KERNEL-BASED INFERENCE IN TIME-VARYING COEFFICIENT COINTEGRATING REGRESSION

By

Degui Li, Peter C. B. Phillips, and Jiti Gao

September 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 2109

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

Kernel-Based Inference in Time-Varying Coefficient Cointegrating Regression*

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

August 22, 2017

Abstract

This paper studies nonlinear cointegrating models with time-varying coefficients and multiple nonstationary regressors using classic kernel smoothing methods to estimate the coefficient functions. Extending earlier work on nonstationary kernel regression to take account of practical features of the data, we allow the regressors to be cointegrated and to embody a mixture of stochastic and deterministic trends, complications which result in asymptotic degeneracy of the kernel-weighted signal matrix. To address these complications new *local* and *global rotation* techniques are introduced to transform the covariate space to accommodate multiple scenarios of induced degeneracy. Under certain regularity conditions we derive asymptotic results that differ substantially from existing kernel regression asymptotics, leading to new limit theory under multiple convergence rates. For the practically important case of endogenous nonstationary regressors we propose a fully-modified kernel estimator whose limit distribution theory corresponds to the prototypical pure (i.e., exogenous covariate) cointegration case, thereby facilitating inference using a generalized Wald-type test statistic. These results substantially generalize econometric estimation and testing techniques in the cointegration literature to accommodate time variation and complications of co-moving regressors. Finally an empirical illustration to aggregate US data on consumption, income, and interest rates is provided.

Keywords: Cointegration; FM-kernel estimation; Generalized Wald test; Global rotation; Kernel degeneracy; Local rotation; Super-consistency; Time-varying coefficients.

Abbreviated Title: Kernel Inference in Cointegrating Regression

---

*Phillips acknowledges the NSF support under Grant Number: SES 12-85258. Gao acknowledges the ARC Discovery Grants support under Grant Numbers: DP150101012 and DP170104421. Computing assistance by Weilun Zhou is also much appreciated.

[†]University of York

[‡]Yale University, University of Auckland, Southampton University, and Singapore Management University; corresponding author email: peter.phillips@yale.edu

[§]Monash University

1 Introduction

Many time series that are encountered in economics and finance are well known to exhibit nonstationary characteristics such as the random wandering behavior of financial asset prices and the secular growth components in aggregate time series data that indicate the presence of some form of deterministic drift. Following the work of Phillips and Durlauf (1986), Engle and Granger (1987), Phillips and Perron (1988); Park and Phillips (1989), Phillips (1988, 1991) and Johansen (1991), substantial investments have been made in econometric methodology to take account of these characteristics in linear and log linear cointegrating regression estimation and inference.

Notwithstanding this body of work many practical implementations reveal that parametric linear cointegration models are often rejected by the data even when there is evident co-movement among the trending series. Acknowledgement of this weakness has led to the recent development of econometric methodology for treating various nonlinear and nonparametric cointegrating models (Park and Phillips, 2001; Karlsen, Myklebust and Tjøstheim, 2007; Cai, Li and Park, 2009; Wang and Phillips, 2009a,b; Xiao, 2009; Gao and Phillips, 2013; Li *et al*, 2017; Phillips, Li and Gao, 2017). For the important case of multivariate integrated covariates, much of this nonparametric research on nonlinear cointegration excludes possible co-movement among the regressors and the presence of deterministic drift. Such restrictions simplify asymptotic theory but limit applicability of the methods to time series without the commonly occuring characteristics that produce co-movement over time and asymptotic degeneracies in the signal matrix.

The primary goal of the present paper is to relax these restrictions by allowing more flexible structures among the covariates, to develop kernel regression asymptotics for a general class of models that accommodate these key features in the data, and to provide inferential machinery that enables convenient estimation and inference in practical work. In developing these methods, our main focus of attention is a multiple regression model with time-varying coefficients of the following form

$$Y_t = \boldsymbol{\beta}_t' X_t + e_{t0}, \quad t = 1, \cdots, T, \tag{1.1}$$

where $\boldsymbol{\beta}_t := \boldsymbol{\beta}(t/T)$ is a d-dimensional vector of coefficients which varies over time, $\boldsymbol{\beta}(\cdot)$ is a d-dimensional vector of functions, $\{X_t\}$ is a d-dimensional nonstationary process, and $\{e_{t0}\}$ is a stationary random error process. The paper studies three generating structures on X_t of increasing complexity: (i) X_t is cointegrated with d_0 cointegrating vectors and no deterministic trend where $0 \le d_0 \le d-1$; (ii) X_t involves a mixture of deterministic and stochastic trends but without any cointegrating structure; and (iii) X_t is cointegrated and has deterministic trend components. Scenario

(iii) is the most general and combines the complications of (i) and (ii). In view of the special technical difficulties involved, it is convenient to treat these generating structures for the regressors X_t individually at first, leading ultimately to a complete set of asymptotics for coefficient function estimation and inference in models of the form (1.1) with time-varying functional coefficients and co-moving endogenous regressors.

Model (1.1) is motivated by the need for a flexible framework that captures structural change via temporal evolution in the functional coefficients in regressions with nonstationary data. The formulation usefully circumvents curse of dimensionality problems that commonly arise in nonparametric regression estimation when the dimension of the covariates is large and that are known to be exacerbated in the nonstationary nonparametric case due to slower convergence rates (Wang and Phillips, 2009a). The modelling framework (1.1) includes and extends many linear and nonlinear cointegration models that have been extensively studied in the literature. For instance, in the constant coefficient case where $\boldsymbol{\beta}_t = \boldsymbol{\beta}$, model (1.1) is a multiple linear regression with integrated regressors in which scenarios (i)-(iii) above may be present in practical work and for which asymptotic linear regression theory was developed in early work by Phillips and Perron (1988); Park and Phillips (1989), Phillips (1988, 1995), and Toda and Phillips (1993). When X_t is not cointegrated and no deterministic drift is involved, model (1.1) reduces to the model studied in Park and Hahn (1999) and Phillips, Li and Gao (2017) where sieve estimation and kernel-based estimation techniques were analyzed, respectively. When the nonstationary regressors X_t are cointegrated, there exist certain linear combinations of X_t (if $d_0 \geq 1$) which can lower the order of integration, leading to the presence of a stationary process component in the regressors. It follows that our modelling framework also relates to work on time-varying coefficient models with stationary (or locally stationary) regressors (c.f., Robinson, 1989; Cai, 2007; Zhou and Wu, 2010; Chen and Hong, 2012; Vogt, 2012; Zhang and Wu, 2012; Giraitis, Kapetanious and Yates, 2014), and may be regarded as an extension of that work to accommodate nonstationary and trending regressor components. The upshot is that the results obtained in the present paper have wide potential applicability to economic time series with stationary, trend stationary, co-moving, and stochastically nonstationary components.

The paper applies standard Nadaraya-Watson kernel method to estimate the coefficient function $\boldsymbol{\beta}(\cdot)$ in the presence of a complicating structure of co-moving and co-trending regressors that raises significant challenges in the development of a limit theory for kernel estimation and inference. The technical challenges may be explained in a heuristic manner as follows. The central difficulty arises from the multiple asymptotic singularities that feature in the kernel-weighted signal matrix – the random matrix that carries the kernel weights and appears in the denominator of the usual kernel estimator. Rotation techniques are used to conform the covariate space to accommodate signals of various orders in developing the asymptotic theory. These techniques extend those that were developed and are now commonly used in the nonstationary linear regression literature (Phillips, 1988; Phillips and Perron, 1988; Park and Phillips, 1989) to the kernel regression environment where both global and local rotators are required. If the regressors are cointegrated, a global rotation technique is applied to separate out the stationary components and nonstationary components, which carry the associated signals in kernel estimation with differing strengths. In the multivariate regressor case, the kernel-weighted random matrix associated with the nonstationary covariate components may have dimension greater than unity, inducing a further signal degeneracy that we refer to as local degeneracy throughout the paper.

When the nonstationary components have only stochastic trends, time-varying coefficient kernel regression naturally concentrates attention on a particular time coordinate of the partial sum process and, in doing so, the associated stochastic process limit process. This focus on a local time coordinate produces a limiting kernel signal matrix of deficient rank unity. On the other hand, when the stochastic trends are themselves asymptotically majorized by deterministic linear trends, the nonstationary components become dominated asymptotically by these linear trends, which reduces asymptotic variability across component variables and leads to further degeneracy in the asymptotics. The local rotation approach used in the present paper addresses this further degeneracy in the nonstationary components and applies whether these components are dominated by stochastic or deterministic trends.

This rotation geometry enables the development of a full asymptotic distribution theory for nonstationary kernel estimation under general regularity conditions. The main results reveal multiple convergence rates in the different directions associated with the rotations. These directions include the usual stationary regressor nonparametric convergence rate ($\sqrt{Th}$), a type 1 super-consistency rate ($T\sqrt{h}$), a type 2 super-consistency rate (Th), and a type 3 rate ($T\sqrt{Th}$) in the direction of the deterministic linear trends. When nonstationary regressors in time-varying coefficient models are endogenous, second-order bias terms are present in the kernel regression limit theory, analogous to the endogeneity bias that occurs in linear cointegrating regression. Although such bias does not affect convergence rates in the asymptotics, it does influence finite sample performance and inference with kernel methods. To address this endogeneity bias, a fully-modified (FM) kernel estimator is developed for which the asymptotic theory corresponds to the pure (exogenous regressor) cointegration case.

A further contribution of the paper is to develop inferential methods for the time-varying coefficient functions in model (1.1). Two different null hypotheses on the coefficient functions are considered, which allow for universal restrictions (that is, restrictions that apply uniformly over time) and local restrictions (that is, restrictions that apply pointwise at some specific time-point corresponding to some sample fraction of the data). Generalized Wald-type test statistics are constructed to test these hypotheses. The limit theory for these tests is developed under both full rank and deficient rank conditions on the covariance structure of the restricted function coefficients, thereby accommodating potential implications for rank arising from signal matrix degeneracies in kernel estimation. The resulting asymptotics involve two types of chi-square limit distributions and possibly divergent degrees of freedom. This limit theory substantially extends existing work on inference in linear cointegrating regressions (particularly, Phillips and Perron, 1988; Park and Phillips, 1989; Toda and Phillips, 1993; Phillips, 1995) to the nonlinear cointegrating model setting.

These contributions combine to bring the limit theory for functional nonparametric nonstationary regression to a similar level of generality as the earlier limit theory for linear cointegating regression, allowing for multiple forms of asymptotic degeneracies in the regressor space and delivering asymptotically chi-square tests that enable inference in nonlinear co-moving systems with multiple covariates under endogeneity. The methods of the paper therefore apply widely and provide a convenient framework for investigators to test hypotheses concerning time evolution and stability in regression coefficients in nonstationary time series environments.

The rest of the paper is organized as follows. Section 2 describes the kernel estimation approach, provides assumptions, develops the double-rotation technique, and derives asymptotic theory when the regressors are cointegrated. Section 3 generalizes the structure and theory to the case where the regressors have a mix of stochastic and deterministic trends and the case when the regressors are cointegrated with deterministic trends. Section 4 introduces the FM kernel estimator and establishes its limit distribution theory. Section 5 explores methods of inference on the coefficient functions. Section 6 provides an empirical illustration to aggregate US data on consumption, income and interest rates. Section 7 concludes the paper. Proofs of the main results are given in Appendix 7. Proofs of some supplementary results and extensive simulation studies to evaluate the finite sample properties of the proposed methods in relation to the asymptotic theory are given in Appendices B and C, respectively, which are included in an online supplementary document.

2 Kernel estimation with cointegrated regressors

In this section, we use kernel smoothing to estimate the coefficient functions in model (1.1) when the nonstationary integrated regressors are cointegrated. We study the effects of the resulting asymptotic signal degeneracy, and introduce rotation techniques to derive the limit theory of the kernel estimates.

2.1 *Model estimation and assumptions*

Smoothness conditions on the coefficient function $\boldsymbol{\beta}(\cdot)$ permit local approximation $\boldsymbol{\beta}(z) \approx \boldsymbol{\beta}(z_0)$ for z in any small neighborhood of $z_0 \in (0,1)$, which motivates Nadaraya-Watson-type local level regression estimation of $\boldsymbol{\beta}(z_0)$ in (1.1) according to the formula

$$\widehat{\boldsymbol{\beta}}(z_0) = \Big[\sum_{t=1}^{T} X_t X_t' K\big(\frac{t-Tz_0}{Th}\big)\Big]^+ \Big[\sum_{t=1}^{T} X_t Y_t K\big(\frac{t-Tz_0}{Th}\big)\Big] =: \boldsymbol{\Lambda}_T^+(z_0)\boldsymbol{\Delta}_T(z_0), \tag{2.1}$$

where $\mathbf{A}^+$ denotes the Moore-Penrose generalised inverse of a matrix $\mathbf{A}$, $K(\cdot)$ is some kernel function, and h is a bandwidth which tends to zero as the sample size T tends to infinity. While the present paper concentrates on this particular kernel estimation method, other kernel methods such as local linear smoothing or local polynomial smoothing approaches (Fan and Gijbels, 1996) may be used in the same way and the methods given here may be suitably modified to accommodate these approaches with similar asymptotic results.

We commence our analysis with the case where the multivariate integrated regressors X_t are cointegrated with d_0 cointegrating vectors, $0 \le d_0 \le d-1$. Letting $d_1 = d - d_0$, there exists a $d \times d$ orthogonal matrix $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$ such that

$$\mathbf{H}_1' X_t = e_{t1}, \quad \Delta(\mathbf{H}_2' X_t) = \mathbf{H}_2'(\Delta X_t) = e_{t2}, \tag{2.2}$$

where the sizes for $\mathbf{H}_1$ and $\mathbf{H}_2$ are $d \times d_0$ and $d \times d_1$, respectively, Δ denotes the first-order difference operator, and $(e_{t1}', e_{t2}')'$ is stationary with e_{t1} being d_0-dimensional and e_{t2} being d_1-dimensional. In view of (2.2), a rotation of the regressor space conveniently separates out the stationary and nonstationary components of the covariates in model (1.1). The transformation matrix $\mathbf{H}$ is not unique and the rank of the cointegrating space d_0 together with the associated directions of cointegration that are embodied in the submatrix $\mathbf{H}_1$ are generally unknown a priori. We emphasize that knowledge of d_0 and $\mathbf{H}_1$ are not needed for application of (2.1) and the methods of the present paper, including the asymptotic results, can be used in practical work without such knowledge, although there are of course well-known parametric and nonparametric methods of testing to determine d_0 and procedures to estimate $\mathbf{H}_1$ in the existing literature (e.g., Johansen, 1991; Phillips, 1996; Cheng and Phillips, 2009). The following example illustrates the formulation (2.2) for a simple cointegrated vector autoregression (VAR) model with general stationary errors (c.f., Cheng and Phillips, 2009).

Example 1. Define the model

$$\Delta X_t = \boldsymbol{\alpha}\boldsymbol{\beta}' X_{t-1} + v_t, \tag{2.3}$$

where $\{v_t\}$ is a covariance stationary time series with mean zero, the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $d \times d_0$ with rank d_0, and $\mathbf{I}_{d_0} + \boldsymbol{\beta}'\boldsymbol{\alpha}$ has latent roots inside the unit circle. Let $\boldsymbol{\beta}_\perp$ be a $d \times (d - d_0)$ full rank matrix complement to $\boldsymbol{\beta}$ so that $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{O}_{d_0 \times d_1}$ where $d_1 = d - d_0$ and $\mathbf{O}_{r\times s}$ is an $r \times s$ null matrix, and define

$$\mathbf{H}_1' = (\boldsymbol{\beta}'\boldsymbol{\beta})^{-1/2}\boldsymbol{\beta}', \quad \mathbf{H}_2' = (\boldsymbol{\beta}_\perp'\boldsymbol{\beta}_\perp)^{-1/2}\boldsymbol{\beta}_\perp',$$

so that $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$ is orthogonal. Following standard arguments (Johansen, 1991; Phillips, 1995), the VAR model (2.3) falls into the framework of (2.2) with transforms

$$\mathbf{H}_1'X_t = (\boldsymbol{\beta}'\boldsymbol{\beta})^{-1/2}(\mathbf{I}_{d_0} + \boldsymbol{\beta}'\boldsymbol{\alpha})\boldsymbol{\beta}'X_{t-1} + \mathbf{H}_1'v_t =: e_{t1}$$

and

$$\Delta(\mathbf{H}_2'X_t) = (\boldsymbol{\beta}_\perp'\boldsymbol{\beta}_\perp)^{-1/2}\boldsymbol{\beta}_\perp'\boldsymbol{\alpha}\boldsymbol{\beta}'X_{t-1} + (\boldsymbol{\beta}_\perp'\boldsymbol{\beta}_\perp)^{-1/2}\boldsymbol{\beta}_\perp'v_t =: e_{t2}.$$

In order to establish limit theory for the kernel estimator in (2.1), we use the following regularity conditions.

Assumption 1. *Let $e_t = (e_{t0}, e_{t1}', e_{t2}')'$ satisfy*

$$e_t = \sum_{j=0}^{\infty}\boldsymbol{\Phi}_j\varepsilon_{t-j} = \sum_{j=0}^{\infty}\boldsymbol{\Phi}_jL^j\varepsilon_t =: \boldsymbol{\Phi}(L)\varepsilon_t,$$

where L is the lag operator, $\{\boldsymbol{\Phi}_j\}$ is a sequence of $(d+1)\times(d+1)$ matrices and $\{\varepsilon_t\}$ is a sequence of i.i.d. $(d+1)$-dimensional random vectors with mean zero, $\boldsymbol{\Omega}_\varepsilon := \mathbb{E}[\varepsilon_t\varepsilon_t']$ being positive definite and $\mathbb{E}[\|\varepsilon_t\|^{4+\delta_0}] < \infty$ for $\delta_0 > 0$, $\|\cdot\|$ denotes the Euclidean norm. In addition, the multivariate linear process coefficient matrices satisfy $\sum_{j=0}^{\infty} j\|\boldsymbol{\Phi}_j\| < \infty$ and the matrix $\boldsymbol{\Omega} := \boldsymbol{\Phi}\boldsymbol{\Omega}_\varepsilon\boldsymbol{\Phi}'$ is positive definite with $\boldsymbol{\Phi} := \sum_{j=0}^{\infty}\boldsymbol{\Phi}_j \neq \mathbf{O}_{(d+1)\times(d+1)}$.

