6-2016

Uniform consistency of nonstationary kernel-weighted sample covariances for nonparametric regression

Degui LI

Peter C. B. PHILLIPS
Singapore Management University, peterphillips@smu.edu.sg

Jiti GAO

Published in Econometric Theory, June 2016, Volume 32 (3), Pages 655-685.
http://doi.org/10.1017/S0266466615000109

Uniform Consistency of Nonstationary Kernel-Weighted Sample Covariances for Nonparametric Regression

Degui Li[*], Peter C. B. Phillips[†] and Jiti Gao[‡]

November 20, 2013

Abstract

We obtain uniform consistency results for kernel-weighted sample covariances in a nonstationary multiple regression framework that allows for both fixed design and random design coefficient variation. In the fixed design case these nonparametric sample covariances have different uniform convergence rates depending on direction, a result that differs fundamentally from the random design and stationary cases. The uniform convergence rates derived are faster than the corresponding rates in the stationary case and confirm the existence of uniform super-consistency. The modelling framework and convergence rates allow for endogeneity and thus broaden the practical econometric import of these results. As a specific application, we establish uniform consistency of nonparametric kernel estimators of the coefficient functions in nonlinear cointegration models with time varying coefficients and provide sharp convergence rates in that case. For the fixed design models, in particular, there are two uniform convergence rates that apply in two different directions, both rates exceeding the usual rate in the stationary case.

Key words and phrases: Cointegration; Functional coefficients; Kernel degeneracy; Nonparametric kernel smoothing; Random coordinate rotation; Super-consistency; Uniform convergence rates; Time varying coefficients.

JEL classification: C13, C14, C32.

[*]University of York
[†]Yale University, University of Auckland, Southampton University, and Singapore Management University
[‡]Monash University

1 Introduction

Uniform consistency results with convergence rates for nonparametric kernel estimators have been extensively studied in the existing literature. These results are important in many kernel-based applications such as semiparametric estimation with first-stage kernel smoothing, kernel-based specification testing, and cross-validation bandwidth selection. Existing studies mainly focus on obtaining uniform consistency results for independent and identically distributed (*i.i.d.*) data or time series that satisfy certain stationarity and mixing conditions. Early statistical studies include Mack and Silverman (1982), Roussas (1990), Liebscher (1996), Masry (1996) and Bosq (1998). Later developments and econometric applications can be found in Hansen (2008), Kristensen (2009) and Li *et al* (2012).

Recent years have witnessed a growing literature on nonparametric kernel smoothing in a nonstationary framework. This work is of practical importance because the stationarity condition is restrictive and unrealistic in many empirical applications as discussed in the literature. Among others, see Phillips and Park (1998), Karlsen and Tjøstheim (2001), Karlsen *et al* (2007), Cai *et al* (2009), Wang and Phillips (2009a, 2009b), Xiao (2009), Chen *et al* (2010), Chen, Gao and Li (2012), and Gao and Phillips (2013a, 2013b). Most recently, there has been interest in obtaining uniform consistency results for nonparametric kernel smoothing under nonstationarity (notably, Chan and Wang, 2012; Wang and Wang, 2013; Gao *et al.,* 2013; Duffy, 2013). This work confirms that uniform convergence rates of kernel-based estimates in nonstationary cases are slower than those in the stationary case. Just as in pointwise convergence, the slower convergence rate is explained by the random wandering character of nonstationary time series (such as those arising in unit root or null recurrent Markov frameworks) so that the amount of time spent by the series in the vicinity of any particular point is of smaller order than the stationary case, thereby reducing the effective sample size in estimation.

This paper develops uniform consistency results for potentially multivariate kernel-weighted sample covariances of the following form

$$Q_n(z) = \sum_{t=1}^{n} K\left(\frac{Z_t - z}{h}\right) X_t e_t, \tag{1.1}$$

where $K(\cdot)$ is a kernel function, $h \equiv h_n$ is a bandwidth which tends to zero as n tends to infinity, X_t is a nonstationary I(1) process with dimension $d \geq 1$, and e_t is stationary. Detailed properties of X_t and e_t are provided in Section 2. Quantities such as the weighted sample covariance (1.1) play a central role in kernel regression and are fundamental in determining the limit theory of such regressions. Interest typically focuses on two cases: (i) $Z_t = \frac{t}{n}$, corresponding to a fixed design structure; and (ii) *i.i.d.* Z_t, corresponding to a random design framework.

For case (ii) we show that the uniform convergence rate of (1.1) is $O_P(n\sqrt{h \log n})$, which exceeds the $O_P(\sqrt{nh \log n})$ rate that holds when both X_t and e_t are stationary. This result can be used to derive a uniform convergence rate for nonparametric kernel-based estimation of the functional coefficients in nonlinear cointegration models where super-consistency exists. Case (i) is much more complicated because kernel weighting produces degeneracy in the signal matrix and this degeneracy introduces a major challenge in developing the asymptotic estimation theory (c.f., Phillips *et al*, 2013). The reason for this "kernel degeneracy" in the limit of the weighted signal matrix is that kernel regression concentrates attention on some time coordinate (say z_0), thereby fixing attention on a particular coordinate of the limit process of the regressor, say $X_{\lfloor nz_0 \rfloor}$, where the floor function $\lfloor \cdot \rfloor$ denotes integer part. In the multivariate case with $d > 1$, this focus on a single time coordinate produces a limit signal matrix (corresponding to the limit of the outer product $\frac{1}{n} X_{\lfloor nz_0 \rfloor} X'_{\lfloor nz_0 \rfloor}$) that is of deficient rank one. Moreover, the zero eigenspace of this limit matrix depends on the (random vector) value of the limit process at that time coordinate. To address such kernel degeneracies Phillips *et al* (2013) transform coordinates to separate the zero and non-zero (random) eigenspaces and provide the convergence rates and limit distribution theory in each of these directions. The present paper extends that analysis to derive uniform consistency with sharp convergence rates in the two directions. Although the uniform convergence rates differ in the two directions, both rates exceed the $O_P(\sqrt{nh \log n})$ rate that applies in the stationary case.

We apply these results to derive the uniform consistency of nonparametric kernel estimates in nonlinear cointegration models with varying coefficients, and confirm the super-consistency rates. Our approach allows for endogeneity between the regressor X_t and the error e_t, which enhances the practical relevance of the results in cointegration analysis: case (i) with the fixed design framework $Z_t = \frac{t}{n}$ relates particularly to cointegration models with time-varying coefficients (Park and Hahn, 1999; Phillips *et al*, 2013); and case (ii) with random design Z_t relates to cointegration models with functional coefficients (Cai *et al*, 2009; Xiao, 2009; Gao and Phillips 2013b). In addition, the uniform consistency results with sharp convergence rates that are obtained here are of some independent interest with other potential applications, such as to semiparametric cointegration models with partially-varying coefficients.

The remainder of the paper is organised as follows. Uniform consistency results for the fixed design case are given in Section 2. Those for the random design case are given in Section 3. Applications of the main results to nonlinear cointegration models with varying coefficients are provided in Section 4. Section 5 concludes. Proofs of the main results are given in the Appendix.

2 Uniform consistency with a fixed design covariate

This section establishes uniform consistency results for $Q_n(z)$ defined in (1.1) with $Z_t = \frac{t}{n}$. The random design case is discussed in Section 3. We start with regularity conditions that characterize the multivariate nonstationary time series X_t and the scalar stationary process e_t. Let X_t be a unit root process with generating mechanism $X_t = X_{t-1} + v_t$, initial value $X_0 = O_P(1)$ and innovations determined by the linear process

$$v_t = \Phi(\mathcal{L})\varepsilon_t = \sum_{j=0}^{\infty} \Phi_j \varepsilon_{t-j}, \tag{2.1}$$

where $\Phi(\mathcal{L}) = \sum_{j=0}^{\infty} \Phi_j \mathcal{L}^j$, Φ_j is a sequence of $d \times d$ matrices, $\mathcal{L}$ is the lag operator and $\{\varepsilon_t\}$ is a sequence of *i.i.d.* innovation vectors with dimension d.

