Functional-coefficient cointegrating regression with endogeneity *

Han-Ying Liang, Yu Shen and Qiying Wang

Tongji University, Tongji University and The University of Sydney

October 13, 2022

Abstract

Joon Y. Park is one of the pioneers in developing nonlinear cointegrating regression. Since his initial work with Phillips (Park and Phillips, 2001) in the area, the past two decades have witnessed a surge of interest in modeling nonlinear nonstationarity in macroeconomic and financial time series, including parametric, nonparametric and semiparametric specifications of such models. These developments have provided a framework of econometric estimation and inference for a wide class of nonlinear, nonstationary relationships. In honor of Joon Y. Park, this paper contributes to this area by exploring nonparametric estimation of functional-coefficient cointegrating regression models where the structural equation errors are serially dependent and the regressor is endogenous. 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 improve those of Cai, Li and Park (2009) and open up inference by conventional nonparametric method to a wide class of potentially nonlinear cointegrated relations.

Key words and phrases: Cointegration, Functional-coefficient model, nonstationary time series, endogeneity, kernel estimation, local linear estimation.

JEL Classification: C14, C22.

*The authors thank the Co-Editor J. I. Miller and the referee for constructive and insightful comments which lead to significant improvement of this paper. Liang acknowledges research support from the National Natural Science Foundation of China (12071348). Wang acknowledges research support from the Australian Research Council. Address correspondence to: Han-Ying Liang and Yu Shen, School of Mathematical Sciences, Tongji University, Shanghai 200092, P. R. China, e-mail: hyliang@tongji.edu.cn; Qiying Wang, School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia, e-mail: qiying.wang@sydney.edu.au.

1 Introduction

Linear cointegrating regression was suggested by Granger (1981) and Engle and Granger (1987) and has attracted extensive researches in both theory and empirical applications. The specification in linear structure is convenient for practical work and software packages have many standard routines for dealing with such a 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, for instance, Park and Phillips (1988), Saikkonen (1995) and Terasvirta, Tjostheim and Granger (2011).

To overcome such deficiencies, various nonlinear cointegrating regression models have been suggested in past two decades. Among many other contributors, we refer to Park and Phillips (2001), Chang, Park and Phillips (2001), Chang and Park (2003), Chan and Wang (2015) and Wang (2021) for parametric nonlinear cointegrating regression; Wang and Phillips (2009a, b, 2016), Kim and Kim (2012), Gao and Phillips (2013a), Duffy (2016, 2020), Dong and Linton (2018) and Wang, Phillips and Kasparis (2021) for nonparametric and semiparametric approaches that can cope with the unknown functional form of the response in a nonstationary time series setting; together with the references cited therein. Nonlinear cointegrating regression with functional-coefficients was introduced in Cai, Li and Park (2009) and Xiao (2009), where the authors suggested a model 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. Extension of the model (1.1) to more general formulations can be found in Gao and Phillips (2013b), Li, Li, Liang and Hsiao (2017), Hirukawa and Sakudo (2018) and Tu and Wang (2019, 2022). Also see Phillips and Wang (2022a, 2022b) for recent developments.

Model (1.1) allows cointegrating relationships 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. Practical examples in forecasting regional long-run energy and electricity demand can be found in Chang, Choi, Kim, Miller and Park (2016, 2021). Asymptotic theory of estimation and inference for model (1.1) and more general related models has been established in the literature. Technical difficulties, however, have confined much of the asymptotic theory to the case of sequential exogeneity where a martingale structure is usually assumed in the models. See, for instance, Cai, et al. (2009), Xiao (2009), Gao and

Phillips (2013a, b), Li, et al. (2017), Hirukawa and Sakudo (2018) and Tu and Wang (2019a, 2022). 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 explanatory variables may be expected to be temporally and contemporaneously correlated[1]. Exogeneity therefore limits 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 on the nonstationary regressor. Generalizing earlier models of Cai, et al. (2009), Xiao (2009) and others, 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. Unlike the papers cited above, our methodology in investigating the asymptotics builds up the techniques recently developed in Wang and Phillips (2009b, 2016), enabling our assumptions to be neat and our proofs to be 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.

In both situations, we allow for $E(\epsilon_t|x_t, z_t) \neq 0$, thereby introducing the endogeneity in model (1.1) and providing an essential extension to previous works in the related fields. It should be mentioned that model (1.1) has been extensively investigated in literature in case that both x_t and z_t are stationary. We only refer to Cai, Fan and Yao (2000), Fan and Zhang (2018) and citations therein. It is of great interest to consider the asymptotics for the nonparametric estimators of $\beta_0(\cdot)$ in model (1.1) when both x_t and z_t are $I(1)$ processes, but there are technical challenges at the moment. See Remark 8 for more details. We conclude in Section 3. The proofs of main results are given in Appendix A. The proofs of some auxiliary results are collected in Appendix B.

[1]To give an example, let us consider the demand for money. It has often been suggested that endogeneity exists in this typical cointgerated system, where the Central Bank supplies base money on demand at its prevailing interest rate, and broad money is created by the banking system. See, for instance, King (1994). A survey paper on the literature related to the demand for money across a range of industrial and developing countries can be found in Sarisam (2000).

Throughout the paper, we make use of the following notation: for $a = (a_{ij}), 1 \le i \le m, 1 \le j \le k$, $||a|| = \sum_{i=1}^m \sum_{j=1}^k |a_{ij}|$.

