Functional-coefficient cointegrating regression with endogeneity

Han-Ying Liang, Yu Shen and Qiying Wang

Tongji University, Tongji University and The University of Sydney

May 12, 2017

Abstract

This paper explores nonparametric estimation of functional-coefficient cointegrating regression models where the structural equation errors are serially dependent and the regressor is endogenours. Generalizing earlier models of Wang and Phillips (2009b, 2016), the self-normalized local kernel and local linear estimators are shown to be asymptotic normal and to be pivotal upon an estimation of co-variances. Our new results open up inference by convenentional nonparametric methods to a wide class of potentially nonlinear cointgerated relations.

Key words and phrases: Cointegration, nonparametric estimation, functional-coefficient model, endogeneity, kernel estimation, local linear estimation.
JEL Classification: C14, C22.

1 Introduction

Since the initial works by Granger (1981) and Engle and Granger (1987), linear cointegrating regression has attracted extensive researches in both theory and empirical applications. The specification in linear structure is convenient for practical work and package software has many standard routines for dealing with such system, encouraging extensive usage of the methods. While common in applications, the linear structure is often too restrictive and linear cointegration models are often rejected by the data even when there is a clear long-run relationship in the series. See, e.g., Park and Phillips (1988), Saikkonen (1995), Terasvirta, et al. (2011), among many others.

To overcome such deficiencies, various nonlinear cointegrating models have been suggested in past decades. For nonlinear parametric cointegrating regression, we refer to Park and Phillips (2001), Chang, et al. (2001), Chang and Park (2003) and Chan and Wang (2015). More recently, Wang and Phillips (2009a, b, 2016), Gao and Phillips (2013a) investigated flexible nonparametric and semiparametric approaches that can cope with the unknown functional form of response in a nonstationary time series setting. A further extension was considered in Cai et al (2009), Xiao (2009) and Trokic

(2014), where the authors suggested a nonlinear cointegrating model with functional-coefficients of the form:

$$y_t = x_t^T \beta_0(z_t) + \epsilon_t, \tag{1.1}$$

where y_t, z_t and ϵ_t are all scalars, $x_t = (x_{t1}, \cdots, x_{td})^T$ is of dimension d, $\beta_0(\cdot)$ is a $d \times 1$ vector of unknown smooth function defined on $\mathbb{R}$ and A^T denotes the transpose of a vector or a matrix A. Extensions of (1.1) to more general nonparametric and semiparametric formulations can be found in Gao and Phillips (2013b) and Li, et al. (2017).

Model (1.1) allows cointegrating relationship that vary or evolve smoothly over time. This framework seems particularly useful in empirical applications where there may be structural evolution in a relationship over time. Asymptotic theory of estimation and inference for model (1.1) and more general related models has been established in the literature. Technical difficulties, however, has confined much of the asymptotic theory to the case of strict exogeneity where the independence is essentially imposed between x_t, z_t and the errors ϵ_t. See, for instance, Cai, et al. (2009), Xiao (2009), Sun, et al. (2013), Gao and Phillips (2013a, b) and Li, et al. (2017). Exogeneity is a natural starting point for a pure cointegrated system and provides some useful insight into the properties of various estimates of nonlinear long run linkages between the system variables. But the assumption is restrictive, especially in a cointegrated framework where the driver variables may be expected to be temporally and contemporaneously correlated. Exogeneity therefore delimits potential applications as well as removing a central technical difficulty in the development of the asymptotics.

The aim of this paper is to remove the exogeneity restriction. Our framework allows for a wider class of regressors and temporal dependence properties within the system, particularly, we may have $E(\epsilon_t|x_t, z_t) \neq 0$, thereby introducing the endogeneity in model (1.1). Another contribution of the present paper is to address the technical difficulties. Our methodology in investigating the asymptotics builds up the techniques currently developed in Wang and Phillips (2009b, 2016), enabling our assumptions neat and our proofs quite straightforward.

The rest of this paper is organized as follows. In Section 2, we investigate the asymptotics for local kernel and local linear nonparametric estimators of $\beta_0(\cdot)$ in model (1.1). The present paper considers two different situations:

(1) x_t is non-stationary and z_t is stationary;

(2) x_t is stationary and z_t is non-stationary.

As mentioned above, in both situations, we allow for $E(\epsilon_t|x_t, z_t) \neq 0$, thereby introducing the endogeneity in model (1.1). Section 3 concludes. The proofs of main results are given in Section 4. The proofs of some auxiliary results are collected in Appendix, i.e., Section 5.

Throughout the paper, we make use of the following notation: for $x = (x_{ij}), 1 \leq i \leq m, 1 \leq j \leq k$, $||x|| = \sum_{i=1}^m \sum_{j=1}^k |x_{ij}|$. We denote constants by $C, C_1, ...$, which may be different at each appearance.

2 Main results

The local kernel estimator of $\beta_0(z)$ in model (1.1) is given by

$$\begin{aligned}\widehat{\beta}_N(z) =&\arg\min_{\beta}\sum_{t=1}^{n}[y_t - x_t^T\beta]^2 K\Big(\frac{z_t - z}{h}\Big)\\ =&\Big[\sum_{t=1}^{n} x_t x_t^T K\Big(\frac{z_t - z}{h}\Big)\Big]^{-1}\sum_{t=1}^{n} x_t y_t K\Big(\frac{z_t - z}{h}\Big),\end{aligned}$$

where $K(x)$ is a nonnegative real function and the bandwidth parameter $h \equiv h_n \to 0$ as $n \to \infty$. The limit behavior of $\widehat{\beta}_N(z)$ has been investigated in past work in some special situations, notably where the error process ϵ_t is a martingale difference sequence and there is no contemporaneous correlation between x_t, z_t and ϵ_t. See, Cai et al.(2009), Xiao (2009), Gao and Phillips (2013a, b) and Li, et al. (2017), for instance.

This work has a similar goal to the previous papers in terms of accommodating endogeneity, but provides more general results with advantages for empirical applications. Our assumptions permit dependence between the error process ϵ_t and the regressors x_t and z_t. These relaxations of the conditions in previous works are particularly important in nonlinear cointegrated systems because finite time horizon dependence between the regressor and the equation error will often be restrictive in practice.

We further consider the local linear estimator $\widehat{\beta}_L(z)$ of $\beta_0(z)$ (e.g., Fan and Gijbels, 1996) defined by

$$\begin{pmatrix}\widehat{\beta}_L(z)\\ \widehat{\beta}_L^{\,\prime}(z)\end{pmatrix} = \arg\min_{\beta,\beta_1}\sum_{t=1}^{n}\big\{y_t - x_t^T[\beta + \beta_1(z_t - z)]\big\}^2 K\Big(\frac{z_t - z}{h}\Big).$$

Namely, we have

$$\widehat{\beta}_L(z) = \Big[\sum_{t=1}^{n} w_t x_t x_t^T K\Big(\frac{z_t - z}{h}\Big)\Big]^{-1}\sum_{t=1}^{n} w_t x_t y_t K\Big(\frac{z_t - z}{h}\Big),$$

where $w_t = V_{n2} - (z_t - z)V_{n1}$, and $V_{nj} = \sum_{t=1}^{n} x_t x_t^T\, K\big(\frac{z_t - z}{h}\big)(z_t - z)^j$.

The asymptotics of $\widehat{\beta}_N(z)$ and $\widehat{\beta}_L(z)$ will be investigated in two different cases mentioned in the introduction part. Since the conditions set on x_t, z_t and ϵ_t are quite different, we consider their theoretical results in Sections 2.1 and 2.2, separately. In Section 2.3, we discuss possible extensions of the model.

2.1 Models with non-stationary x_t and stationary z_t

This section makes use of the following assumptions in the asymptotic development.

A1 (i) $\{z_t, \epsilon_t, \eta_t\}_{t\geq 1}$ (where $\eta_t = x_t - x_{t-1}$) is a stationary α-mixing process of $d + 2$ dimension with $E\eta_t = 0$ and mixing coefficients $\alpha(n) = O(n^{-\gamma})$, where $\gamma > 0$ is specified later;

(ii) $E(\epsilon_1|z_1) = 0$, $E(|\epsilon_1|^3 \big| z_1 = z)$ is bounded and (z_1, ϵ_1) has a joint density function $p(x, y)$ so that $p(x, y)$ is continuous in a neighbourhood of z;

(iii) z_1 has a density function $g(x)$ which is continuous in a neighbourhood of z;

(iv) $\Omega = \lim_{n\to\infty} \frac{1}{n} E x_n\, x_n^T > 0$ and $E||\eta_1||^3 < \infty$.

A2 (i) $K(x)$ is a nonnegative real function having a compact support and $\int_{-\infty}^{\infty} K(x)dx = 1$;

(ii) $\int_{-\infty}^{\infty} xK(x)dx = 0$.

A3 In a neighbourhood of z, for some $\eta > 0$,

(i) $||\beta_0(y + z) - \beta_0(z) - \beta_0'(z)y|| \leq C_z\, |y|^{1+\eta}$;

(ii) $||\beta_0(y + z) - \beta_0(z) - \beta_0'(z)y - \frac{1}{2}\beta''(z)y^2|| \leq C_z\, |y|^{2+\eta}$,

where C_z is a constant depending only on z.

Conditions **A2** and **A3** are standard in literature. See, for instance, Cai, et al. (2000) and Cai, et al. (2009). The smooth condition on $\beta_0(x)$ in **A3** (ii) is stronger than that of **A3** (i), which is required to provide a better bias term in local linear estimator $\hat{\beta}_L(z)$. The conditional mean $E(\epsilon_1|z_1) = 0$ in **A1** (ii) is necessary to make consistency for both estimators $\hat{\beta}_N(z)$ and $\hat{\beta}_L(z)$. We may have $E(\epsilon_1|z_1, x_1) \neq 0$ under **A1**, which introduces endogeneity in the model. This differs from much previous work (e.g. Cai, et al.(2009) and Xiao (2009)) where the model is often assumed to have a martingale structure. The other conditions in **A1** are standard, indicating

$$\Big(\frac{x_{[nt]}}{\sqrt{n}}, \frac{1}{\sqrt{nh}} \sum_{t=1}^{[nt]} K\big[(z_t - z)/h\big]\epsilon_t\Big) \Rightarrow \big\{B_t, \sigma_z\, B_{1t}\big\}, \tag{2.1}$$

on $D_{R^2}[0, 1]$, where

$$\sigma_z^2 = E(\epsilon_1^2|z_1 = z) \int_{-\infty}^{\infty} K^2(x)dx,$$

B_{1t} is a standard Brownian motion independent of B_t, and B_t is a d-dimensional Brownian motion with covariance matrix $\Omega = \lim_{n\to\infty} \frac{1}{n} E x_n x_n^T$. Result (2.1) is vital to establish the asymptotics of $\hat{\beta}_N(z)$ and $\hat{\beta}_L(z)$ in our technical development. For a proof of (2.1), see Lemma 5.1 in Appendix.