Assumption 1. *(i) Let $\{\varepsilon_t\}$ be i.i.d. d-dimensional random vectors with $\mathbb{E}[\varepsilon_t] = 0$, $\Lambda_\varepsilon \equiv \mathbb{E}\big[\varepsilon_t \varepsilon_t'\big]$ positive definite, and $\mathbb{E}\big[\|\varepsilon_t\|^{4+\delta_0}\big] < \infty$ for $\delta_0 > 0$. The linear process coefficient matrices in (2.1) satisfy that $\sum_{j=0}^{\infty} j\|\Phi_j\| < \infty$ and $\Omega_\varepsilon \equiv \Phi\Lambda_\varepsilon\Phi'$ is positive definite with $\Phi = \sum_{j=0}^{\infty} \Phi_j \neq 0$.*

(ii) Let $\{e_t\}$ be generated by the linear process $e_t = \sum_{j=0}^{\infty} \phi_j \eta_{t-j}$, where η_t is an i.i.d. sequence with $\mathbb{E}[\eta_t] = 0$, $\sigma_\eta^2 \equiv \mathbb{E}[\eta_t^2] > 0$, $\mathbb{E}\big[|\eta_t|^{4+\delta_0}\big] < \infty$, $\phi \equiv \sum_{j=0}^{\infty} \phi_j \neq 0$, and $\sum_{j=0}^{\infty} j|\phi_j| < \infty$. In addition, (η_t, ε_t') is independent of $\{(\eta_s, \varepsilon_s') : s \leq t-1\}$, but η_t may be correlated with ε_t.

Assumption 1(i) ensures that a functional law holds for X_t upon standardization. In particular, from Phillips and Solo (1992) we have for $t = \lfloor nx \rfloor$ and $0 < x \leq 1$,

$$\frac{X_t}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{s=1}^{t} v_s + \frac{1}{\sqrt{n}} X_0 = \frac{1}{\sqrt{n}} \sum_{s=1}^{\lfloor nx \rfloor} v_s + o_P(1) \Rightarrow B_x(\Omega_\varepsilon), \tag{2.2}$$

where $B_\cdot(\Omega_\varepsilon)$ is d-dimensional Brownian motion with variance matrix Ω_ε. In a more specialized setting, Assumption 1(ii) might be replaced by a martingale difference structure with $\mathbb{E}\big[e_t|\mathcal{G}_{t-1}\big] = 0$ *a.s.*, where $\mathcal{G}_t = \sigma(e_t, \cdots, e_1, \varepsilon_{t+1}, \varepsilon_t, \cdots)$, and the uniform consistency results developed in this paper still hold. Instead, we allow for a more general linear dependence structure and joint contemporaneous correlation between the innovations η_t and ε_t which builds endogeneity into the regression equation. Uniform consistency continues to hold when e_t and v_t are jointly determined by a multivariate linear process of the form

$$(e_t, v_t')' = \Phi^*(\mathcal{L})\varepsilon_t^* = \sum_{j=0}^{\infty} \Phi_j^* \varepsilon_{t-j}^*,$$

where $\Phi^*(\mathcal{L}) = \sum_{j=0}^{\infty} \Phi_j^* \mathcal{L}^j$ with Φ_j^* a sequence of $d+1$ dimensional coefficient matrices and $\{\varepsilon_t^*\}$ is a sequence of *i.i.d.* random vectors of dimension $d+1$.

We next impose some mild conditions on the kernel function $K(\cdot)$ and the bandwidth h.

Assumption 2. *(i) The kernel function $K(\cdot)$ is continuous, positive, symmetric and has compact support $[-1, 1]$ with $\mu_0 = 1$, where $\mu_j = \int_{-1}^{1} u^j K(u)du$.*

(ii) The bandwidth h satisfies $h \to 0$ and $nh \to \infty$ as n tends to infinity.

A recent paper by Phillips *et al* (2013) shows that for $0 < z \leq 1$,

$$\frac{1}{n^2 h} \sum_{t=1}^{n} X_t X_t' K\left(\frac{t - nz}{nh}\right) \Rightarrow W_z(\Omega_\varepsilon), \tag{2.3}$$

where $W_z(\Omega_\varepsilon) = B_z(\Omega_\varepsilon) B_z(\Omega_\varepsilon)'$ and "$\Rightarrow$" denotes weak convergence. However, the $d \times d$ limit matrix $W_z(\Omega_\varepsilon)$ on the right hand side of (2.3) is singular with rank one when $d > 1$, which indicates that the weighted signal matrix on the left hand side of (2.3) is asymptotically singular whenever the dimension of the regressor X_t exceeds unity. This phenomenon of kernel degeneracy leads to asymptotic singularity in the limit distribution and variance matrix of the kernel-weighted sample covariance $Q_n(z)$ defined in (1.1) when Z_t is a fixed design variable.

To address this kernel degeneracy Phillips *et al* (2013) develop a coordinate transformation to isolate the (random) direction of singularity and use the associated coordinate rotation to obtain the limit distribution theory. We define the quantities $\gamma_n(z) = \lfloor n(z-h) \rfloor$,

$$q_{\gamma_n(z)} = \frac{b_{\gamma_n(z)}}{\left[b_{\gamma_n(z)}' b_{\gamma_n(z)}\right]^{1/2}} = \frac{b_{\gamma_n(z)}}{\|b_{\gamma_n(z)}\|}, \text{ and } b_{\gamma_n(z)} = \frac{1}{\sqrt{n}} X_{\gamma_n(z)},$$

where "$\|\cdot\|$" denotes the Euclidean norm. Let $q_{\gamma_n(z)}^{\perp}$ be an orthogonal complement of $q_{\gamma_n(z)}$, define

$$D_n(z) = \left[q_{\gamma_n(z)}, q_{\gamma_n(z)}^{\perp}\right], \text{ with } D_n(z)' D_n(z) = I_d, \tag{2.4}$$

and introduce the vector

$$R_n = \mathrm{diag}\big\{n\sqrt{h},\ (nh)I_{d-1}\big\}, \tag{2.5}$$

where I_r is the $r \times r$ identity. The matrix $D_n(z)$ is random, path dependent, and localized to the coordinate of concentration at $\gamma_n(z)$.

The following result gives the uniform convergence rates for $Q_n(z)$ when $z \in (h, 1-h)$.

THEOREM 2.1. *Suppose that Assumptions 1 and 2 are satisfied. Let*

$$\frac{n^{2+\delta_0}h^{7+\delta_0}}{(\log n)^{3+\delta_0}} \to \infty, \tag{2.6}$$

where δ_0 is defined as in Assumption 1(i). Then, we have

$$\sup_{h<z<1-h} \big\| R_n^+ D_n(z)' Q_n(z) \big\| = O_P(\sqrt{\log n}), \tag{2.7}$$

where A^+ denotes the Moore-Penrose inverse of A.

From the proof of Theorem 2.1 in the Appendix, it is clear that the same uniform convergence rate as given in (2.7) holds if X_t and e_t are independent. Thus, the existence of correlation between the X_t and e_t does not affect the uniform convergence rate of the kernel-weighted sample covariance. This robustness to endogeneity in the present case arises because the induced asymptotic bias arising from the non-zero mean of $Q_n(z)$ turns out to be a "second order" bias effect as in the linear parametric case (Phillips and Durlauf, 1986; Phillips and Hansen, 1990). Furthermore, from the definitions of $D_n(z)$ and R_n, it is apparent that two different convergence rates obtain for the two directions determined by $q_{\gamma_n(z)}$ and $q^{\perp}_{\gamma_n(z)}$.

COROLLARY 2.1. *Let the assumptions in Theorem 2.1 hold. Then, we have*

$$\sup_{h<z<1-h} \big| q'_{\gamma_n(z)} Q_n(z) \big| = O_P(n\sqrt{h\log n}) \tag{2.8}$$

and

$$\sup_{h<z<1-h} \big\| (q^{\perp}_{\gamma_n(z)})' Q_n(z) \big\| = O_P(nh\sqrt{\log n}). \tag{2.9}$$

Although the uniform convergence rates are different in the two directions, both rates exceeds the usual uniform rate $O_P\left(\sqrt{nh\log n}\right)$ for kernel estimators that applies in stationary models. A detailed discussion of this phenomenon in the point-wise kernel regression case is given in Phillips *et al* (2013). The above results are used in Section 4 to derive uniform convergence rates for nonparametric kernel-based estimators of the time-varying coefficients in nonlinear cointegration models.

3 Uniform consistency with a random design covariate

This section develops uniform consistency for the sample covariance $Q_n(z)$ when Z_t is generated by *i.i.d.* random variables, and compares this result with those of the fixed design case studied in the previous section. For the stationary case, it is well known that the same uniform convergence rates hold for $Q_n(z)$ irrespective of whether Z_t is a random design or fixed design variate. In contrast to Section 2, there is no kernel degeneracy in the random design case and a common uniform convergence rate applies which is the same as that given in (2.8). The next assumption is used in the derivation of the uniform consistency result in Theorem 3.1 below.

Assumption 3. *Let $\{(Z_t, \eta_t, \varepsilon'_t)\}$ be a sequence of i.i.d. random vectors with continuous density function $f(\cdot,\cdot,\cdot)$, and let Z_t be independent of η_t and have compact support, say* $[0, 1]$.

