Estimating Smooth Structural Change in Cointegration Models[1]

by

Peter C. B. Phillips[2], Degui Li[3] and Jiti Gao[4]

Abstract

This paper studies nonlinear cointegration models in which the structural coefficients may evolve smoothly over time. These time-varying coefficient functions are well-suited to many practical applications and can be estimated conveniently by nonparametric kernel methods. It is shown that the usual asymptotic methods of kernel estimation completely break down in this setting when the functional coefficients are multivariate. The reason for this breakdown is a kernel-induced degeneracy in the weighted signal matrix associated with the nonstationary regressors, a new phenomenon in the kernel regression literature. Some new techniques are developed to address the degeneracy and resolve the asymptotics, using a path-dependent local coordinate transformation to re-orient coordinates and accommodate the degeneracy. The resulting asymptotic theory is fundamentally different from the existing kernel literature, giving two different limit distributions with different convergence rates in the different directions (or combinations) of the (functional) parameter space. Both rates are faster than the usual ($\sqrt{nh}$) rate for nonlinear models with smoothly changing coefficients and local stationarity. Hence two types of super-consistency apply in nonparametric kernel estimation of time-varying coefficient cointegration models. The higher rate of convergence ($n\sqrt{h}$) lies in the direction of the nonstationary regressor vector at the local coordinate point. The lower rate (nh) lies in the degenerate directions but is still super-consistent for nonparametric estimators. In addition, local linear methods are used to reduce asymptotic bias and a fully modified kernel regression method is proposed to deal with the general endogenous nonstationary regressor case. Simulations are conducted to explore the finite sample properties of the methods and a practical application is given to examine time varying empirical relationships involving consumption, disposable income, investment and real interest rates. ‘

Key words and phrases: Cointegration; Endogeneity; Kernel degeneracy; Nonparametric regression; Super-consistency; Time varying coefficients.

JEL classification: C13, C14, C32.

Abbreviated title: Smooth Change in Cointegration Models

---

[1]This paper was commenced during November 2012 while the first author visited Monash University. Phillips thanks the Department of Econometrics at Monash for its hospitality during this visit and acknowledges support from a Kelly Felllowship and the NSF under Grant Nos SES 0956687 and SES 1258258. Gao acknowledges support from the ARC under Grant Nos DP1096374 and DP1314229.

[2]Yale University, University of Auckland, Southampton University, and Singapore Management University

[3]University of York

[4]Monash University

1 Introduction

Cointegration models are now one of the most commonly used frameworks for applied research in econometrics, capturing long term relationships among trending macroeconomic time series and present value links between asset prices and fundamentals in finance. These models conveniently combine stochastic trends in individual series with linkages between series that eliminate trending behavior and reflect latent regularities in the data. In spite of their importance and extensive research on their properties (e.g. Park and Phillips, 1988; Johansen, 1988; Phillips, 1991; and Saikkonen, 1995; among many others) linear cointegration models are often rejected by the data even when there is clear co-movement in the series.

Various nonlinear parametric cointegrating models have been suggested to overcome such deficiencies. These models have been the subject of an increasing amount of econometric research following the development of methods for handling nonlinear nonstationary process asymptotics (Park and Phillips, 1999, 2001). Parameter instability and functional form misspecification may also limit the performance of such nonlinear parametric cointegration models in empirical applications (Hong and Phillips, 2010; Kasparis and Phillips, 2012; Kasparis *et al*, 2013). Most recently, therefore, attention has been given to flexible nonparametric and semiparametric approaches that can cope with the unknown functional form of responses in a nonstationary time series setting (Karlsen *et al*, 2007; Wang and Phillips, 2009a, 2009b; Gao and Phillips, 2013). A futher extension of the linear framework allows cointegrating relationships to evolve smoothly over time using time-varying cointegrating coefficients (e.g. Park and Hahn, 1999; Juhl and Xiao, 2005; Cai *et al*, 2009; Xiao, 2009). This framework seems particularly well suited to empirical applications where there may be structural evolution in a relationship over time, thereby tackling one of the main limitations of fixed coefficient linear and nonlinear formulations. It is this framework that is the subject of the present investigation.

Cai *et al* (2009) and Xiao (2009) proposed a nonlinear cointegrating model with functional coefficients of the form

$$y_t = x_t' f(z_t) + u_t, \quad t = 1, \cdots, n, \tag{1.1}$$

where $f(\cdot)$ is a d-dimensional function of coefficients, x_t is an $I(1)$ vector, and both z_t and u_t are scalar processes. By keeping the covariate z_t scalar, the model helps to circumvent the dimensionality difficulties that arise when $d > 1$. Extensions of (1.1) to more general nonparametric and semiparametric formulations are considered in Gao and Phillips (2012) and Li *et al* (2013). Kernel-based nonparametric methods such as Nadaraya-Watson local level regression and local polynomial regression can be used to estimate the functional coefficients in (1.1). While the generality of this model appeals and economic theory may provide some guidance concerning the covariates, it is often difficult in practice to select a single covariate z_t for use in (1.1). When a single variable is not suggested by theory, it will often be of interest to set the model in a different context and explore changes in the cointegrating relationship between the variables y_t and x_t over time. This perspective leads to the following cointegration model with time-varying coefficient functions

$$y_t = x_t' f\left(\frac{t}{n}\right) + u_t = x_t' f_t + u_t, \quad t = 1, \cdots, n. \tag{1.2}$$

Here $f(\cdot)$ is a function of time (a weak trend function) and (1.2) captures potential drifts in the relationship between y_t and x_t over time. Such a modeling structure is especially useful for time series data over long horizons where economic mechanisms are likely to evolve and be subjected to changing institutional or regulatory conditions. For example, firms may change production processes in response to technological innovation and consumers may change consumption and savings behavior in response to new products and new banking regulations. These changes may be captured by temporal evolution in the coefficients through the functional dependence $f\left(\frac{t}{n}\right)$ in (1.2).

Nonparametric inference about time-varying parameters has received attention for modeling stationary or locally stationary time series data - see, for instance, Robinson (1989), Cai (2007), Li *et al* (2011), Chen and Hong (2012), and Zhang and Wu (2012). However, there is little literature on this topic for integrated or cointegrated time series. One exception is Park and Hahn (1999), who considered the time-varying parameter model (1.2) and used sieve methods to transform the nonlinear cointegrating equation to a linear approximation with a sieve basis of possibly diverging dimension. Their asymptotic theory can be seen as an extension of the work by Park and Phillips (1988).

This paper seeks to uncover evolution in the modeling framework for nonstationary time series over a long time horizon by using nonparametric kernel regression methods to estimate $f(\cdot)$. Our treatment shows that estimation of this model by conventional kernel methods encounters a degeneracy problem in the weighted signal matrix, which introduces a major new challenge in developing the limit theory. In fact, kernel degeneracy of this type can arise in many contexts where multivariate time-varying functions are associated with nonstationary regressors. The present literature appears to have overlooked the problem and existing mathematical tools fail to address it. The reason for degeneracy in the limiting weighted signal matrix is that kernel regression concentrates attention on a particular (time) coordinate, thereby fixing attention on a particular coordinate of f and the associated limit process of the regressor. In the multivariate case this focus on a single time coordinate produces a limiting signal matrix of deficient rank one whose zero eigenspace depends on the value of the limit process at that time coordinate. In other words, kernel degeneracy in the signal matrix is random and trajectory dependent.

The problem has two relatives in existing asymptotics. First, it is well understood that cointegrated (or commonly trending) regressors produce degeneracy in the limiting signal matrix of nonstationary data (Park and Phillips, 1988; Phillips, 1989). However, the kernel-induced degeneracy phenomenon is quite different because the null space is determined by the limiting trajectory of the data at the time coordinate of interest, whereas in the cointegrated regressor case the null space is a fixed cointegrating space - a space that by its very nature reduces variability and rank. A more closely related case of signal matrix degeneracy in econometrics occurs when nonstationary regressors have common explosive coefficients - see Phillips and Magdalinos (2008). In that case, the null space is the space orthogonal to the direction vector of the (limit of the standardized) exploding process. The null space is therefore random and determined by the trajectory of the data.

This paper introduces a novel method to accommodate the degeneracy in kernel limit theory. The method transforms coordinates to separate the directions of degeneracy and non-degeneracy and proceeds to establish the kernel limit theory in each of these directions. The asymptotics are fundamentally different from those in the existing literature. As intimated, the transformation is path dependent and local to the coordinate of concentration. Two different convergence rates are obtained for different directions (or combinations) of the multivariate nonparametric estimators, and both of the two rates are faster than the usual ($\sqrt{nh}$) rate of stationary kernel asymptotics. Thus, two types of super-consistency exist for the nonparametric kernel estimation of time-varying coefficient functions, which we refer to as type I and type II super-consistency. The higher rate of convergence ($n\sqrt{h}$) lies in the direction of the nonstationary regressor vector at the local coordinate point and exceeds the usual rate by $\sqrt{n}$ (type I). The lower rate (nh) lies in the degenerate direction but is still super-consistent (type II) for nonparametric estimators and exceeds the usual rate by $\sqrt{nh}$.

The above results are all obtained for the Nadaraya-Watson local level time varying coefficient regression in a cointegrating model. Similar results are shown to apply for local linear time-varying regression which assists in reducing asymptotic bias. The general case of endogenous cointegrating regression is also included in our framework and a fully modified (FM; Phillips and Hansen, 1990) kernel method is proposed to address the endogeneity of the nonstationary regressors. In the use of this method it is interesting to discover that the kernel estimators need to be modified through bias correction only in the degenerate direction as the limit distribution of the estimators is not affected by the possible endogeneity in the direction of the nonstationary regressor vector at the local coordinate point. The limit theory for FM kernel regression also requires new asymptotic results on the consistent estimation of long run covariance matrices, which in turn involve uniform consistency arguments because of the presence of nonparametric regression residuals in these estimates.

The remainder of the paper is organized as follows. Estimation methodology, some technicalities, and assumptions are given in Section 2. This section also introduces the kernel degeneracy problem, explains the phenomenon, and provides intuition for its resolution. Asymptotic properties of the nonparametric kernel estimator are developed in Section 3 with accompanying discussion. A kernel weighted FM regression method is proposed with attendant limit theory in Section 4. Section 5 reports a simulation study that explores the finite sample properties of the developed methods and theory, and Section 6 gives a practical application of the these time varying kernel regression methods to examine empirical relationships involving consumption, disposable income, investment and real interest rates. Section 7 concludes the paper. Proofs of the main theoretical results in the paper are given in Appendix A. Some supplementary technical materials are provided in Appendix B. A nonparametric specification test complete with its asymptotic theory is given in Appendix C.

2 Kernel estimation degeneracy

Set $\tau = \lfloor n\delta \rfloor$ where $\lfloor \cdot \rfloor$ denotes integer part and $\delta \in [0, 1]$ is the sample fraction corresponding to observation t. The functional response in (1.2) allows the regression coefficient to vary over time and kernel regression provides a convenient mechanism for fitting the function locally at a particular (time) coordinate, say $\tau = \lfloor n\delta \rfloor$. At this coordinate the coefficient is the vector $f\left(\lfloor n\delta \rfloor / n\right) \sim f\left(\delta\right)$ and the model response behaves locally around τ as $f\left(\delta\right)' x_{\lfloor n\delta \rfloor}$. Evolution in the response mechanism over time is therefore captured as δ changes through the functional dependence $f\left(\delta\right)' x_{\lfloor n\delta \rfloor}$.

Under certain smoothness conditions on f and for some fixed $\delta_0 \in (0,1)$ we have

$$f\left(\frac{t}{n}\right) = f(\delta_0) + O\left(\frac{t}{n} - \delta_0\right) \approx f(\delta_0)$$

when $\frac{t}{n}$ is in a small neighborhood of δ_0. The Nadaraya-Watson type local level regression estimator of $f(\delta_0)$ has the usual form given by

$$\widehat{f}_n(\delta_0) = \left[\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0)\right]^{+} \left[\sum_{t=1}^{n} x_t y_t K_{th}(\delta_0)\right], \quad K_{th}(\delta_0) = \frac{1}{h} K\left(\frac{\frac{t}{n} - \delta_0}{h}\right), \tag{2.1}$$

where A^{+} denotes the Moore-Penrose inverse of A, $K(\cdot)$ is some kernel function, and h is the bandwidth. Extensions to allow for multiple (distinct) coordinates $\{\delta_i : i = 1, ..., I\}$ of concentration are straightforward.

The weights $K_{th}(\delta_0)$ in the linear regression (2.1) ensure that the primary contributions to the signal matrix $\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0)$ come from observations in the immediate temporal neighborhood of τ. In general, we can expect there to be sufficient variation in x_t within this temporal neighborhood for the signal matrix $\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0)$ to be positive definite in finite samples, i.e. for fixed n and $h > 0$. In the case of stationary and independent generating mechanisms for x_t, the variation in x_t is also sufficient to ensure a positive definite limit as $n \to \infty$ and $h \to 0$ because the second moment matrix $\mathbb{E}(x_t x_t')$ may be assumed to be positive definite. However, in the nonstationary case where x_t converges weakly to a continuous stochastic process upon standardization, localizing the regression around a fixed point such as δ_0 reduces effective variability in the regressor when $n \to \infty$ because of continuity in the limit process and therefore leads to rank degeneracy in the limit of the signal matrix after standardization. The generalized inverse is employed in (2.1) for this reason. This limiting degeneracy in the weighted signal matrix challenges the usual approach to developing kernel asymptotics. As is apparent from the above explanation, limiting degeneracy of this type may be anticipated whenever kernel regression is conducted to fit multivariate time-varying functions that are associated with nonstationary regressors.

To develop the limit theory we start with some regularity conditions to characterize the nonstationary time series x_t and the (scalar) stationary error process u_t. We assume x_t is a unit root process with generating mechanism $x_t = x_{t-1} + v_t$, initial value $x_0 = O_P(1)$, and innovations jointly determined with the equation u_t error according to the linear process

$$w_t = (v_t', u_t)' = \Phi(\mathcal{L})\varepsilon_t = \sum_{j=0}^{\infty} \Phi_j \varepsilon_{t-j}, \tag{2.2}$$

where $\Phi(\mathcal{L}) = \sum_{j=0}^{\infty} \Phi_j \mathcal{L}^j$, Φ_j is a sequence of $(d+1) \times (d+1)$ matrices, $\mathcal{L}$ is the lag operator, and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (*iid*) random vectors with dimension $(d+1)$. Partition Φ_j as $\Phi_j = [\Phi_{j,1}, \ \Phi_{j,2}]'$ so that

$$v_t = \sum_{j=0}^{\infty} \Phi_{j,1}' \varepsilon_{t-j}, \quad \text{and } u_t = \sum_{j=0}^{\infty} \Phi_{j,2}' \varepsilon_{t-j}.$$

Assumption 1. *Let ε_t be iid $(d+1)$-dimensional random vectors with $\mathbb{E}[\varepsilon_t]=0$, $\Lambda_0 \equiv \mathbb{E}[\varepsilon_t\varepsilon_t'] > 0$, and $\mathbb{E}[\|\varepsilon_t\|^{4+\gamma_0}] < \infty$ for $\gamma_0 > 0$. The linear process coefficient matrices in (2.2) satisfy $\sum_{j=0}^{\infty} j\|\Phi_j\| < \infty$.*

By functional limit theory for a standardized linear process (Phillips and Solo, 1992), we have for $t=\lfloor nr \rfloor$ and $0 < r \le 1$,

$$\frac{x_t}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{s=1}^{t} v_s + \frac{1}{\sqrt{n}}x_0 = \frac{1}{\sqrt{n}}\sum_{s=1}^{\lfloor nr \rfloor} v_s + o_P(1) \Rightarrow B_{d,r}(\Omega_v), \tag{2.3}$$

$$n^{-1/2}\sum_{s=1}^{\lfloor nr \rfloor}\varepsilon_s \Rightarrow B_{\varepsilon,r}(\Lambda_0),\ n^{-1/2}\sum_{s=1}^{\lfloor nr \rfloor} w_s \Rightarrow B_{d+1,r}(\Omega),\ n^{-1/2}\sum_{s=1}^{\lfloor nr \rfloor} u_s \Rightarrow B_r(\Omega_u) \tag{2.4}$$

where $B_{d+1,r}(\Omega) = (B_{d,r}(\Omega_v)', B_r(\Omega_u))'$ is $(d+1)$-dimensional Brownian motion (BM) with variance matrix Ω, $B_{\varepsilon,r}(\Lambda_0)$ is $(d+1)$-dimensional BM with variance matrix Λ_0, and

$$\Omega = \Phi(1)'\Lambda_0\Phi(1) = \begin{bmatrix} \Phi_1(1)'\Lambda_0\Phi_1(1) & \Phi_1(1)'\Lambda_0\Phi_2(1) \\ \Phi_2(1)'\Lambda_0\Phi_1(1) & \Phi_2(1)'\Lambda_0\Phi_2(1) \end{bmatrix} \equiv \begin{bmatrix} \Omega_v & \Omega_{vu} \\ \Omega_{uv} & \Omega_u \end{bmatrix}, \tag{2.5}$$

with $\Phi(1) = \sum_{j=1}^{\infty}\Phi_j$, $\Phi_1(1) = \sum_{j=1}^{\infty}\Phi_{j,1}$, and $\Phi_2(1) = \sum_{j=1}^{\infty}\Phi_{j,2}$. Here Ω is the partitioned long run variance matrix of $w_t = (v_t', u_t)'$. The limit theory also involves the partitioned components of the one-sided long run variance matrix

$$\Delta_{ww} \equiv \begin{bmatrix} \Delta_{vv} & \Delta_{vu} \\ \Delta_{uv} & \Delta_{uu} \end{bmatrix} = \sum_{j=0}^{\infty}\mathbb{E}\left(w_{-j}w_0'\right).$$

It is convenient to impose a smoothness condition on the functional coefficient $f(\cdot)$ and some commonly-used conditions on the kernel function and bandwidth. Define $\mu_j = \int_{-1}^{1} u^j K(u)du$ and $\nu_j = \int_{-1}^{1} u^j K^2(u)du$.

Assumption 2. *$f(\cdot)$ is continuous with $|f(\delta_0+z)-f(\delta_0)| = O(|z|^{\gamma_1})$ as $z \to 0$ for some $\frac{1}{2} < \gamma_1 \le 1$.*

Assumption 3. *(i) The kernel function $K(\cdot)$ is continuous, positive, symmetric and has compact support $[-1,1]$ with $\mu_0 = 1$.*

(ii) The bandwidth h satisfies $h \to 0$ and $nh \to \infty$.

In the linear cointegration model with constant coefficients

$$y_t = x_t'\beta + u_t,\quad x_t = x_{t-1} + v_t,\quad t = 1,\cdots,n, \tag{2.6}$$

where v_t and u_t are generated by (2.2) and satisfy Assumption 1, least squares estimation of β gives $\widehat{\beta}_n = \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1}\left(\sum_{t=1}^{n} x_t y_t\right)$. Standard limit theory and super-consistency results for $\widehat{\beta}_n$ involve the following behavior of the signal matrix

$$n^{-2}\sum_{t=1}^{n} x_t x_t' = \frac{1}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}} \Rightarrow \int_0^1 B_{d,r}(\Omega_v)B_{d,r}(\Omega_v)'dr, \tag{2.7}$$

where the limit matrix is positive definite (Phillips and Hansen, 1990). By naive analogy to (2.7) it might be anticipated that the weighted signal matrix appearing in the denominator of the kernel estimator $\widehat{f}_n(\delta_0)$ would have similar properties. However, some simple derivations show this not to be the case, as we now demonstrate.

