Econometric Theory, **32**, 2016, 988–1022.
doi:10.1017/S0266466615000018

A CONSISTENT NONPARAMETRIC TEST ON SEMIPARAMETRIC SMOOTH COEFFICIENT MODELS WITH INTEGRATED TIME SERIES

YIGUO SUN
University of Guelph

ZONGWU CAI
University of Kansas and Xiamen University

QI LI
Texas A&M University and
Capital University of Economics and Business

In this paper, we propose a simple nonparametric test for testing the null hypothesis of constant coefficients against nonparametric smooth coefficients in a semiparametric varying coefficient model with integrated time series. We establish the asymptotic distributions of the proposed test statistic under both null and alternative hypotheses. Moreover, we derive a central limit theorem for a degenerate second order U-statistic, which contains a mixture of stationary and nonstationary variables and is weighted locally on a stationary variable. This result is of independent interest and useful in other applications. Monte Carlo simulations are conducted to examine the finite sample performance of the proposed test.

1. INTRODUCTION

Cointegration has proved to be a powerful tool in studying long-run relationships among integrated time series and is a widely used econometric methodology in macroeconomics and financial time series analysis. Nonetheless, empirical evidence often fails to support the existence of cointegrating relations with fixed cointegrating slope coefficients; e.g., see Taylor and Taylor (2004) for an overview on the purchasing power parity debates. Motivated by empirical findings, researchers propose various flexible specifications to relax the constancy restriction of cointegrating vector(s), including, (i) structural breaks (Gregory and Hansen, 1996), (ii) a smooth transition between different economic regimes

We thank the anonymous referees, Peter Phillips and Qiying Wang for their insightful comments that greatly improved our paper. Sun's research is supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) grant 410-2009-0109. Cai's research is supported, in part, by the National Nature Science Foundation of China grants #71131008 (Key Project) and #70971113. Li's research is partially supported by the National Nature Science Foundation of China grant #71133001. Address correspondence to Zongwu Cai, Department of Economics, University of Kansas, Lawrence, KS 66045, USA; e-mail: caiz@ku.edu.

(Saikkonen and Choi, 2004), (iii) varying coefficient models with the coefficients being functions of some additional variables (Cai, Li, and Park, 2009; Xiao, 2009; Sun and Li, 2011; and Sun, Cai, and Li, 2013), or the coefficients being functions of time (Park and Hahn, 1999; Cai and Wang, 2010; Phillips, Li, and Gao, 2013; and Cai, Wang, and Wang, 2014). Alternatively, some researchers directly seek nonlinear cointegrating relation among integrated time series, see, for example, Granger (1991) and Park and Phillips (2001) for parametric nonlinear cointegrating models and Wang and Phillips (2009a, 2009b) about nonparametric cointegrating models.

In this paper we are interested in testing parameter constancy in the framework of semiparametric varying coefficient models studied by Cai et al. (2009) and Xiao (2009), i.e.

$$Y_t = X_t^T \theta(Z_t) + u_t, \quad 1 \le t \le n, \tag{1}$$

where Y_t, Z_t, and u_t are all scalars, X_t is of dimension d, and $\theta(\cdot)$ is a $d \times 1$ vector of unknown smooth functions,[1] the superscript T denotes the transpose of a matrix. We assume that Z_t and u_t are stationary variables, or I(0) variables, while X_t is allowed to contain some nonstationary components. Also, all the variables are continuously distributed. Of course, Z_t can be t/T so that model (1) becomes the time varying coefficient model discussed in Park and Hahn (1999), Cai and Wang (2010), Phillips et al. (2013), and Cai et al. (2014). We are interested in testing the following null hypothesis:

$$H_0: \qquad \Pr\{\theta(Z_t) \equiv \theta_0\} = 1 \quad \text{for some } \theta_0 \in \Theta \tag{2}$$

against an alternative hypothesis of

$$H_1: \qquad \Pr\{\theta(Z_t) \neq \theta\} > 0 \quad \text{for any } \ \theta \in \Theta, \tag{3}$$

where Θ is a compact subset of $\mathcal{R}^d$. That is, we test whether the coefficient functions in (1), $\theta(\cdot)$, are constant. If the null hypothesis holds true, model (1) becomes a linear cointegrating model; otherwise, model (1) is a semiparametric varying cointegrating model.

There are several new and interesting findings in this paper. First, the power of our test statistic depends on the stochastic property of X_t. Specifically, we consider two cases. In Case (a), X_t is an integrated process of order one (or $I(1)$); in Case (b), $X_t^T = (X_{1,t}^T, X_{2,t}^T)^T$, where $X_{1,t}$ is $I(0)$ and $X_{2,t}$ is $I(1)$.[2] We show that the proposed test is consistent under both cases, although the power of the test varies. If the null hypothesis fails to hold, under Case (a), the test statistic diverges to $+\infty$ at the rate of $n^2\sqrt{h}$; under Case (b), the test diverges at the rate of $n^2 h$ when the coefficients for the I(1) regressors (i.e., $X_{2,t}$) are nonconstant, and the divergence rate reduces to $n\sqrt{h}$ when the coefficients for the I(0) variables are nonconstant but the coefficients for the I(1) variables are constant. These results suggest that the presence of a stationary component (under H_1) reduces the power

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

of the test by an order of $\sqrt{h}$, while the presence of a nonstationary component enhances the power of the test by an order of n.

Another interesting and perhaps surprising finding of the paper is obtained for Case (b). Let $\theta_j(z)$ be the functional coefficient for variable $X_{j,t}$ for $j = 1, 2$. When $\theta_1(z) \equiv \theta_{10}$ (a constant vector) for all z and $\theta_2(z)$ does not equal any constant over a nonempty interval of z, then the least squares estimator $\widehat{\theta}_1$ based on a misspecified linear (null) model diverges to ∞ at the rate of root-n if $Cov\left[X_{1,t}, \theta_2(Z_t)\right] \neq 0$. Therefore, a misspecified linear model leads to inconsistent or divergent OLS estimates from the true parameter value θ_{10}, if the true model is only linear in the stationary covariate $X_{1,t}$, but the coefficient of the nonstationary variable $X_{2,t}$ is a smoothing function of the stationary covariate Z_t. This result suggests that it can be very important to test for the correct model specification when X_t contains both I(0) and I(1) components in model (1). More discussions on this issue are given in Section 3.2 and Appendix A.

Xiao (2009) also independently considered the parameter constancy test for Case (a), where his test statistic is based on the maximum of a sequence of squared (standardized) distances between the kernel estimates calculated at pre-selected points under the alternative hypothesis and the OLS estimate calculated under the null hypothesis. By assuming independency between $\{Z_t\}$ and $\{u_t\}$, Xiao (2009) showed that his proposed test statistic follows a maximum chi-squared distribution under the null hypothesis. With $Z_t = t/n$ in model (1), Park and Hahn (1999) considered the problem of testing for the parameter constancy applying Shin's (1994) residual-based test statistic which was originally used to test the stationarity against nonstationarity of error terms, whereas Cai et al. (2014) proposed a procedure for testing whether $\theta(\cdot)$ has a known parametric functional form for the predictability of asset returns. The test statistics proposed in Park and Hahn (1999) and Cai et al. (2014) do not converge to conventional distributions under the null hypothesis. In contrast, our test statistic given in Section 2 is a consistent test and is asymptotically normally distributed under the null hypothesis.

Against different alternative hypotheses than the semiparametric varying coefficient model considered in this paper, many researchers propose various statistics for testing a linear cointegrating model, including a parameter stability test by Hansen (1992a), a modified RESET test by Hong and Phillips (2010), a nonparametric specification test by Wang and Phillips (2012), and linearity tests of cointegrating smooth transition regressions by Gao, King, Lu, and Tjøstheim (2009) and Choi and Saikkonen (2004), among others. Gao et al. (2009) considered the problem of testing a linear cointegration model, $Y_t = \theta_0 + X_t\theta_1 + u_t$, against a nonlinear cointegration model, $Y_t = g(X_t) + u_t$, where $\{X_t\}_{t=1}^n$ is a random walk process independent of $\{u_t\}_{t=1}^n$. Wang and Phillips (2012) considered a similar testing problem as in Gao et al. (2009) but relaxed many of the restrictive assumptions to allow for a more general nonstationary process for $\{X_t\}_{t=1}^n$. Moreover, Wang and Phillips (2012) did not require $\{X_t\}_{t=1}^n$ to be independent of $\{u_t\}_{t=1}^n$. Choi and Saikkonen (2004) advocated a smooth transition regression model, $Y_t = X_t^T\alpha + \theta^T X_t g(X_{t,j} - c) + u_t$, where X_t is a $p \times 1$ vector of random walk processes, $X_{t,j}$ is the j^{th} component of X_t, and the functional form $g(\cdot)$ is unspecified. Choi and Saikkonen (2004) investigated the problem of testing the null hypothesis of $\theta = 0$, so that the model becomes a linear cointegration model under the null hypothesis. They further applied the Taylor expansion method to the smooth function $g(\cdot)$, so that they essentially tested a linear cointegration model against a parametric nonlinear (with some finite order polynomials in X_t) cointegration model. Our model (1) differs from all the aforementioned models in the sense that under the alternative hypothesis, the coefficients in our model are functions of a stationary variable Z_t with unknown functional forms. While for example, the smooth function $g(\cdot)$ considered in Choi and Saikkonen (2004) is a finite order polynomial function of the nonstationary variable $X_{t,j}$.[3]

Some technical results developed in this paper may be useful in other contexts. For example, Lemmas A.1 and A.6 extend the central limit theorem for degenerate U-statistics of Hall (1984) for i.i.d. data and of Fan and Li (1999) and Gao and Hong (2008) for weakly dependent (absolutely regular) processes to integrated processes. Lemma B.2 in Appendix B gives the convergence results for nondegenerate U-statistics with integrated time series and kernel weights on a stationary variable. In addition, we obtain weak uniform convergence results for a kernel estimator of $\theta(z)$ in model (1), which is used to derive a limiting result of the "asymptotic variance" of the proposed test.[4]

The rest of the paper is organized as follows. Section 2 describes our test statistic. Section 3 studies the asymptotic behaviors of the test statistic under both the null and the alternative hypotheses. Section 4 presents Monte Carlo simulation results to examine the finite sample performance of our test. Section 5 concludes the paper. All the mathematical proofs are relegated to three Appendices.

2. TEST STATISTIC

Following Li , Huang, Li, and Fu (2002), we initiate our test statistic from an L_2 -type test statistic as follows

$$\int \left[\widehat{\theta}(z) - \widehat{\theta}_0\right]^T \left[\widehat{\theta}(z) - \widehat{\theta}_0\right] dz,$$

where $K_t(z) \equiv K((Z_t - z)/h)$, $K(z)$ is a kernel function, h is the bandwidth,

$$\widehat{\theta}(z) = \left[\sum_{t=1}^{n} X_t X_t^T K_t(z)\right]^{-1} \sum_{t=1}^{n} X_t Y_t K_t(z) \tag{4}$$

is the kernel estimator of the unknown smooth coefficient curve $\theta(z)$, and $\widehat{\theta}_0$ is the usual ordinary least squares (OLS) estimator of θ_0 based on the linear (null) model. It is clear that the test statistic has a random denominator. To avoid the random denominator problem, we modify the test statistic with a positive

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

definite weighting matrix $D_n(z) = \sum_{t=1}^{n} X_t X_t^T K_t(z)$, which leads to the following weighted test statistic

$$\int \left[D_n(z)\left(\widehat{\theta}(z) - \widehat{\theta}_0\right)\right]^T \left[D_n(z)\left(\widehat{\theta}(z) - \widehat{\theta}_0\right)\right] dz$$
$$= \sum_{t=1}^{n}\sum_{s=1}^{n} X_t^T X_s \widehat{u}_t \widehat{u}_s \int K_t(z) K_s(z)\, dz, \tag{5}$$

where $\widehat{u}_t = Y_t - X_t^T \widehat{\theta}_0$ is the parametric residual. Define $\bar{K}_{t,s} \equiv \int K_t(z) K_s(z)\, dz$. Then,

$$\bar{K}_{t,s} = h \int K(v) K((Z_s - Z_t)/h + v)\, dv$$

can be regarded as a local weight function selecting (t,s) among all $t \neq s$ such that Z_t and Z_s are close to each other. Therefore, by removing the global center, the term with $t = s$, replacing the local weight function $\bar{K}_{t,s}$ with $K_{t,s} \equiv K((Z_t - Z_s)/h)$ as in Li et al. (2002), and adding a trimming indicator function, we obtain the final test statistic given as follows

$$\widehat{I}_n \equiv \frac{1}{n^3 h} \sum_{t=1}^{n}\sum_{s \neq t}^{n} X_t^T X_s \widehat{u}_t \widehat{u}_s K_{t,s} \mathbf{1}_{n,t,s}, \tag{6}$$

where $\mathbf{1}_{n,t,s} \equiv \mathbf{1}_{n,t}\mathbf{1}_{n,s}$, $\mathbf{1}_{n,t} \equiv \mathbf{1}(Z_t \in \mathcal{S}_n)$, and $\mathbf{1}(A)$ is a trimming indicator function which equals 1 if A holds true and 0 otherwise. The set $\mathcal{S}_n$ trims out the boundary region of the support of Z_t so that we can obtain the weak uniform convergence result for $\widehat{\theta}(z)$ over $z \in \mathcal{S}_n$. Note that $\mathcal{S}_n$ satisfies the condition that $\lim_{n\to\infty} \Pr(Z_t \in \mathcal{S}_n) = 1$ holds uniformly over $t = 1, \ldots, n$; see Lemma C.1 in Appendix C for the construction of the trimming set $\mathcal{S}_n$. Hence, the use of such a trimming function will not affect our test results asymptotically.

Evidently, the proposed test statistic is a second order U-statistic constructed from both I(0) and I(1) variables. To the best of our knowledge, there does not exist any asymptotic result for such a U-statistic, which makes the results of Lemmas A.1 and A.6 be of independent interest and useful in other applications.

3. ASYMPTOTIC RESULTS

For the convenience of readers, we summarize our notation here. (i) For a nondecreasing nonstochastic positive sequence c_n, we use $O_e(c_n)$ to denote an *exact* probability order of c_n; i.e., $A_n = O_e(c_n)$ means that $A_n = O_p(c_n)$ but not $A_n \neq o_p(c_n)$. (ii) Let χ be a finite dimensional matrix of random variables. The L^r-norm of χ is denoted by $\|\chi\|_r = \left(\sum_i \sum_j E\left|\chi_{i,j}\right|^r\right)^{1/r}$, where $\chi_{i,j}$ is the (i,j)-th element of χ, and $\|\cdot\|$ without any subscript denotes the Euclidean norm. (iii) "$\xrightarrow{d}$", "$\xrightarrow{p}$", and "$\xrightarrow{a.s.}$" stand for the convergence in distribution, in probability, and almost surely, respectively, and "$\Rightarrow$" denotes the weak

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

convergence with respect to the Skorohod metric as defined in Billingsley (1999). (iv) We write "$A \equiv B$" to define A by a known or previously defined quantity B, or to assign the quantity A to a new notation B (usually A has a long expression and B is a shorthand notation). (v) $[a]$ is the integer part of a ($a > 0$), and $[0,1]^d = [0,1] \times \cdots \times [0,1]$ refers to the product space of d-multiplication of interval [0,1]. (vi) We denote a generic positive constant by C that may take different values at different places.

Throughout this paper, we assume that $\{Z_t\}_{t=1}^n$ is a strictly stationary, absolutely regular (β-mixing) sequence, and that X_t may contain a constant term but it does not contain any deterministic trend variables. We derive the asymptotic results of the proposed test statistic when X_t is nonstationary in Section 3.1, and when X_t contains both stationary and nonstationary variables in Section 3.2. Below we use $\widehat{I}_n^a$ and $\widehat{I}_n^b$ to denote the test statistic $\widehat{I}_n$ under Case (a) and Case (b), respectively.

3.1. Case (a): X_t is a Vector of Integrated Variables

In this section, X_t is a $d \times 1$ vector of $I(1)$ variables. Below we list some assumptions on the data-generating mechanism of $\{(X_t, Z_t, u_t)\}_{t=1}^n$.

(A1) (i) $X_t = X_{t-1} + \eta_t$ for $1 \le t \le n$, where $X_0 = O_p(1)$ and $\max_{1\le t\le n} E\left(\|\eta_t\|^q\right) \le C < \infty$ for some $q > 8$;

(ii) $\{Z_t\}_{t=1}^n$ is a strictly stationary and absolutely regular (β-mixing) sequence with β-mixing coefficients satisfying $\beta_\tau = O\left(\rho^{-\tau}\right)$ for some $\rho > 1$;

(iii) $\{u_t\}_{t=1}^n$ is independent of $\{(X_t, Z_t)\}_{t=1}^n$; $\left\{\left((u_t, \eta_t), \mathcal{F}_{n,t-1}\right)\right\}_{2\le t\le n}$ and forms a martingale difference sequence with $\sigma_\upsilon^2 = E\left(\upsilon_t^2\right) < \infty$, $\sup_{2\le t\le n}\left|E\left(\upsilon_t^2|\mathcal{F}_{n,t-1}\right) - \sigma_\upsilon^2\right| \overset{a.s.}{\to} 0$, and $\sup_{2\le t\le n} E\left(\upsilon_t^4|\mathcal{F}_{n,t-1}\right) < C < \infty$ for $\upsilon = \eta$ or $\upsilon = u$, where $\mathcal{F}_{n,t} = \sigma\left((u_s, \eta_s, Z_{s+1}), s \le t\right)$ is the smallest σ-field containing all the past history of (u_t, η_t, Z_{t+1}) for $1 \le t \le n$.