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 $n \to \infty$. When one of the regressors x_t and z_t is nonstationary, the limit behavior of $\widehat{\beta}_n(z)$ has been investigated 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 provides more general results with advantages for empirical applications. Our assumptions permit dependence between the error process ϵ_t and the regressors x_t or/and z_t. These relaxations of the conditions in previous works are particularly important in nonlinear cointegrated systems because finite time horizon independence 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 $V_{nj} = \sum_{t=1}^n x_t x_t^T\, K\big(\frac{z_t - z}{h}\big)(z_t - z)^j$ for $j = 0, 1$ and 2.

The asymptotics of $\widehat{\beta}_n(z)$ and $\widehat{\beta}_L(z)$ will be investigated in two different cases mentioned in the introduction. 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 an extension of the model to multivariate settings.

2.1 Model 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 strict 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 $\nu > 0$,

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

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

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 smoothness 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 the local linear estimator $\hat{\beta}_L(z)$. Under **A1**(i), we have $x_t = \sum_{j=1}^{t} \eta_j$, i.e., x_t is a standard $I(1)$ process. It is possible to allow for the x_t to be a nearly $I(1)$ process. Such an extension is omitted since it will involve complicated calculations. As in the situation where both x_t and z_t are stationary, the conditional mean $E(\epsilon_1|z_1) = 0$ in **A1** (ii) is necessary to establish the consistency for both estimators $\hat{\beta}_n(z)$ and $\hat{\beta}_L(z)$. However, under **A1**, we may have $E(\epsilon_t|z_t, x_t) \neq 0$, which introduces endogeneity in the model. This differs from the 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. It should be mentioned that, when the bandwidth parameter h converges to 0 satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$, the condition **A1** with $\gamma \geq \max\{21/2, 6/\delta\}$, together with **A2**(i), implies that

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

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

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

$B_1 = \{B_{1t}\}_{t\geq 0}$ is a standard Brownian motion independent of $B = \{B_t\}_{t\geq 0}$, and B 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 and the condition that $h \to 0$ satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$ is essential to enable the independence between two Brownian motions B_1 and B. To establish (2.1), there is a trade off condition (i.e., $\gamma \geq \max\{21/2, 6/\delta\}$) between the mixing coefficient in **A1**(i) and the convergence rate for the bandwidth parameter h. It is natural for such a joint convergence although it might be slightly stronger than necessary. For a proof of (2.1), see Lemma B.1 in Appendix B.

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:

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

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 cointegrating regression, due to the nonstationarity of the 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 $c_1 = 0$ (take $K(x)$ to be a symmetric kernel, for instance) 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

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

Remark 3. Results (2.2) and (2.3) provide a first order bias $c_1\beta_0'(z)h$. Surprisingly, it is unrealistic to add a deterministic higher order bias term into the result even if we have more smoothness conditions on $\beta_0(z)$. To see this claim, let $\tilde{K}(x) = xK(x)$, $c_1 = \int_{-\infty}^{\infty} xK(x)dx = 0$ and

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

In order to add a deterministic bias term having a order $O(h^2)$ in result (2.3), from the proof of Theorem 2.1 in Section A.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, as $nh^{5/2} = O(1)$,

$$\Lambda_n - c_0 nh^{5/2} = o_P(1), \tag{2.4}$$

for some constant c_0 (c_0 is allowed to be zero). This seems to be impossible except when $\tilde{K}(x) \equiv 0$. 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

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

i.e., $\sqrt{n}h/\Lambda_n = O_P(1)$, indicating that (2.4) is impossible whenever $nh^{5/2} = O(1)$. The similar phenomena has also be noticed in Sun and Li (2011).

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 is always 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):

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

where τ_1 is given in Theorem 2.1.

2.2 Model 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 $\lim_{|t|\to\infty}|t|^{\lambda}|Ee^{it\nu_1}| < \infty$ for some $\lambda > 0$.

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) $E(x_1\epsilon_1) = 0$ 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* $\hat{K}(x) = \int_{-\infty}^{\infty}e^{ixt}K(t)dt$.

Assumption **A4**(i) allows for short (under **SM**) and long (under **LM**) memory innovations ξ_j driving the (near) integrated regressor z_t given in Assumption **A4**(ii). As noticed in Wang and Phillips (2016), this setting is quite general for empirical applications. 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))

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

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 $L_Z(t,s)$ [2] with dual (time and space) parameters (t,s) in $[0,\infty) \times \mathbb{R}$. The characteristic function condition $\lim_{|t|\to\infty} |t|^\lambda |Ee^{it\nu_1}| < \infty$ for some $\lambda > 0$ is not necessary for the establishment of (2.7), but it is required for the convergence to local time in Lemma A.2 in Appendix A. 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)]. We may have $\mathrm{cov}(\epsilon_t, z_t) \neq 0$ under Assumption **A5**(i), which differs from the previous work where the model is often assumed to form a martingale difference sequence structure. In the latter case, $\mathrm{E}(\epsilon_t | x_t, z_t) = 0$. Assumption **A5** (ii) is necessary to establish the 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 $E||x_1 x_1^T|| e_1^2 < \infty$, 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.

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

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

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

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

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 mutually 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 the nonstationary case, the amount of time spent by the process z_t around any particular spatial 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 was first explained 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

$$Ex_{11}\epsilon_1 = Ex_1\epsilon_1 + A_0 E\epsilon_1 = 0.$$

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 Multivariate extension

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 functions defined on $\mathbb{R}^{1+d_1}$. As in the one-dimension situation, the local kernel estimator $\hat{\beta}_0(\cdot, \cdot)$ of $\beta_0(\cdot, \cdot)$ can be similarly defined by

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

where $K(x), L_j(x)$ 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 is 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 consistent estimator, it is also essential to assume that w_t is stationary. 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 assumptions.