Assumption 2. *The d-dimensional coefficient function $\boldsymbol{\beta}(\cdot)$ is continuous with $\|\boldsymbol{\beta}(z_0+z) - \boldsymbol{\beta}(z_0)\| = O(|z|^\gamma)$ as $z \to 0$ for some $\frac{1}{2} < \gamma \le 1$.*

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

(ii) The bandwidth h satisfies $h \to 0$ and $Th \to \infty$ as $T \to \infty$.

Assumption 1 uses a stationary vector linear process specification for $\{e_t\}$ that is common in the literature (c.f., Phillips, 1995; Phillips, Li and Gao, 2017) and includes many popular vector time series processes such as stationary VAR and VARMA models. The linear process dependence structure can be replaced by alternative mixing dependence conditions with some modifications of the proofs. Assumption 1 combined with (2.2) implies that the nonstationary components $\mathbf{H}_2' X_t$ are full rank nonstationary and not cointegrated. In the asymptotic theory developed later, $\mathbf{H}_2' X_t$ may be correlated with e_{t0}, which implies endogeneity in the system. Assumptions 2 and 3 are commonly used conditions in the varying-coefficient and kernel smoothing literature – e.g., Wang and Phillips (2009b) and Phillips, Li and Gao (2017). In particular, if the coefficient function $\boldsymbol{\beta}(\cdot)$ is Lipschitz continuous on $[0,1]$, Assumption 2 is satisfied with $\gamma = 1$. When greater smoothness conditions are imposed on $\boldsymbol{\beta}(\cdot)$ stronger results are possible with local linear and polynomial smoothing methods, and these will be mentioned in what follows.

2.2 *Kernel degeneracy and double-rotation of the covariate space*

By virtue of Assumption 1 and functional limit theory for linear processes (Phillips and Solo, 1992),

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tz \rfloor} e_t \Rightarrow B(z), \quad e_t = \left(e_{t0}, e_{t1}', e_{t2}'\right)', \quad 0 < z \leq 1 \tag{2.4}$$

where $B(z)$ is a $(d+1)$-dimensional Brownian motion with variance matrix $\mathbf{\Omega}$ defined in Assumption 1 and $\lfloor \cdot \rfloor$ denotes the floor function. Partition the $(d+1) \times (d+1)$ matrix $\mathbf{\Omega}$ into cell submatrices $\mathbf{\Omega}_{ij}$ $(i, j = 0, 1, 2)$ conformably with e_t and set $\omega = \mathbf{\Omega}_{00}$. Let $B(z) = \left[B_0(z), B_1'(z), B_2'(z)\right]'$ be component Brownian motion limit process of the following partial sum processes

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tz \rfloor} e_{t0} \Rightarrow B_0(z), \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tz \rfloor} e_{t1} \Rightarrow B_1(z), \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tz \rfloor} e_{t2} \Rightarrow B_2(z), \tag{2.5}$$

where $B_0(\cdot)$, $B_1(\cdot)$ and $B_2(\cdot)$ are univariate, d_0-dimensional and d_1-dimensional Brownian motions with variance matrices ω, $\mathbf{\Omega}_{11}$ and $\mathbf{\Omega}_{22}$, respectively. The limit theory later in the paper also involves partitioned components of the one-sided long run covariance matrix defined by $\mathbf{\Gamma} = \text{lrcov}^+ (e_t, e_t) := \sum_{h=0}^{\infty} \mathbb{E}(e_0 e_h')$ with cell submatrices $\mathbf{\Gamma}_{ij}$ $(i, j = 0, 1, 2)$ that are conformable with the partition of e_t.

When the p-dimensional process $\{Z_t\}$ is stationary and satisfies some standard regularity conditions, it is not difficult to show that

$$\frac{1}{Th} \sum_{t=1}^{T} Z_t Z_t' K\big(\frac{t - Tz_0}{Th}\big) = \mu_0 \mathbb{E}\left[Z_t Z_t'\right] + o_P(1) = \mathbb{E}\left[Z_t Z_t'\right] + o_P(1). \tag{2.6}$$

Furthermore, if $\mathbb{E}\big[Z_t Z_t'\big]$ is positive definite, the limit of the inverse of the kernel weighted sample moment matrix $\frac{1}{Th}\sum_{t=1}^{T} Z_t Z_t' K\big(\frac{t-Tz_0}{Th}\big)$ also exists and conventional asymptotics hold for kernel estimation with Z_t as regressors. However, when $\{Z_t\}$ is generated by a nonstationary full rank unit root process with innovations that satisfy a functional law similar to (2.4), the weighted sample moment matrix behaves very differently. First, we have $T^{-1/2} Z_{\lfloor Tz \rfloor} \Rightarrow B_\diamond(z)$ for $0 < z \leq 1$,where $B_\diamond(\cdot)$ is a p-dimensional Brownian motion with positive definite variance matrix, from which it might be expected that the normalization rate (Th) in (2.6) would simply be replaced by the rate $(T^2 h)$. However, Phillips, Li and Gao (2017) showed that the matrix $\frac{1}{T^2 h}\sum_{t=1}^{T} Z_t Z_t' K\big(\frac{t-Tz_0}{Th}\big)$ is asymptotically singular when the dimension p exceeds unity. The reason for this degeneracy is that time-varying coefficient kernel regression concentrates attention in the nonstationary process on a particular time coordinate (say Tz_0) and the corresponding realized value of the associated limit of the nonstationary process, in contrast to the time average $\mathbb{E}\big[Z_t Z_t'\big]$ in stationary case. When there are multiple nonstationary regressors, this focus on a single time coordinate produces a limiting signal matrix of deficient rank unity whose zero eigenspace depends on the value of the limit process at that time coordinate. In other words, the *kernel induced degeneracy* which occurs in the matrix $\frac{1}{T^2 h}\sum_{t=1}^{T} Z_t Z_t' K\big(\frac{t-Tz_0}{Th}\big)$ for multivariate integrated Z_t is random, trajectory dependent, and localized to the time value z_0. It may therefore be regarded as a form of *local degeneracy*.

To deal with degeneracy in a prototypical case, Phillips, Li and Gao (2017) introduced a novel rotational decomposition for the kernel-weighted signal matrix $\sum_{t=1}^{T} Z_t Z_t' K\big(\frac{t-Tz_0}{Th}\big)$ to develop the limit theory. The rotation involved the use of a random direction based on the regressors. In the present case, this direction takes the form of the (sample size dependent) vector

$$q_T(z_0) = \frac{b_T(z_0)}{\big[b_T(z_0)' b_T(z_0)\big]^{1/2}} = \frac{b_T(z_0)}{\|b_T(z_0)\|}, \quad b_T(z_0) = \frac{1}{\sqrt{T}} Z_{\delta(z_0)}, \quad \delta(z_0) = \lfloor T(z_0 - h) \rfloor,$$

leading to an associated orthogonal matrix

$$\mathbf{Q}_T(z_0) = \big[q_T(z_0), q_T^{\perp}(z_0)\big], \quad \mathbf{Q}_T(z_0)'\mathbf{Q}_T(z_0) = \mathbf{I}_p,$$

where $q_T^{\perp}(z_0)$ is an orthogonal complement to $q_T(z_0)$. Using the standardization matrix $\mathbf{D}_T = \mathsf{diag}\Big\{T\sqrt{h},\ (Th)\mathbf{I}_{p-1}\Big\}$, and Proposition A.1 from Phillips, Li and Gao (2017), we may show that the matrix

$$\mathbf{D}_T^{-1}\mathbf{Q}_T(z_0)' \left[\sum_{t=1}^{T} Z_t Z_t' K\big(\frac{t-Tz_0}{Th}\big)\right] \mathbf{Q}_T(z_0)\mathbf{D}_T^{-1}$$

is of full rank with probability approaching one.

This random rotation technique needs substantial generalization for the setting of the present paper. Here the regressors satisfy the framework (2.2), indicating that three different normalization rates might be needed when $d_0 \geq 1$ and $d_1 \geq 2$, where d_0 and d_1 are the dimensions of the stationary components and nonstationary components. To see this, we first use the orthogonal transformation (2.2) to rotate the regressor space and separate out the stationary and nonstationary components. Define

$$X_{t1} = \mathbf{H}_1' X_t, \quad X_{t2} = \mathbf{H}_2' X_t,$$

where X_{t1} is the d_0-dimensional stationary component and X_{t2} is the d_1-dimensional nonstationary component with unit roots. Then, model (1.1) can be re-written as

$$Y_t = \boldsymbol{\beta}_{t1}' X_{t1} + \boldsymbol{\beta}_{t2}' X_{t2} + e_{t0}, \tag{2.7}$$

with $\boldsymbol{\beta}_{t1}' = \boldsymbol{\beta}_t' \mathbf{H}_1$ and $\boldsymbol{\beta}_{t2}' = \boldsymbol{\beta}_t' \mathbf{H}_2$. Letting $\overline{X}_t = (X_{t1}', X_{t2}')' = \mathbf{H}' X_t$, we transform the Nadaraya-Watson kernel estimate $\widehat{\boldsymbol{\beta}}(z_0)$ to

$$\overline{\boldsymbol{\beta}}(z_0) := \mathbf{H}' \widehat{\boldsymbol{\beta}}(z_0) = \Big[\sum_{t=1}^{T} \overline{X}_t \overline{X}_t' K\big(\frac{t - Tz_0}{Th}\big) \Big]^{+} \Big[\sum_{t=1}^{T} \overline{X}_t Y_t K\big(\frac{t - Tz_0}{Th}\big) \Big], \tag{2.8}$$

which is the estimate of $\mathbf{H}'\boldsymbol{\beta}(z_0)$. The component matrix $\mathbf{H}_1$ generates the stationary components and the convergence rate in this direction will be seen to be the same as the usual convergence rate in stationary kernel regression. In contrast, the component matrix $\mathbf{H}_2$, which is orthogonal to $\mathbf{H}_1$, generates full rank nonstationary variates, leading to faster convergence rates in this direction. However, the above arguments show that the matrix $\frac{1}{T^2 h} \sum_{t=1}^{T} X_{t2} X_{t2}' K\big(\frac{t - Tz_0}{Th}\big)$ is asymptotically singular if its dimension d_1 exceeds unity. Therefore, further transformation of the nonstationary component X_{t2} is required in order to resolve asymptotic behavior.

To proceed, let $q_{T2}(z_0)$ and $\mathbf{Q}_{T2}(z_0)$ be defined just as $q_T(z_0)$ and $\mathbf{Q}_T(z_0)$ above but with Z_t replaced by X_{t2}. Then define

$$\overline{\mathbf{Q}}_T(z_0) = \mathsf{diag}\left\{ \mathbf{I}_{d_0}, \mathbf{Q}_{T2}(z_0) \right\}, \quad \overline{\mathbf{D}}_T = \mathsf{diag}\left\{ \sqrt{Th}\mathbf{I}_{d_0}, \ T\sqrt{h}, \ (Th)\mathbf{I}_{d_1 - 1} \right\}. \tag{2.9}$$

Unlike the transformation matrix $\mathbf{H}$ in the *global rotation* which does not rely on z_0, the matrix $\overline{\mathbf{Q}}_T(z_0)$ used in the further rotation of the nonstationary component space is random and time dependent on

z_0, and is thus called a *local rotation*. Proposition 1 below shows that the matrix

$$\overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\Big[\sum_{t=1}^{T}\overline{X}_t\overline{X}_t'K\big(\frac{t-Tz_0}{Th}\big)\Big]\overline{\mathbf{Q}}_T(z_0)\overline{\mathbf{D}}_T^{-1}$$

is of full rank with probability approaching one as $T\to\infty$.

To complete the statement of the proposition we introduce the following notation. Define the vector

$$q_2(z_0)=\frac{b(z_0)}{\big[b(z_0)'b(z_0)\big]^{1/2}}=\frac{b(z_0)}{\|b(z_0)\|}\quad\text{with}\quad b(z_0)=B_2(z_0),$$

and let $q_2^{\perp}(z_0)$ be a $d_1\times(d_1-1)$ orthogonal complement matrix of $q_2(z_0)$. Define the $d\times d$ matrix

$$\mathbf{\Lambda}(z_0)=\mathsf{diag}\left\{\mathbf{\Lambda}_{11},\mathbf{\Lambda}_2(z_0)\right\}\quad\text{with}\quad\mathbf{\Lambda}_2(z_0)=\begin{bmatrix}\mathbf{\Lambda}_{22}(z_0) & \mathbf{\Lambda}_{23}(z_0)\\ \mathbf{\Lambda}_{32}(z_0) & \mathbf{\Lambda}_{33}(z_0)\end{bmatrix},\tag{2.10}$$

where $\mathbf{\Lambda}_{11}=\mathbb{E}[e_{11}e_{11}']>0$ is independent of z_0, $\mathbf{\Lambda}_{22}(z_0)=\lambda(z_0)=B_2(z_0)'B_2(z_0)$ is a univariate random element, $\mathbf{\Lambda}_{23}(z_0)=\mathbf{\Lambda}_{32}(z_0)'=\sqrt{2}\big[B_2(z_0)'B_2(z_0)\big]^{1/2}\left[\int_{-1}^{1}B_2^*(\frac{z+1}{2})'K(z)dz\right]q_2^{\perp}(z_0)$, $\mathbf{\Lambda}_{33}(z_0)=2q_2^{\perp}(z_0)'\left[\int_{-1}^{1}B_2^*(\frac{z+1}{2})B_2^*(\frac{z+1}{2})'K(z)dz\right]q_2^{\perp}(z_0)$, and $B_2^*(\cdot)$ is an independent copy of the Brownian motion $B_2(\cdot)$.

PROPOSITION 1. *Suppose that Assumptions 1 and 3 are satisfied, $d\geq 3$ with $1\leq d_0\leq d-2$ and $2\leq d_1\leq d-d_0$. Then we have*

$$\begin{aligned}\overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\mathbf{H}'\mathbf{\Lambda}_T(z_0)\mathbf{H}\overline{\mathbf{Q}}_T(z_0)\overline{\mathbf{D}}_T^{-1} &= \overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\left[\sum_{t=1}^{T}\overline{X}_t\overline{X}_t'K\big(\frac{t-Tz_0}{Th}\big)\right]\overline{\mathbf{Q}}_T(z_0)\overline{\mathbf{D}}_T^{-1}\\ &\Rightarrow \mathbf{\Lambda}(z_0)>0\ \ a.s.\end{aligned}\tag{2.11}$$

for fixed $0<z_0<1$, where the notation ">0" denotes positive definiteness.

REMARK 1. This proposition resolves the asymptotic degeneracy of the kernel-weighted signal matrix through a *double-rotation* of the nonstationary regressor space involving the global rotation $\mathbf{H}$ and local rotation $\overline{\mathbf{Q}}_T(z_0)$. This transformation leads to three different normalization rates embodied in the standardization matrix $\overline{\mathbf{D}}_T$. For a special case $d_1=1$, kernel degeneracy is circumvented and the rate (Th) disappears in $\overline{\mathbf{D}}_T$, leaving only the global rotation $\mathbf{H}$. If there is no cointegration among the regressors, the global rotation is not needed in transforming the regressors and the rate $(\sqrt{Th})$ would disappear in $\overline{\mathbf{D}}_T$, specializing the result to Proposition A.1 in Phillips, Li and Gao (2017).

2.3 *Asymptotic theory for cointegrated regressors*

This section derives asymptotic theory for the kernel estimator $\widehat{\boldsymbol{\beta}}(z_0)$ when the nonstationary regressors are cointegrated. We start by introducing notation. Let $\overline{\mathbf{D}}_{T2} = \mathsf{diag}\left\{T\sqrt{h},\ (Th)\mathbf{I}_{d_1-1}\right\}$ and define $\boldsymbol{\Delta}_2(z_0) := \left[\delta(z_0), \boldsymbol{\Delta}_\perp(z_0)'\right]'$ with

$$\begin{aligned}
\delta(z_0) &:= \left[2B_2(z_0)'B_2(z_0)\right]^{1/2}\int_{-1}^{1}K(z)dB_0\left(\frac{z+1}{2}\right),\\
\boldsymbol{\Delta}_\perp(z_0) &:= 2q_2^\perp(z_0)'\left[\int_{-1}^{1}K(z)B_2^*\left(\frac{z+1}{2}\right)dB_0\left(\frac{z+1}{2}\right)+\frac{1}{2}\boldsymbol{\Gamma}_{20}\right],
\end{aligned}$$

where the one sided long run covariance $\boldsymbol{\Gamma}_{20} = \mathrm{lrcov}^+\left(e_{t2}, e_{t0}\right)$ is as defined earlier. The following theorem gives the asymptotic distribution of $\widehat{\boldsymbol{\beta}}(z_0)$.

Theorem 1. *Suppose Assumptions 1–3 are satisfied, $d \geq 3$ with $1 \leq d_0 \leq d-2$ and $2 \leq d_1 \leq d-d_0$, $\boldsymbol{\Lambda}(z_0)$ is non-singular with probability one, and for $s \geq t$, $\mathbb{E}[e_{s0}e_{t1}] = \mathbf{0}_{d_0}$, where $\mathbf{0}_r$ is an r-dimensional vector of zeros.*

(i) If, in addition, $Th^{1+2\gamma} = o(1)$, then as $T \to \infty$

$$\sqrt{Th}\mathbf{H}_1'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] \Rightarrow \boldsymbol{\xi} \tag{2.12}$$

for fixed $z_0 \in (0,1)$, where $\boldsymbol{\xi}$ is a d_0-dimensional normal vector with mean zero and covariance matrix

$$\mathbf{V}_\xi := \nu_0\boldsymbol{\Lambda}_{11}^+\Big\{\sum_{s=-\infty}^{\infty}\mathbb{E}\left[(e_{10}e_{s0})(e_{11}e_{s1}')\right]\Big\}\boldsymbol{\Lambda}_{11}^+,\quad \nu_0 = \int K^2(u)du.$$

(ii) If, in addition, $T^2h^{1+2\gamma} = o(1)$, then as $T \to \infty$

$$\overline{\mathbf{D}}_{T2}\overline{\mathbf{Q}}_{T2}(z_0)'\mathbf{H}_2'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] \Rightarrow \boldsymbol{\Lambda}_2^{-1}(z_0)\boldsymbol{\Delta}_2(z_0) \tag{2.13}$$

for fixed $z_0 \in (0,1)$.

