ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

# Semiparametric Profile Likelihood Estimation of Varying Coefficient Models with Nonstationary Regressors

**Kunpeng Li, Degui Li, Zhongwen Liang and Cheng Hsiao**

**February 2013**

**Working Paper 02/13**

# Semiparametric Profile Likelihood Estimation of Varying Coefficient Models with Nonstationary Regressors

Kunpeng Li, School of Economcs & Management, Capital University of Economics & Business, Beijing, 100071, China

Degui Li, Department of Econometrics & Business Statistics, Monash University, Caulfield East, VIC 3145, Australia

Zhongwen Liang, Department of Economics, University at Albany, SUNY, Albany, NY 12222, U.S.A.

Cheng Hsiao[1], Department of Economics, University of Southern California , Los Angeles, CA 90089-0253, U.S.A.

**Abstract**

We study a partially linear varying coefficient model where the regressors are generated by the multivariate unit root I(1) processes. The influence of the explanatory vectors on the response variable satisfies the semiparametric partially linear structure with the nonlinear component being functional coefficients. The profile likelihood estimation methodology with the first-stage local polynomial smoothing is applied to estimate both the constant coefficients in the linear component and the functional coefficients in the nonlinear component. The asymptotic distribution theory for the proposed semiparametric estimators is established under some mild conditions, from which both the parametric and nonparametric estimators are shown to enjoy the well-known super-consistency property. Furthermore, a simulation study is conducted to investigate the finite sample performance of the developed methodology and results.

**JEL Classifications**: C13, C14, C22.

**Abbreviated Title**: Semi-Varying Coefficient Models

**Keywords**: Functional coefficients, local polynomial fitting, profile likelihood, semiparametric estimation, unit root process.

[1]The corresponding author: Cheng Hsiao, Email: chsiao@usc.edu

## 1. Introduction

In this paper, we are interested in a partially linear varying coefficient model defined by

$$Y_t = X_{1t}^\top \gamma + X_{2t}^\top \beta(Z_t) + u_t, \tag{1.1}$$

where $X_{1t}$ is a $d_1$-dimensional I(1) vector, $X_{2t}$ is a $d_2$-dimensional I(1) vector, $Z_t$ is a scalar stationary (or I(0)) variable, $u_t$ is a stationary error term, $\gamma$ is a $d_1 \times 1$ vector of constant parameters and $\beta(\cdot)$ is a $d_2$-dimensional vector of unspecified smooth functions. The notation "$\top$" denotes transpose of a vector (or matrix). Model (1.1) provides a very flexible framework in nonstationary time series analysis, and it covers various linear and nonlinear time series models with nonstationarity. For example, when $\beta(Z_t) \equiv \beta$, (1.1) reduces to a linear cointegration model which has been systematically investigated by existing literature such as Phillips (1986) and Saikkonen (1995). When $\gamma = 0$, (1.1) becomes a functional coefficient model, which has been studied by Cai *et al.* (2009), Xiao (2009) and Sun and Li (2011). The advantage of the functional coefficient structure in the nonparametric component of model (1.1) is that it could attenuate the so-called "curse of dimensionality" problem in nonparametric estimation when the dimension of the predictors is larger than three.

The main focus of this paper is to construct consistent estimators for both the parameter $\gamma$ and the functional coefficient $\beta(\cdot)$, and then derive their asymptotic theory. In Section 2 below, we will use a profile likelihood approach with first-stage local polynomial fitting to estimate the proposed model. In the case of independent or stationary weakly dependent observations, the profile likelihood methodology has been commonly used to estimate semiparametric models, see, for example, Fan and Huang (2005), Li *et al.* (2011) and the references therein. However, to the best of our knowledge, there is few work of extending such idea to the nonstationary time series case. Chen *et al.* (2012) considered the profile least squares estimation for the partially linear model through the null recurrent Markov chain framework. However, the model in Chen *et al.* (2012) is less general than model (1.1), and it is difficult to verify the null recurrent Markov property in practical applications. The main challenge of deriving the asymptotic theory in our model is the lack of uniform consistency results for the local polynomial estimators in the context of the functional coefficient models with nonstationarity. Hence, in this paper, we establish such uniform consistency result (see, for example, the argument in the proof of Proposition A.1), which is critical in our derivation and of independent interest. Under some mild conditions, we then establish the asymptotic distribution theory for the proposed semiparametric estimators. We show that the estimator for the parameter in the linear component enjoys the well-known super-consistency property with $n$-convergence rate. Such super-consistency result is fundamentally different from the parametric convergence rate in Chen *et al.* (2012), who could only derive the root-$n$ rate. Meanwhile, similar to Cai *et al.* (2009) and Xiao (2009), we can also show that the convergence rate for the nonparametric estimator is faster than the root-$nh$ rate which is common in the stationary case. Our results complement existing literature on nonparametric and semiparametric estimation for nonstationary time series (see, for example, Park and Hahn, 1999; Juhl and Xiao, 2005; Cai *et al.* , 2009; Wang and Phillips, 2009a, 2009b; Xiao 2009; Chen *et al*, 2010; Sun and Li, 2011; and Chen *et al.*, 2012). Furthermore, a simulation study is conducted to illustrate the finite sample performance of the proposed methodology as well as the super-consistency results.

The rest of this paper is organized as follows. The semiparametric profile likelihood estimation methodology is given in Section 2. The asymptotic theory for the proposed method is provided in Section 3. The simulation study is conducted in Section 4. The mathematical proofs of the asymptotic results are given in Appendix.

## 2. Semiparametric estimation

As mentioned above, when $(Y_t, X_t^\top, Z_t)$ is stationary with $X_t^\top = (X_{1t}^\top, X_{2t}^\top)$, the profile likelihood estimation methodology as well as the related asymptotic properties have been extensively studied for model (1.1), see, for example, Fan and Huang (2005). In this paper, we will extend such methodology to the nonstationary time series case, which is an important feature for economic data. To avoid confusion, throughout the paper, we let $\gamma_0$ and $\beta_0(\cdot)$ be the true parameters and functional coefficients.