Take a neighborhood $N_{n,\delta_0}(h) = [\lfloor(\delta_0 - h)n\rfloor, \lfloor(\delta_0 + h)n\rfloor]$ of $\lfloor\delta_0 n\rfloor$ and let $\delta(n) = \lfloor(\delta_0 - h)n\rfloor$. The following representation of the weighted signal matrix is convenient in obtaining the limit behavior

$$\begin{aligned}\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0) &= \sum_{t=1}^{n} x_{\delta(n)} x_{\delta(n)}' K_{th}(\delta_0) + \sum_{t=1}^{n}\left(x_t - x_{\delta(n)}\right) x_{\delta(n)}' K_{th}(\delta_0) \\ &+ \sum_{t=1}^{n} x_{\delta(n)}\left(x_t - x_{\delta(n)}\right)' K_{th}(\delta_0) + \sum_{t=1}^{n}\left(x_t - x_{\delta(n)}\right)\left(x_t - x_{\delta(n)}\right)' K_{th}(\delta_0) \\ &\equiv U_{n1} + U_{n2} + U_{n3} + U_{n4}. \end{aligned} \tag{2.8}$$

Using the BN decomposition as in Phillips and Solo (1992), we have

$$x_t - x_{t-1} = v_t = \overline{v}_t + (\widetilde{v}_{t-1} - \widetilde{v}_t),$$

where $\overline{v}_t = \left(\sum_{j=0}^{\infty} \Phi_{j,1}'\right)\varepsilon_t$, and $\widetilde{v}_t = \sum_{j=0}^{\infty} \widetilde{\Phi}_{j,1}'\varepsilon_{t-j}$ with $\widetilde{\Phi}_{j,1} = \sum_{k=j+1}^{\infty} \Phi_{k,1}$. Then

$$x_{\delta(n)} = \sum_{t=1}^{\delta(n)} v_t + x_0 = \sum_{t=1}^{\delta(n)} \overline{v}_t + \widetilde{v}_0 - \widetilde{v}_{\delta(n)} + x_0. \tag{2.9}$$

By virtue of Assumption 1, we have

$$\begin{aligned}\frac{1}{\delta(n)}\left(\sum_{t=1}^{\delta(n)}\overline{v}_t\right)\left(\sum_{t=1}^{\delta(n)}\overline{v}_t\right)' &= \left(\sum_{j=0}^{\infty}\Phi_{j,1}'\right)\left[\frac{1}{\delta(n)}\left(\sum_{t=1}^{\delta(n)}\varepsilon_t\right)\left(\sum_{t=1}^{\delta(n)}\varepsilon_t'\right)\right]\left(\sum_{j=0}^{\infty}\Phi_{j,1}\right) \\ &\Rightarrow \Phi_1(1)'\mathcal{W}_{d+1}(\Lambda_0)\Phi_1(1), \end{aligned} \tag{2.10}$$

where $\mathcal{W}_{d+1}(\Lambda_0) = B_{\varepsilon,\delta_0}(\Lambda_0)B_{\varepsilon,\delta_0}(\Lambda_0)'$ is a Wishart variate with 1 degree of freedom and mean matrix Λ_0. Note that the summability condition $\sum_{j=0}^{\infty} j\|\Phi_j\| < \infty$ ensures $\sum_{j=0}^{\infty}\|\widetilde{\Phi}_j\| < \infty$ (Phillips and Solo, 1992), so that

$$\left(\widetilde{v}_0 - \widetilde{v}_{\delta(n)} + x_0\right)\left(\widetilde{v}_0 - \widetilde{v}_{\delta(n)} + x_0\right)' = O_P(1), \tag{2.11}$$

and then

$$\left(\sum_{t=1}^{\delta(n)}\overline{v}_t\right)\left(\widetilde{v}_0 - \widetilde{v}_{\delta(n)} + x_0\right)' = O_P(\sqrt{n}) = o_P(n). \tag{2.12}$$

On the other hand, by Assumption 3, we have $n^{-1}\sum_{t=1}^{n} K_{th}(\delta_0) \to \mu_0 = 1$ for $0 < \delta_0 < 1$ which, together with (2.9)–(2.12), implies that

$$\begin{aligned} n^{-2}U_{n1} &= \left(\frac{x_{\delta(n)}x_{\delta(n)}'}{n}\right)\left(\frac{1}{n}\sum_{t=1}^{n}K_{th}(\delta_0)\right) \\ &\Rightarrow \delta_0\Phi_1'(1)\mathcal{W}_{d+1}(\Lambda_0)\Phi_1(1). \end{aligned} \tag{2.13}$$

Next observe that for $t \in N_{n,\delta_0}(h) = [\delta(n), \delta(n) + 2\lfloor hn\rfloor]$ we have $x_t - x_{\delta(n)} = \sum_{s=\delta(n)+1}^{t} v_s$ and then

$$\sup_{t\in N_{n,\delta_0}(h)}\left\|\frac{x_t - x_{\delta(n)}}{\sqrt{2\lfloor nh\rfloor}}\right\| = \sup_{t\in N_{n,\delta_0}(h)}\left\|\frac{\sum_{s=\delta(n)+1}^{t} v_s}{\sqrt{2\lfloor nh\rfloor}}\right\| \Rightarrow \sup_{0<r<1}\|B_{d,r}(\Omega_v)\|, \tag{2.14}$$

where $B_{d,r}(\Omega_v)$ is the Brownian motion with covariance matrix Ω_v defined as in (2.3). Hence, for $h \to 0$ as $n \to \infty$ we have

$$\sup_{t \in N_{n,\delta_0}(h)} \left\| \frac{x_t - x_{\delta(n)}}{\sqrt{nh}} \right\| = O_P(1).$$

For U_{n2}, by Assumption 3 and the fact that $K(\cdot)$ has compact support, we find that

$$\begin{aligned} \|U_{n2}\| &\leq \left\|x_{\delta(n)}\right\| \sum_{t=[(\delta_0-h)n]+1}^{[(\delta_0+h)n]} K_{th}(\delta_0) \left\|x_t - x_{\delta(n)}\right\| \\ &= O_P\left(\sqrt{n}\right) \times O_P(n) \times O_P\left(\sqrt{nh}\right) \\ &= O_P\left(n^2 h^{1/2}\right) = o_P\left(n^2\right), \end{aligned} \tag{2.15}$$

where $\|\cdot\|$ denotes the Euclidean norm. Similarly,

$$\|U_{n3}\| = O_P\left(n^2 h^{1/2}\right) = o_P\left(n^2\right), \tag{2.16}$$

and

$$\|U_{n4}\| = O_P\left(n^2 h\right) = o_P\left(n^2\right). \tag{2.17}$$

In view of (2.8) and (2.13)–(2.17), we deduce that

$$n^{-2} \sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0) \Rightarrow \delta_0 \Phi_1'(1) \mathcal{W}_{d+1}(\Lambda_0) \Phi_1(1), \tag{2.18}$$

which is the limiting signal matrix analogue of (2.7) in the case of nonparametric kernel-weighted least squares. On inspection, the $d \times d$ limit matrix $\Phi_1'(1)\mathcal{W}_{d+1}(\Lambda_0)\Phi_1(1)$ in (2.18) is singular with rank one when $d > 1$. The weighted signal matrix $n^{-2}\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0)$ is therefore asymptotically singular whenever the dimension of the regressor x_t exceeds unity.

The intuition for this limiting degeneracy in the signal matrix is that kernel regression concentrates attention on the time coordinate δ_0 and thereby the realized value of the limit process $B_{d,\delta_0}(\Omega_v)$ of the (standardized) regressor x_t. When x_t is multivariate, this focus on the realization $B_{d,\delta_0}(\Omega_v)$ of the limit process of $n^{-1/2}x_t$ produces a limiting signal matrix of the outer product form $B_{d,\delta_0}(\Omega_v)B_{d,\delta_0}(\Omega_v)'$. In effect, continuity of the limit process $B_{d,r}(\Omega_v)$ ensures that in any shrinking neighborhood of the coordinate δ_0, weighted kernel regression concentrates the signal toward the quantity $B_{d,\delta_0}(\Omega_v)B_{d,\delta_0}(\Omega_v)'$ - as if there were only a single observation of x_t in the limit. Importantly, the limiting form of the weighted signal matrix depends on the realized value $B_{d,\delta_0}(\Omega_v)$ of the limit process at the time coordinate δ_0. So, the kernel degeneracy is random and trajectory dependent.

As pointed out in the Introduction, this phenomenon has two relatives in existing asymptotic theory but seems not before to have arisen in kernel asymptotics. The first relative is a nonstationary linear regression model with many trending and/or cointegrated regressors. In such models the limiting signal matrix of the nonstationary data is degenerate to the extent that the trends do not have full rank - see Park and Phillips (1988) and Phillips (1989). However, in such cases the null space of the limiting signal matrix is a fixed space determined by the parameters that define the direction of the trends and the stochastic nonstationarity and cointegration. The second relative in econometrics occurs in models with nonstationary regressors that have common explosive coefficients - see Phillips and Magdalinos (2008, 2013). Such models can be cointegrated systems with co-moving explosive regressors or vector autoregressions with common explosive roots. In these cases, the null space of the limiting signal matrix is determined by the direction vector of the (limit of the standardized) exploding process and is therefore random and trajectory dependent, as in the present case.

The following section shows how to transform the coordinate system to accommodate the degeneracy and develop limit theory for the kernel regression estimator. This limit theory is operational for practical implementation. However, the asymptotics turn out to be fundamentally different from those in the existing kernel regression literature. Also, unlike the asymptotic theory for linear models with degenerate limits discussed in the last paragraph where the degenerate directions typically have stationary asymptotics with Gaussian limit distributions and conventional $\sqrt{n}$ convergence rates apply, in the kernel regression case both the degenerate and nondegenerate directions give super-consistent estimation and nonstandard asymptotics. Nonstationary kernel regression limit theory therefore has some unusual and rather unexpected properties in the degenerate case induced by time varying coefficient functions.

3 Large sample theory

To simplify presentation define $b \equiv b_{\delta_0} = B_{d,\delta_0}(\Omega_v)$ and set

$$q = \frac{b}{(b'b)^{1/2}} = \frac{b}{\|b\|}.$$

Let $q^{\perp}$ be a $d \times (d-1)$ orthogonal complement matrix such that

$$Q = \left[q, q^{\perp}\right], \quad Q'Q = I_d, \tag{3.1}$$

where I_d is the $d \times d$ identity matrix. Correspondingly, we define the following sample versions of these quantities

$$q_n = \frac{b_n}{(b_n' b_n)^{1/2}} = \frac{b_n}{\|b_n\|}, \quad b_n \equiv b_{n\delta_0} = \frac{1}{\sqrt{n}} x_{\delta(n)},$$

let

$$Q_n = \left[q_n, q_n^{\perp}\right], \quad Q_n' Q_n = I_d, \tag{3.2}$$

and introduce the standardization matrix

$$D_n = \mathrm{diag}\big(n\sqrt{h},\ (nh) I_{d-1}\big). \tag{3.3}$$

The matrices Q and Q_n are random, path dependent, and localized to the coordinate of concentration (at δ_0 and $\delta(n) = \lfloor (\delta_0 - h)\, n \rfloor$, respectively). Write $B_{d+1,r}(\Omega) = \Big[B_{d,r}'(\Omega_v),\ B_r(\Omega_u)\Big]'$ and define

$$\Delta_{\delta_0} = \begin{bmatrix} \Delta_{\delta_0}(1) & \Delta_{\delta_0}(2) \\ \Delta_{\delta_0}(2)' & \Delta_{\delta_0}(3) \end{bmatrix}, \quad \Gamma_{\delta_0} = \begin{bmatrix} \Gamma_{\delta_0}(1) \\ \Gamma_{\delta_0}(2) \end{bmatrix}, \tag{3.4}$$

where the components of the partition are

$$
\begin{aligned}
\Delta_{\delta_0}(1) &= b'b, \\
\Delta_{\delta_0}(2) &= 2\sqrt{2}\left(b'b\right)^{1/2}\left\{\int_{-1}^{1} B_{d,\frac{r+1}{2},*}(\Omega_v)K(r)dr\right\} q^{\perp}, \\
\Delta_{\delta_0}(3) &= 4(q^{\perp})'\left\{\int_{-1}^{1} B_{d,\frac{r+1}{2},*}(\Omega_v)B_{d,\frac{r+1}{2},*}(\Omega_v)'K(r)dr\right\} q^{\perp}, \\
\Gamma_{\delta_0}(1) &= \left(b'b\right)^{1/2}\mathcal{Z}_u^*, \\
\Gamma_{\delta_0}(2) &= 2(q^{\perp})'\left\{\int_{-1}^{1} K(r)B_{d,\frac{r+1}{2},*}(\Omega_v)dB_{\frac{r+1}{2},*}(\Omega_u) + \frac{1}{2}\Delta_{vu}\right\},
\end{aligned}
$$

where $\mathcal{Z}_u^*$ is $\mathrm{N}\left(0, \nu_0\Omega_u\right)$ and independent of $B_{d,\delta_0}(\Omega_v)$, $B_{d,r,*}(\Omega_v)$ is an independent copy of the d-dimensional Brownian motion $B_{d,r}(\Omega_v)$, and $B_{r,*}(\Omega_u)$ is an independent copy of the Brownian motion $B_r(\Omega_u)$. The limit variate $\mathcal{Z}_u^*$ may be correlated with $B_{d,r,*}(\Omega_v)$ under endogeneity of the regressor x_t. The following theorem gives the asymptotic distribution of $\widehat{f}_n(\delta_0)$.

Theorem 3.1. *Suppose Assumptions 1–3 are satisfied and* $n^2h^{1+2\gamma_1} = o(1)$. *Then as* $n \to \infty$

$$
D_n Q_n'\left\{\widehat{f}_n(\delta_0) - f(\delta_0)\right\} \Rightarrow \Delta_{\delta_0}^{+}\Gamma_{\delta_0}, \tag{3.5}
$$

where δ_0 *is fixed* $0 < \delta_0 < 1$ *such that* Δ_{δ_0} *is nonsingular with probability* 1.

From the definition of D_n and (3.5), different convergence rates apply for the directions q_n and $q_n^{\perp}$. In the direction of q_n we have the faster convergence rate given by

$$
q_n'\left\{\widehat{f}_n(\delta_0) - f(\delta_0)\right\} = O_P\left(\frac{1}{n\sqrt{h}}\right). \tag{3.6}
$$

The rate (3.6) exceeds the usual $O_P\left(\sqrt{nh}\right)$ rate for kernel estimators in the stationary case. The $O_P\left(n\sqrt{h}\right)$ rate in (3.6) can be understood as $O_P\left(\sqrt{n^2h}\right)$ so that the effective sample size for estimating $q'f(\delta_0)$ is n^2h, as determined by the signal matrix behavior in this direction, rather than nh. Note that in unstandardized form the signal matrix is $\sum_{t=1}^{n} x_t x_t' K(\frac{t/n-\delta_0}{n})$ which is $O_P\left(n^2h\right)$ by virtue of (2.13) and (2.18). This signal matrix is rank degenerate in the limit. But in the direction q_n we have the non-degenerate signal

$$
q_n'\left\{\sum_{t=1}^{n} x_t x_t' K\left(\frac{\frac{t}{n}-\delta_0}{h}\right)\right\} q_n = O_P\left(n^2h\right).
$$

The replacement of n by n^2 in determining the convergence rate in the nonstationary direction q_n is the result of the stronger signal in the data about the specific component $q'f(\delta_0)$ of the unknown function $f(\delta_0)$ in the direction q. We call this result *type I* super-consistency. The convergence rate $O_P\left(\sqrt{n^2h}\right)$ was obtained by Cai *et al* (2009) and Xiao (2009) in certain functional-coefficient models with multivariate nonstationary regressors and no degeneracies. Type I super-consistency in functional coefficient kernel regression corroborates intuitive ideas from linear parametric models about the additional information in the data about the coefficients of stochastic trends in the direction of those trends.

In the direction of $q_n^{\perp}$, (3.5) gives

$$(q_n^{\perp})'\left\{\widehat{f}_n(\delta_0)-f(\delta_0)\right\}=O_P\left(\frac{1}{nh}\right). \tag{3.7}$$

Interestingly, this rate also exceeds the usual $O_P\left(\sqrt{nh}\right)$ rate for kernel estimators in stationary models. But convergence in the direction $q_n^{\perp}$ is slower than in direction q_n. We call the result in (3.7) *type II* super-consistency. This rate is new to the kernel regression literature. In a functional coefficient cointegrating regression the result indicates that nonstationarity in the regressors increases the rate of convergence in **all** directions, including the components $(q^{\perp})'f(\delta_0)$ of $f(\delta_0)$ in directions that are orthogonal to those of the nonstationary regressor. The reason why the rate exceeds the usual $O_P(\sqrt{nh})$ rate for stationary regression is that the signal in the direction $q_n^{\perp}$ is still stronger than that of a stationary regressor. This feature of the signal is explained by the fact that the signal matrix has order $O_P\left(n^2h^2\right)$ in this direction, viz.,

$$\begin{aligned}(q_n^{\perp})'\left\{\sum_{t=1}^{n}x_tx_t'K\left(\frac{\frac{t}{n}-\delta_0}{h}\right)\right\}q_n^{\perp} &= (q_n^{\perp})'\left\{\sum_{t=1}^{n}\left(x_t-x_{\delta(n)}\right)\left(x_t-x_{\delta(n)}\right)'K\left(\frac{\frac{t}{n}-\delta_0}{h}\right)\right\}q_n^{\perp}\\ &= O_P\left(n^2h^2\right).\end{aligned}$$

So the effective sample size in the estimation of the component $(q^{\perp})'f(\delta_0)$ is n^2h^2, which is smaller than the effective sample size n^2h that applies for estimation of $q'f(\delta_0)$. More specifically, under the compact support condition on the kernel function (as given in Assumption 3), estimation of $(q^{\perp})'f(\delta_0)$ only uses information on $x_t-x_{\delta(n)}$ over the interval of observations $N_{n,\delta_0}(h)=[\lfloor n\delta_0-nh\rfloor,\lfloor\delta_0+nh\rfloor]$. So, the number of observations contributing to nonparametric kernel estimation of $(q^{\perp})'f(\delta_0)$ is only of the order of nh. However, over this interval for $t=\delta(n)+\lfloor 2nhp\rfloor\in N_{n,\delta_0}(h)$ with $p\in[0,1]$ the data increments still manifest nonstationary characteristics. In particular, we have the following weak convergence (c.f. (2.14))

$$\frac{x_t-x_{\delta(n)}}{\sqrt{2\lfloor nh\rfloor}}=\frac{\sum_{s=\delta(n)+1}^{\lfloor 2nhp\rfloor}v_s}{\sqrt{2\lfloor nh\rfloor}}\Rightarrow B_{d,p}(\Omega_v). \tag{3.8}$$

The stronger signal in these observations raises the overall signal in $(q_n^{\perp})'\left\{\sum_{t=1}^{n}x_tx_t'K\left(\frac{\frac{t}{n}-\delta_0}{h}\right)\right\}q_n^{\perp}$ to $O_P\left((\sqrt{nh})^2\right)\times O_P(nh)=O_P\left(n^2h^2\right)$, as distinct from the $O_P(nh)$ signal in conventional stationary kernel regression case. Thus, local nonstationarity in the data around $\lfloor n\delta_0\rfloor$ contributes to greater information about $(q^{\perp})'f(\delta_0)$ than would occur in a stationary kernel regression.

In the pure cointegration case with $\Delta_{vu}=0$, the form of $\Gamma_{\delta_0}(2)$ can be simplified. Define $\overline{\Gamma}_{\delta_0}(2)=2(q^{\perp})'\left\{\int_{-1}^{1}B_{d,\frac{r+1}{2},*}(\Omega_v)dB_{\frac{r+1}{2},*}(\Omega_u)\right\}$ and $\overline{\Gamma}_{\delta_0}$ just as Γ_{δ_0} but with $\Gamma_{\delta_0}(2)$ replaced by $\overline{\Gamma}_{\delta_0}(2)$. The following limit theory applies in this pure cointegration case.

Corollary 3.1. *Suppose that the conditions in Theorem 3.1 are satisfied and $\Delta_{vu}=0$. We then have*

$$D_nQ_n'\left\{\widehat{f}_n(\delta_0)-f(\delta_0)\right\}\Rightarrow\Delta_{\delta_0}^{+}\overline{\Gamma}_{\delta_0}, \tag{3.9}$$

for fixed $0<\delta_0<1$ such that Δ_{δ_0} is nonsingular with probability 1.

To eliminate bias effects in these nonparametric asymptotics we have imposed the bandwidth condition $n^2h^{1+2\gamma_1}=o(1)$ on the bandwidth, which may be somewhat restrictive if γ_1 is close to its lower boundary of $\frac{1}{2}$ (Assumption 2). To relax the restriction in such cases, a higher order kernel function may be considered (e.g., Wand and Jones, 1994) or local polynomial smoothing (e.g., Fan and Gijbels, 1996) can be used. Local linear regression is the most commonly-used local polynomial smoothing method in practical work and has certain advantages over local level regression in stationary regression, although Wang and Phillips (2009b,2011) showed that such bias reduction with local linear methods does not occur (and hence is not an advantage) in nonstationary nonparametric regression.

Assume f has continuous derivatives up to the second order. Then, for fixed $0 < \delta_0 < 1$, the following local linear approximation holds when $\frac{t}{n}$ is in a small neighborhood of δ_0,

$$f\left(\frac{t}{n}\right) = f(\delta_0) + f^{(1)}(\delta_0)\left(\frac{t}{n} - \delta_0\right) + O\left(\left(\frac{t}{n} - \delta_0\right)^2\right),$$

where $f^{(1)}(\delta_0)$ is the first-order derivative of f at δ_0. Define the local loss function

$$L_n(a,b) = \sum_{t=1}^{n}\left[y_t - x_t'a - x_t'b\left(\frac{t}{n} - \delta_0\right)\right]^2 K_{th}(\delta_0), \tag{3.10}$$

where $a = (a_1, \cdots, a_d)'$ and $b = (b_1, \cdots, b_d)'$. The local linear estimator of $f(\delta_0)$ is defined as $\widetilde{f}_n(\delta_0) = \widetilde{a}$, where $(\widetilde{a}, \widetilde{b}) = \arg\min_{(a,b)} L_n(a,b)$. Set

$$\Delta_{\delta_0 *} = \begin{bmatrix} \Delta_{\delta_0 *}(1) & \Delta_{\delta_0 *}(2) \\ \Delta_{\delta_0 *}(2)' & \Delta_{\delta_0 *}(3) \end{bmatrix}, \quad \Gamma_{\delta_0 *} = \begin{bmatrix} \Gamma_{\delta_0 *}(1) \\ \Gamma_{\delta_0 *}(2) \end{bmatrix},$$

where $\Delta_{\delta_0 *}(1) = \Delta_{\delta_0}$, $\Gamma_{\delta_0 *}(1) = \Gamma_{\delta_0}$, $\Delta_{\delta_0 *}(2)$ and $\Delta_{\delta_0 *}(3)$ are defined as in Δ_{δ_0} but with $K(r)$ replaced by $rK(r)$ and $r^2K(r)$, respectively, and $\Gamma_{\delta_0 *}(2)$ is defined as Γ_{δ_0} with $K(r)$ replaced by $rK(r)$. Let $e_d = (I_d,\ O_d)$, where O_d is a $d \times d$ null matrix. The limit theory for the local linear estimator $\widetilde{f}_n(\delta_0)$ is given in the following theorem.

