---

**Article:**
Phillips, Peter C. B., Li, Degui orcid.org/0000-0001-6802-308X and Gao, Jiti (2017) Estimating Smooth Structural Change in Cointegration Models. Journal of Econometrics. pp. 180-195. ISSN 0304-4076

https://doi.org/10.1016/j.jeconom.2016.09.013

---

# Estimating Smooth Structural Change in Cointegration Models[1]

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, and considers time-varying coefficient functions estimated 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 of the (functional) parameter space. Both rates are faster than the usual root-$nh$ rate for nonlinear models with smoothly changing coefficients and local stationarity. 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, which facilitates inference on the time varying functions. The finite sample properties of the methods and limit theory are explored in simulations. A brief empirical application to macroeconomic data shows that a linear cointegrating regression is rejected but finds support for alternative polynomial approximations for the time-varying coefficients in the regression.

*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]The authors thank the Co–Editor, Oliver Linton, an Associate Editor and three anonymous referees for comments and suggestions on the previous version of the paper. The work 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 Numbers: SES 0956687 and SES 1258258. Gao acknowledges support from the ARC under Grant Numbers: DP1314229 and DP150101012.

[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). However, parameter instability and functional form misspecification may 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, 2015; Gao and Phillips, 2013a). 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.

More specifically, we consider the following cointegration model with time-varying coefficient functions

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

where $f(\cdot)$ is a $d$-dimensional function of time (measured as a fraction of the sample size), $x_t$ is an $I(1)$ vector, and $u_t$ is a scalar process. The function $f(t/n)$ is sometimes called a fixed design and, in the present context, may be regarded as a weak trend function so that the model (1.1) captures potential drifts in the cointegrating linkage 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 $f(t/n)$ in the model (1.1). Thus model (1.1) allows the long term relationships among the trending time series to evolve smoothly over time, which provides

a more flexible framework than the parametric linear and nonlinear cointegration models. Some recent papers including Cai *et al* (2009), Xiao (2009), Gao and Phillips (2013b) and Li *et al* (2016) studied a nonlinear cointegrating model with functional coefficients and its generalised version, where the index variable in the functional coefficients is random, and developed the associated asymptotic theory. However, it is often difficult to select an appropriate random covariate as the index variable in practical applications and the requisite data may not be available. Such considerations partly motivate the use of a generic time-varying function to explore potential evolution in the cointegrating relationship between $y_t$ and $x_t$ in model (1.1). 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 (1.1) 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).

The present 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 $f(\cdot)$, and our asymptotic theory is fundamentally different from that in the paper by Park and Hahn (1999). Our treatment shows that estimation of this model by conventional kernel methods encounters a degeneracy problem in the weighted signal matrix (the denominator of the kernel estimator (2.1)), 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.

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 $\sqrt{nh}$-rate by $\sqrt{n}$ (type I super-consistency). The lower rate ($nh$) lies in the degenerate direction but is still super-consistent (type II super-consistency) for nonparametric estimators and exceeds the usual $\sqrt{nh}$-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. Importantly, inference about the time varying coefficient functions is unaffected by the degeneracy once the FM correction is made.

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 simulation findings on the finite sample properties of the methods and limit theory, and gives a practical application of these time-varying kernel regression methods to empirical relationships involving consumption, disposable income, investment and real interest rates. Section 6 concludes the paper. Proofs of the main theoretical results in the paper are given in Appendix A. Some supplementary technical materials and discussions on model specification testing are provided in an online supplement (Phillips, Li and Gao, 2016).

# 2 Kernel estimation degeneracy

Set $\tau = \lfloor n\delta \rfloor$ where the floor function $\lfloor \cdot \rfloor$ denotes integer part and $\delta \in [0,1]$ is the sample fraction corresponding to observation $t$. The functional response in (1.1) 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 $f(\lfloor n\delta \rfloor / n) \sim f(\delta)$ and the model response behaves locally around $\tau$ as $x'_\tau f(\tau/n) \sim x'_{\lfloor n\delta \rfloor} f(\delta)$. Evolution in the response mechanism over time is therefore captured as $\delta$ changes through the

functional dependence $x'_{\lfloor n\delta \rfloor} f(\delta)$.

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

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

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

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

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

The random matrix in the denominator of the local level regression estimation (2.1) is called the signal matrix throughout this paper as it carries the sample signal information in the regressors about the coefficient function $f$ locally in the neighborhood of the fixed point $\delta_0$. The weights $K_{th}(\delta_0)$ in the estimation (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 $\tau$. 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 $\mathbb{E}\left(x_t x_t'\right)$ 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 $\delta_0$ 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 errors $u_t$ 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 $\Phi(\mathcal{L}) = \sum_{j=0}^{\infty} \Phi_j \mathcal{L}^j$, $\Phi_j$ 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

$(d+1)$. Such a generation on $u_t$ and $v_t$ has been commonly used in the literature such as Phillips (1995). Partition $\Phi_j$ as $\Phi_j = [\Phi_{j,1},\ \Phi_{j,2}]'$ so that

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

We use $\|\cdot\|$ to denote the Euclidean norm of a vector or the Frobenius norm of a matrix.

ASSUMPTION 1. *Let $\varepsilon_t$ 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) and noting that

$$n^{-1/2}\sum_{s=1}^{\lfloor nr\rfloor} \varepsilon_s \Rightarrow B_{\varepsilon,r}(\Lambda_0)$$

with $B_{\varepsilon,r}(\Lambda_0)$ being $(d+1)$-dimensional Brownian motion (BM) with variance matrix $\Lambda_0$,, we have for $t = \lfloor nr\rfloor$ and $0 < r \leq 1$,

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

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

where $B_{d+1,r}(\Omega) = [B_{d,r}(\Omega_v)', B_r(\Omega_u)]'$ is $(d+1)$-dimensional BM with variance matrix $\Omega$, 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 $\Omega$ 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 $\frac{1}{2} < \gamma_1 \leq 1$.*

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 $\beta$ 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

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

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 $\mathsf{N}_{n\delta_0}(h) = \left[\lfloor(\delta_0 - h)n\rfloor, \lfloor(\delta_0 + h)n\rfloor\right]$ 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 $\overline{v}_t = \left(\sum_{j=0}^\infty \Phi_{j,1}'\right)\varepsilon_t = \Phi_1(1)'\varepsilon_t$, 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

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

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 $\Lambda_0$. 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

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

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

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

Next observe that for $t \in \overline{\mathsf{N}}_{n\delta_0}(h)$ which is a set of integers in $\mathsf{N}_{n\delta_0}(h)$, we have $x_t - x_{\delta_n} = \sum_{s=\delta_n+1}^{t} v_s$ and then

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

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

$$\sup_{t\in\overline{\mathsf{N}}_{n\delta_0}(h)} \left|\left|\frac{x_t - x_{\delta_n}}{\sqrt{nh}}\right|\right| = O_P(1).$$

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

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

Similarly,

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

and

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

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

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

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 $\frac{1}{n^2}\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0)$ 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 $\delta_0$ and thereby the realized value of the limit process $B_{d,\delta_0}(\Omega_v)$ of the (standardized) regressor $x_t$. When the nonstationary regressor $x_t$ 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 $\delta_0$, 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 $\delta_0$. So, the kernel degeneracy is random and trajectory dependent.

This phenomenon of kernel degeneracy 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

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

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

$$\begin{aligned}
\Delta_{\delta_0}(1) &= b'b, \\
\Delta_{\delta_0}(2) &= \sqrt{2}\left(b'b\right)^{1/2} \left[\int_{-1}^{1} B^*_{d,(r+1)/2}(\Omega_v)' K(r) dr\right] q^{\perp}, \\
\Delta_{\delta_0}(3) &= 2(q^{\perp})' \left[\int_{-1}^{1} B^*_{d,(r+1)/2}(\Omega_v) B^*_{d,(r+1)/2}(\Omega_v)' K(r) dr\right] q^{\perp}, \\
\Gamma_{\delta_0}(1) &= \left(2b'b\right)^{1/2} \int_{-1}^{1} K(r) dB^*_{(r+1)/2}(\Omega_u), \\
\Gamma_{\delta_0}(2) &= 2(q^{\perp})' \left[\int_{-1}^{1} K(r) B^*_{d,(r+1)/2}(\Omega_v) dB^*_{(r+1)/2}(\Omega_u) + \frac{1}{2}\Delta_{vu}\right],
\end{aligned}$$

where the $(d+1)$-dimensional BM $B^*_{d+1,r}(\Omega) = \left[B^*_{d,r}(\Omega_v)', B^*_r(\Omega_u)\right]'$ is an independent copy of $B_{d+1,r}(\Omega) = [B_{d,r}(\Omega_v)', B_r(\Omega_u)]'$. Note that the variate $\int_{-1}^{1} K(r) dB^*_{(r+1)/2}(\Omega_u)$ has the same distribution as $\mathrm{N}(0, \frac{1}{2}\nu_0 \Omega_u)$ and is independent of $B_{d,\delta_0}(\Omega_v)$. The following theorem gives the asymptotic distribution of $\widehat{f}_n(\delta_0)$.

