Semiparametric Estimation of Partially Varying-Coefficient Dynamic Panel Data Models*†

Zongwu Cai[a,b], Linna Chen[c], and Ying Fang[b,‡]

[a]Department of Economics, University of Kansas, Lawrence, KS 66045, USA.

[b]Wang Yanan Institute for Studies in Economics, MOE Key Laboratory of Econometrics, and Fujian Key Laboratory of Statistical Science, Xiamen University, Xiamen, Fujian 361005, China.

[c] School of Economics, Fudan University, Shanghai, Shanghai 200433, China.

July 12, 2013

Abstract

This paper studies a new class of semiparametric dynamic panel data models, in which some of the coefficients are allowed to depend on other informative variables and some of the regressors can be endogenous. To estimate both parametric and nonparametric coefficients, a three-stage estimation method is proposed. A nonparametric GMM is adopted to estimate all coefficients firstly and an average method is used to obtain the root-N consistent estimator of parametric coefficients. At the last stage, the estimator of varying coefficients is obtained by plugging the parametric estimator into the model. The consistency and asymptotic normality of both estimators are derived. Monte Carlo simulations verify the theoretical results and demonstrate that our estimators work well even in a finite sample.

Keywords: Dynamic Panel Data; Varying Coefficients; Nonparametric GMM.

---

*This research was supported, in part, by the National Nature Science Foundation of China grants #71131008 (Key Project), #70971113 and #71271179, as well as the Fundamental Research Funds for the Central Universities grant #2010221092.

†*E-mail addresses*: zongwucai@gmail.com (Z. Cai); linnachen918@gmail.com (L. Chen); yifst1@gmail.com (Y. Fang).

‡*The correspondence author.*

1 Introduction

Dynamic panel data models have received a lot of attentions among both theoretical and empirical economists since the seminal work of Balestra and Nerlove (1966). Based on the early work by Anderson and Hsiao (1981, 1982), there exists a rich literature on using the generalized method of moments (GMM) to estimate the dynamic panel data model and discuss the efficiency of the estimation. For example, Holtz-Eakin, Newey and Rosen (1988) considered the estimation of vector autoregressions with panel data, Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), Hahn (1997, 1999) and among others discussed how to utilize additional instruments to improve the efficiency of GMM estimation. Dynamic panel data models have been widely applied to various empirical studies as well. For example, Baltagi and Levin (1986) estimated the dynamic demand for addictive commodities, Islam (1995) used dynamic panel data approach to study growth empirics, and Park, Sickles and Simar (2007) employed dynamic panel data to analyze the demand between city pairs for some airlines. More references can be found in Arellano (2003), Hsiao (2003) and Baltagi (2005).

It is well known, however, that the aforementioned parametric dynamic panel data models are unable to accommodate sufficient flexibility to catch nonlinear structure and suffer from the model misspecification problem. To deal with this misspecification issue, various nonparametric or semi-parametric static panel data models have been proposed. For example, Horowitz and Markatou (1996), Li, Huang, Li and Fu (2002) and Su and Ullah (2006) studied semiparametric estimation of a partially linear panel data model without including endogenous regressors. Hoover, Rice, Wu and Yang (1998) considered a smoothing spline and a local polynomial estimation for time-varying coefficient panel data models. Lin and Ying (2001) and Lin and Carroll (2001, 2006) examined the semiparametric estimation of a panel data model with random effects. Henderson, Carroll and Li (2008) considered a partially linear panel data model with fixed effects and proposed a consistent estimator based on iterative backfitting procedures and an initial estimator. Finally, Qian and Wang (2011) proposed a marginal integration method to estimate the nonparametric part in a semiparametric panel data with unobserved individual effects.

In recent years, motivated by the increase in the empirical economic growth literature, many studies have paid much attention to the dynamic panel data models. For example, Li and Stengos (1996), Li and Ullah (1998) and Baltagi and Li (2002) considered semiparametric estimation of partially linear dynamic panel data models using instrumental variable methods. Park, Sickles and Simar (2007) focused on constructing a semiparametric efficient estimator in a dynamic panel data model. They considered a linear dynamic panel data model assuming that the error terms are generating from a normal distribution but specifying other parametric distributions nonparametrically. An efficient estimator was established based on a stochastic expansion. However, they ignored the endogenous problem in a dynamic panel data model by assuming all the error terms and the random effect are independent of regressors.

Recently, Cai and Li (2008) proposed a nonparametric GMM estimation of varying-coefficient dynamic panel data models to deal with the potential endogeneity issue. Varying-coefficient models are well known in the statistic literature and also have a lot of applications in economics and finance (Hastie and Tibshirani, 1993; Cai, Fan and Yao, 2000; Cai, Das, Xiong and Wu, 2006; Cai and Hong, 2009; Cai, Gu and Li, 2009; Cai, Chen and Fang, 2012; among others); see Cai (2010) for more details in applications in economics and finance. One of the main advantages of the varying-coefficient models is that it allows the coefficients to depend on some informative variables and then balances the dimension reduction and model flexibility.

In this paper, we consider a new class of partially varying-coefficient dynamic models. It allows for linearity in some regressors but for nonlinearity in other regressors. In other words, some coefficients are constant but others are varying over some variables. The new class model is flexible enough to include many existing models as special cases. By extending from Cai and Li (2008) to a partially varying-coefficient model, we reduce the model dimension without influencing the degree of the model flexibility, and furthermore, root-N consistent estimation of parametric coefficients can be achieved. We propose a three-stage estimation procedure to estimate the constant and varying coefficients. At the first stage, all coefficients are treated as varying coefficients and then the nonparametric GMM proposed by Cai and Li (2008) is adopted. At the second stage, the constant coefficients are estimated by the average method and the root-N consistency and asymptotic normality of the estimators are derived. Finally, the estimators at the second stage are plugged into the original model and then the estimators of varying coefficients are obtained by employing the nonparametric GMM again. The partially varying-coefficient panel data model can be applied to various empirical applications. For example, Lin, Huang and Weng (2006) and Zhou and Li (2011) employed a special case of the partially varying-coefficient models to investigate the so called Kuznet's hypothesis which claims an inverted-U relation between inequality and economic development.

Compared with the existing literature, our three-stage estimation has the following merits. Firstly, in the existing literature, it is common to adopt the Robinson's (1988) framework to estimate a semiparametric panel data model with endogeneity. When endogenous variables appear in the model, a two-stage estimation is required, where a high dimensional nonparametric estimation, in which the dimension depends on the number of excluded instruments and included exogenous variables, is usually employed at the first stage, and then an instrumental variable regression is adopted using first-stage nonparametric estimators as generated regressors. However, the nonparametric GMM adopted in this paper only requires an one-step relatively low dimensional estimation. The dimension of the estimation depends on the number of smoothing variables rather than the included and excluded exogenous variables. Since the nonparametric GMM is adopted at the first stage, some popular semiparametric estimation methods, such as Robinson's (1988) method and profile least squares method, cannot be applied here to estimate the parametric part. Instead, we propose the average method by taking average of all local estimates to obtain the root-N consistent estimation of parametric coefficients. Finally, varying coefficients can be estimated by applying the low dimensional nonparametric GMM using the root-N consistent estimators as generated regressors.

The rest of the paper is organized as follows. Section 2 introduces the model and estimation method. We derive the asymptotic results of the proposed estimators in Section 3. Section 4 reports some Monte Carlo simulations to verify our theoretical results and demonstrates the finite sample performance of our estimators. Finally, Section 5 concludes. All technical proofs are relegated to Appendices.

2 The Model and Estimation Procedures

This paper considers a new class of partially varying-coefficient (dynamic) panel data models as follows:

$$Y_{it} = \boldsymbol{X}'_{it,1}\boldsymbol{\gamma} + \boldsymbol{X}'_{it,2}\boldsymbol{\beta}(U_{it}) + \epsilon_{it}, \quad 1 \le i \le N, \quad 1 \le t \le T, \tag{1}$$

where Y_{it} is a scalar dependent variable, U_{it} is a scalar smoothing variable,[1] $\boldsymbol{X}_{it,1}$ and $\boldsymbol{X}_{it,2}$ are regressors with $d_1 \times 1$ and $d_2 \times 1$ dimensions respectively, $\boldsymbol{\gamma}$ denotes $d_1 \times 1$ constant coefficients and $\boldsymbol{\beta}(\cdot)$ denotes $d_2 \times 1$ varying coefficients, and the random error ϵ_{it} allows to be correlated over periods t but independent over i. We consider a typical panel data model such that N is large but T is relatively short. Moreover, let $\boldsymbol{X}_{it} = (\boldsymbol{X}'_{it,1}, \boldsymbol{X}'_{it,2})'$ with dimension $d \times 1$ where $d = d_1 + d_2$. In particular, in model (1) $\boldsymbol{X}_{it}$ may contain lagged variables of Y_{it} and endogenous variables correlated with the error term so that the classical dynamic panel model can be regarded as a special case. Also, the above setup is quite flexible to capture complex dynamic structures in real applications in economics. For example, Li and Stengos (1996), Li and Ullah (1998) and Baltagi and Li (2002) considered a special case by assuming that $\boldsymbol{X}_{it,2}$ only contains a constant term. When $X_{it,2}$ is a discrete value random variable, the above model reduces to Das (2005). Cai and Li (2008) studied a varying-coefficient model by ignoring the parametric part. Many semiparametric varying coefficient literatures such as Fan and Huang (2005) and Lin, Huang and Weng (2006) studied the above model without dealing with the endogeneity issue.

In model (1), an ordinary least squares estimation cannot be applied since the orthogonality condition fails, i.e., $E[\epsilon_{it}|\boldsymbol{X}_{it}, U_{it}] \neq 0$. Hence, we assume that there exists a $q \times 1$ vector of instruments $\boldsymbol{W}_{it}$ that satisfies $E[\epsilon_{it}|\boldsymbol{W}_{it}, U_{it}] = 0$.[2] By choosing an appropriate vector function

[1] For simplicity, we only consider the univariate case for the smoothing variable. The estimation procedure and asymptotic results still hold for the multivariate case with much complicated notation.