Let $\mathbf{e} = \mathbf{e}_1 \otimes \mathbf{I}_{d_2}$, where $\mathbf{I}_{d_2}$ is a $d_2 \times d_2$ identity matrix, $\mathbf{e}_1 = (1, 0, \ldots, 0)$ is a $q+1$-dimensional row vector with all elements being zeros except that the first element is 1. Define

$$Z_{st,h} = (Z_s - Z_t)/h, \quad K_{h,st} = K(Z_{st,h}), \quad Q_{s,t} = \Big[1, (Z_s - Z_t), \cdots, (Z_s - Z_t)^q\Big]^\top,$$

and $\mathbf{G}_h = \mathrm{diag}\big(1, h, \cdots, h^q\big) \otimes \mathbf{I}_{d_2}$, where $h$ is a bandwidth and $K(\cdot)$ is a kernel function. We next adopt the local polynomial approach (Fan and Gijbels, 1996) to estimate the functional coefficient $\beta_0(\cdot)$ when $\gamma$ is given. Assuming that $\beta_0(\cdot)$ has $q$-th order continuous derivative ($q \geq 1$), we have the following Taylor expansion for the functional coefficient:

$$\beta_0(z) \approx \beta_0(z_0) + \beta_0'(z_0)(z - z_0) + \cdots + \beta_0^{(q)}(z_0)\frac{(z - z_0)^q}{q!}$$

for $z$ in a small neighborhood of $z_0$. Then, the local polynomial estimator of $\beta(Z_t)$ for given $\gamma$, is defined by

$$\begin{aligned}
\widetilde{\beta}(Z_t, \gamma) &= \mathbf{e}\Big[\sum_{s=1}^{n} K_{h,st} Q_{s,t} Q_{s,t}^\top \otimes X_{2s} X_{2s}^\top\Big]^{-1} \sum_{s=1}^{n} K_{h,st} Q_{s,t} \otimes X_{2s}(Y_s - X_{1s}^\top \gamma) \\
&= \mathbf{e}\mathbf{G}_h \Big[\sum_{s=1}^{n} K_{h,st} Q_{s,t} Q_{s,t}^\top \otimes X_{2s} X_{2s}^\top\Big]^{-1} \mathbf{G}_h \mathbf{G}_h^{-1} \sum_{s=1}^{n} K_{h,st} Q_{s,t} \otimes X_{2s}(Y_s - X_{1s}^\top \gamma) \\
&= \mathbf{e}\Big[n^{-2} \sum_{s=1}^{n} K_{h,st} \mathbf{G}_h^{-1} \big(Q_{s,t} Q_{s,t}^\top \otimes X_{2s} X_{2s}^\top\big) \mathbf{G}_h^{-1}\Big]^{-1} \Big[n^{-2} \sum_{s=1}^{n} K_{h,st} \mathbf{G}_h^{-1} (Q_{s,t} \otimes X_{2s})(Y_s - X_{1s}^\top \gamma)\Big] \\
&= \mathcal{A}_{2t} - \mathcal{A}_{1t}\gamma \qquad (2.1)
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{A}_{1t} &= \mathbf{e}\mathcal{S}_t^{-1}\Big[n^{-2} \sum_{s=1}^{n} K_{h,st} \mathbf{G}_h^{-1} (Q_{s,t} \otimes X_{2s}) X_{1s}^\top\Big], \\
\mathcal{A}_{2t} &= \mathbf{e}\mathcal{S}_t^{-1}\Big[n^{-2} \sum_{s=1}^{n} K_{h,st} \mathbf{G}_h^{-1} (Q_{s,t} \otimes X_{2s}) Y_s\Big], \\
\mathcal{S}_t &= n^{-2} \sum_{s=1}^{n} K_{h,st} \mathbf{G}_h^{-1} \big(Q_{s,t} Q_{s,t}^\top \otimes X_{2s} X_{2s}^\top\big) \mathbf{G}_h^{-1},
\end{aligned}$$

and in the second equality above we used the fact that $\mathbf{e}\mathbf{G}_h = \mathbf{e}$.

It is easy to see that $\widetilde{\beta}(Z_t, \gamma)$ can be seen as a function of the unknown parameter $\gamma$. Then, replacing $\beta(Z_t)$ by $\widetilde{\beta}(Z_t, \gamma)$ in model (1.1) and then applying ordinary least squares (OLS) method, we obtain the estimator of $\gamma_0$:

$$\widehat{\gamma} = \left[\sum_{t=1}^{n}(X_{1t} - \mathcal{A}_{1t}^{\top}X_{2t})(X_{1t} - \mathcal{A}_{1t}^{\top}X_{2t})^{\top}\right]^{-1}\sum_{t=1}^{n}(X_{1t} - \mathcal{A}_{1t}^{\top}X_{2t})(Y_t - X_{2t}^{\top}\mathcal{A}_{2t}). \tag{2.2}$$

With $\widehat{\gamma}$ replacing $\gamma$ in (2.1), we obtain a feasible local polynomial estimator of $\beta_0(z)$ by

$$\widehat{\beta}(z) = \mathcal{A}_2(z) - \mathcal{A}_1(z)\widehat{\gamma}, \tag{2.3}$$

where $\mathcal{A}_1(z)$ and $\mathcal{A}_2(z)$ are defined as $\mathcal{A}_{1t}$ and $\mathcal{A}_{2t}$ with $Z_t$ being replaced by $z$. The asymptotic properties of $\widehat{\gamma}$ and $\widehat{\beta}(z)$ will be given in Section 3 below.

## 3. Asymptotic theory

Before discussing the asymptotic distribution theory for both $\widehat{\gamma}$ and $\widehat{\beta}(z)$, we first introduce some regularity conditions. Let $X_{1t} = X_{1,t-1} + x_{1t}$ and $X_{2t} = X_{2,t-1} + x_{2t}$, where $x_{1t}$ and $x_{2t}$ are stationary and weakly dependent random vector processes which will be more specific later. Without loss of generality, we assume that $X_{10} = \mathbf{0}$ and $X_{20} = \mathbf{0}$. Hereafter, let $\|\cdot\| = \|\cdot\|_2$ denote the Euclidean norm.

**Assumption 1**. Let $w_t = (x_t^{\top},\ u_t)$ with $x_t^{\top} = (x_{1t}^{\top},\ x_{2t}^{\top})$. For some $p_1 > p_2 > 2$, $w_t^{\top} = (x_t^{\top}, u_t)$ is a strictly stationary, strongly mixing sequence with zero mean and mixing coefficients $\alpha_m = O\big(m^{-p_1p_2/(p_1-p_2)}\big)$ and $\mathrm{E}\big[\|w_t\|^{p_1}\big] < \infty$. In addition, there exists a positive definite matrix $\Omega$ such that $\frac{1}{n}\mathrm{E}\Big[(\sum_{t=1}^{n} w_t)(\sum_{t=1}^{n} w_t)^{\top}\Big] \to \Omega$.

