$$Y^* = m (X) + \tau V^{-1/2} u,$$

where $m (X) = [m (X_{1,1}), \ldots, m (X_{1,T}), m (X_{2,1}), \ldots, m (X_{2,T}), \ldots, m (X_{n,1}), \ldots, m (X_{n,T})]^T$. Since $V^{-1/2} u$ is an i.i.d. sequence, one can estimate $m (X)$ by the LIWLS estimator with $y^*$ as the working dependent variable with a bandwidth $h$.

As $Y^*$ is unknown, they calculate the LIWLS estimator $\hat{m} (X)$ with a bandwidth $h_0$ and replace $Y^*$ by $\hat{Y}^* = \tau V^{-1/2} Y + (I - \tau V^{-1/2}) \hat{m} (X)$. Assuming the classical nonparametric one-way component random-effects panel data model, Su and Ullah (2007) showed that if $h_0$ converges to zero faster than the optimal rate $n^{-1/5}$, Ruckstuhl et al.’s (2000) two-step estimator has the same asymptotic bias term as the LIWLS estimator $\hat{m} (x)$, where (2.5) and (2.6) are used to estimate the unknown variances. Ruckstuhl et al. (2000, Th.4) indicated that the two-step estimator has a smaller asymptotic variance than the LIWLS estimator if $\tau = \sigma^{-1}_e$.

The existing literature has established that more efficient estimators than the LIWLS estimator can be achieved by taking into account within-group correlation. However, other than the classical
one-way error component case, more researches are required to further understand how to incorporate the within-group correlation for general cases when $V$ is known as well as when $V$ is unknown. For the classical one-way component random-effects models, Henderson and Ullah (2005) presented consistency results of the feasible version of Ullah and Roy’s (1998) and Lin and Carroll’s (2000) estimators via a two-step procedure for large $n$ and small $T$. In the first step, Henderson and Ullah (2005) calculated (2.5) and (2.6). In the second step, the feasible LLWLS estimators are calculated with $V$ replaced by its consistent estimator $\hat{V} = I_n \otimes (\hat{\sigma}_{\mu}^2 I_T + \hat{\sigma}_e^2 I_T)$, where “$\otimes$” stands for the Kronecker product. As $\hat{\sigma}_{\mu}^2 = \sigma_{\mu}^2 + O_p \left( n^{-1/2} \right)$ and $\hat{\sigma}_e^2 = \sigma_e^2 + O_p \left( n^{-1/2} \right)$, $\hat{m}(x) - m(x) = O_p \left( h^2 \right) + O_p \left( (nh)^{-1/2} \right)$, and $n^{-1/2}$ converges to zero faster than $O_p \left( h^2 \right) + O_p \left( (nh)^{-1/2} \right)$. $\sigma_{\mu}^2$ and $\sigma_e^2$ can be considered known without affecting asymptotic results of the LLWLS estimator of $m(x)$.

In a semiparametric varying coefficient partially linear regression framework for longitudinal data with large $n$ and finite $T$, Fan, Huang, and Li (2007) proposed to estimate the unknown covariance matrices $V_i$’s semiparametrically. Specifically, they estimated $\hat{\sigma}_{i,t}^2 \equiv \sigma^2(X_{i,t}) = E \left( u_{i,t}^2 | X_{i,t} \right)$ as in (2.7) by the local linear regression approach proposed by Fan and Yao (1998), where $u_{i,t}$ is replaced by $\tilde{u}_{i,t} = Y_{i,t} - \tilde{m}(X_{i,t})$ with $\tilde{m}(\cdot)$ being the LLLS estimator given above. Secondly, assuming the working correlation matrix $R$ is known up to a finite number of unknown parameters, i.e., $R = R(\theta)$; e.g., $\theta$ can be parameters appearing in a stationary ARMA($p,q$) model of $u_{i,t}$ with $t = 1, ..., T$. Let $\Gamma (\hat{A}, \theta, x)$ be the estimated variance of smoothing spline or penalized regression spline estimator, $\hat{m}(x)$, where $\hat{A}_i = diag(\hat{\sigma}_{i,1}^2, ..., \hat{\sigma}_{i,T}^2)$. Then, in the spirit of Fan et al. (2007), we can estimate $\theta$ by $\hat{\theta} = \text{arg min}_\theta \int \Gamma (\hat{A}, \theta, x) w(x) \, dx$, where $w(x)$ is a weighting function. Let $\tilde{V} = diag \left( \tilde{V}_1, ..., \tilde{V}_n \right)$ and $\hat{V}_i = \tilde{A}_i^{-1/2} R(\hat{\theta}) \tilde{A}_i^{1/2}$. If the parametric assumption on the working correlation matrix $R$ is a good approximation of the true correlation structure, the spline estimators augmented with the estimated covariance $\tilde{V}$ is expected to provide satisfactory results.

Alternatively, Qu and Li (2006) suggested to approximate $R^{-1}$ by a linear combination of basis matrices as in Qu, Lindsay and Li (2000). As $V_i^{-1} = A_i^{-1/2} R A_i^{-1/2}$, they approximate $R^{-1} \approx a_0 I_T + a_1 M_1 + ... + a_r M_r$, where $M_j$’s are symmetric matrices for $j = 1, ..., r \ll T$, and $a_j$’s are unknown constants. Approximating $m(x)$ by series methods first gives $m(x) \approx \sum_{j=0}^p \delta_j P_j(x)$, where $P_j(x)$ are a set of basis functions of a functional space to which $m(x)$ is assumed to belong; e.g., power functions, truncated power splines, B-splines, etc. Then, (2.3) can be expressed as a linear combination of following $(r + 1)$ functions

$$
\overline{g}_n(\delta) \equiv \frac{1}{n} \sum_{i=1}^n g_i(\delta) = \left( \begin{array}{c} n^{-1} \sum_{i=1}^n P^T(X_i) A_i^{-1} \left( Y_i - \sum_{j=0}^p \delta_j P_j(X_i) \right) \\ n^{-1} \sum_{i=1}^n P^T(X_i) A_i^{-1/2} M_1 A_i^{-1/2} \left( Y_i - \sum_{j=0}^p \delta_j P_j(X_i) \right) \\ \vdots \\ n^{-1} \sum_{i=1}^n P^T(X_i) A_i^{-1/2} M_r A_i^{-1/2} \left( Y_i - \sum_{j=0}^p \delta_j P_j(X_i) \right) \end{array} \right)
$$

where $P(X_i) = [P_1(X_i), ..., P_{p+1}(X_i)]$ is a $T$ by $(p + 1)$ matrix and $P_j(X_i) = [P_j(X_{i,1}), ..., P_j(X_{i,T})]^T$ is a $T \times 1$ vector. The model is then estimated by the penalized generalized method of moments.
(PGMM) method

\[ \hat{\delta} = \arg \min_\delta \hat{g}_n(\delta) \Omega^{-1} \hat{g}_n(\delta) + \lambda \delta^T \delta \]  

(2.9)

where \( \Omega = Var(g_i) \) and \( \lambda \geq 0 \) is a regularization parameter that shrinks \( \delta \) towards zero. The purpose of introducing the penalty term in the objective function is to avoid over-parameterization when approximating \( m(x) \) by series methods. It is possible that one applies other regularization methods; e.g., Zou’s (2006) adaptive LASSO penalty and Fan and Li’s (2001) smoothly slipped absolute deviation (SCAD) penalty. Replacing unknown matrices \( A_i \)'s by the nonparametric kernel estimators explained above and estimating the unknown matrix \( \Omega \) by \( \hat{\Omega} = n^{-1} \sum_{i=1}^{n} g_i(\hat{\delta}) g_i^T(\hat{\delta}) \) with \( \hat{\delta} \) being the estimator obtained under working independence, we obtain another estimator of \( m(\cdot) \) whose limiting result is to be derived as our future interest. For the possible choice of the basis matrices, \( M_j \)'s, the readers are referred to Qu and Li (2006, p.382).

The two methods mentioned above hold for finite \( T \) and sufficiently large \( n \). If one allows \( T \) to increase as \( n \) increases, Bickel and Levina’s (2008) covariance regularization method may be used to estimate the true covariance matrix, \( V \) under the condition that \( \ln(T)/n \rightarrow 0 \) as \( n \rightarrow \infty \). To the best of our knowledge, the nonparametric random-effects panel data models are usually studied with a finite \( T \) and more research needs to be done for constructing efficient nonparametric estimator of \( m(x) \) when both \( T \) and \( n \) are large.

All papers discussed above either assume a known or pre-determined or pre-estimated within-group correlation/covariance structure before applying a two-step procedure to estimate \( m(x) \). We conclude this section with a recent paper that estimates \( m(x) \) and the within-group correlation matrix simultaneously.

Yao and Li (2013) proposed a novel estimator that uses Cholesky decomposition and profile least squares techniques. Consider again model (1.1). Let \( u_i = (u_{i,1}, \ldots, u_{i,T})^T, i = 1, \ldots, n \). Suppose that \( \text{Cov}(u_i|X_i) = V \). A Cholesky decomposition of the within-group covariance gives

\[ \text{Cov}(\Phi u_i) = \Phi V \Phi^T = D, \]

where \( \Phi \) is a lower triangle matrix with 1’s on the main diagonal and \( D = \text{diag}(d_1^2, \ldots, d_T^2) \) is a diagonal matrix. Let \( \phi_{t,s} \) be the negative of the \((t, s)\)-element of \( \Phi \) and \( e_i = (e_{i,1}, \ldots, e_{i,T})^T = \Phi u_i \). One can rewrite \( u_i \) as follows:

\[ u_{i,1} = e_{i,1}, \]
\[ u_{i,t} = \phi_{t,1} u_{i,1} + \cdots + \phi_{t,t-1} u_{i,t-1} + e_{i,t}, t = 2, \ldots, T. \]

Since \( D \) is a diagonal matrix, \( e_{i,t} \)'s are uncorrelated with \( \text{Var}(e_{i,t}) = d_t^2, t = 1, \ldots, T. \) If \( \{u_1, \ldots, u_n\} \) were known, we would proceed with the following partially linear model with uncorrelated error term \( e_{i,t} \):

\[ Y_{i,1} = m(X_{i,1}) + e_{i,1}, \]
\[ Y_{i,t} = m(X_{i,t}) + \phi_{t,1} u_{i,1} + \cdots + \phi_{t,t-1} u_{i,t-1} + e_{i,t}, i = 1, \ldots, n, t = 2, \ldots, T. \]  

(2.10)
where $S$ is the original model to the (2.12), wherein the errors are uncorrelated. Yao and Li (2013) proposed an estimate of $h$ as in (2.2) with $K$ the working independence assumption.