Remark 2. (a) The limit theory in Theorem 1 shows that double-rotation of the regressor space is needed to characterize the asymptotics: the global rotator $\mathbf{H}$ addresses potential cointegration among the nonstationary regressors; and the local rotator $\overline{\mathbf{Q}}_T(z_0)$ addresses the kernel degeneracy that arises from the fixed design functional framework. The limit theory in (2.12) and (2.13) encompasses several interesting results from the existing literature. For the case $d_1 = 0$ corresponding to a stationary regressor model, taking $\mathbf{H} = \mathbf{H}_1 = \mathbf{I}_d$ we have from (2.12) that

$$\sqrt{Th}\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] \Rightarrow \boldsymbol{\xi}, \tag{2.14}$$

which delivers results similar to those in the literature on kernel estimation with locally stationary regressors (c.f., Robinson, 1989; Cai, 2007; Vogt, 2012). For the case $d_0 = 0$ corresponding to a full rank integrated regressor model, taking $\mathbf{H} = \mathbf{H}_2 = \mathbf{I}_d$ we obtain Theorem 3.1 in Phillips, Li and Gao (2017) as a corollary of (2.13).

(b) The assumption $\mathbb{E}[e_{s0}e_{t1}] = 0$ for $s \geq t$ implies no contemporaneous or feedforward correlation between the stationary regressor components e_{t1} and the equation error e_{t0}, which ensures kernel estimation is consistent in the direction associated with the stationary components $\mathbf{H}_1' X_t$ (c.f., Park and Phillips, 1989). Theorem 1 does not specify the relationship between the limit distributions of the stationary and nonstationary component estimators in parts (i) and (ii) and to do so we impose the following explicit exogeneity condition.

Assumption 1*. *Let $\mathcal{F}_{t-1} = \sigma(e_{t1}, e_{t2}, e_{t-1}, e_{t-2}, \cdots)$ be the σ-algebra generated by $\{e_{s1}, e_{s2}\}_{s\leq t}$ and $\{e_{s0}\}_{s\leq t-1}$. Then $\{(e_{t0}, \mathcal{F}_t)\}$ is a stationary sequence of martingale differences with $\sigma_e^2 = \mathbb{E}[e_{t0}^2|\mathcal{F}_{t-1}] > 0$ a.s.*

Under Assumption 1*, the asymptotic distribution in the direction $\mathbf{H}_1$ is independent of that in the direction $\mathbf{H}_2$, so that the limit variate $\boldsymbol{\xi}$ is independent of the limit variate $\boldsymbol{\Lambda}_2^{+}(z_0)\boldsymbol{\Delta}_2(z_0)$, which facilitates inference concerning the time varying coefficient function. Further, the one-sided long run covariance matrix $\boldsymbol{\Gamma}_{02}$ is eliminated in the random variate $\boldsymbol{\Delta}_2(z_0)$ for this pure cointegration case. Assumption 1* is common in the literature when stationarity is present and appears, for instance, in both Cai, Li and Park (2009) and Li, Phillips and Gao (2016).

(c) From (2.12) and (2.13) in Theorem 1, we find *three* different convergence rates that apply in different directions. In the direction $\mathbf{H}_1$, by (2.12) we have the well-known stationary rate given by

$$\mathbf{H}_1'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] = O_P\left(\frac{1}{\sqrt{Th}}\right), \tag{2.15}$$

which holds for stationary kernel regression. In the direction $\mathbf{H}_2 q_{T2}(z_0)$, we have the faster rate

$$q_{T2}(z_0)'\mathbf{H}_2'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] = O_P\left(\frac{1}{T\sqrt{h}}\right), \tag{2.16}$$

which is called *type 1 super-consistency* in Li, Phillips and Gao (2016) and Phillips, Li and Gao

(2017). Finally, in direction $\mathbf{H}_2 q_{T2}^{\perp}(z_0)$, we have *type 2 super-consistency* with rate given by

$$q_{T2}^{\perp}(z_0)'\mathbf{H}_2'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] = O_P\left(\frac{1}{Th}\right). \tag{2.17}$$

The type 2 super-consistency rate is slower than the rate in (2.16), but is still faster than the stationary rate in (2.15). Interestingly, therefore, nonstationary regressors raise the rate of convergence over the standard stationary rate in the two relevant directions of nonstationarity in the data.

(d) The bandwidth conditions $Th^{1+2\gamma} = o(1)$ and $T^2h^{1+2\gamma} = o(1)$ in Theorem 1 may appear restrictive. However, if the coefficient function has continuous derivatives up to the second order and if we apply local linear kernel smoothing rather than local costant estimation, then following the proof of Theorem 3.2 in Phillips, Li and Gao (2017), we may relax the above two bandwidth restrictions to $Th^5 = o(1)$ and $T^2h^5 = o(1)$, respectively.

(e) Theorem 1 implies that $\widehat{\boldsymbol{\beta}}(z_0)$ has a degenerate asymptotic normal distribution dominated by the slowest convergent component in (2.12), viz.,

$$\sqrt{Th}\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] \Rightarrow \mathbf{H}_1\boldsymbol{\xi}. \tag{2.18}$$

In spite of this apparent simplification arising from the dominating direction $\mathbf{H}_1$, inference about the full vector of parameters $\boldsymbol{\beta}$ is typically not degenerate and involves the asymptotic behavior of the components of $\boldsymbol{\beta}$ in other directions.

3 Extensions of kernel estimation theory

This section develops kernel estimation theory for the following two cases: (i) the regressors X_t have a mixture of deterministic and stochastic trends but no internal cointegrating structure; and (ii) the regressors X_t have deterministic trends and are cointegrated among themselves.

3.1 *Kernel estimation with stochastic and deterministic trends*

We assume the regressors are generated as stochastic trends with drift according to the scheme

$$X_t = X_{t-1} + \boldsymbol{\mu} + u_t, \tag{3.1}$$

where $\boldsymbol{\mu}$ is a d-dimensional parameter vector representing the accompanying drift of the unit root process, and where $u_t = (e_{t1}', e_{t2}')'$ with e_{t1} and e_{t2} satisfying Assumption 1 in Section 2.1. From (3.1), we have

$$X_t = \sum_{j=1}^{t} u_j + \boldsymbol{\mu} t + X_0 =: S_t + D_t + X_0, \tag{3.2}$$

where $X_0 = O_P(1)$, $S_t := \sum_{j=1}^{t} u_j$ is the stochastic trend and $D_t := \boldsymbol{\mu} t$ is the deterministic drift. Although we consider only a linear trend for D_t in what follows, the method and theory developed in this section are readily extendable to polynomial trends. But general power trends such as t^α with unknown power parameter α involve further complications of asymptotic singularity - see Phillips (2007) and Baek, Cho and Phillips (2015), which are not pursued here.

Since $S_t = O_P\left(t^{1/2}\right)$, the stochastic trend S_t is asymptotically dominated by the deterministic trend D_t. Therefore, we have

$$\sum_{t=1}^{T} X_t X_t' K\left(\frac{t - T z_0}{Th}\right) = \boldsymbol{\mu}\boldsymbol{\mu}' (\lfloor T z_0 \rfloor)^2 Th (1 + o_P(1)). \tag{3.3}$$

When the dimension d exceeds unity, the matrix $\boldsymbol{\mu}\boldsymbol{\mu}'$ is singular, complicating normalization of the kernel-weighted signal matrix

$$\sum_{t=1}^{T} X_t X_t' K\left(\frac{t - T z_0}{Th}\right), \tag{3.4}$$

which is degenerate at the dominating rate $(T^3 h)$ associated with the deterministic component D_t. Degeneracy of this form has long been studied in the linear nonstationary regression literature, where Phillips and Perron (1988) gave a global rotation technique (with non-random transformation matrix) to separate out the stochastic and deterministic trend components with associated standardization rates for the corresponding directions. This global rotation technique cannot be applied in the present setting, however, as will be demonstrated later in this section. Since the kernel-weighted signal matrix (3.4) embodies both stochastic and deterministic trends, a local rotation technique similar to that in Section 2.2 is instead required.

To proceed, define

$$\widetilde{q}_T(z_0) = \frac{\widetilde{b}_T(z_0)}{\left[\widetilde{b}_T(z_0)'\widetilde{b}_T(z_0)\right]^{1/2}} = \frac{\widetilde{b}_T(z_0)}{\|\widetilde{b}_T(z_0)\|}, \quad \widetilde{b}_T(z_0) = X_{\delta(z_0)} = S_{\delta(z_0)} + D_{\delta(z_0)} + X_0,$$

and introduce the orthogonal matrix

$$\widetilde{\mathbf{Q}}_T(z_0) = \left[\widetilde{q}_T(z_0), \widetilde{q}_T^{\perp}(z_0)\right], \quad \widetilde{\mathbf{Q}}_T(z_0)'\widetilde{\mathbf{Q}}_T(z_0) = \mathbf{I}_d,$$

where $\widetilde{q}_T^{\perp}(z_0)$ is a $d \times (d-1)$ orthogonal complement matrix of $\widetilde{q}_T(z_0)$, and define the standardization matrix

$$\widetilde{\mathbf{D}}_T = \mathsf{diag}\left\{T\sqrt{Th},\ (Th)\mathbf{I}_{d-1}\right\}. \tag{3.5}$$

Proposition 2 below shows that asymptotic degeneracy of (3.4) is addressed via application of the local rotator and path-dependent transformation matrix $\widetilde{\mathbf{Q}}_T(z_0)$. Some further notation is needed to state the proposition. Let $\widetilde{\boldsymbol{\mu}}_{\perp}$ be a $d \times (d-1)$ orthogonal complement of $\widetilde{\boldsymbol{\mu}} := (\boldsymbol{\mu}'\boldsymbol{\mu})^{-1/2}\boldsymbol{\mu} = \boldsymbol{\mu}/\|\boldsymbol{\mu}\|$, and $\mathbf{J} = (\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\mu}}_{\perp})$ be the corresponding orthogonal transformation matrix, as used in Phillips and Perron (1988). Define the $d \times d$ matrix

$$\widetilde{\mathbf{\Lambda}}(z_0) = \begin{bmatrix} \widetilde{\mathbf{\Lambda}}_{11}(z_0) & \widetilde{\mathbf{\Lambda}}_{12}(z_0) \\ \widetilde{\mathbf{\Lambda}}_{21}(z_0) & \widetilde{\mathbf{\Lambda}}_{22} \end{bmatrix},$$

where $\widetilde{\mathbf{\Lambda}}_{11}(z_0) = \widetilde{\lambda}(z_0) = \|\boldsymbol{\mu} z_0\|^2$ is non-random and univariate,

$$\widetilde{\mathbf{\Lambda}}_{12}(z_0) = \widetilde{\mathbf{\Lambda}}_{21}(z_0)' = \sqrt{2}\|\boldsymbol{\mu} z_0\| \left[\int_{-1}^{1} \widetilde{B}(\frac{z+1}{2})'K(z)dz\right] \widetilde{\boldsymbol{\mu}}_{\perp},$$
$$\widetilde{\mathbf{\Lambda}}_{22} = 2\widetilde{\boldsymbol{\mu}}_{\perp}' \left[\int_{-1}^{1} \widetilde{B}(\frac{z+1}{2})\widetilde{B}(\frac{z+1}{2})'K(z)dz\right] \widetilde{\boldsymbol{\mu}}_{\perp},$$

$\widetilde{B}(\cdot) = [B_1(\cdot)', B_2(\cdot)']'$, $B_1(\cdot)$ and $B_2(\cdot)$ are defined earlier in (2.5).

Proposition 2. *Suppose Assumptions 1 and 3 are satisfied, $\boldsymbol{\mu} \neq \mathbf{0}_d$ and $d \geq 2$. Then*

$$\widetilde{\mathbf{D}}_T^{-1}\widetilde{\mathbf{Q}}_T(z_0)'\mathbf{\Lambda}_T(z_0)\widetilde{\mathbf{Q}}_T(z_0)\widetilde{\mathbf{D}}_T^{-1} \Rightarrow \widetilde{\mathbf{\Lambda}}(z_0) > 0 \ \ a.s. \tag{3.6}$$

for fixed $0 < z_0 < 1$.

Remark 3. In the proof of (3.6) in Appendix 7, the two random and trajectory-dependent directions $\widetilde{q}_T(z_0)$ and $\widetilde{q}_T^{\perp}(z_0)$ are shown to converge to $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\mu}}_{\perp}$, respectively. Both $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\mu}}_{\perp}$ are non-random, and neither of them rely on z_0. This is unsurprising as the nonstationary process X_t is asymptotically dominated by its linear trend D_t. A natural question in view of this asymptotic behavior is whether the local transformation matrix $\widetilde{\mathbf{Q}}_T(z_0)$ can be replaced by the global matrix $\mathbf{J} = (\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\mu}}_{\perp})$ in (3.6)?

The question is simply answered by examining the special case $d = 2$. Define

$$\widetilde{X}_t := \mathbf{J}' X_t = \left(\widetilde{\boldsymbol{\mu}}' X_t, \widetilde{\boldsymbol{\mu}}_\perp' X_t\right)' =: (\widetilde{X}_{t1}, \widetilde{X}_{t2})'. \tag{3.7}$$

It is easy to see that the univariate component $\widetilde{X}_{t1}$ represents the deterministic trend term, whereas $\widetilde{X}_{t2}$ represents a unit root process without the involvement of the deterministic trend. Defining $\widetilde{\mathbf{D}}_{T*} = \mathsf{diag}\left\{T\sqrt{Th},\ T\sqrt{h}\right\}$, we can show that

$$\widetilde{\mathbf{D}}_{T*}^{-1}\left[\sum_{t=1}^{T} \widetilde{X}_t \widetilde{X}_t' K\big(\frac{t - Tz_0}{Th}\big)\right] \widetilde{\mathbf{D}}_{T*}^{-1} \Rightarrow \left[\begin{array}{cc} \|\boldsymbol{\mu} z_0\|^2 & \|\boldsymbol{\mu} z_0\| \|\widetilde{\boldsymbol{\mu}}_\perp' \widetilde{B}(z_0)\| \\ \|\boldsymbol{\mu} z_0\| \|\widetilde{\boldsymbol{\mu}}_\perp' \widetilde{B}(z_0)\| & \|\widetilde{\boldsymbol{\mu}}_\perp' \widetilde{B}(z_0)\|^2 \end{array}\right]. \tag{3.8}$$

The above result is easily established by noting that the asymptotic leading terms for $\widetilde{X}_{t1}$ and $\widetilde{X}_{t2}$ are $\|\boldsymbol{\mu} z_0\| \cdot T$ and $S_{\delta(z_0)}$, respectively, when $T(z_0 - h) \leq t \leq T(z_0 + h)$. Clearly, the matrix on the right side of (3.8) is singular with probability one. This outcome shows that use of the global (limit) transformation matrix $\mathbf{J} = (\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\mu}}_\perp)$ inadequately deals with the kernel signal matrix degeneracy even though the two relevant directions $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\mu}}_\perp$ figure prominently in the limit. Instead, the local rotator $\widetilde{\mathbf{Q}}_T(z_0)$ and the associated normalization matrix $\widetilde{\mathbf{D}}_T = \mathsf{diag}\left\{T\sqrt{Th},\ (Th)\mathbf{I}_{d-1}\right\}$ in place of $\widetilde{\mathbf{D}}_{T*}$ play key roles in achieving a non-degenerate limit theory.

With Proposition 2 in hand, the limit theory for the kernel estimator $\widehat{\boldsymbol{\beta}}(z_0)$ can now be obtained for stochastic trend with drift regressors, as in (3.1). Let

$$\widetilde{\boldsymbol{\Delta}}(z_0) = \left[\widetilde{\delta}(z_0), \widetilde{\boldsymbol{\Delta}}_\perp'\right]'$$

with

$$\widetilde{\delta}(z_0) = \sqrt{2}\|\boldsymbol{\mu} z_0\| \int_{-1}^{1} K(z) dB_0\big(\frac{z+1}{2}\big), \quad \widetilde{\boldsymbol{\Delta}}_\perp = 2\widetilde{\boldsymbol{\mu}}_\perp' \left[\int_{-1}^{1} K(z) \widetilde{B}\big(\frac{z+1}{2}\big) dB_0\big(\frac{z+1}{2}\big) + \frac{1}{2}\widetilde{\boldsymbol{\Gamma}}\right],$$

where $\widetilde{\boldsymbol{\Gamma}} = (\boldsymbol{\Gamma}_{10}', \boldsymbol{\Gamma}_{20}')'$, and the one-sided long run covariance matrices $\boldsymbol{\Gamma}_{10} = \mathrm{lrcov}^+(e_{t1}, e_{t0})$, $\boldsymbol{\Gamma}_{20} = \mathrm{lrcov}^+(e_{t2}, e_{t0})$ are defined in Section 2.2. The following theorem gives the asymptotic distribution of $\widehat{\boldsymbol{\beta}}(z_0)$.

Theorem 2. *Suppose Assumptions 1–3 are satisfied, $\boldsymbol{\mu} \neq \mathbf{0}_d$, and $d \geq 2$. Then, as $T \to \infty$, we have*

$$\widetilde{\mathbf{D}}_T \widetilde{\mathbf{Q}}_T(z_0)' \left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0) + O_P(h^\gamma)\right] \Rightarrow \widetilde{\boldsymbol{\Lambda}}^{-1}(z_0) \widetilde{\boldsymbol{\Delta}}(z_0), \tag{3.9}$$

for fixed $z_0 \in (0, 1)$.