THEOREM 3.1. *Suppose Assumptions 1–3 are satisfied and $n^2 h^{1+2\gamma_1} = o(1)$. 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 $0 < \delta_0 < 1$ is fixed such that $\Delta_{\delta_0}$ 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 $\sqrt{nh}$ rate for kernel estimators in the stationary case. The $n\sqrt{h}$ rate in (3.6) can be understood as the $\sqrt{n^2 h}$ rate 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 $\sum_{t=1}^{n} x_t x_t' K(\frac{t-n\delta_0}{nh})$ 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

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

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 $\sqrt{n^2h}$ convergence rate was also obtained by Cai *et al* (2009) and Xiao (2009) in certain functional-coefficient models with multivariate nonstationary regressors and no degeneracies. The 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, i.e., the signal matrix in (2.1) has the asymptotic order of $n^2h$ in the direction $q_n$, stronger than the order of $nh$ in the stationary case.

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 $\sqrt{nh}$ 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 $\sqrt{nh}$ 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.,

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

So the effective sample size in the estimation of the component $(q^{\perp})'f(\delta_0)$ has the asymptotic order of $n^2h^2$, which is smaller than the effective sample size with the asymptotic order $n^2h$ 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 $\mathsf{N}_{n\delta_0}(h) = [\lfloor n\delta_0 - nh \rfloor, \lfloor \delta_0 + nh \rfloor]$. 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 $t = \delta_n + \lfloor 2nhp \rfloor \in \mathsf{N}_{n\delta_0}(h)$ with $p \in [0,1]$ the data increments still manifest

nonstationary characteristics. In particular, we have the following weak convergence

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

The stronger signal in these observations raises the overall signal in $(q_n^{\perp})' \left[\sum_{t=1}^n x_t x_t' K\left(\frac{t-n\delta_0}{nh}\right)\right] q_n^{\perp}$ 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.

Although the variate $\int_{-1}^1 K(r) dB^*_{(r+1)/2}(\Omega_u)$ in $\Gamma_{\delta_0}(1)$ has the centred normal distribution, the variate $\int_{-1}^1 K(r) B^*_{d,(r+1)/2}(\Omega_v) dB_{(r+1)/2}(\Omega_u)$ in $\Gamma_{\delta_0}(2)$ has the more complicated mixed normal distribution (Phillips and Hansen, 1990). This further makes the distribution theory in (3.6) different from that in the conventional stationary case which usually has the asymptotic normal distribution in all directions.

In the pure cointegration case with $\Delta_{vu} = 0$ and $\Omega_{vu} = 0$, the form of $\Gamma_{\delta_0}(2)$ can be simplified. Define $\overline{\Gamma}_{\delta_0}(2) = 2(q^{\perp})' \left[\int_{-1}^1 K(r) B^*_{d,(r+1)/2}(\Omega_v) dB^*_{(r+1)/2}(\Omega_u)\right]$ and $\overline{\Gamma}_{\delta_0}$ just as $\Gamma_{\delta_0}$ but with $\Gamma_{\delta_0}(2)$ replaced by $\overline{\Gamma}_{\delta_0}(2)$. Importantly, $\overline{\Gamma}_{\delta_0}(2)$ has a mixed normal distribution in this case in view of the independence of the Brownian motions $B^*_{d,(r+1)/2}(\Omega_v)$ and $B^*_{(r+1)/2}(\Omega_u)$ when $\Omega_{vu} = 0$. The following simplified mixed 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*

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

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

To eliminate bias effects in these nonparametric asymptotics we have imposed the bandwidth condition $n^2 h^{1+2\gamma_1} = o(1)$ on the bandwidth, which may be somewhat restrictive if $\gamma_1$ is close to its lower boundary of 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, 2015) 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 $\delta_0$,

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

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

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

where $\overline{a} = (a_1, \cdots, a_d)'$ and $\overline{b} = (b_1, \cdots, b_d)'$. The local linear estimator of $f(\delta_0)$ is defined as $\widetilde{f}_n(\delta_0) = \widetilde{a}$, where $(\widetilde{a}, \widetilde{b}) = \arg\min_{(\overline{a},\overline{b})} L_n(\overline{a}, \overline{b})$. 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 $\Delta_{\delta_0}$ but with $K(r)$ replaced by $rK(r)$ and $r^2K(r)$, respectively, and $\Gamma_{\delta_0 *}(2)$ is defined as $\Gamma_{\delta_0}$ 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 $\delta_0$ be fixed such that $0 < \delta_0 < 1$ and $\Delta_{\delta_0 *}$ 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 $n^2 h^5 = o(1)$, 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. Note that to eliminate the asymptotic bias of the local linear estimation, we impose the restriction of $n^2 h^5 = o(1)$, which is weaker than the corresponding restriction in Theorem 3.1. As discussed above, the bandwidth condition in Theorem 3.2 might be further relaxed if a higher-order local smoothing technique is applied.

# 4 FM-nonparametric kernel estimation

The one-sided long run covariance $\Delta_{vu}$ which appears in the limit functionals $\Gamma_{\delta_0}$ 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). In addition, there is an endogeneity bias effect arising from the correlation of the limit Brownian motions and these bias effects originate in the correlation between the regressor innovations and the equation error. The effects are second order, 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 the effects 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 $\widehat{\Delta}_{uu}$, $\widehat{\Delta}_{vu}, \widehat{\Delta}_{vv}$, $\widehat{\Omega}_{uv}$ and $\widehat{\Omega}_{vv}$ denote consistent estimates of $\Delta_{uu}$, $\Delta_{vu}, \Delta_{vv}, \Omega_{uv}$ and $\Omega_{vv}$, 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

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

with $\hat{y}_t^{\#} = y_t - \widehat{\Omega}_{uv}\widehat{\Omega}_{vv}^{-1}\Delta x_t$ and

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

and $\widehat{\Delta}_{vu}^{\#} = \widehat{\Delta}_{vu} - \widehat{\Delta}_{vv}\widehat{\Omega}_{vv}^{-1}\widehat{\Omega}_{vu}$. Since$\left(\widehat{\Delta}_{uu}, \widehat{\Delta}_{vu}, \widehat{\Delta}_{vv}, \widehat{\Omega}_{uv}, \widehat{\Omega}_{vv}\right)$= $(\Delta_{uu}, \Delta_{vu}, \Delta_{vv}, \Omega_{uv}, \Omega_{vv})$ + $o_P(1)$, the asymptotic distribution of $\widehat{f}_{n,bc}(\delta_0)$ is obtained in the same manner as the proof of Theorem 3.1 and has a mixed normal limit, just as that of $\widehat{f}_n(\delta_0)$ in the pure cointegration case shown in Corollary 3.1. In the present case, because of the removal of the endogeneity bias, the mixed normal limit theory involves the stochastic integral $\Gamma_{\delta_0}^{\#}(2) = 2(q^{\perp})'\left[\int_{-1}^{1} K(r) B_{d,(r+1)/2}^{*}(\Omega_v) dB_{(r+1)/2}^{*}(\Omega_{u.v})\right]$ where the univariate BM $B_{(r+1)/2}^{*}(\Omega_{u.v})$ has covariance matrix $\Omega_{u.v} = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ and is independent of the $d$-dimensional BM $B_{d,(r+1)/2}^{*}(\Omega_v)$ so that $\Gamma_{\delta_0}^{\#}(2)$ has a mixed normal distribution. We further define $\Gamma_{\delta_0}^{\#}(1) = (2b'b)^{1/2}\left[\int_{-1}^{1} K(r)\, dB_{(r+1)/2}^{*}(\Omega_{u.v})\right]$, which is normally distributed just as $\Gamma_{\delta_0}(1)$ but with the BM $B_{(r+1)/2}^{*}(\Omega_{u.v})$ in place of $B_{(r+1)/2}^{*}(\Omega_u)$. Importantly, these simplifications produce a mixed normal limit theory for $\widehat{f}_{n,bc}(\delta_0)$ which facilitates inference on the time varying coefficient functions, just as in the case of linear FM estimation of fixed coefficient cointegrating relations. Furthermore, we define $\Gamma_{\delta_0}^{\#}$ just as $\Gamma_{\delta_0}$ but with $[\Gamma_{\delta_0}(1), \Gamma_{\delta_0}(2)']$ replaced by $\left[\Gamma_{\delta_0}^{\#}(1), \Gamma_{\delta_0}^{\#}(2)'\right]$.

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

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

*for fixed* $0 < \delta_0 < 1$.

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 the FM-nonparametric kernel regression requires estimation of the long run covariance matrices $\left(\widehat{\Delta}_{vu}, \widehat{\Delta}_{vv}, \widehat{\Omega}_{uv}, \widehat{\Omega}_{vv}\right)$. For the following discussion it will be sufficient

to focus on estimation of the one-sided long run covariance matrix $\Delta_{vu}$. 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 (1.1). 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

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

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 lag 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 Appendix A, for $\tau_*$ 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. *Let the conditions in Theorem 3.1 hold with $\gamma_1 = 1$, $l_n^{10+2\gamma_0+\varpi} = o(n^{5+\gamma_0}h^{9+\gamma_0})$ for arbitrarily small $\varpi > 0$ and $l_n = o\left(\frac{1}{\sqrt{nh}}\right)$. Suppose that the random matrix $\Delta_\delta$ is nonsingular uniformly for $\delta \in [\tau_*,\ 1-\tau_*]$ with probability 1 for any $0 < \tau_* < 1/2$. Then we 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 $\varepsilon_i$. In particular, for $\gamma_0$ 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 $l_n = o\left(\frac{1}{\sqrt{nh}}\right)$ ensures that the asymptotic bias of the kernel estimates does not affect the consistency of $\widehat{\Delta}_{vu}$.