The profile least squares estimate for $\beta$ is then given by

$$Y_i \approx m(X_i) + \hat{F}\phi + e,$$

where $\hat{F} = (\hat{F}_{1,2}, \ldots, \hat{F}_{1,T}, \ldots, \hat{F}_{n,T})^T$. Define $Y^* = Y - \hat{F}\phi$. Then

$$Y^* \approx m(X) + e. \quad (2.12)$$

Therefore if $V$ and thus $\Phi$ were known, we could use Cholesky decomposition to transform the original model to the (2.12), wherein the errors are uncorrelated. Yao and Li (2013) proposed a profile least squares estimator for model (2.12). Define

$$W_d(x) = \text{diag}\{K_h(X_{1,2} - x)/d_2^2, \ldots, K_h(X_{1,T} - x)/d_T^2, \ldots, K_h(X_{n,T} - x)/d_T^2\},$$

where $K_h(t) = h^{-1}K(t/h)$, $K(\cdot)$ is a kernel function and $h$ is the bandwidth, and $d_T^2$ is a consistent estimate of $d_T^2$. It follows that a weighted (by $d_T^2$’s) local linear estimator of $m(\cdot)$ can be constructed as in (2.2) with $W$ replaced by $W_d$, and the estimator is given by

$$\hat{m}(x) = (1,0)(Z^T(x)W_d(x)Z(x))^{-1}Z^T(x)W_d(x)Y^* \equiv S_h(x)Y^*$$

where $Z(x)$ is defined below Eq. (2.2). Thus $\hat{m}(X)$ can be represented by $\hat{m}(X) = S_h(X)Y^*$, where $S_h(X)$ is a $n(T - 1) \times n(T - 1)$ smoothing matrix, depending only on $X$ and the bandwidth $h$. Replacing $m(X)$ in model (2.12) with $\hat{m}(X)$ then yields the following linear model

$$([I - S_h(X)]Y \approx ([I - S_h(X)]\hat{F}\phi + e,$$

where $I$ is the identity matrix. Denote $M_h(X) = I - S_h(X)$ and $\hat{G} = \text{diag}(\hat{d}_2^2, \ldots, \hat{d}_T^2, \ldots, \hat{d}_2^2, \ldots, \hat{d}_T^2)$. The profile least squares estimate for $\phi$ is then given by

$$\hat{\phi} = \left[\hat{F}^TM_h^{-1}(X)\hat{G}^{-1}M_h(X)\hat{F}\right]^{-1}\hat{F}^TM_h^{-1}(X)\hat{G}^{-1}M_h(X)Y. \quad (2.13)$$

Next, letting $\hat{Y}^* = Y - \hat{F}\hat{\phi}$, we have

$$\hat{Y}^* \approx m(X) + e. \quad (2.14)$$

where $e_{i,t}$’s are uncorrelated. We can then estimate (2.14) using the conventional local linear estimator

$$\hat{\beta} = \arg \min_{\beta} \left[\hat{Y}^* - Z(x)\beta\right]^TW_d(x)\left[\hat{Y}^* - Z(x)\beta\right]. \quad (2.15)$$
The local linear estimator of \( m(x) \) is given by \( \hat{m}(x; \hat{\theta}) = \hat{\beta}_0 \).

Yao and Li (2013) established the asymptotic properties of their estimator. In particular, they showed that it is asymptotically efficient as if the true within-group covariance were known. Their Monte Carlo simulations demonstrate that this estimator outperforms the naive local linear estimator with working independence assumption and some existing two-step procedures.

2.2 Fixed effects panel data models

This subsection considers nonparametric estimation of fixed-effects panel data models, where \( X_{i,t} \) in model (1.1) is allowed to be correlated with the error term \( u_{i,t}, i = 1, 2, ..., n; t = 1, ..., T \). Specifically, we consider case (a), where \( u_{i,t} = \epsilon_{i,t} \) and the unobserved individual effects, \( \mu_i \sim i.i.d. (0, \sigma^2) \) are allowed to be correlated with \( \{X_{i,t}\}_{t=1}^T \) in an unspecified way. We consider two cases: (i) \( \{\epsilon_{i,t}\} \) is independent of \( \{X_{i,t}\} \), which excludes the case that \( X_{i,t} \) contains lagged dependent variable(s); (ii) \( X_{i,t} \) includes the lagged dependent variable and \( E(\epsilon_{i,t}|X_{i,t}, X_{i,t-1}, ...) = 0 \).

We consider case (i) first. For large \( n \) and finite \( T > 2 \), Henderson, Carroll and Li (2008) applied a first-differencing method and backfitting algorithm to estimate the unknown function, \( m(\cdot) \). Taking the first difference removes the unobserved fixed effects \( \mu_i \), which gives

\[
Y_{i,t} - Y_{i,1} = m(X_{i,t}) - m(X_{i,1}) + v_{i,t} \tag{2.16}
\]

where \( v_{i,t} = \epsilon_{i,t} - \epsilon_{i,1} \), \( V_t = E(v_{i,t}v_{i,1}^T|X_{i,1}, ..., X_{i,T}) = \sigma^2 (I_{T-1} + \frac{I_T}{T}) + \sigma^2 \), and \( v_i = (v_{i,1}, ..., v_{i,T})^T \). Evidently, the first-differencing method cancels out any time-invariant component in \( m(\cdot) \). As in model (1.1), \( E[m(X_{i,t})] = E(Y_{i,t}) \) and \( m(\cdot) \) is identified if it does not contain time-invariant component other than the non-zero constant mean.

Applying Lin and Carroll’s (2006) profile likelihood method, Henderson, Carroll and Li (2008) constructed a local linear estimator with the backfitting algorithm. Due to the complexity of the backfitting algorithm, they conjectured the asymptotic bias and variance term of the proposed estimator without detailed proofs. As model (2.16) is a nonparametric additive panel data model, Mammen, Stove, and Tjostheim’s (2009) iterative smooth backfitting algorithm is applicable; however, Mammen et al. (2009) did not derive the limit distribution of their proposed estimator, either. Qian and Wang (2012) also considered estimating \( m(\cdot) \) with model (2.16) but used marginal integration method of Linton and Nielsen (1995). They derived the consistency and asymptotic normality results of the proposed estimator for finite \( T \) and large \( n \) not only for cross-sectional fixed-effects models but also for panel data models with both unobserved cross-sectional and time fixed effects.

Alternatively, Su and Ullah (2006), Sun, Carroll and Li (2009) and Li, Lin and Sun (2013) estimated model (1.1) directly, treating unobserved fixed effects \( \mu_i \) as parameters. These estimation methods can be considered as a nonparametric version of least squares dummy variable method. Su and Ullah (2006) assumed \( \sum_{i=1}^n \mu_i = 0 \) for a finite \( T \) and large \( n \). Li, Lin and Sun (2013) showed that this assumption can be dropped for large \( T = O(n^{1/4}) \) and large \( n \). Specifically, introducing \( T \) by \( n \) dummy variables, \( D \), as in parametric fixed-effects panel data models, one considers the
following objective function

\[
\min_{\mu, m(\cdot)} (Y - m(X) - D\mu)^T K_h (x) (Y - m(X) - D\mu).
\]

Assuming \( m(\cdot) \) is known and imposing \( \sum_{i=1}^{n} \mu_i = 0 \), Su and Ullah (2006) obtained \( \hat{\mu}_i(X_{i,t}) \) when \( x = X_{i,t} \) and set \( \hat{Y}_{i,t} = Y_{i,t} - \hat{\mu}_i(X_{i,t}) \) for all \( i \) and \( t \). They then estimated \( m(x) \) by local linear regression approach with the new dependent variable \( \hat{Y}_{i,t} \). Instead, Li, Lin and Sun (2013) removed the impact of \( \mu_i \)'s asymptotically. Although the estimator can be obtained through local polynomial approach, we present here a local constant estimator for simplicity:

\[
\hat{m}(x) = n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \omega_{i,t} Y_{i,t}
\]

(2.17)

where \( \omega_{i,t} = \lambda_{i,t}/\sum_{t=1}^{T} \lambda_{i,t} \) and \( \lambda_{i,t} = K((X_{i,t} - x)/h) \). By construction, \( \hat{m}(x) \) is an average of the dependent variable with a within-group weight \( n^{-1} \omega_{i,t} \) attached to \( Y_{i,t} \) for a given \( i \). Assuming that \( \{ (X_{i,t}, \epsilon_{i,t}) \} \) is strictly stationary with some mixing properties and other regularity conditions, if \( h \to 0, Th \to \infty \) as \( T \to \infty \) and \( nTh^5 = O(1) \) as \( n \to \infty \) and \( T \to \infty \), Li, Lin and Sun (2012, Th.1) showed that at an interior point \( x \),

\[
\sqrt{nTh} \left[ \hat{m}(x) - m(x) - h^2 B(x) - \bar{\mu} \right] \xrightarrow{d} N \left( 0, \frac{\zeta_0 \sigma^2_{\epsilon}}{f(x)} \right),
\]

(2.18)

where \( B(x) = \kappa_2 \left[ 2m' (f' (x) / f (x)) + m^{(2)} (x) \right], \kappa_2 = \int K(v)v^2 dv, \zeta_0 = \int K(v)^2 dv \) and \( \bar{\mu} = n^{-1} \sum_{i=1}^{n} \mu_i \), and \( f(x) \) is the common Lebesgue marginal p.d.f. of \( X_{i,t} \). This result implies that \( \hat{m}(x) = m(x) + h^2 B(x) + \bar{\mu} + O_p \left( nTh^{-1/2} \right) \xrightarrow{p} m(x) \) since \( \bar{\mu} \xrightarrow{p} 0 \) and \( h \to 0 \) as \( n \to \infty \). In addition, the optimal bandwidth \( h_{opt} = O \left( nT^{-1/5} \right) \) with \( T = O \left( n^{1/4} \right) \) allows \( T \) to grow at a slower speed than \( n \).

When the sample size \( n \) is small and/or \( \text{Var}(\mu_i) \) is rather large compared to the variances of \( \epsilon_{it} \) and \( X_{it} \), the existence of the fixed effects term \( \bar{\mu} = O_p \left( n^{-1/2} \right) \) may affect the finite sample performance of the proposed estimator. However, Li et al. (2013, Th.2) found that \( \hat{m}(x) \equiv \hat{m}(x) - \bar{Y} \) is robust to the presence of \( \mu_i \)'s, where \( \bar{Y} = (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} Y_{i,t} \). Since \( E(\bar{Y}_{i,t}) = E[m(X_{i,t})] \equiv c_0 \) with \( E(\mu_i) = E(\epsilon_{i,t}) = 0 \) for all \( i \) and \( t \), they obtained

\[
\sqrt{nTh} \left[ \hat{m}(x) - (m(x) - c_0) - h^2 B(x) \right] \xrightarrow{d} N \left( 0, \frac{\zeta_0 \sigma^2_{\epsilon}}{f(x)} \right),
\]

(2.19)

so that \( \hat{m}(x) \) is a consistent estimator of the demeaned curve \( m(x) - c_0 \).

Next we consider case (ii) where the lagged dependent variable \( Y_{i,t-1} \) is included in the explanatory variables. Again, the first-differencing method is conventionally used to completely remove the impact of the individual effects, which gives

\[
\Delta Y_{i,t} = m(Y_{i,t-1}, X_{i,t}) - m(Y_{i,t-2}, X_{i,t-1}) + v_{i,t},
\]

(2.20)

where \( v_{i,t} \) is defined below eq. (2.16). Since the error terms under this framework are generally correlated with \( Y_{i,t-1} \), traditional kernel methods based on the marginal integration or backfitting
algorithm do not yield a consistent estimator of $m(\cdot)$. Lee (2010) considered the case where $X_{i,t} \equiv Y_{i,t-1}$ and proposed a sieve estimation of the dynamic panel model via within-group transformation. He found that the series estimator is asymptotically biased and proposed a bias-corrected series estimator, which is shown to be asymptotically normal.

An alternative approach is employed by Su and Lu (2013). Their method is motivated by two observations: (i) the two additive terms of (2.20) share the same functional form; (ii) $E[v_{i,t} | Y_{i,t-2}, X_{i,t-1}] = 0$ by assuming $E(\epsilon_{i,t} | X_{i,t}, Y_{i,t-1}, Y_{i,t-2}, X_{i,t-2}, \ldots) = 0$ holds for all $i$ and $t$. Denote $U_{i,t-2} = (Y_{i,t-2}, X_{i,t-1})^T$. It follows that

$$E[\Delta Y_{i,t} - m(Y_{i,t-1}, X_{i,t}) + m(Y_{i,t-2}, X_{i,t-1}) | U_{i,t-2}] = 0.$$  

(2.21)

Denote by $f_{t-2}(u)$ the p.d.f. of $U_{i,t-2}$ and $f_{t-1|t-2}$ the conditional p.d.f. of $U_{i,t-1}$ given $U_{i,t-2}$. Also define $r(u) \equiv -E(\Delta Y_{i,t} | U_{i,t-2} = u)$ and $f(v|u) = f_{t-1|t-2}(v|u)$. One can rewrite (2.21) as follows:

$$m(u) = r(u) + \int m(v)f(v|u)dv.$$  

(2.22)

Thus the parameter of interest, $m$, is implicitly defined as a solution to a Fredholm integral equation of the second kind.