REMARK 4. (a) Since the limit of the direction $\widetilde{q}_T^{\perp}(z_0)$ is independent of z_0, it is interesting to find that both the random matrix $\widetilde{\mathbf{\Lambda}}_{22}$ (the lower-right block matrix of $\widetilde{\mathbf{\Lambda}}(z_0)$) and the random vector $\widetilde{\mathbf{\Delta}}_{\perp}$ are also independent of z_0. In the above theorem, in order to make the bias term asymptotically negligible, we have to impose the strong restriction $T^3 h^{1+2\gamma} = o(1)$, which contradicts the normal bandwidth condition $Th \to \infty$ made in Assumption 3(ii) since $\gamma \in (0.5, 1]$. However, as discussed in Remark 2(d), if local linear smoothing estimation of the coefficient function is used, we may replace the requirement $T^3 h^{1+2\gamma} = o(1)$ by the weaker condition $T^3 h^5 = o(1)$, which is compatible with $Th \to \infty$.

(b) Compared with Theorem 1 in Section 2.3, there is a single rotator matrix $\widetilde{\mathbf{Q}}_T(z_0)$ involved in the limit theory. In consequence, we have *two* different convergence rates. In the direction $\widetilde{q}_T(z_0)$, we have

$$\widetilde{q}_T(z_0)' \left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0) \right] = O_P\left(\frac{1}{T\sqrt{Th}} + h^{\gamma} \right), \tag{3.10}$$

where $1/(T\sqrt{Th})$ is the new super-consistency convergence rate for nonstationary kernel regression that exceeds the rates in (2.15)–(2.17). This fast rate is mainly due to the strong signal from the linear trend of X_t in the direction $\widetilde{q}_T(z_0)$. In contrast, in the direction of $\widetilde{q}_T^{\perp}(z_0)$, from (3.9), we have

$$\widetilde{q}_T^{\perp}(z_0)' \left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0) \right] = O_P\left(\frac{1}{Th} + h^{\gamma} \right), \tag{3.11}$$

which is the same rate as that for type 2 super-consistency in (2.17) if the bias term order is ignored asymptotically. This rate is due to the relatively weaker signal that emerges in the direction $\widetilde{q}_T^{\perp}(z_0)$ as the linear trend cancels out through the transform $\widetilde{q}_T^{\perp}(z_0)$.

3.2 *Kernel cointegrating regression with deterministic trends*

Next, we combine the structures (2.2) and (3.1) and assume that X_t satisfies

$$\mathbf{H}_1' X_t = e_{t1}, \quad \Delta\left(\mathbf{H}_2' X_t\right) = \mathbf{H}_2' \left(\Delta X_t\right) = e_{t2} + \boldsymbol{\mu}, \tag{3.12}$$

where $\mathbf{H}_1$, $\mathbf{H}_2$, e_{t1} and e_{t2} are defined as in Section 2.1 and $\boldsymbol{\mu}$ is defined as in Section 3.1. The following example shows that the structure (3.12) is satisfied for a cointegrated VAR model with a deterministic drift component.

Example 2. Consider the VAR model defined by

$$\Delta X_t = \boldsymbol{\alpha}\boldsymbol{\beta}' X_{t-1} + \boldsymbol{\nu} + v_t, \tag{3.13}$$

where $\{v_t\}$ is a covariance stationary sequence of random vectors with mean zero, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $d \times d_0$ matrices of rank d_0, and $\boldsymbol{\nu}$ is a d-dimensional parameter vector. Letting $\boldsymbol{\beta}_\perp$ and $\boldsymbol{\alpha}_\perp$ be $d \times (d - d_0)$ matrices of full rank satisfying $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{O}_{d_0 \times d_1}$ and $\boldsymbol{\alpha}'\boldsymbol{\alpha}_\perp = \mathbf{O}_{d_0 \times d_1}$, and defining $\mathbf{C} = \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}_\perp' \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}_\perp'$, we have the Granger representation

$$X_t = \mathbf{C} \sum_{j=1}^{t} v_j + \mathbf{C}\boldsymbol{\nu} t + \mathbf{C} X_0 + V_t,$$

where $V_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1} \sum_{i=0}^{\infty} \mathbf{R}^i \boldsymbol{\beta}' v_{t-i}$ is a stationary linear process and the matrix $\mathbf{R} = \mathbf{I}_{d_0} + \boldsymbol{\beta}'\boldsymbol{\alpha}$ has eigenvalues within the unit circle (Johansen, 1991; Cheng and Phillips, 2009). By choosing $\mathbf{H}_1$ and $\mathbf{H}_2$ as in Example 1, it is clear that the cointegrated VAR model (3.13) satisfies the structure (3.12).

To derive the limit theory of $\widehat{\boldsymbol{\beta}}(z_0)$ under the generating mechanism (3.12) for X_t, we first apply the transformation matrix $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$ on the covariate space as in Section 2.2 to separate out stationary and nonstationary components as $X_{t1} = \mathbf{H}_1' X_t$ and $X_{t2} = \mathbf{H}_2' X_t$. For the nonstationary elements X_{t2}, a further local rotation using the matrix $\overline{\mathbf{Q}}_T(z_0)$ defined in (2.9) is applied to overcome kernel degeneracy. Proposition 3 below shows that this double-rotation technique leads to a well defined limit theory for the kernel-weighted signal matrix when X_t is generated by (3.12).

Define the standardization matrix

$$\mathring{\mathbf{D}}_T = \mathsf{diag}\left\{\sqrt{Th}\mathbf{I}_{d_0},\ T\sqrt{Th},\ (Th)\mathbf{I}_{d_1-1}\right\}, \tag{3.14}$$

and the $d \times d$ matrix

$$\mathring{\boldsymbol{\Lambda}}(z_0) = \mathsf{diag}\left\{\boldsymbol{\Lambda}_{11}, \mathring{\boldsymbol{\Lambda}}_2(z_0)\right\} \quad \text{with} \quad \mathring{\boldsymbol{\Lambda}}_2(z_0) = \begin{bmatrix} \mathring{\boldsymbol{\Lambda}}_{22}(z_0) & \mathring{\boldsymbol{\Lambda}}_{23}(z_0) \\ \mathring{\boldsymbol{\Lambda}}_{32}(z_0) & \mathring{\boldsymbol{\Lambda}}_{33} \end{bmatrix},$$

where $\boldsymbol{\Lambda}_{11}$ is defined as in Section 2.2, $\mathring{\boldsymbol{\Lambda}}_{22}(z_0) = \mathring{\lambda}(z_0) = \|\boldsymbol{\mu} z_0\|^2$ is non-random and univariate,

$$\mathring{\boldsymbol{\Lambda}}_{23}(z_0) = \mathring{\boldsymbol{\Lambda}}_{32}(z_0)' = \sqrt{2}\|\boldsymbol{\mu} z_0\| \left[\int_{-1}^{1} B_2\left(\frac{z+1}{2}\right)' K(z) dz\right] \widetilde{\boldsymbol{\mu}}_\perp,$$

$$\mathring{\boldsymbol{\Lambda}}_{33} = 2\widetilde{\boldsymbol{\mu}}_\perp' \left[\int_{-1}^{1} B_2\left(\frac{z+1}{2}\right) B_2\left(\frac{z+1}{2}\right)' K(z) dz\right] \widetilde{\boldsymbol{\mu}}_\perp,$$

$B_2(\cdot)$ is defined in (2.5) and $\widetilde{\boldsymbol{\mu}}_\perp$ is defined as in Section 3.1.

PROPOSITION 3. *Suppose Assumptions 1 and 3 are satisfied, $\boldsymbol{\mu} \neq \mathbf{0}_d$, $d \geq 3$ with $1 \leq d_0 \leq d-2$ and $2 \leq d_1 \leq d - d_0$. Then*

$$\mathring{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\mathbf{H}'\boldsymbol{\Lambda}_T(z_0)\mathbf{H}\overline{\mathbf{Q}}_T(z_0)\mathring{\mathbf{D}}_T^{-1} \Rightarrow \mathring{\boldsymbol{\Lambda}}(z_0) > 0 \quad a.s. \tag{3.15}$$

for fixed $0 < z_0 < 1$.

REMARK 5. The limit result is similar to that in Proposition 1, with two differences. First, the normalization rate $(T\sqrt{h})$ in $\overline{\mathbf{D}}_T$ is replaced by the rate $(T\sqrt{Th})$ in $\mathring{\mathbf{D}}_T$, which is due to the fact that the stochastic trend is asymptotically dominated by the linear trend in the direction of $\mathbf{H}_2$. Second, the limits of the directions $q_{T2}(z_0)$ and $q_{T2}^\perp(z_0)$ in the above proposition are non-random and independent of the point z_0 (i.e., $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\mu}}_\perp$), whereas the corresponding limits in Proposition 1 are random and depend on z_0.

To provide the limit distribution of the kernel estimator under (3.12), we introduce further notation, defining $\mathring{\mathbf{D}}_{T2} = \mathsf{diag}\left\{T\sqrt{Th},\ (Th)\mathbf{I}_{d_1-1}\right\}$ and $\mathring{\boldsymbol{\Delta}}_2(z_0) = \left[\mathring{\delta}(z_0), \mathring{\boldsymbol{\Delta}}_\perp'\right]'$, with

$$\begin{aligned}
\mathring{\delta}(z_0) &= \sqrt{2}\|\boldsymbol{\mu}z_0\| \int_{-1}^{1} K(z)dB_0\left(\frac{z+1}{2}\right), \\
\mathring{\boldsymbol{\Delta}}_\perp &= 2\widetilde{\boldsymbol{\mu}}_\perp' \left[\int_{-1}^{1} K(z)B_2\left(\frac{z+1}{2}\right)dB_0\left(\frac{z+1}{2}\right) + \frac{1}{2}\boldsymbol{\Gamma}_{20}\right].
\end{aligned}$$

The limit theory for $\widehat{\boldsymbol{\beta}}(z_0)$ is as follows.

THEOREM 3. *Suppose that Assumptions 1–3 are satisfied, $\boldsymbol{\mu} \neq \mathbf{0}_d$, $d \geq 3$ with $1 \leq d_0 \leq d-2$ and $2 \leq d_1 \leq d - d_0$, and $\mathring{\boldsymbol{\Lambda}}(z_0)$ is non-singular with probability one. For $s \geq t$, $\mathbb{E}[e_{s0}e_{t1}] = \mathbf{0}_{d_0}$.*

(i) If, in addition, $Th^{1+2\gamma} = o(1)$, then (2.12) holds as $T \to \infty$.

(ii) For fixed $z_0 \in (0,1)$,

$$\mathring{\mathbf{D}}_{T2}\overline{\mathbf{Q}}_{T2}(z_0)'\mathbf{H}_2'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0) + O_P(h^\gamma)\right] \Rightarrow \mathring{\boldsymbol{\Lambda}}_2^{-1}(z_0)\mathring{\boldsymbol{\Delta}}_2(z_0) \tag{3.16}$$

as $T \to \infty$.

REMARK 6. Theorem 3 combines parts of Theorems 1 and 2, showing that application of the rotation matrices $\mathbf{H}$ and $\overline{\mathbf{Q}}_T(z_0)$ resolves the degeneracy issue in kernel-weighted signal matrices in a similar way to the transformations in Section 2. However, due to the presence of deterministic trends, the standardization matrix needs modification and leads to a faster convergence rate in the direction

$\mathbf{H}_2 q_{T2}(z_0)$, viz.,

$$q_{T2}(z_0)'\mathbf{H}_2'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] = O_P\left(\frac{1}{T\sqrt{Th}} + h^\gamma\right), \tag{3.17}$$

analogous to that in (3.10). Furthermore, as noted in Remark 2(b), under the additional Assumption 1^*, the limit distributions in Theorem 3(i) and (ii) are mutually independent.

4 FM-kernel estimation

As is apparent in the limit distributions obtained earlier, second-order bias effects are present in the asymptotics whenever the regressors are endogenous. Just as in linear cointegration regression asymptotics, endogeneity may be addressed by using modified estimation methods, such as those in Phillips and Hansen (1990). This section provides a kernel modification of the Phillips-Hansen approach that is called FM-kernel estimation. This method effectively removes second-order bias effects in the limit distribution associated with the nonstationary direction $\mathbf{H}_2$. To save space, we mainly focus on the case of cointegrated regressors studied in Section 2, and the development of FM-kernel regression is entirely analogous for the cases of mixed stochastic and deterministic trend regressors and cointegrated regressors with deterministic trends considered in Section 3. Methods other than FM-kernel regression may also be designed to resolve endogeneity and serial correlation induced bias issues, just as they are in simple cointegrated regression models. But these are not pursued here.

From Theorem 1(ii), the presence of the one-sided long run covariance $\mathbf{\Gamma}_{20}$ between e_{t0} and e_{t2}, induces a second-order bias effect in the limit distribution in the direction $\mathbf{H}_2 q_{T2}^{\perp}(z_0)$. In addition, there is endogeneity arising from correlation between the limit Brownian motions $B_0(\cdot)$ and $B_2(\cdot)$. These bias effects relate directly to those that are present in linear cointegrating regression limit theory as discussed originally in Phillips and Perron (1988); Park and Phillips (1989) and Phillips and Hansen (1990). Although the existence of this bias does not affect the super-consistency rates of kernel estimation, centering is affected, with consequential impact on statistical inference concerning the coefficient functions. The need to remove these sources of bias and to provide valid inferential machinery motivates the development of an FM-kernel estimator.

FM least squares estimation was introduced by Phillips and Hansen (1990) in the context of traditional linear cointegrating models, and was recently generalized by Phillips, Li and Gao (2017) to nonparametric kernel-based estimation in models with full rank integrated regressors. When nonstationary regressors are cointegrated, they are necessarily of deficient rank asymptotically, therefore complicating the development of FM-kernel estimation methodology. To present the required modifications more clearly, we start the development under the assumption (which is later relaxed) that the cointegration rank d_0 and the transformation matrix $\mathbf{H}$ are known a priori, together with the long-run covariance matrices $\mathbf{\Omega}_{02}$, $\mathbf{\Omega}_{20}$, $\mathbf{\Omega}_{22}$, $\mathbf{\Gamma}_{20}$ and $\mathbf{\Gamma}_{22}$. Correction for endogeneity is achieved by removing the following term in kernel estimation

$$\mathbf{B}^{\star}_{T1}(z_0) = \sum_{t=1}^{T} K\big(\frac{t-Tz_0}{Th}\big) X_t \odot \Big[\mathbf{0}'_{d_0}, \big(\mathbf{\Omega}_{02}\mathbf{\Omega}^{+}_{22}\Delta X_{t2}\mathbf{1}_{d_1}\big)'\Big]', \tag{4.1}$$

where $\mathbf{1}_{d_1}$ is a d_1-dimensional column vector of ones and $\odot$ denotes component-wise product. Correction for serial correlation is achieved by removing the term

$$\mathbf{B}^{\star}_{T2}(z_0) = \mathbf{H}\overline{\mathbf{Q}}_T(z_0)\overline{\mathbf{D}}_T \left\{\mathbf{0}'_{d_0+1}, \left[q^{\perp}_{T2}(z_0)' \left(\mathbf{\Gamma}_{20} - \mathbf{\Gamma}_{22}\mathbf{\Omega}^{+}_{22}\mathbf{\Omega}_{20}\right)\right]'\right\}'. \tag{4.2}$$

Combining (4.1) and (4.2), the (infeasible) FM-kernel estimator is constructed as

$$\widehat{\boldsymbol{\beta}}_{\star}(z_0) = \mathbf{\Lambda}^{+}_{T}(z_0)\mathbf{\Delta}^{\star}_{T}(z_0), \quad \text{with } \mathbf{\Delta}^{\star}_{T}(z_0) = \mathbf{\Delta}_T(z_0) - \mathbf{B}^{\star}_{T1}(z_0) - \mathbf{B}^{\star}_{T2}(z_0), \tag{4.3}$$

where $\mathbf{\Lambda}_T(z_0)$ and $\mathbf{\Delta}_T(z_0)$ are defined in (2.1). Since the quantities d_0, $\mathbf{H}$, $\mathbf{\Omega}_{02}$, $\mathbf{\Omega}_{20}$, $\mathbf{\Omega}_{22}$, $\mathbf{\Gamma}_{20}$ and $\mathbf{\Gamma}_{22}$ are generally unknown, the estimator (4.3) is infeasible in practice. But estimation of these unknown elements has been extensively studied in the literature and similar methods may be utilized in the present context, as we now overview, to produce consistent estimators $\widehat{d}$, $\widehat{\mathbf{H}}$, $\widehat{\mathbf{\Omega}}_{02}$, $\widehat{\mathbf{\Omega}}_{20}$, $\widehat{\mathbf{\Omega}}_{22}$, $\widehat{\mathbf{\Gamma}}_{20}$ and $\widehat{\mathbf{\Gamma}}_{22}$, that may be used to construct a feasible version of the FM-kernel estimator.