(A2) Denote $B_{n,\eta}(r) \equiv n^{-1/2}\sum_{t=1}^{[nr]}\eta_t$. There exists a vector Brownian motion B_η such that

$$B_{n,\eta}(r) \Rightarrow B_\eta(r) \tag{7}$$

on $D[0,1]^d$ as $n \to \infty$, where $D[0,1]^d$ is the space of cadlag functions on $[0,1]^d$ equipped with Skorohod topology, and B_η is a d-dimensional multivariate Brownian motion with a finite positive definite covariance matrix $\Sigma_\eta = \lim_{n\to\infty} \mathrm{Var}\left(n^{-1/2}\sum_{t=1}^n \eta_t\right)$.

Remark 1. The mutual independence between $\{u_t\}_{t=1}^n$ and $\{(X_t, Z_t)\}_{t=1}^n$ in A1 (iii) looks restrictive, but it is not unreasonable given the complexity of our test problem; e.g., Gao et al. (2009) made the same independence assumption when considering a test statistic similar to ours. The martingale difference condition is

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

required for the application of Wang's (2014, Thm. 2.1) generalized martingale central limit theorem when we derive the limiting distribution of our proposed test statistic under the null hypothesis. Taking the weak convergence result in (7) as an assumption is commonly done in the econometrics literature; e.g., Assumption 2.2 in Wang and Phillips (2009a). The conditions for the multivariate functional central limit theorem for partial sums of weakly dependent random vectors can be found in Wooldridge and White (1988) and de Jong and Davidson (2000). Moreover, Assumption A2 assumes that the $d \times 1$ vector X_t is not cointegrated with itself as Σ_η is finite and nonsingular; see Phillips (1986).

It follows clearly by (7) and applying the continuous mapping theorem that $\sup_{0\le r\le 1}\left|B_{n,\eta}(r)\right| \xrightarrow{d} \sup_{0\le r\le 1}\left|B_\eta(r)\right|$. As $\sup_{0\le r\le 1}\left|B_\eta(r)\right| = O_p(1)$, hence, we have

$$\max_{1\le t\le n}|X_t| = O_p\left(\sqrt{n}\right), \tag{8}$$

which is frequently used in our proofs in the appendices.

To derive the limiting distribution of $\widehat{I}_n^a$, we need additional regularity assumptions listed below.

(A3) The sequence $\{Z_t\}_{t=1}^n$ has a common Lebesgue probability density $f(z)$ with bounded uniformly continuous derivatives up to the second order over the support of Z_t. Let $f_{t,s}(z_t, z_s)$ be the joint probability density function of (Z_t, Z_s) for $t \neq s$. Then, $f_{t,s}(z_t, z_s)$ and its first- and second-order partial derivatives are all continuous and uniformly bounded over its support and over all $t \neq s$.

(A4) $\theta(z)$ has bounded uniformly continuous derivatives up to the second order and $\|\theta(Z_t)\|_q < C < \infty$ for some $q \geq 2$.

(A5) (i) The kernel function $K(u)$ is a symmetric (around zero) probability density function on interval $[-1, 1]$;
(ii) $K(\cdot)$ satisfies that $\left|K(u) - K\left(u'\right)\right| \leq C\left|u - u'\right|$ for some $C < \infty$.

(A6) As $n \to \infty$, $h \to 0$ and $nh/\ln n \to \infty$.

Assumptions A3–A6 are regularity assumptions commonly imposed in a nonparametric framework. Next we present the asymptotic properties of our test statistic with the detailed proofs relegated to Appendix A.

THEOREM 3.1. *Under Assumptions A1–A3, A5, and A6, we have,*

(i) under H_0,

$$J_n^a \equiv n\sqrt{h}\widehat{I}_n^a / \sqrt{\widehat{\sigma}_{n,a}^2} \xrightarrow{d} N(0,1), \tag{9}$$

where

$$\widehat{\sigma}_{n,a}^2 \equiv \frac{2}{n^4 h}\sum_{t=1}^n \sum_{s\neq t}^n \tilde{u}_t^2 \tilde{u}_s^2 \left(X_t^T X_s\right)^2 K_{t,s}^2 \mathbf{1}_{n,t,s} \xrightarrow{d} \sigma_a^2, \tag{10}$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

$\tilde{u}_t = Y_t - X_t^T \widehat{\theta}^{(-t)}(Z_t)$ *is the nonparametric residual*[5] *with the leave-one-out semiparametric estimator* $\widehat{\theta}^{(-t)}(Z_t)$ *for all* t, $\sigma_a^2 = 4\sigma_u^4 \upsilon_2 E[f(Z_1)] \int_0^1 \int_0^s \left(B_\eta(s)^T B_\eta(r)\right)^2 dr ds$ *is an almost surely positive random variable, and* $\upsilon_2 = \int K^2(u) du$;

(ii) *under* H_1 *and if Assumption A4 also holds, the test statistic* J_n^a *diverges to* $+\infty$ *at the rate of* $n^2\sqrt{h}$, *viz.*

$$\Pr\left(J_n^a > C_n\right) \to 1 \text{ as } n \to \infty,$$

for any nonstochastic positive sequence $C_n = o(n^2\sqrt{h})$.

Theorem 3.1 indicates that J_n^a is a consistent one-sided test as the leading term of J_n^a diverges to positive infinity at the rate of $n^2\sqrt{h}$ under H_1; see (A.21) in the proof of Theorem 3.1 in Appendix A. The null hypothesis is thus rejected at the significance level α if J_n^a is greater than z_α, the $(1-\alpha)$ 100th percentile of a standard normal distribution for $\alpha \in (0,1)$.

3.2. Case (b): X_t Contains Both Stationary and Integrated Variables

We decompose X_t (a $d \times 1$ vector) into two groups: $X_t = \left(X_{1,t}^T, X_{2,t}^T\right)^T$, where $X_{1,t}$ is of dimension d_1 with its first component unity and the remainder I(0) variables, and $X_{2,t}$ is of dimension d_2 with I(1) variables. Model (1) then becomes $Y_t = X_{1,t}^T \theta_1(Z_t) + X_{2,t}^T \theta_2(Z_t) + u_t$. The null and alternative hypotheses are defined by (2) and (3), respectively, where under the null hypothesis we have a linear cointegrating model, $Y_t = X_{1,t}^T \theta_{10} + X_{2,t}^T \theta_{20} + u_t$. The test statistic is given by (6) and is denoted by $\widehat{I}_n^b$.

Below we only list assumptions that replace the corresponding assumptions listed in Section 3.1.

(B1) (i) We assume that X_1, an $n \times d_1$ matrix containing n observations on $X_{1,t}$, has a full column rank and that $X_{2,t} = X_{2,t-1} + \eta_t$ satisfies Assumption A1(i) with $B_{n,\eta}(r) \equiv n^{-1/2} X_{2,[nr]}$ for $r \in [0,1]$;

(ii) The sequence $\left\{X_{1,t}, \eta_t, Z_t\right\}_{t=1}^n$ is a strictly stationary and absolutely regular (β-mixing) process with β-mixing coefficients satisfying $\beta_\tau = O\left(\rho^{-\tau}\right)$ for some $\rho > 1$;

(iii) Assumption A1 (iii) holds with $\mathcal{F}_{n,t} = \sigma\left(\left(u_s, \eta_s, X_{1,s}, Z_{s+1}\right), s \le t\right)$ being the smallest σ-field containing all the past history of $\left(u_t, \eta_t, X_{1,t}, Z_{t+1}\right)$ for $1 \le t \le n$.

(B3) (i) Assumption A3 holds.

(ii) For some $s > 2$, $E\left(\left\|X_{1,t} X_{1,t}^T\right\|^s\right) \le C < \infty$ and

$$\sup_{z \in \mathcal{S}} E\left(\left\|X_{1,t} X_{1,t}^T\right\|^s |Z_t = z\right) f(z) \le C < \infty, \tag{11}$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

where $\mathcal{S}$ is the support of Z. Also, $\sup_{(z_0,z_t)\in\mathcal{S}\times\mathcal{S}} E\left(\left\|X_{1,t}X_{1,t}^T X_{1,0}X_{1,0}^T\right\| |Z_0 = z_0, Z_t = z_t\right) f_{0,t}(z_0,z_t) \leq C < \infty$, where $f_{0,t}(z_0,z_t)$ is the joint density function of (Z_0, Z_t).

(iii) $E\left(X_{1,t}|Z_t = z\right)$, $E\left(X_{1,t}X_{1,t}^T|Z_t = z\right)$, and $E\left(\left\|X_{1,t}X_{1,t}^T\right\|^{1+\delta_0}|Z_t = z\right)$ all have bounded uniformly continuous derivatives up to the second order for some $\delta_0 > 0$.

Remark 2. Assumptions A5 (ii) and B3 are used to derive the weak uniform convergence rate for the kernel estimators: $\sup_{z\in\mathcal{S}_n}\left\|\widehat{\theta}_1(z) - \theta_1(z)\right\| = o_p(1)$ and $\sup_{z\in\mathcal{S}_n}\left\|\widehat{\theta}_2(z) - \theta_2(z)\right\| = o_p\left(n^{-1/2}\right)$, see Lemma C.1 in Appendix C, where $\mathcal{S}_n$ is an expanding bounded subset of S and is given in Remark 6 at the end of Appendix C.

In the next two theorems, we present the limiting results of the test statistic $\widehat{I}_n^b$ under the null and alternative hypotheses. The proofs are relegated to Appendix A.

THEOREM 3.2. *Under Assumptions B1, A2, B3, A5, and A6, we have, under H_0,*

$$J_n^b \equiv n\sqrt{h}\widehat{I}_n^b/\sqrt{\widehat{\sigma}_{n,b}^2} \xrightarrow{d} N(0,1),$$

where $\widehat{\sigma}_{n,b}^2 \xrightarrow{d} \sigma_b^2$, and $\widehat{\sigma}_{n,b}^2$ and σ_b^2 have respectively the same mathematical representation as $\widehat{\sigma}_{n,a}^2$ and σ_a^2 defined in Theorem 3.1.

In Appendix A we show that the asymptotic property of $\widehat{\sigma}_{n,b}^2$ is dominated by the I(1) covariates so that one can replace $\left(X_t^T X_s\right)^2$ in $\widehat{\sigma}_{n,b}^2$ by $\left(X_{2,t}^T X_{2,s}\right)^2$ if the sample size is sufficiently large. The intuitive explanation to this result is as follows: $\widehat{u}_t$ in equation (6) mimics the stochastic properties of u_t under H_0 so that under H_0, the leading term of $\widehat{I}_n^b$ is

$$\widehat{I}_{1n}^b \equiv \frac{1}{n^3 h}\sum_{t=1}^n\sum_{s\neq t}^n X_t^T X_s u_t u_s K_{t,s}\mathbf{1}_{n,t,s},$$

and further, the leading term of $\widehat{I}_{1n}^b$ is given by $n^{-3}h^{-1}\sum_{t=1}^n\sum_{s\neq t} X_{2,t}^T X_{2,s} u_t u_s K_{t,s}\mathbf{1}_{n,t,s}$ because $X_{2,t}^T X_{2,s}$ is the leading term of $X_t^T X_s = X_{1,t}^T X_{1,s} + X_{2,t}^T X_{2,s}$.

THEOREM 3.3. *Under the assumptions of Theorem 3.2 and if Assumption A4 also holds, we have, under H_1,*

(i) *if* $\Pr[\theta_2(Z_t) \neq \theta_2] > 0$ *for any* $\theta_2 \in \Theta_2$, *then* $J_n^b = O_e(n^2 h)$ *which implies that* $\Pr\left(J_n^b > C_n\right) \to 1$ *as* $n \to \infty$ *for any nonstochastic positive sequence* $C_n = o(n^2 h)$;

(ii) *if* $\Pr[\theta_2(Z_t) \equiv \theta_{20}] = 1$ *for some* $\theta_{20} \in \Theta_2$ *and* $\Pr[\theta_1(Z_t) \neq \theta_1] > 0$ *for any* $\theta_1 \in \Theta_1$, *then* $J_n^b = O_e(n\sqrt{h})$ *which implies that* $\Pr\left(J_n^b > C_n\right) \to 1$ *as*

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

$n \to \infty$ for any nonstochastic positive sequence $C_n = o(n\sqrt{h})$; where Θ_1 and Θ_2 are compact subsets of $\mathcal{R}^{d_1}$ and $\mathcal{R}^{d_2}$, respectively.

Remark 3. Theorem 3.3 shows that under H_1, the test statistic J_n^b diverges to $+\infty$ at different rates depending on whether or not $\theta_2(z) \equiv \theta_{20}$ (a constant vector). Although J_n^b is a consistent test under both cases, more samples are required for the power of the test statistic to approach one under case (ii) than under case (i). Moreover, the proof in Appendix A indicates that when $\theta_1(z) \equiv \theta_{10}$ (a constant vector) for all z and $\theta_2(z) \neq \theta_{20}$ over a nonempty interval of z, the least squares estimator $\widehat{\theta}_1$ of the misspecified linear regression model diverges to ∞ at the rate of root-n if $Cov\left(X_{1,t}, \theta_2\left(Z_t\right)\right) \neq 0$. This result suggests that it is very important to test for the correct model specification when X_t contains both I(0) and I(1) components in model (1).

Specifically, when $\theta_1(z) \equiv \theta_{10}$ (a constant vector) for all z and $\theta_2(z) \neq \theta_{20}$ over a nonempty interval of z, model (1) becomes $Y_t = X_{1,t}^T \theta_{10} + X_{2,t}^T \theta_2(Z_t) + u_t = X_{1,t}^T \theta_{10} + X_{2,t}^T c_0 + \epsilon_t$, where $c_0 = E[\theta_2(Z_t)]$ and $\epsilon_t \equiv e_t + u_t = X_{2,t}^T \{\theta_2(Z_t) - E[\theta_2(Z_t)]\} + u_t$. Applying the partitioned inverse to the OLS estimator gives $\widehat{\theta}_1 = \left(X_1^T M_2 X_1\right)^{-1} X_1^T M_2 Y$, where $M_2 = I_n - X_2\left(X_2^T X_2\right)^{-1} X_2^T$ and I_n is the $n \times n$ identity matrix, so that $\widehat{\theta}_1 - \theta_{10} = \left(X_1^T M_2 X_1\right)^{-1} X_1^T M_2 \epsilon$. It is easy to show that $n^{-1} X_1^T M_2 X_1 = O_e(1)$ and $n^{-1/2} \sum_{t=1}^n u_t = O_p(1)$, so the stochastic order of $\widehat{\theta}_1 - \theta_{10}$ is determined by $n^{-1} X_1^T M_2 e$. If $Cov\left[X_{1,t}, \theta_2^T\left(Z_t\right)\right] \neq 0$, we have $n^{-3/2} \sum_{t=1}^n X_{1,t} X_{2,t}^T \{\theta_2\left(Z_t\right) - E[\theta_2(Z_t)]\} \xrightarrow{d} Cov\left[X_{1,t}, \theta_2^T(Z_t)\right] \times \int_0^1 B_\eta(r)\, dr$, so $\widehat{\theta}_1 - \theta_{10} = O_p\left(\sqrt{n}\right)$. However, $n^{-1} \sum_{t=1}^n X_{1,t} X_{2,t}^T \{\theta_2(Z_t) - E[\theta_2(Z_t)]\} = O_e(1)$ if $Cov\left[X_{1,t}, \theta_2^T\left(Z_t\right)\right] = 0$, which leads to $\widehat{\theta}_1 - \theta_{10} = O_e(1)$. Therefore, it is the correlation between the stationary variables $X_{1,t}$ and $\theta_2(Z_t)$ (the varying coefficients for the integrated variable $X_{2,t}$) that nurtures the inflation of $\widehat{\theta}_1$ by a magnitude of $\sqrt{n}$ when the coefficients for the integrated variables are wrongly specified as constants. Note that $J_n^b = O_e(n^2 h)$ holds true whether $Cov\left[X_{1,t}, \theta_2^T\left(Z_t\right)\right]$ equals zero or not, as terms containing $X_{2,t}$ are the dominating terms of $\widehat{I}_n^b$.

4. MONTE CARLO SIMULATIONS

In this section we use Monte Carlo simulations to examine the finite sample performance of the proposed test. The power simulation results reported below include two cases: (i) the coefficients for the $I(1)$ variables are nonconstant; (ii) the coefficients for the $I(1)$ variables are constant, but the coefficient for the $I(0)$ variable is nonconstant. The test statistic is computed as

$$\widehat{I}_n = \frac{1}{n^3 h} \sum_{t=1}^{n} \sum_{s \neq t}^{n} X_t^T X_s \widehat{u}_t \widehat{u}_s K\left(\frac{Z_t - Z_s}{h}\right). \tag{12}$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

Note that we did not use any trimming indicator function in the simulations. In practice, the data support is always finite. Thus, the trimming indicator carries theoretical importance, but it may not be needed in practice.

The standardized test statistic is given by

$$J_n = n\sqrt{h}\widehat{I}_n / \sqrt{\widehat{\sigma}_n^2}, \tag{13}$$

where $\widehat{\sigma}_n^2 = 2n^{-4}h^{-1}\sum_{t=1}^{n}\sum_{s\neq t}^{n}\tilde{u}_t^2\tilde{u}_s^2\left(X_t^T X_s\right)^2 K_{ts}^2$, $\widehat{u}_t$ in (12) and $\tilde{u}_t$ in $\widehat{\sigma}_n^2$ are the respective OLS and semiparametric residuals calculated under H_0 and H_1. Under H_0, J_n is asymptotically normally distributed with zero mean and unit variance by Theorems 3.1 and 3.2.