A7 (a) The kernels $K(x)$ and $L_1(x)$ have a common compact support satisfying $\int_{-\infty}^{\infty} K(x)dx = \int_{-\infty}^{\infty} L_1(x)dx = 1$ 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

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

where $C > 0$ is an absolute constant, whenever δ_1 and δ_2 are sufficiently small.

A8 (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 $\Gamma_i(.), 1 \le i \le d$, 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 $d_{ij}(w) = E(x_{m_0 i}x_{m_0 j}|w_{m_0} = w)$.

A9 $\{\epsilon_i, \mathcal{F}_{i+1}\}_{i\ge1}$, where $\mathcal{F}_i = \sigma(\eta_i, \eta_{i-1}, ...)$, is a martingale difference sequence 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 $D_n = \sum_{t=1}^n x_t x_t^T K[(z_t - z)/h]L_1[(w_t - w)/h_1]$ and $\tau^2 = \sigma^2 \int_{-\infty}^{\infty} K^2(x)dx \int_{-\infty}^{\infty} L_1^2(x)dx$, $\mathbb{N} \sim N(0, I_d)$ is a standard d-dimensional normal vector.

Remark 7. Under Assumption **A9**, we have $E(\epsilon_t|x_t, z_t, w_t) = 0$. Unlike the asymptotics developed in Theorem 2.3, such a martingale difference restriction seems hard to be reduced in the multivariate extension to model (1.1) due to the technical reasons. This also indicates that a complicated calculation will be involved in investigating the asymptotics of $\hat{\beta}_0(z, w)$, even when the regressor w_t in model (2.10) is stationary.

Remark 8. Theorem 2.4 provides an extension of Theorem 5.7 in Wang (2015) to a functional-coefficient cointegrating regression model. In a related paper, Gao and Phillips (2013b) [also see Sun, et al. (2013)] investigated model (2.10) in the case that 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 interest to remove these restrictions, in particular, to establish the asymptotics without imposing $E(\epsilon_t|x_t, z_t, w_t) = 0$ as those given in Theorems 2.1-2.3.

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, indicating that the model (1.1) may be estimated by kernel regression just as in the case that both x_t and z_t are stationary. Importantly, and in contrast to stationary nonparametric regression, our asymptotic results allow for endogenous regressor in the models, namely, we assume $E(\epsilon_t|x_t, z_t) \neq 0$ in (1.1). As explained in Wang and Phillips (2009b), the reason for this robustness to endogeneity in the regressor is that nonstationary regressors such as unit root processes have a wandering character that assists in tracing out the true regression function. 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. Finally, in both situations described above, we suggest that it is always better to use the local linear non-parametric estimator with symmetric kernel rather than the Nadaraya-Waston estimator in empirical applications, in terms of the bias deduction and uniform asymptotics over wide domains. Furthermore, to ensure our theoretical results work, a unit root pre-test on x_t (or z_t) is essentially necessary.

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 non-stationary 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., Choi, Y., Kim, C.S. Miller, J. I. and Park, J. (2021). Disentangling Temporal Patterns in Elasticities: A Functional Coefficient Panel Analysis of Electricity Demand. *Energy Economics*, **60**, 232-243.

[7] Chang, Y., Choi, Y., Kim, C.S. Miller, J. I. and Park, J. (2021). Forecasting Regional Long-Run Energy Demand: A Functional Coefficient Panel Approach. *Energy Economics*, **96**, 105-117.

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

[9] 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.

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

[11] Dong, C. and Linton, O. (2018). additive nonparametric models with time variables and both stationary and nonstationary regressors. *Econometric Theory*, **34**(4), 754-789.

[12] Duffy, JA (2016). A uniform law for convergence to the local times of linear fractional stable motions. *The Annals of Applied Probability*, **25(1)**, 45-72.

[13] Duffy, JA (2017). Uniform convergence rates over maximal domains in structural nonparametric cointegrating regression. *Econometric Theory*, **33**, 1387-1417.

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

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

[16] Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface*, **1**, 179–195.

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

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

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

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

[21] Hirukawa, M. and Sakudo, M. (2018). Functional-coefficient cointegration models in the presence of deterministic trends. *Econometric Review,* **37**(5), 507-533.

[22] King, M. (1994). The transmission mechanism of monetary policy. *Bank of England Quarterly Bulletin.* August.

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

[24] 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.

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

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

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

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

[29] Phillips, P. C. B. and Wang, Y. (2022a). Functional coefficient panel modeling with communal smoothing covariates, *Journal of Econometrics*, **227**, 371-407.

[30] Phillips, P. C. B and Wang, Y. (2022b). When bias contributes to variance: true limit theory in functional coefficient cointegrating regression. *Econometric Theory,* **25,** 1466-1497.

[31] Sriram, S. S.(2000). A Survey of Recent Empirical Money Demand Studies. *IMF Staff Papers* **47**, 334-365.

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

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

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

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

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

[37] Tu, Y. and Wang, Y. (2019). Functional Coefficient Cointegration Models Subject to Time–Varying Volatility with an Application to the Purchasing Power Parity. *Oxford Bulletin of Economics and Statistics*, **81 (6)**, 1401 - 1423.

[38] Tu, Y. and Wang, Y. (2022). Spurious functional-coefficient regression models and robust inference with marginal integration *Journal of Econometrics*, **229**, pp 396-421.