Theorem 3.2. *Suppose that Assumptions 1 and 3 in Section 2 are satisfied and $f(\cdot)$ has continuous derivatives up to the second order. Let δ_0 be fixed $0 < \delta_0 < 1$ such that $\Delta_{\delta_0 *}$ is nonsingular with probability 1. Then, we have*

$$D_n Q_n' \left\{\widetilde{f}_n(\delta_0) - f(\delta_0) + O_P(h^2)\right\} \Rightarrow e_d\ \Delta_{\delta_0 *}^{+}\Gamma_{\delta_0 *}. \tag{3.11}$$

Furthermore, if $n^2h^5 = o(1)$, we have

$$D_n Q_n' \left\{\widetilde{f}_n(\delta_0) - f(\delta_0)\right\} \Rightarrow e_d\ \Delta_{\delta_0 *}^{+}\Gamma_{\delta_0 *}. \tag{3.12}$$

Just as in the case of Theorem 3.1, types I and II super-consistency apply to the local linear estimator $\widetilde{f}_n(\delta_0)$ according to the directions q_n and $q_n^{\perp}$. The results are entirely analogous, so the details are omitted.

4 FM-nonparametric kernel estimation

The one sided long run covariance Δ_{vu} which appears in the limit functionals Γ_{δ_0} and $\Gamma_{\delta_0 *}$ of Theorems 3.1 and 3.2 induces a "second-order" bias effect just like the bias that appears in linear cointegrating regression limit theory (Park and Phillips, 1988, 1989). The bias effect originates in the correlation between the regressor innovations and the equation error. It is a second order effect, so the two super-consistency rates of the kernel estimator of the functional coefficient shown in Section 3 are unchanged. But, as in the linear cointegration model with constant coefficients, the bias does influence centering of the limit distributions. So its effect can be substantial in finite samples, as is well known in the linear constant coefficient case. This section therefore develops a nonparametric kernel version of the FM regression technique (Phillips and Hansen, 1990) to eliminate the bias effect in this nonstationary case. Although there has been extensive study of this type of correction in linear cointegration models, to the best of our knowledge there is no work on techniques of bias correction for nonparametric kernel estimation of time-varying cointegration models.

Let $\widehat{\Delta}_{vu}$ denote a consistent estimate of Δ_{vu}, whose construction will be considered later in this section. We define the "bias-corrected" FM kernel regression estimator of the functional coefficient $f(\cdot)$ as

$$\widehat{f}_{n,bc}(\delta_0) = \left[\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0)\right]^{+} \left[\sum_{t=1}^{n} x_t y_t K_{th}(\delta_0) - Q_n D_n \widehat{\Gamma}_{n,bc}\right] \tag{4.1}$$

with

$$\widehat{\Gamma}_{n,bc} = \left(0, \left[(q_n^{\perp})' \widehat{\Delta}_{vu}\right]'\right)'. \tag{4.2}$$

Since $\widehat{\Delta}_{vu} = \Delta_{vu} + o_P(1)$, the asymptotic distribution of $\widehat{f}_{n,bc}(\delta_0)$ is obtained in the same manner as the proof of Theorem 3.1 and is the same as that of $\widehat{f}_n(\delta_0)$ in the pure cointegration case shown in Corollary 3.1.

PROPOSITION 4.1. *Suppose that the conditions in Theorem 3.1 are satisfied. We then have*

$$D_n Q_n' \left\{\widehat{f}_{n,bc}(\delta_0) - f(\delta_0)\right\} \Rightarrow \Delta_{\delta_0}^{+} \overline{\Gamma}_{\delta_0} \tag{4.3}$$

for fixed $0 < \delta_0 < 1$*, where* $\overline{\Gamma}_{\delta_0}$ *is defined as in Corollary 3.1.*

From (4.1) and (4.3), it is evident that the bias term of the nonparametric kernel estimator needs only to be corrected in the direction $q_n^{\perp}$, since the limit distribution in the direction q_n remains the same irrespective of whether endogeneity is present. This bias correction technique may similarly be applied to the local linear estimator. Since the derivations and results are the same, the details are omitted.

Practical implementation of FM-nonparametric kernel regression requires estimation of the one-sided long-run covariance matrix Δ_{vu}. The usual approach may be followed here. Let $\widehat{u}_t = y_t - x_t' \widehat{f}_n(t/n)$ be the estimated residuals from applying kernel regression to (1.2). Let $0 < \tau_* < 1/2$, which can be arbitrarily small. Since $v_t = x_t - x_{t-1}$, we may construct the estimated autocovariances

$$\widehat{\Delta}_{vu}(j) = \frac{1}{\lfloor (1-\tau_*)n \rfloor - \lfloor \tau_* n \rfloor} \sum_{t=\lfloor \tau_* n \rfloor + 1}^{\lfloor (1-\tau_*)n \rfloor} v_{t-j} \widehat{u}_t, \quad j = 0, 1, \cdots, l_n, \tag{4.4}$$

which are combined to produce the one-sided long-run covariance estimate

$$\widehat{\Delta}_{vu} = \sum_{j=0}^{l_n} k\left(\frac{j}{l_n}\right)\widehat{\Delta}_{vu}(j), \tag{4.5}$$

where $k(\cdot)$ is a kernel function and $l_n < n$ is the lag truncation number which tends to infinity as $n \to \infty$. To ensure the consistency of $\widehat{\Delta}_{vu}$, the kernel function $k(\cdot)$ is assumed to be bounded with $k(0) = 1$ and $k(-x) = k(x)$ such that $\int_{-1}^{1} k^2(x) < \infty$ and $\lim_{x\to 0} \frac{1-k(x)}{|x|} < \infty$ (e.g. Park and Hahn, 1999). The choice of the truncation number l_n has been discussed in detail in the existing literature on FM regression (e.g. Phillips, 1995).

To avoid possible boundary effects from kernel estimation in the estimated autocovariogram in (4.4), we use only information on $v_{t-j}\widehat{u}_t$ from $\lfloor \tau_* n\rfloor + 1$ to $\lfloor (1-\tau_*)n\rfloor$. This construction differs from usual practice in parametric linear cointegration models where $v_{t-j}\widehat{u}_t$ is summed over the full domain $(j+1, n)$ to estimate the covariance. However, as is evident intuitively and shown rigorously in the proof of Proposition 4.2 in the Appendix, for τ_* close to zero this modification does not affect the asymptotic analysis. In the context of parametric cointegration models, the proof of consistency of $\widehat{\Delta}_{vu}$ is straightforward because the quantities $\widehat{\Delta}_{vu}(j)$ rely on the estimates of residuals that are obtained from coefficients estimated at parametric rates. In the present nonparametric case, kernel methods are used to estimate the time-varying coefficient functions, which in turn complicates the form of the estimated residuals and makes the proof of consistency much more difficult. A particular difficulty in the nonparametric case is that conditions are needed to ensure the nonsingularity of the random denominator of the local level regression estimator $\widehat{f}_n(\delta)$ uniformly over $\delta \in [\tau_*,\ 1-\tau_*]$ for any $0 < \tau_* < 1/2$. The following proposition establishes the consistency of $\widehat{\Delta}_{vu}$ defined in (4.5).

Proposition 4.2. *Suppose that the conditions in Theorem 3.1 are satisfied,* $l_n^{10+2\gamma_0+\varpi} = o(n^{5+\gamma_0}h^{9+\gamma_0})$ *for arbitrarily small* $\varpi > 0$, $l_n = o\left(\frac{1}{\sqrt{nh}}\right)$, *and the random matrix* Δ_δ *is nonsingular uniformly for* $\delta \in [\tau_*,\ 1-\tau_*]$ *with probability* 1 *for any* $0 < \tau_* < 1/2$. *We then have*

$$\widehat{\Delta}_{vu} = \Delta_{vu} + o_P(1). \tag{4.6}$$

The condition $l_n^{10+2\gamma_0+\varpi} = o(n^{5+\gamma_0}h^{9+\gamma_0})$ indicates a trade-off between the restriction on the truncation number l_n and the moment condition on the ε_i. In particular, for γ_0 large enough, we find that the imposed condition is close to $l_n = o(\sqrt{nh})$, which allows the truncation number to increase at a polynomial rate. On the other hand, the restriction $l_n = o\big(\frac{1}{\sqrt{nh}}\big)$ ensures that the asymptotic bias of the kernel estimates does not affect the consistency of $\widehat{\Delta}_{vu}$.

5 Simulations

This section reports simulations designed to investigate the finite sample performance of kernel estimation in multivariate nonstationary settings and examines the adequacy of the asymptotic theory developed earlier in the paper. We are particularly interested in the behavior of multivariate time-varying coefficient function estimators, respective convergence rates, and the effects of endogeneity and serial dependence on these procedures.

EXAMPLE 5.1. We consider a cointegrated system with time-varying coefficient functions

$$y_t = x_t' f_t + u_t, \quad t = 1, \cdots, n, \tag{5.1}$$

where $f_t = (f_{1t}, f_{2t})'$ has the following two functional forms: *M1*: $f_{1t} = f_1\left(\frac{t}{n}\right) = 1 + \frac{t}{n}$ and $f_{2t} = f_2\left(\frac{t}{n}\right) = e^{-\frac{t}{n}}$, and *M2*: $f_{1t} = f_1\left(\frac{t}{n}\right) = \cos\left(\frac{2\pi t}{n}\right)$ and $f_{2t} = f_2\left(\frac{t}{n}\right) = \sin\left(\frac{2\pi t}{n}\right)$, $x_t = (x_{1t}, x_{2t})'$, $x_{i,t} = x_{i,t-1} + v_{i,t}$ for $i = 1$ and 2, $v_{i,t} = \rho_i v_{i,t-1} + \varepsilon_{i,t}$, $u_t = \rho u_{t-1} + \varepsilon_t$, and $(\varepsilon_t, \varepsilon_{1,t}, \varepsilon_{2,t})$ follows

$$\begin{pmatrix} \varepsilon_t \\ \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \overset{iid}{\sim} \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \lambda_1 & \lambda_2 \\ \lambda_1 & 1 & \lambda_3 \\ \lambda_2 & \lambda_3 & 1 \end{pmatrix} \right), \tag{5.2}$$

with $\lambda_i = 0$ or $\lambda_i = 0.5$ for $i = 1, 2$ and 3. The simulation is conducted with sample size $n = 1{,}000$ and $N = 10{,}000$ replications.

The nonparametric kernel estimate of $f(\delta) = \left[f_1(\delta), f_2(\delta)\right]'$ is given by

$$\widehat{f}_n(\delta) = \Big[\sum_{t=1}^{n} x_t x_t' K\big(\frac{t - n\delta}{nh}\big)\Big]^+ \sum_{t=1}^{n} x_t y_t K\big(\frac{t - n\delta}{nh}\big) \equiv \big[\widehat{f}_{1n}(\delta), \widehat{f}_{2n}(\delta)\big]', \tag{5.3}$$

where we use $K(x) = \frac{1}{2} I\{-1 \leq x \leq 1\}$ and choose some possible bandwidth values for h which will be specified later. Before reporting the simulation results, we use the following notation, based partly on earlier definitions. Let $\delta(n) = \lfloor (\delta - h)n \rfloor$, $x_{\delta(n)} = \left(x_{1,\delta(n)}, x_{2,\delta(n)}\right)'$, $b_n(\delta) = \frac{1}{\sqrt{n}} x_{\delta(n)} = \frac{1}{\sqrt{n}}\left(x_{1,\delta(n)}, x_{2,\delta(n)}\right)'$ and $q_n(\delta) = b_n(\delta)/\|b_n(\delta)\| = \left[\frac{x_{1,\delta(n)}}{\sqrt{n}\|b_n(\delta)\|}, \frac{x_{2,\delta(n)}}{\sqrt{n}\|b_n(\delta)\|}\right]' \equiv \left[q_{1n}(\delta), q_{2n}(\delta)\right]'$ with $\|b_n(\delta)\| = \sqrt{\frac{1}{n}[x_{1,\delta(n)}^2 + x_{2,\delta(n)}^2]}$. Let $q_n^{\perp}(\delta) = \left[p_{1n}(\delta), p_{2n}(\delta)\right]'$ be chosen such that $Q_n(\delta) = \left[q_n(\delta), q_n^{\perp}(\delta)\right]$ and $Q_n(\delta)' Q_n(\delta) = I_2$. For this purpose we set $p_{1n} = q_{2n}$ and $p_{2n} = -q_{1n}$.

To evaluate the finite sample performance of the proposed estimators, we introduce the following transformed quantities

$$g_{1n}(\delta) = q_{1n}(\delta)\big[\widehat{f}_{1n}(\delta) - f_1(\delta)\big] + q_{2n}(\delta)\big[\widehat{f}_{2n}(\delta) - f_2(\delta)\big], \tag{5.4}$$

$$g_{2n}(\delta) = p_{1n}(\delta)\big[\widehat{f}_{1n}(\delta) - f_1(\delta)\big] + p_{2n}(\delta)\big[\widehat{f}_{2n}(\delta) - f_2(\delta)\big], \tag{5.5}$$

and compute averages of $g_{1n}(\delta)$ and $g_{2n}(\delta)$ as follows: $g_{in}(\delta) = \frac{1}{N}\sum_{j=1}^{N} g_{in,j}(\delta)$ for $i = 1, 2$ and $N = 10{,}000$, where $g_{in,j}(\delta)$ is the value of $g_{in}(\delta)$ at the j-th replication.

Corresponding results are investigated for the bias-corrected FM kernel regression estimator proposed in equation (4.1) above. Accordingly, we define

$$\begin{aligned} g_{1n}^*(\delta) &= q_{1n}\left(\widehat{f}_{1n,bc}(\delta) - f_1(\delta)\right) + q_{2n}\left(\widehat{f}_{2n,bc}(\delta) - f_2(\delta)\right), \\ g_{2n}^*(\delta) &= p_{1n}\left(\widehat{f}_{1n,bc}(\delta) - f_1(\delta)\right) + p_{2n}\left(\widehat{f}_{2n,bc}(\delta) - f_2(\delta)\right), \end{aligned} \tag{5.6}$$

where $\widehat{f}_{n,bc}(\cdot) = \left[\widehat{f}_{1n,bc}(\cdot), \widehat{f}_{2n,bc}(\cdot)\right]'$ is as defined in (4.1), in which $\widehat{\Delta}_{vu}$ is constructed by equations (4.4) and (4.5) with $\tau^* = \frac{1}{4}$, $k(x) = 1I[|x| \leq 1]$ and $l_n = \left\lfloor \frac{1}{(\sqrt{n}h)\log(n)} \right\rfloor$. This last setting implies that $l_n = o\left(\frac{1}{\sqrt{n}h}\right)$ and the precise expansion rate of l_n depends on h and the restriction on the bandwidth that $\sqrt{n}h \to 0$. As shown in Theorem 3.2, one may relax the condition on h to just ensure that $n^2h^5 = o(1)$ through using the local linear kernel method. In consequence, there is the opportunity of a data–driven version for l_n, a prospect that we leave for future exploration.

Averages of $g^*_{1n}(\delta)$ and $g^*_{2n}(\delta)$ are computed as follows: $g^*_{in}(\delta) = N^{-1}\sum_{j=1}^{N} g^*_{in,j}(\delta)$ for $i = 1, 2$ and $N = 10,000$, where $g^*_{in,j}(\delta)$ is the value of $g^*_{in}(\delta)$ at the j-th replication. The simulation results of point-wise kernel estimation are reported in Tables 5.1 and 5.2, which consider six different parameter constellations for $\{\rho, \rho_i, \lambda_i, (\delta, h)\}$:

$$\text{Case } 1: \rho = \rho_1 = \rho_2 = 0,\ \lambda_1 = \lambda_2 = \lambda_3 = 0,\ (\delta, h) = \left(\frac{1}{4}, \frac{1}{6}\right);$$
$$\text{Case } 2: \ \rho = \rho_1 = \rho_2 = 0,\ \lambda_1 = \lambda_2 = \lambda_3 = 0,\ (\delta, h) = \left(\frac{1}{2}, \frac{1}{3}\right);$$
$$\text{Case } 3: \ \rho = \rho_1 = \rho_2 = 0,\ \lambda_1 = \lambda_2 = \lambda_3 = 0,\ (\delta, h) = \left(\frac{3}{4}, \frac{1}{2}\right);$$
$$\text{Case } 4: \ \rho = 0.5,\ \rho_1 = -0.5,\ \rho_2 = 0.5,\ \lambda_1 = \lambda_2 = \lambda_3 = 0.5,\ (\delta, h) = \left(\frac{1}{4}, \frac{1}{6}\right);$$
$$\text{Case } 5: \ \rho = 0.5,\ \rho_1 = -0.5,\ \rho_2 = 0.5,\ \lambda_1 = \lambda_2 = \lambda_3 = 0.5,\ (\delta, h) = \left(\frac{1}{2}, \frac{1}{3}\right);$$
$$\text{Case } 6: \ \rho = 0.5,\ \rho_1 = -0.5,\ \rho_2 = 0.5,\ \lambda_1 = \lambda_2 = \lambda_3 = 0.5,\ (\delta, h) = \left(\frac{3}{4}, \frac{1}{2}\right).$$

Table 5.1: Absolute averages of $g_{in}(\delta)$ and $g^*_{in}(\delta)$ for
M1: $f_{1t} = 1 + \frac{t}{n}$ and $f_{2t} = e^{-\frac{t}{n}}$

Case 1		Case 2		Case 3	
$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$	$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$	$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$
0.005279	0.007294	0.002083	0.016241	0.001607	0.005815
Case 4		Case 5		Case 6	
$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$	$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$	$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$
0.000895	0.004268	0.000816	0.000458	0.000399	0.011452
$\lvert g^*_{1n}(\delta)\rvert$	$\lvert g^*_{2n}(\delta)\rvert$	$\lvert g^*_{1n}(\delta)\rvert$	$\lvert g^*_{2n}(\delta)\rvert$	$\lvert g^*_{1n}(\delta)\rvert$	$\lvert g^*_{2n}(\delta)\rvert$
0.000870	0.003749	0.000688	0.000185	0.000297	0.011091

Broadly speaking, $|g_{1n}(\delta)|$ is smaller than $|g_{2n}(\delta)|$, which supports the asymptotic theory in Section 3 that $g_{1n}(\delta)$ converges to zero at a faster rate than $g_{2n}(\delta)$. The presence of endogeneity between x_t and u_t does not impose a noticeable impact on the results, corroborating similar findings by Wang and Phillips (2009b) in the context of nonlinear cointegration models with a univariate regressor. The bias-corrected kernel method implies a second-order bias correction for $g_{in}(\cdot)$, as shown in Proposition 4.1. We find that the corresponding values of $|g^*_{1n}(\delta)|$ and $|g^*_{2n}(\delta)|$ are slightly smaller than those for $|g_{1n}(\delta)|$ and $|g_{2n}(\delta)|$ reported in Tables 5.1 and 5.2, providing evidence of bias reduction and supporting the limit theory in Section 4.