We consider the following data-generating process (DGP):

$$Y_t = \theta_1(Z_t)X_{1,t} + \theta_2(Z_t)X_{2,t} + \theta_3(Z_t)X_{3,t} + u_t, \tag{14}$$

where $X_{1,t} \equiv 1$, $X_{2,t} = \sum_{s=1}^{t} v_{1,s}$, $X_{3,t} = \sum_{s=1}^{t} v_{2,s}$, and $\{v_{1,s}\}$ and $\{v_{2,s}\}$ are both randomly drawn from i.i.d. $N(0,1)$, so that $\{X_{1,t}\}$ is an I(0) process and $\{X_{2,t}\}$ and $\{X_{3,t}\}$ are both I(1) processes; $\{Z_t\}$ is randomly drawn from i.i.d. uniform $[0,2]$; $\{u_t\}$ is randomly drawn from i.i.d. $N(0,\sigma_u^2)$ with $\sigma_u = 2$. Also, $\{v_{1,t}\}$, $\{v_{2,t}\}$, $\{Z_t\}$, and $\{u_t\}$ are all mutually independent of each other. We set $\theta_1(z) = \gamma_1 + \gamma_2 z$, $\theta_2(z) = \gamma_3 + \gamma_4 \sin(z)$, and $\theta_3(z) = \gamma_5 + \gamma_6 z$. Three DGPs are considered by setting different values to the six parameters γ_j's for $j = 1,\ldots,6$. They are set as follows:

$$\text{DGP}_1 : (\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5,\gamma_6) = (1,0,0.5,0,0.5,0)$$
$$\text{DGP}_2 : (\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5,\gamma_6) = (1,0.5,0,0.5,0.5,0.3)$$
$$\text{DGP}_3 : (\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5,\gamma_6) = (1,2,0.5,0,0.5,0)$$

where DGP_1 satisfies the null hypothesis with $Y_t = X_{1,t} + 0.5X_{2,t} + 0.5X_{3,t} + u_t$, and both DGP_2 and DGP_3 violate the null hypothesis. None of the three coefficient curves are constant under DGP_2, while the coefficient curves for the integrated variables $X_{2,t}$ and $X_{3,t}$ are constant under DGP_3. To measure the distance between the null and alternative hypotheses, we define

$$\widehat{D}_j = \sum_{l=1}^{3}\frac{1}{n}\sum_{t=1}^{n}\left[\theta_{l,j}(Z_t) - \theta_{j,0}\right]^2 \xrightarrow{p} \sum_{l=1}^{3} E\left[\theta_{l,j}(Z_1) - \theta_{l,0}\right]^2 \equiv D_j,$$

where j refers to the experiment corresponding to DGP_j, $\theta_{l,0}$ and $\theta_{l,j}(z)$ are the lth coefficient under H_0 and DGP_j, respectively. It is easy to calculate that $D_1 = 0$, $D_2 = 0.4979$, and $D_3 = 16/3 \approx 5.33$.

The number of Monte Carlo replications is 1,000, and the sample size $n = 100$, 200, 400, and 600. According to Sun and Li (2011) we use a Gaussian kernel function with $h = c\widehat{\sigma}_z n^{-.5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of $\{Z_t\}_{t=1}^n$, and we choose $c = 0.8$, 1.0, and 1.2 to examine the effects of different degrees of

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

TABLE 1. Estimated sizes

	c = 0.8				c = 1.0				c = 1.2			
n	1%	5%	10%	20%	1%	5%	10%	20%	1%	5%	10%	20%
100	.006	.019	.037	.080	.006	.018	.035	.077	.007	.016	.037	.084
200	.013	.037	.057	.115	.011	.030	.060	.119	.014	.033	.062	.114
400	.010	.044	.073	.143	.011	.045	.079	.145	.011	.046	.084	.143
600	.005	.040	.074	.135	.008	.039	.069	.139	.010	.038	.065	.131

smoothing on test results. Monte Carlo results for DGP_j are reported in Table j, $j = 1, 2, 3$. Using different bandwidths does have mild impacts on the percentage rejection rates–more on estimated powers than estimated sizes. We observe from Table 1 that some downward size distortion of our test. This is quite common in this type of nonparametric test even for independent or weakly dependent data cases. We have done some simulations using a residual-based bootstrap method to generate the null critical values. The results show significant size improvement using the bootstrap method. Since we do not verify the theoretical validity of the bootstrap method in this paper, these results are not reported here. However, the results are available from the authors upon request.

Although our test is under-sized (see Table 1), Table 2 shows that our test is quite powerful against DGP2 where none of the coefficients of the nonstationary covariates are constant. DGP_3 is designed to measure the power of our test when the coefficients for the I(1) variables are constant but the coefficient for the I(0) variable varies with respect to Z_t. Although $D_3 = 16/3$ (the distance between the null DGP_1 and DGP_3) is much larger than D_2 (the distance between DGP_1 and DGP_2), the rejection rates in Table 3 for each given sample size are smaller than those given in Table 2 in most cases. Moreover, the rejection rates in Table 3 grow at speeds slower than those in Table 2. The results given in Tables 2 and 3 are in line with the theory given in Theorem 3.3. Whether the coefficients of the I(1) variables are constant or not overwhelms the distant measure D_j's in the prediction of the power of the test.

TABLE 2. Estimated powers: Varying coefficients for the I(1) variables

	c = 0.8				c = 1.0				c = 1.2			
n	1%	5%	10%	20%	1%	5%	10%	20%	1%	5%	10%	20%
100	.709	.769	.806	.855	.737	.800	.830	.870	.754	.813	.843	.879
200	.964	.978	.984	.991	.968	.982	.988	.993	.970	.985	.992	.994
400	.998	1.00	1.00	1.00	.998	1.00	1.00	1.00	1.00	1.00	1.00	1.00
600	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

TABLE 3. Estimated powers: Constant coefficients for the I(1) variables

	c = 0.8				c = 1.0				c = 1.2			
n	1%	5%	10%	20%	1%	5%	10%	20%	1%	5%	10%	20%
100	.658	.788	.836	.89	.732	.829	.868	.905	.774	.853	.883	.914
200	.930	.945	.955	.969	.94	.953	.963	.972	.944	.96	.966	.976
400	.965	.975	.982	.986	.967	.982	.983	.99	.973	.982	.986	.990
600	.985	.987	.989	.992	.987	.987	.99	.991	.987	.988	.991	.991

5. CONCLUSION

In this paper, we propose a consistent nonparametric test for testing the null hypothesis of constant coefficients against nonparametric smooth coefficients in a semiparametric varying coefficient cointegrating model. We show that the standardized test statistic converges to a standard normal distribution under the null hypothesis.

Although not reported here, we have also done some simulations using a residual-based bootstrap method, the results show that we can have much better estimated sizes for a wide range of smoothing parameter values. We leave the theoretical justification of the bootstrap method as well as the selection of the smoothing parameters balancing the size and the power of the test as future research topics. Finally, we only consider the case that Z_t is a stationary variable. It will be interesting to generalize the result of this paper to the case that Z_t is an I(1) variable, and we leave this as a possible future research topic.

NOTES

1. This paper only deals with the case that Z_t is a scalar for expositional simplicity. However, our results can be easily extended to the case that Z_t contains more than one stationary variable.
2. The test statistic for the case that X_t is stationary was considered by Cai, Fan, and Yao (2000) and Li et al. (2002), among others.
3. It would be desirable if our model can be extended to allow for Z_t to be a nonstationary process, then it will be a general model which covers, for example, the testing problem considered in Gao et al. (2009) as a special case. This extension is beyond the scope of the current paper and is left as a possible future research topic.
4. The "asymptotic variance" does not have the same meaning as in stationary cases as it is a positive random variable, not a constant. The square root of this term is used to scale the test statistic such that the standardized test statistic has a standard normal distribution under H_0. As this scale serves a role similar to the square root of a traditional variance, we abuse the usage of "asymptotic variance" to save creating a new name for this term.
5. Here, the reason for using a nonparametric residual is that (10) holds under both the null and alternative hypotheses.

REFERENCES

Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd ed. Wiley.

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

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

Cai, Z. & Y. Wang (2010) Instability of Predictability of Asset Returns. Working Paper, University of North Carolina at Charlotte.

Cai, Z., Y. Wang, & Y. Wang (2014) Testing instability in a predictive regression model with nonstationary regressors. *Econometric Theory*. doi:10.1017/S0266466614000590.

Choi, I. & P. Saikkonen (2004) Testing linearity in cointegrating smooth transition regressions. *Econometrics Journal* 7, 341–365.

de Jong, R.M. & J. Davidson (2000) The functional central limit theorem and weak convergence to stochastic integrals I. *Econometric Theory* 16, 621–642.

Fan, Y. & Q. Li (1999) Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics* 10, 245–271.

Gao, J. & Y. Hong (2008) Central limit theorems for generalized U-statistics with applications in nonparametric specification. *Journal of Nonparametric Statistics* 20, 61–76.

Gao, J., M. King, Z. Lu, & D. Tjøstheim (2009) Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* 25, 1869–1892.

Granger, C.W.J. (1991) Some recent generalisations of cointegration and the analysis of long-run relationships. In R.F. Engle & C.W.J. Granger (eds.), *Long-Run Economic Relationships*, pp. 277–287. Oxford University Press.

Gregory, A.W. & B.E. Hansen (1996) Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics* 70, 99–126.

Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1–16.

Hansen, B.E. (1992a) Tests for parameter instability in regressions with I(1) processes. *Journal of Business & Economic Statistics* 10, 321–335.

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

Hansen, B.E. (2008) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 1–23.

Hong, S.H. & P.C.B. Phillips (2010) Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business & Economic Statistics* 28, 96–114.

Li, Q., C.J. Huang, D. Li, & T. Fu (2002) Semiparametric smooth coefficient models. *Journal of Business & Economics Statistics* 20, 412–422.

McLeish, D.L. (1975) A maximal inequality and dependent strong laws. *The Annals of Probability* 3, 829–839.

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

Park, J.Y. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117–161.

Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.

Phillips, P.C.B. (1995) Fully modified least squares and vector autoregression. *Econometrica* 63, 1023–1078.

Phillips, P.C.B., D. Li, & J. Gao (2013) Estimating Smooth Structural Change in Cointegration Models. Discussion paper no. 1910, Cowles Foundations, Yale University.

Saikkonen, P. & I. Choi (2004) Cointegrating smooth transition regressions. *Econometric Theory* 20, 301–340.

Shin, Y.C. (1994) A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91–115.
Sun, Y., Z. Cai, & Q. Li (2013) Semiparametric functional coefficient models with integrated covariates. *Econometric Theory* 29, 659–672.
Sun, Y. & Q. Li (2011) Data-driven method selecting smoothing parameters in semiparametric models with integrated time series data. *Journal of Business & Economic Statistics* 29, 541–551.
Taylor, A.M. & M.P. Taylor (2004) The purchasing power parity debate. *Journal of Economic Perspective* 18, 135–158.
Wang, Q. (2014) Martingale limit theorem revisited and nonlinear cointegrating regression. *Econometric Thoeory* 30, 509–535.
Wang, Q. & P.C.B. Phillips (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710–738.
Wang, Q. & P.C.B. Phillips (2009b) Structural nonparametric cointegrating regression. *Econometrica* 77, 1901–1948.
Wang, Q. & P.C.B. Phillips (2012) A specification test for nonlinear nonstationary models. *The Annals of Statistics* 40, 727–758.
White, H. (2001) *Asymptotic Theory for Econometricians*, 2nd ed. Academic Press.
Wooldridge, J.M. & H. White (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4, 210–230.
Xiao, Z. (2009) Functional coefficient co-integration models. *Journal of Econometrics* 152, 81–92.
Yoshihara, K. (1976) Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie view, Gebiete* 35, 237–252.

APPENDIX A: Proofs of Main Theorems

Proof of Theorem 3.1 (i). Under H_0, the OLS residual, $\widehat{u}_t = Y_t - X_t^T\widehat{\theta}_0 = u_t - X_t^T\left(\widehat{\theta}_0 - \theta_0\right)$, and accordingly we decompose $\widehat{I}_n^a$ in (6) as

$$\widehat{I}_n^a = \frac{1}{n^3h}\sum_{t=1}^{n}\sum_{s\neq t} X_t^T X_s\left[u_t u_s + \left(\widehat{\theta}_0-\theta_0\right)^T X_t X_s^T\left(\widehat{\theta}_0-\theta_0\right) - 2u_t X_s^T\left(\widehat{\theta}_0-\theta_0\right)\right]K_{t,s}\mathbf{1}_{n,t,s}$$

$$\equiv I_{1n}^a + \left(\widehat{\theta}_0-\theta_0\right)^T G_{2n}^a\left(\widehat{\theta}_0-\theta_0\right) - 2\left(\widehat{\theta}_0-\theta_0\right)^T G_{3n}^a, \tag{A.1}$$

where

$$I_{1n}^a = \frac{2}{n^3h}\sum_{t=2}^{n}\sum_{s=1}^{t-1} X_t^T X_s u_t u_s K_{t,s}\mathbf{1}_{n,t,s}, \tag{A.2}$$

$$G_{2n}^a = \frac{2}{n^3h}\sum_{t=2}^{n}\sum_{s=1}^{t-1} X_t X_t^T X_s X_s^T K_{t,s}\mathbf{1}_{n,t,s}, \tag{A.3}$$

and

$$G_{3n}^a = \frac{1}{n^3h}\sum_{t=2}^{n}\sum_{s=1}^{t-1} X_t^T X_s\left(u_t X_s + u_s X_t\right) K_{t,s}\mathbf{1}_{n,t,s}. \tag{A.4}$$

Applying the generalized martingale central limit theorem of Wang (2014, Thm. 2.1) we show that $n\sqrt{h}I_{1n}^a/\sqrt{\sigma_{n,a}^2} \xrightarrow{d} N(0,1)$ in Lemma A.1. Also, combining Lemmas A.2 and A.3 gives

$$\sigma_{n,a}^2 = \frac{2}{n^4 h}\sum_{t=1}^{n}\sum_{s\neq t}^{n} u_t^2 u_s^2 \left(X_t^T X_s\right)^2 K_{t,s}^2 \mathbf{1}_{n,t,s} \xrightarrow{d} \sigma_a^2, \tag{A.5}$$

where $\sigma_a^2 = 4\sigma_u^4 \upsilon_2 E\left[f(Z_1)\right]\int_0^1\int_0^s \left(B_\eta(s)^T B_\eta(r)\right)^2 drds > 0$ (almost surely) is independent of a standard normal variate N, and $\upsilon_2 = \int K^2(u)\,du$. In addition, Lemma A.4 shows that $G_{2n}^a = O_p(n)$ and $G_{3n}^a = O_p(1)$. Note that under H_0, model (1) becomes a linear cointegrating model, $Y_t = X_t^T\theta_0 + u_t$, and it is well known that the OLS estimator, $\widehat{\theta}_0$, of the linear cointegrating model, gives $\widehat{\theta}_0 - \theta_0 = O_p(n^{-1})$ (e.g., Phillips, 1995). Taking these results together gives, under H_0, $n\sqrt{h}\widehat{I}_n^a/\sqrt{\sigma_{n,a}^2} = n\sqrt{h}I_{1n}^a/\sqrt{\sigma_{n,a}^2} + O_p\left(\sqrt{h}\right) = n\sqrt{h}I_{1n}^a/\sqrt{\sigma_{n,a}^2} + o_p(1)$, where $h \to 0$ as $n \to \infty$ by Assumption A6. Finally, by Slutsky's lemma, we obtain under H_0 that $J_n^a = n\sqrt{h}\widehat{I}_n^a/\sqrt{\widehat{\sigma}_{n,a}^2} = n\sqrt{h}\widehat{I}_n^a/\sqrt{\sigma_{n,a}^2} \times \sqrt{\sigma_{n,a}^2/\widehat{\sigma}_{n,a}^2} \xrightarrow{d} N(0,1)$ because $\widehat{\sigma}_{n,a}^2 = \sigma_{n,a}^2 + o_p(1)$ by Lemma A.5, where replacing u_t in $\sigma_{n,a}^2$ by the semiparametric residual for all t gives $\widehat{\sigma}_{n,a}^2$. This completes the proof of Theorem 3.1 (i). ■

Below, we present Lemmas A.1–A.5 which are used to prove Theorem 3.1 (i). In particular, Lemma A.1, a limiting result of a degenerate U-statistic with both $I(1)$ and $I(0)$ covariates, should also be useful in other contexts.

LEMMA A.1. *Under Assumptions A1–A3, A5(i), and A6, we have* $n\sqrt{h}I_{1n}^a/\sqrt{\sigma_{n,a}^2} \xrightarrow{d} N(0,1)$.