Much of the existing literature on the limit theory of $Q_n(\cdot)$ for the random design case imposes a martingale difference structure on e_t, which excludes the possibility of correlation between X_t and e_t (c.f., Cai *et al*, 2009; Li *et al*, 2013). However, for consistency with the framework of Section 2, we follow the same structure as Assumption 1 to generate the unit root process X_t and the stationary process e_t, thereby allowing for correlation between X_t and e_t. Hence, the result below has wider applicability than those currently available in the literature.

The uniform convergence rate for $Q_n(z)$ in the random design case is given as follows.

Theorem 3.1. *Suppose that Assumptions 1–3 are satisfied. Let*

$$\frac{n^{2+\delta_0} h^{4+\delta_0}}{(\log n)^{4+\delta_0}} \to \infty, \tag{3.1}$$

where δ_0 is defined in Assumption 1(i). Then, we have

$$\sup_{0<z<1} \|Q_n(z)\| = O_P(n\sqrt{h \log n}). \tag{3.2}$$

This theorem shows that the uniform convergence rate (3.2) is exactly the same as (2.8) and therefore exceeds the stationary rate $O_P(\sqrt{nh \log n})$. This rate is also common across coordinates unlike the different rates that apply in the fixed design model. The result is used in Section 4 to derive a uniform convergence rate for nonparametric kernel-based estimation of the functional coefficients in nonlinear cointegration models under super-consistency.

4 Cointegration models with varying coefficients

In this section we use the results developed earlier to derive corresponding uniform consistency results for nonparametric kernel estimators in a nonlinear cointegration model with varying coefficients. The model has the form

$$Y_t = X_t'\beta(Z_t) + e_t, \quad t = 1, \cdots, n, \tag{4.1}$$

where X_t and e_t satisfy Assumption 1, $\beta(\cdot)$ is a d-dimensional coefficient function, and Z_t is either a fixed design or random design variate. In the fixed design case, model (4.1) is a cointegration model with time-varying coefficients, which was studied in Park and Hahn (1999) and Phillips *et al* (2013). The model can then be regarded as an extension of the locally stationary models used in Robinson (1989) and Cai (2007) where the regressors are stationary. In the random design case, model (4.1) is a cointegration model with functional coefficients of the type studied in Cai *et al* (2009), Xiao (2009) and Gao and Phillips (2013b). These studies provide nonstationary extensions of the models considered in Fan and Zhang (1999) and Cai *et al* (2000). The existing literature in these cases focuses on the development of pointwise asymptotic theory for nonparametric estimators of the coefficient function $\beta(\cdot)$ (c.f., Cai *et al*, 2009; Phillips *et al*, 2013). Uniform consistency results and associated convergence rates in the nonstationary case have so far not been considered due to the technical difficulties involved in the presence of nonstationary regressors. This section aims to fill this gap in the literature.

Under a smoothness condition on $\beta(\cdot)$ and for some fixed z, we have the local approximation

$$\beta\left(Z_t\right) = \beta(z) + O\left(Z_t - z\right) \approx \beta(z)$$

when Z_t is in a small neighborhood of z. The kernel-weighted local level regression estimator of the coefficient $\beta(z)$ at z has the following form

$$\widehat{\beta}_n(z) = \left[\sum_{t=1}^{n} X_t X_t' K\left(\frac{Z_t - z}{h}\right)\right]^{+} \left[\sum_{t=1}^{n} X_t Y_t K\left(\frac{Z_t - z}{h}\right)\right]. \tag{4.2}$$

We provide below a uniform consistency result for the estimator $\widehat{\beta}_n(z)$ over a range of values of z. Other kernel-based approaches such as local polynomial regression are also applicable to estimate the coefficient functions, and similar uniform consistency results as those given here can be obtained with some modification of the proofs.

To establish the limit theory for $\widehat{\beta}_n(\cdot)$, we impose the following commonly used smoothness condition on $\beta(\cdot)$ (c.f., Wang and Phillips, 2009a; Phillips *et al*, 2013).

Assumption 4. *The coefficient function* $\beta(\cdot)$ *is continuous with* $\beta(z+\delta)-\beta(z)| = O(|\delta|^{\alpha_0})$ *as* $\delta \to 0$ *for some* $\alpha_0 > 1/2$ *and any* $z \in (0,1)$.

We start with the fixed design case where $Z_t = \frac{t}{n}$ for $t = 1, \cdots, n$. Let $B_{z,*}(\Omega_\varepsilon)$ be an independent copy of the d-dimensional Brownian motion $B_z(\Omega_\varepsilon)$ which is defined in (2.2), $b_z \equiv b_{\gamma_n(z)}$ and $q_z \equiv q_{\gamma_n(z)}$ and $q_z^\perp = q_{\gamma_n(z)}^\perp$ for $0 < z < 1$. Define

$$\Delta_z = \begin{bmatrix} \Delta_z(1) & \Delta_z(2) \\ \Delta_z(2)' & \Delta_z(3) \end{bmatrix}, \tag{4.3}$$

with $\Delta_z(1) = b_z' b_z$,

$$\Delta_z(2) = 2\sqrt{2}\,(b_z' b_z)^{1/2} \Big\{ \int_{-1}^{1} B_{\frac{z+1}{2},*}(\Omega_\varepsilon) K(z) dz \Big\} q_z^\perp,$$

and

$$\Delta_z(3) = 4(q_z^\perp)' \Big\{ \int_{-1}^{1} B_{\frac{z+1}{2},*}(\Omega_\varepsilon) B_{\frac{z+1}{2},*}(\Omega_\varepsilon)' K(z) dz \Big\} q_z^\perp.$$

For fixed $0 < z < 1$, Proposition A.1 in Phillips *et al* (2013) shows that the standardized denominator matrix of (4.2) converges weakly to the limit

$$R_n^+ D_n(z)' \Big[\sum_{t=1}^{n} X_t X_t' K\Big(\frac{t-nz}{nh}\Big) \Big] D_n(z) R_n^+ \Rightarrow \Delta_z,$$

on which we make the following assumption.

Assumption 5. Δ_z *is non-singular with probability* 1 *uniformly for* $h < z < 1-h$.

Based on Theorem 2.1 and Corollary 2.1, we obtain the following uniform consistency results for the kernel estimator $\widehat{\beta}_n(z)$.

Theorem 4.1. *Suppose that the assumptions in Theorem 2.1 and Assumptions 4 and 5 are satisfied. Then, we have as* $n \to \infty$

$$\sup_{h<z<1-h} \big| q_z' \big[\widehat{\beta}_n(z) - \beta(z) \big] \big| = O_P\left(h^{\alpha_0} + \sqrt{\frac{\log n}{n^2 h}} \right) \tag{4.4}$$

and

$$\sup_{h<z<1-h} \big\| (q_z^\perp)' \big[\widehat{\beta}_n(z) - \beta(z) \big] \big\| = O_P\left(h^{\alpha_0} + \frac{\sqrt{\log n}}{nh} \right). \tag{4.5}$$

The order $O_P(h^{\alpha_0})$ of the asymptotic bias of the nonparametric estimator $\widehat{\beta}_n(z)$ in Theorem 4.1 can be improved to $O_P(h^2)$ if the local linear method (c.f., Fan and Gijbels, 1996) is used to estimate $\beta(\cdot)$. Theorem 4.1 gives different uniform convergence rates for $\widehat{\beta}_n(\cdot)$ in the two directions determined by the kernel degeneracy, just as in Corollary 2.1. In the direction q_z, we have the uniform convergence rate $O_P\big(\sqrt{\frac{\log n}{n^2 h}}\big)$, which we call the type I uniform convergence rate. This rate is faster than the rate $O_P\left(\frac{\sqrt{\log n}}{nh}\right)$ that applies in the other direction (c.f. (4.5)) as well as the usual rate $O_P\big(\sqrt{\frac{\log n}{nh}}\big)$ that applies in the stationary case. In the direction $q_z^{\perp}$, the uniform convergence rate $O_P\big(\frac{\sqrt{\log n}}{nh}\big)$ is slower than the type I uniform convergence rate of (4.4), but is still faster than the stationary rate. The rate $O_P\big(\frac{\sqrt{\log n}}{nh}\big)$ is called the type II uniform convergence rate.

Next consider the random design case where the covariate Z_t is *i.i.d.,* as discussed in Section 3. Define

$$\Lambda_z = f_Z(z)\int B_z(\Omega_\varepsilon)B_z(\Omega_\varepsilon)'dz,$$

where $f_Z(\cdot)$ is the density function of Z_t. It is easy to show that

$$\frac{1}{n^2 h}\sum_{t=1}^{n} X_t X_t' K\left(\frac{Z_t - z}{h}\right) \Rightarrow \Lambda_z$$

for $0 < z < 1$. Using Theorem 3.1 we derive the uniform convergence rate for $\widehat{\beta}_n(\cdot)$ in the following theorem, which shows that a common type I uniform convergence rate is attained in all directions in the random design case.