Table 5.2: Absolute averages of $g_{in}(\delta)$ and $g_{in}^*(\delta)$ for
M2: $f_{1t} = \cos\left(\frac{2\pi t}{n}\right)$ and $f_{2t} = \sin\left(\frac{2\pi t}{n}\right)$

Case 1		Case 2		Case 3	
$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$	$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$	$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$
0.000302	0.026371	0.000504	0.002895	0.042893	0.059356
Case 4		Case 5		Case 6	
$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$	$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$	$\lvert g_{1n}(\delta)\rvert$	$\lvert g_{2n}(\delta)\rvert$
0.006109	0.005456	0.024125	0.049481	0.030661	0.069070
$\lvert g_{1n}^*(\delta)\rvert$	$\lvert g_{2n}^*(\delta)\rvert$	$\lvert g_{1n}^*(\delta)\rvert$	$\lvert g_{2n}^*(\delta)\rvert$	$\lvert g_{1n}^*(\delta)\rvert$	$\lvert g_{2n}^*(\delta)\rvert$
0.005963	0.004695	0.023760	0.049477	0.030607	0.068099

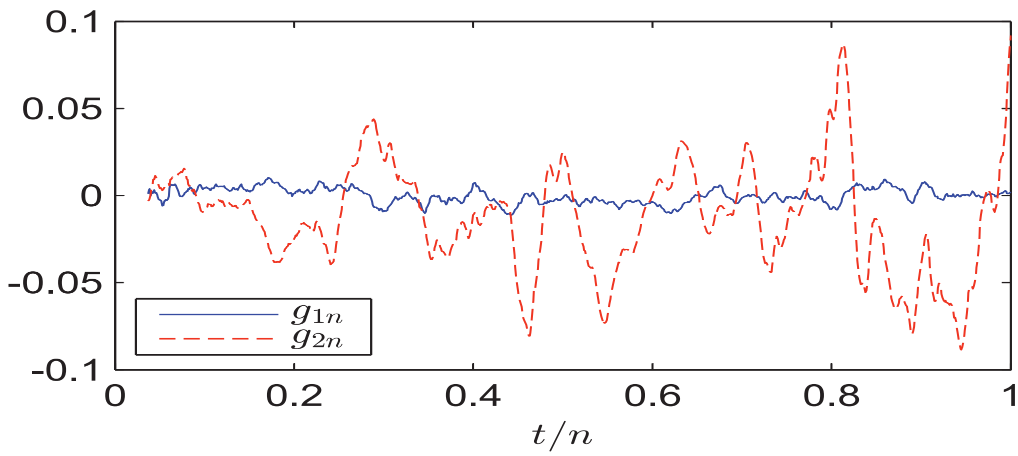

Fig. 5.1: Plots of $g_{1n}(\delta)$ and $g_{2n}(\delta)$ versus δ for functional form *M1*

We next consider the case where $\rho = 0.5$, $\rho_1 = 0.5$ and $\rho_2 = 0.5$ and $\lambda_i = 0.5$ for $i = 1, 2, 3$. For given h, we define the leave-one-out estimate

$$\widehat{f}_t(\delta|h) = \Big[\sum_{s=1,\neq t}^{n} x_s x_s' K\big(\frac{s-n\delta}{nh}\big)\Big]^+ \sum_{s=1,\neq t}^{n} x_s y_s K\big(\frac{s-n\delta}{nh}\big) \equiv \big[\widehat{f}_{1t}(\delta|h), \widehat{f}_{2t}(\delta|h)\big]', \tag{5.7}$$

and the cross-validation function

$$\mathrm{CV}_n(h) = \frac{1}{n}\sum_{t=1}^{n}\Big[y_t - x_t'\widehat{f}_t\big(\frac{t}{n}|h\big)\Big]^2, \tag{5.8}$$

and find an optimal bandwidth of the form

$$\widehat{h}_{\mathrm{cv}} = \arg\min_h \mathrm{CV}_n(h). \tag{5.9}$$

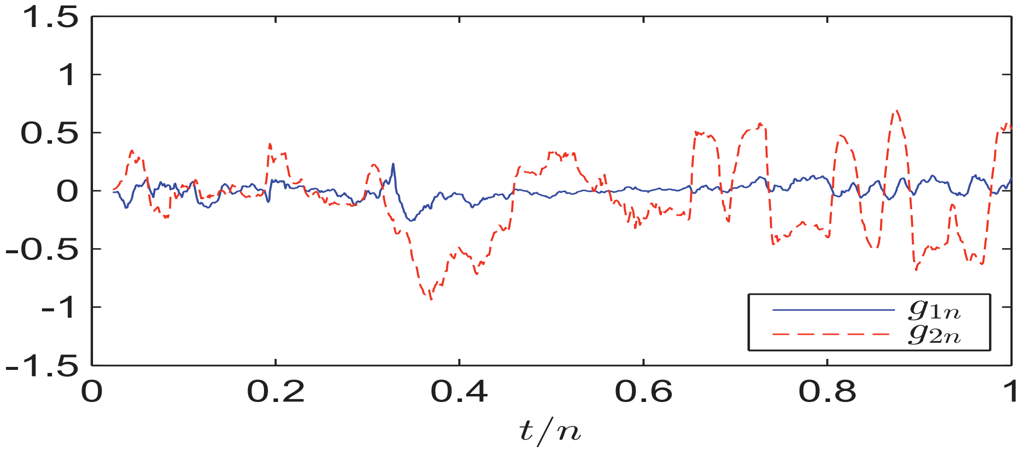

Fig. 5.2: Plots of $g_{1n}(\delta)$ and $g_{2n}(\delta)$ versus δ for functional form *M2*

For $\delta > \widehat{h}_{\mathrm{cv}}$, define $\widehat{\delta}(n) = \lfloor(\delta - \widehat{h}_{\mathrm{cv}})n\rfloor$, $x_{\widehat{\delta}(n)} = \left(x_{1,\widehat{\delta}(n)}, x_{2,\widehat{\delta}(n)}\right)'$, $\widehat{b}_n(\delta) = \frac{1}{\sqrt{n}}x_{\widehat{\delta}(n)} = \frac{1}{\sqrt{n}}\left(x_{1,\widehat{\delta}(n)}, x_{2,\widehat{\delta}(n)}\right)'$ and $\widehat{q}_n(\delta) = \left[\frac{x_{1,\widehat{\delta}(n)}}{\sqrt{n}\|\widehat{b}_n(\delta)\|}, \frac{x_{2,\widehat{\delta}(n)}}{\sqrt{n}\|\widehat{b}_n(\delta)\|}\right]' \equiv \left[\widehat{q}_{1n}(\delta), \widehat{q}_{2n}(\delta)\right]'$ with $\|\widehat{b}_n(\delta)\| = \sqrt{\frac{1}{n}[x^2_{1,\widehat{\delta}(n)} + x^2_{2,\widehat{\delta}(n)}]}$. Let the transformed quantities $g_{1n}(\delta)$ and $g_{2n}(\delta)$ be again defined as in (5.4) and (5.5) but with $q_{in}(\delta)$ and $p_{in}(\delta)$ replaced by $\widehat{q}_{in}(\delta)$ and $\widehat{p}_{in}(\delta)$, respectively, where $\widehat{p}_{in}(\delta)$ is now specified.

In both *M1* and *M2* we construct the components of the direction vector $\widehat{p}_{in}(\delta)$ using $\widehat{p}_{1n}(\delta) = \widehat{q}_{2n}(\delta)$ and $\widehat{p}_{2n}(\delta) = -\widehat{q}_{1n}(\delta)$ (for $\widehat{h}_{\mathrm{cv}} \leq \delta \leq 1$). The plots shown in Figs. 5.1 and 5.2 are based on 500 replications. These plots show clearly that the window of fluctuations of $g_{1n}(\delta)$ is much narrower than that of $g_{2n}(\delta)$, further corroborating the limit theory that the variance of $g_{1n}(\cdot)$ is smaller than that of $g_{2n}(\cdot)$.

6 Empirics

This section applies the time varying coefficient model and estimation methodology to aggregate US data on consumption, income, investment, and interest rates obtained from *Federal Reserve Economic Data (FRED)*[5]. We consider two formulations using data that were studied recently in Athanasopoulos *et al* (2011) using linear VAR and reduced rank regression methods.

Case (i) (Quarterly data over 1960:1–2009:3): y_t is log per-capita real consumption, x_{1t} is log per capita disposable income, and x_{2t} is the real interest rate expressed as a percentage and calculated ex post by deducting the CPI inflation rate over the following quarter from the nominal 90 day Treasury bill rate.

Case (ii) (Quarterly data over 1947:1–2009:4): y_t is log per-capita real consumption, x_{1t} is log per capita real disposable income, and x_{2t} is log per capita real investment.

Application of the nonparametric test in Gao and King (2011) for checking unit root nonstationarity gave p–values of 0.206, 0.217 and 0.112 for y_t, x_{1t} and x_{2t} in case (i), and corresponding

[5]We thank George Athanasopoulos for providing us with the data.

p–values of 0.219, 0.226 and 0.167 for y_t, x_{1t} and x_{2t} in case (ii). The series are plotted in Figs. 6.1 and 6.2.

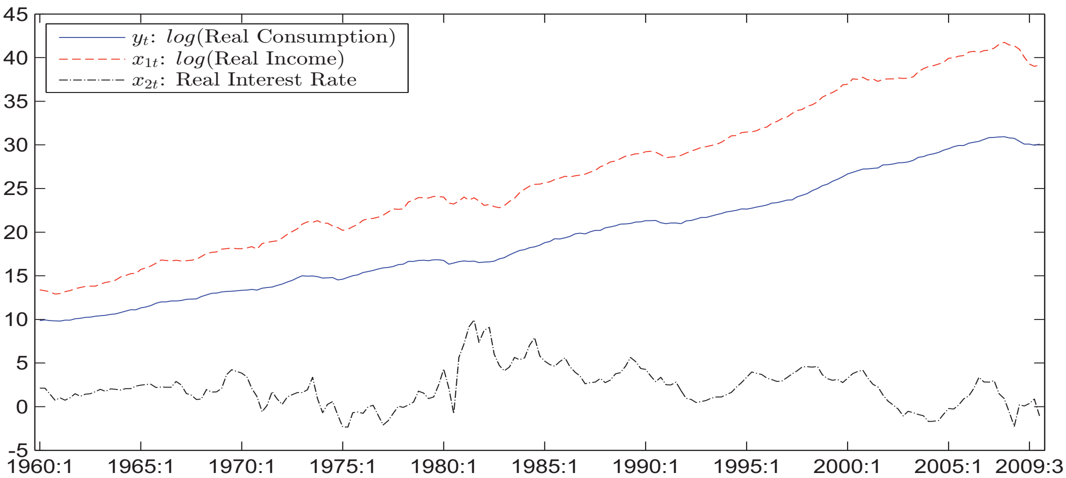

Fig. 6.1: Real consumption, real disposable income, and 90 day T bill rate 1960:1 - 2009:3

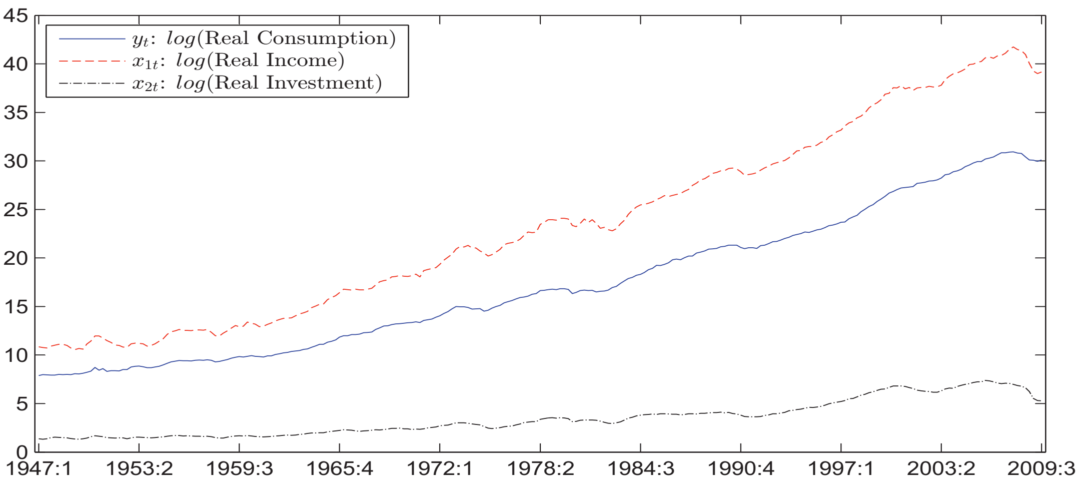

Fig. 6.2: Real consumption, disposable income and investment 1947:1 - 2009:4

In both cases, we fit the following model allowing for a time varying coefficient vector

$$y_t = x_t' f\Big(\frac{t}{n}\Big) + u_t = x_t' f_t + u_t, \quad t = 1, \cdots, n, \tag{6.1}$$

where the regressors and coefficients are partitioned as $x_t = (x_{1t}, x_{2t})'$ and $f_t = (f_{1t}, f_{2t})'$. The coefficient function $f(\cdot) = (f_1(\cdot), f_2(\cdot))'$ is estimated by kernel weighted regression giving

$$\widehat{f}(\delta) = \Big[\sum_{t=1}^{n} x_t x_t' K\Big(\frac{t - n\delta}{nh}\Big)\Big]^{+} \sum_{t=1}^{n} x_t y_t K\Big(\frac{t - n\delta}{nh}\Big), \tag{6.2}$$

where $K(x) = \frac{1}{2}I\{-1 \le x \le 1\}$ as in Section 5, over $\delta \in (0,1]$, and the bandwidth h is chosen by cross-validation as described in (5.9). The nonparametric estimates of the two curves $f_i(\cdot)$ with their 95% confidence bands are shown in Figs. 6.3 and 6.4 for (i), and in Figs. 6.5 and 6.6 for (ii).

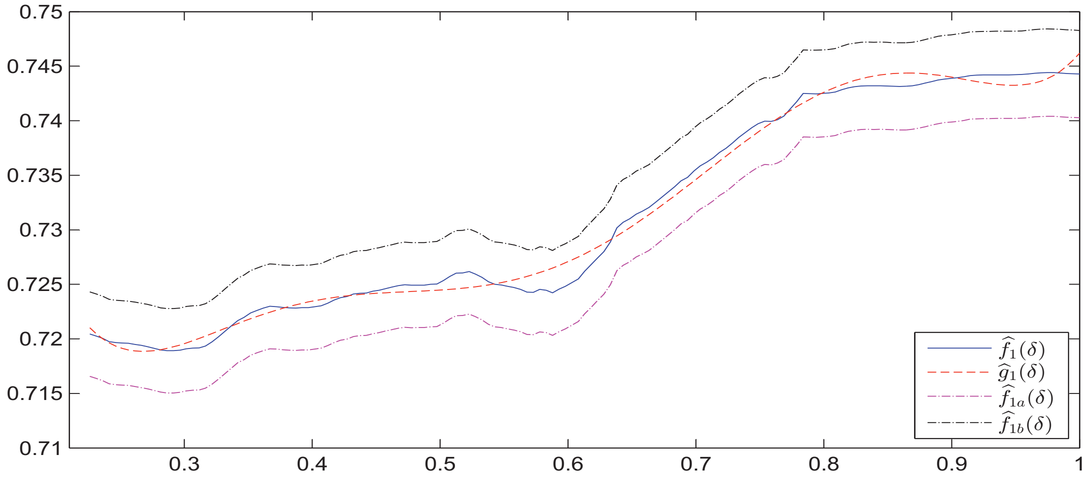

Fig. 6.3: Nonparametric estimate ($\widehat{f}_1$) with confidence bands ($\widehat{f}_{1a}$, $\widehat{f}_{1b}$) together with the parametric polynomial ($\widehat{g}_1$) estimate of f_1 for Case (i)

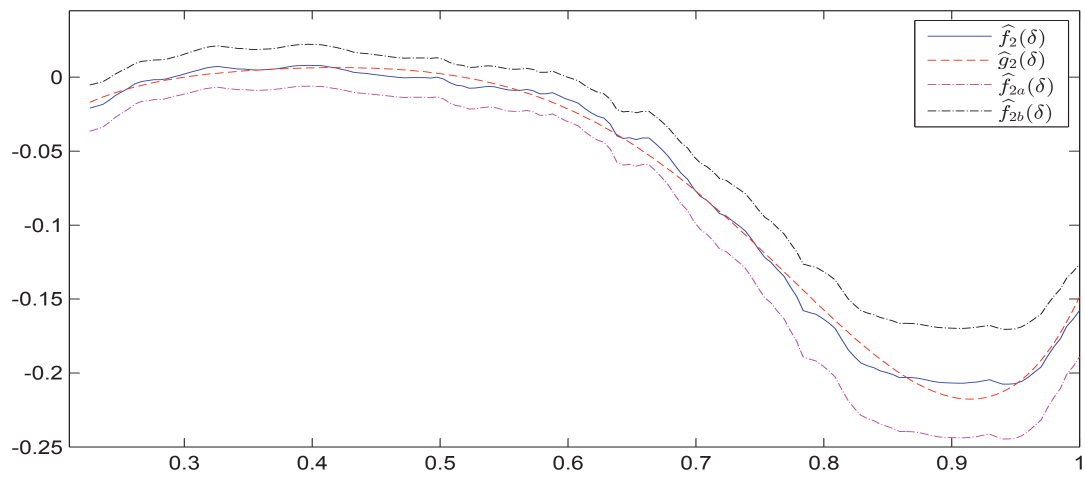

Fig. 6.4: Nonparametric estimate ($\widehat{f}_2$) with confidence bands ($\widehat{f}_{2a}$, $\widehat{f}_{2b}$) together with the parametric polynomial ($\widehat{g}_2$) estimate of f_2 for Case (i)

The plots of $\widehat{f}_1(\delta)$ and $\widehat{f}_2(\delta)$ are strongly indicative of nonlinear functional forms for the coefficients in both cases, but also suggest that the functions $f_i(\delta)$ may be approximated by much simpler parametric functions $g_i(\delta;\theta_i)$, for some parametric θ_i and pre-specified $g_i(\cdot;\cdot)$. For (i), in

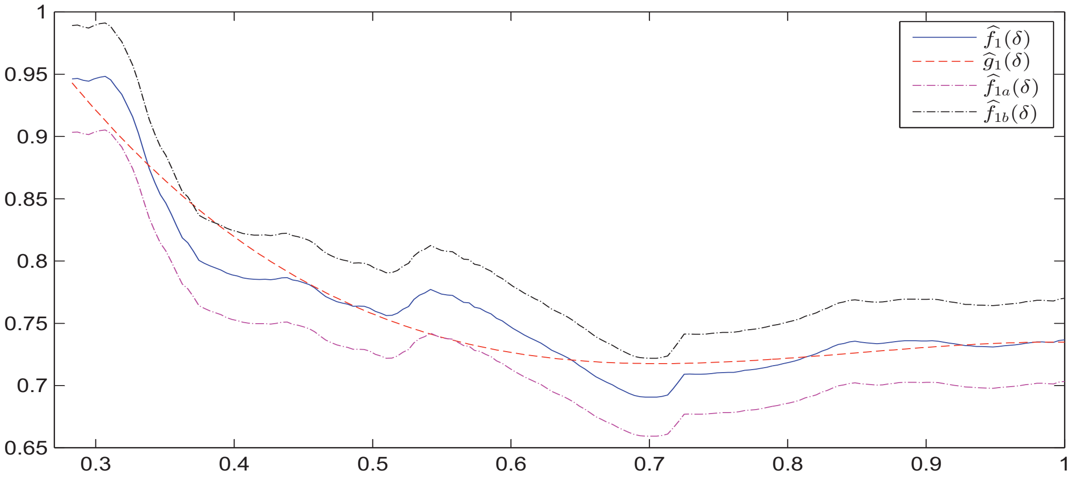

Fig. 6.5: Nonparametric estimate ($\widehat{f}_1$) with confidence bands ($\widehat{f}_{1a}$, $\widehat{f}_{1b}$) together with the parametric polynomial ($\widehat{g}_1$) estimate of f_1 for Case (ii)

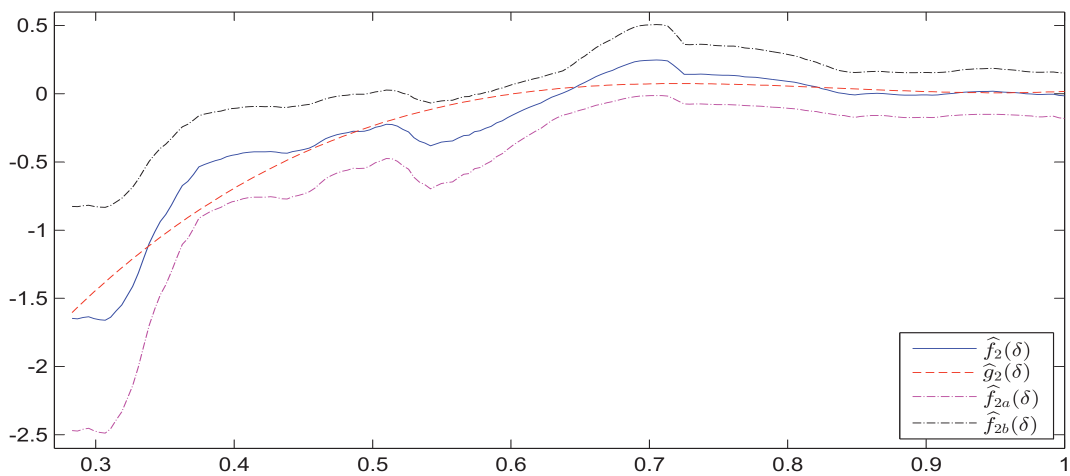

Fig. 6.6: Nonparametric estimate ($\widehat{f}_2$) with confidence bands ($\widehat{f}_{2a}$, $\widehat{f}_{2b}$) together with the parametric polynomial ($\widehat{g}_2$) estimate of f_2 for Case (ii)

Figs. 6.3 and 6.4, we added polynomial specifications of the form

$$g_1(\delta; \widehat{\theta}_1) = \widehat{\theta}_{01} + \sum_{j=1}^{6} \widehat{\theta}_{j1}\delta^j, \tag{6.3}$$

$$g_2(\delta; \widehat{\theta}_2) = \widehat{\theta}_{02} + \sum_{j=1}^{5} \widehat{\theta}_{j2}\delta^j, \tag{6.4}$$

where $\widehat{\theta}_{01} = 1.1036$, $\widehat{\theta}_{11} = -4.9534$, $\widehat{\theta}_{21} = 225.087$, $\widehat{\theta}_{31} = -63.983$, $\widehat{\theta}_{41} = 87.136$, $\widehat{\theta}_{51} = -60.191$, $\widehat{\theta}_{61} = 16.547$; $\widehat{\theta}_{02} = 0.4359$, $\widehat{\theta}_{12} = 4.577$, $\widehat{\theta}_{22} = -19.381$, $\widehat{\theta}_{32} = 41.327$, $\widehat{\theta}_{42} = -43.237$ and $\widehat{\theta}_{52} =$

17.001. Similarly for (ii), in Figs. 6.5 and 6.6, we added the following polynomial specifications:

$$g_1(\delta;\widehat{\theta}_1)=\widehat{\theta}_{01}+\sum_{j=1}^{3}\widehat{\theta}_{j1}\delta^j, \tag{6.5}$$

$$g_2(\delta;\widehat{\theta}_2)=\widehat{\theta}_{02}+\sum_{j=1}^{3}\widehat{\theta}_{j2}\delta^j, \tag{6.6}$$

where $\widehat{\theta}_{01}=1.5525$, $\widehat{\theta}_{11}=-3.0978$, $\widehat{\theta}_{21}=3.7520$, $\widehat{\theta}_{31}=-1.4718$; $\widehat{\theta}_{02}=-6.1002$, $\widehat{\theta}_{12}=-22.890$, $\widehat{\theta}_{22}=-27.873$ and $\widehat{\theta}_{31}=11.100$.