Proof. First, we have

$$\begin{aligned} I_{1n}^a &= \frac{2}{n^3 h}\sum_{t=2}^{n}\sum_{s=1}^{t-1} X_t^T X_s u_t u_s K_{t,s}\mathbf{1}_{n,t,s} \\ &= \frac{2}{n^3 h}\sum_{t=2}^{n}\sum_{s=1}^{t-1} X_{t-1}^T X_s u_t u_s K_{t,s}\mathbf{1}_{n,t,s} + \frac{2}{n^3 h}\sum_{t=2}^{n}\sum_{s=1}^{t-1} \eta_t^T X_s u_t u_s K_{t,s}\mathbf{1}_{n,t,s} \\ &\equiv \Delta_{n1} + \Delta_{n2}, \end{aligned}$$

where Δ_{n1} and Δ_{n2} denote the two terms in the same order as they appear in the second equality line. Under Assumption A1 (iii) we have $E(\Delta_{n2}) = 0$. Without loss of generality we assume that X_t is a scalar for notation simplicity. Under Assumption A1 we have for some $\delta_0 > 0$

$$\begin{aligned} E\left(\Delta_{n2}^2\right) &= \frac{4\sigma_u^4}{n^6 h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E\left[(\eta_t X_s)^2 K_{t,s}^2 \mathbf{1}_{n,t,s}\right] \\ &\le \frac{4\sigma_u^4\sigma_\upsilon^2}{n^6 h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E\left(X_s^2 K_{t,s}^2\right) \end{aligned}$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

$$= \frac{4\sigma_u^4\sigma_v^2}{n^6h^2}\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{i=1}^{s}E\left(\eta_i^2 K_{t,s}^2\right) + \frac{4\sigma_u^4\sigma_v^2}{n^6h^2}\sum_{t=4}^{n}\sum_{s=3}^{t-1}\sum_{i_1=2}^{s}\sum_{i_2=1}^{i_1-1}E\left(\eta_{i_1}\eta_{i_2}K_{t,s}^2\right)$$

$$= O\left(\frac{n^3}{n^6h}\right) + O\left(\frac{1}{n^3h^{1+\delta_0/(1+\delta_0)}}\right) = O\left(\frac{1}{n^3h^{1+\delta_0/(1+\delta_0)}}\right),$$

where applying Lemma B.1 gives $\left|h^{-1}E\left(\eta_{i_1}\eta_{i_2}K_{t,s}^2\right)\right| \le h^{-\delta_0/(1+\delta_0)}\beta_{s-i_1}^{\delta_0/(1+\delta_0)}$ for some $\delta_0 > 0$, and $\sum_{l=1}^{\infty}\beta_l^{\delta_0/(1+\delta_0)} \le C < \infty$ by Assumption A1 (ii). Hence, applying Markov's inequality gives $\Delta_{n2} = O_p\left(\left(n^3h^{1+\delta_0/(1+\delta_0)}\right)^{-1/2}\right)$. Hence, $n\sqrt{h}\Delta_{n2} = O_p\left(\left(nh^{\delta_0/(1+\delta_0)}\right)^{-1/2}\right)$ and is asymptotically ignorable under Assumption A6. ■

Next, denoting $\mathcal{X}_{n,t-1} = 2\left(n^2\sqrt{h}\right)^{-1}X_{t-1}^T\sum_{s=1}^{t-1}X_s u_s K_{t,s}\mathbf{1}_{n,t,s}$ and $S_n^2 = \sigma_u^2\sum_{t=2}^{n}\mathcal{X}_{n,t-1}^2$. Then, $n\sqrt{h}\Delta_{n1} = \sum_{t=2}^{n}u_t\mathcal{X}_{n,t-1}$. Applying Wang (2014, Thm. 2.1) we will show $\left(n\sqrt{h}\Delta_{n1}, S_n^2\right) \xrightarrow{d} \left(\sigma_a^2 N, \sigma_a^2\right)$, which requires that we verify the following results: $\left\{big(\eta_t, u_t), \mathcal{F}_{n,t}\right\}$ forms a martingale difference such that

$$\max_{2\le t\le n} E\left|\left(\eta_t^2|\mathcal{F}_{n,t-1}\right) - E\left(\eta_t^2\right)\right| = o_p(1) \quad \textbf{(A.6)}$$

$$\max_{2\le t\le n}\left|E\left(u_t^2|\mathcal{F}_{n,t-1}\right) - E\left(u_t^2\right)\right| = o_p(1)$$

$$\max_{2\le t\le n}\left|E\left[\eta_t^2 I\left(|\eta_t| \ge C_n\right)|\mathcal{F}_{n,t-1}\right] + E\left[u_t^2 I\left(|u_t| \ge C_n\right)|\mathcal{F}_{n,t-1}\right]\right| = o_p(1) \quad \textbf{(A.7)}$$

for any constant positive sequence $C_n \to \infty$,

$$\max_{2\le t\le n}\left|\mathcal{X}_{n,t-1}\right| = o_p(1) \text{ and } n^{-1/2}\sum_{t=2}^{n}\left|\mathcal{X}_{n,t-1}\right|\left|E\left(\eta_t u_t|\mathcal{F}_{n,t-1}\right)\right| = o_p(1), \quad \textbf{(A.8)}$$

and there exists an almost surely finite functional $g^2\left(B_\eta\right)$ of $B_\eta(r)$, $r \in [0,1]$, such that

$$\left(B_{n,\eta}(r), S_n^2\right) \Rightarrow \left(B_\eta(r), g^2\left(B_\eta\right)\right). \quad \textbf{(A.9)}$$

Evidently, (A.6)–(A.7) hold under Assumption A1 (iii). Also, in Lemma A.2 we have that

$$S_n^2 = \frac{4\sigma_u^4 v_2 E\left[f\left(Z_1\right)\right]}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left(\frac{X_{t-1}^T X_s}{n}\right)^2 + o_p(1)$$

$$= \frac{4\sigma_u^4 v_2 E\left[f\left(Z_1\right)\right]}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left[B_{n,\eta}\left(\frac{t-1}{n}\right)^T B_{n,\eta}\left(\frac{s}{n}\right)\right]^2 + o_p(1)$$

$$\xrightarrow{d} \sigma_a^2.$$

Thus, (A.9) holds by the continuous mapping theorem under Assumption A2 with $g^2\left(B_\eta\right) \equiv 4\sigma_u^4 v_2 E\left[f\left(Z_1\right)\right]\int_0^1\int_0^s\left(B_\eta(s)^T B_\eta(r)\right)^2 dr ds$.

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

Under Assumptions A1 and A3, Hansen (2008, Thm. 2) holds, which gives

$$\sup_{z\in\mathcal{S}_n}\left|\frac{1}{nh}\sum_{t=1}^{n}|u_t|K\left(\frac{Z_t-z}{h}\right)-E\left(|u_t|\right)f(z)\right|=O_p\left(h^2+\sqrt{\frac{\ln n}{nh}}\right).$$

It follows that

$$\max_{2\le t\le n}\left|\mathcal{X}_{n,t-1}\right|=\frac{2}{n^2\sqrt{h}}\max_{2\le t\le n}\left|X_{t-1}^T\sum_{s=1}^{t-1}X_su_sK_{t,s}\mathbf{1}_{n,t,s}\right|$$
$$=O_p\left(\frac{1}{n\sqrt{h}}\right)\max_{2\le t\le n}\sum_{s=1}^{t-1}|u_s|K_{t,s}\mathbf{1}_{n,t,s}=O_p\left(\sqrt{h}\right),$$

where we used (8) that $\max_{1\le t\le n}\|X_t\|=O_p\left(\sqrt{n}\right)$. This gives $\max_{2\le t\le n}\left|\mathcal{X}_{n,t-1}\right|=o_p(1)$ as $h\to 0$ when $n\to\infty$. In addition, the independence between $\{u_t\}_{t=1}^n$ and $\{\eta_t\}_{t=1}^n$ implied in Assumption A1 gives $n^{-1/2}\sum_{t=2}^n\left|\mathcal{X}_{n,t-1}\right|\left|E\left(\eta_tu_t|\mathcal{F}_{n,t-1}\right)\right|=0$. Hence, (A.8) holds true. As all the assumptions required by Wang (2014, Thm. 2.1) hold, we obtain $n\sqrt{h}\Delta_{n1}/\sqrt{S_n^2}\xrightarrow{d}N(0,1)$. Combining this result with Lemma A.3, we obtain $n\sqrt{h}I_{1n}^a/\sqrt{\sigma_{n,a}^2}\xrightarrow{d}N(0,1)$. This completes the proof of Lemma A.1.

Remark 4. Gao and Hong (2008) derived a central limit theorem for a generalized U-statistic of the form of $\sum_{t=1}^n\sum_{s\neq t}^n\psi_n(X_s,X_t)\phi_1(\eta_s,\eta_t)$, where $\psi_n(X_s,X_t)=\sum_{i=1}^{\min(s-1,t-1)}A_{n,i}\phi_2\left(\eta_{s-i},\eta_{t-i}\right)$ is a linear combination of a function of an r-dimensional strictly stationary β-mixing process, $X_t=\left(\eta_{t-1},\ldots,\eta_1\right)$, and $\phi_1(\cdot,\cdot)$ and $\phi_2(\cdot,\cdot)$ are both symmetric functions. The central limit theorem in Gao and Hong (2008) assumes $\psi_n(X_s,X_t)$ to be weakly dependent, while Lemma A.1 considers the case that $\psi_n(X_s,X_t)$ is nonstationary. Therefore, Lemma A.1 can be considered as an extension of Gao and Hong's (2008) result to for integrated time series data.

LEMMA A.2. *Under the assumptions given in Lemma A.1, we have* $S_n^2\xrightarrow{d}\sigma_a^2$.

Proof. First, we have

$$S_n^2=\sigma_u^2\sum_{t=2}^{n}\mathcal{X}_{n,t-1}^2=\frac{4\sigma_u^2}{n^4h}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left(X_{t-1}^TX_s\right)^2u_s^2K_{t,s}^2\mathbf{1}_{n,t,s}$$
$$+\frac{8\sigma_u^2}{n^4h}\sum_{t=3}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}X_{t-1}^TX_{s_1}X_{s_2}^TX_{t-1}u_{s_1}u_{s_2}K_{t,s_1}K_{t,s_2}\mathbf{1}_{n,t,s_1}\mathbf{1}_{n,t,s_2}$$
$$\equiv A_{n1}+A_{n2}. \tag{A.10}$$

Letting $e_{t,s}=\left[u_s^2K_{t,s}^2\mathbf{1}_{n,t,s}-E\left(u_s^2K_{t,s}^2\mathbf{1}_{n,t,s}\right)\right]/h$, we decompose A_{n1} into two terms

$$A_{n1}=\frac{4\sigma_u^2}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left(\frac{X_{t-1}^TX_s}{n}\right)^2E\left(\frac{u_s^2K_{t,s}^2\mathbf{1}_{n,t,s}}{h}\right)+\frac{4\sigma_u^2}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left(\frac{X_{t-1}^TX_s}{n}\right)^2e_{t,s}$$
$$\equiv\Delta_{n1}+\Delta_{n2}.$$

Using Lemma B.1 we have that

$$h^{-1}E\left(u_s^2 K_{t,s}^2 \mathbf{1}_{n,t,s}\right) = \upsilon_2\sigma_u^2 E\left[\mathbf{1}\left(Z_1 \in \mathcal{S}_n\right) f\left(Z_1\right)\right] + O\left(h\right) + O\left(h^{-\delta_0/(1+\delta_0)}\beta_{|t-s|}^{\delta_0/(1+\delta_0)}\right)$$

under Assumption A1 (iii). Then by (8), we have $n^{-4}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left(X_{t-1}^T X_s\right)^2 \left(h + h^{-\delta_0/\left(1+\delta_0\right)}\beta_{|t-s|}^{\delta_0/\left(1+\delta_0\right)}\right) = O_p(h) + O_p\left(n^{-1/(1+\delta_0)}(nh)^{-\delta_0/(1+\delta_0)}\right) = o_p(1)$ by Assumptions A1 (ii) and A6. Therefore, applying the continuous mapping theorem and (7), we obtain

$$\Delta_{n1} = \frac{4\upsilon_2\sigma_u^4 E\left[f\left(Z_1\right)\right]}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left(\frac{X_t^T X_s}{n}\right)^2 + o_p(1) \xrightarrow{d} \sigma_a^2. \tag{A.11}$$

Now, we show that $\Delta_{n2} = o_p(1)$. Let $M_{n,t} = n^{-1}X_t^T \otimes X_t^T$ and $V_{n,s} = vec\left(n^{-1}X_s X_s^T\right)$, where "$\otimes$ " denotes the Kronecker product, and $vec\,(A)$ is an $(nk)\times 1$ vector formed by stacking up the columns of an $n\times k$ matrix A. Then, $n^{-2}X_t^T X_s X_s^T X_t = M_{n,t}V_{n,s}$. Denote $M_n(r) \equiv M_{n,[nr]}$ and $V_n(r) \equiv V_{n,[nr]}$ for any $r \in [0,1]$. By Assumption A2 and the continuous mapping theorem, we have $M_n \Rightarrow M \equiv B_\eta^T \otimes B_\eta^T$ and $V_n \Rightarrow V \equiv vec\left(B_\eta B_\eta^T\right)$. For any small $\epsilon \in (0,1)$, setting $N = [1/\epsilon]$, $s_k = [kn/N]+1$, $s_k^* = s_{k+1}-1$, $N_t^* = [(N-1)(t-1)/n]$, and $s_k^{**} = \min\left(s_k^*, t-1\right)$, we have

$$\begin{aligned}
|\Delta_{n2}| &= \left|n^{-2}\sum_{t=2}^{n}M_{n,t}\sum_{s=1}^{t-1}V_{n,s}e_{t,s}\right| \le \sup_{0\le r\le 1}\|M_n(r)\|\, n^{-2}\sum_{t=2}^{n}\left\|\sum_{s=1}^{t-1}V_{n,s}e_{t,s}\right\| \\
&= \sup_{0\le r\le 1}\|M_n(r)\|\, n^{-2}\sum_{t=2}^{n}\left\|\sum_{k=0}^{N_t^*}\sum_{s=s_k}^{s_k^{**}}V_{n,s}e_{t,s}\right\| \\
&= \sup_{0\le r\le 1}\|M_n(r)\|\, n^{-2}\sum_{t=2}^{n}\left\|\sum_{k=0}^{N_t^*}V_{n,s_k}\sum_{s=s_k}^{s_k^{**}}e_{t,s}\right\| \\
&\quad + \sup_{0\le r\le 1}\|M_n(r)\|\, n^{-2}\sum_{t=2}^{n}\left\|\sum_{k=0}^{N_t^*}\sum_{s=s_k}^{s_k^{**}}\left(V_{n,s}-V_{n,s_k}\right)e_{t,s}\right\| \\
&\le \sup_{0\le r\le 1}\|M_n(r)\| \sup_{0\le r\le 1}\|V_n(r)\|\, n^{-2}\sum_{t=2}^{n}\sum_{k=0}^{N_t^*}\left|\sum_{s=s_k}^{s_k^{*}}e_{t,s}\right| \\
&\quad + \sup_{0\le r\le 1}\|M_n(r)\| \sup_{|r-r'|\le\epsilon}\left\|V_n(r)-V_n\left(r'\right)\right\| n^{-2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|e_{t,s}\right|.
\end{aligned} \tag{A.12}$$

Since M_n and V_n converge to well defined $O_p(1)$ limiting processes under the Skorohod topology, we have $\sup_{0\le r\le 1}\|M_n(r)\| = O_p(1)$ and $\sup_{0\le r\le 1}\|V_n(r)\| = O_p(1)$. In addition, as $n\to\infty$, we have

$$\sup_{|r-r'|\le\epsilon}\left\|V_n(r)-V_n\left(r'\right)\right\| \xrightarrow{d} \sup_{|r-r'|\le\epsilon}\left\|V(r)-V\left(r'\right)\right\| \xrightarrow{p} 0$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

as $\epsilon \to 0$. With $\left|e_{t,s}\right| = O_p(1)$, the second term in (A.12) is $o_p(1)$. Further, we have

$$\frac{1}{n^2}\sum_{t=2}^{n}\sum_{k=0}^{N_t^*} E\left[\left|\sum_{s=s_k}^{s_k^{**}} e_{t,s}\right|\right] \le \frac{1}{n}\sum_{t=2}^{n}\sup_{s+n\epsilon<t} E\left|\frac{1}{n\epsilon}\sum_{i=s}^{s+\epsilon n} e_{t,i}\right| \le C(\epsilon nh)^{-1/2} = o(1),$$

if we set ϵ to be a small positive constant such that $\epsilon nh \to \infty$ as $n \to \infty$. Hence, the first term in (A.12) is also $o_p(1)$. Therefore, we obtain $\Delta_{n2} = o_p(1)$.

Next, we show that $A_{n2} = o_p(1)$. Without loss of generality, we give the proof for a scalar X_t. By Assumption A1 (iii) we have $E(A_{n2}) = 0$ and

$$\begin{aligned} E\left(A_{n2}^2\right) &= \frac{64\sigma_u^8}{n^8h^2}\sum_{t=3}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1} E\left(X_{t-1}^4 X_{s_1}^2 X_{s_2}^2 K_{t,s_1}^2 K_{t,s_2}^2 \mathbf{1}_{n,t,s_1}\mathbf{1}_{n,t,s_2}\right) \\ &\quad + \frac{128\sigma_u^8}{n^8h^2}\sum_{t=4}^{n}\sum_{t'=3}^{t-1}\sum_{s_1=2}^{t'-1}\sum_{s_1=1}^{s_1-1} E\Big(X_{t-1}^2 X_{s_1}^2 X_{s_2}^2 X_{t'-1}^2 K_{t,s_1} \\ &\qquad K_{t,s_2}\mathbf{1}_{n,t,s_1}\mathbf{1}_{n,t,s_2} K_{t',s_1} K_{t',s_2}\mathbf{1}_{n,t',s_1}\mathbf{1}_{n,t',s_2}\Big) \\ &\equiv 64\sigma_u^8(\chi_{n1} + 2\chi_{n2}), \end{aligned} \tag{A.13}$$

where the definition of χ_{n1} and χ_{n2} will be clear from the context below. Consider χ_{n1} first. As $X_t = \sum_{i=1}^{t}\eta_i$, we have $X_{t-1}^4 X_{s_1}^2 X_{s_2}^2 = \sum_{i_1\le t-1}\sum_{i_2\le t-1}\sum_{i_3\le t-1}\sum_{i_4\le t-1}\sum_{i_5\le s_1}\sum_{i_6\le s_1}\sum_{i_7\le s_2}\sum_{i_8\le s_2}\eta_{i_1}\cdots\eta_{i_8}$. Hence, there are totally 11 summations in χ_{n1} over subindexes: t, s_1, s_2, $i_1,\ldots,i_8$. Letting j be the total number of different subindexes, we have $\chi_{n1} \equiv D_{n,1} + \ldots + D_{n,8}$, where

$$D_{n,j} = \frac{1}{n^8h^2}\sum_{t=3}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}\sum_{i_1}\cdots\sum_{i_j} E\left(\eta_{i_1}^{l_1}\ldots\eta_{i_j}^{l_j} K_{t,s_1}^2 K_{t,s_2}^2 \mathbf{1}_{n,t,s_1}\mathbf{1}_{n,t,s_2}\right)$$

sums over $j+3$ different subindexes and $\sum_{s=1}^{j} l_s = 8$. Applying Lemma B.1 to $j \le 4$, it is readily seen that $D_{n,1} = O\left(n^{-4}\right)$, $D_{n,2} = O\left(n^{-3}\right)$, $D_{n,3} = O\left(n^{-2}\right)$, and $D_{n,4} = O\left(n^{-1}\right)$, where $E\left(|\eta_t|^q\right) \le C < \infty$ for some $q > 8$. When $j \ge 5$, there are more than eight different subindexes. Letting $m_n = [C_0 \ln n]$ for some positive constant C_0, we will repeatedly use Lemma B.1 and a *summation splitting* method to obtain the order for $D_{n,5}$ to $D_{n,8}$.