[2] Instruments should be highly correlated to the endogenous variables and uncorrelated to the structural errors. Cai, Fang and Li (2012) and Cai, Fang and Su (2012) studied the instrumental variable estimation using weak instruments in a panel data model. Berkowitz, Caner and Fang (2008, 2012) investigated the impact on estimation and testing when instruments are slightly correlated with random errors.

$Q(\boldsymbol{V}_{it})$ where $\boldsymbol{V}_{it} = (\boldsymbol{W}'_{it}, U_{it})'$, we have the following conditional moment conditions,

$$E[Q(\boldsymbol{V}_{it})\epsilon_{it}|\boldsymbol{V}_{it}] = 0. \tag{2}$$

Instead of using a nonparametric projection of some endogenous components in $\boldsymbol{X}_{it}$ on $Q(\boldsymbol{V}_{it})$, we apply the nonparametric GMM (Cai and Li, 2008) to estimate all varying coefficients at the first stage. We treat all coefficients to be varying so that $\boldsymbol{\gamma} = \boldsymbol{\gamma}(U_{it})$ and $\boldsymbol{\beta} = \boldsymbol{\beta}(U_{it})$. We assume throughout that $\boldsymbol{\beta}(\cdot)$ and $\boldsymbol{\gamma}(\cdot)$ are twice continuously differentiable. We apply the local constant approximations to $\boldsymbol{\gamma}(U_{it})$ and $\boldsymbol{\beta}(U_{it})$ (Lewbel, 2007; Fang, Ren and Yuan, 2011), then model (1) is approximated by the following model in a small neighborhood of u_0:

$$Y_{it} \approx \boldsymbol{X}'_{it}\boldsymbol{\theta} + \epsilon_{it}, \quad 1 \leq i \leq N, \quad 1 \leq t \leq T, \tag{3}$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}(u_0) = (\boldsymbol{\gamma}'(u_0), \boldsymbol{\beta}'(u_0))'$ is a $d \times 1$ vector of parameters. Based on the locally weighted moment conditions $\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{Q}(\boldsymbol{V}_{it})(Y_{it} - \boldsymbol{X}'_{it}\boldsymbol{\theta})K_{h_1}(\boldsymbol{U}_{it} - u_0) = 0$,[3] the nonparametric GMM estimator is given by

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(u_0) = (\boldsymbol{\Omega}'_N\boldsymbol{\Omega}_N)^{-1}\boldsymbol{\Omega}'_N\boldsymbol{\Phi}_N, \tag{4}$$

where $\boldsymbol{\Omega}_N = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{Q}_{it}\boldsymbol{X}'_{it}K_{h_1}(U_{it} - u_0)$ and $\boldsymbol{\Phi}_N = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{Q}_{it}K_{h_1}(U_{it} - u_0)Y_{it}$. We simply choose instruments $\boldsymbol{Q}_{it}$ to be $\boldsymbol{V}_{it}$. Note that we require $q \geq d$ to satisfy the identification condition, and also that $K_{h_1}(\cdot) = h_1^{-1}K(\cdot/h_1)$ where $K(\cdot)$ is a kernel function with a bandwidth $h_1 = h_{1N} > 0$ which controls the degree of smoothing used in the nonparametric GMM estimation.

At the second stage, in order to take advantage of the full sample information to estimate the constant parameters $\boldsymbol{\gamma}$, we employ the average method to achieve the root-N consistent estimator of $\boldsymbol{\gamma}$:

$$\hat{\boldsymbol{\gamma}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{\boldsymbol{\gamma}}(U_{it}). \tag{5}$$

The $\hat{\boldsymbol{\gamma}}(U_{it})$ is the first d_1 components in $\hat{\boldsymbol{\theta}}$.

The last step is to estimate the nonparametric part, the functional coefficients $\boldsymbol{\beta}(U_{it})$, by plugging a root-N consistent estimator $\hat{\boldsymbol{\gamma}}$ into model (1). Define a partial residual $Y^*_{it} = Y_{it} - \boldsymbol{X}'_{it,1}\hat{\boldsymbol{\gamma}}$. Hence,

[3] To obtain a unique $\boldsymbol{\theta}$ satisfying the above moment condition, we follow Cai and Li (2008) by pre-multiplying it by $\boldsymbol{\Omega}'_N$.

model (1) can be approximated by

$$Y_{it}^* \approx \boldsymbol{P}_{it}'\boldsymbol{\delta} + \epsilon_{it}, \quad 1 \le i \le N, \quad 1 \le t \le T, \tag{6}$$

where $\boldsymbol{\delta} = \boldsymbol{\delta}(u_0) = (\boldsymbol{\beta}'(u_0), \dot{\boldsymbol{\beta}}'(u_0))'$, $\dot{\boldsymbol{\beta}}(\cdot)$ denotes the first order derivatives of $\boldsymbol{\beta}(\cdot)$ with respect to U_{it}, and $\boldsymbol{P}_{it} = \begin{pmatrix} \boldsymbol{X}_{it,2} \\ \boldsymbol{X}_{it,2} \otimes (U_{it} - u_0) \end{pmatrix}$ is a $(2d_2) \times 1$ vector. Hence, the nonparametric GMM estimator of the varying coefficients are given by

$$\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}(u_0) = (\boldsymbol{S}_N'\boldsymbol{S}_N)^{-1}\boldsymbol{S}_N'\boldsymbol{T}_N, \tag{7}$$

where $\boldsymbol{S}_N = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{Q}_{it}\boldsymbol{P}_{it}'K_{h_2}(U_{it}-u_0)$ and $\boldsymbol{T}_N = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{Q}_{it}K_{h_2}(U_{it}-u_0)Y_{it}^*$ with $K_{h_2}(\cdot) = h_2^{-1}K(\cdot/h_2)$ and the bandwidth $h_2 = h_{2N} > 0$. Motivated by the local linear fitting, a simple choice of $\boldsymbol{Q}_{it}$ suggested by Cai and Li (2008) is a $(2q) \times 1$ vector

$$\boldsymbol{Q}_{it} = \begin{pmatrix} \boldsymbol{W}_{it} \\ \boldsymbol{W}_{it} \otimes (U_{it} - u_0)/h_2 \end{pmatrix},$$

which is used at the last stage.

3 Asymptotic Theories

In this section, we will derive the asymptotic results of both estimators $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}(u_0)$. The detailed proofs are relegated to Appendices. Firstly, we give some notations and definitions which will be used in the rest of the paper. Denote $\mu_j = \int_{-\infty}^{\infty} u^j K(u)du$ and $\nu_j = \int_{-\infty}^{\infty} u^j K^2(u)du$ with $j \ge 0$. Let $\sigma^2(\boldsymbol{v}) = Var(\epsilon_{it}|\boldsymbol{V}_{it} = \boldsymbol{v})$, $\boldsymbol{\Omega} = \boldsymbol{\Omega}(u_0) = E(\boldsymbol{V}_{it}\boldsymbol{X}_{it}'|u_0)$, $\tilde{\boldsymbol{\Omega}} = \tilde{\boldsymbol{\Omega}}(u_0) = E(\boldsymbol{W}_{it}\boldsymbol{X}_{it,2}'|u_0)$, $\boldsymbol{\Phi} = \boldsymbol{\Phi}(u_0) = Var(\boldsymbol{V}_{it}\epsilon_{it}|u_0)$, $\sigma_{1t}(\boldsymbol{V}_{i1}, \boldsymbol{V}_{it}) = E(\epsilon_{i1}\epsilon_{it}|\boldsymbol{V}_{i1}, \boldsymbol{V}_{it})$, and $\boldsymbol{G}_{1t}(U_{i1}, U_{it}) = E\{\boldsymbol{V}_{i1}\boldsymbol{V}_{it}'\sigma_{1t}|U_{i1}, U_{it}\}$. Moreover, define $\boldsymbol{S} = \boldsymbol{S}(u_0) = \begin{pmatrix} \tilde{\boldsymbol{\Omega}} & \boldsymbol{0} \\ \boldsymbol{0} & \mu_2\tilde{\boldsymbol{\Omega}} \end{pmatrix}$. Next, note that $\boldsymbol{\Phi}_N = \boldsymbol{\Omega}_N\boldsymbol{\theta} + \boldsymbol{\Phi}_N^* + \boldsymbol{\Psi}_N + \boldsymbol{\Lambda}_N$, where

$$\begin{aligned}
\boldsymbol{\Phi}_N^* &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{h_1}(U_{it}-u_0)\boldsymbol{Q}_{it}\epsilon_{it}, \\
\boldsymbol{\Psi}_N &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{h_1}(U_{it}-u_0)\boldsymbol{Q}_{it}\sum_{j=1}^{d}\psi_j(U_{it}, u_0)X_{itj}, \\
\text{and} \quad \boldsymbol{\Lambda}_N &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{h_1}(U_{it}-u_0)\boldsymbol{Q}_{it}\sum_{j=1}^{d}\Lambda_j(U_{it}, u_0)X_{itj}
\end{aligned}$$

with $\psi_j(U_{it}, u_0) = \dot{\theta}_j(u_0)(U_{it} - u_0) + \frac{1}{2}\ddot{\theta}_j(u_0)(U_{it} - u_0)^2$ and $\Lambda_j(U_{it}, u_0) = \theta_j(U_{it}) - \theta_j(u_0) - \dot{\theta}_j(u_0)(U_{it} - u_0) - \frac{1}{2}\ddot{\theta}_j(u_0)(U_{it} - u_0)^2$. Substituting it into (4), we have

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - (\boldsymbol{\Omega}_N'\boldsymbol{\Omega}_N)^{-1}\boldsymbol{\Omega}_N'\boldsymbol{\Psi}_N - (\boldsymbol{\Omega}_N'\boldsymbol{\Omega}_N)^{-1}\boldsymbol{\Omega}_N'\boldsymbol{\Lambda}_N = (\boldsymbol{\Omega}_N'\boldsymbol{\Omega}_N)^{-1}\boldsymbol{\Omega}_N'\boldsymbol{\Phi}_N^*. \quad (8)$$

We will show that the second term on the left side determines the bias, the last term on the left can be asymptotically ignored, and the term on the right follows the asymptotic normality. To establish the asymptotic results for the proposed estimators, following assumptions are needed although they might not be the weakest ones.