Figs. 6.3 – 6.6 show that $f_1(\delta)$ and $f_2(\delta)$ are reasonably well captured by the parametric forms $g_1(\delta;\widehat{\theta}_1)$ and $g_2(\delta;\widehat{\theta}_2)$ in both cases. Interestingly, lower order polynomial approximations were used in case (ii) than those in case (i), even though the data cover a longer period in (ii) than (i). In case (ii) both regressors are macro aggregates (income and investment), and slower moving (i.e., less variable over time) functional responses might be expected. Case (i) involves the interest rate regressor, which displays greater volatility than the macro aggregates, so the functional responses are correspondingly more variable over the sample period and seem to require higher order polynomial approximations to adequately capture the nonparametric fits.

Standard t–tests show that all these coefficients are significant with p–values almost zero. Conventional t–tests are robust to this type of parametric regression under nonstationarity, being equivalent to those from a standardised (weak trend) model of the form $y_t=\widetilde{x}_{d,t}'\widetilde{g}\left(\frac{t}{n};\theta_0\right)+u_t$, where $\widetilde{x}_{d,t}=\frac{x_t}{\sqrt{n}}$ and $\widetilde{g}\left(\frac{t}{n};\theta_0\right)=\sqrt{n}g\left(\frac{t}{n};\theta_0\right)$ giving the same p–values. A formal test of the polynomial specifications may be mounted to test the null hypothesis $H_0:\ y_t=x_t'g\big(\frac{t}{n};\theta_0\big)+u_t$ for a specific parametric form $g(\cdot;\theta_0)$. The test statistic used to assess this (joint) null hypothesis is $L_n(h)$ which is defined in (C.5) in Appendix C. This statistic measures scaled departures of parametrically fitted functional elements from their nonparametric counterparts. A detailed development of the test statistic and its limit theory is provided in Appendix C.

Allowing the equation errors u_t in (C.1) to be weakly dependent, we propose using a block bootstrap method (c.f., Hall, Horowitz and Jing, 1995) to compute p-values of $L_n(h)$ for practical implementation. The procedure is as follows.

STEP 1: For the real data $\{(y_t,x_{1t},x_{2t})\}_{t=1}^n$, compute the statistic $L_n(h)$, where h can be chosen based on the cross-validation method as in (5.9).
STEP 2: Generate $\{\widetilde{u}_t\}$ by $\widetilde{u}_t=\sum_{j=0}^{\infty}\widetilde{\rho}^j\eta_{t-j}$, in which $\{\eta_t\}$ is a sequence of independent observations drawn from N$(0,1)$, and $\widetilde{\rho}$ is estimated based on

$$y_t-x_t'g\big(\frac{t}{n};\widehat{\theta}\big)=\rho\big[y_{t-1}-x_{t-1}'g\big(\frac{t-1}{n};\widehat{\theta}\big)\big]+\varepsilon_t.$$

Let $l=\lfloor n^{\frac{1}{3}}\rfloor$ and choose b such that $bl=n$. Generate $u_{1l}^*(j)=[\widetilde{u}_1(j),\cdots,\widetilde{u}_l(j)]$, $\cdots$, $u_{Nl}^*(j)=\big[\widetilde{u}_{(b-1)l+1}(j),\cdots,\widetilde{u}_{bl}(j)\big]$ in step j for $N=n-l+1$. Replicate the resample $J=250$ times and obtain J bootstrap resamples $\{u_{sl}^*(j):1\le s\le N;1\le j\le J\}$ and then take the average $u_{sl}^*=\frac{1}{J}\sum_{j=1}^{J}u_{sl}^*(j)$ to obtain a block bootstrap version of u_t of the form: $(u_1^*,\cdots,u_n^*)=(u_{1l}^*,\cdots,u_{Nl}^*)$.
STEP 3: Generate $y_t^*=x_t'g\big(\frac{t}{n};\widehat{\theta}\big)+u_t^*$. Re-estimate θ by $\widehat{\theta}^*$ and then compute the corresponding version $L_n^*(h)$ using the new data (y_t^*,x_{1t},x_{2t}).

STEP 4: Repeat the above steps $M = 500$ times to find the bootstrap distribution of $L_n^*(h)$ and compute the proportion of $L_n(h) < L_n^*(h)$. This proportion is an approximate p-value of $L_n(h)$.

For cases (i) and (ii) the calculated p-values are 0.2937 and 0.3178, respectively, confirming that there is insufficient evidence to reject the null hypothesis H_0 in both cases. In other words, a suitable polynomial function provides a reasonable parametric approximation to each coefficient function $f_i(\delta)$ for both data sets over their respective sample periods.

7 Conclusions

Nonlinear cointegrated systems are of particular empirical interest in cases where the data are nonstationary and move together over time yet linear cointegration fails. Time varying coefficient models provide a general mechanism for addressing and capturing such nonlinearities, allowing for smooth structural changes to occur over the sample period. The present paper has explored a general approach to fitting these nonlinear systems using kernel-based structural coefficient estimation which allow the coefficients to evolve smoothly over time.

Our analysis reveals a novel feature of kernel asymptotics that has not been encountered in any previous literature on kernel regression. When the functional coefficient is multivariate, the usual asymptotic methods and limit theory of kernel estimation break down due to a degeneracy in the kernel-weighted signal matrix associated with the nonstationary regressors. This degeneracy does not affect inference but, as we have shown here, it has a major effect on the limit theory. The asymptotics rely on path-dependent local coordinate transformations to re-orient coordinates and accommodate the kernel degeneracy, changing the limit theory in a fundamental way from the existing kernel literature. The degeneracy leads to two different limit distributions with different convergence rates in two complementary directions of the function space. Unexpectedly, and in contradistinction to the case of linear model degeneracy with cointegrated regressors (Park and Phillips, 1988, 1989), both convergence rates are faster than the usual convergence rate for stationary systems – here nonlinear models with smoothly changing coefficients in the conventional setting of local stationarity. The higher rate of convergence ($n\sqrt{h}$) lies in the direction of the nonstationary regressor vector at the local coordinate point of the function and the lower rate (nh) lies in the degenerate direction but this rate is still clearly super-consistent for nonparametric estimators.

Kernel estimation of time varying coefficient cointegration models therefore involves two types of super-consistency and this limit theory differs significantly from other kernel asymptotics for nonlinear systems as well as the limit theory for linear systems with cointegrated regressors. For practical implementation purposes, a local linear estimation approach is developed to reduce asymptotic bias and relax bandwidth restrictions, and a fully modified kernel regression estimator is developed to deal with models where there are endogenous nonstationary regressors. Implementation is illustrated in simulations that study finite sample performance and in an empirical application to explore the linkages over time among aggregate consumption, disposable income, investment and interest rates in the US.

The present paper touches on several topics that deserve further study. Included among these are model specification tests, bandwidth selection methods for kernel smoothing, and the uniform convergence properties of nonparametric kernel estimates in nonstationary time varying coefficient models. Another area of potential importance for empirical research is the case where both deterministic and stochastic trends arise among the regressors. This type of model raises further complications of degeneracy that may be handled by the methods developed here.

A Proofs of the main results

To derive the limit theory for $\widehat{f}_n(\delta_0)$ in (2.1) we start with asymptotics for the denominator involved in $\widehat{f}_n(\delta_0)$. In what follows let $G_{th} = hK_{th}(\delta_0)$, and C be a positive constant whose value may change from line to line.

PROPOSITION A.1. *Suppose that Assumptions 1 and 3 are satisfied. Then, we have, for fixed* $0 < \delta_0 < 1$,

$$D_n^+ Q_n' \Big(\sum_{t=1}^n x_t x_t' G_{th} \Big) Q_n D_n^+ \Rightarrow \Delta_{\delta_0}, \tag{A.1}$$

where Δ_{δ_0} *is defined in (3.4) of Section 3.*

PROOF. For notational economy, let $\Delta = \Delta_{\delta_0}$ and $\Delta(k) = \Delta_{\delta_0}(k)$ for $k = 1, 2, 3$, throughout the proof. Observe that

$$\begin{aligned} & D_n^+ Q_n' \Big(\sum_{t=1}^n x_t x_t' G_{th} \Big) Q_n D_n^+ \\ = & \begin{bmatrix} \frac{1}{nh} \sum_{t=1}^n q_n' \big(\frac{x_t}{\sqrt{n}}\big) \big(\frac{x_t}{\sqrt{n}}\big)' q_n G_{th} & \frac{1}{nh^{3/2}} \sum_{t=1}^n q_n' \big(\frac{x_t}{\sqrt{n}}\big) \big(\frac{x_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} \\ \frac{1}{nh^{3/2}} \sum_{t=1}^n (q_n^{\perp})' \big(\frac{x_t}{\sqrt{n}}\big) \big(\frac{x_t}{\sqrt{n}}\big)' q_n G_{th} & \frac{1}{(nh)^2} \sum_{t=1}^n (q_n^{\perp})' (x_t x_t') q_n^{\perp} G_{th} \end{bmatrix} \\ \equiv & \begin{bmatrix} \Delta_n(1) & \Delta_n(2) \\ \Delta_n(2)' & \Delta_n(3) \end{bmatrix}. \end{aligned} \tag{A.2}$$

We next consider $\Delta_n(1)$, $\Delta_n(2)$ and $\Delta_n(3)$ in turn and prove that $\Delta_n(k) \Rightarrow \Delta(k)$ for $k = 1, 2, 3$, where the submatrices $\Delta(k)$ are defined following (3.4) of Section 3. Let

$$\Delta_n^*(1) = q_n' \big(\frac{x_{\delta(n)}}{\sqrt{n}}\big) \big(\frac{x_{\delta(n)}}{\sqrt{n}}\big)' q_n \Big(\frac{1}{nh} \sum_{t=1}^n G_{th} \Big),$$

where $\delta(n) = \lfloor (\delta_0 - h) n \rfloor$ is defined as in Section 2. By (2.3) and the definition of q_n, we have

$$\Delta_n^*(1) = q_n' \big(\frac{x_{\delta(n)}}{\sqrt{n}}\big) \big(\frac{x_{\delta(n)}}{\sqrt{n}}\big)' q_n (1 + o_P(1)) \Rightarrow B_{d,\delta_0}(\Omega_v)' B_{d,\delta_0}(\Omega_v). \tag{A.3}$$

Following the proof of (2.18), it is easy to show that $\Delta_n^*(1)$ is the leading term of $\Delta_n(1)$. It follows that

$$\Delta_n(1) \Rightarrow B_{d,\delta_0}(\Omega_v)' B_{d,\delta_0}(\Omega_v) \equiv \Delta(1). \tag{A.4}$$

From the BN decomposition (Phillips and Solo, 1992) we have for $t \geq \delta(n)$

$$\begin{aligned} x_t &= \sum_{s=1}^{t} \overline{v}_s + \widetilde{v}_0 - \widetilde{v}_t + x_0 \\ &= \Big[\sum_{s=1}^{\delta(n)} \overline{v}_s + \widetilde{v}_0 - \widetilde{v}_{\delta(n)} + x_0\Big] + \Big[\sum_{s=\delta(n)+1}^{t} \overline{v}_s\Big] + \Big[\widetilde{v}_{\delta(n)} - \widetilde{v}_t\Big] \\ &\equiv x_{\delta(n)} + \eta_t + \xi_t, \end{aligned} \tag{A.5}$$

where $\eta_{\delta(n)} = 0$. Note that $x'_{\delta(n)} q_n^{\perp} = 0$ with probability 1. Hence, $\Delta_n^*(2)$ is asymptotically equivalent to $\Delta_n(2)$, where

$$\Delta_n^*(2) = \frac{1}{nh^{3/2}} \sum_{t=1}^{n} q_n' \big(\frac{x_t}{\sqrt{n}}\big) \big(\frac{\eta_t + \xi_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th}.$$

By using (A.5) again, we have

$$\begin{aligned} \Delta_n^*(2) &= \frac{1}{nh^{3/2}} q_n' \big(\frac{x_{\delta(n)}}{\sqrt{n}}\big) \sum_{t=1}^{n} \big(\frac{\eta_t + \xi_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} \\ &\quad + \frac{1}{nh^{3/2}} \sum_{t=1}^{n} q_n' \big(\frac{\eta_t + \xi_t}{\sqrt{n}}\big) \big(\frac{\eta_t + \xi_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} \\ &\equiv \Delta_n^*(2,1) + \Delta_n^*(2,2). \end{aligned} \tag{A.6}$$

It is easy to show that

$$\begin{aligned} &\Big\| \frac{1}{nh^{3/2}} q_n' \big(\frac{x_{\delta(n)}}{\sqrt{n}}\big) \sum_{t=1}^{n} \big(\frac{\xi_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} \Big\| \\ \leq\;& \frac{1}{(nh)^{1/2}} \cdot \Big\| q_n' \big(\frac{x_{\delta(n)}}{\sqrt{n}}\big) \Big\| \cdot \Big(\frac{1}{nh} \sum_{t=1}^{n} \| \xi_t' q_n^{\perp} \| G_{th} \Big) \\ =\;& O_P\big((nh)^{-1/2}\big) = o_P(1). \end{aligned} \tag{A.7}$$

On the other hand, by Assumption 1, there exist two independent Brownian motions, $B_{d,r}(\Omega_v)$ and $B_{d,r,*}(\Omega_v)$, such that

$$\Big(\frac{1}{\sqrt{\delta(n)}} \sum_{s=1}^{\delta(n)} \overline{v}_s,\ \frac{1}{\sqrt{2\lfloor nh \rfloor}} \sum_{s=\delta(n)+1}^{\delta_r(n)} \overline{v}_s \Big) \Rightarrow \Big[B_{d,1}(\Omega_v),\ B_{d,r,*}(\Omega_v) \Big] \tag{A.8}$$

for $\delta_r(n) = \delta(n) + \lfloor 2rnh \rfloor + 1$ with $0 < r \leq 1$. By using (A.8) and the definition of η_t in (A.5), we can show that

$$\begin{aligned} &\frac{1}{nh^{3/2}} q_n' \big(\frac{x_{\delta(n)}}{\sqrt{n}}\big) \sum_{t=1}^{n} \big(\frac{\eta_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} \\ =\;& 2\sqrt{2} q_n' \big(\frac{x_{\delta(n)}}{\sqrt{n}}\big) \Big[\frac{1}{2nh} \sum_{t=1}^{n} \big(\frac{\eta_t}{\sqrt{2nh}}\big)' G_{th} \Big] q_n^{\perp} \\ \Rightarrow\;& 2\sqrt{2} \big[B_{d,\delta_0}(\Omega_v)' B_{d,\delta_0}(\Omega_v) \big]^{1/2} \Big[\int_{-1}^{1} B_{d,\frac{r+1}{2},*}(\Omega_v) K(r) dr \Big] q^{\perp}, \end{aligned} \tag{A.9}$$

which together with (A.7), indicates that

$$\Delta_n^*(2,1) \Rightarrow \Delta(2) \tag{A.10}$$

with $\Delta(2) = 2\sqrt{2}\big[B_{d,\delta_0}(\Omega_v)' B_{d,\delta_0}(\Omega_v)\big]^{1/2}\big[\int_{-1}^{1} B_{d,\frac{r+1}{2},*}(\Omega_v)K(r)dr\big]q^{\perp}$. Note that

$$\begin{aligned}
\Delta_n^*(2,2) &= \frac{1}{nh^{3/2}}\sum_{t=1}^{n} q_n'\big(\frac{\eta_t}{\sqrt{n}}\big)\big(\frac{\eta_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} + \frac{1}{nh^{3/2}}\sum_{t=1}^{n} q_n'\big(\frac{\eta_t}{\sqrt{n}}\big)\big(\frac{\xi_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} + \\
&\quad \frac{1}{nh^{3/2}}\sum_{t=1}^{n} q_n'\big(\frac{\xi_t}{\sqrt{n}}\big)\big(\frac{\eta_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} + \frac{1}{nh^{3/2}}\sum_{t=1}^{n} q_n'\big(\frac{\xi_t}{\sqrt{n}}\big)\big(\frac{\xi_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} \\
&\equiv \Delta_n^*(2,2,1) + \Delta_n^*(2,2,2) + \Delta_n^*(2,2,3) + \Delta_n^*(2,2,4).
\end{aligned} \tag{A.11}$$

We next show that $\Delta_n^*(2,2,k) = o_P(1)$ for $k = 1,\cdots,4$. To save space, we prove only that $\Delta_n^*(2,2,1) = o_P(1)$ and $\Delta_n^*(2,2,4) = o_P(1)$ as the other two cases follow similarly. By (A.8), we can prove

$$\begin{aligned}
\Delta_n^*(2,2,1) &= \frac{1}{nh^{3/2}}\sum_{t=1}^{n} q_n'\big(\frac{\eta_t}{\sqrt{n}}\big)\big(\frac{\eta_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} \\
&= \frac{2\sqrt{h}}{nh}\sum_{t=1}^{n} q_n'\big(\frac{\eta_t}{\sqrt{2nh}}\big)\big(\frac{\eta_t}{\sqrt{2nh}}\big)' q_n^{\perp} G_{th} \\
&= O_P\big(\frac{\sqrt{h}}{nh}\sum_{t=1}^{n} G_{th}\big) = O_P(\sqrt{h}) = o_P(1),
\end{aligned} \tag{A.12}$$

as $h \to 0$, and

$$\begin{aligned}
\Delta_n^*(2,2,4) &= \frac{1}{nh^{3/2}}\sum_{t=1}^{n} q_n'\big(\frac{\xi_t}{\sqrt{n}}\big)\big(\frac{\xi_t}{\sqrt{n}}\big)' q_n^{\perp} G_{th} \\
&= O_P\big(\frac{1}{n^2h^{3/2}}\sum_{t=1}^{n} G_{th}\big) = O_P\big(\frac{1}{nh^{1/2}}\big) = o_P(1),
\end{aligned} \tag{A.13}$$

as $nh \to \infty$. We have thus proved $\Delta_n^*(2,2) = o_P(1)$ which, together with (A.10), shows that

$$\Delta_n(2) \Rightarrow \Delta(2). \tag{A.14}$$

Finally, consider $\Delta_n(3)$. Noting that $(q_n^{\perp})' x_{\delta(n)} = 0$ with probability 1, we can argue that $\Delta_n^*(3)$ is asymptotically equivalent to $\Delta_n(3)$, where

$$\Delta_n^*(3) = \frac{1}{(nh)^2}\sum_{t=1}^{n} (q_n^{\perp})'\big(\eta_t + \xi_t\big)\big(\eta_t + \xi_t\big)' q_n^{\perp} G_{th}.$$

Furthermore, following the proof of $\Delta_n^*(2,2)$ as above, we can show that

$$\begin{aligned}
\Delta_n^*(3) &= \frac{1}{(nh)^2}\sum_{t=1}^{n} (q_n^{\perp})' \eta_t \eta_t' q_n^{\perp} G_{th} + o_P(1) = (q_n^{\perp})'\Big[\frac{4}{2nh}\sum_{t=1}^{n} \big(\frac{\eta_t}{\sqrt{2nh}}\big)\big(\frac{\eta_t}{\sqrt{2nh}}\big)' G_{th}\Big] q_n^{\perp} \\
&\Rightarrow 4(q^{\perp})'\Big[\int_{-1}^{1} B_{d,\frac{r+1}{2},*}(\Omega_v) B_{d,\frac{r+1}{2},*}(\Omega_v)' K(r)dr\Big] q^{\perp} \equiv \Delta(3).
\end{aligned} \tag{A.15}$$

The proof of (A.1) is now complete in view of (A.4), (A.14) and (A.15). □

Next consider the derivation of the limit behavior of

$$\Gamma_n \equiv \Gamma_{n\delta_0} = D_n^+ Q_n' \Big(\sum_{t=1}^{n} x_t u_t G_{th} \Big). \tag{A.16}$$

Observe that

$$\Gamma_n = D_n^+ Q_n' \Big(\sum_{t=1}^{n} x_t u_t G_{th} \Big) = \left[\begin{array}{c} q_n' \Big(\frac{1}{nh^{1/2}} \sum_{t=1}^{n} x_t u_t G_{th} \Big) \\ (q_n^{\perp})' \Big(\frac{1}{nh} \sum_{t=1}^{n} x_t u_t G_{th} \Big) \end{array} \right] \equiv \left[\begin{array}{c} \Gamma_n(1) \\ \Gamma_n(2) \end{array} \right]. \tag{A.17}$$

We give asymptotic distributions for $\Gamma_n(1)$ and $\Gamma_n(2)$ in the following Propositions A.2 and A.3, respectively.

Proposition A.2. *Suppose that Assumptions 1 and 3 are satisfied. Then,*

$$\Gamma_n(1) \Rightarrow \big[B_{d,\delta_0}(\Omega_v)' B_{d,\delta_0}(\Omega_v) \big]^{1/2} \mathcal{Z}_u^*, \tag{A.18}$$

where $\mathcal{Z}_u^*$ is a normal distribution with zero mean and variance matrix $\nu_0 \Omega_u$, and is independent of $B_{d,\delta_0}(\Omega_v)$, Ω_u is defined in (2.5).