Let $L_2(f)$ be an infinite dimensional Hilbert space. Under certain regularity conditions, we have

$$m = r + Am,$$  

(2.23)

where $A : L_2(f) \rightarrow L_2(f)$ is a bounded linear operator defined by

$$Am(u) = \int m(v)f(v|u)dv.$$  

Suppose $A$ is well behaved, the solution to the above Fredholm integral equation is given by $m = (I - A)^{-1}r = \sum_{j=0}^{\infty} A^j r$, where $I$ is the identity operator. One can then use a successive substitution approach to solve for $m$ numerically. In this case, the sequence of approximations

$$m^{(l)} = r + A^{(l-1)} m, \quad l = 1, 2, 3, \ldots,$$  

is close to the truth from any starting point $m^0$.

Let $\hat{r}$ and $\hat{A}$ be nonparametric estimators of $r$ and $A$. The plug-in estimator $\hat{m}$ is then given by the solution of

$$\hat{m} = \hat{r} + \hat{A}\hat{m}.$$  

Su and Lu (2013) proceeded with an iterative estimator of the parameter of interest $m$. The following steps were proposed:

1. Calculate a consistent initial estimate of $m$, say, $\hat{m}^{(0)}$.

2. Calculate a consistent estimate of $r(u)$, say, $\hat{r}$ by regressing $-\Delta Y_{i,t}$ on $U_{i,t-2}$. (parametric or nonparametric regression?)
3. Calculate a consistent estimate of $A\hat{m}^{(0)}$, say, $\hat{A}\hat{m}^{(0)}$ by regressing $\hat{m}^{(0)}(U_{i,t-1})$ on $U_{i,t-2}$. (parametric or nonparametric regression?)

4. Calculate $\hat{m}^{(1)} = \hat{r} + \hat{A}\hat{m}^{(0)}$.

5. Repeat steps 3-4 until convergence.

Below we provide a brief description of the estimators used in the steps outlined above. As for the initial estimate $\hat{m}^{(0)}$, Su and Lu (2013) used a sieve estimator. Let $\hat{q}_{i;t}^{(1)}$ be a $L$-dimensional vector of sieve basis functions and $\hat{q}_{i;t}^{(2)}$, $\hat{q}_{i;t}^{(3)}$. The estimator takes the form

$$Y_{i,t} = \sum_{i=1}^{n} \sum_{t=2}^{T} [Y_{i,t} - \hat{m}^{(l)}(U_{i,t-1})],$$

where $\hat{m}^{(l)}(u)$ is used in the next iteration.

Su and Lu (2013) established the consistency, convergence rate and asymptotic normality of the proposed estimator under suitable regularity conditions. They further proposed a specification test for linear dynamic panel models. The proposed test is based on the comparison of the restricted linear estimator and the unrestricted estimator. They considered the following smooth functional

$$\Gamma = \int [m(u) - \gamma_0 - \beta_0^T u]^2 a(u)f(u)du,$$

where $a(u)$ is a user-specified nonnegative weighting function defined on the compact support of $u$. A feasible test statistic is then constructed as

$$\Gamma_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} [\hat{m}(U_{i,t-1}) - \hat{\gamma} - \hat{\beta}^T U_{i,t-1}]^2 a(U_{i,t-1}).$$

They showed that after appropriate centering and scaling, $\Gamma_{NT}$ is asymptotically normally distributed under some suitable assumptions under the null hypothesis.

2.3 Panel data models with cross-sectional dependence

In this subsection we discuss models with cross-sectional dependence. Consider a nonparametric panel data model

$$Y_{i,t} = m(X_{i,t}) + u_{i,t},$$

$$u_{i,t} = \mu_i + \delta_i^T g_t + \epsilon_{i,t}, i = 1, ..., n, t = 1, ..., T,$$
where $g_t$ is a $d$ by 1 vector of unobserved common factors, $\delta_i \neq 0$ is a $d$ by 1 vector of the factor loadings, and $\epsilon_{i,t}$ is i.i.d. and independent of $\delta_i^T g_s$ for all $i, j, s,$ and $t$. As $E(u_{i,t}u_{j,t}) = \delta_i^T E(g_tg_t^T)\delta_j$ can be none-zero if $\delta_i \neq 0$ and $\delta_j \neq 0$, the error term exhibits cross-sectional dependence due to unobserved factor $\delta_i^T g_t$. This model is a special case studied by Su and Jin (2012) and Huang (2013), where both papers include a linear component in model (2.24). Also, Eq. (2.25) nests fixed-effects and random-effects two-way error component models with $u_{i,t} = \mu_i + \lambda_t + \epsilon_{i,t}$. Both papers extend Pesaran’s (2006) common correlated effects (CCE) estimation method in linear panel data models to nonparametric panel data models with large $n$ and large $T$.

Closely following Pesaran’s (2006) methodology, Su and Jin (2012) estimated $m(\cdot)$ via sieve estimation method. The validity of this method crucially relies on the following assumption

$$ X_{i,t} = \alpha_i + \gamma_i T g_t + v_{i,t}, \quad (2.26) $$

where $\{(v_{i,t}, \epsilon_{i,t})\}$ is a strictly stationary strong mixing sequence and is independent of another strictly stationary strong mixing sequence $\{g_t\}$, and $\{v_{i,t}\}$ is also independent of $\{\epsilon_{i,t}\}$. Let $\bar{w}_t = n^{-1}\sum_{i=1}^n w_{i,t}$ for any $w$. Taking cross-sectional average to (2.24) and (2.26) gives

$$ \begin{pmatrix} \bar{Y}_t \\ \bar{X}_t \end{pmatrix} = \begin{pmatrix} \bar{\mu} \\ \bar{\alpha} \end{pmatrix} + \bar{\delta} g_t + \begin{pmatrix} \bar{m}_t + \bar{\epsilon}_t \\ \bar{v}_t \end{pmatrix}. \quad (2.27) $$

Letting $\bar{\theta} = (\bar{\delta}, \bar{\gamma})^T$, we have

$$ g_t = (\bar{\theta} \bar{\theta}^T)^{-1} \bar{\theta}^T \left[ \begin{pmatrix} \bar{Y}_t \\ \bar{X}_t \end{pmatrix} - \begin{pmatrix} \bar{\mu} \\ \bar{\alpha} \end{pmatrix} - \begin{pmatrix} \bar{m}_t + \bar{\epsilon}_t \\ \bar{v}_t \end{pmatrix} \right] $$

where $\text{rank}(\bar{\theta} \bar{\theta}^T) \leq d + 1$ as $n \to \infty$. If $\sup_t |\bar{w}_t| \overset{p}{\to} 0$ as $n \to \infty$ for $w = m, \epsilon,$ and $v$, then

$$ g_t - (\bar{\theta} \bar{\theta}^T)^{-1} \bar{\theta}^T \left[ \begin{pmatrix} \bar{Y}_t \\ \bar{X}_t \end{pmatrix} - \begin{pmatrix} \bar{\mu} \\ \bar{\alpha} \end{pmatrix} \right] \overset{p}{\to} 0 \quad \text{as} \quad n \to \infty, $$

which holds uniformly over all $t$. Thus $g_t$ can be well approximated by $(1, \bar{X}_t, \bar{Y}_t)^T$. It follows that model (2.24)-(2.25) can be approximated by

$$ Y_{i,t} \approx \pi_{i,0} + \pi_{i,1} \bar{Y}_t + \pi_{i,2} \bar{X}_t + m(X_{i,t}) + \epsilon_{i,t}, \quad (2.28) $$

which is a partially linear regression model. The identification condition is given by $E[m(X_{i,t})] \equiv 0$. Approximating $m(X_{i,t})$ by a linear combination of basis functions, Su and Jin (2012) derived consistency and asymptotic normality results of the proposed estimator of $m(x)$ when both $n$ and $T$ are sufficiently large and $T/n^s \to 0$ for some $s \in (0, 1)$. When $\delta_i = \gamma_i = 0$ for all $i$, model (2.24)-(2.25) becomes a nonparametric fixed-effects panel data considered by Li, Lin and Sun (2013) in Section 2.3, and the argument based on (2.26) fails to be meaningful.

Huang (2013) followed the same methodology in Pesaran (2006) in the sense that the unobserved time-varying factors $g_t$ are to be instrumented with the cross-section averages, $\bar{Y}_t$ and $\bar{X}_t$. Without
imposing (2.26), Huang (2013) showed that \( \delta \delta^T \) having a full rank is sufficient to obtain a consistent estimator of \( m(\cdot) \). For the fixed-effects panel data model (2.24)-(2.25), Huang (2013) suggested to use the within-group averages, \( \tilde{Y}_i \) and \( \tilde{X}_i \) as instrumental variables for \( \mu_i \) for all \( i \), following similar argument as in using \( \tilde{Y}_t \) and \( \tilde{X}_t \) to predict \( g_t \). In addition, assuming \( \delta_i' s \) are random coefficients independent of \( (g_t, X_{i,t}, \epsilon_{i,t}) \) for all \( i \) and \( t \), Huang (2013) simplified model (2.28) to a model with common parameters across index \( i \\
Y_{i,t} \approx \pi_1 \tilde{Y}_t + \pi_2 \tilde{X}_t + \pi_3 \tilde{Y}_i + \pi_4 \tilde{X}_i + m(\mathbf{X}_{i,t}) + \epsilon_{i,t},
(2.29)
and estimated \( m(x) \) from the model above by local linear regression approach. Both (2.28) and (2.29) approximate the unobserved factors \( g_t \) by linear combination of \( \tilde{Y}_t \) and \( \tilde{X}_t \) globally in Su and Jin (2012) and locally in Huang (2013), respectively. Also, both papers need to assume \( \mathbb{E}[m(\mathbf{X}_{i,t})] = 0 \) for the purpose of identification.

For a large \( T \) and finite \( n \), Robinson (2012) considered the following model
\[
Y_{i,t} = \alpha_i + \beta_t + u_{i,t}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T
(2.30)
\]
where \( \alpha_i 's \) and \( \beta_t 's \) are unknown individual and time effects, \( u_{i,t} \) are unobservable zero-mean random variables that exhibit cross-sectional dependence but cross-time independence. Under the assumption that \( \sum_{i=1}^{n} \alpha_i = 0 \), taking average across units yields
\[
\bar{Y}_t = \beta_t + \bar{u}_t, \quad t = 1, \ldots, T,
(2.31)
\]
where \( \bar{w}_t = n^{-1} \sum_{i=1}^{n} w_{i,t} \) for \( w = Y \) or \( u \). It is assumed that \( \mathbb{E}(u_{i,t}) = 0 \), \( \mathbb{E}(u_{i,t}u_{j,t}) = \omega_{ij} \) and \( \mathbb{E}(u_{i,t}u_{j,s}) = 0 \) for all \( i, j, t, s \). One can then estimate \( \beta_t \) by \( \bar{Y}_t \). However, with small \( n \), the estimated time trend at any given point of time may not be consistent.