Theorem 4.2. *Suppose that the assumptions in Theorem 3.1 and Assumption 4 are satisfied. Let Λ_z be non-singular with probability* 1 *uniformly for $z \in (0,1)$. Then, we have as $n \to \infty$*

$$\sup_{0<z<1} \|\widehat{\beta}_n(z) - \beta(z)\| = O_P\left(h^{\alpha_0} + \sqrt{\frac{\log n}{n^2 h}}\right). \tag{4.6}$$

This uniform consistency result gives a new sharp rate of convergence for the nonlinear cointegration models with functional coefficients and complements the pointwise limit theory developed by Cai *et al* (2009), Xiao (2009) and Gao and Phillips (2013b).

5 Conclusions

This paper derives uniform consistency results for nonparametric kernel-weighted sample covariances and regressions in a nonstationary data framework. This framework has practical application in varying coefficient regressions with coefficient covariates that follow fixed and random designs. In the fixed design case, two different uniform convergence rates apply depending on a certain covariate-sensitive random direction, a result that is quite different from the random design case where a common uniform convergence rate applies. Both results are shown to be robust to endogeneity of the regressors.

A regression application of these results confirms the uniform consistency of nonparametric kernel estimates of the coefficient functions in nonlinear cointegration models with varying coefficients and gives sharp convergence rates in this regression case. In the fixed design framework, two types of uniform convergence rates again apply in the covariate sensitive random directions and both rates exceed the rate in the stationary case. In the random design framework, there is a common uniform convergence rate, which also exceeds that of the stationary case. These uniform consistency results are relevant in estimating semiparametric cointegration models with partially-varying coefficients, long run variance estimation in such models, kernel-based specification testing of nonlinear cointegration models, and the theory for the optimal bandwidth selection in the nonparametric kernel-smoothing under nonstationarity.

6 Acknowledgements

Phillips acknowledges support from the NSF under Grant No. SES 12-58258. Gao acknowledges support from the Australian Research Council Discovery Grants Program under Grant Nos. DP1096374 and DP130104229.

A Proofs of the main results

This appendix provides proofs of the main results in Sections 2–4. To simplify notation, in the sequel we let $q_z = q_{\gamma_n(z)}$ and $q_z^\perp = q_{\gamma_n(z)}^\perp$, and C is used for a positive constant whose value may change from line to line.

Proof of Theorem 2.1. For $0 < z < 1$, define

$$\begin{aligned} Q_n(z,1) &= \frac{q_z'}{n\sqrt{h}} \sum_{t=1}^n K\big(\frac{t-nz}{nh}\big) X_t e_t, \\ Q_n(z,2) &= \frac{(q_z^\perp)'}{nh} \sum_{t=1}^n K\big(\frac{t-nz}{nh}\big) X_t e_t. \end{aligned}$$

Note that

$$Q_n(z,1) = \frac{q_z'}{n\sqrt{h}} X_{\gamma_n(z)} \sum_{t=1}^n K\big(\frac{t-nz}{nh}\big) e_t + \frac{q_z'}{n\sqrt{h}} \sum_{t=1}^n K\big(\frac{t-nz}{nh}\big)(X_t - X_{\gamma_n(z)}) e_t, \tag{A.1}$$

where $\gamma_n(z)$ is defined in Section 2, and

$$Q_n(z,2) = \frac{(q_z^{\perp})'}{nh} \sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)(X_t - X_{\gamma_n(z)})e_t, \tag{A.2}$$

as $q_z^{\perp}$ is orthogonal to $X_{\gamma_n(z)}$ by (2.4) in Section 2. By continuous mapping (e.g. Billingsley, 1968), it is easy to show that

$$\sup_{0<z<1} \big(\|q_z\| + \|q_z^{\perp}\|\big) = O_P(1). \tag{A.3}$$

Then, by (A.1)–(A.3), it is sufficient to show that

$$\sup_{h<z<1-h} \Big|\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)e_t\Big| = O_P(\sqrt{\log n}), \tag{A.4}$$

and

$$\sup_{h<z<1-h} \Big\|\frac{1}{nh} \sum_{t=1}^{n} K\big(\frac{t-nz}{nh}\big)(X_t - X_{\gamma_n(z)})e_t\Big\| = O_P(\sqrt{\log n}), \tag{A.5}$$

which we now prove in turn.

Proof of (A.4). Using the BN decomposition approach of Phillips and Solo (1992), we have

$$e_t = \overline{e}_t + (\widetilde{e}_{t-1} - \widetilde{e}_t), \tag{A.6}$$

where $\overline{e}_t = \big(\sum_{j=0}^{\infty} \phi_j\big)\eta_t = \phi\eta_t$ and $\widetilde{e}_t = \sum_{j=0}^{\infty} \widetilde{\phi}_j \eta_{t-j}$ with $\widetilde{\phi}_j = \sum_{k=j+1}^{\infty} \phi_k$. By (A.6), we can show that

$$\begin{aligned}
\sum_{t=1}^{n} e_t K\big(\frac{t-nz}{nh}\big) &= \sum_{t=1}^{n} \overline{e}_t K\big(\frac{t-nz}{nh}\big) + \sum_{t=1}^{n} \widetilde{e}_{t-1} K\big(\frac{t-nz}{nh}\big) - \sum_{t=1}^{n} \widetilde{e}_t K\big(\frac{t-nz}{nh}\big) \\
&= \sum_{t=1}^{n} \overline{e}_t K\big(\frac{t-nz}{nh}\big) + \sum_{t=1}^{n} \widetilde{e}_{t-1} K\big(\frac{t-1-nz}{nh}\big) - \sum_{t=1}^{n} \widetilde{e}_t K\big(\frac{t-nz}{nh}\big) + \\
&\quad \sum_{t=1}^{n} \widetilde{e}_{t-1} \Big[K\big(\frac{t-nz}{nh}\big) - K\big(\frac{t-1-nz}{nh}\big)\Big] \\
&= \sum_{t=1}^{n} \overline{e}_t K\big(\frac{t-nz}{nh}\big) + \sum_{t=1}^{n} \widetilde{e}_{t-1} \Big[K\big(\frac{t-nz}{nh}\big) - K\big(\frac{t-1-nz}{nh}\big)\Big] + \\
&\quad \widetilde{e}_0 K\big(\frac{-z}{h}\big) - \widetilde{e}_n K\big(\frac{1-z}{h}\big).
\end{aligned}$$

By virtue of Assumption 2(i) and (ii),

$$\widetilde{e}_0 K\big(\frac{-z}{h}\big) = \widetilde{e}_n K\big(\frac{1-z}{h}\big) = 0 \tag{A.7}$$

with probability 1 for any $h < z < 1-h$, which indicates that

$$\sum_{t=1}^{n} e_t K\big(\frac{t-nz}{nh}\big) = \sum_{t=1}^{n} \overline{e}_t K\big(\frac{t-nz}{nh}\big) + \sum_{t=1}^{n} \widetilde{e}_{t-1} \big[K\big(\frac{t-nz}{nh}\big) - K\big(\frac{t-1-nz}{nh}\big)\big] \tag{A.8}$$

uniformly for $0 < z < 1$.

Define $\mathbb{Z}_k = \{z \big| (k-1)nr_n + 1 \le z < knr_n\}$ for $k = 1, 2, \cdots, R_n$, and $\mathbb{Z}_{R_n+1} = \{z \big| nr_nR_n + 1 \le z \le n\}$, where $R_n = \lfloor r_n^{-1} \rfloor$, $r_n = h^{3/2}\log^{1/2}(n)$. Let z_k be the smallest number in the set $\mathbb{Z}_k$ for $k = 1, \cdots, R_n, R_n + 1$. By standard arguments, we have

$$\sup_{h<z<1-h} \big| \sum_{t=1}^{n} \overline{e}_t K\big(\frac{t-nz}{nh}\big)\Big| \le \max_{1\le k\le R_n^*} \sup_{z\in\mathbb{Z}_k} \big| \sum_{t=1}^{n} \overline{e}_t \big[K\big(\frac{t-z}{nh}\big) - K\big(\frac{t-z_k}{nh}\big)\big]\Big| + \max_{1\le k\le R_n^*} \Big| \sum_{t=1}^{n} \overline{e}_t K\big(\frac{t-z_k}{nh}\big)\Big|,$$

where $R_n^* = R_n + 1$. By the Markov inequality, we may show that

$$\max_{1\le k\le R_n^*} \sup_{z\in\mathbb{Z}_k} \Big| \sum_{t=1}^{n} \overline{e}_t \big[K\big(\frac{t-z}{nh}\big) - K\big(\frac{t-z_k}{nh}\big)\big]\Big| = O_P\big(\frac{\sqrt{n}r_n}{h}\big) = O_P\big(\sqrt{nh\log n}\big). \tag{A.9}$$