Let I_d be a d dimensional identity matrix and $z \in R$ be a fixed constant. The next is our first result.

Theorem 2.1 *Under* **A1**, **A2***(i) and* **A3***(i), for any h satisfying $nh^{3/2} = O(1)$ and $nh^{1+\delta} \to \infty$ for some $\delta > 0$, if γ, which is defined in* **A1***(i), satisfies that $\gamma > \max\{21/2, 6/\delta\}$, then*

$$\Big(\sum_{t=1}^{n} x_t x_t^T K\big[(z_t - z)/h\big]\Big)^{1/2} \Big(\widehat{\beta}_N(z) - \beta_0(z) - c_1 \beta_0'(z)h\Big) \to_D \sigma_z\, \mathbb{N}, \tag{2.2}$$

where $c_1 = \int_{-\infty}^{\infty} xK(x)dx$ and $\mathbb{N} \sim N(0, I_d)$ is a standard d-dimensional normal vector.

Remark 1. From the proof of Theorem 2.1, we have also established the following result:

$$nh^{1/2}\Big(\widehat{\beta}_N(z)-\beta_0(z)-c_1\beta_0'(z)h\Big)\to_D \tau_1\,\Big(\int_0^1 B_sB_s^T ds\Big)^{-1/2}\,\mathbb{N}, \tag{2.3}$$

where $\tau_1^2=g^{-1}(z)\,\sigma_z^2$ and $\mathbb{N}$ is independent of $B=\{B_s\}_{s\geq 0}$. As expected in nonparametric cointegraing regression, due to the nonstationarity of regressor x_t, the convergence rate $n\sqrt{h}$ in (2.3) is faster than $\sqrt{nh}$ comparing to the conventional functional coefficient estimators in stationary time series regression [e.g., Cai, et al. (2000)].

Remark 2. In applications, one may choose $c_1=0$ or the bandwidth h satisfying $nh^{3/2}=o(1)$ so that the term $c_1\beta_0'(z)h$ disappears. Consequently, the self-normalized limit (2.2) is pivotal and well-suited to inference and confidence interval construction upon estimation of $E(\epsilon_1^2|z_1=z)$, which can be constructed by

$$\hat{\sigma}_z^2=\frac{\sum_{t=1}^n[y_t-x_t^T\hat{\beta}_N(z_t)]^2K[(z_t-z)/h]}{\sum_{t=1}^n K[(z_t-z)/h]}.$$

Remark 3. Results (2.2) and (2.3) provide a first order bias $c_1\beta_0'(z)h$. Surprisingly, a higher order bias term can not be expected to add in the result (so that the result is more accurate) even there are more smooth conditions on $\beta_0(z)$. To see this claim, let $K_1(x)=xK(x)$, $c_1=\int_{-\infty}^{\infty}xK(x)dx=0$ and

$$\Lambda_n=\frac{h^{1/2}}{n}\sum_{t=1}^n x_tx_t^TK_1\big[(z_t-z)/h\big].$$

In order to consider the bias term having a order $O(h^2)$ in result (2.3), from the proof of Theorem 2.1 in Section 4.1, we have to show that the bandwidth condition that $nh^{3/2}=O(1)$ can be reduced to $nh^{5/2}=O(1)$ and

$$\Lambda_n-c_0nh^{5/2}=o_P(1), \tag{2.4}$$

for some constant c_0 (c_0 allows to be zero), as $nh^{5/2}=O(1)$. This seems to be impossible except $K_1(x)\equiv 0$. Indeed, by letting $d=1$ and $x_t=\sum_{j=1}^t u_j$, where $u_t\sim$ iid $N(0,1)$ and x_t is independent of $z_t\sim$ iid $N(0,1)$, it is readily seen that

$$\begin{aligned} E\Lambda_n^2 &\geq E\big\{K_1\big[(z_1-z)/h\big]-EK_1\big[(z_1-z)/h\big]\big\}^2\frac{h}{n^2}\sum_{t=1}^n Ex_t^4 \\ &= [1+o(1)]\int_{-\infty}^{\infty}K_1^2(x)dx\,nh^2, \end{aligned}$$

i.e., $\sqrt{n}h/\Lambda_n=O_P(1)$, indicating that (2.4) is impossible whenever $nh^{5/2}=O(1)$. The conclusion here contradicts with Theorem 1 of Xiao (2009), where a higher order bias $O(h^q)$ is given. In handing with his bias term $b_{n,k}$, Xiao only provided a rough estimation without serious proof, making the high order bias term in his Theorem 1 to be doubtful.

It is possible to reduce the bias in local linear estimator $\widehat{\beta}_L(z)$, as indicated in the following theorem.

Theorem 2.2 *Under* **A1–A2** *and* **A3** *(ii), for any h satisfying $nh^{5/2} = O(1)$ and $nh^{1+\delta} \to \infty$ for some $\delta > 0$, if γ, which is defined in* **A1***(i), satisfies that $\gamma > \max\{21/2, 6/\delta\}$, then*

$$\Big(\sum_{t=1}^{n} x_t x_t^T K\big[(z_t - z)/h\big]\Big)^{1/2} \Big(\widehat{\beta}_L(z) - \beta_0(z) - c_2\, \beta_0''(z)\, h^2\Big) \to_D \sigma_z\, \mathbb{N}, \tag{2.5}$$

where $c_2 = \frac{1}{2}\int_{-\infty}^{\infty} x^2 K(x)dx$.

Remark 4. As in Remark 2, the self-normalized limit (2.5) is pivotal upon estimation of $E(\epsilon_1^2|z_1 = z)$. Theorem 2.2 also indicates that the local linear estimator always is better in reducing the bias when z_t is stationary in a functional-coefficient cointegrating regression model. In a related paper, under more restrictive conditions (in particular, without consideration of endogeneity), Theorem 2.1 of Cai, et al. (2009) established a similar version of (2.3):

$$nh^{1/2}\Big(\widehat{\beta}_L(z) - \beta_0(z) - c_2\, \beta_0''(z)\, h^2\Big) \to_D \tau_1\, \Big(\int_0^1 B_s B_s^T ds\Big)^{-1/2} \mathbb{N}, \tag{2.6}$$

where τ_1 is given in Theorem 2.1.

2.2 Models with stationary x_t and non-stationary z_t

In this section, let $\eta_i \equiv (\nu_i, \eta_{1i}, ..., \eta_{mi})^T, i \in Z, m \geq 1$ be a sequence of iid random vectors with $E\eta_0 = 0$, $E\left(\eta_0\eta_0^T\right) = \Sigma$ and $E||\eta_0||^4 < \infty$. We further make use of the following assumptions in the asymptotic development.

A4 (i) $\xi_j, j \geq 1$, is a linear process defined by $\xi_j = \sum_{k=0}^{\infty} \phi_k\, \nu_{j-k}$, where the coefficients $\phi_k, k \geq 0$, satisfy one of the following conditions:

LM. $\phi_k \sim k^{-\mu}\, \rho(k)$, where $1/2 < \mu < 1$ and $\rho(k)$ is a function slowly varying at ∞.

SM. $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$;

(ii) $z_k = (1 - c/n)z_{k-1} + \xi_k$, where $z_0 = 0$ and $c \geq 0$ is a constant;

(iii) $E\nu_1^2 = 1$ and $\lim_{|t|\to\infty} |t|^{\eta}|Ee^{it\nu_1}| < \infty$ for some $\eta > 0$.

A5 (i) $\begin{pmatrix} \epsilon_j \\ x_j \end{pmatrix} = \sum_{k=0}^{\infty} \psi_k\, \eta_{j-k}$, *where the coefficient matrix,*

$$\psi_k = \begin{pmatrix} \psi_k^{(1)} \\ \psi_k^{(2)} \\ \cdot \\ \psi_k^{(d+1)} \end{pmatrix} \quad \text{with} \quad \psi_k^{(s)} = (\psi_{k,s1}, \psi_{k,s2}, ..., \psi_{k,s,m+1}),$$

satisfies $\sum_{k=0}^{\infty} k^{1/4}||\psi_k^{(s)}|| < \infty$, for each $1 \leq s \leq d+1$;

(ii) $Ex_1\epsilon_1 = 0$ and $Ex_1x_1^T > 0$, i.e., $Ex_1x_1^T$ is a positive definite matrix.

A6 In addition to **A2**, $\int_{\infty}^{\infty} |\hat{K}(x)|dx < \infty$, *where* $\hat{K}(x) = \int_{-\infty}^{\infty} e^{ixt}K(t)dt$.

As noticed in Wang and Phillips (2016), Assumption **A4**(i) allows for short (under **SM**) and long (under **LM**) memory innovations ξ_j driving the (near) integrated regressor z_t. Set $d_n^2 = \mathbb{E}(\sum_{k=1}^n \xi_k)^2$, $c_\mu = \frac{1}{(1-\mu)(3-2\mu)} \int_0^\infty x^{-\mu}(x+1)^{-\mu}dx$ and denote by $W_\beta(t)$ a fractional Brownian motion with Hurst parameter $0 < \beta < 1$. It is well-known that

$$d_n^2 \sim \begin{cases} c_\mu \, n^{3-2\mu} \, \rho^2(n), & \text{under } \mathbf{LM}, \\ \phi^2 \, n, & \text{under } \mathbf{SM}, \end{cases}$$

and on $D[0,1]$ the following weak convergence applies (e.g., Chapter 2.1.3 of Wang (2015))

$$z_{\lfloor nt \rfloor}/d_n \;\Rightarrow\; Z(t) := \psi(t) + \gamma \int_0^t e^{-\gamma(t-s)}\psi(s)ds, \tag{2.7}$$

where $\psi(t) = \begin{cases} W_{3/2-\mu}(t), & \text{under } \mathbf{LM} \\ W(t), & \text{under } \mathbf{SM}. \end{cases}$ and $W = W_{1/2}$ is Brownian motion. Furthermore, the limit process $Z(t)$ has continuous local time process $L_Z(t,s)$ [1] with dual (time and space) parameters (t,s) in $[0,\infty) \times \mathbb{R}$. The characteristic function condition $\lim_{|t|\to\infty} |t|^\eta |Ee^{it\nu_1}| < \infty$ for some $\eta > 0$ is not necessary for the establishment of (2.7), but it is requited for the convergence to local time in Lemma 4.2 in Section 4. These notations are used throughout the rest of the paper without further explanation.