# 5 Numerical Studies

This section has two numerical examples. The first 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 in the paper. The second provides a practical application of time-varying kernel regression methods to examine empirical relationships involving consumption, disposable income, investment and real interest rates. In the simulations, we are particularly interested in the behavior of multivariate time-varying coefficient function estimators, the 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

$$\begin{aligned} &M1: \quad f_{1t} = f_1(t/n) = 1 + \frac{t}{n}, \quad f_{2t} = f_2(t/n) = e^{-\frac{t}{n}}; \\ &M2: \quad f_{1t} = f_1(t/n) = \cos(2\pi t/n), \quad f_{2t} = f_2(t/n) = \sin(2\pi t/n). \end{aligned}$$

The regressor $x_t = (x_{1,t}, x_{2,t})'$, with $x_{i,t} = x_{i,t-1} + v_{i,t}$ for $i = 1$ and 2, $v_{i,t} = \rho_i v_{i,t-1} + \varepsilon_{i,t}$, and the error $u_t = \rho u_{t-1} + \varepsilon_t$, with innovations $(\varepsilon_t, \varepsilon_{1,t}, \varepsilon_{2,t})$ that follow

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

with $\lambda_i = 0$ or $\lambda_i = 0.5$ for $i = 1, 2$ and 3. Simulations are conducted with sample size $n = 1,000$ and with $R = 10,000$ replications.

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

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

where we use $K(x) = \frac{1}{2} I\{-1 \le x \le 1\}$, with $I\{\cdot\}$ being the indicator function, and choose bandwidth values $h$ that will be specified later. Before reporting the simulation results, we use the following notation, based partly on earlier definitions. Let $\delta_n = \lfloor (\delta - h)n \rfloor$, $x_{\delta_n} = (x_{1,\delta_n}, x_{2,\delta_n})'$, $b_n(\delta) = \frac{1}{\sqrt{n}} x_{\delta_n} = \frac{1}{\sqrt{n}} (x_{1,\delta_n}, x_{2,\delta_n})'$ and $q_n(\delta) = b_n(\delta)/\|b_n(\delta)\| = \left[\frac{x_{1,\delta_n}}{\sqrt{n}\|b_n(\delta)\|}, \frac{x_{2,\delta_n}}{\sqrt{n}\|b_n(\delta)\|}\right]' \equiv [q_{1n}(\delta), q_{2n}(\delta)]'$. 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 $p_{1n}(\delta) = q_{2n}(\delta)$ and $p_{2n}(\delta) = -q_{1n}(\delta)$.

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

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

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

and compute averages of $g_{1n}(\delta)$ and $g_{2n}(\delta)$ as follows: $g_{in}(\delta) = \frac{1}{R}\sum_{j=1}^{R} g_{in,j}(\delta)$ for $i = 1, 2$ and $R = 10,000$, 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

$$g^*_{1n}(\delta) \equiv q_{1n}\left[\widehat{f}_{1n,bc}(\delta) - f_1(\delta)\right] + q_{2n}\left[\widehat{f}_{2n,bc}(\delta) - f_2(\delta)\right], \tag{5.6}$$

$$g^*_{2n}(\delta) \equiv p_{1n}\left[\widehat{f}_{1n,bc}(\delta) - f_1(\delta)\right] + p_{2n}\left[\widehat{f}_{2n,bc}(\delta) - f_2(\delta)\right], \tag{5.7}$$

where $\widehat{f}_{n,bc}(\cdot) \equiv \left[\widehat{f}_{1n,bc}(\cdot), \widehat{f}_{2n,bc}(\cdot)\right]'$ is defined as in (4.1). Averages of $g^*_{1n}(\delta)$ and $g^*_{2n}(\delta)$ are computed as follows: $g^*_{in}(\delta) = \frac{1}{R}\sum_{j=1}^{R} g^*_{in,j}(\delta)$ for $i = 1, 2$ and $R = 10,000$, 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 $\{\rho, \rho_i, \lambda_i, (\delta, h)\}$:

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

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

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

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

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

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

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 5.1: Absolute averages of $g_{in}(\delta)$ and $g^*_{in}(\delta)$ for the functional form *M1*

| 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 |

Table 5.2: Absolute averages of $g_{in}(\delta)$ and $g^*_{in}(\delta)$ for the functional form *M2*

| 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 near here

Fig. 5.2 near here

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

$$\widehat{f}_t(\delta|h) = \left[\sum_{s=1,\neq t}^{n} x_s x_s' K\big(\frac{s-n\delta}{nh}\big)\right]^+ \left[\sum_{s=1,\neq t}^{n} x_s y_s K\big(\frac{s-n\delta}{nh}\big)\right] \equiv \Big[\widehat{f}_{1t}(\delta|h), \widehat{f}_{2t}(\delta|h)\Big]', \tag{5.8}$$

and the cross-validation function

$$\mathrm{CV}_n(h) = \frac{1}{n}\sum_{t=1}^{n}\Big[y_t - x_t'\widehat{f}_t\big(\frac{t}{n}|h\big)\Big]^2, \tag{5.9}$$

and find an optimal bandwidth of the form

$$\widehat{h}_{\rm cv} = \arg\min_{h\in\mathcal{H}_n} {\rm CV}_n(h), \tag{5.10}$$

where $\mathcal{H}_n = \left[n^{-1}, n^{-\frac{2}{3}}\log^{-1}(n)\right]$. For $\delta > \widehat{h}_{\rm cv}$, define $\widehat{\delta}_n = \lfloor(\delta - \widehat{h}_{\rm cv})n\rfloor$, $x_{\widehat{\delta}_n} = \left(x_{1,\widehat{\delta}_n}, x_{2,\widehat{\delta}(n)}\right)'$, $\widehat{b}_n(\delta) = \frac{1}{\sqrt{n}}x_{\widehat{\delta}_n} = \frac{1}{\sqrt{n}}\left(x_{1,\widehat{\delta}_n}, x_{2,\widehat{\delta}_n}\right)'$ and $\widehat{q}_n(\delta) = \left[\frac{x_{1,\widehat{\delta}_n}}{\sqrt{n}\|\widehat{b}_n(\delta)\|}, \frac{x_{2,\widehat{\delta}_n}}{\sqrt{n}\|\widehat{b}_n(\delta)\|}\right]' \equiv [\widehat{q}_{1n}(\delta), \widehat{q}_{2n}(\delta)]'$. 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 $\widehat{p}_{1n}(\delta) = \widehat{q}_{2n}(\delta)$ and $\widehat{p}_{2n}(\delta) = -\widehat{q}_{1n}(\delta)$.

The plots shown in Figs. 5.1 and 5.2 are based on 500 replications. These plots clearly show 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 $g_{2n}(\cdot)$.

Example 5.2. We next apply 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]. Two formulations are considered 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): $c_{1t}$ is log per-capita real consumption, $i_{1t}$ is log per capita disposable income, and $r_t$ 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): $c_{2t}$ is log per-capita real consumption, $i_{2t}$ is log per capita real disposable income, and $z_t$ is log per capita real investment.

The series are plotted in Figs. 5.3 and 5.4, which show that $i_{1t}$, $i_{2t}$ and $z_t$ have trending components. In order to satisfy Assumption 1, we first eliminate the trends by introducing $z_{jt} = i_{jt} - \mu_j t$ with $\mu_j = \frac{1}{n}\sum_{t=1}^{n}(i_{jt} - i_{j,t-1})$ for $j = 1, 2$, and $z_{3t} = z_t - \mu_z t$ with $\mu_z = \frac{1}{n}\sum_{t=1}^{n}(z_t - z_{t-1})$. Figs. 5.3(b), 5.3(c), 5.4(b) and 5.4(c) show that the differenced versions of $z_{kt}$ for $k = 1, 2, 3$ and $r_t$ all appear stationary, leading us to define $y_t = c_{1t}$, $x_{1t} = z_{1t}$ and $x_{2t} = r_t$ for case (i), and $y_t = c_{2t}$, $x_{1t} = z_{2t}$ and $x_{2t} = z_{3t}$ for case (ii). Application of the nonparametric test in Gao and King (2011) for checking unit root nonstationarity gives $p$-values of 0.106 and 0.112 for $x_{1t}$ and $x_{2t}$ in case (i), and corresponding $p$-values of 0.132 and 0.116 for $x_{1t}$ and $x_{2t}$ in case (ii).

| Fig. 5.3(a) near here |
|---|
| Fig. 5.3(b) near here |
| Fig. 5.3(c) near here |
| Fig. 5.4(a) near here |
| Fig. 5.4(b) near here |
| Fig. 5.4(c) near here |