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

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

[41] Wang, Q. (2021). Least squares estimation for nonlinear regression models with heteroscedasticity, *Econometric Theory*, forthcoming.

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

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

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

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

[46] Wang, Q., Phillips, P.C. B. and Kasparis, I. (2021). Latent Variable Nonparametric Cointegrating Regression. *Econometric Theory,* **37**, 138-168.

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

Appendix

A Proofs of main results

Since the methodology is different, the proofs of Theorems 2.1-2.2, 2.3 and 2.4 will be given in Sections A.1, A.2 and A.3, respectively.

A.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. Furthermore, we always assume **A1** and **A2**(i) hold in the following lemmas.

Lemma A.1 *(a) For any* $0 \leq \alpha \leq 3$*, we have*

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

where C_z *is a constant depending only on* z*; (b) As* $h \to 0$*, we have*

$$h^{-1} EK_j\big[(z_1 - z)/h\big] = g(z)\mu_j + o(1), \quad j = 0, 1, 2. \tag{A.2}$$

The proof of Lemma A.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 $||x|| + ||y|| \leq K$,

$$\big|H_1(x) - H_1(y)\big| \leq C_K ||x - y||, \tag{A.3}$$

where C_K is a constant depending only on K.

Lemma A.2 *(i) For any real function* $A_n(x, y)$*,*

(a) we have

$$\frac{1}{n} \sum_{t=1}^{n} H(x_{nt}) A_n(z_t, \epsilon_t) = O_P\big(E|A_n(z_1, \epsilon_1)|\big); \tag{A.4}$$

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

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} H_1(x_{nt}) A_n(z_t, \epsilon_t) \quad = \quad O_P\Big\{\big[E|A_n(z_1, \epsilon_1)|^{2+\alpha}\big]^{1/(2+\alpha)}\Big\}. \tag{A.5}$$

(ii) For any $h \to 0$ satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$, if $\gamma \geq \max\{21/2, 6/\delta\}$ where γ is given in **A1***(i), then*

$$\Big\{\frac{1}{n}\sum_{t=1}^{n} H(x_{nt}),\ \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} H_1(x_{nt})K[(z_t-z)/h]\,\epsilon_t\Big\}$$
$$\to_D\ \Big\{\int_0^1 H(B_s)ds,\ a_1\Big(\int_0^1 H_1^2(B_s)ds\Big)^{1/2} N\Big\}, \tag{A.6}$$

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

The proof of Lemma A.2 will be given in Appendix B. Note that $|K_j(x)| \leq CK(x)$, where $C > 0$ is an absolute constant, as $K(x)$ has a compact support. Result (A.4), together with (A.1), implies that, as $h \to 0$,

$$\frac{1}{nh}\sum_{t=1}^{n} ||x_{nt}x_{nt}^T||\,|K_j\big[(z_t-z)/h\big]| \;=\; O_P(1), \quad j = 0, 1, 2. \tag{A.7}$$

Similarly, by using (A.2) and (A.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

$$\begin{aligned}
\Delta_{nj} \;&:=\; \frac{1}{nh}\sum_{t=1}^{n} x_{nt}x_{nt}^T K_j\big[(z_t-z)/h\big] \\
&=\; \frac{1}{nh}\sum_{t=1}^{n} x_{nt}x_{nt}^T\, EK_j\big[(z_t-z)/h\big] + \frac{1}{nh}\sum_{t=1}^{n} x_{nt}x_{nt}^T\big[K_j\big[(z_t-z)/h\big] - EK_j\big[(z_t-z)/h\big]\big] \\
&=\; [g(z)\mu_j + o(1)]\,\frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + O_P((nh^{1+\alpha/(2+\alpha)})^{-1/2}) \\
&=\; g(z)\mu_j\,\frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + o_P(1), \quad j = 0, 1, 2.
\end{aligned} \tag{A.8}$$

by taking α sufficiently small so that $nh^{1+\alpha/(2+\alpha)} \geq nh^{1+\delta} \to \infty$. Furthermore, it follows from (A.6) that, for any $h \to 0$ satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$,

$$\Big\{(\frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T,\ \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K[(z_t-z)/h]\,\epsilon_t\Big\}$$
$$\to_D\ \Big\{\int_0^1 B_sB_s^T ds,\ a_1\Big(\int_0^1 B_sB_s^T ds\Big)^{1/2}\mathbb{N}\Big\}, \tag{A.9}$$

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

$$n\sqrt{h}\Big(\widehat{\beta}_n(z) - \beta_0(z) - c_1\beta_0'(z)h\Big) = \Delta_{n0}^{-1}\big(S_n + R_{n1} + R_{n2}\big), \tag{A.10}$$

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 (A.7) imply that, for some $\nu > 0$,

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

Write $A_n(z_t, \epsilon_t) = K_1[(z_t - z)/h] - c_1 K[(z_t - z)/h]$. Lemma A.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 (A.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 (A.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 (A.8) - (A.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

$$
n\sqrt{h}\Big(\widehat{\beta}_L(z) - \beta_0(z) - c_2 \beta_0''(z) h^2\Big) \;=\; \Delta_n^{-1} \big(P_n + T_{n1} + \beta_0''(z) n h^{5/2}\, T_{n2}\big), \tag{A.11}
$$

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:

$$
\sum_{t=1}^{n} w_t x_{nt} x_{nt}^T K[(z_t - z)/h](z_t - z) = 0. \tag{A.12}
$$

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

$$n^{-2}h^{1-j}V_{nj} = \Delta_{nj} = g(z)\mu_j \frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + o_P(1), \quad j = 0, 1, 2,$$

by (A.8). It is readily seen from (A.8) and Lemma A.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 (A.11), we obtain