To simplify, it is convenient to consider the case where the integrated regressors are generated from the cointegrated VAR process (2.3) discussed in Example 1. Then, as in Cheng and Phillips (2009), we may use the Bayesian information criterion to consistently estimate the cointegrating rank d_0 and use reduced rank regression to consistently estimate (under normalizing restrictions) the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in (2.3), and thus obtain a consistent estimator of $\mathbf{H}$ that applies in a general semiparametric setting. Denote the corresponding estimates $\widehat{d}$ and $\widehat{\mathbf{H}} = (\widehat{\mathbf{H}}_1, \widehat{\mathbf{H}}_2)$. Feasible FM-kernel estimation further requires estimation of the various long run covariance matrices that appear in (4.1) and (4.2). We illustrate by estimating $\mathbf{\Omega}_{20}$. The remaining long run covariance matrix estimates may be constructed in a similar manner. Let

$$\widehat{e}_{t0} = Y_t - X'_t\widehat{\boldsymbol{\beta}}(t/T) = Y_t - X'_t\widehat{\boldsymbol{\beta}}_t$$

be the estimated equation errors from kernel regression of (1.1). Let $\widehat{e}_{t2} = \widehat{X}_{t2} - \widehat{X}_{t-1,2} = \Delta(\widehat{\mathbf{H}}_2' X_t)$ with $\widehat{X}_{t2} = \widehat{\mathbf{H}}_2' X_t$, and construct estimates of the component autocovariances $\mathbf{\Omega}_{20}(j) = \mathbb{E}[e_{02} e_{j0}]$ using

$$\widehat{\mathbf{\Omega}}_{20}(j) = \frac{1}{\lfloor (1-\tau)T \rfloor - \lfloor \tau T \rfloor} \sum_{t=\lfloor \tau T \rfloor + 1}^{\lfloor (1-\tau)T \rfloor} \widehat{e}_{t-j,2} \widehat{e}_{t0}, \quad j = -l_T, \cdots, 0, \cdots, l_T, \tag{4.4}$$

in which $0 < \tau < 1/2$ is usually close to zero and $l_T \ll T$ is the lag truncation number which tends to infinity as $T \to \infty$. Unlike the existing literature in parametric linear cointegration models where the cross product $\widehat{e}_{t-j,2}\widehat{e}_{t0}$ is summed over the full domain of t (i.e., $1 \le t, t-j \le T$) to estimate the covariance, our method uses only information over the subinterval from $\lfloor \tau T \rfloor + 1$ to $\lfloor (1-\tau)T \rfloor$ to avoid possible boundary effects when applying kernel estimation. Although such construction of covariance estimates may lose some useful information by taking τ close to zero, consistency of the covariance estimate is unaffected. The corresponding long run covariance estimate based on the components (4.4) is

$$\widehat{\mathbf{\Omega}}_{20} = \sum_{j=-l_T}^{l_T} k\left(j/l_T\right) \widehat{\mathbf{\Omega}}_{20}(j), \tag{4.5}$$

where $k(\cdot)$ is a lag kernel function. Using the known uniform consistency of the kernel estimates such as Theorem 4.1 in Li, Phillips and Gao (2016) and following similar arguments to those in the proof of Theorem 4.2 in Phillips, Li and Gao (2017), consistency of $\widehat{\mathbf{\Omega}}_{20}$ can be established under mild conditions on the lag kernel function $k(\cdot)$ and the truncation number l_T (c.f., Phillips, 1995). The proof is standard and details are omitted to save the space.

With consistent estimates of these parameters in hand, we can construct a feasible version of the FM-kernel estimator of the coefficient functions. Define a feasible version of the endogeneity correction as

$$\mathbf{B}_{T1}^{\#}(z_0) = \sum_{t=1}^{T} K\big(\frac{t - Tz_0}{Th}\big) X_t \odot \left[\mathbf{0}_{\widehat{d}}', \left(\widehat{\mathbf{\Omega}}_{02}\widehat{\mathbf{\Omega}}_{22}^{+}\Delta\widehat{X}_{t2}\mathbf{1}_{d-\widehat{d}}\right)'\right]', \tag{4.6}$$

and a feasible version of the serial correlation correction as

$$\mathbf{B}_{T2}^{\#}(z_0) = \widehat{\mathbf{H}}\overline{\mathbf{Q}}_T^{\#}(z_0)\overline{\mathbf{D}}_T \left\{\mathbf{0}_{\widehat{d}+1}', \left[q_{T2}^{\perp\#}(z_0)' \left(\widehat{\mathbf{\Gamma}}_{20} - \widehat{\mathbf{\Gamma}}_{22}\widehat{\mathbf{\Omega}}_{22}^{+}\widehat{\mathbf{\Omega}}_{20}\right)\right]'\right\}', \tag{4.7}$$

where $q_{T2}^{\perp\#}(z_0)$ and $\overline{\mathbf{Q}}_T^{\#}(z_0)$ are defined similarly to $q_{T2}^{\perp}(z_0)$ and $\overline{\mathbf{Q}}_T(z_0)$ but with X_{t2} replaced by $\widehat{X}_{t2}$.

Using the corrections (4.6) and (4.7), we propose the feasible FM-kernel estimator

$$\widehat{\boldsymbol{\beta}}_{\#}(z_0) = \boldsymbol{\Lambda}_T^{+}(z_0)\boldsymbol{\Delta}_T^{\#}(z_0), \quad \boldsymbol{\Delta}_T^{\#}(z_0) = \boldsymbol{\Delta}_T(z_0) - \mathbf{B}_{T1}^{\#}(z_0) - \mathbf{B}_{T2}^{\#}(z_0), \tag{4.8}$$

and proceed to analyze its asymptotic behavior.

Because of the removal of the endogeneity bias via the correction in $\boldsymbol{\Delta}_T^{\#}(z_0)$, the stochastic integral in the limit distribution $\boldsymbol{\Delta}_2(z_0)$ is modified accordingly. We define $\boldsymbol{\Delta}_2^{\#}(z_0) = \left[\delta^{\#}(z_0), \boldsymbol{\Delta}_{\perp}^{\#}(z_0)'\right]'$ with

$$\begin{aligned}\delta^{\#}(z_0) &= \left[2B_2(z_0)'B_2(z_0)\right]^{1/2} \int_{-1}^{1} K(z) dB_0^{\#}\Big(\frac{z+1}{2}\Big), \\ \boldsymbol{\Delta}_{\perp}^{\#}(z_0) &= 2q_2^{\perp}(z_0)' \int_{-1}^{1} K(z) B_2^{*}\Big(\frac{z+1}{2}\Big) dB_0^{\#}\Big(\frac{z+1}{2}\Big),\end{aligned}$$

where the univariate Brownian motion $B_0^{\#}(\cdot)$ has (conditional) variance $\omega_{0|2} = \omega - \boldsymbol{\Omega}_{02}\boldsymbol{\Omega}_{22}^{+}\boldsymbol{\Omega}_{20}$ following the endogeneity correction and is independent of the d_1-dimensional Brownian motions $B_2(\cdot)$ and $B_2^{*}(\cdot)$. Hence, the component $\boldsymbol{\Delta}_2^{\#}(z_0)$ has a mixed normal distribution which facilitates inference on the time-varying coefficient functions in the same way as the usual FM corrections do in linear cointegrating regression. Noting that the bias correction occurs in the direction $\mathbf{H}_2$, the component transform $\mathbf{H}_1'\widehat{\boldsymbol{\beta}}_{\#}(z_0)$ has the same asymptotic distribution as $\mathbf{H}_1'\widehat{\boldsymbol{\beta}}(z_0)$. Hence, we only examine the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{\#}(z_0)$ in the direction $\mathbf{H}_2$. The following result gives the limit theory in this direction. The asymptotic distribution is mixed normal, giving a nonparametric generalization to the kernel regression case of the original finding in Phillips and Hansen (1990). The asymptotic mixed normality in this direction, coupled with the asymptotic normality in the stationary direction open the way to inference using the FM-kernel estimator.

THEOREM 4. *Suppose the conditions of Theorem 1 hold, $T^2h^{1+2\gamma} = o(1)$, and*

$$\left(\widehat{d}_0, \widehat{\mathbf{H}}, \widehat{\boldsymbol{\Omega}}_{02}, \widehat{\boldsymbol{\Omega}}_{20}, \widehat{\boldsymbol{\Omega}}_{22}, \widehat{\boldsymbol{\Gamma}}_{20}, \widehat{\boldsymbol{\Gamma}}_{22}\right) \to_p \left(d_0, \mathbf{H}, \boldsymbol{\Omega}_{02}, \boldsymbol{\Omega}_{20}, \boldsymbol{\Omega}_{22}, \boldsymbol{\Gamma}_{20}, \boldsymbol{\Gamma}_{22}\right). \tag{4.9}$$

Then, as $T \to \infty$

$$\overline{\mathbf{D}}_{T2}\overline{\mathbf{Q}}_{T2}(z_0)'\mathbf{H}_2' \left[\widehat{\boldsymbol{\beta}}_{\#}(z_0) - \boldsymbol{\beta}(z_0)\right] \Rightarrow \boldsymbol{\Lambda}_2^{-1}(z_0)\boldsymbol{\Delta}_2^{\#}(z_0) \tag{4.10}$$

for fixed $z_0 \in (0,1)$.

5 Nonparametric statistical inference

In practical work interest often focuses on whether time-varying coefficients are well approximated by constant coefficients. To provide apparatus for formal consideration of such hypotheses, in this section we develop an inferential framework of tests for the coefficient functions in model (1.1) and derive asymptotics that enable formal testing.

5.1 *Testing the global null hypothesis*

As in Section 4, we concentrate on the case of cointegrated regressors. The methodology and theory are similar for the other cases studied in Section 3, and so the details are omitted here. Specifically, we consider testing the following null hypothesis

$$\mathcal{H}_0: \ \mathbf{R}\big[\boldsymbol{\beta}(z) - \boldsymbol{\beta}_0\big] = \mathbf{0}_r, \quad 0 \le z \le 1,$$

where $\mathbf{R}$ is an $r \times d$ restriction matrix of rank $r < d$ and $\boldsymbol{\beta}_0$ is a d-dimensional vector of unknown parameters. As $\mathbf{R}$ does not rely on z, we refer to $\mathcal{H}_0$ as a *global* null hypothesis.

Before developing a statistic for testing the hypothesis $\mathcal{H}_0$, we derive a useful result from the limit distributions given in Theorem 1. From (2.12) and (2.13) and as in Remark 2(e), we note that

$$\begin{aligned}\sqrt{Th}\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] &= \sqrt{Th}\mathbf{H}\mathbf{H}'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] \\ &= \sqrt{Th}\mathbf{H}_1\mathbf{H}_1'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] + \sqrt{Th}\mathbf{H}_2\mathbf{H}_2'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] \\ &\Rightarrow \mathbf{H}_1\boldsymbol{\xi} + O_P\left(1/\sqrt{Th}\right) = \mathbf{H}_1\boldsymbol{\xi} + o_P(1), \end{aligned} \tag{5.1}$$

under the assumptions of Theorem 1, where $\boldsymbol{\xi}$ is the Gaussian vector defined in Theorem 1(i). Further, under Assumption 1^*, the vector $\boldsymbol{\xi}$ is a d_0-dimensional centred normal vector with covariance matrix $\nu_0\sigma_e^2\boldsymbol{\Lambda}_{11}^+$, where σ_e^2 is defined in Assumption 1^*. The covariance matrix of $\mathbf{H}_1\boldsymbol{\xi}$ is therefore $\nu_0\sigma_e^2\mathbf{H}_1\boldsymbol{\Lambda}_{11}^+\mathbf{H}_1'$, which has degenerate rank.

Construction of a test statistic based on (5.1) requires consistent estimation of the unknown elements σ_e^2 and $\mathbf{H}_1\boldsymbol{\Lambda}_{11}^+\mathbf{H}_1'$ in the covariance structure. From Phillips (1988) and using Proposition 1 in Section 2, we may use the matrix $\left[\frac{1}{Th}\sum_{t=1}^T X_t X_t' K\left(\frac{t-Tz_0}{Th}\right)\right]^+$ as an estimate of $\mathbf{H}_1\boldsymbol{\Lambda}_{11}^+\mathbf{H}_1'$ in view of the fact that

$$\left[\frac{1}{Th}\sum_{t=1}^{T} X_t X_t' K\left(\frac{t-Tz_0}{Th}\right)\right]^{+} = \mathbf{H}_1 \mathbf{\Lambda}_{11}^{+} \mathbf{H}_1' + o_P(1), \quad h \le z_0 \le 1-h. \tag{5.2}$$

The above convergence holds uniformly over $h \le z_0 \le 1-h$. Let the residual $\widehat{e}_{t0}$ be defined as in Section 4, and construct the equation error variance estimate

$$\widehat{\sigma}_e^2 = \frac{1}{\lfloor (1-\tau)T \rfloor - \lfloor \tau T \rfloor} \sum_{t=\lfloor \tau T \rfloor + 1}^{\lfloor (1-\tau)T \rfloor} \widehat{e}_{t0}^2,$$

where τ is defined as in Section 4, which gives a consistent estimate of σ_e^2 in view of Theorem 4.1 in Li, Phillips and Gao (2016), so that

$$\widehat{\sigma}_e^2 = \sigma_e^2 + o_P(1). \tag{5.3}$$

Next let $\{z_k\}_{k=1}^{m}$ be an equidistant grid of points that satisfy $0 < h = z_1 < z_2 < \cdots < z_{m-1} < z_m = 1-h$ and are chosen from the interval $(0,1)$, where the number m is a fixed positive integer. The extension to divergent m will be discussed later in Remark 7. Using (5.1)–(5.3), we construct point-wise Wald test statistics of $\mathcal{H}_0$ as

$$W_T(z_k) = (Th)\left\{\mathbf{R}\left[\widehat{\boldsymbol{\beta}}(z_k) - \widehat{\boldsymbol{\beta}}\right]\right\}' \left\{\widehat{\sigma}_e^2 \nu_0 \mathbf{R}\left[\frac{1}{Th}\sum_{t=1}^{T} X_t X_t' K\left(\frac{t-Tz_k}{Th}\right)\right]^{+} \mathbf{R}'\right\}^{+} \left\{\mathbf{R}\left[\widehat{\boldsymbol{\beta}}(z_k) - \widehat{\boldsymbol{\beta}}\right]\right\}, \tag{5.4}$$

where $\widehat{\boldsymbol{\beta}}$ is a conventional parametric estimate of $\boldsymbol{\beta}_0$ under the global null hypothesis $\mathcal{H}_0$. Assume that

$$\mathsf{rank}\left(\mathbf{R}\mathbf{H}_1 \mathbf{\Lambda}_{11}^{+} \mathbf{H}_1' \mathbf{R}'\right) = r \tag{5.5}$$

and that under the null

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_P(1/\sqrt{Th}). \tag{5.6}$$

It is natural to propose a generalized Wald test statistic by summing the component statistics $W_T(z_k)$ over $k = 1, \cdots, m$ giving

$$W_T = \sum_{k=1}^{m} W_T(z_k). \tag{5.7}$$

The following theorem gives the limit distribution of the generalized Wald test statistic.

Theorem 5. *Suppose the conditions of Theorem 1, Assumption 1*, (5.5), (5.6), and* $Th^{1+2\gamma} = o(1)$

hold. Letting the positive integer m be fixed, we have $W_T \Rightarrow \chi^2_{mr}$ under the null hypothesis $\mathcal{H}_0$, where χ^2_{mr} is a central chi-square distribution with (mr) degrees of freedom. In addition, if (4.9) holds, $W_T^{\#} \Rightarrow \chi^2_{mr}$ under the null hypothesis $\mathcal{H}_0$, where $W_T^{\#}$ is constructed in the same manner as W_T but using FM-kernel estimates of the time-varying coefficients.

REMARK 7. (a) The methodology and theory developed above are applicable if we generalize the global null hypothesis $\mathcal{H}_0$ to

$$\mathcal{H}_0^\star: \ \mathbf{R}\big[\boldsymbol{\beta}(z) - \boldsymbol{\beta}(z, \boldsymbol{\gamma})\big] = \mathbf{0}_r, \quad 0 \le z \le 1,$$

where $\boldsymbol{\beta}(\cdot, \boldsymbol{\gamma}) = \big[\beta_1(\cdot, \boldsymbol{\gamma}_1), \cdots, \beta_d(\cdot, \boldsymbol{\gamma}_d)\big]^\tau$, $\beta_i(\cdot, \boldsymbol{\gamma}_i)$, $i = 1, \cdots, d$, are pre-specified nonlinear functional coefficients with $\boldsymbol{\gamma}_i$ being an unknown parameter vector, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\tau, \cdots, \boldsymbol{\gamma}_d^\tau)^\tau$. We let $\widehat{\boldsymbol{\gamma}}$ be the conventional nonlinear least squares estimator of the parameter vector $\boldsymbol{\gamma}$ and correspondingly construct the generalized Wald test statistic as $\widehat{W}_T = \sum_{k=1}^m \widehat{W}_T(z_k)$, where

$$\begin{aligned} \widehat{W}_T(z_k) &= (Th)\left\{\mathbf{R}\left[\widehat{\boldsymbol{\beta}}(z_k) - \boldsymbol{\beta}(z_k, \widehat{\gamma})\right]\right\}' \left\{\widehat{\sigma}_e^2 \nu_0 \mathbf{R}\left[\frac{1}{Th}\sum_{t=1}^T X_t X_t' K\big(\frac{t - Tz_k}{Th}\big)\right]^+ \mathbf{R}'\right\}^+ \\ & \quad \left\{\mathbf{R}\left[\widehat{\boldsymbol{\beta}}(z_k) - \boldsymbol{\beta}(z_k, \widehat{\gamma})\right]\right\}. \end{aligned} \tag{5.8}$$

By replacing (5.6) by $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = o_P(1/\sqrt{Th})$ and imposing appropriate smoothness condition on the pre-specified functional coefficients $\boldsymbol{\beta}(\cdot, \cdot)$, we may show the limit distribution of $\widehat{W}_T$ is similar to those given in Theorem 5.

(b) We next briefly discuss the case that m diverges to infinity as $T \to \infty$. Using (5.2) and (5.3), we may prove that

$$W_T(z_k) = (Th)\big[\widehat{\boldsymbol{\beta}}(z_k) - \boldsymbol{\beta}_0\big]'\mathbf{R}'\left[\sigma_e^2 \nu_0 \mathbf{R}\mathbf{H}_1 \boldsymbol{\Lambda}_{11}^+ \mathbf{H}_1' \mathbf{R}'\right]^+ \mathbf{R}\big[\widehat{\boldsymbol{\beta}}(z_k) - \boldsymbol{\beta}_0\big](1 + o_P(1)). \tag{5.9}$$

Following the proof of Theorem 1 in Appendix 7 and noting that the kernel function $K(\cdot)$ has the compact support $[0, 1]$, $\widehat{\boldsymbol{\beta}}(z_k) - \boldsymbol{\beta}_0$ is asymptotically determined by (e_{t0}, e_{t1}), $T(z_k - h) \le t \le T(z_k + h)$. By (5.9) and Assumption 1 in Section 2.1, we may show that $\{W_T(z_k)\}_{k=1}^m$ is a sequence of asymptotically independent random elements when $z_{k+1} - z_k \ge 2h$. So the generalized Wald test statistic W_T can be viewed as a sum of asymptotically independent random variables. By appropriately centralizing W_T and using standard central limit arguments, it is clear that the generalized Wald statistic is asymptotically normal when $m \to \infty$.