Assumption **A5** (i) allows (ϵ_t, x_t) to be cross correlated with z_s for all $s \le t$, thereby inducing endogeneity and giving the structural model more natural temporal dependence properties than those used in previous works [e.g., Cai, et al. (2009), Gao and Phillips (2013b)]. As in Section 2.1, we may have $\text{cov}(\epsilon_t, z_t) \neq 0$ under Assumption **A5**(i), which differs from much previous work where the model is often assumed to form a martingale difference sequence structure. In the latter case, $\text{E}(\epsilon_t|x_t, z_t) = 0$. Assumption **A5** (ii) is necessary to make consistency for both estimators $\hat{\beta}_N(z)$ and $\hat{\beta}_L(z)$. These quantities are well-defined due to $E|\eta_0|^4 < \infty$. We further have $E||x_1x_1^T||e_1^2 < \infty$, which is required in the following main result.

Assumption **A6**, which is the same as in Wang and Phillips (2009b), is quite weak, and are easily verified for various kernels $K(x)$.

Let z be a fixed constant in R. We have the following main result in this section.

Theorem 2.3 *Under* **A4**–**A6** *and* **A3***(ii), for any h satisfing $nh^5/d_n = O(1)$ and $nh/d_n \to \infty$, we have*

$$\Big(\sum_{t=1}^n x_t x_t^T K\big[(z_t - z)/h\big]\Big)^{1/2} \Big(\widehat{\beta}_N(z) - \beta_0(z) - c_2\beta_0''(z)h^2\Big) \to_D \sigma\, \mathbb{N}, \tag{2.8}$$

[1]The local time process $L_G(t,s)$ of a stochastic process $G(x)$ is defined by [e.g., Chapter 2 of Wang (2015)]

$$L_G(t,s) = \lim_{\epsilon\to 0} \frac{1}{2\epsilon} \int_0^t I\{|G(r) - s| \le \epsilon\}dr.$$

where $c_2 = \frac{1}{2}\int_{-\infty}^{\infty} x^2 K(x)dx$, $\sigma^2 = [Ex_1x_1^T]^{-1}\, E(\epsilon_1^2 x_1 x_1^T)\, \int_{-\infty}^{\infty} K^2(x)dx$ *and* $\mathbb{N} \sim N(0, I_d)$ *is a standard* d*-dimensional normal vector. Result (2.8) also holds if we replace* $\widehat{\beta}_N(z)$ *by* $\widehat{\beta}_L(z)$*.*

Remark 5. In comparison with Theorem 2.2 where the result is derived under stationary z_t, (2.8) has a similar structure but with different co-variance σ^2, indicating the limit distributions of $\hat{\beta}_N(z)$, also for $\hat{\beta}_L(z)$, is not mutual independent. As in Theorem 2.2, the self-normalized limit (2.8) is pivotal upon estimation of σ^2, which can be constructed by

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^n x_t x_t^T\, [y_t - x_t^T \hat{\beta}_N(z_t)]^2 K^2[(z_t - z)/h]}{\sum_{t=1}^n x_t x_t^T K[(z_t - z)/h]}.$$

We may establish [result also holds if we replace $\widehat{\beta}_N(z)$ by $\widehat{\beta}_L(z)$]

$$\big(\frac{nh}{d_n}\big)^{1/2}\Big(\widehat{\beta}_N(z) - \beta_0(z) - c_2\beta_0''(z)h^2\Big) \to_D \tau_2\, L_Z^{-1/2}(1,0)\,\mathbb{N}, \tag{2.9}$$

where $\tau_2^2 = [Ex_1x_1^T]^{-1}\,\sigma^2$ and $\mathbb{N}$ is independent of $L_Z(1,0)$. Note that (2.9) is quite different from (2.3) or (2.6), indicating that quite different techniques are used in establishing the results. Result (2.9) has a slow convergence rate due to the fact that, in nonstationary case, the amount of time spent by the process z_t around any particular spatial point point z is n/d_n rather than n so that the corresponding convergence rate in such a regression is now $\sqrt{nh/d_n}$. This philosophia was first noticed in Wang and Phillips (2009a, b). Furthermore, unlike Theorem 2.2, the bias reducing advantage of the local linear nonparametric estimator is lost under point-wise estimation as first noticed in Wang and Phillips (2011). In contrast to point-wise estimation, the local linear non-parametric estimator does have superior performance characteristics to the Nadaraya-Watson estimator in terms of uniform asymptotics over wide domains (Chan and Wang, 2014; Duffy, 2017).

Remark 6. Let $x_{1t} = x_t + A_0$, where A_0 is a constant vector. Note that

$$Ex_{11}\epsilon_1 = Ex_1\epsilon_1 + A_0E\epsilon_1 = 0.$$

A routine modification of Theorem 2.3 yields that result (2.8) still holds if we replace x_t by x_{1t} and σ^2 by σ_1^2 defined by

$$\sigma_1^2 = (A_0A_0^T + Ex_1x_1^T)^{-1}\, E\big[\epsilon_1^2(x_1 + A_0)(x_1 + A_0)^T\big] \int_{-\infty}^{\infty} K^2(x)dx.$$

This fact indicates that Theorem 2.3 provides a natural extension of Wang and Phillips (2009b, 2016) to a functional-coefficient cointegrating regression model. As noted in Wang and Phillips (2009b), there is no inverse problem in structural models of nonlinear cointegration of the form (1.1) where the regressor z_t is an endogenously generated integrated process, avoiding the need for instrumentation and completely eliminating ill-posed functional equation inversions. As a consequence, Theorem 2.3 has important implications for applications.

2.3 Possible extensions

In economic applications, it is important to consider multivariate extension of model (1.1), i.e., to consider the model having the form:

$$y_t = x_t^T \beta_0(z_t, w_t) + \epsilon_t, \tag{2.10}$$

where y_t, z_t and ϵ_t are all scalars, $x_t = (x_{t1}, \cdots, x_{td})^T$ and $w_t = (w_{t1}, ..., w_{td_1})$ are of dimension d and d_1, respectively, and $\beta_0(\cdot, \cdots)$ is a $d \times 1$ vector of unknown smooth function defined on $\mathbb{R}^{1+d_1}$. As in one-dimension situation, the local kernel estimator $\hat{\beta}_0(\cdot, \cdot)$ of $\beta_0(\cdot, \cdot)$ can be similarly defined by

$$\hat{\beta}_0(z, w) \quad = \quad \frac{\sum_{t=1}^n x_t\, y_t\, K[(z_t - z)/h]\, \Pi_{j=1}^{d_1} K_j[(w_{tj} - w_j)/h_j]}{\sum_{t=1}^n x_t\, x_t^T K[(z_t - z)/h] \Pi_{j=1}^{d_1} K_j[(w_{tj} - w_j)/h_j]}, \tag{2.11}$$

where $K(x), K_j(x)$ are non-negative kernel functions and the bandwidth $h, h_j \equiv h_{jn} \to 0$ for $j = 1, ..., d_1$.

If x_t is nonstationary and both z_t and w_t are stationary, asymptotics of $\hat{\beta}_0(z, w)$ can be obtained by using similar arguments as in Section 2.1 under some regular settings and hence the details are omitted. We next consider the situation that x_t and w_t are stationary and z_t is an $I(1)$ process. As noticed in Section 5.1.5 of Wang (2015), to enable $\hat{\beta}_0(z, w)$ being a consistency estimator, the stationary assumption on w_t is essentially necessary. We further assume $d_1 = 1$ for the sake of notation convenience. The extension to $d_1 \geq 2$ is straightforward.

To investigate the asymptotics of $\hat{\beta}_0(z, w)$, as in Section 2.2, let $\eta_i \equiv (\nu_i, \eta_{1i}, ..., \eta_{mi})^T, i \in Z, m \geq 1$ is a sequence of iid random vectors with $E\eta_0 = 0$, $E\left(\eta_0 \eta_0^T\right) = \Sigma$ and $E||\eta_0||^4 < \infty$. We also make use of the following assumtions.

A7 [Regression function and Kernel function]

(a) The kernels $K(x)$ and $K_1(x)$ have a common compact support Ω satisfying $\int_\Omega K(x)dx = \int_\Omega K_1(x)dx = 1$ and $\int_{-\infty}^{\infty} |\hat{K}(t)|dt < \infty$, where $\hat{K}(t) = \int_{-\infty}^{\infty} e^{itx} K(x)dx$;

(b) When (x, y) is in a compact set, we have

$$|\beta_0(x + \delta_1, y + \delta_2) - \beta_0(x, y)| \quad \leq \quad C\left(|\delta_1| + |\delta_2|\right), \tag{2.12}$$

whenever δ_1 and δ_2 are sufficiently small.

A8 [Regressors]

(a) z_t is defined as in **A4**(ii) and **A4** holds;

(b) $x_t = (x_{t1}, \cdots, x_{td})^T$, where $x_{ti} = \Gamma_i(\eta_t, ..., \eta_{t-m_0+1})$ for some $m_0 > 0$, where $\Gamma_i(.), 1 \leq i \leq d$, are real measurable functions of its contents;

(c) $w_t = \Gamma_0(\eta_t, ..., \eta_{t-m_0+1})$ for some $m_0 > 0$, where $\Gamma_0(.)$ is a real measurable function of its contents;

(d) For any fixed w and each $1 \le i \le d$, $E(|x_{ti}|^{4+\delta}|w_t = w) < \infty$ with $t = m_0$ for some $\delta > 0$;

(e) For any fixed w and each $1 \le i, j \le d$, (x_{ti}, x_{tj}, w_t) and (x_{ti}, w_t) have joint density functions $p_{ij}(x, y, z)$ and $p_j(x, z)$, respectively, that are continuous in a neighbourhood of w;

(f) For any fixed w, $D_w = \big(d_{ij}(w)\big)_{1\le i,j\le d}$ is a positive-definite matrix, where, with $t = m_0$, $d_{ij}(w) = E(x_{ti}x_{tj}|w_t = w)$.