Noting that $\overline{e}_t = \phi\eta_t$, we next prove

$$\max_{1\le k\le R_n^*} \Big| \sum_{t=1}^{n} \eta_t K\big(\frac{t-z_k}{nh}\big)\Big| = O_P\big(\sqrt{nh\log n}\big) \tag{A.10}$$

by the truncation technique and using the Bernstein inequality (e.g., van der Vaart and Wellner, 1996). Let $\overline{\eta}_t = \eta_t \cdot \mathrm{I}\big(|\eta_t| \le \sqrt{\frac{nh}{\log n}}\big)$ and $\widetilde{\eta}_t = \eta_t - \overline{\eta}_t = \eta_t \cdot \mathrm{I}\big(|\eta_t| > \sqrt{\frac{nh}{\log n}}\big)$, where $\mathrm{I}(\cdot)$ is an indicator function. Noting that

$$\mathbb{P}\Big\{\max_{1\le t\le n} |\eta_t| > \sqrt{\frac{nh}{\log n}}\Big\} \le C \cdot \frac{n(\log n)^{(4+\delta_0)/2}}{(nh)^{(4+\delta_0)/2}} = o(1)$$

as $\frac{n^{2+\delta_0}h^{4+\delta_0}}{(\log n)^{4+\delta_0}} \to \infty$, we can show that

$$\max_{1\le k\le R_n^*} \Big| \sum_{t=1}^{n} \big(\widetilde{\eta}_t - \mathbb{E}[\widetilde{\eta}_t]\big) K\big(\frac{t-z_k}{nh}\big)\Big| = o_P\big(\sqrt{nh\log n}\big). \tag{A.11}$$

On the other hand, note that $\{\eta_t\}$ is a sequence of *i.i.d.* random variables, and the number of non-zero summands in $\sum_{t=1}^{n} \overline{\eta}_t K\big(\frac{t-z_k}{nh}\big)$ is of order (nh) as the compact support of the kernel function is $[-1, 1]$. Letting c_0 be some positive constant and by using the Bernstein inequality, for sufficiently large $M > c_0 > 0$, we have

$$\begin{aligned}
&\mathbb{P}\Big\{\max_{1\le k\le R_n^*} \big| \sum_{t=1}^{n} \big(\overline{\eta}_t - \mathbb{E}[\overline{\eta}_t]\big) K\big(\frac{t-z_k}{nh}\big)\big| > M\sqrt{nh\log n}\Big\} \\
\le\ & \sum_{k=1}^{R_n^*} \mathbb{P}\Big\{\big| \sum_{t=1}^{n} \big(\overline{\eta}_t - \mathbb{E}[\overline{\eta}_t]\big) K\big(\frac{t-z_k}{nh}\big)\big| > M\sqrt{nh\log n}\Big\} \\
\le\ & \sum_{k=1}^{R_n^*} \exp\big\{-\frac{Mnh\log n}{c_0nh}\big\} \le O(r_n^{-1}n^{-M/c_0}) = o(1),
\end{aligned}$$

which indicates that

$$\max_{1\le k\le R_n^*}\Big|\sum_{t=1}^n\big(\overline{\eta}_t-\mathbb{E}[\overline{\eta}_t]\big)K\big(\frac{t-z_k}{nh}\big)\Big|=O_P\big(\sqrt{nh\log n}\big). \tag{A.12}$$

Then, by (A.11) and (A.12), we can prove (A.10), which together with (A.9), leads to

$$\sup_{h<z<1-h}\Big|\sum_{t=1}^n\overline{e}_tK\big(\frac{t-nz}{nh}\big)\Big|=O_P(\sqrt{nh\log n}). \tag{A.13}$$

Noting that $K\big(\frac{t-nz}{nh}\big)-K\big(\frac{t-1-nz}{nh}\big)\le C\frac{1}{nh}$, by a standard derivation, we can also show that

$$\sup_{h<z<1-h}\Big|\sum_{t=1}^n\widetilde{e}_{t-1}\big[K\big(\frac{t-nz}{nh}\big)-K\big(\frac{t-1-nz}{nh}\big)\big]\Big|=O_P(\sqrt{nh\log n}), \tag{A.14}$$

which together with (A.7), (A.8) and (A.13), leads to (A.4).

Proof of (A.5). Using the BN decomposition again, we have

$$X_t-X_{\gamma_n(z)}=\sum_{s=\gamma_n(z)+1}^t v_s=\sum_{s=\gamma_n(z)+1}^t\overline{v}_s+\widetilde{v}_{\gamma_n(z)}-\widetilde{v}_t,$$

where $\overline{v}_t=(\sum_{j=0}^\infty\Phi_j)\varepsilon_t=\Phi\varepsilon_t$ and $\widetilde{v}_t=\sum_{j=0}^\infty\widetilde{\Phi}_j\varepsilon_{t-j}$ with $\widetilde{\Phi}_j=\sum_{k=j+1}^\infty\Phi_k$. Hence, to prove (A.5), we need only prove that

$$\sum_{t=1}^n\Big(\sum_{s=\gamma_n(z)+1}^t\overline{v}_s\Big)e_tK\big(\frac{t-nz}{nh}\big)=O_P(nh\sqrt{\log n}), \tag{A.15}$$

$$\widetilde{v}_{\gamma_n(z)}\sum_{t=1}^n e_tK\big(\frac{t-nz}{nh}\big)=O_P(nh\sqrt{\log n}), \tag{A.16}$$

$$\sum_{t=1}^n\widetilde{v}_te_tK\big(\frac{t-nz}{nh}\big)=o_P(nh\sqrt{\log n}), \tag{A.17}$$

uniformly for $h<z<1-h$.

Note that $\widetilde{v}_t$ and e_t are well defined stationary linear processes, and the numbers of non-zero summands in both $\sum_{t=1}^n\widetilde{v}_te_tK\big(\frac{t-nz}{nh}\big)$ and $\sum_{t=1}^n e_tK\big(\frac{t-nz}{nh}\big)$ are of order (nh). We can thus prove (A.16) and (A.17) easily by standard aguments. This leaves (A.15).

To prove (A.15) we proceed as follows. Let $\overline{v}_t(z)=\sum_{s=\gamma_n(z)+1}^t\overline{v}_s$ and $\overline{v}_t(z)=0$ if $t<\gamma_n(z)+1$.

Using the BN decomposition (A.6), we have

$$\begin{aligned}
\sum_{t=1}^{n}\overline{v}_t(z)e_tK\big(\frac{t-nz}{nh}\big) &= \sum_{t=1}^{n}\overline{v}_t(z)\overline{e}_tK\big(\frac{t-nz}{nh}\big)+\sum_{t=1}^{n}\overline{v}_t(z)\widetilde{e}_{t-1}K\big(\frac{t-nz}{nh}\big)- \\
&\quad \sum_{t=1}^{n}\overline{v}_t(z)\widetilde{e}_tK\big(\frac{t-nz}{nh}\big) \\
&= \sum_{t=1}^{n}\overline{v}_t\overline{e}_tK\big(\frac{t-nz}{nh}\big)+\sum_{t=1}^{n}\overline{v}_{t-1}(z)\overline{e}_tK\big(\frac{t-nz}{nh}\big)+ \\
&\quad \sum_{t=1}^{n}\overline{v}_t\widetilde{e}_{t-1}K\big(\frac{t-nz}{nh}\big)+\sum_{t=1}^{n}\overline{v}_{t-1}(z)\widetilde{e}_{t-1}K\big(\frac{t-nz}{nh}\big)- \\
&\quad \sum_{t=1}^{n}\overline{v}_t(z)\widetilde{e}_tK\big(\frac{t-nz}{nh}\big) \\
&= \sum_{t=1}^{n}\overline{v}_t\overline{e}_tK\big(\frac{t-nz}{nh}\big)+\sum_{t=1}^{n}\overline{v}_{t-1}(z)\overline{e}_tK\big(\frac{t-nz}{nh}\big)+ \\
&\quad \sum_{t=1}^{n}\overline{v}_{t-1}(z)\widetilde{e}_{t-1}\big[K\big(\frac{t-nz}{nh}\big)-K\big(\frac{t-1-nz}{nh}\big)\big]- \\
&\quad \overline{v}_n(z)\widetilde{e}_nK\big(\frac{1-z}{h}\big)+\sum_{t=1}^{n}\overline{v}_t\widetilde{e}_{t-1}K\big(\frac{t-nz}{nh}\big).
\end{aligned}$$

Similar to the proof of (A.14), we may show that

$$\sup_{h<z<1-h}\Big\|\sum_{t=1}^{n}\overline{v}_{t-1}(z)\widetilde{e}_{t-1}\big[K\big(\frac{t-nz}{nh}\big)-K\big(\frac{t-1-nz}{nh}\big)\big]\Big\| = o_P(nh\sqrt{\log n}). \tag{A.18}$$