**Assumption 2**. Let $(u_t, \mathcal{F}_{nt}, 1 \le t \le n)$ be a martingale difference sequence with $\mathrm{E}(u_t|\mathcal{F}_{n,t-1})$=0 a.s. and $E(u_t^2|\mathcal{F}_{n,t-1}) = \sigma_u^2$ a.s., where $\mathcal{F}_{nt} = \sigma\Big\{x_{s1}, Z_{s1}, u_{s2} :\ 1 \le s_1 \le n,\ 1 \le s_2 \le t\Big\}$.

**Assumption 3**. Let $Z_t$ be a strictly stationary and strongly mixing process with mixing coefficients $\alpha_m = O\big(m^{-p_1p_2/(p_1-p_2)}\big)$ and $\mathrm{E}\big[\|Z_t\|^{p_1}\big] < \infty$, where $p_1 > p_2 > 2$ are defined as in Assumption 1. In addition, $Z_t$ has a compact support $\mathcal{S}_Z$.

**Assumption 4**. The function $\beta(z)$ has $(q+1)$-th order continuous derivatives when $z$ is in the compact support of $Z_t$.

**Assumption 5**. The density function of $Z_t$, $f_Z(z)$, is continuous and positive and has second-order continuous derivative when $z$ is in the compact support of $Z_t$. Furthermore, the joint density function of $(Z_1, Z_{s+1})$, $f(u, v; s)$, is bounded for all $s \geq 1$.

**Assumption 6**. $K(\cdot)$ is continuous probability density function with a compact support.

**Assumption 7**. Let $nh^{q+1} \to 0$ and $(nh)/\log n \to \infty$ as $n \to \infty$.

Consider the partial sum process defined by $B_n(r) = n^{-1/2}\sum_{t=1}^{[nr]} w_t$ with $w_t$ being defined in Assumption 1 and $0 \leq r \leq 1$, where $[a]$ denotes the largest integer less than or equal to $a$. By Assumption 1 and the multivariate invariance principle for $B_n(r)$(Phillips and Durlauf, 1986), we have $B_n(r) \Rightarrow B(r)$, where $B(\cdot)$ is a multivariate Brownian motion with $\mathrm{E}\Big[B(1)B^\top(1)\Big] = \Omega$. We can further decompose $B_n(r)$ as $\Big[B_{1n}^\top(r), B_{2n}^\top(r), B_{3n}(r)\Big]^\top$, where $B_{1n}(r) = n^{-1/2}\sum_{t=1}^{[nr]} x_{1t}$, $B_{2n}(r) = n^{-1/2}\sum_{t=1}^{[nr]} x_{2t}$ and $B_{3n}(r) = n^{-1/2}\sum_{t=1}^{[nr]} u_t$. Then, we have $B_{jn}(r) \Rightarrow B_j(r)$ such that $B(r) = \Big[B_1^\top(r), B_2^\top(r), B_3(r)\Big]^\top$. The restriction of martingale differences on the error term $u_t$ in Assumption 2 is to facilitate our proofs, and it can be relaxed at the cost of more lengthy proofs. Assumptions 3–6 ensure some uniform consistency results derived by Masry (1996) and Hansen (2008), which are critical in our proofs and commonly used in the literature on nonparametric estimation such as Fan and Gijbels (1996) and Li and Racine (2007). The bandwidth condition $nh^{q+1} = o(1)$ is imposed to ensure the asymptotic bias of the local polynomial estimators is asymptotically negligible. When the local linear approach ($q = 1$) is applied, we can further relax such condition to $nh^2 = O(1)$.

Let $\mu_j = \int r^j K(r)dr$, $\Delta(\mu)$ be a $(q+1)\times(q+1)$ matrix with the $(i,j)$-th element being

$\mu_{i+j-2}$, and $\Gamma(\mu) = (\mu_0, \mu_1, \cdots, \mu_q)^\top$. Define

$$\Sigma_0 = \int \Big[B_1(r) - W^\top(B_1, B_2)B_2(r)\Big]^{\otimes 2} dr,$$

where $W(B_1, B_2) = \Big[\int_0^1 B_2(r)B_2^\top(r)dr\Big]^{-1} \int_0^1 B_2(r)B_1^\top(r)dr$, $B^{\otimes 2} = BB^\top$ for any matrix $B$ and $\mathbf{e}_1$ is defined as in Section 2. Define

$$\Sigma_1(r) = B_1(r) - W^\top(B_1, B_2)B_2(r).$$

The next theorem gives the asymptotic distribution of $\widehat{\gamma}$ defined in (2.2).

**Theorem 3.1**. *Suppose that Assumptions 1–7 are satisfied and $\Sigma_0$ is non-singular. Then, we have*

$$n(\widehat{\gamma} - \gamma_0) \Rightarrow \Sigma_0^{-1} \int_0^1 \Sigma_1(r)dB_3(r). \tag{3.1}$$

The above theorem shows that the estimator $\widehat{\gamma}$ enjoys the super-consistency property in the context of semi-varying coefficient cointegration models, which can be seen as an extension of some existing results for the parametric cointegration models, see, for example, Park and Phillips (2001). Also, compared with the result in Fan and Huang (2005), we can see that our result is consistent with the semiparametric efficiency property for the profile likelihood estimation of the semiparametric varying-coefficient partially linear models under the stationary setup.

By using Theorem 3.1, we can also derive the asymptotic distribution for $\widehat{\beta}(z)$ defined in (2.3). To simplify the presentation, we only consider the case of $d_2 = 1$. The extension to the case of $d_2 > 1$ is straightforward. Define

$$b_n(z) = \frac{h^{q+1}}{(q+1)!}\mathbf{e}\Delta^{-1}(\mu)(\mu_{q+1}, \cdots, \mu_{2q+1})^\top$$

and

$$\Sigma(z) = \frac{\sigma_u^2 \mathbf{e}\Delta^{-1}(\mu)\Delta(\nu)\Delta^{-1}(\mu)\mathbf{e}^\top}{f_Z(z)\int_0^1 B_2^2(r)dr},$$

where $\mathbf{e}$ is defined as that in Section 2 with $d_2$ replaced by 1 and $\Delta(\nu)$ is a $(q+1)\times(q+1)$ matrix with the $(i,j)$-th element being $\nu_{i+j-2}$, and $\nu_{i+j-2} = \int r^{i+j-2}K^2(r)dr$.

**Theorem 3.2**. *Suppose that the conditions of Theorem 3.1 are satisfied. Then, we have*

$$n\sqrt{h}\left[\widehat{\beta}(z) - \beta_0(z) - b_n(z)\right] \Rightarrow \mathrm{MN}(\Sigma_z), \tag{3.2}$$

*where* $\mathrm{MN}(\Sigma_z)$ *is a mixed normal distribution with zero mean and conditional covariance matrix* $\Sigma_z$*.*

By using the bandwidth condition in Assumption 7, we can further show that the asymptotic bias term in the above theorem is asymptotically negligible. Hence, the asymptotic distribution in (3.2) can be simplified to

$$n\sqrt{h}\left[\widehat{\beta}(z) - \beta_0(z)\right] \Rightarrow \mathrm{MN}(\Sigma_z).$$

We can find that the above convergence rate is faster than the root-$nh$ rate in stationary case, which is consistent with the findings in Cai *et al.* (2009) and Xiao (2009).