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

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

$$y_t = x_t' f\,(t/n) + u_t = x_t' f_t + u_t, \quad t = 1, \cdots, n, \tag{5.11}$$

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

$$\widehat{f}(\delta) = \left[\sum_{t=1}^{n} x_t x_t' K\big(\frac{t-n\delta}{nh}\big)\right]^{+} \left[\sum_{t=1}^{n} x_t y_t K\big(\frac{t-n\delta}{nh}\big)\right] \equiv \left[\widehat{f}_1(\delta), \widehat{f}_2(\delta)\right], \tag{5.12}$$

where $K(x) = \frac{1}{2} I\{-1 \le x \le 1\}$ as in Example 5.1, over $\delta \in (0, 1]$, and the bandwidth $h$ is chosen by cross-validation as described in (5.10). The nonparametric estimates of the two curves $f_i(\cdot)$ with their 95% confidence bands are shown in Figs. 5.5 and 5.6 for case (i), and in Figs. 5.7 and 5.8 for case (ii).

| Fig. 5.5 near here |
|---|
| Fig. 5.6 near here |
| Fig. 5.7 near here |
| Fig. 5.8 near here |

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 values $\theta_i$ and pre-specified functions $g_i(\cdot;\cdot)$. We have done some pre–testing for all possible linear forms and other polynomial approximations before we propose using the parametric polynomial approximations in equations (5.13)–(5.16) below. Therefore, for case (i), in Figs. 5.5 and 5.6, we also consider polynomial fitted specifications of the form:

$$g_1(\delta; \widehat{\theta}_1) = \widehat{\theta}_{01} + \sum_{j=1}^{6} \widehat{\theta}_{j1} \delta^j, \tag{5.13}$$

$$g_2(\delta; \widehat{\theta}_2) = \widehat{\theta}_{02} + \sum_{j=1}^{5} \widehat{\theta}_{j2} \delta^j, \tag{5.14}$$

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 case (ii), in Figs. 5.7 and 5.8, we consider the fitted polynomial specifications:

$$g_1(\delta; \widehat{\theta}_1) = \widehat{\theta}_{01} + \sum_{j=1}^{3} \widehat{\theta}_{j1} \delta^j, \tag{5.15}$$

$$g_2(\delta; \widehat{\theta}_2) = \widehat{\theta}_{02} + \sum_{j=1}^{3} \widehat{\theta}_{j2} \delta^j, \tag{5.16}$$

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. 5.5–5.8 show that $f_1(\delta)$ and $f_2(\delta)$ are reasonably well captured by the parametric forms $g_1(\delta;\widehat{\theta}_1)$ and $g_2(\delta;\widehat{\theta}_2)$. Interestingly, lower order polynomial approximations are 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 $H_0: \; y_t = x'_t g\left(\frac{t}{n};\theta_0\right) + u_t$ for a specific parametric form $g\left(\cdot;\theta_0\right)$. The test statistic used to assess this (joint) null hypothesis is $L_n(h)$, which is defined in Appendix C of the online supplement. This statistic measures scaled departures of parametrically fitted functional elements from their nonparametric counterparts. A detailed development of this test statistic and discussions on its limit theory are provided in Appendix C.

For cases (i) and (ii), by using the block bootstrap method introduced in Appendix C of the online supplement, 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.

This empirical example shows that while co-movement in macroeconomic data may well be supported by data inspection, linear cointegrating regressions with constant coefficients is often rejected in favor of models with time varying coefficients that allow the model to adapt to variations in the relationship over time. These variations are in many cases slowly moving and may be captured, as is done here, by kernel methods or by direct specifications in terms of simple basis functions like time polynomials.

# 6 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 the previous literature. 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.

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. The main asymptotic results in the paper, such as Theorems 3.1 and 3.2, are given for a single fixed time point and extension to the case of finitely many distinct time points is straightforward. However, for a rigorous asymptotic theory of the model specification test and bootstrap procedure discussed in Appendix C of the online supplement, extra work is needed in justification, including the development of uniform convergence results for kernel-based random elements over diverging time points. This work is beyond the scope of the present paper and is left for future research. Some uniform consistency results with sharp convergence rates for kernel estimation in nonstationary time varying coefficient models have been obtained in other work by the authors (Li *et al*, 2016). 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.

# Appendix 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 $\widehat{f}_n(\delta_0)$. An early discussion on weak convergence for a sequence of random matrices can be found in Phillips and Durlauf (1986) and in the overview paper by Phillips (1988). A recent treatment of weak convergence to stochastic integrals with drift is given in Liang *et al* (2016). 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* $\Delta_{\delta_0}$ *is defined in (3.4) of Section 3.*

PROOF. Observe that

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

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. Following the proof of (2.18), it is easy to show that

$$\Delta_n(1) = \Delta_n^*(1) + o_P(1). \tag{A.3}$$

By the definitions of $q_n$ and $b_n$, we have

$$\Delta_n^*(1) = q_n' \big(\frac{x_{\delta_n}}{\sqrt{n}}\big) \big(\frac{x_{\delta_n}}{\sqrt{n}}\big)' q_n + o_P(1) = b_n' b_n + o_P(1). \tag{A.4}$$

Furthermore, letting $b_{nv} = \frac{1}{\sqrt{n}} \sum_{t=1}^{\delta_n} \overline{v}_t$ with $\overline{v}_t$ defined in Section 2 and using (A.3), (A.4) and the BN decomposition (Phillips and Solo, 1992), we may show that

$$\Delta_n(1) = b_{nv}' b_{nv} + o_P(1) \equiv \Delta_n^\diamond(1) + o_P(1). \tag{A.5}$$

From the BN decomposition, we also have for $t \geq \delta_n$

$$\begin{aligned} x_t &= \sum_{s=1}^{t} \overline{v}_s + \widetilde{v}_0 - \widetilde{v}_t + x_0 \\ &= \Big( \sum_{s=1}^{\delta_n} \overline{v}_s + \widetilde{v}_0 - \widetilde{v}_{\delta_n} + x_0 \Big) + \Big( \sum_{s=\delta_n+1}^{t} \overline{v}_s \Big) + (\widetilde{v}_{\delta_n} - \widetilde{v}_t) \\ &\equiv x_{\delta_n} + \eta_t + \xi_t, \end{aligned} \tag{A.6}$$

where $\eta_{\delta_n} = 0$. Note that $(q_n^{\perp})' x_{\delta_n} = 0$ with probability 1. Hence, we have

$$\Delta_n(2) = \Delta_n^*(2) + o_P(1), \tag{A.7}$$

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.6) again, we have

$$\begin{aligned} \Delta_n^*(2) &= \frac{1}{nh^{3/2}} q_n' \big( \frac{x_{\delta_n}}{\sqrt{n}} \big) \sum_{t=1}^{n} \big( \frac{\eta_t + \xi_t}{\sqrt{n}} \big)' q_n^{\perp} G_{th} + \\ &\quad \frac{1}{nh^{3/2}} \sum_{t=1}^{n} q_n' \big( \frac{\eta_t + \xi_t}{\sqrt{n}} \big) \big( \frac{\eta_t + \xi_t}{\sqrt{n}} \big)' q_n^{\perp} G_{th} \\ &\equiv \Delta_n^*(2,1) + \Delta_n^*(2,2). \end{aligned} \tag{A.8}$$

Note that

$$\begin{aligned} \Delta_n^*(2,2) &= \frac{1}{nh^{3/2}} \sum_{t=1}^{n} q_n' \big( \frac{\eta_t}{\sqrt{n}} \big) \big( \frac{\eta_t}{\sqrt{n}} \big)' q_n^{\perp} G_{th} + \frac{1}{nh^{3/2}} \sum_{t=1}^{n} q_n' \big( \frac{\eta_t}{\sqrt{n}} \big) \big( \frac{\xi_t}{\sqrt{n}} \big)' q_n^{\perp} G_{th} + \\ &\quad \frac{1}{nh^{3/2}} \sum_{t=1}^{n} q_n' \big( \frac{\xi_t}{\sqrt{n}} \big) \big( \frac{\eta_t}{\sqrt{n}} \big)' q_n^{\perp} G_{th} + \frac{1}{nh^{3/2}} \sum_{t=1}^{n} q_n' \big( \frac{\xi_t}{\sqrt{n}} \big) \big( \frac{\xi_t}{\sqrt{n}} \big)' q_n^{\perp} G_{th} \\ &\equiv \Delta_n^*(2,2,1) + \Delta_n^*(2,2,2) + \Delta_n^*(2,2,3) + \Delta_n^*(2,2,4). \end{aligned} \tag{A.9}$$

We next show that

$$\Delta_n^*(2,2,k) = o_P(1) \quad \text{for} \quad k = 1, \cdots, 4. \tag{A.10}$$

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. Using the weak convergence results such as (2.3) in Section 2, we can prove