Robinson (2012) proposed to estimate \( \beta_t \) as a smooth function of \( t \) using a kernel estimator, effectively borrowing information from neighboring time periods. Let \( \beta_t = \beta(t/T), \quad t = 1, 2, \ldots, T. \) Define
\[
K_{t,\tau} = K\left(\frac{T \tau - t}{Th}\right) \quad \text{for } \tau \in (0, 1),
\]
where \( K(\cdot) \) is a kernel function. The smooth time trend estimator is then given by
\[
\bar{\beta}(\tau) = \frac{\sum_{t=1}^{T} K_{t,\tau} \bar{Y}_t}{\sum_{t=1}^{T} K_{t,\tau}}.
\]
Under some regularity conditions, he showed that the kernel estimator of \( \bar{\beta}(\tau) \) for \( \tau \in (0, 1) \) from model (2.31) is consistent if \( h \to 0 \) and \( Th \to \infty \) as \( T \to \infty \). Let \( \Omega \) be the \( n \times n \) cross-sectional covariance matrix with the \((i,j)\)-th element being \( \omega_{i,j} \). The mean squared error of \( \bar{\beta}(\tau) \) is given by
\[
\text{MSE}(\bar{\beta}(\tau)) \sim \frac{\zeta_0}{Th} \frac{\epsilon_n^* \Omega_n \epsilon_n}{n^2} + \frac{\kappa_2 h^4 [\beta''(\tau)]^2}{4}, \quad \text{as } T \to \infty,
(2.32)
\]
where \( \epsilon_n \) is a \( n \times 1 \) vector of 1’s, \( \zeta_0 = \int K^2(u)du \), and \( \kappa_2 = \int u^2 K(u)du \).
The smooth time trend estimator, $\tilde{\beta}(\tau)$, has the standard convergence rate of kernel estimators. However, it only utilizes the average of each time period. Robinson (2012) then proceeded to show that $\tilde{\beta}(\tau)$ is dominated by an improved estimator that takes advantage of cross-sectional data, which offers the possibility of variance reduction. Let $w = (w_1, \ldots, w_n)^T$ be a vector of weights with $w^T 1_n = 1$ and $Y_t = (Y_{1,t}, \ldots, Y_{n,t})^T$. A weighted kernel estimator of the time trend is constructed as

$$
\tilde{\beta}^{(w)}(\tau) = \frac{\sum_{t=1}^{T} K_{t,\tau} w^T Y_t}{\sum_{t=1}^{T} K_{t,\tau}}, \quad \tau \in (0, 1).
$$

Robinson (2012) showed that the weight $w$ affects the variance of $\tilde{\beta}^{(w)}(\tau)$ but not its bias and that the corresponding mean squared error is given by

$$
\text{MSE}(\tilde{\beta}^{(w)}(\tau)) \sim \frac{c_0}{h} w^T \Omega w + \frac{\kappa_2 h^4 [\beta''(\tau)]^2}{4}, \quad \text{as } T \to \infty. \tag{2.33}
$$

Minimizing (2.33) with respect to $w$ subject to $w^T 1_n = 1$ then yields the optimal weight

$$
w^* = \frac{\Omega^{-1} 1_n}{1_n^T \Omega 1_n}.
$$

It follows that

$$
\text{MSE}(\tilde{\beta}^{(w^*)}(\tau)) \sim \frac{c_0}{h} (1_n^T \Omega^{-1} 1_n)^{-1} + \frac{\kappa_2 h^4 [\beta''(\tau)]^2}{4}, \quad \text{as } T \to \infty. \tag{2.34}
$$

Thus $\tilde{\beta}^{(w^*)}$ generally dominates $\tilde{\beta}$. As expected, the cross-sectional variation of the panel data provides useful information to improve the estimation of time trend in terms of efficiency. Robinson (2012) further derived the optimal bandwidth of the proposed estimator and reported Monte Carlo simulations to confirm its improved performance. Lee and Robinson (2012) extended Robinson (2012) to allow for stochastic covariates, conditional heteroskedasticity and intertemporal dependence.

Chen et al. (2012a) considered a semiparametric partially linear panel model with cross-sectional dependence,

$$
Y_{i,t} = X_{i,t}^T \beta + f_t + \mu_i + \epsilon_{i,t}, \quad (2.35)
$$

$$
X_{i,t} = g_t + x_i + v_{i,t}, \quad (2.36)
$$

where $X_{i,t}$ is a $d$-dimensional covariates that are allowed to be nonstationary with a trending component and correlated with $\mu_i$ and $\epsilon_{i,t}$, and $f_t \equiv f(t/T)$ is an unknown smooth measurable function of a time trend. For the purpose of identification, they assumed $\sum_{i=1}^{n} \mu_i = 0$ and $\sum_{i=1}^{n} x_i = 0$. They proposed to estimate (2.35) via a semiparametric profile least squares method, wherein the individual effects are treated as unknown parameters and the time trend is approximated by local linear estimator.
Define the following quantities:

\[ Y = (Y_{i,1}, \ldots, Y_{i,T}, Y_{2,1}, \ldots, Y_{2,T}, \ldots, Y_{n,1}, \ldots, Y_{n,T})^T, \]
\[ X = (X_{i,1}, \ldots, X_{i,T}, X_{2,1}, \ldots, X_{2,T}, \ldots, X_{n,1}, \ldots, X_{n,T})^T, \]
\[ \mu = (\mu_2, \ldots, \mu_n)^T, \quad D = (-t_{n-1}, I_{n-1})^T \otimes \nu_T, \]
\[ f = t_n \otimes (f_1, \ldots, f_T)^T, \]
\[ \epsilon = (\epsilon_{i,1}, \ldots, \epsilon_{i,T}, \epsilon_{2,1}, \ldots, \epsilon_{2,T}, \ldots, \epsilon_{n,1}, \ldots, \epsilon_{n,T})^T, \]

where \( I_n \) is the \( n \times n \) identity matrix and \( \nu_T \) is a \( T \times 1 \) vector of ones. Model (2.35) can then be rewritten as

\[ Y = X\beta + f + D\mu + \epsilon. \]

Denote

\[ z(\tau) = \begin{pmatrix} 1 & \frac{1 - \tau T}{Th} \\ \vdots & \vdots \\ 1 & \frac{T - \tau T}{Th} \end{pmatrix}, \tau \in (0, 1), \]

and \( Z(\tau) = \nu_n \otimes z(\tau) \). Then by Taylor expansion we have \( f(\tau) \approx Z(\tau)[f(\tau), h f'(\tau)]^T. \)

Let \( K(\cdot) \) be a kernel function and \( h \) be a bandwidth. Define \( w(\tau) = \text{diag}(K(\frac{1-\tau T}{Th}), \ldots, K(\frac{T-\tau T}{Th})) \)

and \( W(\tau) = \nu_n \otimes w(\tau) \). A pooled local linear estimator is then defined by the minimization of the following objective function

\[ (Y - X\beta - D\mu - Z(\tau)(a,b)^T)^T W(\tau) [Y - X\beta - D\mu - Z(\tau)(a,b)^T]. \]

Chen et al. (2012a) proceeded by first concentrating out the time trend. That is, for given \( \mu \) and \( \beta \), they estimated \( f(\tau) \) and \( f'(\tau) \) by

\[(\hat{f}_{\mu,\beta}(\tau), \hat{f}'_{\mu,\beta}(\tau)) = \arg \min_{a,b} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[ Y_{i,t} - X_{i,t}^T \beta - \mu_i - a - b \left( \frac{t}{T} - \tau \right) \right]^2 K \left( \frac{t - \tau T}{Th} \right). \]

Define \( S(\tau) = [Z^T W(\tau) Z(\tau)]^{-1} Z^T W(\tau) \), one can show that

\[ \hat{f}_{\mu,\beta} = (1,0)S(\tau)(Y - X\beta - D\mu) = s(\tau)(Y - X\beta - D\mu), \] (2.37)

where \( s(\tau) = (1,0)S(\tau) \). And, the profile least square estimator of \( \mu \) and \( \beta \) is then defined by

\[ (\hat{\mu}^T, \hat{\beta}^T)^T = \arg \min_{\mu,\beta} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[ Y_{i,t} - X_{i,t}^T \beta - \mu_i - \hat{f}_{\mu,\beta} \left( \frac{t}{T} \right) \right]^2. \] (2.38)

Further define \( R^* = (I_{nT} - S)R \) for \( R = Y, X \) or \( D \) and \( M^* = I_{nT} - D^*(D^{*T} D^*)^{-1} D^{*T} \). Then, (2.38) is simplified to the usual least squares problem of modified data \((Y^*, X^*, D^*)\), which gives:

\[ \hat{\beta} = (X^{*T} M^* X^{*T})^{-1} X^{*T} M^* Y^*, \]
\[ \hat{\mu} = (D^{*T} D^*)^{-1} D^{*T} (Y^* - X^{*T} \hat{\beta}). \]
Finally plugging $\hat{\beta}$ and $\hat{\mu}$ into (2.37) yields the local linear estimates of the trend function

$$\hat{f}(\tau) = s(\tau)(Y - X\hat{\beta} - D\hat{\mu}).$$

Under fairly general conditions, Chen et al. (2012a) showed that the parametric parameters are asymptotically normally distributed with a $\sqrt{nT}$-convergence rate as the time series length $T$ and cross-sectional size $n$ both tend to infinity, while the nonparametrically estimated trend function converges at a $\sqrt{nT}/nT$-rate. They also suggested that if the cross-sectional averaging approach of Robinson (2012) was employed, the convergence rate of the estimator of the parametric component is $\sqrt{\frac{n}{nT}}$ while that of the nonparametric trend estimator remains the same. These observations were confirmed in their Monte Carlo simulations. Li et al. (2011) generalized Chen et al. (2012a) to time-varying coefficient panel data models with a common time trend and cross-sectional dependence. Chen et al. (2013a, 2013b) considered single index panel models with fixed effects and unknown link functions.

While this survey is written, many new research works in this area have been growing quietly. For example, for large $n$ and finite $T$, Freyberger (2012) considered nonparametric panel data models with interactive fixed effects in the form of $Y_{i,t} = m_t(X_{i,t}, \delta_i g_t + \epsilon_{i,t})$, where $m_t(\cdot, \cdot)$ is an unknown, time-varying, smooth measurable function that is strictly increasing in the second argument, while for large $n$ and large $T$, Su and Zhang (2013) focused on the sieve estimation and specification testing for nonparametric dynamic panel data models with interactive fixed effects. We thank the referee for providing us these two references.

### 3 Conditional Quantile Regression Models

Section 2 focuses on studying conditional mean regression models, which can be used to predict the average response of a dependent variable with respect to changes in covariates. However, with heteroskedastic data, mean regression models disguise heterogeneity and are not robust to tail reactions. In the presence of strong data heteroskedasticity, one may find it attractive to estimate extreme causal relationship of an explanatory variable to a dependent variable, which can be suitably studied in quantile regression framework.

Since the seminal works of Koenker and Bassett (1978, 1982), quantile regression models have gradually appeared on the radar of both econometricians and applied economists in the presence of data heteroskedasticity where quantile estimations across different probability masses provide a better view of causal relationship across a conditional distribution other than the average information revealed by conditional mean regression studies. For example, Yu, Lu, and Stander (2003), Buchinsky (1998) and a special issue of Empirical Economics (2001, Vol. 26) presented empirical applications of parametric quantile regressions to labor, microeconomics, macroeconomics, and finance. Koenker (2005) gave a full length survey on parametric quantile regressions for estimation, tests and computational issues. Early works on nonparametric estimation include Bhattacharya
and Gangopadhyay (1990), Chaudhuri (1991), Fan et al. (1994), Koenker, Ng and Portnoy (1994), Yu and Jones (1998), Cai (2002), Lee (2003), and Sun (2005). Zheng (1998) extended the residual-based nonparametric test to test for parametric linear quantile regression (QR) models against nonparametric QR models. All these research works considered either cross-sectional or time series data. In this section, we review some recent development of quantile regression models for panel data.

Let \( F_{i,t}(y|x) = \Pr(Y_{i,t} \leq y|X_{i,t} = x) \) be the conditional distribution of \( Y_{i,t} \) given \( X_{i,t} = x \). Let \( Q_p(x) \) be the 100th conditional quantile function of \( Y_{i,t} \) given \( X_{i,t} = x \) for \( p \in (0,1) \). Define \( Q_p(x) = \inf\{y : F_{i,t}(y|x) \geq p\} \). When \( F_{i,t}(y|x) \) is absolutely continuous in \( y \) for all \( x \), \( Q_p(x) \) is the unique solution to \( F_{i,t}(y|x) = p \).

This section contains two subsections. Section 3.1 considers nonparametric pooled panel data quantile regression (QR) models when \( u_{i,t} = \sigma(X_{i,t})\varepsilon_{i,t} \) and \( \varepsilon_{i,t} \sim i.i.d.(0,1) \) is independent of \( \{X_{i,t}\} \). Section 3.2 considers fixed-effects panel data QR models with \( u_{i,t} = \mu_i + \sigma(X_{i,t})\varepsilon_{i,t} \), where \( \mu_i \sim i.i.d.(0,\sigma^2_\mu) \), and \( \varepsilon_{i,t} \sim i.i.d.(0,1) \) is independent of \( \{X_{i,t}\} \).