A9 [Error processes] $\{\epsilon_i, \mathcal{F}_i\}_{i\ge1}$, where $\mathcal{F}_{i+1} = \sigma(\eta_i, \eta_{i-1}, ...)$, is a martingale difference such that, as $i \to \infty$, $\mathrm{E}\left(\epsilon_i^2 \mid \mathcal{F}_{i-1}\right) \to_{a.s.} \sigma^2 > 0$, and, as $A \to \infty$,

$$\sup_{i\ge1} \mathrm{E}\left[\epsilon_i^2 \mathrm{I}(|\epsilon_i| \ge A) \mid \mathcal{F}_{i-1}\right] = o_{\mathrm{P}}(1).$$

Theorem 2.4 *Under Assumptions* **A7**–**A9***, for any h and h_1 satisfying $nhh_1/d_n \to \infty$ and $(h + h_1)^2 nhh_1/d_n \to 0$, we have*

$$D_n^{1/2}\left[\hat\beta_0(z, w) - \beta_0(z, w)\right] \to_D \tau\, \mathbb{N}, \tag{2.13}$$

where $D_n = \sum_{t=1}^n x_t x_t^T K[(z_t - z)/h] K_1[(w_t - w)/h_1]$ and $\tau^2 = \sigma^2 \int_\Omega K_1^2(x)dx \int_\Omega K_2^2(x)dx$, $\mathbb{N} \sim N(0, I_d)$ is a standard d-dimensional normal vector.

Remark 7. Theorem 2.4 provides an extension of Theorem 5.7 in Wang (2015) to a functional-coefficient cointegrating regression model by imposing the restriction $E(\epsilon_t|x_t, z_t, w_t) = 0$ on the error processes. In a related paper, Gao and Phillips (2013b) [also see Sun, et al. (2013)] investigated the model (2.10) with both x_t and z_t are $I(1)$ processes under some similar conditions. Their main theorems made use of a result established by Phillips (2009), where the independence between x_t and z_t, w_t is essentially required. In terms of possible empirical applications, it is of interests to remove these restrictions, in particular, to establish the asymptotics without imposing $E(\epsilon_t|x_t, z_t, w_t) = 0$ in the model as in Theorems 2.1-2.3. There are some technical challenges in the investigation of general model (2.10), and hence the extensions will be left for future work.

3 Conclusion

This paper studies nonparametric estimation for functional-coefficient cointegrating regression models of the form (1.1) in two different situations: (1) x_t is nonstationary and z_t is stationary and (2) x_t is stationary and z_t is nonstationary. Both self-normalized local kernel and local linear estimators are shown to be asymptotic normal and to be pivotal upon an estimation of co-variances. Importantly, our asymptotic results allow for endogenous regressor in the models, namely, we assume $E(\epsilon_t|x_t, z_t) \ne 0$ in

(1.1). These structural models differ from various previous works and open up some interesting possibilities for functional-coefficient regression in empirical research with integrated processes. In terms of many possible empirical applications, some extensions of the ideas presented here to other useful models involving nonlinear functions of integrated processes seems to be interesting. In particular, partial linear cointegration models (e.g., Gao and Phillips, 2003b)) may be treated in a similar way to (1.1), but there are difficulties for multiple non-stationary regression models, due to the nonrecurrence of the limit processes in high dimensions (c.f. Park and Phillips, 2001). It will also be of interest in exploring the functional-coefficient cointegration models by the use of instrumental variables in the present nonstationary context. We plan to report on some of these extensions in later work.

4 Proofs

Since methodology is different, the proofs of Theorems 2.1-2.2, 2.3 and 2.4 will be Sections 4.1, 4.2 and 4.3, respectively.

4.1 Proofs of Theorems 2.1 and 2.2

We start with the some preliminaries. Write $x_{nt} = x_t/\sqrt{n}$, $K_j(x) = x^j K(x)$ and $\mu_j = \int_{-\infty}^{\infty} K_j(x)dx$ for $j \geq 0$. Other notation is the same as in previous sections except mentioned explicitly.

Lemma 4.1 *(a) Under* **A1** *(ii), for any* $0 \leq \alpha \leq 3$*, we have*

$$EK\big[(z_1 - z)/h\big](1 + |\epsilon_1|^{\alpha}) \quad \leq \quad C(z)h, \tag{4.1}$$

where $C(z)$ *is a constant depending only on* z*; (b) Under* **A1** *(iii), as* $h \to 0$*, we have*

$$h^{-1} EK_j\big[(z_1 - z)/h\big] = g(z)\mu_j + o(1). \tag{4.2}$$

The proof of Lemma 4.1 is routine, and hence the details are omitted. In the next lemma, suppose that $H(x)$ and $H_1(x)$ are locally bounded real functions on R^d and $H_1(x)$ satisfies the local lipschitz condition, i.e., for any $||x|| + ||y|| \leq K$,

$$\big|H_1(x) - H_1(y)\big| \leq C_K ||x - y||, \tag{4.3}$$

where C_K is a constant depending only on K.

Lemma 4.2 *(i) For any real function* $A_n(x, y)$*,*

(a) we have

$$\frac{1}{n}\sum_{t=1}^{n} H(x_{nt})A_n(z_t, \epsilon_t) = O_P\big(E|A_n(z_1, \epsilon_1)|\big); \tag{4.4}$$

(b) If $EA_n(z_1, \epsilon_1) = 0$ for each $n \geq 1$, then for any $\alpha > 0$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} H_1(x_{nt}) A_n(z_t, \epsilon_t) \quad = \quad O_P\Big\{ \big[E|A_n(z_1, \epsilon_1)|^{2+\alpha}\big]^{1/(2+\alpha)} \Big\}. \tag{4.5}$$

(ii) For any $h \to 0$ satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$, we have

$$\Big\{ \frac{1}{n} \sum_{t=1}^{n} H(x_{nt}), \; \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} H_1(x_{nt}) K[(z_t - z)/h]\, \epsilon_t \Big\}$$
$$\to_D \; \Big\{ \int_0^1 H(B_s) ds, \; a_1 \big(\int_0^1 H_1^2(B_s) ds \big)^{1/2} N \Big\}, \tag{4.6}$$

where $a_1^2 = g(z)\, \sigma_z^2$, N is a standard normal variate independent of B_s.

The proof of Lemma 4.2 will be given in Appendix. Note that $|K_j(x)| \leq CK(x)$ as $K(x)$ has a compact support. Result (4.4), together with (4.1), implies that, as $h \to 0$,

$$\frac{1}{nh} \sum_{t=1}^{n} ||x_{nt} x_{nt}^T|| \, |K_j\big[(z_t - z)/h\big]| \quad = \quad O_P(1). \tag{4.7}$$

Similarly, by using (4.2) and (4.5) with $A_n(z_t, \epsilon_t) = K_j\big[(z_t - z)/h\big] - EK_j\big[(z_t - z)/h\big]$, we have

$$\begin{aligned}
\Delta_{nj} \quad &:= \quad \frac{1}{nh} \sum_{t=1}^{n} x_{nt} x_{nt}^T K_j\big[(z_t - z)/h\big] \\
&= \quad \frac{1}{nh} \sum_{t=1}^{n} x_{nt} x_{nt}^T \, EK_j\big[(z_t - z)/h\big] + \frac{1}{nh} \sum_{t=1}^{n} x_{nt} x_{nt}^T \big[K_j\big[(z_t - z)/h\big] - EK_j\big[(z_t - z)/h\big]\big] \\
&= \quad [g(z)\mu_j + o(1)] \, \frac{1}{n} \sum_{t=1}^{n} x_{nt} x_{nt}^T + O_P((nh^{1+\alpha/(2+\alpha)})^{-1/2}) \\
&= \quad g(z)\mu_j \, \frac{1}{n} \sum_{t=1}^{n} x_{nt} x_{nt}^T + o_P(1),
\end{aligned} \tag{4.8}$$

by taking α sufficiently small so that $nh^{1+\alpha/(2+\alpha)} \geq nh^{1+\delta} \to \infty$. Furthermore, it follows from (4.6) and the continuous mapping theorem that, for any $h \to 0$ satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$,

$$\Big(\frac{1}{n} \sum_{t=1}^{n} x_{nt} x_{nt}^T, \; \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_{nt} K[(z_t - z)/h]\, \epsilon_t \Big\}$$
$$\to_D \; \Big\{ \int_0^1 B_s B_s^T ds, \; a_1 \big(\int_0^1 B_s B_s^T ds \big)^{1/2} \mathbb{N} \Big\}, \tag{4.9}$$

where $\mathbb{N} \sim N(0, I_d)$ is a d dimensional normal vector independent of B_s with covariance I_d.

We are now ready to prove the main results.

Proof of Theorem 2.1. We may write

$$n\sqrt{h}\Big(\widehat{\beta}_N(z) - \beta_0(z) - c_1 \beta_0'(z) h \Big) = \Delta_{n0}^{-1} \big(S_n + R_{n1} + R_{n2} \big), \tag{4.10}$$

where

$$\begin{aligned} S_n &= \frac{1}{\sqrt{nh}} \sum_{t=1}^n x_{nt} K(z_t - z) \epsilon_t, \\ R_{n1} &= h^{-1/2} \sum_{t=1}^n x_{nt} x_{nt}^T K[(z_t - z)/h][\beta_0(z_t) - \beta_0(z) - \beta_0'(z)(z_t - z)], \\ R_{n2} &= h^{1/2} \beta_0'(z) \sum_{t=1}^n x_{nt} x_{nt}^T \big(K_1[(z_t - z)/h] - c_1 K[(z_t - z)/h]\big). \end{aligned}$$

A3 (i) and (4.7) imply that, for some $\eta > 0$,

$$||R_{n1}|| \leq C(1 + |z|^\beta)\, h^{1/2+\eta} \sum_{t=1}^n ||x_{nt} x_{nt}^T||\, K[(z_t - z)/h] = O_P(nh^{3/2+\eta}) = o_P(1).$$

Write $A_n(z_t, \epsilon_t) = K_1[(z_t - z)/h] - c_1 K[(z_t - z)/h]$. Lemma 4.1 implies that $h^{-1} E A_n(z_1, \epsilon_1) = o(1)$ and $E|A_n(z_1, \epsilon_1)|^{2+\alpha} = O(h)$. It is readily seen from (4.5) that

$$||R_{n2}|| = o_P(1)\, nh^{3/2} + O_P(1)\, \sqrt{n} h^{1/2+1/(2+\alpha)} = o_P(1)$$

whenever $nh^{3/2} = O(1)$. Taking these estimates into (4.10), we get

$$\Big(\sum_{t=1}^n x_t x_t^T K\big[(z_t - z)/h\big]\Big)^{1/2} \Big(\widehat{\beta}_N(z) - \beta_0(z) - c_1 \beta_0'(z) h\Big) = \Delta_{n0}^{-1/2} \big[S_n + o_P(1)\big] \to_D \sigma_z \mathbb{N},$$

due to (4.8) - (4.9) and the continuous mapping theorem. Theorem 2.1 is now proved. ■