## 4. Simulation study

In this section, we give a simulated example to illustrate the proposed methodology and theory. Consider the model

$$Y_t = X_{1t}\gamma_0 + X_{2t}\beta_0(Z_t) + u_t, \quad t = 1, 2, \cdots, n, \tag{4.1}$$

where $\gamma_0 = 2$ and $\beta_0(z) = \sin(\pi z)$, $u_t \overset{i.i.d.}{\sim} \mathrm{N}(0, 0.5^2)$, and $\{Z_t\}$ is generated by the AR(1) model:

$$Z_t = 0.5Z_{t-1} + z_t \ \text{ with } \ z_t \overset{i.i.d.}{\sim} \mathrm{N}(0, 0.5^2),$$

and $\{z_t\}$ is independent of $\{u_t\}$. It is easy to check that $\{Z_t\}$ is stationary and $\alpha$-mixing dependent with geometric decaying coefficient. For the generation of $\{X_t\}$ with $X_t = (X_{1t}, X_{2t})^\top$, we consider the following two cases:

(i) $\{X_t\}$ is generated by $X_t = 0.5X_{t-1}+x_t$, where $x_t = (x_{1t}, x_{2t})^\top \overset{i.i.d.}{\sim} \mathrm{N}\left((0,0)^\top, \mathrm{diag}(1,1)\right)$.

(ii) $\{X_t\}$ is generated by $X_t = X_{t-1} + x_t$, where $x_t$ is generated as those in Case (i).

It is easy to show that $\{X_t\}$ defined in Case (i) is stationary and $\alpha$-mixing dependent, whereas $\{X_t\}$ defined in Case (ii) is nonstationary I(1). In this simulation, we consider the sample size $n = 300$ and 600 with replication number $N = 200$. For simplicity, we use the local linear smoother (corresponds to the local polynomial smoother with $q = 1$) to estimate the coefficient function $\beta_0(\cdot)$ with the standard normal kernel function.

To investigate the performance of the proposed semiparametric estimation methods for the above two cases, we apply the measurement of mean squared errors for both the parametric and nonparametric estimators. Let

$$\mathrm{MSE}(\gamma) = \frac{1}{N}\sum_{j=1}^{N}\left[\widehat{\gamma}(j) - \gamma_0\right]^2 \tag{4.2}$$

and

$$\mathrm{MSE}(\beta) = \frac{1}{N}\sum_{j=1}^{N}\mathrm{MSE}_j(\beta), \quad \mathrm{MSE}_j(\beta) = \frac{1}{n}\sum_{t=1}^{n}\left[\widehat{\beta}(Z_t, j) - \beta_0(Z_t)\right]^2, \tag{4.3}$$

where $\widehat{\gamma}(j)$ and $\widehat{\beta}(\cdot, j)$ are the resulting parametric and nonparametric estimators in the $j$-th simulation. The simulation results are reported in Table 1. From the table, we have the following conclusions: (1) the performance of the semiparametric estimators improve as the sample size increases, and the convergence rate of the parametric estimator is faster than that of the nonparametric estimator; (2) both the parametric and nonparametric estimators in Case (ii) outperform those in Case (i), which is a strong evidence of the existence of super-consistency results.

**Table 1.** *Means and standard errors (SE) of MSEs of the estimators*

| | n=300 | | n=600 | |
|---|---|---|---|---|
| | Case (i) | Case (ii) | Case (i) | Case (ii) |
| MSE($\gamma$) | 7.7801($\times 10^{-4}$) | 3.6322 ($\times 10^{-5}$) | 3.2147 ($\times 10^{-4}$) | 5.9910 ($\times 10^{-6}$) |
| SE($\gamma$) | 10.676($\times 10^{-4}$) | 7.5918 ($\times 10^{-5}$) | 4.1579 ($\times 10^{-4}$) | 9.9302 ($\times 10^{-6}$) |
| MSE($\beta$) | 6.4643($\times 10^{-2}$) | 1.1077 ($\times 10^{-2}$) | 3.6157 ($\times 10^{-2}$) | 3.1700 ($\times 10^{-3}$) |
| SE($\beta$) | 1.4846($\times 10^{-2}$) | 1.5852 ($\times 10^{-3}$) | 6.9910 ($\times 10^{-3}$) | 6.3283 ($\times 10^{-3}$) |

## 5. Acknowledgement

The second author was financially supported by the Australian Research Council Discovery Early Career Researcher Award (DE120101130), the Monash Researcher Accelerator Plan, and the DECRA Support Grant from Faculty of Business and Economics at Monash University.

## Appendix: Proofs of the main results

In this appendix, we give the proofs of the theoretical results given in Section 3. Let

$$M_\beta(z) = \left[\beta(z), \beta'(z), \cdots, \frac{\beta^{(q)}(z)}{q!}\right]^\top,$$

and $\widehat{M_\beta}(z)$ be the $q$-order local polynomial estimated value of $M_\beta(z)$.

**Proof of Theorem 3.1**. Note that

$$
\begin{aligned}
\hat{\gamma}-\gamma_0 &= \Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})^{\top}\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})(Y_t-X_{2t}^{\top}\mathcal{A}_{2t})-\gamma_0 \\
&= \Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})^{\top}\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})(Y_t-X_{1t}^{\top}\gamma_0) \\
&\quad+\Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})^{\top}\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})X_{2t}^{\top}\big(\mathcal{A}_{1t}\gamma_0-\mathcal{A}_{2t}\big) \\
&= \Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})^{\top}\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})u_t \\
&\quad-\Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})^{\top}\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})\widetilde{u}_t \\
&\quad+\Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})^{\top}\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})X_{2t}^{\top}\big[\beta_0(Z_t)-\overline{\beta}(Z_t)\big],
\end{aligned}
$$

where $\widetilde{u}_t=\mathbf{e}\mathcal{S}_t^{-1}\Big[n^{-2}\sum_{s=1}^{n}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})u_s\Big]$, and $\overline{\beta}(Z_t)=\mathbf{e}\overline{M}_{\beta}(Z_t)$ with $\overline{M}_{\beta}(Z_t)=\mathcal{S}_t^{-1}\Big[n^{-2}\sum_{s=1}^{n}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})X_{2s}^{\top}\beta_0(Z_s)\Big]$.