### 3.1 Pooled panel data quantile regression models

Consider a nonparametric panel data mean regression model

\[
Y_{i,t} = m(X_{i,t}) + \sigma(X_{i,t})\epsilon_{i,t}, i = 1,...,n, t = 1,...,T,
\]

where \( \epsilon_{i,t} \sim i.i.d.(0,1) \). Suppose that \( \epsilon_{i,t} \) has a common probability function \( f_\epsilon(z) > 0 \) for all \( z \in R \). Let \( \gamma_p \) be the 100th quantile of \( \epsilon_{i,t} \). Then, we can define a 100th nonparametric panel data QR model by

\[
Y_{i,t} = m_p(X_{i,t}) + \sigma(X_{i,t})v_{i,t}
\]

where \( v_{i,t} = \epsilon_{i,t} - \gamma_p \) with \( \Pr(v_{i,t} \leq 0|X_{i,t}) = p \in (0,1) \), and \( m_p(x) \equiv m(x) + \gamma_p\sigma(x) \). If \( \sigma(x) \equiv \sigma_0 \) almost surely for all \( x \in R \), \( m_p(x) \) is parallel to each other for different values of \( p \).

When \( m_p(\cdot) \) is known up to a finite number of unknown parameters, say \( m_p(x) \equiv m_p(x,\theta) \), where \( m_p(\cdot) \) has a known functional form and \( \theta \) is a \( d \times 1 \) unknown parameter vector with a finite integer \( d \geq 1 \), a consistent estimator of \( \theta \) can be obtained from the following optimization problem

\[
\hat{\theta}_p = \arg\min_{\theta \in \Theta} \sum_{i=1}^N \sum_{t=1}^T \omega_{i,t}\rho_p(Y_{i,t} - m_p(X_{i,t},\theta)),
\]

where the first element of \( X_{i,t} \) is 1, \( \Theta \subset R^d \) is a compact subset of \( R^d \), \( \rho_p(u) \) denotes the “check” function \( \rho_p(u) = u[p - I(u < 0)] \), \( I(A) \) is the indicator function equal to 1 if event \( A \) occurs, and \( \omega_{i,t} \) is a weight attached to the \((i,t)\)th observation. The objective function (3.2) gives the following estimating equation

\[
\sum_{i=1}^N \sum_{t=1}^T \omega_{i,t} \frac{\partial m_p(X_{i,t},\theta)}{\partial \theta^i} [p - I(Y_{i,t} \leq m_p(X_{i,t},\theta))] = 0.
\]
For a linear regression case, Eq. (3.3) becomes

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \omega_{i,t} X_{i,t} \xi_{i,t} = 0,$$

where $\xi_{i,t} = p - I(Y_{i,t} \leq m_p(X_{i,t}, \theta))$. If one assumes $\omega_{i,t} \equiv 1$, one ignores possible within-group correlation and cross-sectional heteroskedasticity. Although the resulted estimator under working independence and homoskedasticity is consistent and asymptotically normally distributed if $m_p(x, \theta)$ is correctly specified for $p \in (0, 1)$, with a stationary AR(1) error term, the bootstrap results given in Karlsson (2009) indicated that the quantile estimator with $\omega_{i,t} \equiv 1$ exhibits larger bias than with $\omega_{i,t} \equiv \tilde{\sigma}_{i,t}$, the $i$th group’s sample standard deviation of estimated residuals. Assuming $\Pr(v_{i,t} \leq 0 \text{ and } v_{i,s} \leq 0) = \delta$, a constant for any $t \neq s$, Fu and Wang (2012) obtained an exchangeable covariance matrix for $\xi_i = (\xi_{i1}, ..., \xi_{iT})^T$; i.e.,

$$V_i = \text{Var} (\xi_i | X_i) = p(1-p) [(1-\gamma) I_T + \gamma I_T^T]$$

(3.4)

where $\gamma = \text{Corr} (\xi_{it}, \xi_{is}) = (\delta - p^2) / (p - p^2)$ for any $t \neq s$. Following Jung’s (1996) quasi-likelihood approach, Fu and Wang (2012) then introduced a weighted estimating equation

$$M_p(\theta) = \sum_{i=1}^{n} X_i^T V_i^{-1} \xi_i.$$ 

Setting $M_p(\theta) = 0$ gives a weighted estimator of $\theta$ denoted by $\hat{\theta}_w$, which would be consistent and asymptotically normally distributed if $\gamma$ were known. Fu and Wang (2012) considered the case with large $n$ and finite $T$. Our survey on estimating working correlation matrix in Section 2.1 can be relevant here; e.g., Qu et al. (2000) and Fan et al. (2007). When $p = 0.5$, He, Fu and Fung (2003) compared the estimator with working independence, Jung’s (1996) quasi-likelihood estimator and Huggins’ (1993) robust estimator and found that the weighted estimator under exchangeable covariance assumption significantly more efficient than the estimator under working independence. However, the exchangeable variance structure imposed by Fu and Wang (2012) is rather restrictive for panel data models. For seemingly unrelated quantile regression models with dependent cross-equation errors conditional on regressors, Jun and Pinkse (2009) used k-nearest neighbor methods to estimate the unknown optimal weight appearing in the estimating equation and showed that the resulted estimator is asymptotically efficient in the sense that the proposed estimator achieves corresponding semiparametric efficient bound.\footnote{Zhao (2001) derived an asymptotically efficient estimator for linear median regression model in the presence of unknown heteroskedasticity, using nearest neighbour method. Oberhofer (1982), Weiss (1991), and Koenker and Park (1996) are among earlier contributions in estimating nonlinear regression quantiles. Mukherjee (2000) and Oberhofer and Haupt (2006) derived the limiting results of nonlinear quantile estimator for long range dependent and weakly dependent errors, respectively. Komunjer and Vuong (2010) derived efficient semiparametric estimator for dynamic QR models.}

Following the local linear regression approach of Fan, Hu, and Truong (1994), Sun (2006) estimated the unknown 100th quantile curve $m_p(x)$ at an interior point $x$ as follows:

$$\left(\hat{\theta}_0, \hat{\theta}_1\right) = \arg \min_{\theta \in \Theta} \sum_{i=1}^{n} \sum_{t=1}^{T} \rho_p(Y_{i,t} - \theta_0 - \theta_1 (X_{i,t} - x)) K \left( \frac{X_{i,t} - x}{h} \right)$$

(3.5)
where \( \rho_p(u) \) is “check” function defined above, \( K(\cdot) \) is a second-order kernel function, and \( h \) is the bandwidth parameter. Also, \( \tilde{m}_p(x) = \hat{\theta}_0 \) estimates \( m_p(x) \), and \( \tilde{\theta}_1 \) estimates \( m_p^1(x) \), the first-order derivative of \( m_p(x) \). Under some regularity conditions, Sun (2006, Lemma 1) showed that \( \tilde{m}_p(x) \overset{p}{\rightarrow} m_p(x) \) if \( h \to 0 \) and \( nh \to \infty \) as \( n \to \infty \) for a finite \( T \). She did not derive the asymptotic normality result for the estimator as the primary interest of Sun (2006) is to conduct a nonparametric poolability test (see Section 4).

He, Zhu and Fung (2002) and Wang, Zhu and Zhou (2009) studied partially linear panel data QR models, where the latter also considers semiparametric varying coefficient models. Both papers assume an unknown smooth function of time and take nonparametric panel data QR models as a special case. Applying series approximation method, both papers derived the consistency result of the nonparametric estimator for large \( n \) and finite \( T \). Parallel to our review in Section 2.1, we conjecture that the kernel-based estimator is better calculated with working independence, while the sieve estimator benefits from a properly chosen working dependence matrix as a weighting matrix.

Further, combining information from different quantile regressions, a composite quantile regression (CQR) method is showed to be asymptotically more efficient than the quantile curve estimators for a given probability mass when \( \sigma(x) \equiv \sigma_0 \); see, Koenker (1984), Zou and Yuan (2008) and Kai, Li, and Zou (2010).

### 3.2 Fixed effects panel data quantile regression models

We will show that fixed effects panel data quantile regression (QR) models naturally result from panel data mean regression models with unobserved individual factors,

\[
Y_{i,t} = m(X_{i,t}) + \mu_i + \sigma(X_{i,t}) \epsilon_{i,t}, \ i = 1, \ldots, n, \ t = 1, \ldots, T \tag{3.6}
\]

where \( \mu_i \) is the unobserved factor from \( i \)th unit and is independent of the error term \( \{\epsilon_{i,t}\}_{t=1}^T \), \( \epsilon_{i,t} \sim \text{i.i.d.} \ (0, 1) \) has a common probability function \( f_\epsilon(z) > 0 \) for all \( z \in R \), and \( \{\epsilon_{i,t}\}_{t=1}^T \) is independent of \( \{X_{i,t}\}_{t=1}^T \). Let \( \gamma_p \) be the \( 100p \)th quantile of \( \epsilon_{i,t} \). Then, we obtain a 100pth fixed-effects panel data QR model

\[
Y_{i,t} = m_p(X_{i,t}) + \mu_i + \sigma(X_{i,t}) v_{i,t}, \tag{3.7}
\]

where \( v_{i,t} = \epsilon_{i,t} - \gamma_p \) and \( \Pr(v_{i,t} \leq 0 | X_{i,t}) \equiv p \in (0, 1) \), and \( m_p(x) \equiv m(x) + \gamma_p \sigma(x) \). For individual unit \( i \), the \( 100p \)th regression quantile of \( Y_{i,t} \) given \( X_{i,t} \) and \( \mu_i \) is

\[
m_{p,i}(X_{i,t}) \equiv m_p(X_{i,t}) + \mu_i, \tag{3.8}
\]

where unit \( i \) shares the same individual unobservable factor \( \mu_i \).

Naturally, one attempts to remove the unobserved fixed effects via first-differencing method. However, the first-differencing method does not work for panel data QR models with fixed effects because the quantile operator is nonlinear in nature and does not preserve additive property.
If one applies the first difference to model (3.7), he/she obtains

$$Y_{i,t} - Y_{i,t-1} = m_p(X_{i,t}) - m_p(X_{i,t-1}) + \sigma(X_{i,t})v_{i,t} - \sigma(X_{i,t-1})v_{i,t-1}$$  (3.9)

which cancels out not only $\mu_i$ but also time-invariant terms in $m_p(X_{i,t})$. Letting $\varepsilon_{i,t} = \sigma(X_{i,t})v_{i,t} - \sigma(X_{i,t-1})v_{i,t-1}$, this model looks like a nonparametric additive panel data model. However, it is generally untrue that $E(\varepsilon_{i,t}|X_{i,t}, X_{i,t-1}) = -[\sigma(X_{i,t}) - \sigma(X_{i,t-1})] \gamma_p$ equals a constant or $\Pr(\varepsilon_{i,t} \leq 0|X_{i,t}, X_{i,t-1}) = p$. In addition, it is not clear how to set the identification condition in model (3.9) so that $m_p(x)$ is monotonic in $p \in (0, 1)$; see, Rosen (2012) for identification issue with large $n$ and fixed $T$.

The concern given above also holds for a special case like linear regression models. Assuming $\sigma(X_{i,t}) = X_{i,t}^T\delta$ and letting $\beta_p = \beta + \gamma_p \delta$ and $\varepsilon_{i,t} = (X_{i,t}^T\delta)v_{i,t} - (X_{i,t-1}^T\delta)v_{i,t-1}$, model (3.7) becomes

$$Y_{i,t} = X_{i,t}^T\beta_p + \mu_i + (X_{i,t}^T\delta)v_{i,t},$$  (3.10)

and the first-differencing model becomes $Y_{i,t} - Y_{i,t-1} = (X_{i,t} - X_{i,t-1})^T\beta_p + \varepsilon_{i,t}$, where $E(\varepsilon_{i,t}|X_{i,t}, X_{i,t-1}) = -(X_{i,t} - X_{i,t-1})^T \delta \gamma_p$. Unless $\delta \gamma_p = 0$, one can not identify $\beta_p$ if OLS estimator is applied to the first-differencing model. In addition, the $100p$th quantile of $v_{i,t}$ equal to 0 does not automatically ensure the $100p$th conditional quantile of $\varepsilon_{i,t}$ equal to 0 since the quantile operator is not a linear operator.