Proof. Define

$$u_{t\blacktriangle} = \sum_{j=0}^{\rho(n)} \Phi_{j,2}' \varepsilon_{t-j}, \quad u_{t\blacktriangledown} = \sum_{j=\rho(n)+1}^{\infty} \Phi_{j,2}' \varepsilon_{t-j},$$

where $\rho(n) \to \infty$ and will be specified later. Note that

$$\begin{aligned} \Gamma_n(1) &= q_n' \Big(\frac{1}{nh^{1/2}} \sum_{t=1}^{n} x_t u_t G_{th} \Big) \\ &= q_n' \Big(\frac{x_{\delta(n)}}{\sqrt{n}} \cdot \frac{1}{(nh)^{1/2}} \sum_{t=1}^{n} u_{t\blacktriangle} G_{th} \Big) + q_n' \Big(\frac{x_{\delta(n)}}{\sqrt{n}} \cdot \frac{1}{(nh)^{1/2}} \sum_{t=1}^{n} u_{t\blacktriangledown} G_{th} \Big) + \\ & \quad q_n' \Big[\frac{1}{nh^{1/2}} \sum_{t=1}^{n} \big(x_t - x_{\delta(n)} \big) u_t G_{th} \Big] \\ &\equiv \Gamma_n(1,1) + \Gamma_n(1,2) + \Gamma_n(1,3). \end{aligned} \tag{A.19}$$

By Assumptions 1 and 3 we have

$$\begin{aligned} \big| \Gamma_n(1,3) \big| &\leq \frac{1}{nh^{1/2}} \cdot \| q_n \| \cdot \Big\| \sum_{t=1}^{n} \big(x_t - x_{\delta(n)} \big) u_t G_{th} \Big\| \\ &= \frac{1}{nh^{1/2}} \cdot O_P(1) \cdot O_P(nh) = O_P(\sqrt{h}) = o_P(1). \end{aligned} \tag{A.20}$$

As $\{\varepsilon_t\}$ is a sequence of *iid* random vectors, we have, for any t,

$$\mathbb{E}\big[u_{t\blacktriangledown}^2 \big] \leq \sum_{j=\rho(n)+1}^{\infty} \| \Phi_{j,2} \|^2 = o_P(\rho^{-3}(n))$$

by Assumption 1. Hence, we have

$$\left\|x_{\delta(n)}\sum_{t=1}^{n}u_{t\blacktriangledown}G_{th}\right\| = o_P\big(\sqrt{n}\cdot nh\cdot\rho^{-3/2}(n)\big) = o_P(n\sqrt{h}) \tag{A.21}$$

by letting $\rho(n)=(nh)^{\frac{1}{3}+\epsilon_*}$, $0<\epsilon_*<\frac{\gamma_0}{3(6+2\gamma_0)}$, where γ_0 is defined in Assumption 1. We therefore have

$$\left|\Gamma_n(1,2)\right| = o_P(1). \tag{A.22}$$

Let $\varsigma_n=\rho(n)/n$ and $\widetilde{\delta}(n)=\lfloor(\delta_0-h-\varsigma_n)n\rfloor$. Observe that

$$\begin{aligned}\frac{1}{n\sqrt{h}}q_n'x_{\delta(n)}\sum_{t=1}^{n}u_{t\blacktriangle}G_{th} &= \frac{1}{n\sqrt{h}}q_n'x_{\widetilde{\delta}(n)}\Big(\sum_{t=1}^{n}u_{t\blacktriangle}G_{th}\Big)+\frac{1}{n\sqrt{h}}q_n'\big[x_{\delta(n)}-x_{\widetilde{\delta}(n)}\big]\Big(\sum_{t=1}^{n}u_{t\blacktriangle}G_{th}\Big)\\ &\equiv \Gamma_n(1,1,1)+\Gamma_n(1,1,2).\end{aligned} \tag{A.23}$$

As $\varsigma_n\to 0$, it is easy to see that $\Gamma_n(1,1,2)$ is dominated by $\Gamma_n(1,1,1)$, which is the leading term of $\Gamma_n(1,1)$. We next establish asymptotics for $\Gamma_n(1,1,1)$. Note that $\varsigma_n n=\rho(n)$ and

$$\sum_{t=1}^{n}u_{t\blacktriangle}G_{th}=\sum_{t=\delta(n)+1}^{\delta_1(n)}u_{t\blacktriangle}G_{th},$$

by Assumption 3(i), where $\delta_1(n)$ is defined in the proof of Proposition A.1. Thus, by Assumption 1 $x_{\widetilde{\delta}(n)}$ is independent of $\sum_{t=1}^{n}u_{t\blacktriangle}G_{th}$. Similar to the proof of (A.3) we can show that

$$\frac{1}{\sqrt{n}}q_n'x_{\widetilde{\delta}(n)}\Rightarrow\Delta_{\delta_0}^{1/2}(1). \tag{A.24}$$

On the other hand, note that $\{z_t:\ \delta(n)\le t\le\delta_1(n)\}$ is a sequence of $\rho(n)$-dependent random vectors, where $z_t=u_{t\blacktriangle}G_{th}$. By Assumption 1, it is easy to see that condition (i) in Lemma B.1 (stated in Appendix B) is satisfied with $\varrho=2+\gamma_0$. As $\rho(n)=(nh)^{\frac{1}{3}+\epsilon_*}$ with $0<\epsilon_*<\frac{\gamma_0}{3(6+2\gamma_0)}$, we can easily show that

$$\rho^{2+\frac{2}{\varrho}}(n)=\rho^{\frac{6+2\gamma_0}{2+\gamma_0}}(n)=o(nh),$$

which indicates that condition (iv) in Lemma B.1 is also satisfied. We next show that condition (iii) in Lemma B.1 holds. Observe that

$$\mathbb{E}\Big[\Big(\sum_{t=\delta(n)+1}^{\delta_1(n)}z_t\Big)^2\Big]=\sum_{t=\delta(n)+1}^{\delta_1(n)}\mathbb{E}[z_t^2]+\sum_{t=\delta(n)+1}^{\delta_1(n)}\sum_{0<|s-t|\le\rho(n)}\mathbb{E}[z_tz_s]. \tag{A.25}$$

Letting $\Delta_{s,t}(K) = K\big(\frac{s-\delta_0 n}{nh}\big) - K\big(\frac{t-\delta_0 n}{nh}\big)$, by Assumption 3(i), we can prove that

$$\begin{aligned}\sum_{t=\delta(n)+1}^{\delta_1(n)} \sum_{0<|s-t|\le\rho(n)} \mathbb{E}[z_t z_s] &= \sum_{t=\delta(n)+1}^{\delta_1(n)} \sum_{0<|s-t|\le\rho(n)} G_{th}G_{sh}\mathbb{E}[u_{t\blacktriangle}u_{s\blacktriangle}] \\ &= \sum_{t=\delta(n)+1}^{\delta_1(n)} G_{th}^2 \sum_{0<|s-t|\le\rho(n)} \mathbb{E}[u_{t\blacktriangle}u_{s\blacktriangle}] \\ &\quad + \sum_{t=\delta(n)+1}^{\delta_1(n)} G_{th} \sum_{0<|s-t|\le\rho(n)} \Delta_{s,t}(K)\mathbb{E}[u_{t\blacktriangle}u_{s\blacktriangle}] \\ &= \sum_{t=\delta(n)+1}^{\delta_1(n)} G_{th}^2 \sum_{0<|s-t|\le\rho(n)} \mathbb{E}[u_{t\blacktriangle}u_{s\blacktriangle}] + o(nh). \qquad \text{(A.26)}\end{aligned}$$

On the other hand, by elementary calculations we have, as $n \to \infty$,

$$\sum_{s=t-\rho(n)}^{t+\rho(n)} \mathbb{E}[u_{t\blacktriangle}u_{s\blacktriangle}] \longrightarrow \Omega_u \qquad \text{(A.27)}$$

for $\delta(n) \le t \le \delta_1(n)$. Equations (A.25)–(A.27) lead to condition (iii) in Lemma B.1 with $\sigma^2 = \nu_0\Omega_u$. Analogously, we can also prove that condition (ii) in Lemma B.1 is satisfied. Then, using (A.16) in Lemma B.1, we can prove that

$$\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} u_{t\blacktriangle} G_{th} \Rightarrow \mathcal{Z}_u^*, \qquad \text{(A.28)}$$

where $\mathcal{Z}_u^*$ is normal distribution with zero mean and variance matrix $\nu_0\Omega_u$, and is independent of $B_{d,\delta_0}(\Omega_v)$. It follows that

$$\Gamma_n(1,1) \Rightarrow \big[B_{d,\delta_0}(\Omega_v)' B_{d,\delta_0}(\Omega_v)\big]^{1/2} \mathcal{Z}_u^*. \qquad \text{(A.29)}$$

By using (A.19), (A.20), (A.22) and (A.29), we can establish (A.18). So the proof of Proposition A.2 is complete. □

Proposition A.3. *Suppose Assumptions 1 and 3 hold. Then*

$$\Gamma_n(2) \Rightarrow 2(q^\perp)' \Big\{ \int_{-1}^{1} B_{d,\frac{r+1}{2},*}(\Omega_v) dB_{\frac{r+1}{2},*}(\Omega_u) + \frac{1}{2}\Delta_{vu} \Big\}, \qquad \text{(A.30)}$$

which is defined as $\Gamma_{\delta_0}(2)$ in Section 3.

Proof. By the definition of $q_n^\perp$ and following the proof of (A.10), we can show that the leading term of $\Gamma_n(2)$ asymptotically is

$$\Gamma_n^*(2) = (q_n^\perp)' \Big[\frac{1}{nh} \sum_{t=1}^{n} (x_t - x_{\delta(n)}) u_t G_{th}\Big]. \qquad \text{(A.31)}$$

In view of the weak convergence (3.8) we have

$$\frac{1}{\sqrt{2nh}} \sum_{t=\delta(n)+1}^{\delta(n)+\lfloor 2nhp\rfloor} \big(x_t - x_{\delta(n)}\big) = \frac{1}{\sqrt{2nh}} \sum_{t=\lfloor\delta_0 n\rfloor-\lfloor nh\rfloor+1}^{\lfloor\delta_0 n\rfloor+\lfloor 2nhp\rfloor} v_t \Rightarrow B_{d,p,*}(\Omega_v) =_d B_{d,\frac{r+1}{2},*}(\Omega_v),$$

for $p = \frac{r+1}{2} \in [0,1]$ with $r \in [-1,1]$, and $\delta(n) = \lfloor \delta_0 n \rfloor - \lfloor nh \rfloor$. Then, using Assumption 3 and Lemma B.2 in Appendix B we have

$$\Gamma_n^*(2) = 2(q_n^{\perp})' \left(\sum_{t=1}^{n} \frac{x_t - x_{\delta(n)}}{\sqrt{2nh}} \frac{u_t}{\sqrt{2nh}} h K_{th}(\delta_0) \right) = 2(q_n^{\perp})' \left(\sum_{t=1}^{n} \frac{x_t - x_{\delta(n)}}{\sqrt{2nh}} \frac{u_t}{\sqrt{2nh}} K\left(\frac{\frac{t}{n} - \delta_0}{h} \right) \right)$$

$$\begin{aligned}
&= 2(q_n^{\perp})' \left(\sum_{t=1}^{n} \frac{x_t - x_{\delta(n)}}{\sqrt{2nh}} \frac{u_t}{\sqrt{2nh}} K\left(\frac{t - \delta_0 n}{nh} \right) \right) \\
&= 2(q_n^{\perp})' \left\{ \sum_{t=\lfloor \delta_0 n \rfloor - \lfloor nh \rfloor + 1}^{\lfloor \delta_0 n \rfloor + \lfloor nh \rfloor} \frac{x_t - x_{\delta(n)}}{\sqrt{2nh}} \frac{u_t}{\sqrt{2nh}} K\left(\frac{t - \delta_0 n}{nh} \right) \right\} \\
&\sim 2(q_n^{\perp})' \left(\sum_{t=\lfloor \delta_0 n \rfloor - \lfloor nh \rfloor + 1}^{\lfloor \delta_0 n \rfloor + \lfloor nh \rfloor} \frac{x_t - x_{\delta(n)}}{\sqrt{2nh}} \frac{u_t}{\sqrt{2nh}} K\left(\frac{\lfloor 2nhp \rfloor - \lfloor nh \rfloor}{nh} \right) \right), \quad t = \delta(n) + \lfloor 2nhp \rfloor \\
&\sim 2(q_n^{\perp})' \left\{ \int_0^1 K(2p-1) B_{d,p,*}(\Omega_v) dB_{p,*}(\Omega_u) + \Delta_{vu} \int_0^1 K(2p-1)\, dp \right\} \\
&= 2(q_n^{\perp})' \left\{ \int_{-1}^1 K(r) B_{d,\frac{r+1}{2},*}(\Omega_v) dB_{\frac{r+1}{2},*}(\Omega_u) + \frac{1}{2} \Delta_{vu} \int_{-1}^1 K(r)\, dr \right\},
\end{aligned}$$

giving

$$\Gamma_n^*(2) \Rightarrow 2(q^{\perp})' \Big\{ \int_{-1}^1 K(r) B_{d,\frac{r+1}{2},*}(\Omega_v) dB_{\frac{r+1}{2},*}(\Omega_u) + \frac{1}{2} \Delta_{vu} \Big[\int_{-1}^1 K(r) dr \Big] \Big\}. \tag{A.32}$$

Noting that $\int_{-1}^1 K(r)dr = 1$ in Assumption 3(i), the proof of Proposition A.3 is complete. □

With Propositions A.1–A.3 in hand, we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. Observe that

$$\begin{aligned}
\widehat{f}_n(\delta_0) - f(\delta_0) &= \Big[\sum_{t=1}^n x_t x_t' G_{th}(\delta_0) \Big]^+ \Big\{ \sum_{t=1}^n x_t x_t' \big[f\big(\frac{t}{n}\big) - f(\delta_0) \big] G_{th}(\delta_0) \Big\} + \\
&\quad \Big[\sum_{t=1}^n x_t x_t' G_{th}(\delta_0) \Big]^+ \Big[\sum_{t=1}^n x_t u_t G_{th}(\delta_0) \Big].
\end{aligned} \tag{A.33}$$

By Taylor expansion of $f(\cdot)$, and Assumption 2, we can show that

$$f\big(\frac{t}{n}\big) - f(\delta_0) = O(h^{\gamma_1}) \tag{A.34}$$

when $\left| \frac{t}{n} - \delta_0 \right| \le Ch$. By (A.34) and following the proof of Proposition A.1, we can easily prove that

$$\Big[\sum_{t=1}^n x_t x_t' G_{th}(\delta_0) \Big]^+ \Big\{ \sum_{t=1}^n x_t x_t' \big[f\big(\frac{t}{n}\big) - f(\delta_0) \big] G_{th}(\delta_0) \Big\} = O_P(h^{\gamma_1}). \tag{A.35}$$

Then, using Propositions A.1–A.3, (A.35) in conjunction with the condition $n^2 h^{1+2\gamma_1} = o(1)$, we can prove (3.5) in Theorem 3.1. □

Proof of Theorem 3.2. Let $D_{n*} = I_2 \otimes D_n$, $Q_{n*} = I_2 \otimes Q_n$,

$$\Delta_{n*} \equiv \Delta_{n\delta_0 *} = D_{n*}^+ Q_{n*}' \begin{pmatrix} \sum_{t=1}^n x_t x_t' G_{th} & \sum_{t=1}^n x_t x_t' G_{th*} \\ \sum_{t=1}^n x_t x_t' G_{th*} & \sum_{t=1}^n x_t x_t' G_{th**} \end{pmatrix} Q_{n*} D_{n*}^+$$

and

$$\Gamma_{n*} \equiv \Gamma_{n\delta_0 *} = D_{n*}^{+} Q_{n*}' \begin{pmatrix} \sum_{t=1}^{n} x_t u_t G_{th} \\ \sum_{t=1}^{n} x_t u_t G_{th*} \end{pmatrix},$$

where

$$G_{th*} = \Big(\frac{t-\delta_0 n}{nh}\Big) K\Big(\frac{t-\delta_0 n}{nh}\Big), \quad G_{th**} = \Big(\frac{t-\delta_0 n}{nh}\Big)^2 K\Big(\frac{t-\delta_0 n}{nh}\Big).$$

Following the proofs of Propositions A.1–A.3, we can establish that

$$\Delta_{n*} \Rightarrow \Delta_{\delta_0 *}, \quad \Gamma_{n*} \Rightarrow \Gamma_{\delta_0 *}, \tag{A.36}$$

where both $\Delta_{\delta_0 *}$ and $\Gamma_{\delta_0 *}$ are defined in Section 3. By some elementary calculations for the local linear fitting, we obtain

$$\begin{aligned} D_n Q_n' \big[\widetilde{f}_n(\delta_0) - f(\delta_0)\big] &= e D_{n*} Q_{n*}' \Big[\begin{array}{c} \widetilde{f}_n(\delta_0) - f(\delta_0) \\ h\widetilde{f}_n'(\delta_0) - hf(\delta_0) \end{array} \Big] \\ &= e \Delta_{n*} \Gamma_{n*} + O_P(h^2 D_n Q_n). \end{aligned} \tag{A.37}$$

Equations (A.36) and (A.37) lead to (3.11) in Theorem 3.2. Meanwhile, (3.11) and the bandwidth condition $n^2 h^5 = o(1)$ together imply that (3.12) holds. The proof of Theorem 3.2 is then complete. □

Proof of Proposition 4.1. Note that

$$\begin{aligned} D_n Q_n' \big\{\widehat{f}_{n,bc}(\delta_0) - f(\delta_0)\big\} &= D_n Q_n' \Big\{ \Big[\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0)\Big]^{+} \sum_{t=1}^{n} x_t y_t K_{th}(\delta_0) - f(\delta_0) \Big\} \\ &\quad - D_n Q_n' \Big[\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0)\Big]^{+} Q_n D_n \widehat{\Gamma}_{n,bc} \\ &= D_n Q_n' \big\{\widehat{f}_n(\delta_0) - f(\delta_0)\big\} - D_n Q_n' \Big[\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0)\Big]^{+} Q_n D_n \widehat{\Gamma}_{n,bc}. \end{aligned}$$

Note that $\widehat{\Delta}_{vu}$ is assumed to be a consistent estimate of Δ_{vu}. By the definition of $\widehat{\Gamma}_{n,bc}$ in (4.2) and using Theorem 3.1 and Proposition A.1, we can show that the second-order bias of $\widehat{f}_n(\delta_0)$ in the direction $q_n^{\perp}$ can be eliminated, and (4.3) can thus be proved. □

Proof of Proposition 4.2. Let $\widehat{f}_{nt} = \widehat{f}\big(\frac{t}{n}\big)$ and recall that $f_t = f\big(\frac{t}{n}\big)$. Observe that

$$\widehat{u}_t = y_t - x_t' \widehat{f}_{nt} = u_t - x_t'\big(\widehat{f}_{nt} - f_t\big),$$

which implies that

$$\begin{aligned} \widehat{\Delta}_{vu}(j) &= \frac{1}{\tau_n^* - \tau_n} \sum_{t=\tau_n+1}^{\tau_n^*} v_{t-j} \widehat{u}_t \\ &= \frac{1}{\tau_n^* - \tau_n} \sum_{t=\tau_n+1}^{\tau_n^*} v_{t-j} u_t - \frac{1}{\tau_n^* - \tau_n} \sum_{t=\tau_n+1}^{\tau_n^*} v_{t-j} x_t' \big(\widehat{f}_{nt} - f_t\big) \\ &\equiv \overline{\Delta}_{vu}(j) - \widetilde{\Delta}_{vu}(j), \end{aligned} \tag{A.38}$$

for $j = 1, \cdots, l_n$, where $\tau_n = \lfloor \tau_* n \rfloor$ and $\tau_n^* = \lfloor (1-\tau_*)n \rfloor$. Using (A.38), we have

$$\widehat{\Delta}_{vu} = \sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big)\widehat{\Delta}_{vu}(j) = \sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big)\overline{\Delta}_{vu}(j) - \sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big)\widetilde{\Delta}_{vu}(j). \tag{A.39}$$

We first prove the second term on the right hand side of (A.39) is asymptotically negligible. By the definition of $\widehat{f}_{nt}$ in (2.1) and letting $G_{sh}(t/n) = hK_{st}(t/n)$, we have

$$\begin{aligned}
\widehat{f}_{nt} - f_t &= \left[\sum_{s=1}^{n} x_s x_s' G_{sh}(t/n)\right]^+ \left[\sum_{s=1}^{n} x_s y_s G_{sh}(t/n)\right] - f_t \\
&= \left[\sum_{s=1}^{n} x_s x_s' G_{sh}(t/n)\right]^+ \left[\sum_{s=1}^{n} x_s u_s G_{sh}(t/n)\right] + \\
&\quad \left[\sum_{s=1}^{n} x_s x_s' G_{sh}(t/n)\right]^+ \left[\sum_{s=1}^{n} x_s f_s G_{sh}(t/n)\right] - f_t \\
&\equiv \Theta_{nt}(u) + \Theta_{nt}(f)
\end{aligned}$$

for $t = \tau_n + 1, \cdots, \tau_n^*$. By transforming coordinates, we can show that

$$\begin{aligned}
\Theta_{nt}(u) &= Q_{nt} D_n^+ \left[D_n^+ Q_{nt}' \sum_{s=1}^{n} x_s x_s' G_{sh}(t/n) Q_{nt} D_n^+\right]^+ \left[D_n^+ Q_{nt}' \sum_{s=1}^{n} x_s u_s G_{sh}(t/n)\right] \\
&\equiv Q_{nt} D_n^+ \Theta_{nt,1}^+ \Theta_{nt,2}(u),
\end{aligned} \tag{A.40}$$

where $Q_{nt} = \big[q_{nt}, q_{nt}^{\perp}\big]$ with

$$q_{nt} = \frac{b_{nt}}{(b_{nt}' b_{nt})^{1/2}} = \frac{b_{nt}}{\|b_{nt}\|}, \quad b_{nt} = \frac{1}{\sqrt{n}} x_{\lfloor t-nh \rfloor}$$

and $q_{nt}^{\perp}$ such that $Q_{nt}' Q_{nt} = I_d$.