A1. $\{(\boldsymbol{W}_{it}, \boldsymbol{X}_{it}, Y_{it}, U_{it}, \epsilon_{it})\}$ are independently and identically distributed across the i index for each fixed t and strictly stationary over t for each fixed i, $E||\boldsymbol{W}_{it}\boldsymbol{X}_{it}'||^2 < \infty$, $E||\boldsymbol{W}_{it}\boldsymbol{W}_{it}'||^2 < \infty$, $E(\epsilon_{it}) = 0$, and $E|\epsilon_{it}|^4 < \infty$, where $||\cdot||^2$ is the standard L_2-norm for a finite-dimensional matrix.

A2. For each $t \geq 1$, $\boldsymbol{G}_{1t}(U_{i1}, U_{it})$ is continuous at (U_{i1}, U_{it}). Also, for each u_0, $\boldsymbol{\Omega}(u_0) > 0$ and $f(u_0) > 0$, which is the density function of U_{it} at u_0. Further, $sup_{t\geq 1}|\boldsymbol{G}_{1t}(u_0, u_0)f(u_0)| \leq \boldsymbol{M}(u_0) < \infty$ for some function $\boldsymbol{M}(u_0)$. Finally, $\boldsymbol{\beta}(u_0)$ and $f(u_0)$ are both two times continuously differentiable.

A3. The kernel function $K(\cdot)$ is a symmetric, bounded density with a bounded support region.

A4. The instrumental variable $\boldsymbol{V}_{it}$ satisfies the instrument exogeneity condition that $E(\epsilon_{it}|\boldsymbol{V}_{it}) = 0$.

A5. $h_1 \to 0$, $h_2 \to 0$, $Nh_1 \to \infty$ and $Nh_2 \to \infty$ as $N \to \infty$. Furthermore, $h_1 = o(h_2)$.

To derive the asymptotic properties for $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}}$, we first prove the following preliminary results.

Proposition 1. Under Assumptions A1-A5, we have

$$\begin{aligned}
&(i) \quad \boldsymbol{\Omega}_N = f(u_0)\boldsymbol{\Omega}[1 + o_p(1)],\\
&(ii) \quad \boldsymbol{\Psi}_N = \frac{h_1^2}{2} f(u_0)\mu_2[2(\dot{\boldsymbol{\Omega}} + \boldsymbol{\Omega}\frac{\dot{f}(u_0)}{f(u_0)})\dot{\boldsymbol{\theta}} + \boldsymbol{\Omega}\ddot{\boldsymbol{\theta}}] + o_p(h_1^2),\\
&(iii) \quad \boldsymbol{\Lambda}_N = o_p(h_1^2),\\
&(iv) \quad Nh_1 Var(\boldsymbol{\Phi}_N^*) \to \frac{1}{T} f(u_0)\boldsymbol{\Phi}.
\end{aligned}$$

Clearly, by Proposition 1 and (8), we can obtain

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - bias_{\boldsymbol{\theta}} = f^{-1}(u_0)(\boldsymbol{\Omega}'\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}'\boldsymbol{\Phi}_N^*[1 + o_p(1)], \tag{9}$$

where $bias_{\boldsymbol{\theta}} = \frac{h_1^2}{2} f(u_0)\mu_2[2((\boldsymbol{\Omega}'\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}'\dot{\boldsymbol{\Omega}} + \frac{\dot{f}(u_0)}{f(u_0)})\dot{\boldsymbol{\theta}} + \ddot{\boldsymbol{\theta}}] + o_p(h_1^2)$. The next two theorems demonstrate the consistency and asymptotic normality of $\hat{\boldsymbol{\gamma}}$, respectively.

THEOREM 1. Under Assumptions A1-A5, we have

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - bias_{\boldsymbol{\theta}} = o_p(h_1^2) + O_p(\frac{1}{\sqrt{Nh_1}}), \tag{10}$$

which implies the consistency of $\hat{\boldsymbol{\theta}}$.

Remark 1: As defined earlier, $\boldsymbol{\theta} = \boldsymbol{\theta}(u_0) = (\boldsymbol{\gamma}'(u_0), \boldsymbol{\beta}'(u_0))'$ so that $\hat{\boldsymbol{\gamma}}$ is the first d_1 component in $\hat{\boldsymbol{\theta}}$. Thus, we have

$$\begin{aligned} \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_1'[(\hat{\boldsymbol{\theta}}(U_{it}) - \boldsymbol{\theta}(U_{it})) - bias_{\boldsymbol{\theta}}] + O_p(\frac{1}{\sqrt{N}}) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[\hat{\boldsymbol{\gamma}}(U_{it}) - \boldsymbol{\gamma}(U_{it}) - bias_{\boldsymbol{\gamma}}(u_0)] + O_p(\frac{1}{\sqrt{N}}), \end{aligned} \tag{11}$$

where the selection matrix $e_1' = (\boldsymbol{I}_{d_1}, \boldsymbol{0}_{d_1 \times d_2})$ and $bias_{\boldsymbol{\gamma}}(u_0) = e_1' bias_{\boldsymbol{\theta}}(u_0)$.

THEOREM 2. Under Assumptions A1-A5, we have

$$\sqrt{N}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} - bias_{\boldsymbol{\gamma}}) \xrightarrow{D} N(0, \frac{1}{T}\boldsymbol{\Sigma}_{\gamma}), \tag{12}$$

where $\boldsymbol{\Sigma}_{\gamma} = E\{e_1'[\boldsymbol{D}(U_{it})\boldsymbol{\Phi}(U_{it})\boldsymbol{D}'(U_{it}) + \frac{2}{T}\sum_{t=2}^{T}(T-t+1)\boldsymbol{D}(U_{i1})\boldsymbol{G}_{\boldsymbol{1t}}(U_{i1}, U_{it})\boldsymbol{D}'(U_{it})]e_1\}$ with $\boldsymbol{D}(U_{it}) = (\boldsymbol{\Omega}'(U_{it})\boldsymbol{\Omega}(U_{it}))^{-1}\boldsymbol{\Omega}'(U_{it})$ and $bias_{\boldsymbol{\gamma}} = E[bias_{\boldsymbol{\gamma}}(U_{it})]$.

Remark 2: As $Nh_1^4 \to 0$, the bias term in the above theorem shrinks toward zero, which implies that we need to under-smooth at the first step to reduce the influence of the bias term that may be brought to the second step, while in the meantime, the effect of the first-step bandwidth selection on the variance can be smoothed out by using the average method.

Finally, we focus on the nonparametric estimation of $\boldsymbol{\beta}(u_0)$. Similar to the decomposition of $\boldsymbol{\Phi}_N$, we have $\boldsymbol{T}_N = \tilde{\boldsymbol{S}}_N \boldsymbol{H}\boldsymbol{\delta} + \boldsymbol{T}_N^* + \boldsymbol{B}_N + \boldsymbol{R}_N$ where

$$\begin{aligned}
\boldsymbol{T}_N^* &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_2}(U_{it}-u_0)\boldsymbol{Q}_{it}\epsilon_{it}, \\
\boldsymbol{B}_N &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_2}(U_{it}-u_0)\boldsymbol{Q}_{it}\frac{1}{2}\sum_{j=1}^{d_2}\ddot{\beta}_j(u_0)(U_{it}-u_0)^2 X_{it,2j}, \\
\text{and} \quad \boldsymbol{R}_N &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_2}(U_{it}-u_0)\boldsymbol{Q}_{it}\sum_{j=1}^{d_2} Rj(U_{it},u_0)X_{it,2j},
\end{aligned}$$

where $R_j(U_{it}, u_0) = \beta_j(U_{it}) - a_j - b_j(U_{it}-u_0) - \frac{1}{2}\ddot{\beta}_j(u_0)(U_{it}-u_0)^2$. Hence,

$$\boldsymbol{H}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}) - [\tilde{\boldsymbol{S}}_N'\tilde{\boldsymbol{S}}_N]^{-1}\tilde{\boldsymbol{S}}_N'\boldsymbol{B}_N - [\tilde{\boldsymbol{S}}_N'\tilde{\boldsymbol{S}}_N]^{-1}\tilde{\boldsymbol{S}}_N'\boldsymbol{R}_N = [\tilde{\boldsymbol{S}}_N'\tilde{\boldsymbol{S}}_N]^{-1}\tilde{\boldsymbol{S}}_N'\boldsymbol{T}_N^*, \tag{13}$$

where $\boldsymbol{H} = (\boldsymbol{I}_{d_2}, h_2\boldsymbol{I}_{d_2})$ and $\tilde{\boldsymbol{S}}_N = \boldsymbol{S}_N\boldsymbol{H}^{-1} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{Q}_{it}\tilde{\boldsymbol{P}}_{it}' K_{h_2}(U_{it}-u_0)$ with $\tilde{\boldsymbol{P}}_{it} = \boldsymbol{H}^{-1}\boldsymbol{P}_{it}$. Similar to Proposition 1, we have the following preliminary results.

Proposition 2. Under Assumptions A1-A5, we have

$$\begin{aligned}
&(i) \quad \tilde{\boldsymbol{S}}_N = f(u_0)\boldsymbol{S}[1+o_p(1)], \\
&(ii) \quad \boldsymbol{B}_N = \frac{h_2^2}{2}f(u_0)\begin{pmatrix} \mu_2\tilde{\boldsymbol{\Omega}}\ddot{\boldsymbol{\beta}} \\ 0 \end{pmatrix} + o_p(h_2^2), \\
&(iii) \quad \boldsymbol{R}_N = o_p(h_2^2), \\
&(iv) \quad Nh_2 Var(\boldsymbol{T}_N^*) \to \frac{1}{T}f(u_0)\boldsymbol{S}^*,
\end{aligned}$$

where $e_2' = (\boldsymbol{I}_q, \boldsymbol{0}_{q\times 1})$ is a selecting matrix, $\boldsymbol{S}^* = \boldsymbol{S}^*(u_0) = \begin{pmatrix} \nu_0 e_2'\boldsymbol{\Phi}e_2 & \boldsymbol{0} \\ \boldsymbol{0} & \nu_2 e_2'\boldsymbol{\Phi}e_2 \end{pmatrix}$.