5.2 *Testing the local null hypothesis*

We next turn to the more challenging case where the rank condition (5.5) fails. Our approach follows Phillips (1995) in the application of FM regression to models with cointegrated variates where Wald test statistics suffer from rank condition failure asymptotically. As will become apparent, kernel FM regression tests involve further complications under rank condition failure in (5.5).

To proceed, we consider the *localized* version of the null hypothesis:

$$\mathcal{H}_0^{\diamond}: \ \mathbf{R}(z)\left[\boldsymbol{\beta}(z)-\boldsymbol{\beta}_0\right]=\mathbf{0}_r, \quad 0\leq z\leq 1, \tag{5.10}$$

where $\mathbf{R}(z)$ is an $r\times d$ restriction matrix. As discussed in Phillips (1995) rank condition failure occurs when the restriction matrix isolates some of the nonstationary variable coefficients, thereby necessarily involving estimates of these coefficients in the limit theory of Wald-type test statistics. Motivated by Phillips (1995), under such rank condition failure, the restriction matrix can be written in the following form

$$\begin{aligned}\mathbf{R}(z)' &= \left[\mathbf{R}_1(z) \vdots \mathbf{R}_2(z)\right]=(\mathbf{H}_1,\mathbf{H}_2)\begin{bmatrix}\mathbf{S}_1 & \mathbf{S}_{h1} & \vdots & \mathbf{O}_{d_0\times r_2}\\ \mathbf{O}_{d_1\times r_0} & \overline{\mathbf{Q}}_{T2}(z)\mathbf{S}_{h2} & \vdots & \overline{\mathbf{Q}}_{T2}(z)\mathbf{S}_2\end{bmatrix}\\ &= \left[\mathbf{H}_1\mathbf{S}_1,\mathbf{H}_1\mathbf{S}_{h1}+\mathbf{H}_2\overline{\mathbf{Q}}_{T2}(z)\mathbf{S}_{h2} \vdots \mathbf{H}_2\overline{\mathbf{Q}}_{T2}(z)\mathbf{S}_2\right],\end{aligned} \tag{5.11}$$

which in the present case involves the localized structure where $\mathbf{R}_1(z)$ and $\mathbf{R}_2(z)$ are $d\times r_1$ and $d\times r_2$ matrices with $r_1+r_2=r$, $\mathbf{S}_1$, $\mathbf{S}_{h1}$, $\mathbf{S}_{h2}$ and $\mathbf{S}_2$ are the matrices with sizes $d_0\times r_0$, $d_0\times(r_1-r_0)$, $d_1\times(r_1-r_0)$ and $d_1\times r_2$, respectively, $\mathbf{S}_{h1}$ has full column rank. From (5.11), we have that (5.5) reduces as follows

$$\mathbf{R}(z)\mathbf{H}_1\mathbf{\Lambda}_{11}^{+}\mathbf{H}_1'\mathbf{R}(z)'=\begin{bmatrix}\mathbf{R}_1(z)\mathbf{H}_1\mathbf{\Lambda}_{11}^{+}\mathbf{H}_1'\mathbf{R}_1(z)' & \mathbf{O}_{r_1\times r_2}\\ \mathbf{O}_{r_2\times r_1} & \mathbf{O}_{r_2\times r_2}\end{bmatrix} \tag{5.12}$$

whose rank is smaller than r.

The rank deficiency in (5.12) implies that the arguments used above to prove Theorem 5 no longer apply to the generalized Wald statistic for testing the $\mathcal{H}_0^{\diamond}$ and different methods are required. Instead of using Theorem 1 in Section 2, we make use of Theorem 4 in Section 4. In the remainder of this section, we apply the test statistic constructed from the FM-kernel estimates for which the mixed normal distribution derived in Theorem 4 plays an important role in achieving the limit theory. Further, to simplify derivations, we use the uniform kernel $K(u)=I(-1\leq u\leq 1)$ where $I(\cdot)$ denotes the indicator function. We remark that $\mu_0 = 2$ in this case, which differs from the unit normalization condition used in Assumption 3(ii).

Again, we start by defining the following point-wise Wald statistics based on FM-kernel estimation

$$\begin{aligned} W_T^{\diamond}(z_k) &= (Th)\left\{\mathbf{R}(z_k)\left[\widehat{\boldsymbol{\beta}}_{\#}(z_k)-\widehat{\boldsymbol{\beta}}\right]\right\}'\left\{\widehat{\sigma}_e^2\nu_0\mathbf{R}(z_k)\left[\frac{1}{Th}\sum_{t=1}^{T}X_tX_t'K\big(\frac{t-Tz_k}{Th}\big)\right]^{+}\mathbf{R}(z_k)'\right\}^{+} \\ &\quad \left\{\mathbf{R}(z_k)\left[\widehat{\boldsymbol{\beta}}_{\#}(z_k)-\widehat{\boldsymbol{\beta}}\right]\right\}, \end{aligned} \tag{5.13}$$

where $\nu_0 = 2$ since $K(u) = I(-1 \le u \le 1)$. As in (5.7), we construct the generalized Wald test statistic by summing $W_T^{\diamond}(z_k)$ over $k = 1, \cdots, m$, i.e.,

$$W_T^{\diamond} = \sum_{k=1}^{m} W_T^{\diamond}(z_k). \tag{5.14}$$

Define

$$\overline{\mathbf{D}}_T^{\diamond} = \mathsf{diag}\left\{\mathbf{I}_{d_0}, \sqrt{T}, (\sqrt{Th})\mathbf{I}_{d_1-1}\right\}$$

and assume

$$\sqrt{Th}\,\overline{\mathbf{D}}_T^{\diamond}\mathbf{H}'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = o_P(1). \tag{5.15}$$

The following limit theorem provides the asymptotic distribution of $W_T^{\diamond}$ defined in (5.14) under the local null hypothesis, which differs from the earlier limit distribution given in (5.10) and can be viewed as a nonparametric kernel-FM test generalization of Theorem 4.5 in Phillips (1995).

Theorem 6. *Suppose that the conditions of Theorem 1, Assumption 1*, (5.15) and* $T^2h^{1+2\gamma} = o(1)$ *are all satisfied. Letting the positive integer m be fixed, we have*

$$W_T^{\diamond} \Rightarrow \chi^2_{mr_1} + \frac{\omega_{0|2}}{\sigma_e^2}\chi^2_{mr_2,*}, \tag{5.16}$$

under the null hypothesis $\mathcal{H}_0^{\diamond}$ with (5.11), where $\chi^2_{mr_1}$ and $\chi^2_{mr_2,*}$ are two independent chi-square distributions with degrees of freedom mr_1 and mr_2, respectively, and $\omega_{0|2} = \omega - \mathbf{\Omega}_{02}\mathbf{\Omega}_{22}^{+}\mathbf{\Omega}_{20}$.

Remark 8. Note that $\omega_{0|2} = \omega - \mathbf{\Omega}_{02}\mathbf{\Omega}_{22}^{+}\mathbf{\Omega}_{20} \le \omega$ so that the ratio in (5.16) can be written in the form

$$\frac{\omega_{0|2}}{\sigma_e^2} = \frac{\omega - \mathbf{\Omega}_{02}\mathbf{\Omega}_{22}^{+}\mathbf{\Omega}_{20}}{\omega}\frac{\omega}{\sigma_e^2} \le \frac{\omega}{\sigma_e^2}.$$

It follows that

$$\mathsf{P}\left(W_T^{\diamond} > w\right) \to \mathsf{P}\left(\chi^2_{mr_1} + \frac{\omega_{0|2}}{\sigma_e^2}\chi^2_{mr_2,*} > w\right) \le \mathsf{P}\left(\chi^2_{mr_1} + \frac{\omega}{\sigma_e^2}\chi^2_{mr_2,*} > w\right),$$

so that a test of $\mathcal{H}_0^{\diamond}$ based on critical values of the distribution of $\chi^2_{mr_1} + \frac{\omega}{\sigma_e^2}\chi^2_{mr_2,*}$ would be an asymptotically conservative test if r_1 and r_2 were known and consistent estimates of $\omega = \text{lrvar}\left(e_{t0}\right)$ and σ_e^2 were employed in calculating critical values. Further, under Assumption 1*, we have $\omega_{0|2} = \sigma_e^2 - \mathbf{\Omega}_{02}\mathbf{\Omega}_{22}^{+}\mathbf{\Omega}_{20} \le \sigma_e^2$ and then

$$\chi^2_{mr_1} + \frac{\omega}{\sigma_e^2}\chi^2_{mr_2,*} \le \chi^2_{mr_1} + \chi^2_{mr_2,*} =_d \chi^2_{mr},$$

so that a conservative test can be computed directly by using critical values from a χ^2_{mr} distribution.

6 Empirics: time-varying consumption behavior

We next apply the time-varying coefficient model and kernel estimation methodology to aggregate US data on consumption, income, and nominal interest rate obtained from *Federal Reserve Economic Data (FRED)*. We consider a quarterly data set over the first quarter of 1960 to the last quarter of 2009 with 200 observations: c_t is log consumption expenditure, i_t is log disposable income, and nr_t is the nominal interest rate expressed as a percentage. All the three series are plotted in Figure 1, which shows that c_t and i_t have co-moving trend components. The unit root tests confirm nonstationarity for all variables.[1]

Set $Y_t = c_t$ and $X_t = (i_t, i_{t-1}, nr_t)'$, where i_t and i_{t-1} are cointegrated regressors, and nr_t follows a unit root process. Noting that $i_t - i_{t-1}$ is stationary, as in Section 2.1 we may apply the global transformation matrix $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$ with

$$\mathbf{H}_1 = (\sqrt{2}/2, -\sqrt{2}/2, 0)' \quad \text{and} \quad \mathbf{H}_2 = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}',$$

on the covariate space to separate out stationary and nonstationary components as $X_{t1} = \mathbf{H}_1' X_t$ and $X_{t2} = \mathbf{H}_2' X_t$, respectively. We first fit the following time-varying coefficient model:

$$Y_t = \boldsymbol{\beta}_t' \overline{X}_t + e_t, \quad \boldsymbol{\beta}_t = \boldsymbol{\beta}(t/T), \quad t = 1, \cdots, T, \tag{6.1}$$

[1] The PP tests proposed by Phillips and Perron (1988) with fitted mean and linear trend were conducted for c_t and i_t, giving p-values of 0.7248 and 0.7603. The PP test with fitted mean for nr_t gave a p-value of 0.2661.

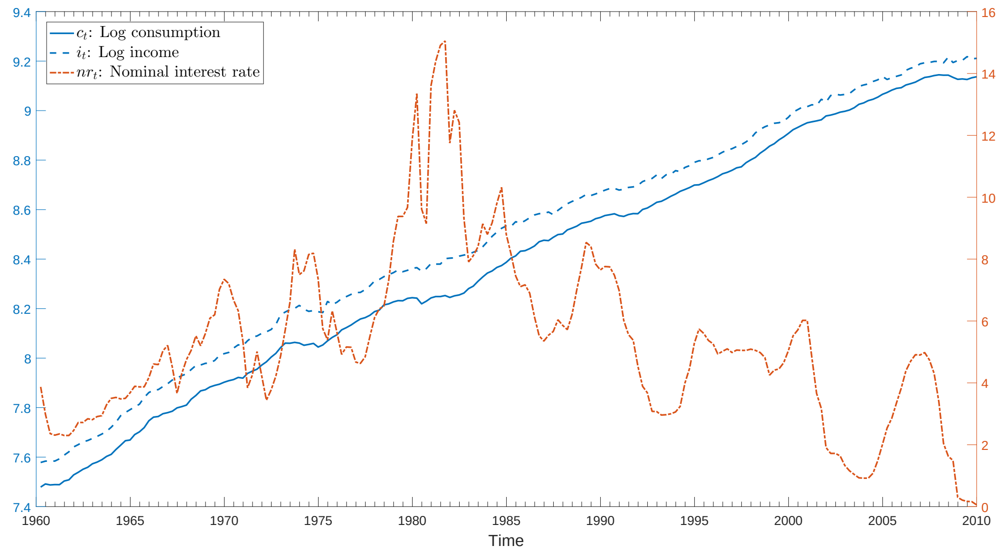

Figure 1: Aggregate US data on consumption, income and nominal interest rate over 1960–2009.

where $\overline{X}_t = \mathbf{H}'X_t = (X_{t1}, X_{t2}')'$ and $T = 200$. For given $0 < \tau < 1$, the coefficient function $\boldsymbol{\beta}(\tau) = [\beta_1(\tau), \beta_2(\tau), \beta_3(\tau)]'$ is estimated by local level regression as

$$\widehat{\boldsymbol{\beta}}(\tau) = \left[\widehat{\beta}_1(\tau), \widehat{\beta}_2(\tau), \widehat{\beta}_3(\tau)\right]' = \left[\sum_{t=1}^{T} \overline{X}_t \overline{X}_t' K\left(\frac{t - T\tau}{Th}\right)\right]^{+} \left[\sum_{t=1}^{T} \overline{X}_t Y_t K\left(\frac{t - T\tau}{Th}\right)\right], \quad (6.2)$$

where $K(u) = \frac{3}{4}(1 - u^2)I(-1 \leq u \leq 1)$ and the bandwidth h is chosen by cross-validation. The three nonparametrically estimated curves $\widehat{\beta}_i(\cdot)$ with their 95% confidence intervals are exhibited in Figures 2–4, where the confidence intervals are computed using the bootstrap approach.

For comparision, we also consider a traditional linear consumption function of the following form

$$Y_t = \boldsymbol{\gamma}'\overline{X}_t + v_t, \quad \boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)', \quad (6.3)$$

whose constant coefficients are estimated as $\widehat{\boldsymbol{\gamma}} = (-0.4099, 0.7019, -0.0065)'$. The constant coefficient specification (6.3) fails to capture any time-varying components in the coefficients, whereas plots of the fitted functions $\widehat{\beta}_1(\cdot)$, $\widehat{\beta}_2(\cdot)$ and $\widehat{\beta}_3(\cdot)$ in Figures 2–4 strongly support the presence of nonlinear functional forms for these coefficients. Based on the observed patterns of $\widehat{\beta}_j(\cdot)$, a high-order polynomial function might be a good parametric candidate for approximating the estimated time-varying coefficient functions. Accordingly, we fitted 4^{th} order polynomial functions for each of the coefficient functions.

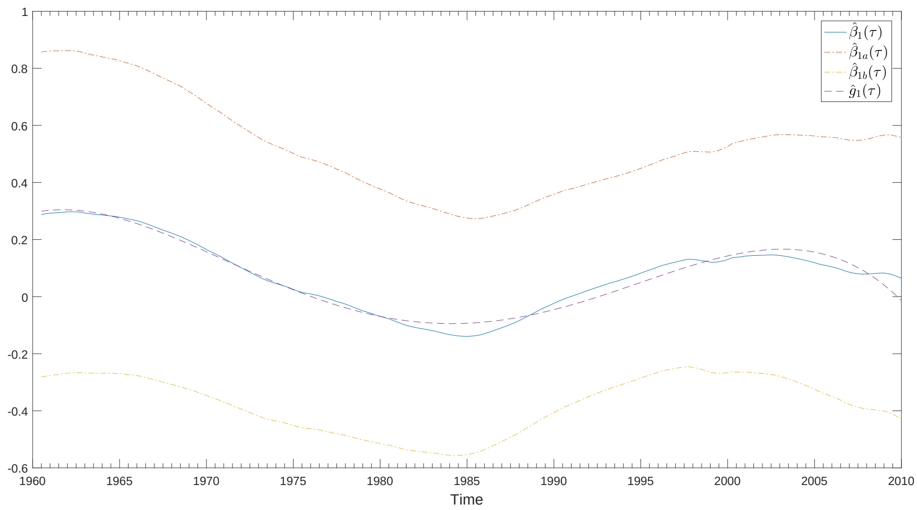

Figure 2: Nonparametric function estimate $\widehat{\beta}_1$ with confidence intervals $(\widehat{\beta}_{1a}, \widehat{\beta}_{1b})$ together with the 4^{th} order parametric polynomial $\widehat{g}_1$ estimate of β_1.

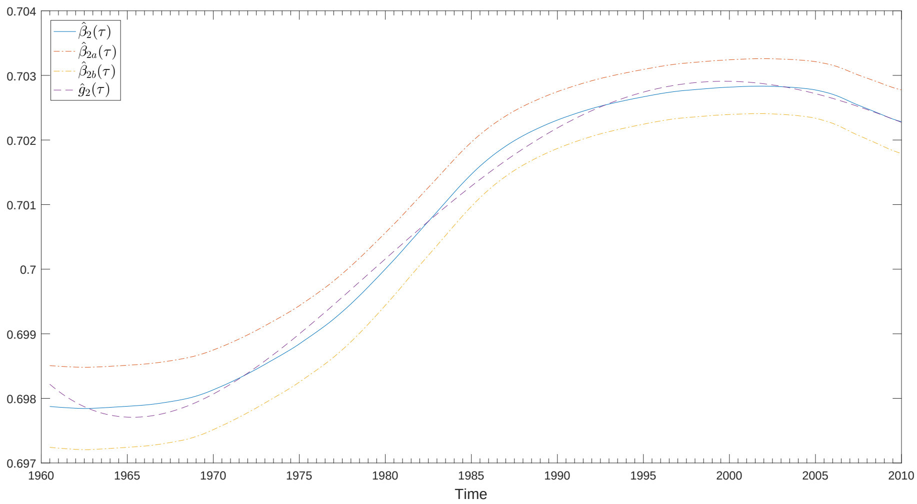

Figure 3: Nonparametric function estimate $\widehat{\beta}_2$ with confidence intervals $(\widehat{\beta}_{2a}, \widehat{\beta}_{2b})$ together with the 4^{th} order parametric polynomial $\widehat{g}_2$ estimate of β_2.