Following the proof of (A.4), we can also show that

$$\begin{aligned}
\sup_{h<z<1-h}\Big\|\sum_{t=1}^{n}\overline{v}_t\overline{e}_tK\big(\frac{t-nz}{nh}\big)\Big\| &\leq \sup_{h<z<1-h}\Big\|\sum_{t=1}^{n}\big\{\overline{v}_t\overline{e}_t-\mathbb{E}[\overline{v}_t\overline{e}_t]\big\}K\big(\frac{t-nz}{nh}\big)\Big\|+ \\
&\quad \sup_{h<z<1-h}\Big\|\sum_{t=1}^{n}\mathbb{E}[\overline{v}_t\overline{e}_t]K\big(\frac{t-nz}{nh}\big)\Big\| \\
&= O_P(\sqrt{nh\log n})+O(nh) \\
&= o_P(nh\sqrt{\log n}). \qquad\qquad (A.19)
\end{aligned}$$

Noting that $\overline{v}_t$ and $\widetilde{e}_t$ are stationary, and the compact support of the kernel function is $[-1,1]$, we can prove that

$$\sup_{h<z<1-h}\Big\|\overline{v}_n(z)\widetilde{e}_nK\big(\frac{1-z}{h}\big)\Big\| = o_P(nh\sqrt{\log n}), \tag{A.20}$$

$$\sup_{h<z<1-h}\Big\|\sum_{t=1}^{n}\overline{v}_t\widetilde{e}_{t-1}K\big(\frac{t-nz}{nh}\big)\Big\| = o_P(nh\sqrt{\log n}). \tag{A.21}$$

By (A.18)–(A.21), to complete the proof of (A.15), we need only prove that

$$\sup_{h<z<1-h}\Big\|\sum_{t=1}^{n}\overline{v}_{t-1}(z)\overline{e}_tK\big(\frac{t-nz}{nh}\big)\Big\| = O_P(nh\sqrt{\log n}). \tag{A.22}$$

Let $\mathbb{Z}_k$, z_k, R_n, R_n^* and r_n be defined as above. By standard arguments, we have

$$\begin{aligned}\sup_{h<z<1-h}\Big\|\sum_{t=1}^{n}\overline{v}_{t-1}(z)\overline{e}_tK\big(\frac{t-nz}{nh}\big)\Big\| \quad \leq \quad & \max_{1\leq k\leq R_n^*}\sup_{z\in\mathbb{Z}_k}\Big\|\sum_{t=1}^{n}\overline{v}_{t-1}(z)\overline{e}_t\big[K\big(\frac{t-z}{nh}\big)-K\big(\frac{t-z_k}{nh}\big)\big]\Big\| \\ & + \max_{1\leq k\leq R_n^*}\sup_{z\in\mathbb{Z}_k}\Big\|\sum_{t=1}^{n}\big[\overline{v}_{t-1}(z)-v_{t-1}(z_k)\big]\overline{e}_tK\big(\frac{t-z_k}{nh}\big)\Big\| \\ & + \max_{1\leq k\leq R_n^*}\Big\|\sum_{t=1}^{n}\overline{v}_{t-1}(z_k)\overline{e}_tK\big(\frac{t-z_k}{nh}\big)\Big\|,\end{aligned}$$

where $\overline{v}_t(z)\equiv\overline{v}_t(z/n)$ on the right hand side of the inequality and in the sequel. To prove (A.22), we need to show that

$$\max_{1\leq k\leq R_n^*}\sup_{z\in\mathbb{Z}_k}\Big\|\sum_{t=1}^{n}\overline{v}_{t-1}(z)\overline{e}_t\big[K\big(\frac{t-z}{nh}\big)-K\big(\frac{t-z_k}{nh}\big)\big]\Big\| = O_P(nh\sqrt{\log n}), \tag{A.23}$$

$$\max_{1\leq k\leq R_n^*}\sup_{z\in\mathbb{Z}_k}\Big\|\sum_{t=1}^{n}\big[\overline{v}_{t-1}(z)-x_{t-1}(z_k)\big]\overline{e}_tK\big(\frac{t-z_k}{nh}\big)\Big\| = O_P(nh\sqrt{\log n}), \tag{A.24}$$

and

$$\max_{1\leq k\leq R_n^*}\Big\|\sum_{t=1}^{n}\overline{v}_{t-1}(z_k)\eta_tK\big(\frac{t-z_k}{nh}\big)\Big\| = O_P(nh\sqrt{\log n}). \tag{A.25}$$

We provide the proof of (A.25) and the proofs for (A.23) and (A.24) are entirely analogous. Let $w_t(z_k)=\overline{v}_{t-1}(z_k)\eta_t$, $\mathcal{F}_t=\{(\eta_s,\varepsilon_s'):s\leq t\}$, and

$$\overline{w}_t(z_k)=w_t(z_k)\cdot \mathrm{I}\big(\|\overline{v}_{t-1}(z_k)\|\leq\frac{(nh)^{3/4}}{(\log n)^{1/4}},|\eta_t|\leq\frac{(nh)^{1/4}}{(\log n)^{1/4}}\big),\quad \widetilde{w}_t(z_k)=w_t(z_k)-\overline{w}_t(z_k).$$

Noting that

$$\mathbb{P}\Big\{\max_{1\leq k\leq R_n^*}\max_{z_k-nh\leq t\leq z_k+nh}\|\widetilde{w}_t(z_k)\|>0\Big\}\leq C\cdot\frac{nR_n^*(\log n)^{(4+\delta_0)/2}}{(nh)^{(4+\delta_0)/2}}=o(1),$$

as $\frac{n^{2+\delta_0}h^{7+\delta_0}}{(\log n)^{3+\delta_0}}\to\infty$, we can show that

$$\max_{1\leq k\leq R_n^*}\Big\|\sum_{t=1}^{n}\big(\widetilde{w}_t(z_k)-\mathbb{E}[\widetilde{w}_t(z_k)|\mathcal{F}_{t-1}]\big)K\big(\frac{t-z_k}{nh}\big)\Big\| = o_P\big(nh\sqrt{\log n}\big). \tag{A.26}$$

On the other hand, note that $\{(w_t(z_k), \mathcal{F}_t):\ t\geq 1\}$ is a sequence of martingale differences. Then, by the exponential inequality for martingale differences (c.f., de la Pena, 1999) and letting c_1 be some positive constant, we have for sufficiently large $M > c_1 > 0$,

$$
\begin{aligned}
& \mathbb{P}\Big\{\max_{1\leq k\leq R_n^*}\|\sum_{t=1}^{n}\big(\overline{w}_t(z_k)-\mathbb{E}[\overline{w}_t(z_k)|\mathcal{F}_{t-1}]\big)K\big(\frac{t-z_k}{nh}\big)\|>Mnh\sqrt{\log n}\Big\}\\
\leq\ & \sum_{k=1}^{R_n^*}\mathbb{P}\Big\{\|\sum_{t=1}^{n}\big(\overline{w}_t(z_k)-\mathbb{E}[\overline{w}_t(z_k)|\mathcal{F}_{t-1}]\big)K\big(\frac{t-z_k}{nh}\big)\|>Mnh\sqrt{\log n}\Big\}\\
\leq\ & \sum_{k=1}^{R_n^*}\exp\big\{-\frac{M(nh)^2\log n}{c_1(nh)^2}\big\}\\
\leq\ & O(r_n^{-1}n^{-M/c_1})=o(1),
\end{aligned}
$$

which indicates that

$$
\max_{1\leq k\leq R_n^*}\Big\|\sum_{t=1}^{n}\big(\overline{w}_t(z_k)-\mathbb{E}[\overline{w}_t(z_k)|\mathcal{F}_{t-1}]\big)K\big(\frac{t-z_k}{nh}\big)\Big\|=O_P\big(nh\sqrt{\log n}\big). \tag{A.27}
$$

Then, by (A.26) and (A.27), we can prove (A.25) and this complete the proof of (A.15) and (A.5). Theorem 2.1 then follows. □

Proof of Theorem 3.1. Note that

$$
\begin{aligned}
Q_n(z) &= \sum_{t=1}^{n}K\big(\frac{Z_t-z}{h}\big)X_{t-1}e_t+\sum_{t=1}^{n}K\big(\frac{Z_t-z}{h}\big)v_te_t\\
&\equiv Q_{n1}(z)+Q_{n2}(z).
\end{aligned} \tag{A.28}
$$