By Proposition 2 and (12), we can obtain

$$\boldsymbol{H}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}) - \frac{h_2^2}{2}\begin{pmatrix} \mu_2\ddot{\boldsymbol{\beta}} \\ \boldsymbol{0} \end{pmatrix} + o_p(h_2^2) = f^{-1}(u_0)(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{T}_N^*[1+o_p(1)]. \tag{14}$$

The next two theorems depict the consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}(u_0)$, respectively.

THEOREM 3. Under Assumptions A1-A5, we have

$$\boldsymbol{H}\begin{pmatrix} \hat{\boldsymbol{\beta}}-\boldsymbol{\beta} \\ \hat{\dot{\boldsymbol{\beta}}}-\dot{\boldsymbol{\beta}} \end{pmatrix} - \frac{h_2^2}{2}\begin{pmatrix} \mu_2\ddot{\boldsymbol{\beta}} \\ \boldsymbol{0} \end{pmatrix} = o_p(h_2^2) + O_p(\frac{1}{\sqrt{Nh_2}}). \tag{15}$$

Also, we have the following asymptotic normality,

$$\sqrt{Nh_2}[\boldsymbol{H}\begin{pmatrix}\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\\ \hat{\dot{\boldsymbol{\beta}}}-\dot{\boldsymbol{\beta}}\end{pmatrix}-\frac{h_2^2}{2}\begin{pmatrix}\mu_2\ddot{\boldsymbol{\beta}}\\ \mathbf{0}\end{pmatrix}+o_p(h_2^2)]\xrightarrow{D}N(0,\frac{1}{T}f^{-1}(u_0)\boldsymbol{\Sigma}_\beta), \tag{16}$$

where $\boldsymbol{\Sigma}_\beta=(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{S}^*\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}$.

4 A Monte Carlo Study

In this section, Monte Carlo simulations are conducted to verify theoretical results in Section 3 and demonstrate the finite sample performance of both estimators. The mean absolute deviation errors (MADE) of the estimators are computed to measure the estimation performance. The MADE is defined by

$$\text{MADE}_j=\frac{1}{G}\sum_{i=g}^{G}|\hat{\delta}_j(U_g)-\delta_j(u_g)|,$$

where $\delta(\cdot)$ is either γ_Y, γ_Z or $\beta(\cdot)$ in (17) and $\{u_g\}_{g=1}^G$ are the gird points within the domain of U_{it}. Note that for both γ_Y and γ_Z, their MADE becomes the absolute deviation error (ADE).

We consider the following data generating process:

$$Y_{it}=Y_{it-1}\gamma_Y+Z_{it}\gamma_Z+\tilde{X}_{it}\beta(U_{it})+\epsilon_{it},\qquad \tilde{X}_{it}=W_{it}+\eta_{it}, \tag{17}$$

where the smoothing variable U_{it} and the exogenous variable Z_{it} are generated from uniform distributions $U(-3,3)$ and $U(-2,2)$, respectively. The excluded instruments W_{it} is generated independently from a uniform distribution $U(-2,2)$. The error terms ϵ_{it} and η_{it} are generated jointly from a standard bivariate normal distribution with the correlation coefficient 0.3. The coefficients are set by $\gamma_Y=0.5$, $\gamma_Z=3$ and $\beta(U_{it})=1.5e^{-U_{it}^2}$. We fix $T=10$ and let $N=200$, 500 and 1000 respectively. When generating the series of Y_{it}, we set the initial value to be zero and drop the first 100 observations to reduce the impact of initial values. For a given sample size, we repeat 500 times to calculate the MADE. The bandwidth in the first step is undersmoothed and we find the estimation of γ is not very sensitive to the bandwidth selection when it is chosen within a reasonable range.

Table 1 reports the medians and the standard deviations (in parentheses) of the MADE for different estimators under different sample sizes. When the sample size increases, the medians of ADE values for $\hat{\gamma}_Y$ and $\hat{\gamma}_Z$ shrink from 0.004 to 0.001 and from 0.016 to 0.006, respectively. The standard deviations also shrink quickly when the sample size is enlarged. For $\hat{\gamma}_Y$, the standard deviation shrinks from 0.003 to 0.001, and for $\hat{\gamma}_Z$, it decreases from 0.012 to 0.005. The nonparametric estimator of $\beta(\cdot)$ shows similar results. The median of the MADE values decreases from 0.076 to 0.030 when the sample size increases from 200 to 1000. At the same time, the standard deviation of the estimator also shrinks from 0.014 to 0.006. Compared with parametric estimations in Columns 2 and 3, the convergence speed of nonparametric estimator is relatively slow. All results show that the estimators proposed in the paper are consistent estimators and all outcomes in the simulations are consistent with the theoretical results in the previous section.

Figure 1 demonstrates the estimated curve of $\beta(\cdot)$ with a sample size $N = 500$ for a typical sample. The typical example is chosen such that its MADE_β value equals to the median of the 500 MADE_β values in the repeated experiments of the case $N = 500$. The solid line represents the true curve and the dotted line denotes the estimated one. Figure 1 shows that the nonparametric GMM estimation works very well even in a small sample.

Table 1: Median and standard deviation of the MADE values.

N	γ_Y	γ_Z	$\beta(\cdot)$
200	0.004144002	0.01629416	0.0768469
	(0.003479078)	(0.01242114)	(0.01431995)
500	0.002379212	0.009720373	0.04487873
	(0.002281685)	(0.008604812)	(0.00858833)
1000	0.001707388	0.006097411	0.03057437
	(0.001506411)	(0.005766932)	(0.006049484)

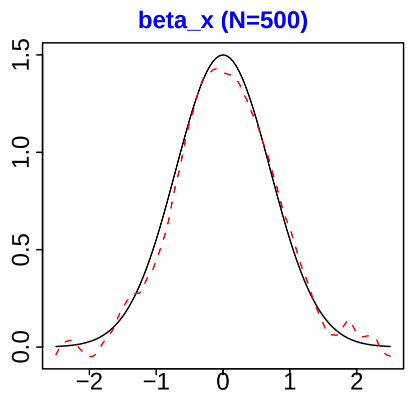

Figure 1: Functional Coefficient of $\beta(\cdot)$. The solid line represents the true curve and the dotted line denotes the estimated one.

5 Conclusion

This paper proposes a three-stage estimation procedure for a new class of partially varying coefficients dynamic panel data models, which, as expected, has many applications in applied economics particularly in empirical growth literature. The asymptotic properties of both constant and varying coefficients are established. The Monte Carlo simulations demonstrate that the proposed estimators work very well even in small samples. However, the cross sectional independence may be a restrictive assumption for some applications in real data. Therefore, it would be an interesting future research topic to work on a partially varying coefficients dynamic panel data model with cross sectional dependence.

Reference

Ahn, S.C., Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68: 5-27.

Anderson, T.W., Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76: 598-606.

Anderson, T.W., Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18: 47-82.

Arellano, M. (2003). Panel data econometrics. Oxford: Oxford University Press.

Arellano, M., Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277-297.

Arellano, M., Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68: 29-51.

Balestra, P., Nerlove, M. (1966). Pooling cross-section and time series data in the estimation of a dynamic model: the demand for nature gas. *Econometrica* 34: 585-612.

Baltagi, B.H. (2005). Econometric analysis of panel data. West Sussex: John Wiley & Sons Ltd.

Baltagi, B.H., Li, Q. (2002). On instrumental variable estimation of semiparametric dynamic panel data models. *Economics Letters* 76: 1-9.

Baltagi, B.H., Levin, D. (1986). Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation and advertising reconsidered. *The Review of Economics and Statistics* 68: 148-155.

Berkowitz, D., Caner, M., Fang, Y. (2008). Are nearly exogenous instruments reliable? *Economics Letters* 101: 20-28.

Berkowitz, D., Caner, M., Fang, Y. (2012). The validity of instruments revisited. *Journal of Econometrics* 166: 255-266.

Cai, Z. (2010). Functional coefficient models for economic and financial data. In *Oxford Handbook of Functional Data Analysis (Eds: F. Ferraty and Y. Romain)*. Oxford University Press, Oxford, UK, pp.166-186.

Cai, Z., Chen, L., Fang, Y. (2012). A new forecasting model for USA/CNY exchange rate. *Studies in Nonlinear Dynamics & Econometrics* 16: issue 3.

Cai, Z., Das, M., Xiong, H., Wu, X. (2006).Functional-coefficient instrumental variables models. *Journal of Econometrics* 133: 207-241.

Cai, Z., Fan, J. (2000). Average regression surface for dependent data. *Journal of Multivariate Analysisi* 75: 112-142.

Cai, Z., Fan, J., Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* 95: 941-956.

Cai, Z., Fang, Y., Li, H. (2012). Weak instrumental variables models for longitudinal data. *Econometric Reviews* 31: 361-389.

Cai, Z., Fang, Y., Su, J. (2012). Reducing Asymptotic bias of weak instrumental estimation using independently repeated cross-sectional information. *Statistics & Probability Letters* 82: 180-185.

Cai, Z., Gu, J., Li, Q. (2009). Some recent developments on nonparametric econometrics. *Advances in Econometrics* 25: 495-549.

Cai, Z., Hong, Y. (2009). Some recent developments in nonparametric finance. *Advances in Econometrics* 25: 379-432.

Cai, Z., Li, Q. (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory* 24: 1321-1342.

Das, M. (2005). Instrumental variables estimators for nonparametric models with discrete endogenous regressors. *Journal of Econometrics* 124: 335-361.

Fan, J., Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* 11: 1031-1057.

Fang, Y., Ren, Y., Yuan, Y. (2011). Nonparametric estimation and testing of stochastic discount factor. *Finance Research Letters* forthcoming, 8:196-205.

Hahn, J. (1997). Efficient estimation of panel data models with sequential moment restrictions. *Journal of Econometrics* 79: 1-21.

Hahn, J. (1999). How informative is the initial condition in nthe dynamic panel model with fixed effects? *Journal of Econometrics* 93: 309-326.

Hastie, T.J., Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* 55: 757-796.

Henderson, D., Carroll, R., Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics* 144: 257-275.