To further simplify the presentation, we define

$$
\begin{aligned}
\mathcal{B}_{1n} &= \frac{1}{n^2}\sum_{t=1}^{n}\big(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t}\big)\big(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t}\big)^{\top}, \\
\mathcal{B}_{2n} &= \frac{1}{n}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})u_t, \\
\mathcal{B}_{3n} &= \frac{1}{n}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})X_{2t}^{\top}\widetilde{u}_t, \\
\mathcal{B}_{4n} &= \frac{1}{n}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})X_{2t}^{\top}\big[\beta_0(Z_t)-\overline{\beta}(Z_t)\big].
\end{aligned}
$$

Then, Theorem 3.1 can be proved through the following three propositions. ■

**Proposition A.1**. *Under the conditions of Theorem 3.1, we have*

$$
\mathcal{B}_{1n}\Rightarrow\int\Big[B_1(r)-W^{\top}(B_1,B_2)B_2(r)\Big]^{\otimes 2}dr=\Sigma_0, \tag{A.1}
$$

*where $W(B_1,B_2)$ is defined in Section 3.*

**Proof**. Let $\mathbf{K}_{h,sz}(z) = \frac{1}{h} K_{h,sz} Q_{s,z} Q_{s,z}^{\top}$, where

$$Z_{sz,h} = (Z_s - z)/h, \quad K_{h,sz} = K(Z_{sz,h}), \quad Q_{s,z} = \Big[1, (Z_s - z), \cdots, (Z_s - z)^q\Big]^{\top}.$$

We have

$$\begin{aligned}
\frac{1}{h}\mathcal{S}_{tz} \quad &=: \quad \mathbf{G}_h^{-1}\Big[n^{-2}\sum_{s=1}^{n} \mathbf{K}_{h,sz}(z) \otimes X_{2s}X_{2s}^{\top}\Big]\mathbf{G}_h^{-1} \\
&= \quad \mathbf{G}_h^{-1}\Big\{n^{-2}\sum_{s=1}^{n} \mathrm{E}\big[\mathbf{K}_{h,sz}(z)\big] \otimes X_{2s}X_{2s}^{\top}\Big\}\mathbf{G}_h^{-1} \\
&\quad\; + \mathbf{G}_h^{-1}\Big(n^{-2}\sum_{s=1}^{n} \big\{\mathbf{K}_{h,sz}(z) - \mathrm{E}\big[\mathbf{K}_{h,sz}(z)\big]\big\} \otimes X_{2s}X_{2s}^{\top}\Big)\mathbf{G}_h^{-1}. \\
&=: \quad \mathcal{S}_{1t,z} + \mathcal{S}_{2t,z},
\end{aligned}$$

where $\mathcal{S}_{1t,z} = \mathbf{G}_h^{-1}\Big\{n^{-2}\sum_{s=1}^{n} \mathrm{E}\big[\mathbf{K}_{h,sz}(z)\big] \otimes X_{2s}X_{2s}^{\top}\Big\}\mathbf{G}_h^{-1}$, $\mathcal{S}_{2t,z} = \mathbf{G}_h^{-1}\Big(n^{-2}\sum_{s=1}^{n} \eta_{h,sz}(z) \otimes X_{2s}X_{2s}^{\top}\Big)\mathbf{G}_h^{-1}$, and $\eta_{h,sz}(z) = \mathbf{K}_{h,sz}(z) - \mathrm{E}\big[\mathbf{K}_{h,sz}(z)\big]$. By Assumptions 5 and 6, we have, uniformly in $z \in \mathcal{S}_Z$,

$$\mathrm{E}\big[\mathbf{K}_{h,sz}(z)\big] = f_Z(z)\Delta(\mu) + o(1),$$

where $\Delta(\mu)$ is a $(q+1)\times(q+1)$ matrix with the $(i,j)$-th element being $\mu_{i+j-2}$. On the other hand, by Assumption 1, we can prove that

$$n^{-2}\sum_{s=1}^{n} X_{2s}X_{2s}^{\top} = n^{-1}\sum_{s=1}^{n} \frac{X_{2s}}{\sqrt{n}} \cdot \frac{X_{2s}^{\top}}{\sqrt{n}} \Rightarrow \int_0^1 B_2(r)B_2^{\top}(r)dr = O_P(1).$$

Hence, we have, uniformly in $z \in \mathcal{S}_Z$,

$$\frac{1}{f_Z(z)}\mathcal{S}_{1t,z} - \Delta(\mu) \otimes \Big(n^{-2}\sum_{s=1}^{n} X_{2s}X_{2s}^{\top}\Big) = o_P(1). \tag{A.2}$$

We next prove that $\mathcal{S}_{2t,z}$ is $o_P(1)$ uniformly for $z \in \mathcal{S}_Z$. Let

$$Q_{s,z}^{*} = \Big[1, \frac{Z_s - z}{h}, \cdots, \frac{(Z_s - z)^q}{h^q}\Big]^{\top}$$

and $\eta_{h,sz}^{*}(z)$ be defined as $\eta_{h,sz}(z)$ with $Q_{s,z}$ replaced by $Q_{s,z}^{*}$. Then, we can show that $\mathcal{S}_{2t,z} = n^{-2}\sum_{s=1}^{n} \eta_{h,sz}^{*}(z) \otimes X_{2s}X_{2s}^{\top}$. As in Theorem 1 of Masry (1996), we can prove that

$$\sup_{l\geq 0}\sup_{z\in\mathcal{S}_Z} \mathrm{Var}\Big[\sum_{s=l+1}^{l+m} \eta_{h,sz}^{*}(z)\Big] = O\big(\frac{m}{h}\big)$$

for all $m \geq 1$. For some $0 < \delta < 1$, set $N = [1/\delta]$, $s_k = [kn/N] + 1$, $s_k^* = s_{k+1} - 1$, and $s_k^{**} = \min\{s_k^*, n\}$. Let $U_{n,s} = X_{2s}X_{2s}^\top/n$ for any $1 \leq s \leq n$ and $U_n(r) = U_{n,[nr]}$ for any $r \in [0,1]$. Following the proof of Theorem 3.3 of Hansen (1992), we have