Proof of Theorem 2.2. Similarly to the proof of $\widehat{\beta}_N(z)$, we may write

$$n\sqrt{h}\Big(\widehat{\beta}_L(z) - \beta_0(z) - c_2 \beta_0''(z) h^2\Big) = \Delta_n^{-1} \big(P_n + T_{n1} + \beta_0''(z) nh^{5/2}\, T_{n2}\big), \tag{4.11}$$

where, by letting $v_t = K[(z_t - z)/h][\beta_0(z_t) - \beta_0(z) - \beta_0'(z)(z_t - z) - \frac{1}{2}\beta_0''(z)(z_t - z)^2]$,

$$\begin{aligned} \Delta_n &= \frac{1}{nh} \sum_{t=1}^n w_t x_{nt} x_{nt}^T K[(z_t - z)/h], \\ P_n &= \frac{1}{\sqrt{nh}} \sum_{t=1}^n w_t x_{nt} K[(z_t - z)/h] \epsilon_t, \\ T_{n1} &= \frac{1}{\sqrt{h}} \sum_{t=1}^n w_t x_{nt} x_{nt}^T v_t, \\ T_{n2} &= \frac{1}{nh} \sum_{t=1}^n w_t x_{nt} x_{nt}^T \Big\{\frac{1}{2} K_2[(z_t - z)/h] - c_2 K[(z_t - z)/h]\Big\}, \end{aligned}$$

where we have used the fact:

$$\sum_{t=1}^n w_t x_{nt} x_{nt}^T K[(z_t - z)/h](z_t - z) = 0. \tag{4.12}$$

Note that, as $h \to 0$ and $nh^{1+\delta} \to \infty$ for some $\delta > 0$,

$$n^{-2}h^{1-j}V_{nj} = \Delta_{nj} = g(z)\mu_j \frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + o_P(1),$$

by (4.8). It is readily seen from (4.8) and Lemma 4.2 that, by recalling $\mu_1 = 0$ and $\mu_0 = 1$,

$$\begin{aligned}
\Delta_n &= V_{n2}\,\Delta_{n0} - hV_{n1}\,\Delta_{n1} = V_{n2}\Big[\frac{g(z)}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + o_P(1)\Big];\\
P_n &= V_{n2}\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K[(z_t - z)/h]\epsilon_t - hV_{n1}\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K_1[(z_t - z)/h]\epsilon_t\\
&= V_{n2}\Big[\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K[(z_t - z)/h]\epsilon_t + o_P(1)\Big];\\
\|T_{n1}\| &\le Ch^{3/2+\delta}\sum_{t=1}^{n}|w_t|\,||x_{nt}x_{nt}^T||\,K[(z_t - z)/h]\\
&\le Ch^{3/2+\delta}\Big(|V_{n2}|\sum_{t=1}^{n}||x_{nt}x_{nt}^T||\,K[(z_t - z)/h] + h|V_{n1}|\sum_{t=1}^{n}||x_{nt}x_{nt}^T||\,K_1[(z_t - z)/h]\Big)\\
&= O_P(nh^{5/2+\delta})\,V_{n2};\\
T_{n2} &= V_{n2}\frac{1}{nh}\sum_{t=1}^{n} x_{nt}x_{nt}^T\Big\{\frac{1}{2}K_2[(z_t - z)/h] - c_2K[(z_t - z)/h]\Big\}\\
&\quad -hV_{n1}\frac{1}{nh}\sum_{t=1}^{n} x_{nt}x_{nt}^T\Big\{\frac{1}{2}K_3[(z_t - z)/h] - c_2K_1[(z_t - z)/h]\Big\}\\
&= o_P(1)V_{n2}.
\end{aligned}$$

Taking these facts into (4.11), we obtain

$$\begin{aligned}
&\Big(\sum_{t=1}^{n} x_tx_t^T K\big[(z_t - z)/h\big]\Big)^{1/2}\Big(\widehat{\beta}_L(z) - \beta_0(z) - c_2\beta_0''(z)h^2\Big)\\
&= \Big[\frac{g(z)}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + o_P(1)\Big]^{-1/2}\Big[\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K[(z_t - z)/h]\epsilon_t + o_P(1)nh^{5/2}\Big]\\
&\to_D \sigma_z\,\mathbb{N}
\end{aligned}$$

as $nh^{5/2} = O(1)$ and $nh^{1+\delta} \to \infty$ for some $\delta > 0$, due to (4.9) and the continuous mapping theorem. Theorem 2.2 is now proved. ■

4.2 Proof of Theorem 2.3

As in Section 4.1, let $K_j(x) = x^jK(x)$ and $\mu_j = \int_{-\infty}^{\infty} K_j(x)dx$ for $j \ge 0$. Let

$$u_k = \sum_{l,m=0}^{\infty} \varphi_l^T\,\eta_{k-l}\,\eta_{k-m}^T\,\widetilde{\varphi}_m, \tag{4.1}$$

where coefficient constants φ_l and $\widetilde{\varphi}_m$ are the $d+1$ dimensional vectors satisfying $\sum_{l=0}^{\infty} l^{1/4}\|\varphi_l\| < \infty$ and $\sum_{m=0}^{\infty} m^{1/4}\|\widetilde{\varphi}_m\| < \infty$. We start with the following lemma. The proof of Lemma 4.3 is similar to Lemma 2.2 and Theorem 3.16 of Wang (2015). A outline will be given in the appendix. For a proof of Lemma 4.4, we refer to Theorem 3.18 of Wang (2015). See, also, Wang and Phillips (2011).

Lemma 4.3 *Let z be a fixed constant. For any $1 \le s, t \le d+1$ and any h satisfying $h \log n \to 0$ and $nh/d_n \to \infty$, we have*

$$\sum_{k=1}^{n}(1+|u_k|)K\Big(\frac{z_k-z}{h}\Big) = O_P\big(nh/d_n\big); \tag{4.2}$$

$$\sum_{k=1}^{n}\big(u_k - Eu_k\big)K\Big(\frac{z_k-z}{h}\Big) = O_P\Big(\big(\frac{nh}{d_n}\big)^{1/2}\Big)\sum_{l,m=0}^{\infty}\Big(l^{1/4}m^{1/4}\|\varphi_l\|\,\|\widetilde{\varphi}_m\|\Big). \tag{4.3}$$

and

$$\begin{aligned}\Big\{\frac{d_n}{nh}\sum_{t=1}^{n}K\Big(\frac{z_t-z}{h}\Big),\ \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n}\big(u_k - Eu_k\big)K\Big(\frac{z_t-z}{h}\Big)\Big\} \\ \to_D \ \Big\{L_Z(1,0), a_2\, L_Z^{1/2}(1,0)\, N\Big\},\end{aligned} \tag{4.4}$$

where $a_2^2 = E(u_k^2)\int_{-\infty}^{\infty} K^2(t)dt$, and N is standard normal variate independent of $L_Z(1,0)$;

Lemma 4.4 *Let $g(x)$ be a real function having a compact support. If $\int_{-\infty}^{\infty} g(x)dx = 0$, then*

$$\sum_{k=1}^{n} g\Big(\frac{z_k-z}{h}\Big) = O_P\Big(\big(nh/d_n\big)^{1/2}\Big), \tag{4.5}$$

for any h satisfying $nh/d_n \to \infty$.

Since, due to Assumption **A5**, each element of $x_t x_t^T$ and $x_t\epsilon_t$ can be representated as u_k for some specified φ_l and $\widetilde{\varphi}_m$, it follows from Lemma 4.3 that

$$\begin{aligned} D_{nj} &:= \frac{d_n}{nh}\sum_{t=1}^{n} x_t x_t^T K_j\Big(\frac{z_t-z}{h}\Big) \\ &= \frac{d_n}{nh}\sum_{t=1}^{n} E(x_t x_t^T)K_j\Big(\frac{z_t-z}{h}\Big) + \frac{d_n}{nh}\sum_{t=1}^{n}\big[x_t x_t^T - E(x_t x_t^T)\big]K_j\Big(\frac{z_t-z}{h}\Big) \\ &= E(x_1 x_1^T)\frac{d_n}{nh}\sum_{t=1}^{n} K_j\Big(\frac{z_t-z}{h}\Big) + O_P\Big(\Big(\frac{d_n}{nh}\Big)^{1/2}\Big) \\ &= E(x_1 x_1^T)\frac{d_n}{nh}\sum_{t=1}^{n} K_j\Big(\frac{z_t-z}{h}\Big) + o_P(1), \end{aligned} \tag{4.6}$$

as $nh/d_n \to \infty$. Furthermore, due to $E(x_1\epsilon_1) = 0$, it follows from (4.4) and the continuous mapping theorem that

$$\Big\{\frac{d_n}{nh}\sum_{t=1}^{n}K\Big(\frac{z_t-z}{h}\Big),\ \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n}x_t\epsilon_t K\Big(\frac{z_t-z}{h}\Big)\Big\}$$

$$\to_D \ \Big\{L_Z(1,0), a_2\, L_Z^{1/2}(1,0)\, \mathbb{N}\Big\}, \tag{4.7}$$

where $a_2^2 = E(\epsilon_1^2 x_1 x_1^T)\int_{-\infty}^{\infty} K^2(x)dx$, and $\mathbb{N} \sim N(0, I_d)$ is a d dimensional normal vector independent of $L_Z(1,0)$.