Holtz-Eakin, D., Newey, W., Rosen, H.S. (1988). Estimating vector autoregressions with panel data. *Econometrica* 56: 1371-1395.

Hoover, D., Rice, J., Wu, C., Yang, L. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 85: 809-822.

Horowitz, J.L., Markatou, M. (1996). Semiparametric estimation of regression models for panel data. *Review of Economic Studies* 63: 145-168.

Hsiao, C. (2003). Analysis of panel data. Cambridge: Cambridge University Press.

Islam, N. (1995). Grwoth empirics: a panel data approach. *Quarterly Journal of Economics* 110: 1127-1170.

Lewbel, A. (2007). A local generalized method of moments estimator. *Economics Letters* 94: 124C128.

Li, Q., Huang, D., Li, Fu, F. (2002). Semiparametric smooth coefficient model. *Journal of Business and Economic Statistics* 71: 389-397.

Li, Q., Stengos, T. (1996). Semiparametric estimation of partially linear panel data models. *Journal of Econometrics* 71: 389-397.

Li, Q., Ullah, A. (1998). Estimating partially linear models with one-way error components. *Econometric Reviews* 17: 145-166.

Lin, X., Carroll, R.J. (2001). Semiparametric regression for clustered data using generalised estimation equations. *Journal of the American Statistical Association* 96: 1045-1056.

Lin, X., Carroll, R.J. (2006). Semiparametric estimation in general repeated measures problems. *Journal of the Royal Statistical Society, Series B* 68: 68-88.

Lin, S.C., Huang, H.C., Weng, H.W. (2006). A semiparametric partially linear investigation of the Kuznets' hypothesis. *Journal of Comparative Economics* 34: 634-647.

Lin, D., Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *Journal of the American Statistical Association* 96: 103C126.

Park, B.U., Sickles, R., Simar, L. (2007). Semiparametric efficient estimation of dynamic panel data models. *Journal of Econometrics* 136: 281-301.

Powell, J.L., Stock, J.H., Stocker, T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica* 51: 1403-1430.

Qian, J., Wang, L. (2011). Estimating semiparametric panel data models by marginal integration. *Journal of Econometrics* forthcoming, DOI: 10.1016/j.jeconom.2011.09.030.

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

Su, L., Ullah, A. (2006). Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters* 92: 75-81.

Zhou, X., Li, K.W. (2011). Inequality and development: evidence from semiparametric estimation with panel data. *Economics Letters* 113: 203-207.

Appendix A: Proofs of Propositions

It is clear that $\boldsymbol{\Phi}_N = \boldsymbol{\Omega}_N\boldsymbol{\theta} + \boldsymbol{\Phi}_N^* + \boldsymbol{\Psi}_N + \boldsymbol{\Lambda}_N$ and $\boldsymbol{T}_N = \tilde{\boldsymbol{S}}_N\boldsymbol{H}\boldsymbol{\delta} + \boldsymbol{T}_N^* + \boldsymbol{B}_N + \boldsymbol{R}_N$. Indeed,

$$\begin{aligned}
\boldsymbol{\Omega}_N\boldsymbol{\theta} + \boldsymbol{\Phi}_N^* + \boldsymbol{\Psi}_N + \boldsymbol{\Lambda}_N &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_1}(U_{it}-u_0)\boldsymbol{W}_{it}[\boldsymbol{X}_{it}'\boldsymbol{\theta} + \epsilon_{it} + \boldsymbol{X}_{it}'\boldsymbol{\theta}(U_{it}) - \boldsymbol{X}_{it}'\boldsymbol{\theta}] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_1}(U_{it}-u_0)\boldsymbol{W}_{it}Y_{it} \\
&= \boldsymbol{\Phi}_N
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\boldsymbol{S}}_N\boldsymbol{H}\boldsymbol{\delta} + \boldsymbol{T}_N^* + \boldsymbol{B}_N + \boldsymbol{R}_N &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_2}(U_{it}-u_0)\boldsymbol{Q}_{it}[\boldsymbol{P}_{it}'\boldsymbol{\delta} + \epsilon_{it} + \boldsymbol{X}_{it}'\boldsymbol{g}(U_{it}) - \boldsymbol{P}_{it}'\boldsymbol{\delta}] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_2}(U_{it}-u_0)\boldsymbol{Q}_{it}Y_{it} \\
&= \boldsymbol{T}_N.
\end{aligned}$$

Proof of Propositions 1(i) **and 2**(i). It is clear that

$$\begin{aligned}
& E[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{V}_{it}\boldsymbol{X}'_{it}(\frac{U_{it}-u_0}{h_1})^j K_{h_1}(U_{it}-u_0)] \\
= & E[\boldsymbol{V}_{it}\boldsymbol{X}'_{it}(\frac{U_{it}-u_0}{h_1})^j K_{h_1}(U_{it}-u_0)] \\
= & \int \boldsymbol{\Omega}(U_{it})(\frac{U_{it}-u_0}{h_1})^j K_{h_1}(U_{it}-u_0) f(U_{it}) dU_{it} \\
= & [\boldsymbol{\Omega}(u_0)+O(h_1)]\int u^j K(u)du[f(u_0)+O(h_1)] \\
= & \boldsymbol{\Omega}(u_0)f(u_0)\mu_j + O(h_1)
\end{aligned}$$

Hence, we have

$$E(\boldsymbol{\Omega}_N) = f(u_0)\boldsymbol{\Omega}[1+o(1)].$$

Now, we consider $\tilde{\boldsymbol{S}}_N$. Indeed,

$$\begin{aligned}
\tilde{\boldsymbol{S}}_N &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{Q}_{it}\tilde{\boldsymbol{P}}'_{it}K_{h_2}(U_{it}-u_0) \\
&= \begin{pmatrix} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}K_{h_2}(U_{it}-u_0) & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\frac{U_{it}-u_0}{h_2}K_{h_2}(U_{it}-u_0) \\ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\frac{U_{it}-u_0}{h_2}K_{h_2}(U_{it}-u_0) & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\frac{(U_{it}-u_0)^2}{h_2^2}K_{h_2}(U_{it}-u_0) \end{pmatrix}.
\end{aligned}$$

For any $j = 0, 1, 2$,

$$\begin{aligned}
& E[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}(\frac{U_{it}-u_0}{h_2})^j K_{h_2}(U_{it}-u_0)] \\
= & E[\boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}(\frac{U_{it}-u_0}{h_2})^j K_{h_2}(U_{it}-u_0)] \\
= & \int \tilde{\boldsymbol{\Omega}}(U_{it})(\frac{U_{it}-u_0}{h_2})^j K_{h_2}(U_{it}-u_0) f(U_{it}) dU_{it} \\
= & [\tilde{\boldsymbol{\Omega}}(u_0)+O(h_2)]\int u^j K(u)du[f(u_0)+O(h_2)] \\
= & \tilde{\boldsymbol{\Omega}}(u_0)f(u_0)\mu_j + O(h_2).
\end{aligned}$$

Hence, we have

$$E(\tilde{\boldsymbol{S}}_N) = f(u_0)\boldsymbol{S}[1+o(1)].$$

And, for $1 \le l \le q$ and $1 \le m \le d$, let

$$s_{N,lmj} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}W_{itl}X_{itm}(\frac{U_{it}-u_0}{h_N})^j K_h(U_{it}-u_0),$$

where $h = h_N = h_1$ or $h = h_N = h_2$, W_{itl} is the lth element of $\boldsymbol{W}_{it}$, and X_{itm} is the mth element of $\boldsymbol{X}_{it}$. Then, by the stationary assumptions, we have

$$\begin{aligned}
Var(s_{N,lmj}) &= Var[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} W_{itl}X_{itm}(\frac{U_{it}-u_0}{h_N})^j K_h(U_{it}-u_0)] \\
&= \frac{1}{NT^2} Var[\sum_{t=1}^{T} W_{itl}X_{itm}(\frac{U_{it}-u_0}{h_N})^j K_h(U_{it}-u_0)] \\
&= \frac{1}{NT} Var[W_{itl}X_{itm}(\frac{U_{it}-u_0}{h_N})^j K_h(U_{it}-u_0)] \\
&+ \frac{2}{NT^2}\sum_{t=2}^{T}(T-t+1)Cov(W_{i1l}X_{i1m}(\frac{U_{i1}-u_0}{h_N})^j K_h(U_{i1}-u_0), \\
&\quad W_{itl}X_{itm}(\frac{U_{it}-u_0}{h_N})^j K_h(U_{it}-u_0)) \\
&\equiv I_{s1} + I_{s2}.
\end{aligned}$$

By assumptions and Cauchy-Schwarz inequality, $I_{s1} \le \frac{C}{Nh_N}$ and $|I_{s2}| \le \frac{C}{N}$. Thus, $Var(s_{N,lmj}) \to 0$. It follows that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{V}_{it}\boldsymbol{X}'_{it}(\frac{U_{it}-u_0}{h_1})^j K_{h_1}(U_{it}-u_0) = \boldsymbol{\Omega}(u_0)f(u_0)\mu_j + O_p(h_1)$$

and

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}(\frac{U_{it}-u_0}{h_2})^j K_{h_2}(U_{it}-u_0) = \tilde{\boldsymbol{\Omega}}(u_0)f(u_0)\mu_j + O_p(h_2).$$

Therefore, we have

$$\boldsymbol{\Omega}_N = f(u_0)\boldsymbol{\Omega}[1+o_p(1)] \quad \text{and} \quad \tilde{\boldsymbol{S}}_N = f(u_0)\boldsymbol{S}[1+o_p(1)].$$

The proof is complete.