$$\begin{aligned}
\sup_{z\in\mathcal{S}_Z} \|\mathcal{S}_{2t,z}\| &= \sup_{z\in\mathcal{S}_Z} \Big\| \frac{1}{n}\sum_{s=1}^{n} \eta_{h,sz}^*(z) \otimes U_{n,s} \Big\| = \sup_{z\in\mathcal{S}_Z} \Big\| \frac{1}{n}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s_k^{**}} \eta_{h,sz}^*(z) \otimes U_{n,s} \Big\| \\
&\leq \sup_{z\in\mathcal{S}_Z} \Big\| \frac{1}{n}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s_k^{**}} \eta_{h,sz}^*(z) \otimes U_{n,s_k} \Big\| + \sup_{z\in\mathcal{S}_Z} \Big\| \frac{1}{n}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s_k^{**}} \eta_{h,sz}^*(z) \otimes \big(U_{n,s} - U_{n,s_k}\big) \Big\| \\
&\leq \sup_{z\in\mathcal{S}_Z} \frac{1}{n}\sum_{k=0}^{N-1} \Big\| \sum_{s=s_k}^{s_k^{**}} \eta_{h,sz}^*(z) \Big\| \cdot \big\|U_{n,s_k}\big\| + \sup_{z\in\mathcal{S}_Z} \frac{1}{n}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s_k^{**}} \big\|\eta_{h,sz}^*(z)\big\| \cdot \big\|U_{n,s} - U_{n,s_k}\big\| \\
&\leq \frac{1}{n}\sum_{k=0}^{N-1} \sup_{z\in\mathcal{S}_Z} \Big\| \sum_{s=s_k}^{s_k^{**}} \eta_{h,sz}^*(z) \Big\| \sup_{0\leq r\leq 1} \|U_n(r)\| \\
&\quad + \sup_{|r_1-r_2|\leq\delta} \|U_n(r_1) - U_n(r_2)\| \cdot \sup_{z\in\mathcal{S}_Z} \frac{1}{n}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s_k^{**}} \big\|\eta_{h,sz}^*(z)\big\| \\
&=: \mathcal{S}_{n,21} + \mathcal{S}_{n,22}.
\end{aligned}$$

Note that $\sup_{0\leq r\leq 1} \|U_n(r)\| = O_p(1)$ as $U_n(r) \Rightarrow B_2(r)B_2^\top(r)$ by Assumption 1. Furthermore, following the argument in the proof of Theorem in Masry (1996), we have

$$\begin{aligned}
&\frac{1}{n}\sum_{k=0}^{N-1} \sup_{z\in\mathcal{S}_Z} \Big\| \sum_{s=s_k}^{s_k^*} \eta_{h,sz}(z) \Big\| \leq \frac{N}{n} \sup_{0\leq k\leq N-1} \sup_{z\in\mathcal{S}_Z} \Big\| \sum_{s=s_k}^{s_k^*} \eta_{h,sz}(z) \Big\| \\
&\leq \sup_{1\leq s\leq n} \sup_{z\in\mathcal{S}_Z} \Big\| \frac{1}{\delta n} \sum_{i=s}^{s+\delta n} \eta_{h,sz}(z) \Big\| = O_P\left((\delta n h)^{-1/2}\right) = o_P(1),
\end{aligned}$$

which implies that

$$\mathcal{S}_{n,21} = \sup_{0\leq r\leq 1} \|U_n(r)\| \cdot o_P(1) = o_P(1). \tag{A.3}$$

It is easy to see that uniformly for $z \in \mathcal{S}_Z$,

$$\frac{1}{n}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s_k^{**}} \big\|\eta_{h,sz}(z)\big\| = O_P(1),$$

which implies that

$$\mathcal{S}_{n,22} = \sup_{|r_1-r_2|\leq\delta} \Big\|U_n(r_1) - U_n(r_2)\Big\| \cdot O_P(1) = o_P(1) \tag{A.4}$$

by letting $\delta \to 0$.

By using (A.3) and (A.4), we have shown that $\mathcal{S}_{2t,z} = o_P(1)$ uniformly in $z \in \mathcal{S}_Z$, which, together with (A.2), leads to

$$\frac{1}{hf_Z(Z_t)}\mathcal{S}_t - \Delta(\mu) \otimes \left(n^{-2}\sum_{s=1}^{n} X_{2s}X_{2s}^\top\right) = o_P(1) \tag{A.5}$$

Similarly, we can also prove that, for any $t = 1, \cdots, n$,

$$\frac{1}{n^2hf_Z(Z_t)}\sum_{s=1}^{n} K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t} \otimes X_{2s})X_{1s}^\top - \Gamma(\mu) \otimes \left(n^{-2}\sum_{s=1}^{n} X_{2s}X_{1s}^\top\right) = o_P(1), \tag{A.6}$$

where $\Gamma(\mu) = (\mu_0, \mu_1, \cdots, \mu_q)^\top$.

Then, by the definition of $\mathcal{A}_{1t}$, we can show that

$$\mathcal{A}_{1,t} - W_n(B_1, B_2) = o_P(1), \quad \text{uniformly in } t = 1, \cdots, n, \tag{A.7}$$

where $W_n(B_1, B_2) = \left(\mathbf{e}_1\Delta(\mu)^{-1}\Gamma(\mu)\right) \otimes \left[\left(n^{-2}\sum_{s=1}^{n} X_{2s}X_{2s}^\top\right)^{-1}\left(n^{-2}\sum_{s=1}^{n} X_{2s}X_{1s}^\top\right)\right]$. By standard algebraic calculation, we have $\mathbf{e}_1\Delta(\mu)^{-1}\Gamma(\mu) \equiv 1$. Noting that $W_n(B_1, B_2) \Rightarrow W(B_1, B_2)$, by the definition of $\mathcal{B}_{1n}$ and (A.7), we can complete the proof of Proposition A.1. ■

**Proposition A.2**. *Under the conditions of Theorem 3.1, we have $\mathcal{B}_{4,n} = O_P(nh^{q+1})$.*

**Proof**. For $\overline{M}_\beta(Z_t) = \mathcal{S}_t^{-1}\left[n^{-2}\sum_{s=1}^{n} K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t} \otimes X_{2s})X_{2s}^\top\beta_0(Z_s)\right]$ and

$$M_\beta(Z_t) = \left[\beta(Z_t), \beta'(Z_t), \cdots, \frac{\beta^{(q)}(Z_t)}{q!}\right]^\top,$$

we have that

$$\begin{aligned}
&\left\|\overline{M}_\beta(Z_t) - M_\beta(Z_t)\right\| \\
= &\left\|S_t^{-1}\left[n^{-2}\sum_{s=1}^{n} K_{h,st}G_h^{-1}(Q_{s,t} \otimes X_{2s})X_{2s}^\top\left(\beta_0(Z_s) - \sum_{i=0}^{q}\frac{\beta_0^{(i)}(Z_t)}{i!}(Z_s - Z_t)^i\right)\right]\right\| \\
\le &\|hS_t^{-1}\|n^{-2}\sum_{s=1}^{n}\|G_h^{-1}(Q_{s,t} \otimes X_{2s})\|\|X_{2s}^\top\|\left\|\left(\beta_0(Z_s) - \sum_{i=0}^{q}\frac{\beta_0^{(i)}(Z_t)}{i!}(Z_s - Z_t)^i\right)h^{-1}K_{h,st}\right\| \\
= &O_p(h^{q+1}),
\end{aligned}$$