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

as $nh^{5/2} = O(1)$ and $nh^{1+\delta} \to \infty$ for some $\delta > 0$, due to (A.9) and the continuous mapping theorem. Theorem 2.2 is now proved. ■

A.2 Proof of Theorem 2.3

As in Section A.1, let $K_j(x) = x^jK(x)$ and $\mu_j = \int_{-\infty}^{\infty} K_j(x)dx$ for $j \ge 0$. Let

$$u_k = \sum_{l,m=0}^{\infty} \varphi_l^T\,\eta_{k-l}\,\eta_{k-m}^T\,\widetilde{\varphi}_m, \tag{A.13}$$

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 lemmas. Under the conditions **A4** and **A6**, the proof of Lemma A.3 is similar to Lemma 2.2 and Theorem 3.16 of Wang (2015). A outline will be given in Appendix B. For a proof of Lemma A.4, we refer to Theorem 3.18 of Wang (2015). See, also, Wang and Phillips (2011).

Lemma A.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*

$$\sum_{k=1}^{n}(1+|u_k|)K_j\Big(\frac{z_k - z}{h}\Big) = O_P\big(nh/d_n\big); \tag{A.14}$$

$$\sum_{k=1}^{n}\big(u_k - Eu_k\big)K_j\Big(\frac{z_k - z}{h}\Big) = O_P\Big(\big(\frac{nh}{d_n}\big)^{1/2}\Big)\sum_{l,m=0}^{\infty}\Big(l^{1/4}m^{1/4}\|\varphi_l\|\,\|\widetilde{\varphi}_m\|\Big), \tag{A.15}$$

$j = 0, 1, 2$, and

$$\begin{aligned}&\Big\{\frac{d_n}{nh}\sum_{t=1}^{n}K\Big(\frac{z_t - z}{h}\Big),\ \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^{n}\big(u_k - Eu_k\big)K\Big(\frac{z_t - z}{h}\Big)\Big\}\\ &\qquad \to_D\ \Big\{L_Z(1,0), a_2\, L_Z^{1/2}(1,0)\, N\Big\},\end{aligned} \tag{A.16}$$

where $a_2^2 = E(u_1^2)\int_{-\infty}^{\infty} K^2(t)dt$, and N is standard normal variate independent of $L_Z(1,0)$;

Lemma A.4 *Let $f(x)$ be a real function having a compact support. If $\int_{-\infty}^{\infty} f(x)dx = 0$, then*

$$\sum_{k=1}^{n} f\Big(\frac{z_k - z}{h}\Big) = O_P\Big(\big(nh/d_n\big)^{1/2}\Big), \tag{A.17}$$

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 A.3 that

$$\begin{aligned}D_{nj} &:= \frac{d_n}{nh}\sum_{t=1}^{n} x_t x_t^T K_j\Big(\frac{z_t - z}{h}\Big)\\ &= \frac{d_n}{nh}\sum_{t=1}^{n} E(x_t x_t^T)K_j\Big(\frac{z_t - z}{h}\Big) + \frac{d_n}{nh}\sum_{t=1}^{n}\big[x_t x_t^T - E(x_t x_t^T)\big]K_j\Big(\frac{z_t - z}{h}\Big)\\ &= E(x_1 x_1^T)\frac{d_n}{nh}\sum_{t=1}^{n} K_j\Big(\frac{z_t - z}{h}\Big) + O_P\Big(\big(\frac{d_n}{nh}\big)^{1/2}\Big)\\ &= E(x_1 x_1^T)\frac{d_n}{nh}\sum_{t=1}^{n} K_j\Big(\frac{z_t - z}{h}\Big) + o_P(1), \quad j = 0, 1, 2,\end{aligned} \tag{A.18}$$

as $nh/d_n \to \infty$. Furthermore, due to $E(x_1\epsilon_1) = 0$, it follows from (A.16) and the continuous mapping theorem that

$$\Big\{\frac{d_n}{nh}\sum_{t=1}^n K\Big(\frac{z_t - z}{h}\Big),\ \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t\epsilon_t K\Big(\frac{z_t - z}{h}\Big)\Big\}$$
$$\to_D \ \Big\{L_Z(1,0), a_2\, L_Z^{1/2}(1,0)\,\mathbb{N}\Big\}, \tag{A.19}$$

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

$$\big(\frac{nh}{d_n}\big)^{1/2}\Big(\widehat{\beta}_n(z) - \beta_0(z) - c_2\beta_0''(z)h^2\Big) = D_{n0}^{-1}\Big(S_n + R_{n1} + \beta_0'(z)R_{n2} + \beta_0''(z)R_{n3}\Big), \tag{A.20}$$

where

$$\begin{aligned}
S_n &= \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t\epsilon_t K\Big(\frac{z_t - z}{h}\Big),\\
R_{n1} &= \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t x_t^T v_t,\\
R_{n2} &= h\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t x_t^T K_1\Big(\frac{z_t - z}{h}\Big),\\
R_{n3} &= h^2\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t x_t^T\Big\{\frac{1}{2}K_2[(z_t - z)/h] - c_2K[(z_t - z)/h]\Big\}.
\end{aligned}$$

From **A3**(ii) and (A.14), as $nh^5/d_n = O(1)$ we have

$$\|R_{n1}\| \le Ch^{2+\eta}\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n \|x_t x_t^T\|\, K\Big(\frac{z_t - z}{h}\Big) = O_P\Big(\big(\frac{nh}{d_n}\big)^{1/2} h^{2+\eta}\Big) = o_P(1). \tag{A.21}$$