We are now ready to prove Theorems 2.3. By letting $v_t = K\big(\frac{z_t - z}{h}\big)[\beta_0(z_t) - \beta_0(z) - \beta_0'(z)(z_t - z) - \frac{1}{2}\beta_0''(z)(z_t - z)^2]$, we may write

$$\big(\frac{nh}{d_n}\big)^{1/2}\Big(\widehat{\beta}_N(z) - \beta_0(z) - c_2\beta_0''(z)h^2\Big) = D_{n0}^{-1}\Big(S_n + R_{n1} + \beta_0'(z)R_{n2} + \beta_0''(z)R_{n3}\Big), \tag{4.8}$$

where

$$\begin{aligned} S_n &= \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t\epsilon_t K\Big(\frac{z_t - z}{h}\Big), \\ R_{n1} &= \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t x_t^T v_t, \\ R_{n2} &= h\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t x_t^T K_1\Big(\frac{z_t - z}{h}\Big), \\ R_{n3} &= h^2\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t x_t^T\Big\{\frac{1}{2}K_2[(z_t - z)/h] - c_2 K[(z_t - z)/h]\Big\}. \end{aligned}$$

From **A3**(ii) and (4.2), as $nh^5/d_n = O(1)$ we have

$$\|R_{n1}\| \le Ch^{2+\eta}\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n ||x_t x_t^T||\, K\Big(\frac{z_t - z}{h}\Big) = O_P\Big(\big(\frac{nh}{d_n}\big)^{1/2} h^{2+\eta}\Big) = o_P(1). \tag{4.9}$$

From (4.3) and (4.5), as $h \to 0$ we have

$$\begin{aligned} R_{n2} &= h\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n \big[x_t x_t^T - E(x_t x_t^T)\big]K_1\Big(\frac{z_t - z}{h}\Big) \\ &\quad + h\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n E(x_t x_t^T)K_1\Big(\frac{z_t - z}{h}\Big) \\ &= O_P(h) = o_P(1). \end{aligned} \tag{4.10}$$

Note that $\int_{-\infty}^{\infty}\big[\frac{1}{2}K_2(x) - c_2K(x)\big]dx = 0$. It follows from (4.3) and (4.5) again that, as $nh^5/d_n = O(1)$,

$$\begin{aligned} R_{n3} &= h^2\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n \big[x_t x_t^T - E(x_t x_t^T)\big]\Big\{\frac{1}{2}K_2[(z_t - z)/h] - c_2K[(z_t - z)/h]\Big\} \\ &\quad + \Big(\frac{nh^5}{d_n}\Big)^{1/2}\frac{d_n}{nh}\sum_{t=1}^n E(x_t x_t^T)\Big\{\frac{1}{2}K_2[(z_t - z)/h] - c_2K[(z_t - z)/h]\Big\} \\ &= O_P(h^2) + O_p\Big(\big(nh^5/d_n\big)^{1/2}\Big)\, o_P(1) = o_P(1). \end{aligned} \tag{4.11}$$

Combining (4.6) and (4.8)-(4.11), we obtain

$$\Big(\sum_{t=1}^n x_t x_t^T K\big[(z_t - z)/h\big]\Big)^{1/2}\,\Big(\widehat{\beta}_N(z) - \beta_0(z) - c_2\beta_0''(z)h^2\Big)$$

$$= \Big[E(x_1x_1^T)\frac{d_n}{nh}\sum_{t=1}^n K\Big(\frac{z_t-z}{h}\Big)+o_P(1)\Big]^{-1/2}\Big[\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t\epsilon_t K\Big(\frac{z_t-z}{h}\Big)+o_P(1)\Big]\to_D \sigma\,\mathbb{N},$$

due to (4.7) and the continuous mapping theorem. This proves (2.8).

We next prove that(2.8) still holds if $\widehat{\beta}_N(z)$ is replaced by $\widehat{\beta}_L(z)$. In fact, as in the proof of Theorem 2.2, we may write

$$\big(\frac{nh}{d_n}\big)^{1/2}\Big(\widehat{\beta}_L(z)-\beta_0(z)-c_2\beta_0''(z)h^2\Big)=D_n^{-1}\big(P_n+T_{n1}+\beta_0''(z)T_{n2}\big), \tag{4.12}$$

by virtue of (4.12), where

$$\begin{aligned}
D_n &= \frac{d_n}{nh}\sum_{t=1}^n w_t x_t x_t^T K\Big(\frac{z_t-z}{h}\Big),\\
P_n &= \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n w_t x_t\epsilon_t K\Big(\frac{z_t-z}{h}\Big),\\
T_{n1} &= \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n w_t x_t x_t^T v_t,\\
T_{n2} &= h^2\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n w_t x_t x_t^T\Big\{\frac{1}{2}K_2[(z_t-z)/h]-c_2K[(z_t-z)/h]\Big\}.
\end{aligned}$$

Noting that from (4.6), it can be obtained

$$\frac{d_n}{nh}h^{-j}V_{nj}=D_{nj}=E(x_1x_1^T)\frac{d_n}{nh}\sum_{t=1}^n K_j\Big(\frac{z_t-z}{h}\Big)+o_P(1).$$

Since $\mu_1=0$ and $\mu_0=1$, from Lemmas 4.3 and 4.4 we have

$$\begin{aligned}
D_n &= V_{n2}D_{n0}-hV_n1D_{n1}=V_{n2}[D_{n0}+o_P(1)];\\
P_n &= V_{n2}\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_t\epsilon_t K\Big(\frac{z_t-z}{h}\Big)-hV_{n1}\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_t\epsilon_t K_1\Big(\frac{z_t-z}{h}\Big)\\
&= V_{n2}\Big[\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_t\epsilon_t K\Big(\frac{z_t-z}{h}\Big)+o_P(1)\Big];\\
\|T_{n1}\| &\le V_{n2}\|R_{n1}\|+h|V_{n1}|\sum_{t=1}^n x_t x_t^T|v_t(z_t-z)|\\
&= O_P\Big(\big(\frac{nh}{d_n}\big)^{1/2}h^{2+\eta}\Big)V_{n2};\\
T_{n2} &= V_{n2}R_{n3}-V_{n1}\,h^3\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^n x_t x_t^T\Big\{\frac{1}{2}K_3[(z_t-z)/h]-c_2K_1[(z_t-z)/h]\Big\}\\
&= o_P(1)\,V_{n2}.
\end{aligned}$$

Taking these facts into (4.12), the claim follows from

$$\Big(\sum_{t=1}^n x_t x_t^T K\big[(z_t-z)/h\big]\Big)^{1/2}\Big(\widehat{\beta}_L(z)-\beta_0(z)-c_2\beta_0''(z)h^2\Big)$$

$$
\begin{aligned}
&= \Big[E(x_1x_1^T)\frac{d_n}{nh}\sum_{t=1}^n K\Big(\frac{z_t-z}{h}\Big)+o_P(1)\Big]^{-1/2}\Big[\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_t\epsilon_t K\Big(\frac{z_t-z}{h}\Big)+o_P(1)\Big]\\
&\to_D \sigma\,\mathbb{N},
\end{aligned}
$$

due to (4.7) and the continuous mapping theorem. Theorem 2.3 is now proved. ■

4.3 Proof of Theorem 2.4

Let $V_t = x_t K[(z_t - z)/h]\, K_1[(w_t - w)/h_1]$. We may write

$$
D_n^{1/2}\big[\hat\beta_0(z,w)-\beta_0(z,w)\big] = D_n^{-1/2}S_n + R_n. \tag{4.13}
$$

where $S_n = \sum_{t=1}^n \epsilon_t V_t$ and, by (2.12),

$$
\begin{aligned}
||R_n|| &= \sum_{t=1}^n |\beta_0(z_t,w_t)-\beta_0(z,w)|\,||D_n^{-1/2}V_t||\\
&\le C(|h_1|+|h_2|)\sum_{t=1}^n ||D_n^{-1/2}V_t||.
\end{aligned}
$$

By the continuous mapping theorem, result (2.13) will follow if we prove

$$
||R_n|| = o_P(1), \tag{4.14}
$$

and for any $A=(A_1,...,A_d)^T\in R^d$,

$$
\begin{aligned}
&\Big\{\frac{d_n}{nhh_1}A^TD_nA,\quad (\frac{d_n}{nhh_1})^{1/2}A^TS_n\Big\}\\
&\to_D \big\{(A^TD_wA)\,L_Z(1,0),\ \tau^{1/2}\,(A^TD_wA)\,N\,L_Z^{1/2}(1,0)\big\},
\end{aligned} \tag{4.15}
$$

where $L_Z((1,0)$ is given as in Section 2.2 and $N\sim N(0,1)$ is independent of $L_Z(1,0)$.

We start with some preliminaries. Set $\Delta_t = \sum_{k,j=1}^d A_kA_jx_{tk}x_{tj}K_1[(w_t-w)/h_1]$. Since, by **A8**(d) and some standard arguments,

$$
\begin{aligned}
Ex_{ti}x_{tj}\,K_1^\gamma[(w_t-w)/h_1] &= h_1E(x_{ti}x_{tj}|w_t=w)\int_\Omega K_1^\gamma(x)dx+o(h_1),\\
E|x_{ti}|^\beta\,K_1^\gamma[(w_t-w)/h_1] &= h_1E(|x_{ti}|^\beta|w_t=w)\int_\Omega K_1^\gamma(x)dx+o(h_1)=O(h_1),
\end{aligned}
$$

for any $\gamma>0$, $0\le\beta\le 4+\delta$ and uniformly for all $t\ge m_0$, we have $E\Delta_t = h_1[A^TDA+o(1)]$ and $E\Delta_t^2 = O(h_1)$. Now it follows from Lemma 2.2 (ii) of Wang (2015) that, for any $A=(A_1,...,A_d)^T\in R^d$,

$$
\begin{aligned}
\frac{d_n}{nhh_1}A^TD_nA &= \frac{d_n}{nhh_1}\sum_{t=1}^n K[(z_t-z)/h]\,E\Delta_t+\frac{d_n}{nhh_1}\sum_{t=1}^n K[(z_t-z)/h]\,(\Delta_t-E\Delta_t)\\
&= \big[A^TD_wA+o(1)\big]\,\frac{d_n}{nh}\sum_{t=1}^n K[(z_t-z)/h]+O_P\big[(\frac{d_n}{nhh_1})^{1/2}\big]
\end{aligned}
$$

$$= \left[A^T D_w A + o(1)\right] \frac{d_n}{nh} \sum_{t=1}^n K[(z_t - z)/h] + o_P(1), \tag{4.16}$$

due to $nhh_1/d_n \to \infty$. Similarly, we have

$$\begin{aligned} \frac{d_n}{nhh_1} \sum_{t=1}^n (A^T V_t)^2 &= \left[A^T D_w A + o(1)\right] \int_\Omega K_1^2(x)dx \, \frac{d_n}{nh} \sum_{t=1}^n K^2[(z_t - z)/h] \\ &\quad + o_P(1), \end{aligned} \tag{4.17}$$

and

$$\begin{aligned} \sum_{t=1}^n \left(||V_t|| + ||V_t||^{2+\delta/2}\right) &= \sum_{t=1}^n (||x_t|| + ||x_t||^{2+\delta}) K[(z_t - z)/h] \, K_1[(w_t - w)/h_1] \\ &\leq O(h_1) \sum_{t=1}^n K[(z_t - z)/h] + O_P[(nhh_1/d_n)^{1/2}] \\ &= O_P(nhh_1/d_n), \end{aligned} \tag{4.18}$$

where we have used the fact that $\sum_{t=1}^n K[(z_t - z)/h] = O_P(nh/d_n)$. By virtue of Theorem 2.21 of Wang (2015), results (4.16)-(4.17) imply that

$$\begin{aligned} &\Big\{ \frac{\sum_{j=1}^{[nt]} \nu_j}{\sqrt{n}}, \ \frac{\sum_{j=1}^{[nt]} \nu_{-j}}{\sqrt{n}}, \ \frac{d_n}{nhh_1} A^T D_n A, \ \frac{d_n}{nhh_1} \sum_{t=1}^n (A^T V_t)^2 \Big\} \\ \Rightarrow \ &\big\{ B_t, \ B_{-t}, \ (A^T D_w A) \, L_Z(1,0), \ \tau_1 \, (A^T D_w A) \, L_Z(1,0) \big\}, \end{aligned} \tag{4.19}$$

on $D_{R^4}[0,\infty)$, where $B = \{B_t\}_{t \in R}$ is a standard Brown motion and

$$\tau_1 = \int_\Omega K_1^2(x)dx \int_\Omega K^2(x)dx.$$

We are now ready to prove (4.14) and (4.15). By noting that D_w is positive-definite, it is readily seen from (4.16) that $D_n^{-1} = O_P(d_n/nhh_1)$. This, together with (4.18), yields that

$$||R_n|| = (|h| + |h_1|) \, O_P[(d_n/nhh_1)^{1/2}] \sum_{t=1}^n ||V_t|| = O_P\big[(|h| + |h_1|) \, (nhh_1/d_n)^{1/2}\big] = o_P(1),$$

implying (4.14).