First consider $Q_{n1}(z)$, which is the leading term of $Q_n(z)$. Decompose $Q_{n1}(z)$ as

$$
\begin{aligned}
Q_{n1}(z) &= \sum_{t=1}^{n}\mathbb{E}\big[K\big(\frac{Z_t-z}{h}\big)\big]X_{t-1}e_t+\sum_{t=1}^{n}\big\{K\big(\frac{Z_t-z}{h}\big)-\mathbb{E}\big[K\big(\frac{Z_t-z}{h}\big)\big]\big\}X_{t-1}e_t\\
&\equiv Q_{n3}(z)+Q_{n4}(z).
\end{aligned} \tag{A.29}
$$

Noting that

$$
\mathbb{E}\big[K\big(\frac{Z_t-z}{h}\big)\big]\sim hf_Z(z)\mu_0,
$$

uniformly for $0<z<1$, and

$$
\sum_{t=1}^{n}X_{t-1}e_t=O_P(n),
$$

by using the functional limit theorem for the partial sum of the linear process (Phillips and Solo, 1992) and continuous mapping (Billingsley, 1968), we can prove that

$$
\sup_{0<z<1}\big\|Q_{n3}(z)\big\|=O_P(nh)=o_P(n\sqrt{h\log n}). \tag{A.30}
$$

For $Q_{n4}(z)$, it is easy to check that $\big\{(u_t(K,z)X_{t-1}e_t, \mathcal{F}_t^*)\big\}$ is a sequence of martingale differences, where

$$u_t(K,z) = K\big(\frac{Z_t - z}{h}\big) - \mathbb{E}\big[K\big(\frac{Z_t - z}{h}\big)\big], \quad \mathcal{F}_t^* = \sigma\big\{\eta_{t+1}, (Z_s, \eta_s, \varepsilon_s) : s \le t\big\}.$$

The following proof is similar to the proof of (A.22) with some modifications. We cover the interval $(0,1)$ by a finite number of disjoint intervals $\mathbb{S}_k$ with centre point s_k and radius $r_{n*} = h^{3/2}\sqrt{\log n}/\sqrt{n}$, and the number of these intervals is $N_n = O(r_{n*}^{-1})$. By some standard arguments, we have

$$\begin{aligned}\sup_{0<z<1}\Big\|\sum_{t=1}^n u_t(K,z)X_{t-1}e_t\Big\| &\le \max_{1\le k\le N_n}\sup_{s\in\mathbb{S}_k}\Big\|\sum_{t=1}^n X_{t-1}e_t\big[u_t(K,s) - u_t(K,s_k)\big]\Big\| \\ &\quad + \max_{1\le k\le N_n}\Big\|\sum_{t=1}^n u_t(K,s_k)X_{t-1}e_t\Big\|.\end{aligned}$$

Noting that

$$|u_t(K,s) - u_t(K,s_k)| = O_P(r_{n*}h^{-1}),$$

and $\max_{1\le t\le n}\|X_t\| = O_P(\sqrt{n})$, we can show that

$$\max_{1\le k\le N_n}\sup_{s\in\mathbb{S}_k}\Big\|\sum_{t=1}^n X_{t-1}e_t\big[u_t(K,s) - u_t(K,s_k)\big]\Big\| = O_P(n^{3/2}r_{n*}h^{-1}) = O_P(n\sqrt{h\log n}). \quad \text{(A.31)}$$

We next prove that

$$\max_{1\le k\le N_n}\Big\|\sum_{t=1}^n u_t(K,s_k)X_{t-1}e_t\Big\| = O_P(n\sqrt{h\log n}). \quad \text{(A.32)}$$

As $\frac{n^{2+\delta_0}h^{4+\delta_0}}{(\log n)^{4+\delta_0}} \to \infty$, there exists a positive function $l(n)$ such that

$$l(n) \to \infty \quad \text{and} \quad \frac{n^{2+\delta_0}h^{4+\delta_0}}{l(n)(\log n)^{4+\delta_0}} \to \infty. \quad \text{(A.33)}$$

Let $W_t(s_k) = u_t(K,z)X_{t-1}e_t$, $L(n) = \big[l(n)\big]^{\frac{1}{4+\delta_0}}$, and

$$\overline{W}_t(s_k) = W_t(z_k)\cdot \mathrm{I}\Big(\|X_{t-1}\| \le \sqrt{nL(n)}, |e_t| \le \sqrt{\frac{nh}{L(n)\log n}}\Big), \quad \widetilde{W}_t(s_k) = W_t(s_k) - \overline{W}_t(s_k).$$

From the definition of $\widetilde{W}_t(s_k)$, it is easy to see that if the two events $\Big\{\|X_{t-1}\| \le \sqrt{nL(n)}, t = 1,\cdots,n\Big\}$ and $\Big\{|e_t| \le \sqrt{\frac{nh}{L(n)\log n}}, t = 1,\cdots,n\Big\}$ hold simultaneously, $\Big\|\sum_{t=1}^n \widetilde{W}_t(s_k)\Big\| = 0$ for any $1 \le k \le N_n$. In other words, if $\Big\|\sum_{t=1}^n \widetilde{W}_t(s_k)\Big\| > 0$, we must have either $\Big\{\|X_{t-1}\| > \sqrt{nL(n)}\Big\}$ for at least one $1 \le t \le n$, or $\Big\{|e_t| > \sqrt{\frac{nh}{L(n)\log n}}\Big\}$ for at least one $1 \le t = 1 \le n$. Hence, we have for any $\varepsilon > 0$,

$$
\begin{aligned}
& \mathbb{P}\Big\{\max_{1\le k\le N_n}\Big\|\sum_{t=1}^{n}\widetilde{W}_t(s_k)\Big\| > \varepsilon n\sqrt{h\log n}\Big\} \\
\le\ & \mathbb{P}\Big\{\max_{1\le t\le n}\|X_{t-1}\| > \sqrt{nL(n)}\Big\} + \mathbb{P}\Big\{\max_{1\le t\le n}|e_t| > \sqrt{\frac{nh}{L(n)\log n}}\Big\} \\
=\ & o(1) + O\Big(\frac{n[L(n)\log n]^{(4+\delta_0)/2}}{(nh)^{(4+\delta_0)/2}}\Big) = o(1), \qquad \text{(A.34)}
\end{aligned}
$$

by (A.33), and we can show that

$$
\max_{1\le k\le N_n}\Big\|\sum_{t=1}^{n}\widetilde{W}_t(s_k)\Big\| = o_P\big(n\sqrt{h\log n}\big). \qquad \text{(A.35)}
$$

On the other hand, by the exponential inequality for martingale differences and letting c_2 be some positive constant, we have for sufficiently large $M > c_2 > 0$,

$$
\begin{aligned}
& \mathbb{P}\Big\{\max_{1\le k\le N_n}\|\sum_{t=1}^{n}\overline{W}_t(s_k)\| > Mn\sqrt{h\log n}\Big\} \\
\le\ & \sum_{k=1}^{N_n}\exp\Big\{-\frac{Mnh^2\log n}{c_2nh^2}\Big\} \\
\le\ & O(r_{n*}^{-1}n^{-M/c_2}) = o(1),
\end{aligned}
$$

which indicates that

$$
\max_{1\le k\le N_n}\Big\|\sum_{t=1}^{n}\overline{W}_t(s_k)\Big\| = O_P\big(n\sqrt{h\log n}\big). \qquad \text{(A.36)}
$$

In view of (A.35) and (A.36), we can complete the proof of (A.32), which together with (A.31), indicates that

$$
\sup_{0<z<1}\|Q_{n4}(z)\| = O_P\big(n\sqrt{h\log n}\big). \qquad \text{(A.37)}
$$

Then, by (A.30) and (A.37), we can show that

$$
\sup_{0<z<1}\|Q_{n1}(z)\| = O_P\big(n\sqrt{h\log n}\big). \qquad \text{(A.38)}
$$

We next consider $Q_{n2}(z)$, which is relatively simpler. Let $v_t = \sum_{j=0}^{\infty}\Phi_j\varepsilon_{t-j} = \varepsilon_t + \sum_{j=1}^{\infty}\Phi_j\varepsilon_{t-j} \equiv \varepsilon_t + \widehat{v}_t$ and $e_t = \sum_{j=0}^{\infty}\phi_j\eta_{t-j} = \eta_t + \sum_{j=1}^{\infty}\phi_j\eta_{t-j} \equiv \eta_t + \widehat{e}_t$. Note that

$$
\begin{aligned}
Q_{n2}(z) &= \sum_{t=1}^{n}K\Big(\frac{Z_t-z}{h}\Big)\varepsilon_t\eta_t + \sum_{t=1}^{n}K\Big(\frac{Z_t-z}{h}\Big)\widehat{v}_t\eta_t + \\
&\quad \sum_{t=1}^{n}K\Big(\frac{Z_t-z}{h}\Big)\varepsilon_t\widehat{e}_t + \sum_{t=1}^{n}K\Big(\frac{Z_t-z}{h}\Big)\widehat{v}_t\widehat{e}_t \\
&\equiv \sum_{k=5}^{8}Q_{nk}(z). \qquad \text{(A.39)}
\end{aligned}
$$