Note that $x_t = x_{\lfloor t-nh \rfloor} + x_t - x_{\lfloor t-nh \rfloor}$ and $x_{\lfloor t-nh \rfloor}' q_{nt}^{\perp} = 0$ with probability 1. Then, using Lemmas B.3 and B.4 in Appendix B and by Taylor expansion of $f(\cdot)$, we can prove that

$$\begin{aligned}
\sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big)\widetilde{\Delta}_{vu}(j) &= \frac{1}{\tau_n^* - \tau_n} \sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big) \sum_{t=\tau_n+1}^{\tau_n^*} v_{t-j} x_t' \big(\widehat{f}_{nt} - f_t\big) \\
&= \frac{1}{\tau_n^* - \tau_n} \sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big) \sum_{t=\tau_n+1}^{\tau_n^*} v_{t-j} x_t' \big[\Theta_{nt}(u) + \Theta_{nt}(f)\big] \\
&= o_P(1) + O_P(\sqrt{n} h l_n) = o_P(1)
\end{aligned} \tag{A.41}$$

as $l_n = o\big(\frac{1}{\sqrt{nh}}\big)$.

We finally consider $\sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big)\overline{\Delta}_{vu}(j)$. Since $\tau_n^* - \tau_n \to \infty$ when $\tau_* \in \big(0, \frac{1}{2}\big)$, it follows as in Park and Phillips (1988, 1989) that

$$\sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big)\overline{\Delta}_{vu}(j) = \sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big)\Big(\frac{1}{\tau_n^* - \tau_n} \sum_{t=\tau_n+1}^{\tau_n^*} v_{t-j} u_t\Big) = \Delta_{vu} + o_P(1). \tag{A.42}$$

Using (A.41) and (A.42), we can complete the proof of Proposition 4.2. □

B Lemmas and supplemental proofs

This appendix gives some technical lemmas which play a critical role in the proofs in Appendix A, and provides supplemental proofs of key steps in the proof of Proposition 4.2. The first result is a central limit theorem for m-dependent random variables from Berk (1973).

Lemma B.1. *Let $\{z_i:\ i\geq 1\}$ be a sequence of m_n-dependent random variables with zero mean, where m_n may tend to infinity as n tends to infinity. Suppose that*

(i) For some $\varrho>0$, $\mathbb{E}[|z_i|^{2+\varrho}]\leq C$, where C is positive and bounded, $\varrho>0$;

(ii) $\mathbb{E}[(z_i+\cdots+z_j)^2]\leq C(j-i)$ for any $i<j$ and $0<C<\infty$;

(iii) $\sigma^2\equiv\frac{1}{n}\mathbb{E}[(\sum_{i=1}^n z_i)^2]$ exists and is positive;

(iv) $m_n^{2+\frac{2}{\varrho}}=o(n)$. Then,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i\Rightarrow \mathrm{N}(0,\sigma^2). \tag{B.1}$$

as $n\to\infty$.

Let $g(\cdot)$ be a continuous function on $[0,1]$. The next result follows as in Ibragimov and Phillips (2008, theorem 3.1) and is useful in establishing the limit distribution of $\Gamma_n(2)$ which can been seen from the proof of Proposition A.3 in Appendix A.

Lemma B.2. *Suppose Assumption 1 holds. Then*

$$\frac{1}{n}\sum_{t=1}^n x_t u_t g_{nt}\Rightarrow\int_0^1 g(r)B_{d,r}(\Omega_v)dB_r(\Omega_u)+\Delta_{vu}\big[\int_0^1 g(r)dr\big], \tag{B.2}$$

where $g_{nt}=g\big(\frac{t}{n}\big)$ and Δ_{vu} is defined in Section 2.

The next two lemmas play a crucial role in the proof of Proposition 4.2. Their proofs have potential application in deriving sharp uniform convergence rates of nonparametric kernel estimators of the time-varying coefficient functions in cointegration models, a subject that will be studied in future work.

Lemma B.3. *Suppose that Assumptions 1 and 3 hold, and $l_n^{10+2\gamma_0+\varpi}=o(n^{5+\gamma_0}h^{9+\gamma_0})$ where $\varpi>0$ is arbitrarily small and γ_0 is defined in Assumption 1. Then,*

$$\big\|\Theta_{nt,2}(u)\big\|=o_P(\sqrt{nh}l_n^{-1}) \tag{B.3}$$

uniformly for $t=\tau_n+1,\cdots,\tau_n^*$, where $\Theta_{nt,2}(u)$ is defined in (A.40).

Proof. For $t=\tau_n+1,\cdots,\tau_n^*$, define

$$\begin{aligned}
\Theta_{nt,2}^*(u,1) &= q_{nt}'\big(\frac{x_{\lfloor(t-nh\rfloor}}{\sqrt{n}}\big)\Big[\frac{1}{\sqrt{nh}}\sum_{s=1}^n u_s G_{sh}(t/n)\Big],\\
\Theta_{nt,2}^*(u,2) &= 2(q_{nt}^{\perp})'\Big[\frac{1}{2nh}\sum_{s=1}^n\big(x_s-x_{\lfloor t-nh\rfloor}\big)u_s G_{sh}(t/n)\Big],
\end{aligned}$$

where q_{nt}, $q_{nt}^{\perp}$ and $G_{sh}(t/n)$ are defined as in the proof of Proposition 4.2. Following the proofs of Propositions A.2 and A.3 in Appendix A, $\big[\Theta_{nt,2}^*(u,1),\Theta_{nt,2}^*(u,2)'\big]'$ is the leading term of $\Theta_{nt,2}(u)$.

By the continuous mapping theorem (e.g. Billingsley, 1968), it is easy to show that $q_{nt}'\big(\frac{x_{\lfloor t-nh\rfloor}}{\sqrt{n}}\big) = \|b_{nt}'b_{nt}\|^{1/2} = O_P(1)$ and $q_{nt}^{\perp} = O_P(1)$ uniformly for $t = \tau_n + 1, \cdots, \tau_n^*$. Hence, to prove (B.3), we only need to prove

$$\sum_{s=1}^{n} u_s G_{sh}(t/n) = o_P(nhl_n^{-1}) \tag{B.4}$$

and

$$\sum_{s=1}^{n} \big(x_s - x_{\lfloor t-nh\rfloor}\big) u_s G_{sh}(t/n) = o_P((nh)^{3/2} l_n^{-1}) \tag{B.5}$$

uniformly for $t = \tau_n + 1, \cdots, \tau_n^*$.

Proof of (B.4): Note that

$$u_s = \overline{u}_s + (\widetilde{u}_{s-1} - \widetilde{u}_s), \tag{B.6}$$

where $\overline{u}_s = \big(\sum_{j=0}^{\infty} \Phi_{j,2}'\big)\varepsilon_s$ and $\widetilde{u}_s = \sum_{j=0}^{\infty} \widetilde{\Phi}_{j,2}'\varepsilon_{t-j}$ with $\widetilde{\Phi}_{j,2} = \sum_{k=j+1}^{\infty} \Phi_{k,2}$. Using the BN decomposition (B.6), we have

$$\begin{aligned}
\sum_{s=1}^{n} u_s K\big(\frac{s-t}{nh}\big) &= \sum_{s=1}^{n} \overline{u}_s K\big(\frac{s-t}{nh}\big) + \sum_{s=1}^{n} \widetilde{u}_{s-1} K\big(\frac{s-t}{nh}\big) - \sum_{s=1}^{n} \widetilde{u}_s K\big(\frac{s-t}{nh}\big) \\
&= \sum_{s=1}^{n} \overline{u}_s K\big(\frac{s-t}{nh}\big) + \sum_{s=1}^{n} \widetilde{u}_{s-1} K\big(\frac{s-1-t}{nh}\big) - \sum_{s=1}^{n} \widetilde{u}_s K\big(\frac{s-t}{nh}\big) + \\
&\quad \sum_{s=1}^{n} \widetilde{u}_{s-1}\Big[K\big(\frac{s-t}{nh}\big) - K\big(\frac{s-1-t}{nh}\big)\Big] \\
&= \sum_{s=1}^{n} \overline{u}_s K\big(\frac{s-t}{nh}\big) + \sum_{s=1}^{n} \widetilde{u}_{s-1}\Big[K\big(\frac{s-t}{nh}\big) - K\big(\frac{s-1-t}{nh}\big)\Big] + \\
&\quad \widetilde{u}_0 K\big(\frac{-t}{nh}\big) - \widetilde{u}_n K\big(\frac{n-t}{nh}\big).
\end{aligned}$$

By virtue of Assumption 3(i), for large enough n we have

$$\widetilde{u}_0 K\big(\frac{-t}{nh}\big) = \widetilde{u}_n K\big(\frac{n-t}{nh}\big) = 0$$

with probability 1 for any $t = \tau_n + 1, \cdots, \tau_n^*$, which indicates that

$$\sum_{s=1}^{n} u_s K\big(\frac{s-t}{nh}\big) = \sum_{s=1}^{n} \overline{u}_s K\big(\frac{s-t}{nh}\big) + \sum_{s=1}^{n} \widetilde{u}_{s-1}\Big[K\big(\frac{s-t}{nh}\big) - K\big(\frac{s-1-t}{nh}\big)\Big] \tag{B.7}$$

with probability 1 uniformly for $t = \tau_n + 1, \cdots, \tau_n^*$.

Define $T_k = \{\tau_n + (k-1)\lfloor r_n n\rfloor + 1, \cdots, \tau_n + k\lfloor r_n n\rfloor\}$ for $k = 1, 2, \cdots, R_n$, where

$$R_n = \lfloor \frac{\tau_n^* - \tau_n}{r_n n} \rfloor + 1, \quad r_n = \sqrt{n} h^2 l_n^{-(1+\varpi)},$$

in which $\varpi > 0$ is arbitrarily small. Let t_k be the smallest number in the set T_k (in fact t_k can be any number in T_k). By standard arguments, we have

$$\begin{aligned}
\max_{\tau_n+1\le t\le\tau_n^*} \Big|\sum_{s=1}^{n} \overline{u}_s K\big(\frac{s-t}{nh}\big)\Big| &\le \max_{1\le k\le R_n} \max_{t\in T_k} \Big|\sum_{s=1}^{n} \overline{u}_s \big[K\big(\frac{s-t}{nh}\big) - K\big(\frac{s-t_k}{nh}\big)\big]\Big| + \\
&\quad \max_{1\le k\le R_n} \Big|\sum_{s=1}^{n} \overline{u}_s K\big(\frac{s-t_k}{nh}\big)\Big|.
\end{aligned}$$

By the Markov inequality, we may show that

$$\max_{1\le k\le R_n}\max_{t\in T_k}\Big|\sum_{s=1}^{n}\overline{u}_s\big[K\big(\frac{s-t}{nh}\big)-K\big(\frac{s-t_k}{nh}\big)\big]\Big| = O_P\big(\frac{\sqrt{n}r_n}{h}\big) = O_P\big(\frac{nh}{l_n^{1+\varpi}}\big). \tag{B.8}$$

As $l_n\to\infty$ and $\varpi>0$, (B.8) implies that

$$\max_{1\le k\le R_n}\max_{t\in T_k}\Big|\sum_{s=1}^{n}\overline{u}_s\big[K\big(\frac{s-t}{nh}\big)-K\big(\frac{s-t_k}{nh}\big)\big]\Big| = o_P(nhl_n^{-1}). \tag{B.9}$$

On the other hand, by the Markov inequality and the Marcinkiewicz–Zygmund inequality for independent random variables (e.g., Theorem 2 in Section 10.3 of Chow and Teicher, 2003), we have for any $\epsilon>0$

$$\begin{aligned}
&\mathbb{P}\Big\{\max_{1\le k\le R_n}\Big|\sum_{s=1}^{n}\overline{u}_sK\big(\frac{s-t_k}{nh}\big)\Big|>\epsilon nhl_n^{-1}\Big\}\\
\le\;& \sum_{k=1}^{R_n}\mathbb{P}\Big\{\Big|\sum_{s=1}^{n}\overline{u}_sK\big(\frac{s-t_k}{nh}\big)\Big|>\epsilon nhl_n^{-1}\Big\}\\
\le\;& l_n^{4+\gamma_0}\sum_{k=1}^{R_n}\frac{\mathbb{E}\big|\sum_{s=1}^{n}\overline{u}_sK\big(\frac{s-t_k}{nh}\big)\big|^{4+\gamma_0}}{(\epsilon nh)^{4+\gamma_0}}\\
\le\;& \frac{CR_nl_n^{4+\gamma_0}(nh)^{\frac{4+\gamma_0}{2}}}{(nh)^{4+\gamma_0}} \le \frac{Cl_n^{5+\gamma_0+\varpi}}{n^{\frac{5+\gamma_0}{2}}h^{\frac{8+\gamma_0}{2}}} = o(1),
\end{aligned}$$

as $l_n^{10+2\gamma_0+\varpi}=o(n^{5+\gamma_0}h^{8+\gamma_0})$. Thus, we have

$$\max_{1\le k\le R_n}\Big|\sum_{s=1}^{n}\overline{u}_sK\big(\frac{s-t_k}{nh}\big)\Big| = o_P(nhl_n^{-1}). \tag{B.10}$$

By (B.9) and (B.10), we can show that

$$\max_{\tau_n\le t\le\tau_n^*}\Big|\sum_{s=1}^{n}\overline{u}_sK\big(\frac{s-t}{nh}\big)\Big| = o_P(nhl_n^{-1}). \tag{B.11}$$

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

$$\max_{\tau_n+1\le t\le\tau_n^*}\Big|\sum_{s=1}^{n}\widetilde{u}_{s-1}\Big[K\big(\frac{s-t}{nh}\big)-K\big(\frac{s-1-t}{nh}\big)\Big]\Big| = o_P(nhl_n^{-1}). \tag{B.12}$$

We can then complete the proof of (B.4) in view of (B.11) and (B.12). □

Proof of (B.5): The proof of (B.5) is more complicated than the proof of (B.4) because of the involvement of $x_s-x_{\lfloor t-nh\rfloor}$. Using the BN decomposition again as in (A.5), we have, for $s\ge\lfloor t-nh\rfloor+1$,

$$x_s-x_{\lfloor t-nh\rfloor} = \sum_{k=\lfloor t-nh\rfloor+1}^{s}v_k = \sum_{k=\lfloor t-nh\rfloor+1}^{s}\overline{v}_k+\widetilde{v}_{\lfloor t-nh\rfloor}-\widetilde{v}_s,$$

where the definitions of $\overline{v}_k$ and $\widetilde{v}_k$ are given in Section 2. Thus, to prove (B.5), we need to prove that

$$\sum_{s=1}^{n}\Big(\sum_{k=\lfloor t-nh\rfloor+1}^{s}\overline{v}_k\Big)u_sG_{sh}(t/n)=o_P((nh)^{3/2}l_n^{-1}), \tag{B.13}$$

$$\widetilde{v}_{\lfloor t-nh\rfloor}\sum_{s=1}^{n}u_sG_{sh}(t/n)=o_P((nh)^{3/2}l_n^{-1}), \tag{B.14}$$

$$\sum_{s=1}^{n}\widetilde{v}_su_sG_{sh}(t/n)=o_P((nh)^{3/2}l_n^{-1}) \tag{B.15}$$

uniformly for $t=\tau_n+1,\cdots,\tau_n^*$.

Noting that $\widetilde{v}_s$ and u_s are well defined stationary linear processes, by (B.4) and Assumptions 1 and 3, we can prove (B.14) and (B.15) easily. We next turn to the proof of (B.13). Define $\overline{x}_s(t)=\sum_{k=\lfloor t-nh\rfloor+1}^{s}\overline{v}_k$ to simplify notation, and $\overline{x}_s(t)=0$ if $s<\lfloor t-nh\rfloor+1$. Using the BN decomposition (B.6) again, we have

$$\begin{aligned}
\sum_{s=1}^{n}\overline{x}_s(t)u_sK\big(\frac{s-t}{nh}\big) &= \sum_{s=1}^{n}\overline{x}_s(t)\overline{u}_sK\big(\frac{s-t}{nh}\big)+\sum_{s=1}^{n}\overline{x}_s(t)\widetilde{u}_{s-1}K\big(\frac{s-t}{nh}\big)-\sum_{s=1}^{n}\overline{x}_s(t)\widetilde{u}_sK\big(\frac{s-t}{nh}\big)\\
&= \sum_{s=1}^{n}\overline{v}_s\overline{u}_sK\big(\frac{s-t}{nh}\big)+\sum_{s=1}^{n}\overline{x}_{s-1}(t)\overline{u}_sK\big(\frac{s-t}{nh}\big)+\sum_{s=1}^{n}\overline{x}_{s-1}(t)\widetilde{u}_{s-1}K\big(\frac{s-t}{nh}\big)+\\
&\quad \sum_{s=1}^{n}\overline{v}_s\widetilde{u}_{s-1}K\big(\frac{s-t}{nh}\big)-\sum_{s=1}^{n}\overline{x}_s(t)\widetilde{u}_sK\big(\frac{s-t}{nh}\big)\\
&= \sum_{s=1}^{n}\overline{x}_{s-1}(t)\overline{u}_sK\big(\frac{s-t}{nh}\big)+\sum_{s=1}^{n}\overline{x}_{s-1}(t)\widetilde{u}_{s-1}\Big[K\big(\frac{s-t}{nh}\big)-K\big(\frac{s-1-t}{nh}\big)\Big]+\\
&\quad \sum_{s=1}^{n}\overline{v}_s\overline{u}_sK\big(\frac{s-t}{nh}\big)+\sum_{s=1}^{n}\overline{v}_s\widetilde{u}_{s-1}K\big(\frac{s-t}{nh}\big).
\end{aligned}$$

Noting that $\widetilde{v}_s$, $\overline{u}_s$ and $\widetilde{u}_s$ are stationary, by Assumptions 1 and 3, we can prove that

$$\max_{\tau_n+1\le t\le\tau_n^*}\Big|\sum_{s=1}^{n}\overline{v}_s\overline{u}_sK\big(\frac{s-t}{nh}\big)\Big|=o_P((nh)^{3/2}l_n^{-1}), \tag{B.16}$$

$$\max_{\tau_n+1\le t\le\tau_n^*}\Big|\sum_{s=1}^{n}\overline{v}_s\widetilde{u}_{s-1}K\big(\frac{s-t}{nh}\big)\Big|=o_P((nh)^{3/2}l_n^{-1}). \tag{B.17}$$

Noting that $K\big(\frac{s-t}{nh}\big)-K\big(\frac{s-1-t}{nh}\big)\le C\frac{1}{nh}$, we may also show that

$$\max_{\tau_n+1\le t\le\tau_n^*}\Big|\sum_{s=1}^{n}\overline{x}_{s-1}(t)\widetilde{u}_{s-1}\Big[K\big(\frac{s-t}{nh}\big)-K\big(\frac{s-1-t}{nh}\big)\Big]\Big|=o_P((nh)^{3/2}l_n^{-1}). \tag{B.18}$$

By (B.16)–(B.18), to complete the proof of (B.5), we need only to show that

$$\max_{\tau_n+1\le t\le\tau_n^*}\Big|\sum_{s=1}^{n}\overline{x}_{s-1}(t)\overline{u}_sK\big(\frac{s-t}{nh}\big)\Big|=o_P((nh)^{3/2}l_n^{-1}). \tag{B.19}$$

Define $T_{k*} = \{\tau_n + (k-1)\lfloor r_{n*}n\rfloor + 1, \cdots, \tau_n + k\lfloor r_{n*}n\rfloor\}$ for $k = 1, 2, \cdots, R_{n*}$, where

$$R_{n*} = \lfloor \frac{\tau_n^* - \tau_n}{r_{n*}n} \rfloor + 1, \quad r_{n*} = \sqrt{n}h^{5/2}l_n^{-(1+\varpi)},$$

in which $\varpi > 0$ is arbitrarily small. Let t_{k*} be the smallest number in the set T_{k*}. By some standard arguments, we have

$$\begin{aligned}
\max_{\tau_n+1\le t\le\tau_n^*} \Big|\sum_{s=1}^n \overline{x}_{s-1}(t)\overline{u}_s K\big(\frac{s-t}{nh}\big)\Big| &\le \max_{1\le k\le R_{n*}} \max_{t\in T_{k*}} \Big|\sum_{s=1}^n \overline{x}_{s-1}(t)\overline{u}_s \big[K\big(\frac{s-t}{nh}\big) - K\big(\frac{s-t_{k*}}{nh}\big)\big]\Big| \\
&+ \max_{1\le k\le R_{n*}} \max_{t\in T_{k*}} \Big|\sum_{s=1}^n \big[\overline{x}_{s-1}(t) - x_{s-1}(t_{k*})\big]\overline{u}_s K\big(\frac{s-t_{k*}}{nh}\big)\Big| \\
&+ \max_{1\le k\le R_{n*}} \Big|\sum_{s=1}^n \overline{x}_{s-1}(t_{k*})\overline{u}_s K\big(\frac{s-t_{k*}}{nh}\big)\Big|.
\end{aligned}$$