Proof of Propositions 1(ii) **and 2**(ii).

$$\begin{aligned}
\boldsymbol{\Psi}_N &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_1}(U_{it}-u_0)\boldsymbol{V}_{it}\sum_{j=1}^{d}[\dot{\theta}_j(u_0)(U_{it}-u_0) + \frac{1}{2}\ddot{\theta}_j(u_0)(U_{it}-u_0)^2]X_{itj} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_1}(U_{it}-u_0)\boldsymbol{V}_{it}\boldsymbol{X}'_{it}\dot{\boldsymbol{\theta}}(U_{it}-u_0) + \frac{1}{2NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_1}(U_{it}-u_0)\boldsymbol{V}_{it}\boldsymbol{X}'_{it}\ddot{\boldsymbol{\theta}}(U_{it}-u_0)^2 \\
&\equiv \boldsymbol{\Psi}_N^1 + \boldsymbol{\Psi}_N^2.
\end{aligned}$$

Then,

$$\begin{aligned}
E(\boldsymbol{\Psi}_N^1) &= h_1 E[\boldsymbol{V}_{it}\boldsymbol{X}'_{it}\frac{U_{it}-u_0}{h_1}K_{h_1}(U_{it}-u_0)]\dot{\boldsymbol{\theta}} \\
&= h_1\int \boldsymbol{\Omega}(U_{it})\frac{U_{it}-u_0}{h_1}K_{h_1}(U_{it}-u_0)f(U_{it})dU_{it}\dot{\boldsymbol{\theta}} \\
&= h_1\int [\boldsymbol{\Omega}(u_0)+\dot{\boldsymbol{\Omega}}(u_0)uh_1+o(h_1)]uK(u)[f(u_0)+\dot{f}(u_0)uh_1+o(h_1)]du\dot{\boldsymbol{\theta}} \\
&= h_1^2\mu_2[f(u_0)\dot{\boldsymbol{\Omega}}(u_0)+\dot{f}(u_0)\boldsymbol{\Omega}(u_0)]\dot{\boldsymbol{\theta}}+o(h_1^2),
\end{aligned}$$

and

$$\begin{aligned}
E(\boldsymbol{\Psi}_N^2) &= \frac{h_1^2}{2}E[\boldsymbol{V}_{it}(\boldsymbol{D}'_i,\boldsymbol{X}'_{it})(\frac{U_{it}-u_0}{h_1})^2K_{h_1}(U_{it}-u_0)]\ddot{\boldsymbol{\theta}} \\
&= \frac{h_1^2}{2}\int \boldsymbol{\Omega}(U_{it})(\frac{U_{it}-u_0}{h_1})^2K_{h_1}(U_{it}-u_0)f(U_{it})dU_{it}\ddot{\boldsymbol{\theta}} \\
&= \frac{h_1^2}{2}[\boldsymbol{\Omega}(u_0)+O(h_1)]\int u^2K(u)du[f(u_0)+O(h_1)]\ddot{\boldsymbol{\theta}} \\
&= \frac{h_1^2}{2}[\boldsymbol{\Omega}(u_0)\mu_2 f(u_0)+O(h_1)]\ddot{\boldsymbol{\theta}} \\
&= \frac{h_1^2}{2}f(u_0)\mu_2\boldsymbol{\Omega}\ddot{\boldsymbol{\theta}}+o(h_1^2).
\end{aligned}$$

Hence,

$$h_1^{-2}E(\boldsymbol{\Psi}_N)=\frac{1}{2}f(u_0)\mu_2[2(\dot{\boldsymbol{\Omega}}+\boldsymbol{\Omega}\frac{\dot{f}(u_0)}{f(u_0)})\dot{\boldsymbol{\theta}}+\boldsymbol{\Omega}\ddot{\boldsymbol{\theta}}]+o(1).$$

Now,

$$\begin{aligned}
\boldsymbol{B}_N &= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T K_{h_2}(U_{it}-u_0)\boldsymbol{Q}_{it}\frac{1}{2}\sum_{j=1}^{d_2}\ddot{\beta}_j(u_0)(U_{it}-u_0)^2X_{it,2j} \\
&= \frac{h_2^2}{2}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \boldsymbol{Q}_{it}\boldsymbol{X}'_{it,2}\ddot{\boldsymbol{\beta}}(u_0)(\frac{U_{it}-u_0}{h_2})^2K_{h_2}(U_{it}-u_0) \\
&= \frac{h_2^2}{2}\begin{pmatrix} \frac{1}{NT}\sum\limits_{i=1}^N\sum\limits_{t=1}^T \boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\ddot{\boldsymbol{\beta}}(u_0)(\frac{U_{it}-u_0}{h_2})^2K_{h_2}(U_{it}-u_0) \\ \frac{1}{NT}\sum\limits_{i=1}^N\sum\limits_{t=1}^T \boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\ddot{\boldsymbol{\beta}}(u_0)(\frac{U_{it}-u_0}{h_2})^3K_{h_2}(U_{it}-u_0) \end{pmatrix}.
\end{aligned}$$

For $j = 2, 3$,

$$
\begin{aligned}
& E[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\ddot{\boldsymbol{\beta}}(u_0)(\frac{U_{it}-u_0}{h_2})^j K_{h_2}(U_{it}-u_0)] \\
= \; & E[\boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\ddot{\boldsymbol{\beta}}(u_0)(\frac{U_{it}-u_0}{h_2})^j K_{h_2}(U_{it}-u_0)] \\
= \; & \int \tilde{\boldsymbol{\Omega}}(U_{it})(\frac{U_{it}-u_0}{h_2})^j K_{h_2}(U_{it}-u_0)f(U_{it})dU_{it}\ddot{\boldsymbol{\beta}}(u_0) \\
= \; & [\tilde{\boldsymbol{\Omega}}(u_0)+O(h_2)]\int u^j K(u)du[f(u_0)+O(h_2)]\ddot{\boldsymbol{\beta}}(u_0) \\
= \; & f(u_0)\mu_j\tilde{\boldsymbol{\Omega}}\ddot{\boldsymbol{\beta}}+O(h_2).
\end{aligned}
$$

Hence,

$$
h_2^{-2}E(\boldsymbol{B}_N) = \frac{1}{2}f(u_0)\begin{pmatrix} \mu_2\tilde{\boldsymbol{\Omega}}\ddot{\boldsymbol{\beta}} \\ 0 \end{pmatrix} + o(1).
$$

Similar to (i), any component of the variance of $h_1^{-2}\boldsymbol{\Psi}_N$ and $h_2^{-2}\boldsymbol{B}_N$ converges to zero. Therefore, we have

$$
\boldsymbol{\Psi}_N = \frac{h_1^2}{2}f(u_0)\mu_2[2(\dot{\boldsymbol{\Omega}}+\boldsymbol{\Omega}\frac{\dot{f}(u_0)}{f(u_0)})\dot{\boldsymbol{\theta}}+\boldsymbol{\Omega}\ddot{\boldsymbol{\theta}}]+o_p(h_1^2) \quad \text{and} \quad \boldsymbol{B}_N = \frac{h_2^2}{2}f(u_0)\begin{pmatrix} \mu_2\tilde{\boldsymbol{\Omega}}\ddot{\boldsymbol{\beta}} \\ 0 \end{pmatrix} + o_p(h_2^2).
$$

Therefore, this proves the results.

Proof of Propositions 1(iii) and 2(iii).

$$
\begin{aligned}
E(\boldsymbol{\Lambda}_N) &= E[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{h_1}(U_{it}-u_0)\boldsymbol{V}_{it}\boldsymbol{X}'_{it}\boldsymbol{\Lambda}(U_{it},u_0)] \\
&= E[\boldsymbol{V}_{it}\boldsymbol{X}'_{it}\boldsymbol{\Lambda}(U_{it},u_0)K_{h_1}(U_{it}-u_0)] \\
&= \int \boldsymbol{\Omega}(U_{it})\boldsymbol{\Lambda}(U_{it},u_0)K_{h_1}(U_{it}-u_0)f(U_{it})dU_{it} \\
&= [f(u_0)+O(h_1)]\int \boldsymbol{\Omega}(u_0)\boldsymbol{\Lambda}(u_0+uh_1,u_0)K(u)du.
\end{aligned}
$$

And, for any $1 \le j \le d$,

$$
\Lambda_j(u_0+uh_1,u_0) = \theta_j(u_0+uh_1)-\theta_j(u_0)-h_1\dot{\theta}_j(u_0)u-\frac{h_1^2}{2}\ddot{\theta}_j(u_0)u^2 = O(h_1^3).
$$

Therefore, $\boldsymbol{\Lambda}_N = o_p(h_1^2)$, and

$$
\begin{aligned}
\boldsymbol{R}_N &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{h_2}(U_{it}-u_0)\boldsymbol{Q}_{it}\frac{1}{2}\sum_{j=1}^{d_2} R_j(U_{it},u_0)X_{it,2j} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{Q}_{it}\boldsymbol{X}'_{it,2}\boldsymbol{R}(U_{it},u_0)K_{h_2}(U_{it}-u_0) \\
&= \begin{pmatrix} \frac{1}{NT}\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T} \boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\boldsymbol{R}(U_{it},u_0)K_{h_2}(U_{it}-u_0) \\ \frac{1}{NT}\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T} \boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\boldsymbol{R}(U_{it},u_0)\frac{U_{it}-u_0}{h_2}K_{h_2}(U_{it}-u_0) \end{pmatrix}.
\end{aligned}
$$

For any component in the above vector, $j = 0, 1$,

$$
\begin{aligned}
& E[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\boldsymbol{R}(U_{it},u_0)(\frac{U_{it}-u_0}{h_2})^j K_{h_2}(U_{it}-u_0)] \\
= \ & E[\boldsymbol{W}_{it}\boldsymbol{X}'_{it,2}\boldsymbol{R}(U_{it},u_0)(\frac{U_{it}-u_0}{h_2})^j K_{h_2}(U_{it}-u_0)] \\
= \ & \int \tilde{\boldsymbol{\Omega}}(U_{it})\boldsymbol{R}(U_{it},u_0)(\frac{U_{it}-u_0}{h_2})^j K_{h_2}(U_{it}-u_0)f(U_{it})dU_{it} \\
= \ & [f(u_0)+O_p(h_2)]\int \tilde{\boldsymbol{\Omega}}(u_0)\boldsymbol{R}(u_0+uh_2,u_0)u^j K(u)du.
\end{aligned}
$$

And, for any $1 \le j \le d_2$,

$$
R_j(u_0+uh_2,u_0) = \beta_j(u_0+uh_2) - \beta_j(u_0) - h_2\dot{\beta}_j(u_0)u - \frac{h_2^2}{2}\ddot{\beta}_j(u_0)u^2 = O(h_2^3).
$$

Therefore, $\boldsymbol{R}_N = o_p(h_2^2)$. This completes the proof.