(i) For $j = 5$, $D_{n,5}$ contains 8 summations, where $1 \le l_s \le 4$ for $s = 1,\ldots,5$, $\sum_{s=1}^{5} l_s = 8$, and there are at least two l 's equal to one. As an illustration, we take the case that $l_5 = l_4 = 1$ and $t > i_1 > s_1 > i_2 > i_3 > s_2 > i_4 > i_5$, and apply Lemma B.1 to obtain

$$\left|h^{-2} E\left(\eta_{i_5}\eta_{i_4}\eta_{i_3}^{l_3}\eta_{i_2}^{l_2}\eta_{i_1}^{l_1} K_{t,s_1}^2 K_{t,s_2}^2 \mathbf{1}_{n,t,s_1}\mathbf{1}_{n,t,s_2}\right)\right| \le \frac{C}{h^{2\delta_0/(1+\delta_0)}}\begin{cases}\beta_{m_n}^{\delta_0/(1+\delta_0)}, & \text{if } s_2 - i_4 > m_n \\ \beta_{i_4-i_5}^{\delta_0/(1+\delta_0)}, & \text{if } s_2 - i_4 \le m_n\end{cases}.$$

Applying this method to other combinations of summations, we obtain

$$E\left(D_{n,5}\right) = O\left(\frac{n^8\beta_{m_n}^{\delta_0/(1+\delta_0)}}{n^8h^{2\delta_0/(1+\delta_0)}}\right) + O\left(\frac{n^6 m_n}{n^8h^{2\delta_0/(1+\delta_0)}}\right).$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

(ii) For $j = 6$, $D_{n,6}$ contains 9 summations, where $1 \le l_s \le 3$ for $s = 1, \ldots, 6$, $\sum_{s=1}^{6} l_s = 8$, and there are at least four l 's equal to one. We take a case that $l_6 = l_5 = l_4 = l_3 = 1$ and $t > i_1 > i_2 > s_1 > i_3 > i_4 > s_2 > i_5 > i_6$ and apply Lemma B.1 to obtain

$$\left| h^{-2} E\left(\eta_{i_6}\eta_{i_5}\eta_{i_4}\eta_{i_3}\eta_{i_2}^{l_2}\eta_{i_1}^{l_1} K_{t,s_1}^2 K_{t,s_2}^2 \mathbf{1}_{n,t,s_1}\mathbf{1}_{n,t,s_2}\right)\right| \le \frac{C}{h^{2\delta_0/(1+\delta_0)}} \begin{cases} \beta_{m_n}^{\delta_0/(1+\delta_0)}, & \text{if } i_3 - i_4 > m_n \\ \beta_{m_n}^{\delta_0/(1+\delta_0)}, & \text{if } i_3 - i_4 \le m_n \text{ and } s_2 - i_5 > m_n \\ \beta_{i_5 - i_6}^{\delta_0/(1+\delta_0)}, & \text{if } i_3 - i_4 \le m_n \text{ and } s_2 - i_5 \le m_n \end{cases}.$$

Applying this method to other combinations of summations, we obtain

$$E\left(D_{n,6}\right) = O\left(\frac{n^9 \beta_{m_n}^{\delta_0/(1+\delta_0)}}{n^8 h^{2\delta_0/(1+\delta_0)}}\right) + O\left(\frac{n^6 m_n^2}{n^8 h^{2\delta_0/(1+\delta_0)}}\right).$$

(iii) Applying the same method to the cases with $j = 7$ and $j = 8$ we obtain

$$E\left(D_{n,7}\right) = O\left(\frac{n^{10} \beta_{m_n}^{\delta_0/(1+\delta_0)}}{n^8 h^{2\delta_0/(1+\delta_0)}}\right) + O\left(\frac{n^6 m_n^3}{n^8 h^{2/(1+\delta_0)}}\right),$$

$$E\left(D_{n,8}\right) = O\left(\frac{n^{11} \beta_{m_n}^{\delta_0/(1+\delta_0)}}{n^8 h^{2\delta_0/(1+\delta_0)}}\right) + O\left(\frac{n^6 m_n^4}{n^8 h^{2\delta_0/(1+\delta_0)}}\right).$$

Therefore, we obtain

$$\chi_{n1} = O\left(\frac{1}{n}\left(1 + \frac{n^4 \beta_{m_n}^{\delta_0/(1+\delta_0)}}{h^{2\delta_0/(1+\delta_0)}} + \frac{m_n^4}{n h^{2\delta_0/(1+\delta_0)}}\right)\right) = O\left(\frac{1}{n}\right) \quad \textbf{(A.14)}$$

if $C_0 > \left[4 + 2\delta_0 (2+\alpha)\right] / (\delta_0 \ln \rho)$ for some $\delta_0 \in (0,1)$ as Assumption A6 implies $h \sim n^{-\alpha}$ for some $\alpha \in (0,1)$. Similarly, we can show

$$\chi_{n2} = O\left(h\left(1 + \frac{n^4 \beta_{m_n}^{\delta_0/(1+\delta_0)}}{h^{3\delta_0/(1+\delta_0)}} + \frac{m_n^4}{n h^{3\delta_0/(1+\delta_0)}}\right)\right) = O\left(h\right) \quad \textbf{(A.15)}$$

if $C_0 > \left[5 + \delta_0 (5 + 3\alpha)\right] / (\delta_0 \ln \rho)$ for some $\delta_0 \in (0, 1/2)$. Hence, we obtain $A_{n2} = O_p\left(n^{-1/2} + \sqrt{h}\right)$ by Markov's inequality. Combining this result with (A.11) gives $S_n^2 \xrightarrow{d} \sigma_a^2$. This completes the proof of Lemma A.2. ■

LEMMA A.3. *Under the assumptions given in Lemma A.1, we have* $\sigma_{n,a}^2 = S_n^2 + o_p(1)$.

Proof. Simple calculations lead to

$$\sigma_{n,a}^2 - S_n^2 = \frac{4}{n^4 h}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left(u_t^2 - \sigma_u^2\right) u_s^2 \left(X_t^T X_s\right)^2 K_{t,s}^2 \mathbf{1}_{n,t,s}.$$

By Assumption A1 (iii) we have $E\left(\sigma_{n,a}^2 - S_n^2\right) = 0$ and

$$E\left[\left(\sigma_{n,a}^2 - S_n^2\right)^2\right] \leq \frac{C}{n^8h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E\left[\left(X_t^T X_s\right)^4 K_{t,s}^4 \mathbf{1}_{n,t,s}\right]$$
$$+\frac{C}{n^8h^2}\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{s'=1}^{s-1} E\left[\left(X_t^T X_s X_{s'}^T X_t\right)^2 K_{t,s}^2 K_{t,s'}^2 \mathbf{1}_{n,t,s}\mathbf{1}_{n,t,s'}\right]$$
$$= O\left(\frac{1}{n^2h}\left(1+\frac{n^4\beta_{m_n}^{\delta_0/(1+\delta_0)}}{h^{\delta_0/(1+\delta_0)}}+\frac{m_n^4}{nh^{\delta_0/(1+\delta_0)}}\right)\right)$$
$$+O\left(\frac{1}{n}\left(1+\frac{n^4\beta_{m_n}^{\delta_0/(1+\delta_0)}}{h^{2\delta_0/(1+\delta_0)}}+\frac{m_n^4}{nh^{2\delta_0/(1+\delta_0)}}\right)\right)$$
$$= O\left(n^{-2}h^{-1}\right) + O\left(n^{-1}\right),$$

where we use the same proof method as used in the proof of Lemma A.2, and we apply Assumption A1 (ii) to obtain the last line for properly chosen $\delta_0 \in (0,1)$ and $C_0 > 0$. Therefore, we obtain $\sigma_{n,a}^2 - S_n^2 = O_p\left(n^{-1/2}\right)$ by Markov's inequality as $nh \to \infty$ when $n \to \infty$. This completes the proof of this lemma. ■

LEMMA A.4. *Under the assumptions given in Theorem 3.1, we obtain $G_{2n}^a = O_e(n)$ and $G_{3n}^a = O_p(1)$, where G_{2n}^a and G_{3n}^a are defined in (A.3) and (A.4), respectively.*

Proof. Using exactly the same arguments as those used in the proof of Lemma A.2, we obtain

$$n^{-1}G_{2n}^a = \frac{2}{n}\sum_{t=2}^{n}\frac{X_t^T}{\sqrt{n}}\frac{1}{n}\sum_{s=1}^{t-1}\frac{X_sX_s^T}{n}\;\frac{X_t}{\sqrt{n}}\frac{E\left(K_{t,s}\mathbf{1}_{n,t,s}\right)}{h}$$
$$+o_p(1) \xrightarrow{d} 2E[f(Z_1)]\int_0^1\int_0^s (B_\eta(s)^T B_\eta(r))^2 dr ds = O_e(1).$$

Hence, $G_{2n}^a = O_e(n)$. Next, we write $G_{3n}^a = G_{3n,1}^a + G_{3n,2}^a$, where

$$G_{3n,1}^a = \left(n^3h\right)^{-1}\sum_{t=2}^{n} u_t \sum_{s=1}^{t-1} X_t^T X_s X_s K_{t,s}\mathbf{1}_{n,t,s}$$

and $G_{3n,2}^a = \left(n^3h\right)^{-1}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n} X_t^T X_s u_s X_t K_{t,s}\mathbf{1}_{n,t,s}$. By Assumption A1 (iii) we have $EG_{3n,j}^a = 0$ for $j = 1,2$, and applying the same proof method used in the proof of Lemma A.2 gives

$$E\left[G_{3n,1}^a\left(G_{3n,1}^a\right)^T\right] = \frac{\sigma_u^2}{n^6h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E\left(X_t^T X_s X_s X_s^T X_s^T X_t K_{t,s}^2\mathbf{1}_{n,t,s}\right)$$
$$+\frac{\sigma_u^2}{n^6h^2}\sum_{t=3}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1} E\left(X_t^T X_{s_1} X_{s_1} X_{s_2}^T X_{s_2}^T X_t K_{t,s_1} K_{t,s_2}\mathbf{1}_{n,t,s_1}\mathbf{1}_{n,t,s_2}\right)$$
$$= O\left(n^{-1}h^{-1}\right) + O(1) = O(1).$$

Then, applying Markov's inequality gives $G_{3n,1}^a = O_p(1)$. Similarly, we can show $G_{3n,2}^a = O_p(1)$. This completes the proof of Lemma A.4. ∎

LEMMA A.5. *Under the assumptions given in Theorem 3.1 (i), we have* $\widehat{\sigma}_{n,a}^2 - \sigma_{n,a}^2 = o_p(1)$, *where* $\widehat{\sigma}_{n,a}^2$ *is defined by (10).*

Proof. Replacing u_t in $\sigma_{n,a}^2$ by $\tilde{u}_t = Y_t - X_t^T \widehat{\theta}^{(-t)}(Z_t) = u_t - X_t^T \left[\widehat{\theta}^{(-t)}(Z_t) - \theta(Z_t)\right]$ gives $\widehat{\sigma}_{n,a}^2$. Applying Lemma C.1 we verify Lemma A.5. ∎

Remark 5. Here we emphasize that it is important to use the nonparametric residuals in computing $\widehat{\sigma}_{na}^2$. If the nonparametric residual $\tilde{u}_t$ is replaced by the parametric residual $\widehat{u}_t = Y_t - X_t^T \widehat{\theta}_0 = u_t + X_t^T \left[\widehat{\theta}_0 - \theta(Z_t)\right]$, then under H_1, $\widehat{u}_t = O_p(\sqrt{n})$ and Lemma A.5 does not hold; the resulting test may have only trivial power even as $n \to \infty$. The same argument also applies to Theorem 3.2.

Proof of Theorem 3.1 (ii). Under H_1, we express $\widehat{\theta}_0$ as $\widehat{\theta}_0 = \widehat{\theta}_0 - E[\theta(Z_t)] + E[\theta(Z_t)]$, where

$$\widehat{\theta}_0 - E[\theta(Z_t)] = \left(\sum_t X_t X_t^T\right)^{-1} \sum_t X_t X_t^T e_t + \left(\sum_t X_t X_t^T\right)^{-1} \sum_t X_t u_t \quad \textbf{(A.16)}$$

with $e_t = \theta(Z_t) - E\left[\theta(Z_t)\right]$. By Assumption A1 (ii), White (2001, Thm. 3.49), McLeish (1975, Lem. 2.1) and the fact that a β-mixing sequence is also an α -mixing sequence, we have as $m \to \infty$,

$$\sup_t E\left|E_{t-m}(e_t)\right| \le 6\beta_m^{1-1/q} \|e_t\|_q = o(1) \text{ for some } q > 1, \quad \textbf{(A.17)}$$

which implies that $\max_{1\le t\le n} \left\| n^{-2} \sum_{s=1}^t X_s X_s^T e_s \right\| = o_p(1)$ by Hansen (1992b, Thm. 3.3). Hence, we have

$$\widehat{\theta}_0 - E\left[\theta(Z_t)\right] = o_p(1) + O_p\left(n^{-1}\right) = o_p(1), \quad \textbf{(A.18)}$$

which means that the OLS estimator converges to the mean value of the random coefficient under the alternative hypothesis. Since $\widehat{u}_t = Y_t - X_t^T \widehat{\theta}_0 = u_t - X_t^T \left(\widehat{\theta}_0 - \theta(Z_t)\right)$, simple calculations lead to

$$\begin{aligned}\widehat{I}_n^a &= \frac{1}{n^3 h} \sum_{t=1}^n \sum_{s\ne t} X_t^T X_s \left[u_t u_s + \left(\widehat{\theta}_0 - \theta(Z_t)\right)^T X_t X_s^T \left(\widehat{\theta}_0 - \theta(Z_s)\right)\right. \\ &\qquad \left. -2u_t X_s^T \left(\widehat{\theta}_0 - \theta(Z_s)\right)\right] K_{t,s} \mathbf{1}_{n,t,s} \\ &\equiv I_{1n}^a + I_{2n}^a - 2I_{3n}^a, \end{aligned} \quad \textbf{(A.19)}$$

where the definitions of I_{jn}^a $(j = 1,2,3)$ should be apparent. As I_{1n}^a is the same as that defined under H_0, we obtain $I_{1n}^a = O_p\left(n^{-1}h^{-1/2}\right)$ by Lemma A.1. Next, we consider

$$n^{-1} I_{2n}^a \equiv \frac{2}{n^4 h} \sum_{t=2}^n \sum_{s=1}^{t-1} X_t^T X_s e_t^T X_t X_s^T e_s K_{t,s} \mathbf{1}_{n,t,s}$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

$$+n^{-1}\left\{\widehat{\theta}_0 - E\left[\theta\left(Z_t\right)\right]\right\}^T G_{2n}^a \left\{\widehat{\theta}_0 - E\left[\theta\left(Z_t\right)\right]\right\}$$
$$+\left\{\widehat{\theta}_0 - E\left[\theta\left(Z_t\right)\right]\right\}^T \frac{2}{n^4 h} \sum_{t=1}^{n} \sum_{s \neq t} X_t^T X_s e_t^T X_t X_s K_{t,s} \mathbf{1}_{n,t,s}, \qquad \textbf{(A.20)}$$

where the second term equals $o_p(1)$ by (A.18) and $G_{2n}^a = O_p(n)$ by Lemma A.4. Applying $vec(ABC) = \left(C^T \otimes A\right) vec(B)$ and Lemma B.2, we have

$$\frac{1}{n}\sum_{t=2}^{n}\frac{1}{n}\sum_{s=1}^{t-1}\frac{X_t^T X_s}{n} e_t^T \frac{X_t X_s^T}{n} e_s \frac{K_{t,s}\mathbf{1}_{n,t,s}}{h}$$
$$= \frac{1}{n}\sum_{t=2}^{n}\frac{1}{n}\sum_{s=1}^{t-1}\frac{X_t^T X_s}{n}\left(\frac{X_s^T \otimes X_t^T}{n}\right) vec\left(e_t e_s^T\right)\frac{K_{t,s}\mathbf{1}_{n,t,s}}{h}$$
$$= \frac{1}{n}\sum_{t=2}^{n}\frac{1}{n}\sum_{s=1}^{t-1}\frac{X_t^T X_s}{n}\left(\frac{X_s^T \otimes X_t^T}{n}\right) E\left[vec\left(e_t e_s^T\right)\frac{K_{t,s}\mathbf{1}_{n,t,s}}{h}\right] + o_p(1)$$
$$\xrightarrow{d} \int_0^1 \int_0^r B_\eta(r)^T B_\eta(s) B_\eta(r)^T E\left[e_1 e_1^T f(Z_1)\right] B_\eta(s)\, ds\, dr, \qquad \textbf{(A.21)}$$

which is an almost surely positive random variable. Similarly, we have

$$n^{-4}h^{-1}\sum_{t=1}^{n}\sum_{s\neq t} X_t^T X_s e_t^T X_t X_s K_{t,s}\mathbf{1}_{n,t,s} = O_p(1).$$

Combining the above result with (A.18), one can easily see that the third term in (A.20) is of order $o_p(1)$. Hence, we obtain $I_{2n}^a = O_e(n)$.