Assuming $m(\cdot)$ and $\sigma(\cdot)$ are linear functions, Koenker (2004) considered a linear fixed-effects panel data $100p$th QR model

$$Y_{i,t} = X_{i,t}^T\beta_p + \mu_i + e_{p,i,t},$$  (3.11)

where $\Pr(e_{p,i,t} \leq 0|X_{i,t}) = p \in (0, 1)$. He proposed a penalized quantile regression fixed effects (PQRFE) estimator

$$\left(\hat{\beta}, \{\hat{\mu}_i\}_{i=1}^n\right) = \arg \min_{\beta, \{\mu_i\}_{i=1}^n} \sum_{k=1}^q \sum_{i=1}^n w_k \sum_{t=1}^T \rho_{p_k}(Y_{i,t} - X_{i,t}^T\beta_{p_k} - \mu_i) + \lambda \sum_{i=1}^n |\mu_i|$$  (3.12)

where $\beta = \{\beta_{p_k}\}_{k=1}^q$, $w_k$ is a subjectively-chosen weight assigned to the $k$th quantile and controls the relative influence of the quantiles at probability mass ($p_1, ..., p_k$) on the estimation of the common individual effects, $\mu_i$. The $l_1$ penalty function serves to shrink $\mu_i$'s toward zero and other penalty functions may also be used. $\lambda \geq 0$ is called a tuning parameter or a regularization parameter. When $\lambda \to 0$, $\hat{\mu}$'s are not shrunk and we have a fixed effects model; when $\lambda \to \infty$, $\hat{\mu}$ is close to zero and we have a pooled panel data QR model. Lamarche (2010) proposed to choose $\lambda$ under the assumption that $\mu_i$ and $X_i = (X_{i,1}, ..., X_{i,T})^T$ are independent for all $i$. When setting $\lambda = 0$, both Koenker (2004) and Kato et al. (2012) showed consistency of $\hat{\beta}$ if $n/T^s \to 0$ for some $s > 0$ and $T \to \infty$ as $n \to \infty$. However, Kato et al. (2012) found that the asymptotic normality result of $\hat{\beta}$ requires a stronger condition on $T$ (i.e., $n^2 (\log n)^3 / T \to 0$ as $n \to \infty$ and $T \to \infty$).

Galvao (2011) considered fixed effects dynamic panel data linear QR models,

$$Y_{i,t} = Y_{i,t-1}^T \delta_p + X_{i,t}^T \beta_p + \mu_i + e_{p,i,t},$$  (3.13)
where $X_{i,t}$, $\mu_i$, and $e_{p,i,t}$ under the same assumptions given in model (3.10). For fixed effects dynamic panel data linear mean regression models, the fixed-effects least squared estimator may be biased for a moderate $T$ and its consistency depends critically on the assumptions on initial conditions; see Hsiao (2003), Arellano (2003), Phillips and Sul (2007) for examples. One way to construct a consistent estimator independent of initial conditions is to apply instrumental variables approach; e.g., Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991). Similar arguments also apply to the PQRFE estimator; see Monte Carlo evidence in Galvao (2011). Following the methodology in Chernozhukov and Hansen (2008) and Harding and Lamarche (2009), Galvao (2011) proposed an instrumental variables quantile regression with fixed effects (IVQRFE) estimator

$$
\hat{\delta} = \arg \min_\delta \| \hat{\gamma}(\delta) \|_A
$$

$$
\left( \hat{\beta}(\delta), \{ \hat{\mu}_i(\delta) \}_{i=1}^n, \hat{\gamma}(\delta) \right) = \arg \min_{(\beta, \{ \mu_i \}_{i=1}^n, \gamma, \{ \delta \}_{i=1}^n)} \sum_{k=1}^q w_k \sum_{i=1}^n \sum_{t=1}^T \rho_{p_k} (Y_{i,t} - Y_{i,t-1} \delta_{p_k} - X_{i,t}^T \beta_{p_k} - \mu_i - Z_{i,t}^T \gamma_{p_k})
$$

where $\alpha = \{ \alpha_{pk} \}_{k=1}^q$ for $\alpha = \beta, \delta$ or $\gamma$, $Z_{i,t}$’s are proper instrumental variables correlated with $Y_{i,t-1}$ but uncorrelated with $e_{p,i,t}$ for all $i$ and $t$, and $\| x \|_A = \sqrt{x^T A x}$ and $A$ is a positive definite matrix. Through a grid search over $\delta$, one obtains $\hat{\delta}$ then defines $\hat{\beta} = \hat{\beta}(\hat{\delta})$. If $T \to \infty$ as $n \to \infty$ and $n^a / T \to 0$ for some $a > 0$ and under some other conditions, Galvao (2011) showed consistency and asymptotic normality of the proposed estimator.

Galvao and Montes-Rojas (2010, p.3479) commented that “......Unfortunately, in estimation of dynamic panel data models IVs become less informative when the autoregressive parameter increases toward one, and also when the variability of the FE increases......” Hence, combining Koenker (2004) and Galvao (2011), Galvao and Montes-Rojas (2010) introduced a penalized IVQRFE (IVQRFE) estimator

$$
\hat{\delta}(\lambda) = \arg \min_\delta \| \hat{\gamma}(\delta, \lambda) \|_A
$$

$$
\left( \hat{\beta}(\delta, \lambda), \{ \hat{\mu}_i(\delta, \lambda) \}_{i=1}^n, \hat{\gamma}(\delta, \lambda) \right) = \arg \min_{(\beta, \{ \mu_i \}_{i=1}^n, \gamma, \{ \delta \}_{i=1}^n)} \sum_{k=1}^q w_k \sum_{i=1}^n \sum_{t=1}^T \rho_{p_k} (Y_{i,t} - Y_{i,t-1} \delta_{p_k} - X_{i,t}^T \beta_{p_k} - \mu_i - Z_{i,t}^T \gamma_{p_k}) + \lambda \sum_{i=1}^n | \mu_i |
$$

and the final parameter estimators are $\hat{\delta}(\lambda)$ and $\hat{\beta}(\lambda) = \hat{\beta}(\hat{\delta}(\lambda), \lambda)$. With the same assumption on $\lambda$, $T$ and $n$ and other conditions, Galvao and Montes-Rojas (2010) showed consistency and delivered asymptotic normality result for the IVQRFE estimator. For the choice of $\lambda$ and the comparative performance of the PQRFE, IVQRFE and PIVQRFE estimators, interested readers are referred to their original papers for details.

Alternatively, Canay (2011) proposed a different estimator than Koenker (2004) for model (3.10), basing on the following observation: when both $n \to \infty$ and $T \to \infty$, the least-squared dummy variable approach can be used to estimate both $\beta$ and $\mu_i$’s consistently from the fixed-effects panel data linear mean regression model. Let us call these estimates $\hat{\mu}_i$’s. Canay (2011) then proposed to estimate $\beta_p$ after subtracting $\hat{\mu}_i$ from $Y_{i,t}$ and showed that the estimator of $\beta_p$ is
consistent and asymptotically normally distributed with a convergence rate of \((nT)^{-1/2}\) if \(n/T^s \to 0\) for some \(s \in (1, \infty)\) as \(n \to \infty\) and \(T \to \infty\).

To the best of our knowledge, we did not find published research estimating nonparametric fixed-effects panel data QR models. However, for both large \(T\) and \(n\), we conjecture that Canay’s (2011) methodology is applicable to nonparametric framework through two steps. In the first step, Henderson, Carroll and Li (2008) or Li, Lin and Sun (2013) method can be used to estimate the unobserved fixed effects \(\mu_i\)’s. Naming the estimator \(\hat{\mu}_i\) and we have \(Y_{i,t} - \hat{\mu}_i = m_p(X_{i,t}) + v_{i,t}\) with \(v_{i,t} = \sigma(X_{i,t}) e_{i,t} + \mu_i - \hat{\mu}_i\). Intuitively, \(\sup_i |\mu_i - \hat{\mu}_i| = o_p\left(n^{\xi-1/2}\right)\) holds for some small \(\xi \in (0, 1/2)\).

In the second step, one can estimate \(m_p(x)\) from the pooled model \(Y_{i,t} - \hat{\mu}_i = m_p(X_{i,t}) + v_{i,t}\) as long as one can show that the impact of \(\mu_i - \hat{\mu}_i\) is asymptotically ignorable as both \(n\) and \(T\) grow.

### 3.3 Other issues

Our survey does not include the literature of generalized QR regression models with nonlinearly transformed dependent variable, and we intuitively expect that the methodology discussed above can be extended to the generalized QR models through the so-called ‘equivariance to monotone transformation’ property of quantile operator. Specifically, let \(w(\cdot)\) be a nondecreasing function along the real line, and for any random variable \(Y\), let \(Q^w_p(x)\) and \(Q_p(x)\) be the respective 100\(p\)th conditional quantile of \(w(Y)\) and \(Y\) given \(X = x\). The ‘equivariance to monotone transformations’ property means that

\[
Q^w_p(x) = w\left(Q_p(x)\right) .
\]

If \(w(\cdot)\) is a known monotonic function, then the above-mentioned methods can be used to consistently estimate panel data QR models with nonlinearly transformed dependent variables (e.g., Box-Cox transformed dependent variable; Powell (1986) for censored data, Koenker and Bilias (2001) for duration data). When it is unknown, \(w(\cdot)\) can be estimated consistently by nonparametric methods; see, Mu and Wei (2009).

A word of caution is given here: fitted quantile curves, by both parametric and nonparametric method, can suffer quantile crossing problem; that is, a 100\(p\)th (nonparametric) quantile curve can cross over a 100\(q\)th (nonparametric) quantile curve for some \(p \neq q\). The quantile crossing may result from model misspecification, colinearity or outliers. Neocleous and Portnoy (2008) explained how to correct this problem for linear quantile regression models, while He (1997), Dette and Volgushev (2008), Chernozhukov et al. (2010) and Bondell, Reich and Wang (2010) suggested solutions for nonparametric quantile regression models for independent data.

### 4 Nonseparable Models

All models reviewed in the previous sections assume that the individual effects and idiosyncratic error terms enter the models additively. In this section we review methods on nonseparable non-parametric panel data models that relax this assumption. Depending on the treatment of the error...
term, there are two types of nonseparable models. The first type is the partially separable model

\[ Y_{i,t} = m(X_{i,t}, \mu_i) + \epsilon_{i,t}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \]

where the individual effect \( \mu_i \) enters the conditional mean in an unknown form, while the idiosyncratic error term is assumed to be additive. The second type is the fully nonseparable model

\[ Y_{i,t} = m(X_{i,t}, \mu_i, \epsilon_{i,t}), \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \]

where neither \( \mu_i \) nor \( \epsilon_{i,t} \) is separable from the conditional mean.

Su and Ullah (2011, Section 7) provided an in-depth review of nonseparable nonparametric models, focusing on two papers: Evdokimov (2009) and Altonji and Matzkin (2005). The former considered a partially nonseparable model. Assuming that \( \{X_i, \mu_i, \epsilon_i\} \) are i.i.d. sequence and that the conditional density of \( \epsilon_{i,t} \) given \( (X_i, \mu_i) \) equals that of \( \epsilon_{i,t} \) given \( X_{i,t} \) and other regularity conditions, Evdokimov (2009) proposed a kernel-based conditional deconvolution method to estimate the structural function \( m(x, \mu) \). Altonji and Matzkin (2005) considered the fully nonseparable model. They assumed a conditional density restriction that

\[ f(\mu, \epsilon|X', Z') = f(\mu, \epsilon|X'', Z'') \]

for specific values \((X', Z')\) and \((X'', Z'')\) and a conditional independence condition that

\[ f(\mu, \epsilon|X, Z) = f(\mu, \epsilon|Z). \]

Under these assumptions, they proposed a nonparametric control function estimator to estimate the local average response function

\[ \beta(x) = \int \frac{\partial}{\partial x} m(x, \mu, \epsilon) dF(\mu, \epsilon), \]

where \( F(\mu, \epsilon) \) is the joint distribution of \( \mu \) and \( \epsilon \). Interested readers are referred to the original papers and Su and Ullah (2011) for discussions of relevant work.