From (A.15) with $j = 1$ and (A.17) with $f(x) = K_1(x)$, as $h \to 0$ we have

$$\begin{aligned}
R_{n2} &= h\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n \big[x_t x_t^T - E(x_t x_t^T)\big]K_1\Big(\frac{z_t - z}{h}\Big)\\
&\quad + h\,\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n E(x_t x_t^T)K_1\Big(\frac{z_t - z}{h}\Big)\\
&= O_P(h) = o_P(1).
\end{aligned} \tag{A.22}$$

Note that $\int_{-\infty}^{\infty}\big[\frac{1}{2}K_2(x) - c_2K(x)\big]dx = 0$. It follows from (A.15) and (A.17) again that, as $nh^5/d_n = O(1)$,

$$R_{n3} = h^2\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n \big[x_t x_t^T - E(x_t x_t^T)\big]\Big\{\frac{1}{2}K_2[(z_t - z)/h] - c_2K[(z_t - z)/h]\Big\}$$

$$
\begin{aligned}
&+\Big(\frac{nh^5}{d_n}\Big)^{1/2}\frac{d_n}{nh}\sum_{t=1}^n E(x_tx_t^T)\Big\{\frac12 K_2[(z_t-z)/h]-c_2K[(z_t-z)/h]\Big\}\\
=\;& O_P(h^2)+O_p\Big(\big(nh^5/d_n\big)^{1/2}\Big)\,o_P(1)=o_P(1). \qquad \text{(A.23)}
\end{aligned}
$$

Combining (A.18) and (A.20)-(A.23), we obtain

$$
\begin{aligned}
&\Big(\sum_{t=1}^n x_tx_t^T K\big[(z_t-z)/h\big]\Big)^{1/2}\Big(\widehat{\beta}_n(z)-\beta_0(z)-c_2\beta_0''(z)h^2\Big)\\
=\;&\Big[E(x_1x_1^T)\frac{d_n}{nh}\sum_{t=1}^n K\Big(\frac{z_t-z}{h}\Big)+o_P(1)\Big]^{-1/2}\Big[\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n x_t\epsilon_tK\Big(\frac{z_t-z}{h}\Big)+o_P(1)\Big]\to_D\sigma\,\mathbb{N},
\end{aligned}
$$

due to (A.19) 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

$$
\big(\frac{nh}{d_n}\big)^{1/2}\Big(\widehat{\beta}_L(z)-\beta_0(z)-c_2\beta_0''(z)h^2\Big)=D_n^{-1}\big(P_n+T_{n1}+\beta_0''(z)T_{n2}\big), \qquad \text{(A.24)}
$$

by virtue of (A.12), where

$$
\begin{aligned}
D_n &= \frac{d_n}{nh}\sum_{t=1}^n w_tx_tx_t^TK\Big(\frac{z_t-z}{h}\Big),\\
P_n &= \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n w_tx_t\epsilon_tK\Big(\frac{z_t-z}{h}\Big),\\
T_{n1} &= \big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n w_tx_tx_t^Tv_t,\\
T_{n2} &= h^2\big(\frac{d_n}{nh}\big)^{1/2}\sum_{t=1}^n w_tx_tx_t^T\Big\{\frac12 K_2[(z_t-z)/h]-c_2K[(z_t-z)/h]\Big\}.
\end{aligned}
$$

Noting that from (A.18), 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 A.3 and A.4 we have

$$
\begin{aligned}
D_n &= V_{n2}D_{n0}-hV_n1D_{n1}=V_{n2}[D_{n0}+o_P(1)];\\
P_n &= V_{n2}\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_t\epsilon_tK\Big(\frac{z_t-z}{h}\Big)-hV_{n1}\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_t\epsilon_tK_1\Big(\frac{z_t-z}{h}\Big)\\
&= V_{n2}\Big[\Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{t=1}^n x_t\epsilon_tK\Big(\frac{z_t-z}{h}\Big)+o_P(1)\Big];
\end{aligned}
$$

$$
\begin{aligned}
\|T_{n1}\| &\le V_{n2}\|R_{n1}\| + h|V_{n1}|\sum_{t=1}^{n} x_t x_t^T |v_t(z_t - z)| \\
&= O_P\Big((\frac{nh}{d_n})^{1/2}\, h^{2+\eta}\Big) V_{n2}; \\
T_{n2} &= V_{n2}R_{n3} - V_{n1}\, h^3 \left(\frac{d_n}{nh}\right)^{1/2} \sum_{t=1}^{n} x_t x_t^T \Big\{ \frac{1}{2} K_3[(z_t - z)/h] - c_2 K_1[(z_t - z)/h] \Big\} \\
&= o_P(1)\, V_{n2}.
\end{aligned}
$$

Taking these facts into (A.24), the claim follows from

$$
\begin{aligned}
& \Big(\sum_{t=1}^{n} x_t x_t^T K\big[(z_t - z)/h\big] \Big)^{1/2} \Big(\widehat{\beta}_L(z) - \beta_0(z) - c_2 \beta_0''(z) h^2 \Big) \\
= \quad & \Big[E(x_1 x_1^T) \frac{d_n}{nh} \sum_{t=1}^{n} K\Big(\frac{z_t - z}{h}\Big) + o_P(1) \Big]^{-1/2} \Big[\Big(\frac{d_n}{nh}\Big)^{1/2} \sum_{t=1}^{n} x_t \epsilon_t K\Big(\frac{z_t - z}{h}\Big) + o_P(1) \Big] \\
\to_D \quad & \sigma\, \mathbb{N},
\end{aligned}
$$

due to (A.19) and the continuous mapping theorem. Theorem 2.3 is now proved. ■

A.3 Proof of Theorem 2.4

Let $V_t = x_t K[(z_t - z)/h]\, L_1[(w_t - w)/h_1]$. We may write

$$
D_n^{1/2}\left[\hat{\beta}_0(z,w) - \beta_0(z,w)\right] \quad = \quad D_n^{-1/2} S_n + R_n. \tag{A.25}
$$