Finally, we consider $I_{3n}^a \equiv \left\{\widehat{\theta}_0 - E\left[\theta\left(Z_1\right)\right]\right\}^T G_{3n}^a - I_{3n,2}^a$, where G_{3n}^a is defined by (A.4) and $I_{3n,2}^a \equiv \left(n^3 h\right)^{-1}\sum_{t=1}^{n}\sum_{s\neq t} X_t^T X_s u_t X_s^T e_s K_{t,s}\mathbf{1}_{n,t,s}$. Following the proof of Lemma A.4 we can show that $I_{3n,2}^a = O_p(1)$. As $G_{3n}^a = O_p(1)$ by Lemma A.4. Also, by (A.18), we have $I_{3n}^a = O_p(1)$. Therefore, under H_1, $I_{2n}^a = O_e(n)$ is the leading term of $\widehat{I}_n^a$. Consequently, $n\sqrt{h}\widehat{I}_n^a = n\sqrt{h}O_e(n)$ diverges to $+\infty$ at the rate of $n^2\sqrt{h}$. Combining this result with Lemmas A.2 and A.5 completes the proof of Theorem 3.1 (ii). ■

Proof of Theorem 3.2. Under H_0, we have $\widehat{u}_t = Y_t - X_t^T\widehat{\theta}_0 = u_t - X_t^T\left(\widehat{\theta}_0 - \theta_0\right)$. Then $\widehat{I}_n^b$ has the same decomposition as $\widehat{I}_n^a$ given by (A.1), viz.

$$\widehat{I}_n^b = I_{1n}^b + (\widehat{\theta}_0 - \theta_0)^T G_{2n}^b (\widehat{\theta}_0 - \theta_0) - 2(\widehat{\theta}_0 - \theta_0)^T G_{3n}^b,$$

where I_{1n}^b, G_{2n}^b, and G_{3n}^b are defined the same as in (A.2), (A.3), and (A.4), respectively. Lemma A.6 below shows that, under H_0, $n\sqrt{h}I_{1n}^b/\sqrt{\sigma_{n,b}^2} \xrightarrow{d} N(0,1)$, where $\sigma_{n,b}^2$ and σ_b^2 have exactly the same mathematical formula as $\sigma_{n,a}^2$ and σ_a^2, respectively. Lemma A.7 below shows that $(\widehat{\theta}_0 - \theta)^T G_{2n}^b (\widehat{\theta}_0 - \theta) = O_p\left(n^{-1}\right)$ and $(\widehat{\theta}_0 - \theta)^T G_{3n}^b = O_p\left(n^{-1}\right)$. Hence,

$$n\sqrt{h}\widehat{I}_n^b/\sqrt{\sigma_{n,b}^2} = n\sqrt{h}I_{1n}^b/\sqrt{\sigma_{n,b}^2} + O_p(\sqrt{h}) \xrightarrow{d} N(0,1).$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

As $\widehat{\sigma}_{n,b}^2 = \sigma_{n,b}^2 + o_p(1)$ by Lemma A.8 below, we have $n\sqrt{h}\widehat{I}_n^b/\sqrt{\widehat{\sigma}_{n,b}^2} \xrightarrow{d} N(0,1)$ by Slutsky's lemma. In addition, $\sigma_{n,b}^2 = \sigma_b^2 + o_p(1)$ by Lemmas B.2, A.2, and A.3. This completes the proof of Theorem 3.2. ∎

The following lemma gives the asymptotic distribution of a degenerate U-statistic when X_t contains both I(0) and I(1) variables.

LEMMA A.6. *Under Assumptions B1, A2, B3, A5 (i), and A6, we obtain* $n\sqrt{h}I_{1n}^b/\sqrt{\sigma_{n,b}^2} \xrightarrow{d} N(0,1)$.

Proof. A simple calculation gives

$$\begin{aligned} I_{1n}^b &= \frac{2}{n^3 h}\sum_{t=2}^{n}\sum_{s=1}^{t-1} X_{1,t}^T X_{1,s} u_t u_s K_{t,s}\mathbf{1}_{n,t,s} + \frac{2}{n^3 h}\sum_{t=2}^{n}\sum_{s=1}^{t-1} X_{2,t}^T X_{2,s} u_t u_s K_{t,s}\mathbf{1}_{n,t,s} \\ &\equiv I_{1n,1}^b + I_{1n,2}^b, \end{aligned}$$

where $E\left(I_{1n,j}^b\right) = 0$ for $j = 1,2$ by Assumption A1 (iii). Applying Lemma B.1, one can show that $Var\left(I_{1n,1}^b\right) = O\left(n^{-4}h^{-1}\right)$. Hence, $n\sqrt{h}I_{1n,1}^b = n\sqrt{h}O_p\left(n^{-2}h^{-1/2}\right) = O_p\left(n^{-1}\right) = o_p(1)$. Following closely the proof of Lemma A.1, we have $n\sqrt{h}I_{1n,2}^b/\sqrt{\sigma_{n,b,2}^2} \xrightarrow{d} N(0,1)$ under Assumptions B1, A2, B3, A5 (i), and A6, where $\sigma_{n,b,2}^2 = 2\left(n^4 h\right)^{-1}\sum_{t=1}^n\sum_{s\neq t}^n u_t^2 u_s^2 \left(X_{2,t}^T X_{2,s}\right)^2 K_{t,s}^2 \mathbf{1}_{n,t,s}$. Note that

$$\begin{aligned} \sigma_{n,b}^2 &= \frac{2}{n^4 h}\sum_{t=1}^{n}\sum_{s\neq t}^{n} u_t^2 u_s^2 \left(X_t^T X_s\right)^2 K^2\left(\frac{Z_t - Z_s}{h}\right)\mathbf{1}_{n,t,s} \\ &= \frac{2}{n^4 h}\sum_{t=1}^{n}\sum_{s\neq t}^{n} u_t^2 u_s^2 \left[\left(X_{1,t}^T X_{1,s}\right)^2 + 2X_{1,t}^T X_{1,s} X_{2,t}^T X_{2,s} + \left(X_{2,t}^T X_{2,s}\right)^2\right] \\ &\quad \times K^2\left(\frac{Z_t - Z_s}{h}\right)\mathbf{1}_{n,t,s} \\ &= \sigma_{n,b,2}^2 + O_p\left(n^{-1}\right) \end{aligned}$$

by Lemma B.2. This completes the proof of this lemma. ∎

LEMMA A.7. *Under the assumptions given in Lemma A.6, we have*

$$\left(\widehat{\theta}_0 - \theta_0\right)^T G_{2n}^b \left(\widehat{\theta}_0 - \theta_0\right) = O_p\left(n^{-1}\right),$$

and

$$\left(\widehat{\theta}_0 - \theta_0\right)^T G_{3n}^b = O_p\left(n^{-1}\right),$$

where G_{2n}^b and G_{3n}^b are defined in (A.3) and (A.4), respectively.

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

Proof. A simple calculation gives

$$
\begin{aligned}
G_{2n}^{b} &= \frac{1}{n^3 h}\sum_{t=2}^{n}\sum_{s=1}^{t-1} X_{1,t}^{T} X_{1,s}\begin{pmatrix} X_{1,t}X_{1,s}^{T} & X_{1,t}X_{2,s}^{T} \\ X_{2,t}X_{1,s}^{T} & X_{2,t}X_{2,s}^{T}\end{pmatrix} K_{t,s}\mathbf{1}_{n,t,s} \\
&\quad + \frac{1}{n^3 h}\sum_{t=2}^{n}\sum_{s=1}^{t-1} X_{2,t}^{T} X_{2,s}\begin{pmatrix} X_{1,t}X_{1,s}^{T} & X_{1,t}X_{2,s}^{T} \\ X_{2,t}X_{1,s}^{T} & X_{2,t}X_{2,s}^{T}\end{pmatrix} K_{t,s}\mathbf{1}_{n,t,s} \\
&= \begin{pmatrix} O_P\left(n^{-1}\right) & O_P\left(n^{-1/2}\right) \\ O_P\left(n^{-1/2}\right) & O_P(1)\end{pmatrix} + \begin{pmatrix} O_P(1) & O_P\left(\sqrt{n}\right) \\ O_P\left(\sqrt{n}\right) & O_P(n)\end{pmatrix}, \qquad \textbf{(A.22)}
\end{aligned}
$$

where the last line can be obtained by following the proof of Lemma B.2. As the model becomes $Y_t = X_{1,t}^{T}\theta_{10} + X_{2,t}^{T}\theta_{20} + u_t$ under H_0, it is well known that $\widehat{\theta}_{10} - \theta_{10} = O_P\left(n^{-1/2}\right)$ and $\widehat{\theta}_{20} - \theta_{20} = O_P\left(n^{-1}\right)$ given the assumptions imposed in this paper. We therefore obtain $\left(\widehat{\theta}_0 - \theta_0\right)^{T} G_{2n}^{b}\left(\widehat{\theta}_0 - \theta_0\right) = O_P\left(n^{-1}\right)$. Next, we consider

$$
\begin{aligned}
G_{3n}^{b} &= \left(n^3 h\right)^{-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1} u_t X_t^{T} X_s X_s K_{t,s}\mathbf{1}_{n,t,s} \\
&\quad + \left(n^3 h\right)^{-1}\sum_{s=1}^{n-1}\sum_{t=s+1}^{n} u_s X_t^{T} X_s X_t K_{t,s}\mathbf{1}_{n,t,s} \equiv G_{3n,1}^{b} + G_{3n,2}^{b}.
\end{aligned}
$$

Below we will only calculate the stochastic order of $G_{3n,1}^{b}$ in details as the proof for $G_{3n,2}^{b}$ is similar. First, by Assumption A1 (iii) we have $E\left(G_{3n,j}^{b}\right) = 0$ for $j = 1, 2$. Applying the same method used to prove Lemma A.2 gives

$$
\begin{aligned}
E\left[G_{3n,1}^{b}\left(G_{3n,1}^{b}\right)^{T}\right] &= \frac{\sigma_u^2}{n^6 h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E\left[X_t^{T} X_s X_s X_s^{T} X_t^{T} X_s K_{t,s}^2 \mathbf{1}_{n,t,s}\right] \\
&\quad + \frac{\sigma_u^2}{n^6 h^2}\sum_{t=2}^{n}\sum_{s_1=1}^{t-1}\sum_{s_2=1, s_2\neq s_1}^{t-1} E\Big[X_t^{T} X_{s_1} X_{s_1} X_{s_2}^{T} X_t^{T} \\
&\qquad X_{s_2} K_{t,s_1} K_{t,s_2}\mathbf{1}_{n,t,s_1}\mathbf{1}_{n,t,s_2}\Big] \\
&= \begin{bmatrix} O\left(n^{-1}\right) & O\left(\left(n h^{2\delta_0/(1+\delta_0)}\right)^{-1}\right) \\ O\left(\left(n h^{2\delta_0/(1+\delta_0)}\right)^{-1}\right) & O(1)\end{bmatrix}
\end{aligned}
$$

for some $\delta_0 \in (0,1)$. Hence, we have $G_{3n,1}^{b} = \left(O_P\left(n^{-1/2}\right), O_P(1)\right)^{T}$ by Markov's inequality. Similarly, $G_{3n,2}^{b}$ has the same order as $G_{3n,1}^{b}$. Hence, $(\widehat{\theta}_0 - \theta_0)^{T} G_{3n}^{b} = O_P(n^{-1})$. This completes the proof of this lemma. ■

LEMMA A.8. *Under the assumptions given in Theorem 3.2, we obtain*

$$
\widehat{\sigma}_{n,b}^2 = \frac{2}{n^4 h}\sum_{t=1}^{n}\sum_{s\neq t}^{n} \tilde{u}_t^2 \tilde{u}_s^2 \left(X_t^{T} X_s\right)^2 K^2\left(\frac{Z_t - Z_s}{h}\right)\mathbf{1}_{n,t,s} = \sigma_{n,b}^2 + o_P(1),
$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

where $\tilde{u}_t = Y_t - X_t^T \widehat{\theta}^{(-t)}(Z_t)$ *is the semiparametric residual, and replacing* $\tilde{u}_t$ *by* u_t *for all* t *in* $\widehat{\sigma}^2_{n,b}$ *gives* $\sigma^2_{n,b}$.

Proof. Note that $\tilde{u}_t = Y_t - X_t^T \widehat{\theta}^{(-t)}(Z_t) = u_t - X_t^T \left[\widehat{\theta}^{(-t)}(Z_t) - \theta(Z_t)\right]$. The result in Lemma C.1 implies that $\sigma^2_{n,b}$ is the leading term of $\widehat{\sigma}^2_{n,b}$. This completes the proof of this lemma. ■

Proof of Theorem 3.3. Under H_1 we decompose the least squares estimator $\widehat{\theta}_0$ as

$$\widehat{\theta}_0 = \left(\sum_{t=1}^n X_t X_t^T\right)^{-1} \sum_{t=1}^n X_t X_t^T \theta(Z_t) + \left(\sum_{t=1}^n X_t X_t^T\right)^{-1} \sum_{t=1}^n X_t u_t. \tag{A.23}$$

It is well established that (from the linear regression model with $I(0)$ and $I(1)$ regressors)

$$\left(\sum_{t=1}^n X_t X_t^T\right)^{-1} \sum_{t=1}^n X_t u_t = \begin{pmatrix} O_p(n^{-1/2}) \\ O_p(n^{-1}) \end{pmatrix}, \tag{A.24}$$

which has a smaller order compared with the first term on the right-hand side of (A.23). Therefore, we only consider the leading term of $\widehat{\theta}_0$ and show that the stochastic order of $\widehat{\theta}_0$ depends on whether $\theta_2(Z_t)$ is a constant vector or not. For notational simplicity, we will only consider the case that both $X_{1,t}$ and $X_{2,t}$ are scalars. We denote $\theta_{1t} \equiv \theta_1(Z_t)$, $\theta_{2t} \equiv \theta_2(Z_t)$, and $D_n = \sum_{t=1}^n X_t X_t^T = \begin{pmatrix} d_{1n} & d_{2n} \\ d_{2n} & d_{3n} \end{pmatrix}$, where $d_{1n} = \sum_{t=1}^n X_{1,t}^2$, $d_{2n} = \sum_{t=1}^n X_{1,t} X_{2,t}$ and $d_{3n} = \sum_{t=1}^n X_{2,t}^2$. It is straightforward to show that

$$\begin{aligned}
&\begin{bmatrix} n^{-1/2} & 0 \\ 0 & 1 \end{bmatrix} \left(\sum_{t=1}^n X_t X_t^T\right)^{-1} \sum_{t=1}^n X_t X_t^T \theta(Z_t) \\
&= \frac{1}{det(D_n)} \begin{pmatrix} n^{-1/2} d_{3n} & -n^{-1/2} d_{2n} \\ -d_{2n} & d_{1n} \end{pmatrix} \begin{pmatrix} \sum_{t=1}^n X_{1,t}^2 \theta_{1t} + \sum_{t=1}^n X_{1,t} X_{2,t} \theta_{2t} \\ \sum_{t=1}^n X_{1,t} X_{2,t} \theta_{1t} + \sum_{t=1}^n X_{2,t}^2 \theta_{2t} \end{pmatrix} \\
&= \frac{1}{det(D_n)} \begin{pmatrix} n^{-1/2} d_{3n} \left(\sum_{t=1}^n X_{1,t}^2 \theta_{1t} + \sum_{t=1}^n X_{1,t} X_{2,t} \theta_{2t}\right) \\ -n^{-1/2} d_{2n} \left(\sum_{t=1}^n X_{1,t} X_{2,t} \theta_{1t} + \sum_{t=1}^n X_{2,t}^2 \theta_{2t}\right) \\ d_{1n} \left(\sum_{t=1}^n X_{1,t} X_{2,t} \theta_{1t} + \sum_{t=1}^n X_{2,t}^2 \theta_{2t}\right) \\ -d_{2n} \left(\sum_{t=1}^n X_{1,t}^2 \theta_{1t} + \sum_{t=1}^n X_{1,t} X_{2,t} \theta_{2t}\right) \end{pmatrix} \\
&= \frac{1}{det(D_n)} \begin{pmatrix} n^{-1/2} d_{3n} \sum_{t=1}^n X_{1,t} X_{2,t} \theta_{2t} - n^{-1/2} d_{2n} \sum_{t=1}^n X_{2,t}^2 \theta_{2t} \\ d_{1n} \sum_{t=1}^n X_{2,t}^2 \theta_{2t} - d_{2n} \sum_{t=1}^n X_{1,t} X_{2,t} \theta_{2t} \end{pmatrix} + O_p\left(n^{-1/2}\right) \\
&\xrightarrow{d} \begin{pmatrix} D^{-1} G_1 \\ D^{-1} G_2 \end{pmatrix},
\end{aligned} \tag{A.25}$$

where $D = \mu_2 W_{(2)} - (\mu_1 W_{(1)})^2$, $G_1 = W_{(1)} W_{(2)} \left(\mu_{1,\theta_2} - \mu_1 \mu_{\theta_2}\right)$, $G_2 = \mu_2 \mu_{\theta_2} W_{(2)} - \mu_1 \mu_{1,\theta_2} W_{(1)}^2$, and for $j = 1,2$ and $s = 1,2$, we denote

$$W_{(j)} = \int_0^1 B_\eta(r)^j dr, \quad \mu_j = E\left(X_{1,t}^j\right), \quad \mu_{s,\theta_j} = E[X_{1,t}^s \theta_j(Z_t)], \text{ and } \mu_{\theta_j} = E[\theta_j(Z_t)].$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