In this section, we focus on a recent paper by Chernozhukov et al. (2013) that considers both average and quantile effects in fully nonseparable panel models. A key assumption of this paper is a time homogeneity condition which prescribes the stationarity of the conditional distribution of \( \epsilon_{i,t} \) given \( X_i \) and \( \mu_i \). Formally, this condition is given by

\[ \epsilon_{i,t} \mid (X_i, \mu_i) \overset{d}{=} \epsilon_{i,1} \mid (X_i, \mu_i), \text{ for all } t. \quad (4.1) \]

This condition, as suggested by the authors, "... it is like ‘time is randomly assigned’ or ‘time is an instrument’ with the distribution of factors other than \( x \) not varying over time, so that changes in \( x \) over time can help identify the effect of \( x \) on \( y \)."

The authors focused on two objects of interest, the average structural function (ASF) and the quantile structural function (QSF). The ASF is given by

\[ g(x) = E[m(x, \mu_i, \epsilon_{i,t})] = \int m(x, \mu, \epsilon) dF(\mu, \epsilon). \]
The average treatment effect (ATE), as in the treatment effect literature, of changing \( x \) from \( x^b \) (before) to \( x^a \) (after) is then
\[
\Delta = g(x^a) - g(x^b).
\]
The authors showed that the ATE obtained this way is identical to that of the conditional mean model, which is given by \( \int [m(x^a, \mu) - m(x^b, \mu)]dF(\mu) \).

The QSF, \( Q(p, x) \), is the 100th quantile of \( m(x, \mu_i, \epsilon_{i,t}) \), given by
\[
Q(p, x) = G^{-1}(p, x) \text{ and } G(y, x) = E[I(m(x, \mu_i, \epsilon_{i,t}) \leq y)].
\]
The 100th quantile treatment effect (QTE) of changing \( x \) from \( x^b \) to \( x^a \) is given by
\[
\Delta_p = Q(p, x^a) - Q(p, x^b).
\]
The authors further assumed that the regressors are discrete with finite support. Let \( I(X_{i,t} = x) \) denote the indicator function that is equal to one when \( X_{i,t} = x \) and zero otherwise and \( T_i(x) = \sum_{t=1}^T I(X_{i,t} = x) \). Also let \( D_i = I(T_i(x^a) > 0)I(T_i(x^b) > 0) \) be the indicator that \( X_i \) includes both \( x^a \) and \( x^b \) for some period. The ATE is then defined as
\[
\delta = E[m(x^a, \mu_i, \epsilon_{i,1}) - m(x^b, \mu_i, \epsilon_{i,1})|D_i = 1].
\]
A simple estimator of the conditional ATE \( \delta \) is
\[
\hat{\delta} = \frac{\sum_{i=1}^n D_i[\bar{Y}_i(x^a) - \bar{Y}_i(x^b)]}{\sum_{i=1}^n D_i},
\]
where
\[
\bar{Y}_i(x) = \begin{cases} 
T_i(x)^{-1} \sum_{t=1}^T I(X_{i,t} = x)Y_{i,t}, & T_i(x) > 0; \\
0, & T_i(x) = 0.
\end{cases}
\]
Consistency of this ATE estimator can be established based on
\[
E \left\{ D_i \left[ \bar{Y}_i(x^a) - \bar{Y}_i(x^b) \right] \right\} = E \left\{ D_i \left[ m(x^a, \mu_i, \epsilon_{i,1}) - m(x^b, \mu_i, \epsilon_{i,1}) \right] \right\}.
\]
A consistent estimator of the asymptotic variance of \( \sqrt{n}(\hat{\delta} - \delta) \) is given by \( n^{-1} \sum_{i=1}^n \hat{\psi}_i^2 \) where \( \hat{\psi}_i = nD_i[\bar{Y}_i(x^a) - \bar{Y}_i(x^b) - \delta] / \sum_{i=1}^n D_i \). The authors also showed that the usual panel data within (linear fixed effects) estimator is not a consistent estimator of \( \delta \) because the within estimator constrains the slope coefficients to be the same for each \( i \) when the slope is actually varying with \( i \).

We can identify and estimate the conditional QTE in a similar manner. Let \( G(y, x | D_i = 1) = \Pr(m(x, \mu_i, \epsilon_{i,1}) \leq y | D_i = 1) \) denote the CDF of \( m(x, \mu_i, \epsilon_{i,1}) \) conditional on \( D_i = 1 \). The QTE conditional on \( D_i = 1 \)
\[
\delta_p = G^{-1}(p, x^a | D_i = 1) - G^{-1}(p, x^b | D_i = 1).
\]
To calculate this estimator, we need an estimator of \( G(y, x | D_i = 1) \). Let \( \Phi \) be the standard Gaussian CDF and \( h \) a bandwidth. A smoothed estimator of \( G(y, x | D_i = 1) \) is given by
\[
\bar{G}(p, x | D_i = 1) = \frac{\sum_{i=1}^n D_i \bar{G}_i(y, x)}{\sum_{i=1}^n D_i},
\]
and method to estimate their bounds nonparametrically. Suppose for all continuous regressors if they are properly discretized. The restriction is rather innocuous. We conjecture that the methods can be used for models with discrete regressors. Given their focus on treatment effect estimation, this restriction is rather innocuous. We conjecture that the methods can be used for models with continuous regressors if they are properly discretized.

A consistent estimator of the asymptotic variance of $\sqrt{n}(\hat{\delta}_p - \delta_p)$ is given by $n^{-1} \sum_{i=1}^{n} \hat{\psi}_{p,i}$ where

$$\hat{\psi}_{p,i} = \left( n D_i \sum_{i=1}^{n} \left[ \frac{G_i(q^a_i, x^a_i) - p}{G_i(q^b_i, x^a_i D_i = 1)} - \frac{G_i(q^b_i, x^b_i)}{G_i(q^b_i, x^b_i D_i = 1)} \right] \right)^2$$

and $G'(y, x | D_i = 1) = \partial G(y, x | D_i = 1)/\partial y$.

For $x$’s that do not appear in $X_i$, their effects can not be identified. The authors proposed a method to estimate their bounds nonparametrically. Suppose for all $x$, $B_t \leq m(x, \mu_i, \epsilon_{i,1}) \leq B_u$ for known constants $B_l$ and $B_u$. Let $\tilde{P}(x) = \sum_{i=1}^{n} I(T_i(x) = 0)/n$ be the sample frequency of $x$ not occurring in any time period. Estimated lower and upper bounds for $g(x)$ are

$$\hat{g}_l(x) = n^{-1} \sum_{i=1}^{n} \tilde{Y}_i(x) + \tilde{P}(x)B_l,$$

$$\hat{g}_u(x) = \hat{g}_l(x) + \tilde{P}(x)(B_u - B_l).$$

Corresponding estimated lower and upper bounds for the ATE are $\hat{\Delta}_l = \hat{g}_l(x^a) - \hat{g}_u(x^b)$ and $\hat{\Delta}_u = \hat{g}_u(x^a) - \hat{g}_l(x^b)$. The width of these estimated bounds is then given by

$$\hat{\Delta}_u - \hat{\Delta}_l = [\tilde{P}(x^a) + \tilde{P}(x^b)](B_u - B_l).$$

The bounds for the QSF can be obtained in a similar manner, based on the known lower bound of 0 and upper bound of 1 for the CDF. For both types of bounds, the authors proposed consistent estimators of their asymptotic variance and established asymptotic normality.

The static model presented above is based on the time homogeneity condition (4.1). The authors also considered a dynamic panel model under the condition

$$\epsilon_{i,t} | (X_{i,t}, \ldots, X_{i,1}, \mu_i) \overset{d}{=} \epsilon_{i,1} | (X_{i,1}, \mu_i), \quad \text{for all } t.$$ 

This is a “predetermined” version of time homogeneity under which the conditional distribution given only current and lagged regressors must be time invariant. The conditioning on $X_{i,1}$ is a way to account for the initial conditions of dynamic models.

The authors also considered other extensions such as incorporation of time effects and semiparametric multinomial models. To save space, we do not discuss these extensions. Interested readers are referred to the original paper for details.

A remarkable feature of the methods proposed in this paper is their simplicity. No optimizations or iterations are required to calculate most of the quantities of interest. However, the methods are only applicable to discrete regressors. Given their focus on treatment effect estimation, this restriction is rather innocuous. We conjecture that the methods can be used for models with continuous regressors if they are properly discretized.
5 Nonparametric Tests

In nonparametric test literature, it is a common practice to test for functional form specification or conditional variance structure, although tests for random-effects models against fixed-effects models, for data poolability, for cross-sectional independence, and for within-group independence are only relevant for panel/longitudinal data.

Henderson et al. (2008) and Li, Lin and Sun (2013) constructed nonparametric tests to test for a linear fixed-effects panel data mean regression model against a nonparametric fixed-effects panel data mean regression model, where the former is based on an $L^2$-distance in the spirit of Härdel and Mammen (1993) and the latter is essentially a residual-based test in the spirit of Zheng (1996). Henderson et al. (2008) relied on bootstrap methods to remove asymptotic centers and to obtain critical values. Li et al.’s (2013) test converges to a standard normal distribution under the linear fixed effects panel data model and is explosive under the alternative hypothesis. Another advantage of Li et al.’s (2013) model specification test is that it works under both fixed-effects and random-effects model and it is not necessary to pre-test whether the model is a fixed-effects model. However, if readers are interested in testing a nonparametric random-effects panel data model against a nonparametric fixed-effects panel data model, Sun, Carroll and Li’s (2009) residual-based test can be used. In a semiparametric varying coefficient panel data model setup, Sun et al. (2009) showed that their proposed test is a consistent test and converges to a standard normal distribution under the random-effects null hypothesis.

For the rest of this section, our interests lie in nonparametric poolability test in Section 5.1 and nonparametric test for cross-sectional independence in Section 5.2.

5.1 Poolability tests

Baltagi, Hidalgo and Li (1996) tested for no structural changes across time for a panel data with large $n$ and finite $T$. Specifically, they considered the following nonparametric panel data model

$$Y_{i,t} = m_t (X_{i,t}) + u_{i,t}, i = 1, 2, ..., n, t = 1, ..., T$$

(5.1)

where $(X_{i,t}, Y_{i,t})$ are independent across index $i$ with no restriction across index $t$, $m_t (\cdot)$ is an unknown smooth measurable function that may vary across time $t$, $E (u_{i,t}|X_{i,t}) = 0$, and $\sigma^p (x) = E (u_{i,t}^p | X_{i,t} = x)$ (CHECK! ) is continuous and bounded for $p \leq 4$ with a finite $E (u_{i,t}^4)$. The null and alternative hypotheses are

$$H_0 : \Pr \{ m_t (X) = m_s (X) \} = 1 \text{ for all } t \neq s.$$  

(5.2)

$$H_1 : \Pr \{ m_t (X) \neq m_s (X) \} > 0 \text{ for some } t \neq s.$$  

(5.3)

Define $m (x) \equiv T^{-1} \sum_{t=1}^T m_t (x)$. Under $H_0$, $m (x) \equiv m_t (X_{i,t})$ almost surely for all $t$, while under $H_1$, $m (x) \neq m_t (X_{i,t})$ with a positive probability over some interval for some $t$. Let $e_{i,t} \equiv Y_{i,t} - m (X_{i,t})$. Then, $E [e_{i,t} | E (e_{i,t} | X_{i,t})] = E [m_t (X_{i,t}) - m (X_{i,t})]^2 \geq 0$ so that the null hypothesis holds true if and only if $E [e_{i,t} | E (e_{i,t} | X_{i,t})] = 0$ for all $t$. Let $\hat{m} (x)$ be the kernel estimator estimated from
the pooled model $Y_{i,t} = m(X_{i,t}) + e_{i,t}$. The test statistic is constructed by estimating the unconditional mean $E(\cdot)$ by the sample average and replacing $E(e_{i,t}|X_{i,t})$ by its kernel estimator and $e_{i,t}$ by $\hat{e}_{i,t} = Y_{i,t} - \hat{m}(X_{i,t})$. To remove the random denominator due to the kernel estimation of $E(e_{i,t}|X_{i,t})$, the final test is based on $E[e_{i,t}E(e_{i,t}|X_{i,t}) f_t(X_{i,t})]$, where $f_t(x)$ is the Lebesgue probability density function of $X_{i,t}$ for all $i$ and a given $t$. The final test statistic, which is a residual-based test statistic in the spirit of Zheng (1996), is shown to be asymptotically pivotal and consistent and has a standard normal distribution under the null hypothesis.