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

$$
||R_n|| \quad = \quad o_P(1), \tag{A.26}
$$

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

$$
\begin{aligned}
& \Big\{ \frac{d_n}{nhh_1}\, A^T D_n A, \quad (\frac{d_n}{nhh_1})^{1/2}\, A^T S_n \Big\} \\
\to_D \quad & \big\{ (A^T D_w A)\, L_Z(1,0), \ \tau^{1/2}\, (A^T D_w A)\, N\, L_Z^{1/2}(1,0) \big\},
\end{aligned} \tag{A.27}
$$

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 $\Delta_t = \sum_{k,j=1}^d A_k A_j x_{tk} x_{tj} L_1[(w_t - w)/h_1]$. Since, by **A8**(d) and some standard arguments,

$$\begin{aligned} E x_{ti} x_{tj}\, L_1^\gamma[(w_t-w)/h_1] &= h_1 E(x_{ti}x_{tj}|w_t=w) \int_\Omega L_1^\gamma(x)dx + o(h_1), \\ E|x_{ti}|^\beta\, L_1^\gamma[(w_t-w)/h_1] &= h_1 E(|x_{ti}|^\beta|w_t=w) \int_\Omega L_1^\gamma(x)dx + o(h_1) = O(h_1), \end{aligned}$$

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

$$\begin{aligned} \frac{d_n}{nhh_1} A^T D_n A &= \frac{d_n}{nhh_1} \sum_{t=1}^n K[(z_t-z)/h]\, E\Delta_t + \frac{d_n}{nhh_1} \sum_{t=1}^n K[(z_t-z)/h]\, (\Delta_t - E\Delta_t) \\ &= \big[A^T D_w A + o(1)\big] \frac{d_n}{nh} \sum_{t=1}^n K[(z_t-z)/h] + O_P\big[(\frac{d_n}{nhh_1})^{1/2}\big] \\ &= \big[A^T D_w A + o(1)\big] \frac{d_n}{nh} \sum_{t=1}^n K[(z_t-z)/h] + o_P(1), \qquad \text{(A.28)} \end{aligned}$$

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

$$\begin{aligned} \frac{d_n}{nhh_1} \sum_{t=1}^n (A^T V_t)^2 &= \big[A^T D_w A + o(1)\big] \int_\Omega L_1^2(x)dx\, \frac{d_n}{nh} \sum_{t=1}^n K^2[(z_t-z)/h] \\ &\quad + o_P(1), \qquad \text{(A.29)} \end{aligned}$$

and

$$\begin{aligned} \sum_{t=1}^n \big(||V_t|| + ||V_t||^{2+\delta/2}\big) &= \sum_{t=1}^n (||x_t|| + ||x_t||^{2+\delta}) K[(z_t-z)/h]\, L_1[(w_t-w)/h_1] \\ &\le O(h_1) \sum_{t=1}^n K[(z_t-z)/h] + O_P[(nhh_1/d_n)^{1/2}] \\ &= O_P(nhh_1/d_n), \qquad \text{(A.30)} \end{aligned}$$

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 (A.28)-(A.29) imply that

$$\begin{aligned} &\Big\{ \frac{\sum_{j=1}^{[nt]} \nu_j}{\sqrt{n}},\ \frac{\sum_{j=1}^{[nt]} \nu_{-j}}{\sqrt{n}},\ \frac{d_n}{nhh_1} A^T D_n A,\ \frac{d_n}{nhh_1} \sum_{t=1}^n (A^T V_t)^2 \Big\} \\ &\Rightarrow\ \big\{ B_t,\ B_{-t},\ (A^T D_w A)\, L_Z(1,0),\ \tau_1\, (A^T D_w A)\, L_Z(1,0) \big\}, \qquad \text{(A.31)} \end{aligned}$$

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

$$\tau_1 = \int_\Omega L_1^2(x)dx \int_\Omega K^2(x)dx.$$

We are now ready to prove (A.26) and (A.27). By noting that D_w is positive-definite, it is readily seen from (A.28) that $D_n^{-1} = O_P(d_n/nhh_1)$. This, together with (A.30), 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 (A.26).