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

as $h \to 0$, and

$$\begin{aligned}\Delta_n^*(2,2,4) &= \frac{1}{nh^{3/2}}\sum_{t=1}^{n} q_n'\big(\frac{\xi_t}{\sqrt{n}}\big)\big(\frac{\xi_t}{\sqrt{n}}\big)'q_n^{\perp}G_{th} \\ &= O_P\big(\frac{1}{n^2h^{3/2}}\sum_{t=1}^{n} G_{th}\big) = O_P\big(\frac{1}{nh^{1/2}}\big) = o_P(1), \end{aligned} \tag{A.12}$$

as $nh \to \infty$. Then, the proof of (A.10) has been completed for $k = 1$ and 4. On the other hand, it is easy to show that

$$\begin{aligned} &\Big\|\frac{1}{nh^{3/2}}q_n'\big(\frac{x_{\delta_n}}{\sqrt{n}}\big)\sum_{t=1}^{n}\big(\frac{\xi_t}{\sqrt{n}}\big)'q_n^{\perp}G_{th}\Big\| \\ \le\ & \frac{1}{(nh)^{1/2}}\cdot\Big\|q_n'\big(\frac{x_{\delta_n}}{\sqrt{n}}\big)\Big\|\cdot\Big(\frac{1}{nh}\sum_{t=1}^{n}\|\xi_t'q_n^{\perp}\|G_{th}\Big) \\ =\ & O_P\big((nh)^{-1/2}\big) = o_P(1), \end{aligned}$$

which indicates that

$$\begin{aligned}\Delta_n^*(2,1) &= \frac{1}{nh^{3/2}}q_n'\big(\frac{x_{\delta_n}}{\sqrt{n}}\big)\sum_{t=1}^{n}\big(\frac{\eta_t}{\sqrt{n}}\big)'q_n^{\perp}G_{th} + o_P(1) \\ &= \sqrt{2}\left[b_{nv}'b_{nv}\right]^{1/2}\Big[\frac{1}{nh}\sum_{t=1}^{n}\big(\frac{\eta_t}{\sqrt{2nh}}\big)'G_{th}\Big]q_{nv}^{\perp} + o_P(1) \\ &\equiv \Delta_n^{\diamond}(2) + o_P(1), \end{aligned} \tag{A.13}$$

where $q_{nv}$ and $q_{nv}^{\perp}$ are defined as $q_n$ and $q_n^{\perp}$ in Section 3 but with $b_n$ replaced by $b_{nv}$. By (A.7)–(A.10) and (A.13), we have

$$\Delta_n(2) = \Delta_n^{\diamond}(2) + o_P(1). \tag{A.14}$$

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 $\Delta_n^*(2)$ as above, we can show that

$$\Delta_n^*(3) = \frac{1}{(nh)^2}\sum_{t=1}^{n}(q_n^{\perp})'\eta_t\eta_t'q_n^{\perp}G_{th} + o_P(1) = (q_{nv}^{\perp})'\Big[\frac{2}{nh}\sum_{t=1}^{n}\big(\frac{\eta_t}{\sqrt{2nh}}\big)\big(\frac{\eta_t}{\sqrt{2nh}}\big)'G_{th}\Big]q_{nv}^{\perp} + o_P(1),$$

which leads to

$$\Delta_n(3) = (q_{nv}^{\perp})'\Big[\frac{2}{nh}\sum_{t=1}^{n}\big(\frac{\eta_t}{\sqrt{2nh}}\big)\big(\frac{\eta_t}{\sqrt{2nh}}\big)'G_{th}\Big]q_{nv}^{\perp} + o_P(1) \equiv \Delta_n^{\diamond}(3) + o_P(1). \tag{A.15}$$

By (A.5), (A.14) and (A.15), we have

$$D_n^+ Q_n' \Big( \sum_{t=1}^{n} x_t x_t' G_{th} \Big) Q_n D_n^+ = \begin{bmatrix} \Delta_n(1) & \Delta_n(2) \\ \Delta_n(2)' & \Delta_n(3) \end{bmatrix} = \begin{bmatrix} \Delta_n^\diamond(1) & \Delta_n^\diamond(2) \\ \Delta_n^\diamond(2)' & \Delta_n^\diamond(3) \end{bmatrix} + o_P(1). \tag{A.16}$$

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

$$\Big( \frac{1}{\sqrt{n}} \sum_{s=1}^{\delta_n} \overline{v}_s, \ \frac{1}{\sqrt{2\lfloor nh \rfloor}} \sum_{s=\delta_n+1}^{\delta_n(r)} \overline{v}_s \Big) \Rightarrow \Big[ B_{d,\delta_0}(\Omega_v), \ B_{d,r,*}(\Omega_v) \Big] \tag{A.17}$$

for $\delta_n(r) = \delta_n + \lfloor 2rnh \rfloor + 1$ with $0 < r \leq 1$. By using (A.16), (A.17) and the continuous mapping theorem (Billingsley, 1968), we can complete the proof of (A.1). □

Next consider the derivation of the limit behavior of

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

We derive asymptotic distribution for $\Gamma_n$ through the following Propositions A.2 and A.3.

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

$$\begin{aligned} \Gamma_n &= \begin{bmatrix} (2b_n' b_n)^{1/2} \cdot \frac{1}{\sqrt{2nh}} \sum_{t=1}^{n} u_t G_{th} \\ (2q_n^\perp)' \cdot \frac{1}{2nh} \sum_{t=1}^{n} (x_t - x_{\delta_n}) u_t G_{th} \end{bmatrix} + o_P(1) \\ &\equiv \begin{bmatrix} (2b_n' b_n)^{1/2} \cdot \Gamma_n^\diamond(1) \\ (2q_n^\perp)' \cdot \Gamma_n^\diamond(2) \end{bmatrix} + o_P(1), \end{aligned} \tag{A.18}$$

*where $\Gamma_n^\diamond(1)$ and $\Gamma_n^\diamond(2)$ are asymptotically independent of $b_n$.*

PROOF. Observe that $\Gamma_n = [\Gamma_n(1), \Gamma_n(2)']'$, where

$$\begin{aligned} \Gamma_n(1) &= q_n' \cdot \frac{1}{nh^{1/2}} \sum_{t=1}^{n} x_t u_t G_{th}, \\ \Gamma_n(2) &= (q_n^\perp)' \cdot \frac{1}{nh} \sum_{t=1}^{n} x_t u_t G_{th}. \end{aligned}$$

For $\Gamma_n(1)$, note that

$$\begin{aligned} \Gamma_n(1) &= \sqrt{2} q_n' \cdot \frac{x_{\delta_n}}{\sqrt{n}} \cdot \frac{1}{\sqrt{2nh}} \sum_{t=1}^{n} u_t G_{th} + q_n' \cdot \frac{1}{nh^{1/2}} \sum_{t=1}^{n} \big( x_t - x_{\delta_n} \big) u_t G_{th} \\ &= (2b_n' b_n)^{1/2} \cdot \Gamma_n^\diamond(1) + o_P(1), \end{aligned} \tag{A.19}$$

as

$$\Big|q_n' \cdot \frac{1}{nh^{1/2}} \sum_{t=1}^{n} \big(x_t - x_{\delta_n}\big) u_t G_{th}\Big| \le \frac{1}{nh^{1/2}} \cdot \|q_n\| \cdot \Big\| \sum_{t=1}^{n} \big(x_t - x_{\delta_n}\big) u_t G_{th}\Big\|$$
$$= \frac{1}{nh^{1/2}} \cdot O_P(1) \cdot O_P(nh) = O_P(\sqrt{h}) = o_P(1) \tag{A.20}$$

using Assumptions 1 and 3.

We next show that $\Gamma_n^{\diamond}(1)$ is asymptotically independent of $b_n$. 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 $\rho(n) = (nh)^{\frac{1}{3}+\epsilon_*}$, $0 < \epsilon_* < \frac{\gamma_0}{3(6+2\gamma_0)}$. Note that

$$\begin{aligned}(2b_n'b_n)^{1/2} \cdot \Gamma_n^{\diamond}(1) &= (2b_n'b_n)^{1/2} \cdot \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} u_{t\blacktriangle} G_{th} + (2b_n'b_n)^{1/2} \cdot \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} u_{t\blacktriangledown} G_{th} \\ &\equiv (2b_n'b_n)^{1/2} \cdot \Gamma_n^{\diamond}(1,1) + (2b_n'b_n)^{1/2} \cdot \Gamma_n^{\diamond}(1,2). \end{aligned} \tag{A.21}$$

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

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

by Assumption 1. Hence, we have

$$\Big\| \sum_{t=1}^{n} u_{t\blacktriangledown} G_{th}\Big\| = o_P\big(nh \cdot \rho^{-3/2}(n)\big) = o_P\big(\sqrt{nh}\big), \tag{A.22}$$

which indicates that $\Gamma_n^{\diamond}(1,2)$ is asymptotically dominated by $\Gamma_n^{\diamond}(1,1)$. On the other hand, let $\varsigma_n = \rho(n)/n$, $\widetilde{\delta}_n = \lfloor(\delta_0 - h - \varsigma_n)n\rfloor$ and $\widetilde{b}_n = \frac{1}{\sqrt{n}} x_{\widetilde{\delta}_n}$. We may show that $b_n = \widetilde{b}_n + o_P(1)$. It is obvious that $\Gamma_n^{\diamond}(1,1)$ is independent of $\widetilde{b}_n$. Thus, we have proved that $\Gamma_n^{\diamond}(1)$ is asymptotically independent of $b_n$.