Therefore, combining (A.23), (A.24), and (A.25) gives

$$n^{-1/2}\widehat{\theta}_{10} \overset{d}{=} D^{-1}G_1 + O_p\left(n^{-1/2}\right) \text{ and } \widehat{\theta}_{20} \overset{d}{=} D^{-1}G_2 + O_p\left(n^{-1/2}\right), \quad \textbf{(A.26)}$$

where $\widehat{\theta}_0 = \left(\widehat{\theta}_{10}, \widehat{\theta}_{20}\right)^T$ and $a_n \overset{d}{=} b_n$ means that the two random sequences a_n and b_n have the same distribution asymptotically. Evidently, $\widehat{\theta}_{20} = O_p(1)$. If $Cov\left(X_{1,t}, \theta_2\left(Z_t\right)\right) = \mu_{1,\theta_2} - \mu_1\mu_{\theta_2} \neq 0$, we have $\widehat{\theta}_{10} = O_e\left(\sqrt{n}\right)$ as $\Pr(G_1 = 0) = \Pr(W_{(1)}W_{(2)} = 0) = 0$ by the fact that $W_{(1)}W_{(2)}$ is a continuous random variable. We have $\widehat{\theta}_{10} = O_e(1)$ if $Cov\left(X_{1,t}, \theta_2\left(Z_t\right)\right) = 0$ holds true. However, if $\theta_{2t} \equiv \theta_{20}$, a constant, for all t, simple calculations lead to

$$\begin{aligned}\widehat{\theta}_0 &= \frac{1}{det(D_n)}\begin{pmatrix} d_{3n}\sum_{t=1}^n X_{1,t}^2\theta_{1t} - d_{2n}\sum_{t=1}^n X_{1,t}X_{2,t}\theta_{1t} \\ d_{1n}\sum_{t=1}^n X_{1,t}X_{2,t}\theta_{1t} - d_{2n}\sum_{t=1}^n X_{1,t}^2\theta_{1t} + (d_{1n}d_{3n} - d_{2n}^2)\theta_{20}\end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \theta_{20}\end{pmatrix} + \frac{1}{det(D_n)}\begin{pmatrix} d_{3n}\sum_{t=1}^n X_{1,t}^2\theta_{1t} - d_{2n}\sum_{t=1}^n X_{1,t}X_{2,t}\theta_{1t} \\ d_{1n}\sum_{t=1}^n X_{1,t}X_{2,t}\theta_{1t} - d_{2n}\sum_{t=1}^n X_{1,t}^2\theta_{1t}\end{pmatrix}, \quad \textbf{(A.27)}\end{aligned}$$

which implies, by Lemma B.2,

$$\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{n}\end{bmatrix}\begin{pmatrix} \widehat{\theta}_{10} \\ \widehat{\theta}_{20} - \theta_{20}\end{pmatrix} \overset{d}{\to} \begin{pmatrix} D^{-1}G_3 \\ D^{-1}G_4\end{pmatrix}, \quad \textbf{(A.28)}$$

where $G_3 = W_{(2)}\mu_{2,\theta_1} - \mu_1\mu_{1,\theta_1}W_{(1)}^2$ and $G_4 = \left(\mu_2\mu_{1,\theta_1} - \mu_1\mu_{2,\theta_1}\right)W_{(1)}$. Therefore, combining (A.23), (A.24), and (A.28) gives

$$\widehat{\theta}_{10} \overset{d}{=} D^{-1}G_3 + O_p\left(n^{-1/2}\right) \text{ and } \sqrt{n}\left(\widehat{\theta}_{20} - \theta_{20}\right) \overset{d}{=} D^{-1}G_4 + O_p\left(n^{-1/2}\right), \quad \textbf{(A.29)}$$

which implies $\widehat{\theta}_{10} = O_e(1)$ and $\widehat{\theta}_{20} = \theta_{20} + O_p\left(n^{-1/2}\right)$. Compared with the OLS estimator when the true model has a varying coefficient for the integrated variable, the OLS estimator here has a stochastic order lowered by a factor of $n^{-1/2}$. Consequently, the OLS estimator for the coefficient of the integrated variable is $\sqrt{n}$-consistent if $\theta_1(z)$ is not constant over nonnegligible intervals, and the OLS estimator for the coefficient for the stationary covariate is not explosive any more although it is still inconsistent. Moreover, (A.29) indicates that the OLS estimator of the coefficient for the integrated variable is not super-consistent any more if the stationary covariate has a varying coefficient.

Below we will show that $I_{2n}^b = O_e(n)$ in case (I) and that $I_{2n}^b = O_e(1)$ in case (II). Therefore, the leading term of $n\sqrt{h}\widehat{I}_n^b$ is $n\sqrt{h}I_{2n}^b = O_e(n^2\sqrt{h})$ for case (I), and it becomes $O_e(n\sqrt{h})$ for case (II).

Case (I): $\Pr\{\theta_2(Z_t) \neq \theta_2\} > 0$ for any $\theta_2 \in \Theta_2 \subset R$. Define $\bar{\theta}_1 = D^{-1}G_1$ and $\bar{\theta}_2 = D^{-1}G_2$. Then obviously, $\bar{\theta}_j = O_p(1)$ for $j = 1, 2$. Then, by (A.26), we have $\widehat{\theta}_{10} \overset{d}{=} n^{1/2}\bar{\theta}_1$ and $\widehat{\theta}_{20} \overset{d}{=} \bar{\theta}_2$. Hence, the leading term in I_{2n}^b can be obtained by replacing $\widehat{\theta}_{10}$ and $\widehat{\theta}_{20}$ by $n^{1/2}\bar{\theta}_1$ and $\bar{\theta}_2$, respectively. We have

$$n^{-1}I_{2n}^b = \frac{2}{n^4h}\sum_{t=2}^n\sum_{s=1}^{t-1} X_t^T X_s X_t^T(\widehat{\theta}_0 - \theta_t)X_s^T(\widehat{\theta}_0 - \theta_s)K_{t,s}\mathbf{1}_{n,t,s}$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

$$\stackrel{d}{=} \frac{2}{n^3h}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left(X_{1,t}X_{1,s}+X_{2,t}X_{2,s}\right)[X_{1,t}(\bar{\theta}_1-\frac{\theta_{1t}}{\sqrt{n}})+\frac{X_{2,t}}{\sqrt{n}}(\bar{\theta}_2-\theta_{2t})]$$
$$\times[X_{1,s}(\bar{\theta}_1-\frac{\theta_{1s}}{\sqrt{n}})+\frac{X_{2,s}}{\sqrt{n}}(\bar{\theta}_2-\theta_{2s})]K_{t,s}\mathbf{1}_{n,t,s}\left[1+o_p(1)\right]$$
$$=\frac{2}{n^3h}\sum_{t=2}^{n}\sum_{s=1}^{t-1}X_{2,t}X_{2,s}\left(X_{1,t}\bar{\theta}_1+\frac{X_{2,t}}{\sqrt{n}}\bar{\theta}_2-\frac{X_{2,t}}{\sqrt{n}}\theta_{2t}\right)$$
$$\times\left(X_{1,s}\bar{\theta}_1+\frac{X_{2,s}}{\sqrt{n}}\bar{\theta}_2-\frac{X_{2,s}}{\sqrt{n}}\theta_{2s}\right)K_{t,s}\mathbf{1}_{n,t,s}\left[1+o_p(1)\right], \tag{A.30}$$

where the leading term of (A.30) equals a summation of six distinct components, and each component has exactly the same order of $O_e(1)$; for example, $\left(n^4h\right)^{-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1}X_{2,t}^2 X_{2,s}^2\theta_{2t}\theta_{2s}K_{t,s}\mathbf{1}_{n,t,s} \stackrel{d}{\to} \int_0^1\int_0^r\left(B_\eta(r)B_\eta(s)\right)^2 ds dr E\left[\theta_2^2(Z_1)f(Z_1)\right] = O_e(1)$ by Lemma B.2. Therefore, $I_{2n}^b = O_e(n)$. In addition, by the symmetry of (A.30) and the fact that

$$\left(n^3h\right)^{-1}\sum_{t=1}^{n}X_{2,t}^2\left(X_{1,t}\bar{\theta}_1+\frac{X_{2,t}}{\sqrt{n}}\bar{\theta}_2-\frac{X_{2,t}}{\sqrt{n}}\theta_{2t}\right)^2K(0)=O_e\left(n^{-1}h^{-1}\right)=o_p(1)$$

imply that $n^{-1}I_{2n}^b$ converges in distribution to a positive random variable. Similarly, one can show that $I_{3n}^b = o_p(n)$. Because we have already shown $I_{1n}^b = o_p(1)$, we have $n\sqrt{h}\widehat{I}_n^b = O_e(n^2\sqrt{h})$, which diverges to $+\infty$ at the rate of $n^2\sqrt{h}$. Evidently, whether $Cov\left(X_{1,t},\theta_2(Z_t)\right)=0$ or not does not change the result as $X_{2,t}\bar{\theta}_2/\sqrt{n}$ dominates the order of both $n^{-1}I_{2n}^b$ and I_{3n}^b.

Case (II): $\Pr\{\theta_2(Z_t)\equiv\theta_{20}\}=1$. Define $\tilde{\theta}_1 = D^{-1}G_3$ and $\tilde{\theta}_2 = D^{-1}G_4$, where $\tilde{\theta}_j = O_p(1)$ for $j=1,2$. Then, by (A.29), we get $\widehat{\theta}_{10}\stackrel{d}{=}\tilde{\theta}_1$ and $\sqrt{n}\left(\widehat{\theta}_{20}-\theta_{20}\right)\stackrel{d}{=}\tilde{\theta}_2$. Hence, the leading term in I_{2n}^b can be obtained by replacing $\widehat{\theta}_{10}$ and $\widehat{\theta}_{20}-\theta_{20}$ by $\tilde{\theta}_1$ and $n^{-1/2}\tilde{\theta}_2$, respectively. Therefore, we have

$$I_{2n}^b=\frac{2}{n^3h}\sum_{t=2}^{n}\sum_{s=1}^{t-1}X_t^TX_s(\widehat{\theta}_0-\theta_t)X_tX_s^T(\widehat{\theta}_0-\theta_s)K_{t,s}\mathbf{1}_{n,t,s}$$
$$\stackrel{d}{=}\frac{2}{n^3h}\sum_{t=2}^{n}\sum_{s=1}^{t-1}(X_{1,t}X_{1,s}+X_{2,t}X_{2,s})[X_{1,t}(\tilde{\theta}_1-\theta_{1t})+X_{2,t}\tilde{\theta}_2/\sqrt{n})]$$
$$\times[X_{1,s}(\tilde{\theta}_1-\theta_{1s})+X_{2,s}\tilde{\theta}_2/\sqrt{n})]K_{t,s}\mathbf{1}_{n,t,s}\left[1+o_p(1)\right]$$
$$=\frac{2}{n^2h}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\frac{X_{2,t}}{\sqrt{n}}\frac{X_{2,s}}{\sqrt{n}}\left[X_{1,t}(\tilde{\theta}_1-\theta_{1t})+\frac{X_{2,t}}{\sqrt{n}}\tilde{\theta}_2\right]$$
$$\times\left[X_{1,s}(\tilde{\theta}_1-\theta_{1s})+\frac{X_{2,s}}{\sqrt{n}}\tilde{\theta}_2\right]K_{t,s}\mathbf{1}_{n,t,s}\left[1+o_p(1)\right], \tag{A.31}$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

where the leading term of equation (A.31) equals a summation of four components, and each component has exactly the same order of $O_e(1)$; e.g., applying Lemma B.2 gives $\left(n^4 h\right)^{-1} \sum_{t=2}^{n} \sum_{s=1}^{t-1} X_{2,t}^2 X_{2,s}^2 K_{t,s} \mathbf{1}_{n,t,s} \xrightarrow{d} \int_0^1 \int_0^r \left[B_\eta(r) B_\eta(s)\right]^2 ds dr E\left[f\left(Z_1\right)\right]$ $= O_e(1)$. Therefore, $I_{2n}^b = O_e(1)$. In addition, by the symmetry of (A.31) and the fact that

$$\left(n^4 h\right)^{-1} \sum_{t=1}^{n} X_{2,t} X_{2,s} \left[X_{1,t}(\tilde{\theta}_1 - \theta_{1t}) + \frac{X_{2,t}}{\sqrt{n}} \tilde{\theta}_2\right]^2 K(0) = O_e\left(n^{-2} h^{-1}\right) = o_p(1),$$

which implies that I_{2n}^b converges in distribution to a positive random variable. Similarly, one can show that $I_{3n}^b = o_p(1)$. Taking these results together lead to $n\sqrt{h}\,\widehat{I}_n^b = n\sqrt{h} I_{2n}^b$ $+ o_p(n\sqrt{h}) = O_e(n\sqrt{h})$, which diverges to $+\infty$ at the rate of $n\sqrt{h}$.

Finally, when $X_{1,t}$ and $X_{2,t}$ are vectors of dimensions $d_1 \times 1$ and $d_2 \times 1$, respectively, it is easy to show that the conclusion in Theorem 3.3 still holds true. All one needs to do is to replace $B_\eta(r)^2$ by $B_\eta(r) B_\eta^T(r)$, $E\left(X_{1,t} X_{2,t}^T\right)$ rather than $E(X_{1,t} X_{2,t})$, and so on. This completes the proof of Theorem 3.3. ■

APPENDIX B: Some Useful Lemmas

LEMMA B.1. *Suppose that* $\{\xi_i\}$ *is a* q*-dimensional strictly stationary process satisfying the* β*-mixing condition with coefficients* β_τ. *For any* j $(1 \le j \le k-1)$ *and arbitrary integers* $i_1 < i_2 < \cdots < i_k$, $(\xi_{i_1}, \ldots, \xi_{i_k})$, $(\xi_{i_1}, \ldots, \xi_{i_j})$, *and* $(\xi_{i_j+1}, \ldots, \xi_{i_k})$ *have cumulative distribution functions* $F(x_1, \ldots, x_k)$, $F^{(1)}(x_1, \ldots, x_j)$, *and* $F^{(2)}(x_{j+1}, \ldots, x_k)$, *respectively. Let* $G(x_1, \ldots, x_k)$ *be a Borel measurable function such that for some* $\delta_0 > 0$,

$$\int \cdots \int_{R^{qk}} |G(x_1, \ldots, x_k)|^{1+\delta_0} dF^{(1)}(x_1, \ldots, x_j) dF^{(2)}(x_{j+1}, \ldots, x_k) \le C < \infty.$$

Then

$$\begin{aligned}&\left| \int \cdots \int_{R^{qk}} G(x_1, \ldots, x_k) dF(x_1, \ldots, x_k) \right.\\ &\qquad \left. - \int \cdots \int_{R^{qk}} G(x_1, \ldots, x_k) dF^{(1)}(x_1, \ldots, x_j) dF^{(2)}(x_{j+1}, \ldots, x_k) \right| \\ &\quad \le 4C^{1/(1+\delta_0)} \beta_\tau^{\delta_0/(1+\delta_0)}, \quad \textit{where } \tau = i_{j+1} - i_j.\end{aligned}$$

Proof. This is Lemma 1 in Yoshihara (1976). ■

To simplify notation, the following lemma takes $X_{1,t}$ and $X_{2,t}$ as scalars.

LEMMA B.2. *Let* $g(\cdot)$ *and* $m(\cdot)$ *be Borel measurable functions. Denote* $\mu_g(z) = E\left[g(X_{1,t})|Z_t = z\right]$, $\mu_m(z) = E\left[m(X_{1,t})|Z_t = z\right]$, $\psi_{g,\delta}(z) = E\left[\left|g\left(X_{1,t}\right)\right|^{1+\delta} |Z_t = z\right]$, *and* $\psi_{m,\delta}(z) = E\left[\left|m\left(X_{1,t}\right)\right|^{1+\delta} |Z_t = z\right]$ *for some* $\delta > 0$. *If* $\mu_g(z)$, $\mu_m(z)$, $\psi_{g,\delta}(z)$, *and*

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

$\psi_{m,\delta}(z)$ all have bounded uniformly continuous derivatives up to the second order, under Assumptions B1, A2, B3, A5(i), and A6, we obtain, for any positive integers j, j', and l,

$$A_n = \frac{1}{n^2 n^{(j+j')/2} h} \sum_{t=2}^{n} \sum_{s=1}^{t-1} X_{2,t}^{j} X_{2,s}^{j'} g(X_{1,t}) m(X_{1,s}) K_{t,s}^{l} \mathbf{1}_{n,t,s}$$

$$\xrightarrow{d} \nu_l E[\mu_g(Z_1)\mu_m(Z_1) f(Z_1)] \int_0^1 \int_0^r B_\eta^j(r) B_\eta^{j'}(s) ds dr,$$

where $\nu_l = \int K^l(u) du$.

Proof. Applying the same proof method used in the proof of Lemma A.2, we obtain

$$A_n = \frac{1}{n^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \left(\frac{X_{2,t}}{\sqrt{n}}\right)^j \left(\frac{X_{2,s}}{\sqrt{n}}\right)^{j'} E[h^{-1} g(X_{1,t}) m(X_{1,s}) K_{t,s}^{l} \mathbf{1}_{n,t,s}] + o_p(1),$$

where $\mathbf{1}_{n,t,s} = \mathbf{1}_{n,t}\mathbf{1}_{n,s}$ and $\mathbf{1}_{n,t} = \mathbf{1}(Z_t \in \mathcal{S}_n)$, and $\mathbf{1}(A)$ is a trimming indicator function which equals 1 if A holds and 0 otherwise.