To prove (A.27), write $u_{nt} = \big(\frac{d_n}{nhh_1}\big)^{1/2} A^T V_t$, namely, we have $\big(\frac{d_n}{nhh_1}\big)^{1/2} A^T S_n = \sum_{t=1}^n \epsilon_t\, u_{nt}$. By using (A.28), routine calculations show that $\frac{1}{\sqrt{n}} \sum_{t=1}^n |u_{nt}| = o_P(1)$ and

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

Now, by recalling **A8** and (A.31), (A.27) 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. □

B Proofs of auxiliary results

Throughout this section, we denote an absolute positive constant by C, which may be different at each appearance. A sequence $\{\xi_k, k \ge 1\}$ is said to be α mixing if the α mixing coefficient

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

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

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 $1 \le p \le r \le \infty$,

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

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

$$|EXY - EXEY| \le 8\|X\|_p \|Y\|_q \{\alpha(k-i)\}^{1-p^{-1}-q^{-1}}. \tag{B.2}$$

Lemma B.1 *For any $h \to 0$ satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$, the condition* **A1** *with $\gamma \ge \max\{21/2, 6/\delta\}$, together with* **A2***(i), implies (2.1).*

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

$$\frac{x_{[nt]}}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \eta_k \Rightarrow B_t,$$

on $D_R[0,1]$, namely, $\{x_{[nt]}/\sqrt{n}\}_{n\ge 1}$ 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 those of $(B_t, \sigma_z B_{1t})$;

(ii) $\{R_n(t)\}_{n\ge1}$ 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_{1t} comes from the fact that, for any $0 < t \le 1$, the covariance of $x_{[nt]}/\sqrt{n}$ and $R_n(t)$ converges to zero in probability. Indeed, by using (B.2) and Lemma A.1, we have

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

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

$$\sup_{n\ge1} E R_n^2(t) \le C\, t, \tag{B.3}$$

indicating that $\{R_n(t)\}_{n\ge1}$, for any $0 < t \le 1$, is uniformly integrable. This fact will be used later.

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

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

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

$$E \max_{1\le k\le n} |\beta_{nk}| = o_P(1). \tag{B.4}$$

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 (B.3), (B.4) and the fact $R_n(t) = \sum_{k=1}^{[nt]} w_{nk} + \beta_{n,[nt]}$ again.

It remains to prove (B.4). Note that $EA_1 = 0$ and $E|A_1|^r \le C_z h$ for any $1 \le r \le 3$ by (A.1). Standard arguments by using (B.1), together with $\alpha(n) = O(n^{-\gamma})$ for some $\gamma > 0$, show that, for any $1 \le p < 3$ and $0 < \alpha \le 3 - p$,

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

and

$$(E|\beta_{ni}|^p)^{1/p} \le (nh)^{-1/2} \sum_{k=1}^{\infty} \left(E|E(A_{i+k}|\mathcal{F}_i)|^p\right)^{1/p} \le C\,(nh)^{-1/2}\, h^{1/(p+\alpha)} \sum_{k=1}^{\infty} k^{-\gamma\alpha/p(p+\alpha)}. \quad \text{(B.5)}$$

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

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

We have now established (B.4) by taking $\gamma > 6/\delta$, and α sufficiently small so that

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

The proof of Lemma B.1 is now complete. $\square$

Proof of Lemma A.2. We only prove (A.6). Due to the local boundedness of $H(x)$, (A.4) is obvious. The proof of (A.5) follows from similar arguments as in the proof of (A.6). We omit the details.

As in the proof of Lemma B.1 for (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 $s \le 0$. By writing

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

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

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

$$= \sum_{k=1}^{n} H_1(x_{nk})v_k + R(n), \quad \text{say}, \tag{B.6}$$

where we recall $x_{n,t} = x_{nt} = x_t/\sqrt{n}$. As in the proof of (B.4), we have

$$\max_{1\le i\le n} |A_i - v_i| \;\; \le \;\; 2 \max_{1\le i\le n} |u_i| = o_P(\sqrt{nh}). \tag{B.7}$$

This, together with (A.4) and Lemma B.1 (i.e., (2.1) holds), implies that

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

on $D_{R^2}[0,1]$. Now, by recalling that B_{1t} 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 (B.6), (A.6) will follow if we prove

$$|R(n)| \;\; = \;\; o_P(\sqrt{nh}). \tag{B.8}$$

To this end, write $\Omega_K = \{x_{ni} : \max_{1\le i\le n+1} |x_{ni}| \le K\}$. Note that (B.5) implies $E|u_j|^p \le Ch^{p/(p+\alpha)}$ for any $\alpha > 0$, $1 \le p \le 3$ and for any $j \ge 1$. It follows from (A.3) and $E||\eta_1||^3 < \infty$ that

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

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 A.2 is complete. ■

Proof of Lemma A.3. We only provide a outline. Results (A.14) and (A.17) 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 $l, m \ge 0$, we have

$$E\Big(\sum_{k=1}^{n} \big[\eta_{k-l}\eta_{k-m}^T - E(\eta_{k-l}\eta_{k-m}^T)\big] K[(z_k - z)/h]\Big)^2$$

$$\leq \; C\,(1+\max\{l^{1/2}, m^{1/2}\}+h\log n)\,\big[E(\eta_1\eta_1^T)+E(\eta_1\eta_2^T)\big]\,nh/d_n.$$

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 (A.15). To see (A.16), 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

$$\begin{aligned}
& \Big\{\frac{d_n}{nh}\sum_{t=1}^n K\Big(\frac{z_t - z}{h}\Big),\; \big(\frac{d_n}{nh}\big)^{1/2} \sum_{t=1}^n \bar{u}_{kM}\, K\Big(\frac{z_t - z}{h}\Big)\Big\} \\
& \qquad \to_D \; \Big\{L_Z(1,0), a_{M2}\, L_Z^{1/2}(1,0)\, N\Big\},
\end{aligned} \tag{B.9}$$

where $a_{M2}^2 = E(u_{kM}^2)\int_{-\infty}^{\infty} K^2(t)dt$. Since $a_{M2}^2 \to a_2^2$ as $M \to \infty$, (A.16) follows easily from (B.9) and the fact:

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

as $n \to \infty$ first and then $M \to \infty$. The proof of Lemma A.3 is now complete. ■