Proof of Propositions 1(iv) **and 2**(iv). Under the above assumptions, we have

$$
\begin{aligned}
Nh_1Var(\boldsymbol{\Phi}_N^*) &= Nh_1Var[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{V}_{it}\epsilon_{it}K_{h_1}(U_{it}-u_0)] \\
&= \frac{h_1}{T^2}Var[\sum_{t=1}^{T}\boldsymbol{V}_{it}\epsilon_{it}K_{h_1}(U_{it}-u_0)] = \frac{h_1}{T}Var[\boldsymbol{V}_{it}\epsilon_{it}K_{h_1}(U_{it}-u_0)] \\
&+ \frac{2h_1}{T^2}\sum_{t=2}^{T}(T-t+1)Cov(\boldsymbol{V}_{i1}\epsilon_{i1}K_{h_1}(U_{i1}-u_0), \boldsymbol{V}_{it}\epsilon_{it}K_{h_1}(U_{it}-u_0)) \\
&= \frac{h_1}{T}E[\boldsymbol{V}_{it}\boldsymbol{V}'_{it}\epsilon_{it}^2K_{h_1}^2(U_{it}-u_0)] \\
&+ \frac{2h_1}{T^2}\sum_{t=2}^{T}(T-t+1)E[\boldsymbol{V}_{i1}\boldsymbol{V}'_{it}\epsilon_{i1}\epsilon_{it}K_{h_1}(U_{i1}-u_0)K_{h_1}(U_{it}-u_0)] \\
&\equiv \boldsymbol{I}_3 + \boldsymbol{I}_4, \qquad (18)
\end{aligned}
$$

and

$$
\begin{aligned}
Nh_2Var(\boldsymbol{T}_N^*) &= Nh_2Var\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{Q}_{it}\epsilon_{it}K_{h_2}(U_{it}-u_0)\} \\
&= \frac{h_2}{T^2}Var\{\sum_{t=1}^{T}\boldsymbol{Q}_{it}\epsilon_{it}K_{h_2}(U_{it}-u_0)\} = \frac{h_2}{T}Var\{\boldsymbol{Q}_{it}\epsilon_{it}K_{h_2}(U_{it}-u_0)\} \\
&+ \frac{2h_2}{T^2}\sum_{t=2}^{T}(T-t+1)Cov(\boldsymbol{Q}_{i1}\epsilon_{i1}K_{h_2}(U_{i1}-u_0),\boldsymbol{Q}_{it}\epsilon_{it}K_{h_2}(U_{it}-u_0)) \\
&\equiv \boldsymbol{I}_5+\boldsymbol{I}_6. \qquad (19)
\end{aligned}
$$

For the first term in (19),

$$
\begin{aligned}
&\frac{h_2}{T}Var[\boldsymbol{Q}_{it}\epsilon_{it}K_{h_2}(U_{it}-u_0)] = \frac{h_2}{T}E[\boldsymbol{Q}_{it}\boldsymbol{Q}'_{it}\epsilon_{it}^2K_{h_2}^2(U_{it}-u_0)] \\
= \ &\frac{h_2}{T}E\begin{pmatrix} \boldsymbol{W}_{it}\boldsymbol{W}'_{it}\epsilon_{it}^2K_{h_2}^2(U_{it}-u_0) & \boldsymbol{W}_{it}\boldsymbol{W}'_{it}\epsilon_{it}^2K_{h_2}^2(U_{it}-u_0)\frac{U_{it}-u_0}{h_2} \\ \boldsymbol{W}_{it}\boldsymbol{W}'_{it}\epsilon_{it}^2K_{h_2}^2(U_{it}-u_0)\frac{U_{it}-u_0}{h_2} & \boldsymbol{W}_{it}\boldsymbol{W}'_{it}\epsilon_{it}^2K_{h_2}^2(U_{it}-u_0)(\frac{U_{it}-u_0}{h_2})^2 \end{pmatrix}.
\end{aligned}
$$

For any component in the above matrix, $j=0,1,2$, $h=h_N=h_1$ or $h=h_N=h_2$,

$$
\begin{aligned}
&E[\boldsymbol{V}_{it}\boldsymbol{V}'_{it}\epsilon_{it}^2K_h^2(U_{it}-u_0)(\frac{U_{it}-u_0}{h_N})^j] \\
= &\int \boldsymbol{\Phi}(U_{it})K_h^2(U_{it}-u_0)(\frac{U_{it}-u_0}{h_N})^j f(U_{it})dU_{it} \\
= &\frac{1}{h_N}[f(u_0)+O(h_N)][\boldsymbol{\Phi}(u_0)+O(h_N)]\int K^2(u)u^j du \\
= &\frac{1}{h_N}[f(u_0)\boldsymbol{\Phi}(u_0)\nu_j+O(h_N)].
\end{aligned}
$$

Hence,

$$
\boldsymbol{I}_3 \to \frac{1}{T}f(u_0)\boldsymbol{\Phi} \text{ and } \boldsymbol{I}_5 \to \frac{1}{T}f(u_0)\boldsymbol{S}^*.
$$

For the second term in (19),

$$
\begin{aligned}
&\frac{2h_2}{T^2}\sum_{t=2}^{T}(T-t+1)Cov(\boldsymbol{Q}_{i1}\epsilon_{i1}K_{h_2}(U_{i1}-u_0),\boldsymbol{Q}_{it}\epsilon_{it}K_{h_2}(U_{it}-u_0)) \\
= \ &\frac{2h_2}{T^2}\sum_{t=2}^{T}(T-t+1)E[\boldsymbol{Q}_{i1}\boldsymbol{Q}'_{it}\epsilon_{i1}\epsilon_{it}K_{h_2}(U_{i1}-u_0)K_{h_2}(U_{it}-u_0)] \\
= \ &\frac{2h_2}{T^2}\sum_{t=2}^{T}(T-t+1)E\begin{pmatrix} \mathbb{W} & \mathbb{W}\frac{U_{it}-u_0}{h_2} \\ \mathbb{W}\frac{U_{i1}-u_0}{h_2} & \mathbb{W}\frac{U_{i1}-u_0}{h_2}\frac{U_{it}-u_0}{h_2} \end{pmatrix},
\end{aligned}
$$

where $\mathbb{W} = \boldsymbol{W}_{i1}\boldsymbol{W}'_{it}\epsilon_{i1}\epsilon_{it}K_{h_2}(U_{i1}-u_0)K_{h_2}(U_{it}-u_0)$. For any component in the above matrix, $j=0,1$ and $i=0,1$, $h=h_N=h_1$ or $h=h_N=h_2$,

$$\begin{aligned}
& E[\boldsymbol{V}_{i1}\boldsymbol{V}'_{it}\epsilon_{i1}\epsilon_{it}K_h(U_{i1}-u_0)K_h(U_{it}-u_0)(\frac{U_{i1}-u_0}{h_N})^i(\frac{U_{it}-u_0}{h_N})^j] \\
= \ & E[E(\boldsymbol{V}_{i1}\boldsymbol{V}'_{it}\epsilon_{i1}\epsilon_{it}|U_{i1},U_{it})K_h(U_{i1}-u_0)K_h(U_{it}-u_0)(\frac{U_{i1}-u_0}{h_N})^i(\frac{U_{it}-u_0}{h_N})^j] \\
= \ & E[\boldsymbol{G}_{1t}(U_{i1},U_{it})K_h(U_{i1}-u_0)K_h(U_{it}-u_0)(\frac{U_{i1}-u_0}{h_N})^i(\frac{U_{it}-u_0}{h_N})^j] \\
= \ & f(u_0,u_0)\boldsymbol{G}_{1t}(u_0,u_0)\nu_i\nu_j + O_p(h_N).
\end{aligned}$$

Hence,

$$\boldsymbol{I}_4 \to \frac{2h_1}{T^2}f(u_0)\sum_{t=2}^{T}(T-t+1)\boldsymbol{G}_{1t}(u_0,u_0) \text{ and } \boldsymbol{I}_6 \to \frac{2h_2}{T^2}f(u_0)\sum_{t=2}^{T}(T-t+1)\boldsymbol{G}^*_{1t}(u_0,u_0),$$

where $\boldsymbol{G}^*_{1t} = \boldsymbol{G}^*_{1t}(u_0,u_0) = \begin{pmatrix} \nu_0^2 e_2'\boldsymbol{G}_{1t}(u_0,u_0)e_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$. Therefore,

$$Nh_1 Var(\boldsymbol{\Phi}^*_N) \to \frac{1}{T}f(u_0)\boldsymbol{\Phi}$$

and

$$Nh_2 Var(\boldsymbol{T}^*_N) \to \frac{1}{T}f(u_0)\boldsymbol{S}^*.$$

Then, the proof is complete.