Similar to the proof of (B.9) we can show that

$$\max_{1\le k\le R_{n*}} \max_{t\in T_{k*}} \Big|\sum_{s=1}^n \overline{x}_{s-1}(t)\overline{u}_s \big[K\big(\frac{s-t}{nh}\big) - K\big(\frac{s-t_{k*}}{nh}\big)\big]\Big| = o_P((nh)^{3/2}l_n^{-1}), \tag{B.20}$$

and

$$\max_{1\le k\le R_{n*}} \max_{t\in T_{k*}} \Big|\sum_{s=1}^n \big[\overline{x}_{s-1}(t) - x_{s-1}(t_{k*})\big]\overline{u}_s K\big(\frac{s-t_{k*}}{nh}\big)\Big| = o_P((nh)^{3/2}l_n^{-1}). \tag{B.21}$$

On the other hand, note that $\{(\overline{x}_{s-1}(t)\overline{u}_s, \mathcal{F}_s) : s \ge 1\}$ is a sequence of martingale differences, where $\mathcal{F}_s = \sigma(\varepsilon_s, \varepsilon_{s-1}, \cdots)$. Then, by the Markov inequality and Burkholder's inequality for martingale differences (e.g. Theorem 2.10 in Hall and Heyde, 1980), we have for any $\epsilon > 0$

$$\begin{aligned}
&\mathbb{P}\Big\{\max_{1\le k\le R_{n*}} \Big|\sum_{s=1}^n \overline{x}_{s-1}(t_{k*})\overline{u}_s K\big(\frac{s-t_{k*}}{nh}\big)\Big| > \epsilon(nh)^{3/2}l_n^{-1}\Big\} \\
\le\ & \sum_{k=1}^{R_{n*}} \mathbb{P}\Big\{\Big|\sum_{s=1}^n \overline{x}_{s-1}(t_{k*})\overline{u}_s K\big(\frac{s-t_{k*}}{nh}\big)\Big| > \epsilon(nh)^{3/2}l_n^{-1}\Big\} \\
\le\ & l_n^{4+\gamma_0} \sum_{k=1}^{R_{n*}} \frac{\mathbb{E}\big|\sum_{s=1}^n \overline{x}_{s-1}(t_{k*})\overline{u}_s K\big(\frac{s-t_{k*}}{nh}\big)\big|^{4+\gamma_0}}{(\epsilon nh)^{3(4+\gamma_0)/2}} \\
\le\ & \frac{CR_n l_n^{4+\gamma_0}(nh)^{4+\gamma_0}}{(nh)^{3(4+\gamma_0)/2}} \le \frac{Cl_n^{5+\gamma_0+\varpi}}{n^{\frac{5+\gamma_0}{2}}h^{\frac{9+\gamma_0}{2}}} = o(1),
\end{aligned}$$

as $l_n^{10+2\gamma_0+\varpi} = o(n^{5+\gamma_0}h^{9+\gamma_0})$. Thus, we have

$$\max_{1\le k\le R_{n*}} \Big|\sum_{s=1}^n \overline{x}_{s-1}(t_{k*})\overline{u}_s K\big(\frac{s-t_{k*}}{nh}\big)\Big| = o_P((nh)^{3/2}l_n^{-1}). \tag{B.22}$$

We can then complete the proof of (B.19) in view of (B.20)–(B.22). □

LEMMA B.4. *Suppose that Assumptions 1 and 3 hold, and Δ_δ is assumed to be nonsingular uniformly over $[\tau_*,\ 1-\tau_*]$ with probability 1 for any $0 < \tau_* < 1/2$. Then, the random matrix $\Theta_{nt,1}$ is nonsingular (in probability) uniformly over $t = \tau_n + 1, \cdots, \tau_n^*$, where $\Theta_{nt,1}$ is defined in (A.40).*

Proof. We first define

$$\Theta^*_{nt,1}(1) = q'_{nt}\big(\frac{x_{\lfloor t-nh\rfloor}}{\sqrt{n}}\big)\big(\frac{x_{\lfloor t-nh\rfloor}}{\sqrt{n}}\big)' q_{nt}\Big[\frac{1}{nh}\sum_{s=1}^{n} G_{sh}(t/n)\Big],$$

$$\Theta^*_{nt,1}(2) = 2\sqrt{2}q'_{nt}\big(\frac{x_{\lfloor t-nh\rfloor}}{\sqrt{n}}\big)\Big[\frac{1}{2nh}\sum_{s=1}^{n}\big(\frac{\eta_s}{\sqrt{2nh}}\big)' G_{sh}(t/n)\Big] q^{\perp}_{nt},$$

$$\Theta^*_{nt,1}(3) = 4(q^{\perp}_{nt})'\Big[\frac{1}{(2nh)^2}\sum_{s=1}^{n}\eta_s\eta'_s G_{sh}(t/n)\Big] q^{\perp}_{nt}$$

for $t=\tau_n+1,\cdots,\tau^*_n$, where η_s, $s\geq 1$, are defined as in (A.5). Let ε^*_t be an independent copy of ε_t and satisfy Assumption 1 in Section 2, $\overline{v}^*_t = (\sum_{j=0}^{\infty}\Phi'_{j,1})\varepsilon^*_t$ and $\eta^*_s = \sum_{t=1}^{s}\overline{v}^*_t$. Define

$$\Theta^*_n(1) = \frac{1}{nh}\sum_{s=1}^{2nh} K\big(\frac{s-nh}{nh}\big) = \frac{1}{nh}\sum_{s=1}^{n} G_{sh}(t/n), \tag{B.23}$$

$$\Theta^*_n(2) = \frac{1}{2nh}\sum_{s=1}^{n}\big(\frac{\eta^*_s}{\sqrt{2nh}}\big)' K\big(\frac{s-nh}{nh}\big) =_d \frac{1}{2nh}\sum_{s=1}^{n}\big(\frac{\eta_s}{\sqrt{2nh}}\big)' G_{sh}(t/n), \tag{B.24}$$

$$\Theta^*_n(3) = \frac{1}{(2nh)^2}\sum_{s=1}^{n}\eta^*_s(\eta^*_s)' K\big(\frac{s-nh}{nh}\big) =_d \frac{1}{(2nh)^2}\sum_{s=1}^{n}\eta_s\eta'_s G_{sh}(t/n). \tag{B.25}$$

It is easy to see that $\Theta^*_n(k)$, $k=1,2,3$, do not rely on t, and are independent of q_{nt} and $q^{\perp}_{nt}$. Furthermore, by Assumption 1, the BN decomposition, and the strong approximation result (e.g. Csörgö and Révész, 1981), there exists $B_{d,\delta}(\Omega_v)$ such that $\sup_{0<\delta<1}\left\|q_{n\lfloor n\delta\rfloor}-q_\delta\right\| = O_P(n^{-\frac{2+\gamma_0}{8+2\gamma_0}})$ and $q^{\perp}_{n\lfloor n\delta\rfloor}$ converges in probability to $q^{\perp}_\delta$ uniformly for $\delta\in(0,\ 1)$, where $q_\delta = B_{d,\delta}(\Omega_v)/\|B_{d,\delta}(\Omega_v)\|$ and $q^{\perp}_\delta$ is a $d\times(d-1)$ orthogonal complement random matrix such that $Q_\delta = \left[q_\delta,\ q^{\perp}_\delta\right]$ and $Q'_\delta Q_\delta = I_d$. Hence, by (B.23)–(B.25) and following the argument in the proof of Proposition A.1, we can show that, uniformly for $\delta\in(0,\ 1)$,

$$\begin{aligned}
\Theta^*_{n\lfloor n\delta\rfloor,1}(1) &=_d (q'_\delta q_\delta)\Theta^*_n(1),\\
\Theta^*_{n\lfloor n\delta\rfloor,1}(2) &=_d 2\sqrt{2}(q'_\delta q_\delta)^{1/2}\Theta^*_n(2) q^{\perp}_\delta,\\
\Theta^*_{n\lfloor n\delta\rfloor,1}(3) &=_d 4(q^{\perp}_\delta)'\Theta^*_n(3) q^{\perp}_\delta.
\end{aligned}$$

Then, applying the continuous mapping theorem to $\Theta^*_n(2)$ and $\Theta^*_n(3)$ and noting that Δ_δ is assumed to be nonsingular uniformly over $[\tau_*,\ 1-\tau_*]$ with probability 1, the random matrix $\Theta_{nt,1}$ is nonsingular (in probability) uniformly for $t=\tau_n+1,\cdots,\tau^*_n$. This completes the proof of Lemma B.4. □

C Model specification

In practical work, parametric forms of the time-varying coefficient function f_t are often convenient. We therefore consider the following parametric hypotheses

$$H_0:\ f(\delta) = g(\delta;\theta_0) \quad \text{versus} \quad H_1:\ f(\delta)\neq g(\delta;\theta_0), \tag{C.1}$$

where $g(\delta;\theta_0)$ is a pre-specified function indexed by a p-dimensional vector of unknown parameters θ_0 with $\theta_0 \in \Theta_0$, a compact parameter space in $\mathbb{R}^p$. In what follows we develop a statistic for testing H_0. We make the following assumption.

ASSUMPTION C. *Suppose that $g(\delta;\theta)$ is continuous in $\delta \in [0,1]$, is twice differentiable with respect to θ, and the matrix $\Lambda(\theta_0) = \frac{\partial^2 g(\delta;\theta)}{\partial\theta\partial\theta'}|_{\theta=\theta_0}$ is positive definite.*

Under the null H_0, the model has the explicit form

$$y_t = x_t' g\big(\frac{t}{n};\theta_0\big) + u_t, \tag{C.2}$$

and the parameter vector θ_0 may be estimated by various methods such as nonlinear least squares, giving $\widehat{\theta} = \arg\min_{\theta\in\Theta_0}\left\{\sum_{t=1}^n \left(y_t - x_t' g(\tau_t;\theta)\right)^2\right\}$. It follows from Assumptions 1 and C that $\widehat{\theta}-\theta_0 = O_P(n^{-1})$. Recall the definition of $\widehat{f}_n(\delta)$ in Section 2 and define

$$\overline{f}_n(\delta) = \Big[\sum_{t=1}^n x_t x_t' K\big(\frac{t-n\delta}{nh}\big)\Big]^+ \sum_{t=1}^n x_t x_t' g(\delta;\widehat{\theta})\ K\big(\frac{t-n\delta}{nh}\big) = \big[\overline{f}_{1n}(\delta), \overline{f}_{2n}(\delta)\big]'. \tag{C.3}$$

As in Section 5, let $x_{\delta(n)} = \big[x_{1,\delta(n)}, x_{2,\delta(n)}\big]'$, $b_n = \frac{1}{\sqrt{n}} x_{\delta(n)} = \frac{1}{\sqrt{n}}\big[x_{1,\delta(n)}, x_{2,\delta(n)}\big]'$ and $q_n = \big[\frac{x_{1,\delta(n)}}{\sqrt{n}||b_n||}, \frac{x_{2,\delta(n)}}{\sqrt{n}||b_n||}\big]' \equiv \big[q_{1n}, q_{2n}\big]'$ with $||b_n|| = \sqrt{\frac{1}{n}\big[x^2_{1,\delta(n)} + x^2_{2,\delta(n)}\big]}$. Define

$$\begin{aligned} R_{1n}(\delta) &= n\sqrt{h}\ q_{1n}\left(\widehat{f}_{1n}(\delta) - \overline{f}_{1n}(\delta)\right) + n\sqrt{h}\ q_{2n}\left(\widehat{f}_{2n}(\delta) - \overline{f}_{2n}(\delta)\right), \\ R_{2n}(\delta) &= nh\ p_{1n}\left(\widehat{f}_{1n}(\delta) - \overline{f}_{1n}(\delta)\right) + nh\ p_{2n}\left(\widehat{f}_{2n}(\delta) - \overline{f}_{2n}(\delta)\right), \end{aligned} \tag{C.4}$$

where $p_{1n} = q_{2n}$ and $p_{2n} = -q_{1n}$ as in Section 5.

The components $R_{in}(\delta)$ in (C.4) measure scaled departures of the parametrically fitted elements $\bar{f}_{in}(\delta)$ from the nonparametric estimates $\hat{f}_{in}(\delta)$. These components may be used jointly (or individually) to test the validity of the parametric specification (C.2). We introduce a joint test statistic of the form

$$L_n(h) = \int_0^1 (R_{1n}, R_{2n})(R_{1n}, R_{2n})'\, d\delta = \int_0^1 \left(R^2_{1n}(\delta) + R^2_{2n}(\delta)\right) d\delta, \tag{C.5}$$

where h can be chosen by a suitable method such as the cross-validation method. The following proposition gives the limit theory for the statistic $L_n(h)$. Recall the definitions of $\Delta_\delta = \Delta(\delta)$ and $\Gamma_\delta = \Gamma(\delta)$ given in Section 3. Critical values for the implementation of this test statistic in empirical work may be obtained by bootstrap methods, as discussed in Section 6.

PROPOSITION C.1. *Suppose that the conditions of Theorem 3.1 and Assumption C are satisfied. Under the null hypothesis H_0, we have as $n \to \infty$*

$$L_n(h) \Rightarrow L(\Delta, \Gamma), \tag{C.6}$$

where $L(\Delta,\Gamma) = \int_0^1 \Gamma'(\delta)\Delta^{-2}(\delta)\Gamma(\delta)d\delta$.

PROOF: The proof follows directly from Theorem 3.1 in Section 3 and its proof in Appendix A. In fact

$$\begin{aligned}
\widehat{f}_n(\delta) - \overline{f}_n(\delta) &= \Big[\sum_{t=1}^{n} x_t x_t' K\big(\frac{t-n\delta}{nh}\big)\Big]^{+} \sum_{t=1}^{n} x_t \big[y_t - x_t'\widehat{g}(\delta;\widehat{\theta})\big]\ K\big(\frac{t-n\delta}{nh}\big) \\
&= \Big[\sum_{t=1}^{n} x_t x_t' K\big(\frac{t-n\delta}{nh}\big)\Big]^{+} \sum_{t=1}^{n} x_t u_t + \Big[\sum_{t=1}^{n} x_t x_t' K\big(\frac{t-n\delta}{nh}\big)\Big]^{+} \times \\
&\quad \sum_{t=1}^{n} x_t x_t' \big[g(\delta;\theta_0) - g(\delta;\widehat{\theta})\big]\ K\big(\frac{t-n\delta}{nh}\big) \\
&\equiv J_{1n} + J_{2n}. \qquad \text{(C.7)}
\end{aligned}$$

Note that J_{1n} is the same as the leading term in the first equation of the proof of Theorem 3.1. Note also that $J_{2n} = O_P\left(n^{-1}\right)$ by Assumption C. The fact that $n^2 h J_{2n}^2 = O_P(h) = o_P(1)$, along with the rest of the proof of Theorem 3.1, completes the proof of Proposition C.1. □

References

Athanasopoulos, G., O. T. de Carvalho Guill'en, J. V. Issler, and F. Vahid (2011): "Model Selection, Estimation and Forecasting in VAR Models with Short–run and Long–run Restrictions", *Journal of Econometrics*, 164, 116–129.

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

Berk, K. N. (1973): "A Central Limit Theorem for m-Dependent Random Variables with Unbounded m", *Annals of Probability*, 1, 352-354.

Cai, Z. (2007): "Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors", *Journal of Econometrics*, 136, 163-188.

Cai, Z., Q. Li, and J. Y. Park (2009): "Functional-Coefficient Models for Nonstationary Time Series Data", *Journal of Econometrics*, 148, 101-113.

Chen, B., and Y. Hong (2012): "Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression", *Econometrica*, 80, 1157-1183.

Chow, Y. S., and H. Teicher (2003): *Probability Theory: Independence, Interchangeability, Martingales* (third edition). Springer-Verlag, New York.

Csörgö, M., and P. Révész (1981): *Strong Approximations in Probability and Statistics*, Academic Press.

Engle, R., and C. W. J. Granger (1987): "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251-276.

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

Gao, J., and M. L. King (2011): "A New Test in Parametric Linear Models against Nonparametric Autoregressive Errors," *Working paper* available at http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011/wp20-11.pdf.

Gao, J., and P. C. B. Phillips (2012): “Trending Time Series with Non– and Semi–Parametric Cointegration”, *Working paper* available at http://www.jitigao.com/working-papers.html.

Gao, J., and P. C. B. Phillips (2013): “Semiparametric Estimation in Triangular Simultaneous Equations with Nonstationarity”, *Journal of Econometrics*, 176, 59–79.

Hall, P., and C. C. Heyde (1980): *Martingale Limit Theory and Its Applications*, Academic Press.

Hall, P., J. Horowitz, and B. Jing (1995): “On Blocking Rules for the Bootstrap with Dependent Data,” *Biometrika* 82, 561–574.

Hong, S. H., and P. C. B. Phillips (2010): “Testing Linearity in Cointegrating Relations with an Application to PPP”, *Journal of Business and Economic Statistics*, 28, 96-114.

Ibragimov, R., and P. C. B. Phillips (2008): “Regression Asymptotics using Martingale Convergence Methods”, *Econometric Theory*, 24, 888-947.

Johansen, S. (1988): “Statistical Analysis of Cointegration Vectors”, *Journal of Economic Dynamics and Control* 12, 231–254.

Juhl, T., and Z. Xiao (2005): “Partially Linear Models with Unit Roots”, *Econometric Theory*, 21 877-906.

Karlsen, H. A., T. Myklebust, and D. Tjøstheim (2007): “Nonparametric Estimation in a Nonlinear Cointegration Type Model”, *Annals of Statistics*, 35, 252-299.

Kasparis, I. (2008): “Detection of Functional Form Misspecification in Cointegrating Relations”, *Econometric Theory*, 24,1373-1403.

Kasparis, I., and P. C. B. Phillips (2012): “Dynamic Misspecification in Nonparametric Cointegrating Regression”, *Journal of Econometrics*, 168, 270-284.

Kasparis, I., P. C. B. Phillips, and T. Magdalinos (2013): “Nonlinearity Induced Weak Instrumentation”, *Econometric Reviews* (forthcoming).

Li, D., J. Chen, and Z. Lin (2011): “Statistical Inference in Partially Time-Varying Coefficient Models”, *Journal of Statistical Planning and Inference*, 141, 995-1013.

Li, K., D. Li, Z. Liang, and C. Hsiao (2013): “Semiparametric Profile Likelihood Estimation of Varying Coefficient Models with Nonstationary Regressors”, Working paper, Department of Econometrics and Business Statistics, Monash University.

Moon, H. R., and B. Perron (2004): “Testing for a Unit Root in Panels with Dynamic Factor”, *Journal of Econometrics* ,122, 81-126.

Park, J. Y. (1992): “Canonical Cointegrating Regressions”, *Econometrica*, 60, 119-143.

Park, J. Y., and S. B. Hahn (1999): “Cointegrating Regressions with Time Varying Coefficients”, *Econometric Theory*, 15, 664-703.

Park, J. Y., and P. C. B. Phillips (1988): “Statistical Inference in Regression with Integrated Processes: Part 1”, *Econometric Theory*, 4, 468-497.

——— (1989): "Statistical Inference in Regression with Integrated Processes: Part 2", *Econometric Theory*, 5, 95-131.

——— (2001): "Nonlinear Regressions with Integrated Time Series", *Econometrica*, 69, 117-161.

Phillips, P. C. B. (1989): "Partially Identified Econometric Models", *Econometric Theory*, 5, 181-240.

——— (1991): "Optimal Inference in Cointegrated Systems", *Econometrica*, 59, 283-306.

——— (1995): "Fully modified least squares and vector autoregression", *Econometrica*, 63, 1023-1078.

Phillips, P. C. B., and B. Hansen (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes", *Review of Economic Studies*, 57, 99-125.

Phillips, P. C. B., and T. Magdalinos (2008): "Limit theory for explosively cointegrated systems", *Econometric Theory*, 24, 865-887.

Phillips, P. C. B., and V. Solo (1992): "Asymptotics for Linear Processes", *Annals of Statistics*, 20, 971-1001.

Robinson, P. M. (1989): "Nonparametric Estimation of Time-Varying Parameters", in *Statistical Analysis and Forecasting of Economic Structural Change* ed. by P. Hackl. Springer, Berlin, pp. 164-253.

Saikkonen, P. (1995): "Problems with the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems", *Econometric Theory*, 11, 888-911.

Wand, M. P., and M. C. Jones (1995): *Kernel Smoothing*, Chapman and Hall.

Wang, Q., and P. C. B. Phillips (2009a): "Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression", *Econometric Theory*, 25, 710-738.

——— (2009b): "Structural Nonparametric Cointegrating Regression", *Econometrica*, 77, 1901-1948.

——— (2011): "Asymptotic Theory for Zero Energy Density Estimation with Nonparametric Regression Applications", *Econometric Theory*, 27, 236-269.

Xiao, Z. (2009): "Functional-Coefficient Cointegrating Regression", *Journal of Econometrics*, 152, 81-92.

Zhang, T., and W. B. Wu (2012): "Inference of Time Varying Regression Models", *Annals of Statistics*, 40, 1376-1402.