By Lemma B.1, we have

$$\left| h^{-1} E[g(X_{1,t}) m(X_{1,s}) K_{t,s}^{l} \mathbf{1}_{n,t,s}] - \chi_{t,s} \right| \le C h^{\frac{-\delta}{1+\delta}} \beta_{|t-s|}^{\delta/(1+\delta)},$$

where letting $\omega = (Z_t - Z_s)/h$ and applying the change of variables gives

$$\chi_{t,s} = h^{-1} \int\int \mu_g(Z_t)\mu_m(Z_s) K^l\left(\frac{Z_t - Z_s}{h}\right) f(Z_t) f(Z_s) \mathbf{1}(Z_t \in \mathcal{S}_n)\mathbf{1}(Z_s \in \mathcal{S}_n) dZ_t dZ_s$$

$$= \int\int \mu_g(h\omega + Z_s)\mu_m(Z_s) K^l(\omega) f(h\omega + Z_s) f(Z_s) \mathbf{1}(h\omega + Z_s \in \mathcal{S}_n)\mathbf{1}(Z_s \in \mathcal{S}_n) d\omega dZ_s$$

$$= \int K^l(\omega) d\omega E\left[\mu_g(Z_1)\mu_m(Z_1) f(Z_1) \mathbf{1}(Z_1 \in \mathcal{S}_n)\right] + O(h),$$

and

$$h^{-(1+\delta)} \int\int \psi_{g,\delta}(Z_t)\psi_{m,\delta}(Z_s) K^{l(1+\delta)}\left(\frac{Z_t - Z_s}{h}\right) f(Z_t) f(Z_s) dZ_t dZ_s$$

$$= h^{-\delta} \int\int \psi_{g,\delta}(h\omega + Z_s)\psi_{m,\delta}(Z_s) K^{l(1+\delta)}(\omega) f(h\omega + Z_s) f(Z_s) d\omega dZ_s$$

$$= h^{-\delta} \int K^{l(1+\delta)}(\omega) d\omega E\left[\psi_{g,\delta}(Z_1)\psi_{m,\delta}(Z_1) f(Z_1)\right] + O\left(h^{2-\delta}\right).$$

Therefore, we have

$$\frac{1}{n^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \left(\frac{X_{2,t}}{\sqrt{n}}\right)^j \left(\frac{X_{2,s}}{\sqrt{n}}\right)^{j'} \left(h + h^{\frac{-\delta}{1+\delta}} \beta_{|t-s|}^{\delta/(1+\delta)}\right) = O_p(h) + O_p\left((nh)^{\frac{-\delta}{1+\delta}} n^{-\frac{1}{1+\delta}}\right)$$
$$= o_p(1)$$

as $\max_{1 \le t \le n} \|X_{2,t}\| = O_p(\sqrt{n})$ by (8), $\sum_{l=1}^{\infty} \beta_l^{\delta/(1+\delta)} \le C < \infty$ by Assumption A1 (ii), and $nh \to \infty$ and $h \to 0$ as $n \to \infty$ by Assumption A6. Since $\Pr(Z_1 \in \mathcal{S}_n) \to 1$ as $n \to \infty$, we complete the proof of this lemma. ■

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

APPENDIX C: Proof of Weak Uniform Convergence under Case (b)

In this appendix we extend Hansen's (2008, Thm. 2) weak uniform convergence results of kernel estimator derived for absolutely regular (β-mixing) time series to time series with both integrated and absolutely regular variables. The following proof is derived for the case that Z_t has an unbounded support.

LEMMA C.1. *Under Assumptions B1, A2, B3, and A4-A6, we obtain* $\sup_{z\in\mathcal{S}_n}||\widehat{\theta}_1(z)-\theta_1(z)||=o_p(1)$ *and* $\sup_{z\in\mathcal{S}_n}||\widehat{\theta}_2(z)-\theta_2(z)||=o_p\left(n^{-1/2}\right)$*, where* $\mathcal{S}_n=\mathcal{S}\cap[-c_n,c_n]$ *and* $c_n=O\left(n^{\phi}\ln n\right)$ *is a sequence of nondecreasing positive numbers for any* $\phi>0$.

Proof. The kernel estimator of $\theta(z)$ in (4) can be rewritten as $\widehat{\theta}(z)-\theta(z)=\Psi_{n1}(z)^{-1}\Psi_{n2}(z)$, where

$$\Psi_{n1}(z)=\frac{1}{nh}\sum_{t=1}^{n}\begin{pmatrix}X_{1,t}X_{1,t}^T & X_{1,t}X_{2,t}^T\\ X_{2,t}X_{1,t}^T & X_{2,t}X_{2,t}^T\end{pmatrix}K_t(z)\equiv\begin{pmatrix}\Psi_{n1,11}(z) & \Psi_{n1,12}(z)\\ \Psi_{n1,12}(z)^T & \Psi_{n1,22}(z)\end{pmatrix}, \tag{C.1}$$

and $\Psi_{n2}(z)=(nh)^{-1}\sum_{t=1}^{n}X_t\left[X_t^T\Pi(Z_t,z)+u_t\right]K_t(z)$ with $\Pi(Z_t,z)=\theta\left(Z_t\right)-\theta(z)$. Applying Hansen (2008, Thm. 2) to $\Psi_{n1,11}(z)$ gives $\sup_{z\in\mathcal{S}_n}\left|\Psi_{n1,11}(z)-E\Psi_{n1,11}(z)\right|=O_p(b_n)$, where $b_n=\sqrt{\ln n/(nh)}$. By Assumptions B1 and B3 and applying the change of variables gives $E\Psi_{n1,11}(z)=g_1(z)+O\left(h^2\right)$, which holds uniformly over $z\in\mathcal{S}_n$ with

$$g_1(z)=E\left(X_{1,t}X_{1,t}^T|Z_t=z\right)f(z).$$

Therefore, we have

$$\sup_{z\in\mathcal{S}_n}\left\|\Psi_{n1,11}(z)-g_1(z)\right\|=O_p\left(b_n+h^2\right). \tag{C.2}$$

As for $\Psi_{n1,12}(z)$, it can be expressed as

$$\begin{aligned}\Psi_{n1,12}(z)&=(nh)^{-1}\sum_{t=1}^{n}X_{1,t}X_{2,t}^TK_t(z)\\ &=(nh)^{-1}E\left[X_{1,t}K_t(z)\right]\sum_{t=1}^{n}X_{2,t}^T+n^{-1}\sum_{t=1}^{n}e_t(z)X_{2,t}^T,\end{aligned}$$

where $e_t(z)=h^{-1}\left\{X_{1,t}K_t(z)-E\left[X_{1,t}K_t(z)\right]\right\}$. Denoting

$$g_2(z)=E\left(X_{1,t}|Z_t=z\right)f(z), \tag{C.3}$$

we obtain that $h^{-1}E\left[X_{1,t}K_t(z)\right]=g_2(z)+O\left(h^2\right)$ holds uniformly over $z\in\mathcal{S}_n$. Hence,

$$(nh)^{-1}E\left[X_{1,t}K_t(z)\right]\sum_{t=1}^{n}X_{2,t}=g_2(z)n^{-1}\sum_{t=1}^{n}X_{2,t}+O_p\left(h^2\sqrt{n}\right) \tag{C.4}$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

as $\max_{1\le t\le n}\left\|X_{2,t}\right\| = O_p\left(\sqrt{n}\right)$. For the second term of $\Psi_{n1,12}(z)$, we set $\tau_n \in (0,1)$, $N = [1/\tau_n]$, $t_k = [kn/N]+1$, $t_k^* = t_{k+1}-1$, and $t_k^{**} = \min\left(t_k^*, n\right)$. To simplify notation and without loss of generality, we give the proof of (C.5) for scalar case. We have

$$\begin{aligned}
&\left|n^{-3/2}\sum_{t=1}^{n} X_{2,t}e_t(z)\right| \\
&\quad = \left|n^{-3/2}\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^{**}} X_{2,t}e_t(z)\right| \\
&\quad \le \left|n^{-3/2}\sum_{k=0}^{N-1} X_{2,t_k}\sum_{t=t_k}^{t_k^{**}} e_t(z)\right| + \left|n^{-3/2}\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^{**}}\left(X_{2,t}-X_{2,t_k}\right)e_t(z)\right| \\
&\quad \le \sup_{r\in[0,1]}\left|B_{n,\eta}(r)\right| n^{-1}\sum_{k=0}^{N-1}\left|\sum_{t=t_k}^{t_k^{**}} e_t(z)\right| + \sup_{|r-r'|\le\tau_n}\left|B_{n,\eta}(r)-B_{n,\eta}\left(r'\right)\right| n^{-1}\sum_{t=1}^{n}\left|e_t(z)\right|,
\end{aligned}$$

where $\sup_{r\in[0,1]}\left|B_{n,\eta}(r)\right| = O_p(1)$ by Assumption A2 and $\sup_{z\in\mathcal{S}} n^{-1}\sum_{t=1}^{n}|e_t(z)| = O_p(1)$. Simple calculations give

$$\begin{aligned}
\sup_{z\in\mathcal{S}_n} n^{-1}\sum_{k=0}^{N-1}\left|\sum_{t=t_k}^{t_k^{**}} e_t(z)\right| &\le \frac{N}{n}\sup_{z\in\mathcal{S}_n}\sup_{0\le k\le N-1}\left|\sum_{t=t_k}^{t_k^{**}} e_t(z)\right| \\
&\le \sup_{z\in\mathcal{S}_n}\sup_{t+\tau_n n\le n}\left|\frac{1}{\tau_n n}\sum_{i=t}^{t+\tau_n n} e_i(z)\right| = O_p\left(b_{\tau_n,n}\right),
\end{aligned}$$

where $b_{\tau_n,n} = \sqrt{\ln(n\tau_n)/(n\tau_n h)}$ and the last equality follows from Hansen (2008, Thm. 2). As $\sup_{|r-r'|\le\tau_n}\left|B_{n,\eta}(r)-B_{n,\eta}\left(r'\right)\right| = O_p\left(\sqrt{\tau_n}\right)$, we have

$$\sup_{z\in\mathcal{S}_n}\left|n^{-3/2}\sum_{t=1}^{n} X_{2,t}e_t(z)\right| = O_p\left(b_n^*\right), \tag{C.5}$$

where $b_n^* = \sqrt{\ln(n\tau_n)/(n\tau_n h)} + \sqrt{\tau_n}$. Combining (C.4) and (C.5) gives

$$\sup_{z\in\mathcal{S}_n}\frac{1}{\sqrt{n}}\left\|\Psi_{n1,12}(z) - g_2(z)n^{-1}\sum_{t=1}^{n} X_{2,t}^T\right\| = O_p\left(h^2\right) + O_p\left(b_n^*\right). \tag{C.6}$$

Next, we consider $\Psi_{n1,22}(z) = (nh)^{-1}\sum_{t=1}^{n} X_{2,t}X_{2,t}^T K_t(z) = h^{-1}E\left[K_t(z)\right]n^{-1}\sum_{t=1}^{n} X_{2,t}X_{2,t}^T + n^{-1}\sum_{t=1}^{n} X_{2,t}X_{2,t}^T e_t(z)$, where $e_t(z) = h^{-1}\left[K_t(z) - E\left(K_t(z)\right)\right]$ and $h^{-1}E\left[K_t(z)\right] = f(z) + O\left(h^2\right)$. Applying the same method used above, we obtain

$$\sup_{z\in\mathcal{S}_n} n^{-1}\left\|\Psi_{n1,22}(z) - f(z)n^{-1}\sum_{t=1}^{n} X_{2,t}X_{2,t}^T\right\| = O_p\left(h^2\right) + O_p\left(b_n^*\right). \tag{C.7}$$

Letting $D_n = \text{diag}\{I_{d_1}, \sqrt{n} I_{d_2}\}$ and combining (C.2), (C.6), and (C.7), we obtain

$$\sup_{z \in \mathcal{S}_n} \left\| D_n^{-1} \Psi_{n1}(z) D_n^{-1} - S(z) \right\| = O_p\left(h^2\right) + O_p(b_n) + O_p\left(b_n^*\right),$$

where

$$S(z) = \begin{pmatrix} g_1(z) & g_2(z) n^{-3/2} \sum_{t=1}^n X_{2,t}^T \\ n^{-3/2} \sum_{t=1}^n X_{2,t} g_2(z)^T & f(z) n^{-2} \sum_{t=1}^n X_{2,t} X_{2,t}^T \end{pmatrix}.$$

Finally, we consider $\Psi_{n2}(z)$. Simple mathematical manipulations give

$$\begin{aligned} D_n^{-1} \Psi_{n2}(z) &= \frac{1}{nh} \sum_{t=1}^n D_n^{-1} X_t X_t^T D_n^{-1} D_n \Pi(Z_t, z) K_t(z) + \frac{1}{nh} \sum_{t=1}^n D_n^{-1} X_t u_t K_t(z) \\ &= \frac{1}{nh} \sum_{t=1}^n \begin{pmatrix} X_{1,t} X_{1,t}^T \Pi_1(Z_t, z) + X_{1,t} \frac{X_{2,t}^T}{\sqrt{n}} \sqrt{n} \Pi_2(Z_t, z) \\ \frac{X_{2,t}}{\sqrt{n}} X_{1,t}^T \Pi_1(Z_t, z) + \frac{X_{2,t} X_{2,t}^T}{n} \sqrt{n} \Pi_2(Z_t, z) \end{pmatrix} K_t(z) \\ &\quad + \frac{1}{nh} \sum_{t=1}^n \begin{pmatrix} X_{1,t} u_t \\ \frac{X_{2,t}}{\sqrt{n}} u_t \end{pmatrix} K_t(z) \\ &\equiv \begin{pmatrix} \Psi_{n2,11}(z) + \Psi_{n2,12}(z) \\ \Psi_{n2,12}^T(z) + \Psi_{n2,22}(z) \end{pmatrix} + \begin{pmatrix} \Psi_{n2,13}(z) \\ \Psi_{n2,23}(z) \end{pmatrix}. \end{aligned}$$

Again, applying the same method used in the proof of the weak uniform convergence rate for $\Psi_{n1}(z)$, we have that

$$\begin{aligned} &\begin{pmatrix} \Psi_{n2,11}(z) + \Psi_{n2,12}(z) \\ \Psi_{n2,12}^T(z) + \Psi_{n2,22}(z) \end{pmatrix} - h^2 \mu_2(K) \left[S^{(1)}(z) D_n \theta^{(1)}(z) + S(z) D_n \theta^{(2)}(z)/2 \right] \\ &\quad = O_p\left(\sqrt{n}\left(h^4 + h b_n^*\right)\right) \end{aligned}$$

holds uniformly over $z \in \mathcal{S}_n$, where $\mu_2(K) = \int u^2 K(u)\, du$, and that

$$\sup_{z \in \mathcal{S}_n} \left\| \Psi_{n2,13}(z) \right\| = O_p(b_n) \quad \text{and} \quad \sup_{z \in \mathcal{S}_n} \left\| \Psi_{n2,23}(z) \right\| = O_p\left(b_n^*\right). \tag{C.8}$$

Therefore, we obtain that

$$\begin{aligned} &D_n \left\{ \widehat{\theta}(z) - \theta(z) - h^2 \mu_2(K) D_n^{-1} S(z)^{-1} \left[S^{(1)}(z) D_n \theta^{(1)}(z) + S(z) D_n \theta^{(2)}(z)/2 \right] \right\} \\ &\quad = O_p\left(\sqrt{n} \delta_n^{-1}\left(h^4 + h b_n^*\right)\right) + O_p\left(\delta_n^{-1}\left(b_n + b_n^*\right)\right) \end{aligned}$$

or

$$\begin{aligned} &D_n \left[\widehat{\theta}(z) - \theta(z) - h^2 \mu_2(K) \theta^{(2)}(z)/2 \right] \\ &\quad = h^2 \mu_2(K) S(z)^{-1} S^{(1)}(z) D_n \theta^{(1)}(z) + O_p\left(\sqrt{n} \delta_n^{-1}\left(h^4 + h b_n^*\right)\right) + O_p\left(\delta_n^{-1}\left(b_n + b_n^*\right)\right) \\ &\quad = O_p\left(\delta_n^{-1} \sqrt{n} h^2\right) + O_p\left(\delta_n^{-1} \sqrt{n} h b_n^*\right) + O_p\left(\delta_n^{-1}\left(b_n + b_n^*\right)\right) \end{aligned} \tag{C.9}$$

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press

holds uniformly over $z \in \mathcal{S}_n$, where $\delta_n = \inf_{z \in \mathcal{S}_n} f(z)$. Let $\tau_n = O\left(n^{-\varsigma}\right)$, $\delta_n = O\left(n^{-\epsilon}\right)$, and $h = O\left(n^{-\alpha}\right)$. Taking $\alpha \in (1/3, 1)$, $\varsigma \in (1-2\alpha, \min(1-\alpha, \alpha))$, and $\epsilon \in \min\{2\alpha - 1/2, (\alpha - \varsigma)/2, (\varsigma - 1 + 2\alpha)/2, (1 - \varsigma - \alpha)/2, \varsigma/2\}$, we can show that $D_n\left[\widehat{\theta}(z) - \theta(z) - h^2 \mu_2(K)\theta^{(2)}(z)/2\right] = o_p(1)$. This completes the proof of this lemma. ■

Remark 6. When Z_t has an unbounded support, $\delta_n \to 0$ as $c_n \to \infty$, and the rate at which c_n can go to $+\infty$ is determined by the tail behavior of the density function $f(z)$. When Z_t has a bounded support, $\mathcal{S}_n$ trims out the data within ς_n distance to the boundary of $\mathcal{S}$, where $\varsigma_n > 0$ and $\varsigma_n \to 0$ as $n \to \infty$; for example, if $\mathcal{S} = [0, 1]$, we can choose $\mathcal{S}_n = [\varsigma_n, 1 - \varsigma_n]$, where $\varsigma_n \to 0$ and $h/\varsigma_n \to 0$ as $n \to \infty$.

https://doi.org/10.1017/S0266466615000018 Published online by Cambridge University Press