To prove (4.15), write $u_{nt} = \left(\frac{d_n}{nhh_1}\right)^{1/2} A^T V_t$, namely, we have $\left(\frac{d_n}{nhh_1}\right)^{1/2} A^T S_n = \sum_{t=1}^n \epsilon_t \, u_{nt}$. By using (4.16), routine calculations show that $\frac{1}{\sqrt{n}} \sum_{t=1}^n |u_{nt}| = o_P(1)$ and

$$\max_{1 \leq t \leq n} |u_{nt}| \ \leq \ \left(\frac{d_n}{nhh_1}\right)^{1+\delta/4} \sum_{t=1}^n |A^T V_t|^{2+\delta/2} = o_P(1).$$

Now, by recalling **A8** and (4.19), (4.15) follows from Wang's extended martingale limit theorem, e.g., Wang (2014 or Theorem 3.14 of Wang (2015) . The proof of Theorem 2.4 is complete. □

5 Appendix

A sequence $\{\xi_k, k \geq 1\}$ is said to be α mixing if the α mixing coefficient

$$\alpha(n) \;:=\; \sup_{k\geq 1} \sup \{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{F}_{n+k}^{\infty}, B \in \mathcal{F}_1^k\}$$

converges to zero as $n \to \infty$, where $\mathcal{F}_l^m$ denoted the $\sigma-$algebra generated by $\xi_l, ..., \xi_m$ with $l \leq m$.

The following results for the moment properties of α-mixing sequence are well-known (e.g., McLeish,1975 or Hall and Heyde, 1980, page 278), which will be used in the proofs of other results.

Suppose $X \in \mathcal{F}_k^{\infty}$ and $Y \in \mathcal{F}_{-\infty}^{i}$), where $k > i$. Then,

(a). for any $1 \leq p \leq r \leq \infty$,

$$\|E(X|\mathcal{F}_{-\infty}^{i}) - EX\|_p \leq 2(2^{1/p}+1)\{\alpha(k-i)\}^{1/p-1/r}\|X\|_r; \tag{5.1}$$

(b) for any p, $q > 1$, $p^{-1} + q^{-1} < 1$,

$$|EXY - EXEY| \leq 8\|X\|_p\|Y\|_q\{\alpha(k-i)\}^{1-p^{-1}-q^{-1}}. \tag{5.2}$$

Lemma 5.1 *Under* **A1**, *for any $h \to 0$ satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$, result (2.1) holds .*

Proof. Write $A_k = K\big[(z_k - z)/h\big]\,\epsilon_k$, $W_{nk} = A_k/\sqrt{nh}$ and $R_n(t) = \sum_{k=1}^{[nt]} W_{nk}$. It is well-known (see, e.g., Davidson (1994)) that

$$\frac{x_{[nt]}}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} \eta_k \Rightarrow B(t),$$

namely, $\{x_{[nt]}/\sqrt{n}\}_{n\geq 1}$ is tight. As a consequence, to prove (2.1), it suffices to show that

(i) the finite dimensional distributions of $\big(x_{[nt]}/\sqrt{n},\ R_n(t)\big)$ converges to that of $(B(t), \sigma_z B_1(t))$;

(ii) $\{R_n(t)\}_{n\geq 1}$ is tight.

The proof of the finite dimensional convergence is of somewhat standard. See, for instance, Cai, et al. (2000) with some routine modification. The independence between $B(t)$ and $B_1(t)$ comes from the fact that, for any $0 < t \leq 1$, the covariance of $x_{[nt]}/\sqrt{n}$ and $R_n(t)$ converges to zero in probability. Indeed, by using (5.2) and Lemmas 4.1, we have

$$\begin{aligned}
|\mathrm{Cov}\big(x_{[nt]}/\sqrt{n}, R_n(t)\big)| &\leq \frac{1}{n\sqrt{h}}\sum_{k=1}^{n} E|\eta_k A_k| + \frac{2}{n\sqrt{h}}\sum_{k=1}^{n}\sum_{j=0}^{n-k} |E(\eta_k A_{k+j})| \\
&\leq 8\,h^{-1/2}\,(E|A_1|^{7/4})^{4/7}\,(E|\eta_1|^3)^{1/3}\sum_{j=0}^{\infty} j^{-2\gamma/21} \\
&\leq C\,h^{1/14} \to 0,
\end{aligned}$$

for any $0 < t \leq 1$ and $\gamma > 21/2$. Simple calculations by using similar arguments [see, e.g., Lemma A1 (c) of Cai, et al. (2000)] also yield that

$$\sup_{n\geq 1} E R_n^2(t) \leq C\, t, \tag{5.3}$$

indicating that $\{R_n(t)\}_{n\geq 1}$, for any $0 < t \leq 1$, is uniformly integrable. This fact will be used later.

We next prove the tightness. To this end, let $\mathcal{F}_k = \sigma(z_i, \epsilon_i; i \leq k)$,

$$\beta_{nk} \;=\; \sum_{i=1}^{\infty} E(W_{n,i+k}|\mathcal{F}_k), \quad w_{nk} = \sum_{i=0}^{\infty} \left[E(W_{n,i+k}|\mathcal{F}_k) - E(W_{n,i+k}|\mathcal{F}_{k-1})\right.$$

It is well-known that β_{nk} and w_{nk} are well defined and $W_{nk} = w_{nk} + \beta_{n,k-1} - \beta_{nk}$. Since $R_n(t) = \sum_{k=1}^{[nt]} w_{nk} + \beta_{n,[nt]}$, the tightness of $R_n(t)$ will follow if we prove that $\sum_{k=1}^{[nt]} w_{nk}$ is tight and

$$E \max_{1\leq k\leq n} |\beta_{nk}| \;=\; o_P(1). \tag{5.4}$$

Note that $\{w_{nk}, \mathcal{F}_k\}$ forms a sequence of martingale differences and the finite dimensional distribution converges to a joint normal distribution. To prove $\sum_{k=1}^{[nt]} w_{nk}$ is tight, it suffice to show that, for any $t > 0$, $\sum_{k=1}^{[nt]} w_{nk}$ is uniformly integrable [see, e.g., Proposition 1.2 of Aldous (1989)], which follows from (5.3), (5.4) and the fact $R_n(t) = \sum_{k=1}^{[nt]} w_{nk} + \beta_{n,nt}$ again.

It remains to prove (5.4). Note that $EA_1 = 0$ and $E|A_1|^r \leq C(z)h$ for any $1 \leq r \leq 3$ by (4.1). Standard arguments by using (5.1), together with $\alpha(n) = O(n^{-\gamma})$ for some $\gamma > 0$, show that, for any $1 \leq p < 3$ and $0 < \alpha \leq 3 - p$,

$$\Big(E|E(A_{i+k}|\mathcal{F}_i)|^p\Big)^{1/p} \leq C\alpha(k)^{\alpha/p(p+\alpha)} \left(E|A_1|^{p+\alpha}\right)^{1/(p+\alpha)}$$

and

$$(E|\beta_{ni}|^p)^{1/p} \leq (nh)^{-1/2} \sum_{k=1}^{\infty} \left(E|E(A_{i+k}|\mathcal{F}_i)|^p\right)^{1/p} \leq C\,(nh)^{-1/2}\, h^{1/(p+\alpha)} \sum_{k=1}^{\infty} k^{-\gamma\alpha/p(p+\alpha)}. \tag{5.5}$$

This implies that, for any $2 < p < 3$, $0 < \alpha \leq 3 - p$ and $\gamma\alpha/p(p+\alpha) > 1$,

$$E \max_{1\leq k\leq n} |\beta_{nk}| \;\leq\; \Big[\sum_{i=1}^{n} E|\beta_{ni}|^p\Big]^{1/p} \leq C\left(nh^{1+\alpha/[(p+\alpha)(p/2-1)]}\right)^{(1-p/2)/p}.$$

We now establish (5.4) by taking $\gamma > 6/\delta$, and α sufficiently small so that

$$nh^{1+\alpha/[(p+\alpha)(p/2-1)]} \geq nh^{1+\delta} \to \infty.$$

The proof of Lemma 5.1 is now complete. □

Proof of Lemma 4.2. We only prove (4.6). Due to the local boundedness of $H(x)$, (4.4) is obvious. The proof of (4.5) follows from similar arguments as in proof of (4.6). We omit the details.