For $\Gamma_n(2)$, by the definition of $q_n^{\perp}$ and following the argument in the proof of Proposition A.1, we can show that

$$\Gamma_n(2) = (2q_n^{\perp})' \cdot \frac{1}{2nh} \sum_{t=1}^{n} (x_t - x_{\delta_n}) u_t G_{th} + o_P(1) \equiv (2q_n^{\perp})' \Gamma_n^{\diamond}(2) + o_P(1). \tag{A.23}$$

Applying the truncation technique used in the above argument, we can similarly prove that $\Gamma_n^{\diamond}(2)$ is asymptotically independent of $b_n$.

The proof of Proposition A.2 has been completed. □

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

$$\Gamma_n^{\diamond} \equiv \begin{bmatrix} \Gamma_n^{\diamond}(1) \\ \Gamma_n^{\diamond}(2) \end{bmatrix} \Rightarrow \begin{bmatrix} \int_{-1}^{1} K(r) dB^*_{(r+1)/2}(\Omega_u) \\ \int_{-1}^{1} K(r) B_{d,\frac{r+1}{2},*}(\Omega_v) dB_{\frac{r+1}{2},*}(\Omega_u) + \frac{1}{2}\Delta_{vu} \end{bmatrix}. \tag{A.24}$$

PROOF. Observe that

$$\Gamma_n^{\diamond}(1) = \sum_{t=1}^{n} \frac{u_t}{\sqrt{2nh}} K\left(\frac{t-n\delta_0}{nh}\right) = \sum_{t=\lfloor \delta_0 n\rfloor - \lfloor nh\rfloor +1}^{\lfloor \delta_0 n\rfloor + \lfloor nh\rfloor} \frac{u_t}{\sqrt{2nh}} K\left(\frac{t-\delta_0 n}{nh}\right)$$

and

$$\Gamma_n^{\diamond}(2) = \sum_{t=1}^{n} \frac{x_t - x_{\delta_n}}{\sqrt{2nh}} \frac{u_t}{\sqrt{2nh}} K\left(\frac{t-n\delta_0}{nh}\right) = \sum_{t=\lfloor \delta_0 n\rfloor - \lfloor nh\rfloor +1}^{\lfloor \delta_0 n\rfloor + \lfloor nh\rfloor} \frac{x_t - x_{\delta_n}}{\sqrt{2nh}} \frac{u_t}{\sqrt{2nh}} K\left(\frac{t-\delta_0 n}{nh}\right).$$

Then, using the weak convergence result of

$$\frac{1}{\sqrt{2nh}}\Big( \sum_{t=\delta_n+1}^{\delta_n(p)} v_t', \sum_{t=\delta_n+1}^{\delta_n(p)} u_t \Big) \Rightarrow \Big( B_{d,p,*}(\Omega_v)',\ B_{p,*}(\Omega_u) \Big), \quad 0 \le p \le 1$$

with $\delta_n(p) = \delta_n + \lfloor 2nhp \rfloor + 1$, Lemma B.1 in Appendix B and the continuous mapping theorem, we may complete the proof of (A.24). □

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

PROOF OF THEOREM 3.1. Observe that

$$\begin{aligned} \widehat{f}_n(\delta_0) - f(\delta_0) \quad = \quad & \Big[ \sum_{t=1}^{n} x_t x_t' G_{th}(\delta_0) \Big]^{+} \Big\{ \sum_{t=1}^{n} x_t x_t' \big[ f\big(t/n\big) - f(\delta_0) \big] G_{th}(\delta_0) \Big\} + \\ & \Big[ \sum_{t=1}^{n} x_t x_t' G_{th}(\delta_0) \Big]^{+} \Big[ \sum_{t=1}^{n} x_t u_t G_{th}(\delta_0) \Big]. \end{aligned} \tag{A.25}$$

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

$$f\big(t/n\big) - f(\delta_0) = O(h^{\gamma_1}) \tag{A.26}$$

when $\big|\frac{t}{n} - \delta_0\big| \le h$. By (A.26) and following the proof of Proposition A.1, we can easily prove that

$$\Big[ \sum_{t=1}^{n} x_t x_t' G_{th}(\delta_0) \Big]^{+} \Big\{ \sum_{t=1}^{n} x_t x_t' \big[ f\big(\frac{t}{n}\big) - f(\delta_0) \big] G_{th}(\delta_0) \Big\} = O_P(h^{\gamma_1}). \tag{A.27}$$

Then, using Propositions A.1–A.3, (A.27) 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 $D_{n*} = I_2 \otimes D_n$, $Q_{n*} = I_2 \otimes Q_n$,

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

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

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

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

$$\big(\Delta_{n*}, \Gamma_{n*}\big) \Rightarrow \big(\Delta_{\delta_0*}, \Gamma_{\delta_0*}\big), \tag{A.28}$$

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

$$\begin{aligned} D_n Q_n' \big[\widetilde{f}_n(\delta_0) - f(\delta_0)\big] &= e_d D_{n*} Q_{n*}' \begin{bmatrix} \widetilde{f}_n(\delta_0) - f(\delta_0) \\ h\widetilde{f}_n'(\delta_0) - hf(\delta_0) \end{bmatrix} \\ &= e_d \Delta_{n*}^{+} \Gamma_{n*} + O_P(h^2 D_n Q_n). \end{aligned} \tag{A.29}$$

Equations (A.28) and (A.29) 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

$$\hat{y}_t^{\#} = y_t - \widehat{\Omega}_{uv}\widehat{\Omega}_{vv}^{-1}\Delta x_t = y_t - \Omega_{uv}\Omega_{vv}^{-1} v_t + \Big(\Omega_{uv}\Omega_{vv}^{-1} - \widehat{\Omega}_{uv}\widehat{\Omega}_{vv}^{-1}\Big) v_t$$

and let

$$y_t^{\#} = y_t - \Omega_{uv}\Omega_{vv}^{-1} v_t = f_t' x_t + u_t - \Omega_{uv}\Omega_{vv}^{-1} v_t \equiv f_t' x_t + u_t^{\#}.$$

Following the proof of Theorem 3.1 and using the fact that

$$\Big(\widehat{\Delta}_{uu}, \widehat{\Delta}_{vu}, \widehat{\Delta}_{vv}, \widehat{\Omega}_{uv}, \widehat{\Omega}_{vv}\Big) = (\Delta_{uu}, \Delta_{vu}, \Delta_{vv}, \Omega_{uv}, \Omega_{vv}) + o_P(1),$$

we may show that

$$D_n Q_n' \Big[\sum_{t=1}^{n} x_t x_t' K_{th}(\delta_0)\Big]^{+} \sum_{t=1}^{n} x_t \Big(\Omega_{uv}\Omega_{vv}^{-1} - \widehat{\Omega}_{uv}\widehat{\Omega}_{vv}^{-1}\Big) v_t K_{th}(\delta_0) = o_P(1).$$

Now we have

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

Note that $n^{-1/2}\sum_{s=1}^{\lfloor nr\rfloor} u_s^{\#} \Rightarrow B_r(\Omega_{u.v})$ which is independent of the Brownian motion $B_{d,r}(\Omega_v)$. By virtue of the definition of $\widehat{\Gamma}_{n,bc}$ in (4.2) and proceeding as in Theorem 3.1, Proposition A.1, and the FM regression arguments of Phillips and Hansen (1990), we can show that the two components of the second-order bias of $\widehat{f}_n(\delta_0)$ in the direction $q_n^{\perp}$ are eliminated, and the limit theory given in (4.3) follows. □

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

$$\begin{aligned}
\widehat{\Delta}_{vu}(j) &= \frac{1}{\tau_n^* - \tau_n}\sum_{t=\tau_n+1}^{\tau_n^*} v_{t-j}\widehat{u}_t \\
&= \frac{1}{\tau_n^* - \tau_n}\sum_{t=\tau_n+1}^{\tau_n^*} v_{t-j}u_t - \frac{1}{\tau_n^* - \tau_n}\sum_{t=\tau_n+1}^{\tau_n^*} v_{t-j}x_t'\big(\widehat{f}_{nt} - f_t\big) \\
&\equiv \overline{\Delta}_{vu}(j) - \widetilde{\Delta}_{vu}(j), \qquad \text{(A.30)}
\end{aligned}$$

for $j = 1, \cdots, l_n$, where $\tau_n = \lfloor \tau_* n\rfloor$ and $\tau_n^* = \lfloor (1-\tau_*)n\rfloor$. Using (A.30), we have

$$\widehat{\Delta}_{vu} = \sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big)\widehat{\Delta}_{vu}(j) = \sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big)\overline{\Delta}_{vu}(j) - \sum_{j=0}^{l_n} k\big(\frac{j}{l_n}\big)\widetilde{\Delta}_{vu}(j). \qquad \text{(A.31)}$$