Applying the decompositions:

$$\begin{aligned}
Q_{n7}(z) &= \sum_{t=1}^{n}\mathbb{E}\big[K\big(\frac{Z_t-z}{h}\big)\varepsilon_t\big]\widehat{e}_t+\sum_{t=1}^{n}\big\{K\big(\frac{Z_t-z}{h}\big)\varepsilon_t-\mathbb{E}\big[K\big(\frac{Z_t-z}{h}\big)\varepsilon_t\big]\big\}\widehat{e}_t \\
Q_{n8}(z) &= \sum_{t=1}^{n}\mathbb{E}\big[K\big(\frac{Z_t-z}{h}\big)\big]\widehat{v}_t\widehat{e}_t+\sum_{t=1}^{n}\big\{K\big(\frac{Z_t-z}{h}\big)-\mathbb{E}\big[K\big(\frac{Z_t-z}{h}\big)\big]\big\}\widehat{v}_t\widehat{e}_t,
\end{aligned}$$

and following the proof of (A.37), we may show that

$$\begin{aligned}
\sup_{0<z<1}\|Q_{n7}(z)\| &= o_P(n\sqrt{h\log n}), &\text{(A.40)}\\
\sup_{0<z<1}\|Q_{n8}(z)\| &= o_P(n\sqrt{h\log n}). &\text{(A.41)}
\end{aligned}$$

Meanwhile, following the proof of the uniform consistency results in the stationary case (*i.i.d.* or stationary martingale differences), we can also prove that

$$\sup_{0<z<1}\|Q_{n5}(z)\| = o_P(n\sqrt{h\log n}), \tag{A.42}$$

$$\sup_{0<z<1}\|Q_{n6}(z)\| = o_P(n\sqrt{h\log n}). \tag{A.43}$$

In view of (A.40)–(A.43), we can show that

$$\sup_{0<z<1}\|Q_{n2}(z)\| = o_P(n\sqrt{h\log n}). \tag{A.44}$$

Then, the proof of Theorem 3.1 can be completed by (A.28), (A.38) and (A.44). □

PROOF OF THEOREM 4.1. Note that

$$\begin{aligned}
\widehat{\beta}_n(z)-\beta(z) &= \Big[\sum_{t=1}^{n}X_tX_t'K\big(\frac{t-nz}{nh}\big)\Big]^{+}\Big\{\sum_{t=1}^{n}X_tX_t'\big[\beta\big(\frac{t}{n}\big)-\beta(z)\big]K\big(\frac{t-nz}{nh}\big)\Big\}+\\
&\quad \Big[\sum_{t=1}^{n}X_tX_t'K\big(\frac{t-nz}{nh}\big)\Big]^{+}\Big[\sum_{t=1}^{n}X_te_tK\big(\frac{t-nz}{nh}\big)\Big]\\
&\equiv \Pi_{n1}(z)+\Pi_{n2}(z). &\text{(A.45)}
\end{aligned}$$

By Lemma B.4 in Phillips *et al* (2013) and Assumption 5, we may show that the matrix

$$R_n^{+}D_n(z)'\Big[\sum_{t=1}^{n}X_tX_t'K\big(\frac{t-nz}{nh}\big)\Big]D_n(z)R_n^{+}$$

is not-singular with probability 1 uniformly for $z\in(h,1-h)$. Then, by Theorem 2.1, we can prove

$$\sup_{h<z<1-h}\|R_nD_n(z)'\Pi_{n2}(z)\| = O_P(\sqrt{\log n}). \tag{A.46}$$

By Taylor expansion of $\beta(\cdot)$ and Assumption 4, we can show that

$$\beta\big(\frac{t}{n}\big) - \beta(z) = O(h^{\alpha_0}), \quad \big|\frac{t}{n} - z\big| \leq h. \tag{A.47}$$

By (A.47) and standard arguments it readily follows that

$$\sup_{h<z<1-h} \|\Pi_{n1}(z)\| = O_P(h^{\alpha_0}). \tag{A.48}$$

The proof of Theorem 4.1 can be completed in view of (A.45), (A.46), and (A.48) in conjunction with the definitions of R_n and $D_n(z)$. □

Proof of Theorem 4.2. The proof is similar to the proof of Theorem 4.1 above. As in (A.45), we have

$$\begin{aligned}
\widehat{\beta}_n(z) - \beta(z) &= \Big[\sum_{t=1}^{n} X_t X_t' K\big(\frac{Z_t - z}{h}\big)\Big]^{+} \Big\{\sum_{t=1}^{n} X_t X_t' \big[\beta\big(Z_t\big) - \beta(z)\big] K\big(\frac{Z_t - z}{h}\big)\Big\} + \\
&\quad \Big[\sum_{t=1}^{n} X_t X_t' K\big(\frac{Z_t - z}{h}\big)\Big]^{+} \Big[\sum_{t=1}^{n} X_t e_t K\big(\frac{Z_t - z}{h}\big)\Big] \\
&\equiv \Pi_{n3}(z) + \Pi_{n4}(z).
\end{aligned} \tag{A.49}$$

Following the proof of Proposition A.1 in Li *et al* (2013), we can show that the random denominator $\frac{1}{n^2 h}\sum_{t=1}^{n} X_t X_t' K\big(\frac{t-nz}{nh}\big)$ is non-singular with probability 1 uniformly for $z \in (0,1)$. Then, by Theorem 3.1, we can prove

$$\sup_{0<z<1} \|\Pi_{n4}(z)\| = O_P\Big(\sqrt{\frac{\log n}{n^2 h}}\Big). \tag{A.50}$$

On the other hand, by Taylor expansion of $\beta(\cdot)$ and Assumption 4, it follows easily that

$$\sup_{0<z<1} \|\Pi_{n3}(z)\| = O_P(h^{\alpha_0}). \tag{A.51}$$

The proof of Theorem 4.2 is completed by using (A.49)–(A.51). □

References

Billingsley, P. (1968). *Convergence of Probability Measure*. Wiley, New York.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction* (2nd ed.). Lecture Notes in Statistics 110. Springer–Verlag.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, **136**, 163–188.

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying–coefficient models. *Journal of American Statistical Association* **95**, 888–902.

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

Chan, N. and Wang, Q. (2012). Uniform convergence for Nadaraya-Watson estimators with non-stationary data. Working paper, School of Mathematics and Statistics, University of Sydney.

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

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

Duffy, J. (2013). Uniform convergence rates, on a maximal domain, for structural nonparametric cointegrating regression. *Unpublished paper, Yale University*

Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics* **27**, 1491–1518.

Gao, J., Kanaya, S., Li, D. and Tjøstheim, D. (2013). Uniform consistency for nonparametric estimators in null recurrent time series. Forthcoming in *Econometric Theory*.

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

Gao, J. and Phillips, C. B. P. (2013b). Functional coefficient nonstationary regression. *Cowles Foundation Discussion Paper* **1911**. Cowles Foundation for Research in Economics, Yale University.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.

Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Annals of Statistics* **29**, 372–416.

Karlsen, H. A., Myklebust, T. and Tj¿stheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* **35**, 252–299.

Kristensen, D. (2009). Uniform convergence rates of kernel estimators with heterogenous dependent data. *Econometric Theory* **25**, 1433–1445.

Li, D., Lu, Z. and Linton, O. (2012). Local linear fitting under near epoch dependence: uniform consistency with convergence rate. *Econometric Theory* **28**, 935–958.

Li, K., Li, D., Liang, Z. and Hsiao, C. (2013). Estimation of semi-varying coefficient models with nonstationary regressors. Working paper, Department of Econometrics and Business Statistics, Monash University.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications* **65**, 69–80.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift fur Wahrscheinlichskeittheorie und verwandte Gebiete* **61**, 405-415.

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

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

de la Pena, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *Annals of Probability* **27**, 537–564.

Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *Review of Economic Studies*, 53, 473-496.

Phillips, P. C. B. and Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies*, 57, 99-125.

Phillips, P. C. B., Li, D. and Gao, J. (2013). Estimating smooth structural change in cointegration mdoels. Cowles Foundation Discussion Paper 1910.

Phillips, P. C. B. and Park J. (1998). Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper 1181.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics* **20**, 971–1001.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. *Statistical Analysis and Forecasting of Economic Structural Change* (ed. by P. Hackl). Springer, Berlin, pp. 164–253.

Roussas, G. G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Processes and Their Applications* **36**, 107–116.

van der Vaart, A. W. and Wellner, J. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer.

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

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

Wang, Q. and Wang, Y. (2013). Nonparametric cointegrating regression with NNH errors. *Econometric Theory* **29**, 1–27.

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