In nonparametric panel data QR framework, Sun (2006) considered a poolability test for survey data sets comprising $J$ subgroups and each subgroup has $n_j$ members obeying a nonparametric QR model

$$Y_{i,j} = m_j(X_{i,j}) + u_{i,j}, \quad i = 1, ..., n_j, \quad j = 1, ..., J$$

(5.4)

where $Pr(u_{i,j} \leq 0|X_{i,j}) = p \in (0, 1)$ in model (5.1), the total sample size $n = \sum_{j=1}^{J} n_j$, and $n_j$ is large such that $\lim_{n \to \infty} n_j/n = a_j \in (0, 1)$ for all $j$ and $J$ is a finite positive integer. The observations $(X_{i,j}, Y_{i,j})$, $i = 1, ..., n_j$, $j = 1, ..., J$ are assumed to be i.i.d. across index $i$ and independence across index $j$. A relevant empirical exercise can be testing median food Engle curves equality across different family types. Zheng (1998) studied the residual-based nonparametric test to test for a parametric QR model against a nonparametric one for cross-sectional data, while Jeong et al. (2012) studied the same test but for weakly dependent time series data. Following Sun’s (2006), Dette, Wagener, and Volgushev (2011) developed poolability tests based on an $L_2$-distance between non-crossing nonparametric quantile curve estimators of Dette and Volgushev (2008), which is measured by

$$M^2 \equiv \sum_{j_1=1}^{J} \sum_{j_2=1}^{J} \int [m_{j_1}(x) - m_{j_2}(x)]^2 w_{j_1,j_2}(x) \, dx \geq 0,$$

where $w_{j_1,j_2}(\cdot) > 0$ are subjectively-chosen weight functions. The null hypothesis of equality holds if and only if $M^2 = 0$. Both Sun’s (2006) and Dette et al.’s (2011) test statistics are shown to be consistent and asymptotically normal under the null hypothesis of data poolability across subgroups.

Jin and Su (2013) constructed a nonparametric poolability test for a panel data model with cross-sectional dependence that can be applied to test a null model (2.24)-(2.25)-(2.26) from an alternative model $Y_{i,t} = m_i(X_{i,t}) + u_{i,t}$ with (2.25) and (2.26), where $\{v_{i,t}\}_{t=1}^{T}$ is a strictly stationary $\alpha$-mixing sequence and is independent of $\{\epsilon_{i,t}\}$, $\{g_t\}$ is independent of $\{(\epsilon_{i,t}, v_{i,t})\}$, and $\{(\epsilon_{i,t}, v_{i,t})\}$ is independently distributed across index $i$. The authors are interested in testing whether $m_1(x) \equiv m_2(x) \equiv \ldots \equiv m_n(x)$ or homogeneous relationship across different cross-sectional units for large $n$. 

28
and large $T$. The measure used to differentiate the null hypothesis from the alternative hypothesis is

$$
\Gamma_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \int [m_i(x) - m_j(x)]^2 w(x) \, dx,
$$

where $w(x)$ is a known weight function. With a large $n$, $\Gamma_n$ can not tell $H_0$ from the case that only ignorable number of units having different curves, so the null and alternative hypotheses are revised as

$$
H_0 : \Delta_m = 0 \text{ vs. } H_1 : \Delta_m > 0,
$$

where $\Delta_m = \lim_{n \to \infty} \Gamma_n$. Applying the sieve estimation method presented in Jin and Su (2013), they first estimated the unknown functions $m_i(\cdot), i = 1, ..., n$, under the alternative hypothesis and then calculated the test statistic $\hat{\Gamma}_n$ with $m_i(x)$ replaced by its estimator. The standardized version of the test statistic $\hat{\Gamma}_n$ is shown to be consistent and converge to a standard normal distribution under $H_0$.

### 5.2 Tests for cross-sectional independence

Assuming the panel data $\{(X_{i,t}, Y_{i,t})\}$ to be independent in index $i$ is a common practice in empirical analysis. However, it is reasonable to believe that common economic status and other unobserved factors may induce cross-sectional dependence of $Y_{i,t}$ even after controlling $X_{i,t}$. For example, one household’s spending decision may not be totally independent of others’ spending decisions at a given time as they all live through the same economy, so that a panel data model with unobserved factors changing across units and time is more proper than a simple cross-sectional fixed-effects panel data model. Statistically, ignoring cross-sectional dependence structure may lead to inconsistent parametric and nonparametric estimators constructed from cross-sectional independent models. Sarafidis and Wansbeek (2012) provided an excellent overview on recent development for parametric panel data models with cross-section dependence, including the conceptual measurement of the degree of cross-sectional dependence, estimation methods under both strict and weak exogeneity and tests for cross-section dependence.

Chen, Gao and Li (2012b) constructed a kernel-based test statistic to test for cross-sectional correlation and considered the following model

$$
Y_{i,t} = m_i(X_{i,t}) + \sigma_i(X_{i,t}) \epsilon_{i,t}, \quad i = 1, ..., n, t = 1, ..., T,
$$

where both $m_i(\cdot)$ and $\sigma_i(\cdot)$ are unknown smooth measurable functions and can be different for different unit $i$, and $\{\epsilon_{i,t}\}$ is independent of $\{X_{i,t}\}$ with $E(\epsilon_{i,t}) = 0$ and $E(\epsilon_{i,t}^2) = 1$. The null and alternative hypotheses are

$$
H_0 : E(\epsilon_{i,t}\epsilon_{j,t}) = 0 \text{ for all } t \geq 1 \text{ and all } i \neq j;
$$

$$
H_1 : E(\epsilon_{i,t}\epsilon_{j,t}) \neq 0 \text{ for some } t \geq 1 \text{ and some } i \neq j.
$$

With sufficiently large $T$, they firstly calculated the LLLS estimator $\hat{m}_i(x)$ using data only from unit $i$ and defined $\hat{u}_{i,t} = Y_{i,t} - \hat{m}_i(X_{i,t})$. Following Pesaran (2004) for parametric linear panel
data models, Chen et al. (2012b) proposed a nonparametric cross-sectional uncorrelatedness test statistic

\[ NCU = \sqrt{\frac{T}{n(n-1)}} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{\rho}_{i,j} \] (5.6)

where \( \hat{\rho}_{i,j} = \frac{T^{-1} \sum_{t=1}^{T} \tilde{u}_{i,t} \tilde{u}_{j,t}}{\sqrt{\sum_{t=1}^{T} \tilde{u}_{i,t}^{2}} \sqrt{\sum_{t=1}^{T} \tilde{u}_{j,t}^{2}}} \)
is the sample correlation between \( \tilde{u}_{i,t} \) and \( \tilde{u}_{j,t} \) with \( \tilde{u}_{i,t} = \hat{u}_{it} \hat{f}_{i}(X_{i,t}) \) and \( \hat{f}_{i}(X_{i,t}) \) equal to the random denominator of the LLLS estimator \( \hat{m}_{i}(X_{i,t}) \). The test is applied to models with strictly exogenous explanatory variables and is asymptotically normally distributed under \( H_0 \).

However, for a sufficiently large \( n \), the comment on Jin and Su’s (2013) null and alternative hypotheses seems also applicable here: The test cannot distinguish an alternative model with an ignorable number of pairs of \((i,t), (j,t)\) being correlated from the null model as \( \hat{\rho}_{i,j} \leq 1 \) in (5.6).

With strictly exogenous variable, \( X_{i,t} \), in model (5.5), Su and Zhang (2010) tested for pairwise cross-sectional independence by comparing the joint and marginal density functions of \( u_{i,t} \) and \( u_{j,t} \). Assuming that \( f_{i,j}(u_{i,t}, u_{j,t}) \) has a joint probability density function \( f_{i,j}(\cdot, \cdot) \), the null and alternative hypotheses are

\[ H_0 : \Pr \{ f_{i,j}(u_{i,t}, u_{j,t}) = f_{i}(u_{i,t}) f_{j}(u_{j,t}) \text{ for all } i \neq j \} = 1 \]
\[ H_1 : \text{not } H_0. \]

The measurement used to differentiate the null against alternative hypothesis is an \( L_2 \)-distance

\[ \Gamma_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \int |f_{i,j}(u, v) - f_{i}(u) f_{j}(v)|^2 du dv \geq 0 \]

and the null hypothesis holds true if and only if \( \Gamma_n \equiv 0 \). As Su and Zhang (2010) considered the case that both \( n \) and \( T \) are sufficiently large, for a given \( i \), the unknown curve \( m_i(\cdot) \) is consistently estimated via the local polynomial approach and the test is conducted over nonparametric residuals \( \tilde{u}_{i,t} = Y_{i,t} - \hat{m}_{i}(X_{i,t}) \) with \( \hat{f}_i(\cdot) \) and \( f_{i,j}(\cdot, \cdot) \) replaced by their corresponding kernel density estimators.

With sufficiently large \( n \) and large \( T \), for each given \( i \), \( X_{i,t} \) is assumed to be a strictly exogenous variable in both Chen et al. (2012b) and Su and Zhang (2010), under which condition the presence of cross-sectional correlation or dependence only affects the estimation efficiency of nonparametric estimation of the unknown curves not the consistency. The strict exogeneity assumption, to some extent, limits the value of testing for cross-sectional dependence in practice. In addition, if one assumes that \( m_i(\cdot) \) are the same across units, our review in Section 2.1 may be extended to cross-sectional dependence case: the kernel estimator of the unknown curve is expected to be asymptotically more efficient with working independence across units and time than with working
dependence across units and time, so that it is not a primary interest to test cross-sectional dependence when both \( n \) and \( T \) are large and the unknown curves are the same across time and units for strictly exogeneous cases.

In linear fixed-effects panel data model framework, Sarafidis, Yamagata, and Robertson (2009) tested cross-sectional dependence for large \( n \) and finite \( T \), using the generalized method of moments (GMM). It would be interesting to see tests for cross-sectional dependence for large \( n \) but small \( T \) in nonparametric panel data model framework.

6 Conclusion

The chapter selectively reviews some recent results on nonparametric panel data analysis, where the panel data are assumed to be stationary across time if \( T \) is large. For panel unit roots tests and panel cointegration literature, Banerjee (1999) and Phillips and Moon (2000) are two excellent survey papers in parametric regression framework. Also, our literature search on estimation and test of nonparametric panel data quantile regression models are not very fruitful, although there are many publications on estimation and tests of non-/semi-parametric panel data quantile regression models for independent and time series data; e.g., Lee (2003), Honda (2004), Sun (2005), Kim (2007), Cai and Xu (2008) and Cai and Xiao (2012), among which Cai and Xiao’s (2012) semiparametric dynamic quantile regressions models with partially varying coefficients have the most general form. However, there is no parallel work extended to panel data models yet, to the best of our knowledge. Lastly, there is a large body of work on nonparametric panel data models using spline estimators. Interested readers are referred to Wu and Zhang’s (2006) book on nonparametric methods for longitudinal data analysis for a general treatment of these methods.
References