We first prove the second term on the right hand side of (A.31) 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 $t = \tau_n + 1, \cdots, \tau_n^*$. Letting $Q_{nt} = \big[q_{nt}, q_{nt}^{\perp}\big]$ 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 $Q_{nt}'Q_{nt} = I_d$, by transforming coordinates, we have

$$\begin{aligned}
\sum_{s=1}^{n} x_s x_s' G_{sh}(t/n) &= \Big[Q_{nt}D_n\Big]\Big[D_n^{+}Q_{nt}'\sum_{s=1}^{n} x_s x_s' G_{sh}(t/n) Q_{nt} D_n^{+}\Big]\Big[D_n Q_{nt}'\Big] \\
&\equiv \Theta_{nt}^{\diamond}(1)\Theta_{nt}^{\diamond}(2)\Theta_{nt}^{\diamond}(1)'.
\end{aligned}$$

Furthermore, by Lemma B.3 in Appendix B of the online supplement and the definitions of $Q_{nt}$ and $D_n$, we may show that $\Theta^{\diamond}_{nt}(1)$ and $\Theta^{\diamond}_{nt}(2)$ are invertible with probability 1, and thus the generalized inverse becomes the conventional inverse. Then we have

$$\begin{aligned}\Theta_{nt}(u) &= Q_{nt}D_n^{+}\left[D_n^{+}Q_{nt}'\sum_{s=1}^{n}x_s x_s' G_{sh}(t/n)Q_{nt}D_n^{+}\right]^{+}\left[D_n^{+}Q_{nt}'\sum_{s=1}^{n}x_s u_s G_{sh}(t/n)\right] \\ &\equiv Q_{nt}D_n^{+}\Theta^{+}_{nt,1}\Theta_{nt,2}(u). \end{aligned} \tag{A.32}$$

Note that $x_t = x_{\lfloor t-nh\rfloor} + x_t - x_{\lfloor t-nh\rfloor}$ and $x'_{\lfloor t-nh\rfloor}q^{\perp}_{nt} = 0$ with probability 1. Then, using Lemmas B.2 and B.3 in Appendix B and by Taylor expansion of $f(\cdot)$, we can prove that

$$\begin{aligned}\sum_{j=0}^{l_n}k\big(\frac{j}{l_n}\big)\widetilde{\Delta}_{vu}(j) &= \frac{1}{\tau_n^{*}-\tau_n}\sum_{j=0}^{l_n}k\big(\frac{j}{l_n}\big)\sum_{t=\tau_n+1}^{\tau_n^{*}}v_{t-j}x_t'\big(\widehat{f}_{nt}-f_t\big) \\ &= \frac{1}{\tau_n^{*}-\tau_n}\sum_{j=0}^{l_n}k\big(\frac{j}{l_n}\big)\sum_{t=\tau_n+1}^{\tau_n^{*}}v_{t-j}x_t'\big[\Theta_{nt}(u)+\Theta_{nt}(f)\big] \\ &= o_P(1)+O_P(\sqrt{n}hl_n) = o_P(1) \end{aligned} \tag{A.33}$$

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

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

$$\sum_{j=0}^{l_n}k\big(\frac{j}{l_n}\big)\overline{\Delta}_{vu}(j) = \sum_{j=0}^{l_n}k\big(\frac{j}{l_n}\big)\Big(\frac{1}{\tau_n^{*}-\tau_n}\sum_{t=\tau_n+1}^{\tau_n^{*}}v_{t-j}u_t\Big) = \Delta_{vu}+o_P(1). \tag{A.34}$$

Using (A.33) and (A.34), we can complete the proof of Proposition 4.2. □

# 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.

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.

Fan, J., and I. Gijbels, 1996. Local Polynomial Modelling and Its Applications. Chapman and Hall, London.

Gao, J., and M. L. King, 2011. A new test in parametric linear models against nonparametric autoregressive errors. Working paper at http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011/wp20-11.pdf.

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

Gao, J., and P. C. B. Phillips, 2013b. Trending time series with non– and semi–parametric cointegration. Working paper at http://ideas.repec.org/p/msh/ebswps/2013-16.html.

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.

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., 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. Forthcoming in Econometric Reviews.

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, D., P. C. B. Phillips, and J. Gao, 2016. Uniform consistency of nonstationary kernel-weighted sample covariances for nonparametric regression. Econometric Theory, 32, 655–685.

Li, K., D. Li, Z. Liang, and C. Hsiao, 2016. Estimation of semi–varying coefficient models with nonstationary regressors. Forthcoming in Econometric Reviews.

Liang, H., P. C. B. Phillips, H. Wang, and Q. Wang. 2016. Weak convergence to stochastic integrals for econometric applications. Econometric Theory, 32, 359-401.

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.

Park, J. Y., and P. C. B. Phillips, 1989. Statistical inference in regression with integrated processes: part 2. Econometric Theory, 5, 95–131.

Park, J. Y., and P. C. B. Phillips, 1989. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory, 15, 269–98.

Park, J. Y., and P. C. B. Phillips, 2001. Nonlinear regressions with integrated time series. Econometrica, 69, 117-161.

Phillips, P. C. B., 1988. Multiple regression with integrated processes, in *Statistical Inference from Stochastic Processes* ed. by N. U. Prabhu. Contemporary Mathematics 80, 79–106.

Phillips, P. C. B., 1989. Partially identified econometric models. Econometric Theory, 5, 181-240.

Phillips, P. C. B., 1991. Optimal inference in cointegrated systems. Econometrica, 59, 283-306.

Phillips, P. C. B., 1995. Fully modified least squares and vector autoregression. Econometrica, 63, 1023-1078.

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

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., D. Li, and J. Gao, 2016. Online supplement to: 'Estimating smooth structural change in cointegration models'. Submitted working document.

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

Phillips, P. C. B., and T. Magdalinos, 2013. Inconsistent VAR regression with common explosive roots. Econometric Theory, 29, 808-837.

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, London.

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

Wang, Q., and P. C. B. Phillips, 2009b. Structural nonparametric cointegrating regression. Econometrica, 77, 1901–1948.

Wang, Q., and P. C. B. Phillips, 2011. Asymptotic theory for zero energy density estimation with nonparametric regression applications. Econometric Theory, 27, 236-269.

Wang, Q. and P. C. B. Phillips, 2015. Nonparametric cointegrating regression with endogeneity and long memory. Forthcoming in Econometric Theory.

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.

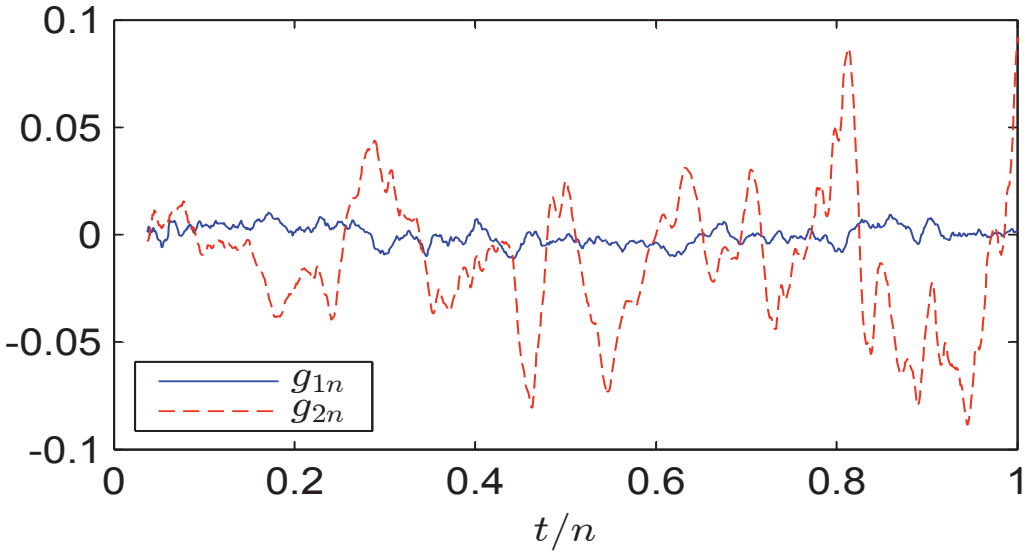



Fig. 5.1: Plots of $g_{1n}(\delta)$ and $g_{2n}(\delta)$ versus $\delta$ for the functional form *M1*

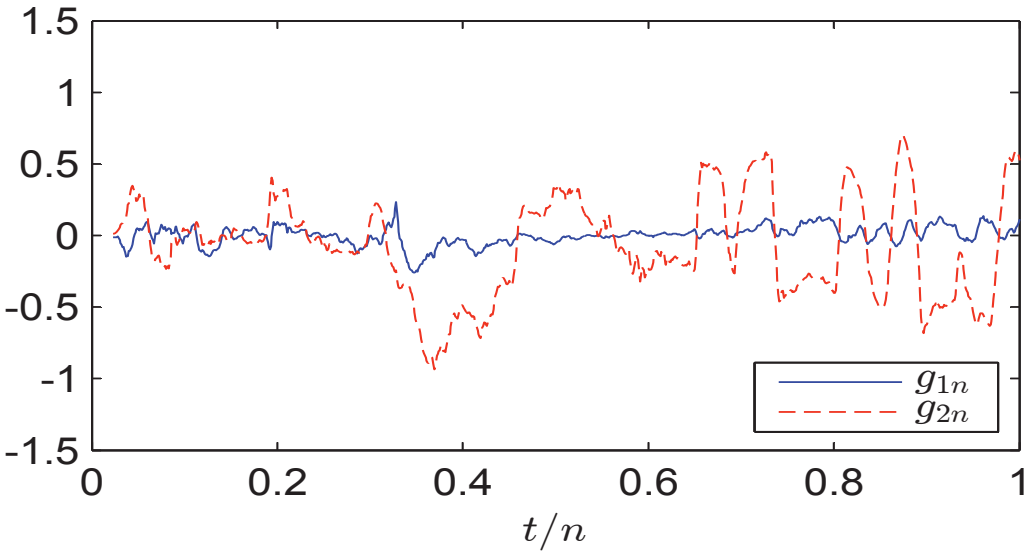



Fig. 5.2: Plots of $g_{1n}(\delta)$ and $g_{2n}(\delta)$ versus $\delta$ for the functional form *M2*