The plots of these fitted polynomial functions[2] are shown in Figures 2–4. Standard t-tests were used to select the chosen specifications of the polynomial functions and, although not detailed here, the coefficients in the selected specifications were significant with p-values close to zero. Figures 2–4 show that the nonparametric fits are well captured by the 4^{th} order parametric polynomial approximations

[2]The fitted functions are $\widehat{g}_1(\tau) = 0.2968 + 0.5368\tau - 9.8343\tau^2 + 19.8220\tau^3 - 10.8344\tau^4$, $\widehat{g}_2(\tau) = 0.6983 - 0.0120\tau + 0.0715\tau^2 - 0.0858\tau^3 + 0.0303\tau^4$, $\widehat{g}_3(\tau) = -0.0028 + 0.0150\tau - 0.0763\tau^2 + 0.0762\tau^3 - 0.0159\tau^4$.

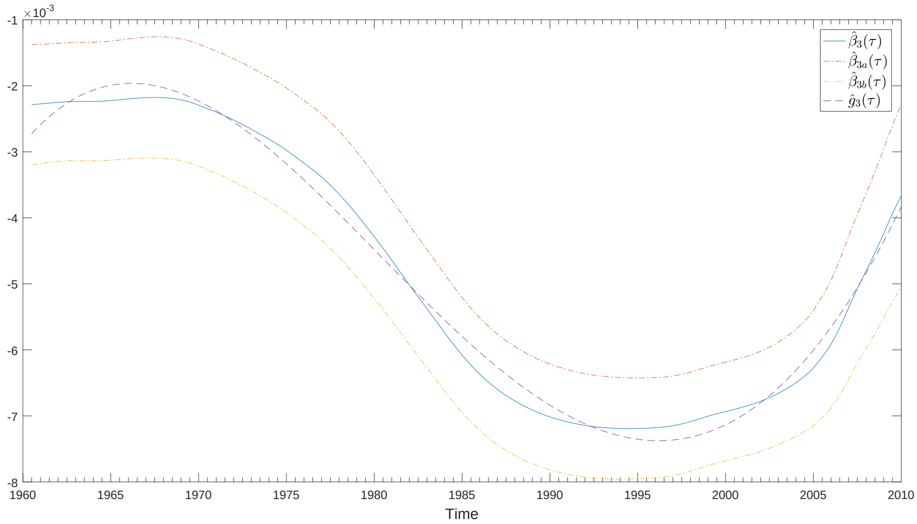

Figure 4: Nonparametric function estimate $\widehat{\beta}_3$ with confidence intervals $(\widehat{\beta}_{3a}, \widehat{\beta}_{3b})$ together with the 4^{th} order parametric polynomial $\widehat{g}_3$ estimate of β_3.

with no need for higher order specifications.

Proceeding further we analyzed residuals from the time-varying coefficient model (6.1) and the linear model (6.3), which are plotted in Figure S.1 available in Appendix C of the Online Supplementary Document. The residuals $\widehat{v}_t$ from the linear consumption function show a clear upward drift when compared with the residuals $\widehat{e}_t$ of the time-varying coefficient model. Standard residual based unit root tests[3], shown in Table 1, indicate stronger evidence for stationarity in $\widehat{e}_t$ than $\widehat{v}_t$. For example, when the PP test is applied, the null hypothesis is rejected at the 1% level for $\widehat{e}_t$ but the null fails to be rejected at the 5% level for $\widehat{v}_t$. In addition, the KPSS test suggests that $\widehat{v}_t$ may have a unit root at the 1% level. Based on these results, we conclude that $\widehat{e}_t$ is stationary, but $\widehat{v}_t$ is nonstationary, indicating that a time-varying coefficient consumption function is more appropriate in capturing cointegrating links between the variables than a linear model for consumption behavior.

In order to capture the drift presented in $\widehat{v}_t$, we fitted a fixed design nonparametric specification $\widehat{v}_t = m(t/T) + u_t$, $t = 1, \cdots, T$, to the residuals using local level kernel estimation

$$\widehat{m}(\tau) = \left(\sum_{t=1}^{T} K\Big(\frac{t - T\tau}{Th}\Big)\widehat{v}_t\right) \Big/ \left(\sum_{t=1}^{T} K\Big(\frac{t - T\tau}{Th}\Big)\right), \quad 0 < \tau < 1.$$

[3]Formal residual based unit root tests (c.f. Phillips and Ouliaris (1990)) are unavailable for specifically testing residuals from time varying coefficient cointegrating regressions and are presently under development by the authors in other work. Standard unit root tests are used here instead.

Table 1: Unit root tests for the residuals

	ADF	DF-GLS	PP	KPSS
$\widehat{e}_t$	-3.85***	-2.82***	-4.03***	0.33
$\widehat{v}_t$	-2.99**	-2.74***	-2.85*	1.36***

*, **, and *** imply rejection of the null hypothesis at 10%, 5%, and 1% level.

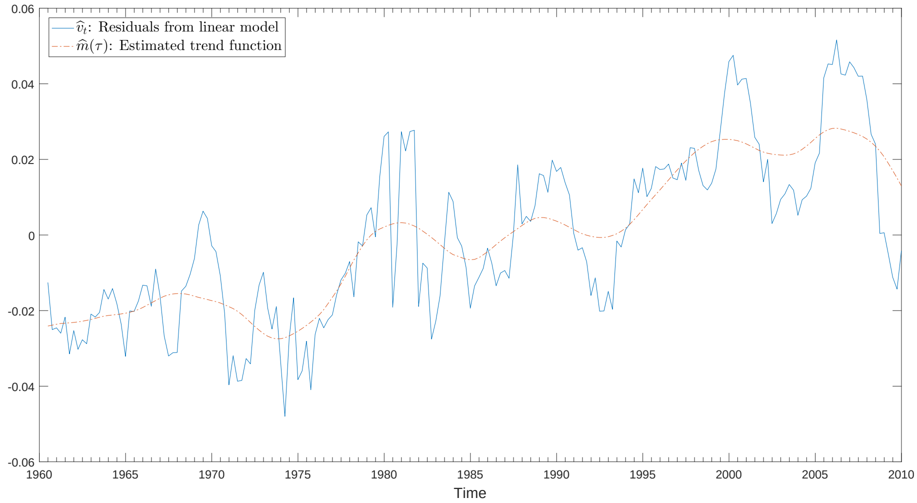

Figure 5: The estimated nonlinear trend of the residuals $\widehat{v}_t$ from the linear model

to estimate the trend function $m(\cdot)$. The estimated trend is shown in Figure 5, which is strongly indicative of a nonlinear trend in $\widehat{v}_t$. The detrended residuals $\widehat{u}_t := \widehat{v}_t - \widehat{m}(t/T)$ from this nonparametric regression are plotted in Figure S.2 (available in Appendix C of the Online Supplementary Document) against the residuals $\widehat{e}_t$ from the time varying coefficient consumption function. The close correspondence of these residuals provides further confirmation of the presence of time variation in the consumption function.

7 Conclusions

Nonparametric methods offer empirical researchers considerable flexibility in model specifications, allowing for time dependent formulations that are useful when models with constant coefficients prove inadequate. In time series regressions, this flexibility is particularly useful when series move together over time but fail cointegration tests because of evolving coefficients.

The kernel estimation approach studied in the present paper allows empirical research with time-varying coefficient cointegrating models when the regressors are multivariate and embody a mixture of stochastic and deterministic trends combined with potential co-movement among themselves. This structure is sufficiently rich to accommodate many empirical applications with co-moving nonstationary time series. Standard local level kernel regression technique forms the basis of the approach and the FM-kernel methodology extends to nonparametric regression the FM-OLS method of estimating linear cointegrating regressions with endogenous regressors and serially dependent error processes.

The methods are straightforward to implement and have the advantage that conventional limit theory can be used in a way that facilitates inference, even though the model complexities imply signal matrix degeneracies that lead to multiple convergence rates in different directions of the parameter space. In particular, the usual kernel convergence rate ($\sqrt{Th}$) applies in the stationary direction, a type 1 super-consistency rate ($T\sqrt{h}$) and a type 2 super-consistency rate (Th) apply in nonstationary directions, and a type 3 rate ($T\sqrt{Th}$) applies in the direction of the deterministic linear trends. The local and global rotation techniques used in the paper to address these challenges are a technical device only. While they produce new asymptotic theory for kernel estimation techniques that differs considerably from standard kernel limit theory, the rotation methods are not needed in empirical research with these kernel estimators or with the test statistics that are based on them.

In addition to the estimation methodology and new limit theory for time-varying parameter cointegrating regression, a generalized Wald-type statistic is introduced to provide a statistical test of whether the time-varying coefficients can be approximated by constant coefficients. That methodology also allows for testing the adequacy of specific functional forms such as polynomial time-varying parameter specifications. These specification tests enable researchers to evaluate whether greater flexibility is needed in the formulation of cointegration regression models to allow for the coefficients in these models to evolve over time. Empirical application of these methods to aggregate consumption behavior in the US is strongly indicative of the need for such flexibility.

Appendix A: Proofs of the main results

This appendix provides proofs of the main results in the paper.

Proof of Proposition 1. Define

$$\begin{aligned}
\mathbf{\Lambda}_{T11}(z_0) &= \frac{1}{Th}\sum_{t=1}^{T} X_{t1}X_{t1}'K_{th}(z_0),\\
\mathbf{\Lambda}_{T12}(z_0) &= \mathbf{\Lambda}_{T12}(z_0)' = \frac{1}{T^{3/2}h}\sum_{t=1}^{T} X_{t1}X_{t2}'q_{T2}(z_0)K_{th}(z_0),\\
\mathbf{\Lambda}_{T13}(z_0) &= \mathbf{\Lambda}_{T13}(z_0)' = \frac{1}{T^{3/2}h^{3/2}}\sum_{t=1}^{T} X_{t1}X_{t2}'q_{T2}^{\perp}(z_0)K_{th}(z_0),\\
\mathbf{\Lambda}_{T22}(z_0) &= \frac{1}{T^2h}\sum_{t=1}^{T} q_{T2}(z_0)'X_{t2}X_{t2}'q_{T2}(z_0)K_{th}(z_0),\\
\mathbf{\Lambda}_{T23}(z_0) &= \mathbf{\Lambda}_{T32}(z_0)' = \frac{1}{T^2h^{3/2}}\sum_{t=1}^{T} q_{T2}(z_0)'X_{t2}X_{t2}'q_{T2}^{\perp}(z_0)K_{th}(z_0),\\
\mathbf{\Lambda}_{T33}(z_0) &= \frac{1}{T^2h^2}\sum_{t=1}^{T} q_{T2}^{\perp}(z_0)'X_{t2}X_{t2}'q_{T2}^{\perp}(z_0)K_{th}(z_0),
\end{aligned}$$

where $K_{th}(z_0) = K\left(\frac{t-Tz_0}{Th}\right)$, $q_{T2}(z_0)$ is defined as in Section 2.2 and $q_{T2}^{\perp}(z_0)$ is the $d_1 \times (d_1-1)$ orthogonal complement of $q_{T2}(z_0)$. Observe that

$$\overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\mathbf{H}'\mathbf{\Lambda}_T(z_0)\mathbf{H}\overline{\mathbf{Q}}_T(z_0)\overline{\mathbf{D}}_T^{-1} = \begin{bmatrix} \mathbf{\Lambda}_{T11}(z_0) & \mathbf{\Lambda}_{T12}(z_0) & \mathbf{\Lambda}_{T13}(z_0) \\ \mathbf{\Lambda}_{T21}(z_0) & \mathbf{\Lambda}_{T22}(z_0) & \mathbf{\Lambda}_{T23}(z_0) \\ \mathbf{\Lambda}_{T31}(z_0) & \mathbf{\Lambda}_{T32}(z_0) & \mathbf{\Lambda}_{T33}(z_0) \end{bmatrix}. \tag{A.1}$$

We next prove that, as $T \to \infty$,

$$\mathbf{\Lambda}_{T11}(z_0) \xrightarrow{P} \mathbf{\Lambda}_{11}, \quad \mathbf{\Lambda}_{T1k}(z_0) = o_P(1), \quad \mathbf{\Lambda}_{Tk1}(z_0) = o_P(1) \tag{A.2}$$

for $k = 2, 3$, and

$$\mathbf{\Lambda}_{T2}(z_0) \Rightarrow \mathbf{\Lambda}_2(z_0), \tag{A.3}$$

where $\mathbf{\Lambda}_{11}$ and $\mathbf{\Lambda}_2(z_0)$ are defined in Section 2.2, and

$$\mathbf{\Lambda}_{T2}(z_0) = \begin{bmatrix} \mathbf{\Lambda}_{T22}(z_0) & \mathbf{\Lambda}_{T23}(z_0) \\ \mathbf{\Lambda}_{T32}(z_0) & \mathbf{\Lambda}_{T33}(z_0) \end{bmatrix}.$$

Note that $X_{t1} = \mathbf{H}_1' X_t = e_{t1}$ is a stationary linear process by Assumption 1(i), and

$$\begin{aligned}
\mathbf{\Lambda}_{T11}(z_0) &= \frac{1}{Th}\sum_{t=1}^{T} X_{t1}X_{t1}'K_{th}(z_0) = \frac{1}{Th}\sum_{t=1}^{T} e_{t1}e_{t1}'K_{th}(z_0) \\
&= \mathbb{E}\left[e_{11}e_{11}'\right]\cdot\left[\frac{1}{Th}\sum_{t=1}^{T}K_{th}(z_0)\right] + \frac{1}{Th}\sum_{t=1}^{T}\left\{e_{t1}e_{t1}' - \mathbb{E}[e_{t1}e_{t1}']\right\}K_{th}(z_0) \\
&=: \mathbf{\Lambda}_{T11}(z_0,1) + \mathbf{\Lambda}_{T11}(z_0,2). \qquad \text{(A.4)}
\end{aligned}$$

As shown in Appendix B, by applying a truncation technique for the linear process e_{t1}, we may prove that

$$\mathbf{\Lambda}_{T11}(z_0,2) = o_P(1), \qquad \text{(A.5)}$$

which implies that the leading term of $\mathbf{\Lambda}_{T11}(z_0)$ is $\mathbf{\Lambda}_{T11}(z_0,1)$. The detailed proof of (A.5) is given in Appendix B of the supplementary document. On the other hand, by Assumption 3 and some basic calculation, we have

$$\mathbf{\Lambda}_{T11}(z_0,1) \xrightarrow{P} \mu_0\mathbb{E}[e_{11}e_{11}'] = \mathbb{E}[e_{11}e_{11}'] =: \mathbf{\Lambda}_{11}. \qquad \text{(A.6)}$$

Thus, (A.5) and (A.6) lead to the first assertion in (A.2).

We next consider $\mathbf{\Lambda}_{T12}(z_0)$. Note that

$$\begin{aligned}
\mathbf{\Lambda}_{T12}(z_0) &= \frac{1}{T^{3/2}h}\sum_{t=1}^{T} X_{t1}X_{t2}'q_{T2}K_{th}(z_0) \\
&= \left[q_{T2}(z_0)'q_{T2}(z_0)\right]\cdot\left[\frac{1}{Th}\sum_{t=1}^{T}X_{t1}K_{th}(z_0)\right] + \frac{1}{T^{3/2}h}\sum_{t=1}^{T}X_{t1}\left[X_{t2} - X_{\delta(z_0)2}\right]'q_{T2}(z_0)K_{th}(z_0) \\
&=: \mathbf{\Lambda}_{T12}(z_0,1) + \mathbf{\Lambda}_{T12}(z_0,2). \qquad \text{(A.7)}
\end{aligned}$$

By the weak convergence results (2.4) and (2.5), and the standardization $|q_{T2}(z_0)'q_{T2}(z_0)| = \|q_{T2}(z_0)\| = 1$, we have

$$\frac{1}{Th}\sum_{t=1}^{T}X_{t1}\left[X_{t2} - X_{\delta(z_0)2}\right]'K_{th}(z_0) \Rightarrow 2\int_{-1}^{1}K(z)\left[dB_1(\frac{z+1}{2})\right]\left[B_2(\frac{z+1}{2})\right]' + \mathbf{\Gamma}_{12}. \qquad \text{(A.8)}$$

The proof of (A.8) is provided in Appendix B. The above results, together with the fact that

$$\frac{1}{Th}\sum_{t=1}^{T}X_{t1}K_{th}(z_0) = O_P(1/\sqrt{Th}) = o_P(1),$$

imply that

$$\left\|\mathbf{\Lambda}_{T12}(z_0)\right\| = \left\|\mathbf{\Lambda}_{T12}(z_0,1)\right\| + \left\|\mathbf{\Lambda}_{T12}(z_0,2)\right\| = O_P(1/\sqrt{Th}) + O_P(1/\sqrt{T}) = o_P(1), \tag{A.9}$$

and thus the second assertion in (A.2) holds with $k = 2$. The third assertion in (A.2) with $k = 2$ is proved in exactly the same way.