As in the proof of Lemma 5.1 [or (2.1)], let $A_i = K\big[(z_i - z)/h\big]\,\epsilon_i$, $\mathcal{F}_t = \sigma(\eta_{i+1}, z_i, \epsilon_i, 0 < i \le t)$, and $\mathcal{F}_s = \sigma(\phi, \Omega)$ be the trivial σ-field for $s \le 0$, and by putting

$$u_i = \sum_{k=1}^{\infty} E(A_{i+k}|\mathcal{F}_i) \quad \text{and} \quad v_i = \sum_{k=0}^{\infty} [E(A_{i+k}|\mathcal{F}_i) - E(A_{i+k}|\mathcal{F}_{i-1})],$$

$\{v_i, \mathcal{F}_i\}_{i\ge1}$ forms a sequence of martingale differences and, as in Liang, et al. (2016),

$$\begin{aligned}\sum_{k=1}^{n} H_1(x_{nk})A_k = & \sum_{k=1}^{n} H_1(x_{nk})(v_k + u_{k-1} - u_k) \\ = & \sum_{k=1}^{n} H_1(x_{nk})v_k + \sum_{k=1}^{n}[H_1(x_{n,k+1}) - H_1(x_{n,k})]u_k - H_1(x_{n,n+1})u_n \\ = & \sum_{k=1}^{n} H_1(x_{nk})v_k + R(n), \quad \text{say.}\end{aligned} \tag{5.6}$$

As in the proof of (5.4), we have

$$\max_{1\le i\le n} |A_i - v_i| \;\le\; 2 \max_{1\le i\le n} |u_i| = o_P(\sqrt{nh}). \tag{5.7}$$

This, together with (2.1) and (4.4), implies that

$$\begin{aligned}\Big(x_{n,[nt]}, \frac{1}{\sqrt{nh}} \sum_{k=1}^{[nt]} v_k\Big) \;=\; & \Big\{\frac{x_{[nt]}}{\sqrt{n}},\ \frac{1}{\sqrt{nh}} \sum_{t=1}^{[nt]} K\big[(z_t - z)/h\big]\epsilon_t\Big\} + o_P(1) \\ \Rightarrow\; & \big\{B(t), \sigma_z B_1(t)\big\},\end{aligned}$$

on $D_{R^2}[0,1]$. Now, by recalling that $B_1(t)$ is independent of $B(t)$, standard argument on the convergence to stochastic integrals yields that

$$\Big(\frac{1}{n}\sum_{t=1}^{n} H(x_{nt}), \frac{1}{\sqrt{nh}} \sum_{k=1}^{n} H_1(x_{nk})v_k\Big) \to_D \Big\{\int_0^1 H(B_s)ds, \sigma_z \Big(\int_0^1 H_1^2(B_s)ds\Big)^{1/2} N\Big\},$$

where $N \sim N(0,1)$ independent of $B(t)$. Taking this estimation into (5.6), (4.6) will follow if we prove

$$|R(n)| \;=\; o_P(\sqrt{nh}). \tag{5.8}$$

To this end, write $\Omega_K = \{x_{ni} : \max_{1\le i\le n+1} |x_{ni}| \le K\}$. Note that (5.5) implies $E|u_1|^p \le C h^{p/(p+\alpha)}$ for any $\alpha > 0$ and $1 \le p \le 3$. It follows from (4.3) and $E||\eta_1||^3 < \infty$ that

$$\begin{aligned}E\Big[|R(n)|I(g(\Omega_K)\Big] \;\le\; & C_K \Big(\sum_{k=1}^{n} E\big(||x_{n,k+1} - x_{n,k}||\,|u_k|\big) + E|u_n|\Big) \\ \le\; & \frac{C_K}{\sqrt{n}} \sum_{k=1}^{n} E(||\eta_k||\,|u_k|) + o(1) \\ \le\; & C_K \sqrt{n}\,\big(E||\eta_1||^3\big)^{1/3} (E|u_1|^{3/2})^{2/3} + o(1)\end{aligned}$$

$$\leq \; C_K \sqrt{n}\, h^{2/(3+2\alpha)} + o(1) = o(\sqrt{nh}),$$

by taking $\alpha < 1/2$. This implies that $R(n) = o_P(\sqrt{nh})$ due to $P(\Omega_K) \to 1$ as $K \to \infty$.

The proof of Lemma 4.2 is complete. ■

Proof of Lemma 4.3. We only provide a outline. Results (4.2) and (4.5) follow from (2.94) and Theorem 3.18 of Wang (2015), respectively. By using similar arguments as in proof of (2.96) in Wang (2015), for any $l, m \geq 0$, we have

$$\begin{aligned} & E\Big(\sum_{k=1}^{n} \big[\eta_{k-l}\eta_{k-m}^T - E(\eta_{k-l}\eta_{k-m}^T)\big] K[(z_k - z)/h]\Big)^2 \\ \leq \; & C\,(1 + \max\{l^{1/2}, m^{1/2}\} + h\log n)\,\big[E(\eta_1\eta_1^T) + E(\eta_1\eta_2^T)\big]\, nh/d_n. \end{aligned}$$

This, together with Hólder's inequality, yields that

$$\begin{aligned} & E\Big|\sum_{k=1}^{n}(u_k - Eu_k)K\Big(\frac{z_k - z}{h}\Big)\Big| \\ \leq \; & \sum_{l,m=0}^{\infty} \|\varphi_l\|\, \|\widetilde{\varphi}_m\|\, E\Big|\sum_{k=1}^{n} \big[\eta_{k-l}\eta_{k-m}^T - E(\eta_{k-l}\eta_{k-m}^T)\big] K[(z_k - z)/h]\Big| \\ = \; & O\big[\big(nh/d_n\big)^{1/2}\big] \sum_{l,m=0}^{\infty} \Big(l^{1/4}m^{1/4}\|\varphi_l\|\, \|\widetilde{\varphi}_m\|\Big), \end{aligned}$$

implying (4.3). To see (4.4), let $u_{kM} = \sum_{l,m=0}^{M} \varphi_l^T\, \eta_{k-l}\, \eta_{k-m}^T\, \widetilde{\varphi}_m$ and $\bar{u}_k = u_k - Eu_k, \bar{u}_{kM} = u_{kM} - Eu_{kM}$. For any $M \geq 1$, (3.8) of Wang and Phillips (2009b) implies that

$$\begin{aligned} & \Big\{\frac{d_n}{nh}\sum_{t=1}^{n} K\Big(\frac{z_t - z}{h}\Big),\; \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n} \bar{u}_{kM}\, K\Big(\frac{z_t - z}{h}\Big)\Big\} \\ & \qquad \to_D \; \Big\{L_Z(1,0), a_{M2}\, L_Z^{1/2}(1,0)\, N\Big\}, \end{aligned} \tag{5.9}$$

where $a_{M2}^2 = E(u_{kM}^2)\int_{-\infty}^{\infty} K^2(t)dt$. Since $a_{M2}^2 \to a_2^2$ as $M \to \infty$, (4.4) follows easily from (5.9) and the fact:

$$\begin{aligned} & \big(\frac{d_n}{nh}\big)^{1/2}\sum_{k=1}^{n}(\bar{u}_k - \bar{u}_{kM})\, K\Big(\frac{z_t - z}{h}\Big) \\ = \; & O_P(1)\Big(\sum_{l=M,m=0}^{\infty} + \sum_{l=0,m=M}^{\infty}\Big) l^{1/4}m^{1/4}\|\varphi_l\|\, \|\widetilde{\varphi}_m\| = o_P(1), \end{aligned}$$

as $n \to \infty$ first and then $M \to \infty$. The proof of Lemma 4.3 is now complete. ■

Acknowledgments. Liang acknowledges research support from the National Natural Science Foundation of China (11671299). Shen acknowledges research support from the international exchange program for graduate students, Tongji University (2016020028). Wang acknowledges research support from the Australian Research Council.

References

[1] Aldous, D.J. (1989). Stopping times and tightness, II, *Ann. Probab.* **17**, 586-595.

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

[3] Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series, *Journal of Amer. Assoc.*, **95**, 941-956.

[4] Chan, N. and Wang, Q. (2014). Uniform convergence for nonparametric estimators with nonstationary data. *Econometric Theory*, 30, 1110-1133.

[5] Chan, N. and Wang, Q. (2015) Nonlinear regressions with nonstationary time series. *Journal of Econometrics* **185(1)**, 182-195.

[6] Chang, Y. and Park, J.Y. (2003) Index models with integrated time series. *Journal of Econometrics* **114(1)**, 73-106.

[7] Chang, Y., Park, J.Y. and Phillips, P.C.B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *Econometrics Journal* **4**, 1-36.

[8] Davidson, J. (1994). *Stochastic Limit Theory*, Oxford University Press.

[9] Duffy, J. (2017). Uniform convergence rates on a maximal domain for structural nonparametric cointegrating regression. *Econometrics Journal* , forthcoming.

[10] Engle, R.F. and Granger, C.W.J. (1987). Cointegration and error correction: Representation, estimation, and testing. *Econometrica* **55**, 251-276.

[11] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications.* Chapman and Hall, London.

[12] Gao, J. and Phillips, P.C.B. (2013a). Semiparametric estimation in triangular system equations with nonstationarity, *Journal of Econometrics,* **176**, 59–79.

[13] Gao, J. and Phillips, P.C.B. (2013b). Functional coefficient nonstationary regression. *Cowles fundation discussion paper No 1911.*

[14] Granger, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* **16**, 121-130.

[15] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and its Application.* Academic Press.

[16] Li, K. Li, D. Liang, Z and Hsiao, C. (2017). Estimation of semi-varying coefficient models with nonstationary regressors. *Econometric Review*, **36**, 354–369.

[17] Liang H.-Y., Phillips, P.C.B., Wang, H. and Wang, Q. (2016). Weak convergence to stochastic integrals for econometric applications. *Econometric Theory* **32**, 1349-1375.

[18] McLeish, D.L. (1975) A maximal inequality and dependent strong laws. *Annals of Probability* **3**, 829-839.

[19] Park, J.Y. and Phillips, P.C.B. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* **4**, 468-497.

[20] Park, J.Y. and Phillips, P.C.B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69**, 117-161.

[21] Phillips, P.C.B (2009). Local Limit theorem and spurious regressions, *Econometric Theory*, **15**, 269-298.

[22] Sun, Y. Cai, Z and Li, Q. (2013). Semiparametric functional coefficient models with integrated covariaates, *Econometric Theory*, **29**, 659–672.

[23] Saikkonen P. (1995). Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems. *Econometric Theory* **11(5)**, 888-911.

[24] Terasvirta, T. Tjostheim, D. and Granger, C. W. J. (2011). *Modelling nonlinear economic time series*, Oxford University Press.

[25] Trokic, M. (2014). Functional Coefficient Models for Nearly (Possibly Weakly) $I(1)$ Processe, online, $https://papers.ssrn.com/sol3/papers.cfm?abstract_id = 2374495$.

[26] Wang, Q. (2014). Martingale limit theorems revisited and non-linear cointegrating regression. *Econometric Theory*, 30, 509-535.

[27] Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression.* World Scientific.

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

[29] Wang, Q. and Phillips, P.C.B. (2009b). Structural Nonparametric Cointegrating Regression, *Econometrica*, **25**, 1901-1948.

[30] Wang, Q. and Phillips, P. C. B. (2011). Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory*, 27, 235-259.

[31] Wang, Q. and Phillips, P.C. B.(2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory,* **32,** 359–401.

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