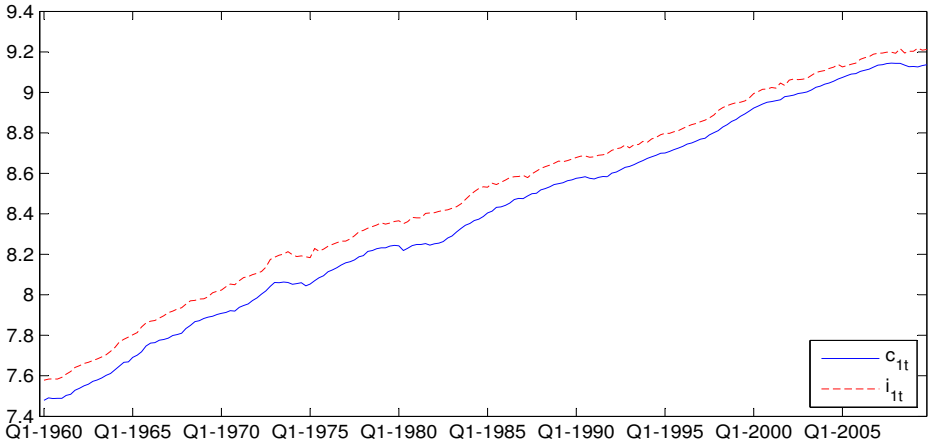
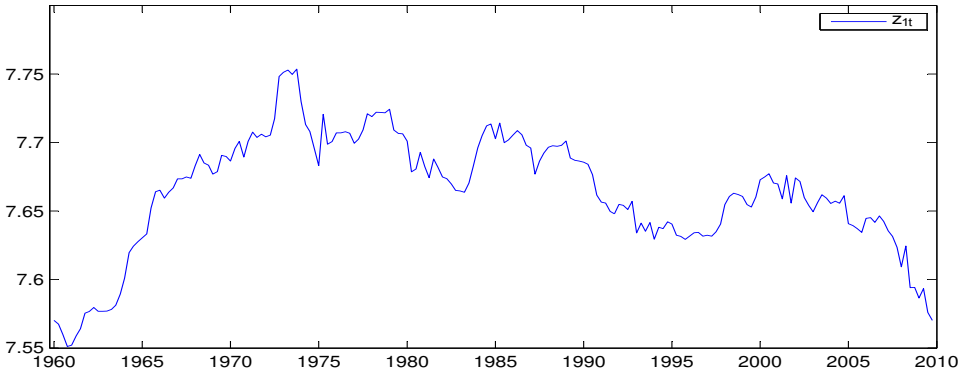


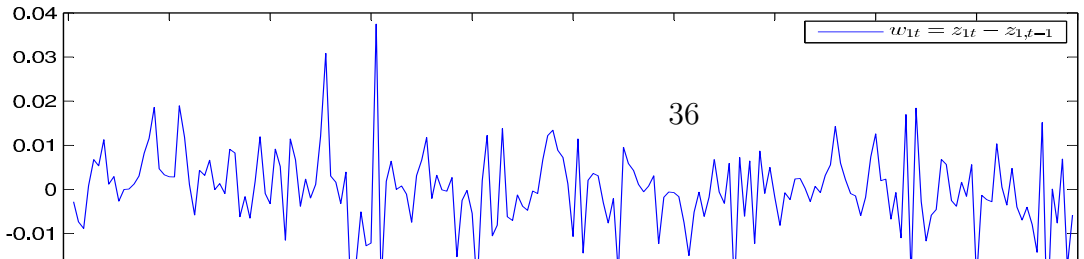

Fig. 5.3(a): Real consumption and real disposable income for 1960:1 – 2009:3




(A)

(B)

Fig. 5.3(b): (A) The detrended series $z_{1t}$; and (B) its differenced version for 1960:1 – 2009:3



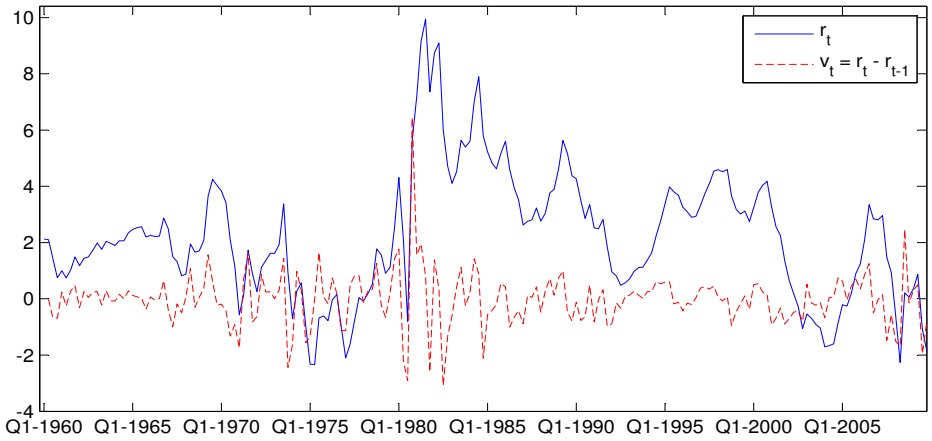



Fig. 5.3(c): The real interest rate for 1960:1 – 2009:3 and the differenced version

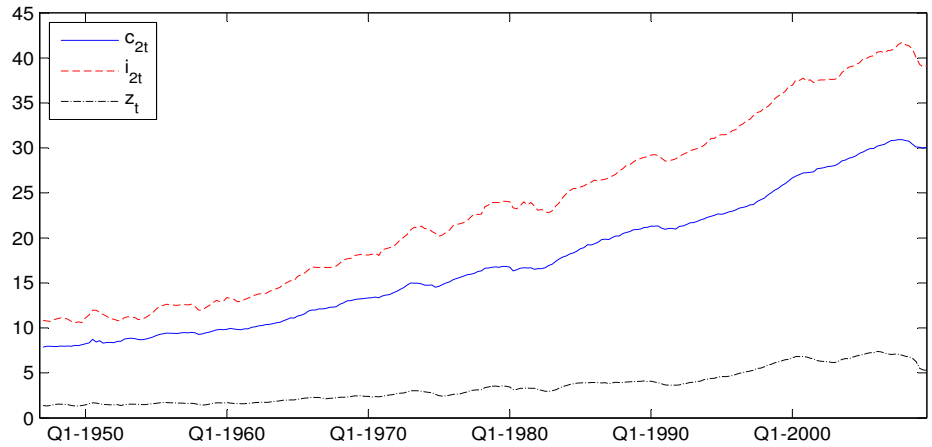



Fig. 5.4(a): Real consumption, disposable income and investment for 1947:1 – 2009:4

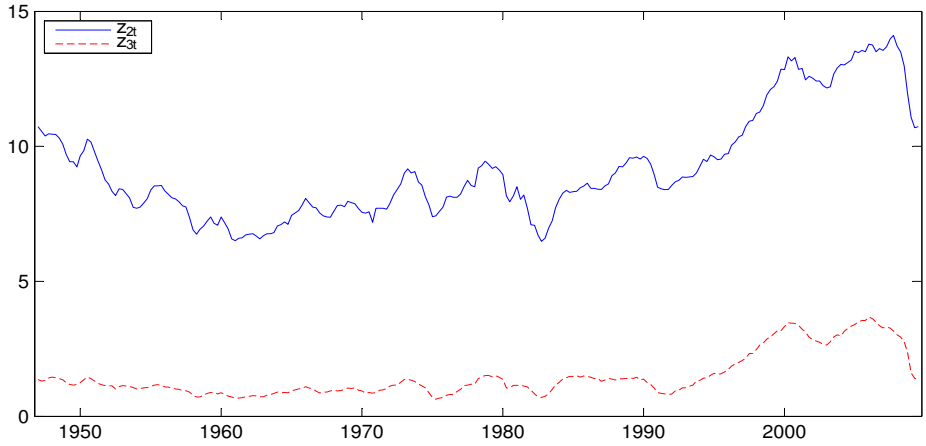



Fig. 5.4(b): The detrended series $z_{2t}$ and $z_{3t}$ for 1947:1 – 2009:4

Fig. 5.4(c): The differenced versions of $z_{2t}$ and $z_{3t}$ for 1947:1 – 2009:4