where we use

$$E\left\|\left(\beta_0(Z_s)-\sum_{i=0}^{q}\frac{\beta_0^{(i)}(Z_t)}{i!}(Z_s-Z_t)^i\right)h^{-1}K_{h,st}\right\|=O(h^{q+1})$$

in the last equality. Thus, we obtain that

$$\overline{M}_\beta(Z_t)-M_\beta(Z_t)=O_P(h^{q+1}) \tag{A.8}$$

uniformly for $t=1,\cdots,n$. Thus, we can further prove that

$$\sup_{1\le t\le n}\left\|\beta_0(Z_t)-\overline{\beta}(Z_t)\right\|=\sup_{1\le t\le n}\left\|\mathbf{e}M_\beta(Z_t)-\mathbf{e}\overline{M}_\beta(Z_t)\right\|=O_P(h^{q+1}). \tag{A.9}$$

Then, we have

$$\begin{aligned}\|\mathcal{B}_{4n}\| &\le \sup_{1\le t\le n}\left\|\beta_0(Z_t)-\overline{\beta}(Z_t)\right\|\frac{1}{n}\sum_{t=1}^{n}\left\|X_{1t}-\mathcal{A}_{1t}^\top X_{2t}\right\|\cdot\left\|X_{2t}\right\| \\ &= O_P(h^{q+1})\cdot O_P(n)=O_P(nh^{q+1}),\end{aligned}$$

as $\left\|X_{1t}-\mathcal{A}_{1t}^\top X_{2t}\right\|+\left\|X_{2t}\right\|=O_P(n^{1/2})$. We then complete the proof of Proposition A.2. ■

**Proposition A.3**. *Under the conditions of Theorem 3.1, we have*

$$\mathcal{B}_{2n}-\mathcal{B}_{3n}\Rightarrow\int_0^1\left[B_1(r)-W^\top(B_1,B_2)B_2(r)\right]dB_3(r), \tag{A.10}$$

*where $W(B_1,B_2)$ is defined in Section 3.*

**Proof**. Observe that

$$\begin{aligned}\mathcal{B}_{3n} &= \frac{1}{n}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})X_{2t}^\top\widetilde{u}_t, \\ &= \frac{1}{n}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})X_{2t}^\top\Big\{\mathbf{e}\mathcal{S}_t^{-1}\Big[\frac{1}{n^2}\sum_{s=1}^{n}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})u_s\Big]\Big\} \\ &= \frac{1}{n}\sum_{s=1}^{n}\Big[\frac{1}{n^2}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})X_{2t}^\top\left(\mathbf{e}\mathcal{S}_t^{-1}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})\right)\Big]u_s. \end{aligned} \tag{A.11}$$

Similar to the proof of Proposition A.1, we can prove that, uniformly in $s = 1, \cdots, n$,

$$
\begin{aligned}
& \frac{1}{n^2}\sum_{t=1}^{n}(X_{1t} - \mathcal{A}_{1t}^{\top}X_{2t})X_{2t}^{\top}\Big[\mathbf{e}\mathcal{S}_t^{-1}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes\frac{X_{2s}}{\sqrt{n}})\Big] \\
= \quad & \frac{1}{n^2}\sum_{t=1}^{n}\big[X_{1t} - W_n^{\top}(B_1,B_2)X_{2t}\big]X_{2t}^{\top}\Big[\mathbf{e}\mathcal{S}_t^{-1}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes\frac{X_{2s}}{\sqrt{n}})\Big] + o_P(1) \\
= \quad & \frac{1}{n^2}\sum_{t=1}^{n}\big[X_{1t} - W_n^{\top}(B_1,B_2)X_{2t}\big]X_{2t}^{\top}\Big\{\mathbf{e}\Big[\Delta(\mu)\otimes\Big(n^{-2}\sum_{t=1}^{n}X_{2t}X_{2t}^{\top}\Big)\Big]^{-1}(\Gamma(\mu)\otimes\frac{X_{2s}}{\sqrt{n}})\Big\} + o_P(1).
\end{aligned}
\tag{A.12}
$$

Let

$$
V_{ns} = \frac{1}{n^2}\sum_{t=1}^{n}\big[X_{1t} - W_n^{\top}(B_1,B_2)X_{2t}\big]X_{2t}^{\top}\Big\{\mathbf{e}\Big[\Delta(\mu)\otimes\Big(n^{-2}\sum_{t=1}^{n}X_{2t}X_{2t}^{\top}\Big)\Big]^{-1}(\Gamma(\mu)\otimes\frac{X_{2s}}{\sqrt{n}})\Big\},
$$

and

$$
\Theta_{ns} = \frac{1}{n^2}\sum_{t=1}^{n}(X_{1t} - \mathcal{A}_{1t}^{\top}X_{2t})X_{2t}^{\top}\Big[\mathbf{e}\mathcal{S}_t^{-1}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes\frac{X_{2s}}{\sqrt{n}})\Big] - V_{ns}.
$$

By (A.11) and (A.12), we can write that

$$
\mathcal{B}_{3n} = \frac{1}{\sqrt{n}}\sum_{s=1}^{n}V_{ns}u_s + \frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Theta_{ns}u_s, \tag{A.13}
$$

where $\Theta_{ns} = o_P(1)$ uniformly in $s = 1, \cdots, n$.