Appendix B: Useful Lemmas for Theorem 2

By (9), we know that for any U_{it},

$$(\hat{\boldsymbol{\theta}}(U_{it}) - \boldsymbol{\theta}(U_{it})) - bias_{\boldsymbol{\theta}} \simeq f^{-1}(U_{it})\boldsymbol{D}(U_{it})\frac{1}{NT}\sum_{j=1}^{N}\sum_{k=1}^{T}K_{h_1}(U_{jk}-U_{it})\boldsymbol{V}_{jk}\epsilon_{jk},$$

where $\boldsymbol{D}(U_{it}) = (\boldsymbol{\Omega}'(U_{it})\boldsymbol{\Omega}(U_{it}))^{-1}\boldsymbol{\Omega}'(U_{it})$. Denote ξ_i as the information set of individual i. Thus,

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_1'[(\hat{\boldsymbol{\theta}}(U_{it}) - \boldsymbol{\theta}(U_{it})) - bias_{\boldsymbol{\theta}}] \\
&\simeq \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{T} e_1' f^{-1}(U_{it})\boldsymbol{D}(U_{it})K_{h_1}(U_{jk} - U_{it})\boldsymbol{V}_{jk}\epsilon_{jk} \\
&= \frac{1}{N^2}\sum_{1\le i<j\le N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{k=1}^{T}[e_1' f^{-1}(U_{it})\boldsymbol{D}(U_{it})K_{h_1}(U_{jk} - U_{it})\boldsymbol{V}_{jk}\epsilon_{jk} \\
&+ e_1' f^{-1}(U_{jk})\boldsymbol{D}(U_{jk})K_{h_1}(U_{it} - U_{jk})\boldsymbol{V}_{it}\epsilon_{it}] \\
&= \frac{1}{N^2}\sum_{1\le i<j\le N}[\frac{1}{T}\sum_{t=1}^{T} e_1' f^{-1}(U_{it})\boldsymbol{D}(U_{it})\boldsymbol{A}(U_{it},\xi_j) + \frac{1}{T}\sum_{k=1}^{T} e_1' f^{-1}(U_{jk})\boldsymbol{D}(U_{jk})\boldsymbol{A}(U_{jk},\xi_i)] \\
&\equiv \frac{N-1}{2N}\mathbb{U}_N,
\end{aligned}$$

where $e_1' = (\boldsymbol{I}_{d_1}, \boldsymbol{0}_{d_1\times d_2})$, which is used to extract the parametric part from the estimates of nonparametric GMM procedure using local constant fitting scheme, $\boldsymbol{A}(U_{it},\xi_j) = \frac{1}{T}\sum\limits_{k=1}^{T} K_{h_1}(U_{jk} - U_{it})\boldsymbol{V}_{jk}\epsilon_{jk}$, and $\mathbb{U}_N = \frac{2}{N(N-1)}\sum\limits_{1\le i<j\le N} p_N(\xi_i,\xi_j)$ is a U-statistic with

$$p_N(\xi_i,\xi_j) = \frac{1}{T}\sum_{t=1}^{T} e_1' f^{-1}(U_{it})\boldsymbol{D}(U_{it})\boldsymbol{A}(U_{it},\xi_j) + \frac{1}{T}\sum_{k=1}^{T} e_1' f^{-1}(U_{jk})\boldsymbol{D}(U_{jk})\boldsymbol{A}(U_{jk},\xi_i).$$

Following Theorem 3.1 in Powell, Stock amd Stocker (1989), we define

$$r_N(\xi_i) = E[p_N(\xi_i,\xi_j)|\xi_i],$$

$$\theta_N = E[r_N(\xi_i)] = E[p_N(\xi_i,\xi_j)],$$

$$\hat{\mathbb{U}}_N = \theta_N + \frac{2}{N}\sum_{i=1}^{N}[r_N(\xi_i) - \theta_N].$$

In order to establish the asymptotic normality of $\hat{\mathbb{U}}_N$, the condition of Lemma 3.1 in Powell, Stock and Stoker (1989) should be satisfied. It is easy to prove that $E[||p_N(\xi_i,\xi_j)||^2] = O(h^{-1}) = O[N(Nh_1)^{-1}]$. Thus, we have$E[||p_N(\xi_i,\xi_j)||^2] = o(N)$ if and only if $Nh_1 \to \infty$ as $h_1 \to 0$.

Lemma 1: Under assumptions A1-A5,

$$r_N(\xi_i) = e_1'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{D}(U_{it})\boldsymbol{V}_{it}\epsilon_{it}[1 + o(1)].$$

Lemma 2: Under assumptions A1-A5,

$$\begin{aligned}
&(i) \quad E[r_N(\xi_i)] = 0, \\
&(ii) \quad Var[r_N(\xi_i)] = \frac{1}{T}\Sigma_\gamma[1 + o(1)].
\end{aligned}$$

The detailed proofs of the above three lemmas are given in Appendix C.

Appendix C: Proofs of Lemmas

Proof of Lemma 1. Firstly,

$$
\begin{aligned}
& E[f^{-1}(U_{jk})\boldsymbol{D}(U_{jk})\boldsymbol{A}(U_{jk},\xi_i)|\xi_i] \\
= \; & E[f^{-1}(U_{jk})\boldsymbol{D}(U_{jk})\frac{1}{T}\sum_{t=1}^{T}K_{h_1}(U_{it}-U_{jk})\boldsymbol{V}_{it}\epsilon_{it}|\xi_i] \\
= \; & \frac{1}{T}\sum_{t=1}^{T}E[f^{-1}(U_{jk})\boldsymbol{D}(U_{jk})K_{h_1}(U_{it}-U_{jk})|\xi_i]\boldsymbol{V}_{it}\epsilon_{it} \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\int f^{-1}(U_{jk})\boldsymbol{D}(U_{jk})K_{h_1}(U_{it}-U_{jk})f(U_{jk})dU_{jk}\boldsymbol{V}_{it}\epsilon_{it} \\
= \; & \frac{1}{T}\sum_{t=1}^{T}[\boldsymbol{D}(U_{it})+o(1)]\boldsymbol{V}_{it}\epsilon_{it} \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{D}(U_{it})\boldsymbol{V}_{it}\epsilon_{it}[1+o(1)]
\end{aligned}
$$

and

$$
E[\boldsymbol{A}(U_{it},\xi_j)|\xi_i] = E[\frac{1}{T}\sum_{k=1}^{T}K_{h_1}(U_{jk}-U_{it})\boldsymbol{V}_{jk}\epsilon_{jk}|\xi_i] = \frac{1}{T}\sum_{k=1}^{T}E[K_{h_1}(U_{jk}-U_{it})E(\boldsymbol{V}_{jt}\epsilon_{jt}|U_{jk})|\xi_i] = 0.
$$

By the definition of $p_N(\xi_i,\xi_j)$,

$$
\begin{aligned}
& E[p_N(v,\xi_j)|\xi_i] \\
= \; & E[\frac{1}{T}\sum_{t=1}^{T}e_1'f^{-1}(U_{it})\boldsymbol{D}(U_{it})\boldsymbol{A}(U_{it},\xi_j)+\frac{1}{T}\sum_{k=1}^{T}e_1'f^{-1}(U_{jk})\boldsymbol{D}(U_{jk})\boldsymbol{A}(U_{jk},\xi_i)|\xi_i] \\
= \; & \frac{1}{T}\sum_{t=1}^{T}e_1'f^{-1}(U_{it})\boldsymbol{D}(U_{it})E[\boldsymbol{A}(U_{it},\xi_j)|\xi_i]+\frac{1}{T}\sum_{k=1}^{T}e_1'E[f^{-1}(U_{jk})\boldsymbol{D}(U_{jk})\boldsymbol{A}(U_{jk},\xi_i)|\xi_i] \\
= \; & e_1'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{D}(U_{it})\boldsymbol{V}_{it}\epsilon_{it}[1+o(1)].
\end{aligned}
$$

The proof of the lemma is complete.

Proof of Lemma 2. It is easy to see that

$$
\begin{aligned}
E[r_N(\xi_i)] &= E\{E[p_N(\xi_i,\xi_j)|\xi_i]\} \\
&= E\{e_1'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{D}(U_{it})\boldsymbol{V}_{it}\epsilon_{it}[1+o_p(1)]\} \\
&= E\{e_1'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{D}(U_{it})E(\boldsymbol{V}_{it}\epsilon_{it}|U_{it})[1+o_p(1)]\} = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
& Var[r_N(\xi_i)] = E[r_N(\xi_i)]^2 \\
= \; & E[e_1'\frac{1}{T}\sum_{t_1=1}^{T}\boldsymbol{D}(U_{it_1})\boldsymbol{V}_{it_1}\epsilon_{it_1}\frac{1}{T}\sum_{t_2=1}^{T}\boldsymbol{V}'_{it_2}\epsilon_{it_2}\boldsymbol{D}'(U_{it_2})e_1][1+o(1)] \\
= \; & E[e_1'E(\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\boldsymbol{D}(U_{it_1})\boldsymbol{V}_{it_1}\epsilon_{it_1}\boldsymbol{V}_{it_2}\epsilon_{it_2}\boldsymbol{D}'(U_{it_2})|U_{it_1},U_{it_2})e_1][1+o(1)] \\
= \; & E\{e_1'[\frac{1}{T}\boldsymbol{D}(U_{it})\boldsymbol{\Phi}(U_{it})\boldsymbol{D}'(U_{it})+\frac{2}{T^2}\sum_{t=2}^{T}(T-t+1)\boldsymbol{D}(U_{i1})\boldsymbol{G_{1t}}(U_{i1},U_{it})\boldsymbol{D}'(U_{it})]e_1\}[1+o(1)] \\
\equiv \; & \frac{1}{T}\boldsymbol{\Sigma}_\gamma[1+o(1)].
\end{aligned}
$$

This concludes the proof of the lemma.

Appendix D: Proofs of Theorems

Proof of Theorem 1 and (15) in Theorem 3: By the assumptions, it is easy to see that $E(\boldsymbol{T}_N^*)=0$ and $E(\boldsymbol{\Phi}_N^*)=0$. Hence, the proofs are straightforward from Proposition 1(iv) and 2(iv), (9) and (14). This completes the proof of the theorems.

Proof of Theorem 2: Applying Theorem 3.1 in Powell, Stock and Stoker (1989), we have

$$\sqrt{N}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}-bias_{\boldsymbol{\gamma}}) \xrightarrow{D} N(0,\frac{1}{T}\boldsymbol{\Sigma}_\gamma).$$

This completes the proof of the theorem.

Proof of (16) in Theorem 3: We use the Cramer-Wold device to derive the asymptotic normality. Denote $\omega_{it}=(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{Q}_{it}\epsilon_{it}K_{h_2}(U_{it}-u_0)f^{-1}(u_0)$,$\omega_{N,it}=\sqrt{\frac{h_2}{T}}\boldsymbol{d}'\omega_{it}$, and $\omega_{N,i}^*=$

$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\omega_{N,it}}$, which only contains the information of individual i. Then,

$$\sqrt{Nh_2}\boldsymbol{d}'f^{-1}(u_0)(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{T}_N^* = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\omega_{N,it} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_{N,i}^*.$$

According to our model setting, the information between individuals are IID. Hence, $\omega_{N,i}^*$ series is an iid series. By Lindeberg-Lévy central limit theorem for iid case, the normality of $\sqrt{Nh_2}\boldsymbol{d}'\boldsymbol{T}_N^*$ is verified. This concludes the proof of the theorem.