For any $\epsilon_1 > 0$ and $\epsilon_2 > 0$, we have

$$
\begin{aligned}
& \mathrm{P}\Big\{\Big\|\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Theta_{ns}u_s\Big\| > \epsilon_1\Big\} \\
= \quad & \mathrm{P}\Big\{\Big\|\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Theta_{ns}u_s\Big\| > \epsilon_1, \quad \max_s\|\Theta_{ns}\| > \epsilon_2\Big\} \\
& + \mathrm{P}\Big\{\Big\|\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Theta_{ns}u_s\Big\| > \epsilon_1, \quad \max_s\|\Theta_{ns}\| \le \epsilon_2\Big\} \\
\le \quad & \mathrm{P}\Big\{\Big\|\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Theta_{ns}u_s\Big\| > \epsilon_1, \quad \max_s\|\Theta_{ns}\| \le \epsilon_2\Big\} + \mathrm{P}\Big\{\max_s\|\Theta_{ns}\| > \epsilon_2\Big\} \\
= \quad & \frac{\mathrm{E}\Big\{\big\|\sum_{s=1}^{n}\Theta_{ns}u_s\big\|^2 I\big(\max_s\|\Theta_{ns}\| \le \epsilon_2\big)\Big\}}{n\epsilon_1^2} + o(1).
\end{aligned}
\tag{A.14}
$$

By Assumptions 1 and 2, we can show that

$$\frac{1}{n}\mathrm{E}\Big\{\|\sum_{s=1}^{n}\Theta_{ns}u_s\big\|^2 I\big(\max_s\|\Theta_{ns}\|\le\epsilon_2\big)\Big\}=o(1) \tag{A.15}$$

by letting $\epsilon_2\to 0$. Further, we can see that

$$\begin{aligned} V_{ns} &= \frac{1}{n^2}\sum_{t=1}^{n}\big[X_{1t}-W_n^\top(B_1,B_2)X_{2t}\big]X_{2t}^\top\mathbf{e}\Big[\Delta(\mu)\otimes\Big(\frac{1}{n^2}\sum_{t=1}^{n}X_{2t}X_{2t}^\top\Big)\Big]^{-1}(\Gamma(\mu)\otimes\frac{X_{2s}}{\sqrt{n}})\\ &= \frac{1}{n^2}\sum_{t=1}^{n}\big[X_{1t}-\Big[\Big(\frac{1}{n^2}\sum_{s=1}^{n}X_{2s}X_{2s}^\top\Big)^{-1}\Big(\frac{1}{n^2}\sum_{s=1}^{n}X_{2s}X_{1s}^\top\Big)\Big]^\top X_{2t}\big]X_{2t}^\top\Big[\Big(\frac{1}{n^2}\sum_{t=1}^{n}X_{2t}X_{2t}^\top\Big)^{-1}\frac{X_{2s}}{\sqrt{n}}\Big]\\ &= \Big[\frac{1}{n^2}\sum_{t=1}^{n}X_{1t}X_{2t}^\top-\Big(\frac{1}{n^2}\sum_{s=1}^{n}X_{1s}X_{2s}^\top\Big)\Big(\frac{1}{n^2}\sum_{s=1}^{n}X_{2s}X_{2s}^\top\Big)^{-1}\frac{1}{n^2}\sum_{t=1}^{n}X_{2t}X_{2t}^\top\Big]\Big[\Big(\frac{1}{n^2}\sum_{t=1}^{n}X_{2t}X_{2t}^\top\Big)^{-1}\frac{X_{2s}}{\sqrt{n}}\Big]\\ &\equiv 0. \end{aligned} \tag{A.16}$$

By using (A.13)–(A.16), we prove that

$$\mathcal{B}_{3n}=o_P(1). \tag{A.17}$$

Since

$$\begin{aligned} \mathcal{B}_{2n} &= \frac{1}{n}\sum_{s=1}^{n}(X_{1s}-\mathcal{A}_{1s}^\top X_{2s})u_s\\ &= \frac{1}{\sqrt{n}}\sum_{s=1}^{n}(\frac{X_{1s}}{\sqrt{n}}-W_n(B_1,B_2)^\top\frac{X_{2s}}{\sqrt{n}})u_s+\frac{1}{n}\sum_{s=1}^{n}\Phi_{ns}u_s\\ &=: \frac{1}{\sqrt{n}}\sum_{s=1}^{n}U_{ns}u_s+\frac{1}{n}\sum_{s=1}^{n}\Phi_{ns}u_s, \end{aligned}$$

where $U_{ns}=\frac{X_{1s}}{\sqrt{n}}-W_n(B_1,B_2)^\top\frac{X_{2s}}{\sqrt{n}}$, $\Phi_{ns}=\frac{X_{1s}}{\sqrt{n}}-\mathcal{A}_{1t}^\top\frac{X_{2s}}{\sqrt{n}}-U_{ns}$, and by (A.7) we have $\|\Phi_{ns}\|=o_P(1)$ uniformly in $s=1,\cdots,n$, similar as (A.14) and (A.15), we have that $\frac{1}{n}\sum_{s=1}^{n}\Phi_{ns}u_s=o_P(1)$.

By (A.17) and the definition of $\mathcal{B}_{2n}$, we have

$$\mathcal{B}_{2n}-\mathcal{B}_{3n}=\frac{1}{\sqrt{n}}\sum_{s=1}^{n}U_{ns}u_s+o_P(1),$$

which leads to (A.10). We thus complete the proof of Proposition A.3. ■

**Proof of Theorem 3.2**. By Theorem 3.1, and following the standard argument in local polynomial estimators, we can directly prove (3.2). Details are omitted here. ■

## References


Cai, Z., Q. Li, and J. Park (2009) Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* 148, 101-113.

Chen, J., D. Li, and L. Zhang (2010) Robust estimator in a nonlinear cointegration model. *Journal of Multivariate Analysis* 101, 706-717.

Chen, J., J. Gao, and D. Li (2012) Estimation in semiparametric regression with nonstationary regressors. *Bernoulli* 18, 678-702.

Fan, J., and I. Gijbels (1996) *Local Polynomial Modelling and Its Applications.* London: Chapman & Hall/CRC.

Fan. J., and T. Huang (2005) Profile likelihood inference on semiparametric varying coefficient partially linear models. *B*ernoulli 11, 1031-1059.

Hansen, B.E. (1992) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489-500.

Juhl, T. and Z. Xiao (2005) Partially linear models with unit roots. *Econometric Theory* 21, 877-906.

Li, D., J. Chen and Z. Lin (2011) Statistical inference in partially time-varying coefficient models. *Journal of Statistical Planning and Inference* 141, 995-1013.

Masry, E. (1996) Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571-599.

Park, J.Y., and S.B. Hahn (1999). Cointegrating regressions with time varying coefficients. *E*conometric Theory 15, 664-703.

Phillips, P.C.B. (1986) Optimal inference in cointegrated systems. *Econometrica* 59, 283-306.

Phillips, P.C.B. and S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473-495.

Saikkonen, P. (1995) Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems. *Econometric Theory* 11, 888-911.

Sun, Y., and Q. Li (2011) Data-driven bandwidth selection for nonstationary semiparametric models. *Journal of Business and Economic Statistics* 29, 541-551.

Wang, Q. and P.C.B. Phillips (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710-738.

Wang, Q. and P.C.B. Phillips (2009b) Structural nonparametric cointegrating regression. *Econometrica* 77, 1901-1948.

Xiao, Z. (2009) Functional-coefficient cointegration models. *Journal of Econometrics* 152, 81-92.