For $\mathbf{\Lambda}_{T13}(z_0)$, we observe that

$$\begin{aligned}\mathbf{\Lambda}_{T13}(z_0) &= \frac{1}{T^{3/2}h^{3/2}}\sum_{t=1}^{T} X_{t1}X_{t2}'q_{T2}^{\perp}(z_0)K_{th}(z_0) \\ &= \frac{1}{T^{3/2}h^{3/2}}\sum_{t=1}^{T} X_{t1}\left[X_{t2} - X_{\delta(z_0)2}\right]' q_{T2}^{\perp}(z_0)K_{th}(z_0) \end{aligned}\tag{A.10}$$

as $X_{\delta(z_0)2}' q_T^{\perp}(z_0) = \mathbf{0}$ by virtue of the construction of $q_{T2}^{\perp}(z_0)$ in Section 2.2. By (A.8), (A.10) and the fact that $\|q_{T2}^{\perp}(z_0)\| = O_P(1)$, we can easily prove that the second assertion in (A.2) holds with $k = 3$. Similarly, we can also prove the third assertion in (A.2) with $k = 3$. By Proposition A.1 in Phillips, Li and Gao (2017), we can prove (A.3), where $\mathbf{\Lambda}_2(z_0)$ is positive definite almost surely by virtue of Lemma A in Appendix B, thereby completing the proof of Proposition 1. □

Proof of Theorem 1. Letting

$$\overline{\mathbf{S}}_T(z_0) = \overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\left[\sum_{t=1}^{T}\overline{X}_t\overline{X}_t'K_{th}(z_0)\right]\overline{\mathbf{Q}}_T(z_0)\overline{\mathbf{D}}_T^{-1},$$

we observe that

$$\overline{\mathbf{D}}_T\overline{\mathbf{Q}}_T(z_0)'\mathbf{H}'\widehat{\boldsymbol{\beta}}(z_0) = \overline{\mathbf{S}}_T^{+}(z_0)\left[\overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\sum_{t=1}^{T}\overline{X}_tY_tK_{th}(z_0)\right]$$

as $\overline{\mathbf{D}}_T\overline{\mathbf{D}}_T^{-1} = \mathbf{I}_d$, $\overline{\mathbf{Q}}_T(z_0)'\overline{\mathbf{Q}}_T(z_0) = \mathbf{I}_d$ and $\mathbf{H}'X_t = \overline{X}_t$. Hence, we have the following decomposition:

$$\overline{\mathbf{D}}_T\overline{\mathbf{Q}}_T(z_0)'\mathbf{H}'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] = \overline{\mathbf{S}}_T^{+}(z_0)\overline{\mathbf{V}}_T(z_0) + \overline{\mathbf{S}}_T^{+}(z_0)\overline{\mathbf{B}}_T(z_0), \tag{A.11}$$

where

$$\begin{aligned}\overline{\mathbf{V}}_T(z_0) &= \overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\sum_{t=1}^{T}\overline{X}_te_{t0}K_{th}(z_0), \\ \overline{\mathbf{B}}_T(z_0) &= \overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\sum_{t=1}^{T}\overline{X}_tX_t'\left[\boldsymbol{\beta}(\frac{t}{T}) - \boldsymbol{\beta}(z_0)\right]K_{th}(z_0).\end{aligned}$$

We proceed to establish the limit distributions given in parts (i) and (ii).

(i) By Assumption 2, note that $\boldsymbol{\beta}(\frac{t}{T}) = \boldsymbol{\beta}(z_0) + O(h^\gamma)$ when $T(z_0 - h) \leq t \leq T(z_0 + h)$. Then, following the proof of Proposition 1, we can show that

$$\overline{\mathbf{S}}_T^+(z_0)\overline{\mathbf{B}}_T(z_0) = O_P(\overline{\mathbf{D}}_T h^\gamma), \tag{A.12}$$

which indicates that the first d_0 elements of $\overline{\mathbf{S}}_T^+(z_0)\overline{\mathbf{B}}_T(z_0)$ have asymptotic order of $O_P\left(\sqrt{Th}h^\gamma\right)$. This order is asymptotically negligible because of the condition $Th^{1+2\gamma} = o(1)$ in Theorem 1(i). On the other hand, note that

$$\overline{\mathbf{V}}_T(z_0) = \left[\overline{\mathbf{V}}_{T1}(z_0)', \overline{\mathbf{V}}_{T2}(z_0)'\right]' \tag{A.13}$$

with

$$\begin{aligned} \overline{\mathbf{V}}_{T1}(z_0) &= \frac{1}{\sqrt{Th}} \sum_{t=1}^T X_{t1} e_{t0} K_{th}(z_0), \\ \overline{\mathbf{V}}_{T2}(z_0) &= \overline{\mathbf{D}}_{T2}^{-1} \overline{\mathbf{Q}}_{T2}(z_0)' \sum_{t=1}^T X_{t2} e_{t0} K_{th}(z_0). \end{aligned}$$

Using the central limit theorem for the kernel-weighted sum of a locally stationary process (c.f., Zhou and Wu, 2010), we may show that

$$\overline{\mathbf{V}}_{T1}(z_0) = \frac{1}{\sqrt{Th}} \sum_{t=1}^T X_{t1} e_{t0} K_{th}(z_0) \Rightarrow \mathsf{N}\left(\mathbf{0}, \nu_0 \sum_{s=-\infty}^{\infty} \mathbb{E}\left[(e_{10} e_{s0})(e_{11} e_{s1}')\right]\right). \tag{A.14}$$

By (A.12) and (A.14), we readily have (2.12), completing the proof of part (i) of Theorem 1.

(ii) From (A.12), the last d_1 elements of $\overline{\mathbf{S}}_T^+(z_0)\overline{\mathbf{B}}_T(z_0)$ have order $O_P\left(T\sqrt{h}h^\gamma\right)$, which is asymptotically negligible because of the condition $T^2 h^{1+2\gamma} = o(1)$ in Theorem 1(ii). By (2.4), Proposition 1, and continuous mapping we have

$$\overline{\mathbf{V}}_{T2}(z_0) = \overline{\mathbf{D}}_{T2}^{-1} \overline{\mathbf{Q}}_{T2}(z_0)' \sum_{t=1}^T X_{t2} e_{t0} K_{th}(z_0) \Rightarrow \boldsymbol{\Delta}_2(z_0). \tag{A.15}$$

The nonsingularity of $\boldsymbol{\Delta}_2(z_0)$ follows from Proposition 1 and Lemma A in Appendix B, and this completes the proof of Theorem 1(ii). $\square$

Proof of Proposition 2. Observe that

$$\widetilde{\mathbf{D}}_T^{-1} \widetilde{\mathbf{Q}}_T(z_0)' \boldsymbol{\Lambda}_T(z_0) \widetilde{\mathbf{Q}}_T(z_0) \widetilde{\mathbf{D}}_T^{-1} = \begin{bmatrix} \widetilde{\boldsymbol{\Lambda}}_{T11}(z_0) & \widetilde{\boldsymbol{\Lambda}}_{T12}(z_0) \\ \widetilde{\boldsymbol{\Lambda}}_{T21}(z_0) & \widetilde{\boldsymbol{\Lambda}}_{T22}(z_0) \end{bmatrix}, \tag{A.16}$$

where

$$\begin{aligned}
\widetilde{\mathbf{\Lambda}}_{T11}(z_0) &= \frac{1}{T^3h}\sum_{t=1}^{T}\widetilde{q}_T(z_0)'X_tX_t'\widetilde{q}_T(z_0)K_{th}(z_0),\\
\widetilde{\mathbf{\Lambda}}_{T12}(z_0) &= \widetilde{\mathbf{\Lambda}}_{T21}(z_0)' = \frac{1}{T^{5/2}h^{3/2}}\sum_{t=1}^{T}\widetilde{q}_T(z_0)'X_tX_t'\widetilde{q}_T^{\perp}(z_0)K_{th}(z_0),\\
\widetilde{\mathbf{\Lambda}}_{T22}(z_0) &= \frac{1}{T^2h^2}\sum_{t=1}^{T}\widetilde{q}_T^{\perp}(z_0)'X_tX_t'\widetilde{q}_T^{\perp}(z_0)K_{th}(z_0).
\end{aligned}$$

Noting that the asymptotic leading term of $\widetilde{q}_T(z_0)'X_t$ is $(\boldsymbol{\mu}'\boldsymbol{\mu})^{1/2}\lfloor Tz_0\rfloor = \|\boldsymbol{\mu}\|\lfloor Tz_0\rfloor$ for any $T(z_0-h)\leq t\leq T(z_0+h)$, we may show that

$$\widetilde{\mathbf{\Lambda}}_{T11}(z_0) = \frac{\boldsymbol{\mu}'\boldsymbol{\mu}\cdot z_0^2}{Th}\sum_{t=1}^{T}K_{th}(z_0)+o_P(1) = \|\boldsymbol{\mu}z_0\|^2+o_P(1) =: \widetilde{\mathbf{\Lambda}}_{11}(z_0)+o_P(1). \tag{A.17}$$

By (3.2) and the construction of $\widetilde{q}_T^{\perp}(z_0)$, we have for any $T(z_0-h)\leq t\leq T(z_0+h)$,

$$\begin{aligned}
X_t'\widetilde{q}_T^{\perp}(z_0) &= X_{\delta(z_0)}'\widetilde{q}_T^{\perp}(z_0)+\left[X_t-X_{\delta(z_0)}\right]'\widetilde{q}_T^{\perp}(z_0)\\
&= \left[S_t-S_{\delta(z_0)}\right]'\widetilde{q}_T^{\perp}(z_0)+\left[D_t-D_{\delta(z_0)}\right]'\widetilde{q}_T^{\perp}(z_0)\\
&= \left[S_t-S_{\delta(z_0)}\right]'\widetilde{q}_T^{\perp}(z_0)+\frac{t-\delta(z_0)}{\delta(z_0)}\left[\boldsymbol{\mu}\cdot\delta(z_0)+S_{\delta(z_0)}+X_0\right]'\widetilde{q}_T^{\perp}(z_0)\\
&\quad -\frac{t-\delta(z_0)}{\delta(z_0)}\left[S_{\delta(z_0)}+X_0\right]'\widetilde{q}_T^{\perp}(z_0)\\
&= \left[S_t-S_{\delta(z_0)}\right]'\widetilde{q}_T^{\perp}(z_0)-\frac{t-\delta(z_0)}{\delta(z_0)}\left[S_{\delta(z_0)}+X_0\right]'\widetilde{q}_T^{\perp}(z_0)\\
&= \left[S_t-S_{\delta(z_0)}\right]'\widetilde{q}_T^{\perp}(z_0)+O_P\left(T^{1/2}h\right).
\end{aligned} \tag{A.18}$$

Then, using the fact that $\widetilde{q}_T(z_0)'X_t = \|\boldsymbol{\mu}z_0\|T(1+o_P(1))$ for any $T(z_0-h)\leq t\leq T(z_0+h)$, we may show that

$$\widetilde{\mathbf{\Lambda}}_{T12}(z_0) = \frac{\|\boldsymbol{\mu}z_0\|}{T^{3/2}h^{3/2}}\sum_{t=1}^{T}\left[S_t-S_{\delta(z_0)}\right]'\widetilde{q}_T^{\perp}(z_0)K_{th}(z_0)+O_P\left(h^{1/2}\right). \tag{A.19}$$

Similarly, using (A.18), we can further prove that

$$\widetilde{\mathbf{\Lambda}}_{T22}(z_0) = \frac{1}{T^2h^2}\sum_{t=1}^{T}\widetilde{q}_T^{\perp}(z_0)'\left[S_t-S_{\delta(z_0)}\right]\left[S_t-S_{\delta(z_0)}\right]'\widetilde{q}_T^{\perp}(z_0)K_{th}(z_0). \tag{A.20}$$

Then, by (A.14), (A.19), (A.20), the weak convergence result (2.4), the continuous mapping theorem and the fact that $\widetilde{q}_T(z_0) = \boldsymbol{\mu}/\|\boldsymbol{\mu}\| + o_P(1)$, we may complete the proof of Proposition 2. Positive definiteness of the limit matrix $\widetilde{\boldsymbol{\Lambda}}(z_0)$ follows as in Lemma A in Appendix B and Proposition 1. □

Proof of Theorem 2. Define

$$\begin{aligned}
\widetilde{\mathbf{S}}_T(z_0) &= \widetilde{\mathbf{D}}_T^{-1}\widetilde{\mathbf{Q}}_T(z_0)'\left[\sum_{t=1}^T X_tX_t'K_{th}(z_0)\right]\widetilde{\mathbf{Q}}_T(z_0)\widetilde{\mathbf{D}}_T^{-1},\\
\widetilde{\mathbf{V}}_T(z_0) &= \widetilde{\mathbf{D}}_T^{-1}\widetilde{\mathbf{Q}}_T(z_0)'\sum_{t=1}^T X_te_{t0}K_{th}(z_0),\\
\widetilde{\mathbf{B}}_T(z_0) &= \widetilde{\mathbf{D}}_T^{-1}\widetilde{\mathbf{Q}}_T(z_0)'\sum_{t=1}^T X_tX_t'\left[\boldsymbol{\beta}(\frac{t}{T}) - \boldsymbol{\beta}(z_0)\right]K_{th}(z_0).
\end{aligned}$$

Note that

$$\widetilde{\mathbf{D}}_T\widetilde{\mathbf{Q}}_T(z_0)'\left[\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right] = \widetilde{\mathbf{S}}_T^{+}(z_0)\widetilde{\mathbf{V}}_T(z_0) + \widetilde{\mathbf{S}}_T^{+}(z_0)\widetilde{\mathbf{B}}_T(z_0). \tag{A.21}$$

Similar to the proof of Theorem 1, we may show that the asymptotic bias term has the order of $O_P(\widetilde{\mathbf{D}}_T h^\gamma)$. Hence, we need only derive the limiting distribution theory for the first term on the right hand side of (A.21).

Note that

$$\widetilde{\mathbf{V}}_T(z_0) = \left[\widetilde{\mathbf{V}}_{T1}(z_0), \widetilde{\mathbf{V}}_{T2}(z_0)'\right]', \tag{A.22}$$

where

$$\begin{aligned}
\widetilde{\mathbf{V}}_{T1}(z_0) &= \frac{1}{T\sqrt{Th}}\sum_{t=1}^T \widetilde{q}_T(z_0)'X_te_{t0}K_{th}(z_0),\\
\widetilde{\mathbf{V}}_{T2}(z_0) &= \frac{1}{Th}\sum_{t=1}^T \widetilde{q}_T^{\perp}(z_0)'X_te_{t0}K_{th}(z_0).
\end{aligned}$$

Following the argument in the proof of Proposition 2, we may show that

$$\widetilde{\mathbf{V}}_{T1}(z_0) = \|\boldsymbol{\mu} z_0\|\frac{1}{\sqrt{Th}}\sum_{t=1}^T e_{t0}K_{th}(z_0) + o_P(1) \tag{A.23}$$

and

$$\widetilde{\mathbf{V}}_{T2}(z_0) = \frac{1}{Th}\sum_{t=1}^T \widetilde{q}_T^{\perp}(z_0)'\left[S_t - S_{\delta(z_0)}\right]e_{t0}K_{th}(z_0) + o_P(1). \tag{A.24}$$

Then, by (2.4), (A.23), (A.24), Proposition 2, and the continuous mapping theorem, we have

$$\widetilde{\mathbf{V}}_T(z_0) \Rightarrow \widetilde{\boldsymbol{\Delta}}(z_0), \tag{A.25}$$

where $\widetilde{\mathbf{\Delta}}(z_0)$ is defined as in Section 3.1, thereby completing the proof of Theorem 2. □

Proof of Proposition 3. The proof is a combination of the relevant arguments in the proofs of Propositions 1 and 2 above. Details are omitted to save space. □

Proof of Theorem 3. By using Proposition 3, the proof is similar to the relevant arguments in the proofs of Theorems 1 and 2 above. Details are omitted to save space. □

Proof of Theorem 4. By (4.9), we may argue that $\mathbf{H}_2'\widehat{\boldsymbol{\beta}}_{\#}(z_0)$ is asymptotically equivalent to $\mathbf{H}_2'\widehat{\boldsymbol{\beta}}_{\star}(z_0)$, i.e.,

$$\overline{\mathbf{D}}_{T2}\overline{\mathbf{Q}}_{T2}(z_0)'\mathbf{H}_2'\left[\widehat{\boldsymbol{\beta}}_{\#}(z_0)-\widehat{\boldsymbol{\beta}}_{\star}(z_0)\right]=o_P(1),$$

which indicates that we need only show

$$\overline{\mathbf{D}}_{T2}\overline{\mathbf{Q}}_{T2}(z_0)'\mathbf{H}_2'\left[\widehat{\boldsymbol{\beta}}_{\star}(z_0)-\boldsymbol{\beta}(z_0)\right]\Rightarrow\mathbf{\Lambda}_2^{-1}(z_0)\mathbf{\Delta}_2^{\#}(z_0). \tag{A.26}$$

From the definition of $\mathbf{B}_{T2}^{\star}(z_0)$, we have

$$\overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\mathbf{H}'\mathbf{B}_{T2}^{\star}(z_0)=\left\{\mathbf{0}_{d_0+1}',\left[q_{T2}^{\perp}(z_0)'\left(\mathbf{\Gamma}_{20}-\mathbf{\Gamma}_{22}\mathbf{\Omega}_{22}^{+}\mathbf{\Omega}_{20}\right)\right]'\right\}'. \tag{A.27}$$

On the other hand, by standard arguments, we may show that

$$\overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_0)'\mathbf{H}'\left[\sum_{t=1}^{T}X_te_{t0}K_{th}(z_0)-\mathbf{B}_{T1}^{\star}(z_0)\right]=\left[\mathbf{0}_{d_0}',\overline{\mathbf{V}}_{T2}^{\star}(z_0)'\right]' \tag{A.28}$$

with

$$\overline{\mathbf{V}}_{T2}^{\star}(z_0)=\overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_{T2}(z_0)'\sum_{t=1}^{T}X_{t2}e_{t0}^{\star}K_{th}(z_0),\quad e_{t0}^{\star}=e_{t0}-\mathbf{\Omega}_{02}\mathbf{\Omega}_{22}^{+}e_{t2}.$$

Combining (A.27) and (A.28), using the continuous mapping theorem and the arguments in the proof of Theorem 1, we can prove (A.26), completing the proof of Theorem 4. □

Proof of Theorem 5. By (5.2) and (5.3), we have

$$\widehat{\sigma}_e^2\mathbf{R}\left[\frac{1}{Th}\sum_{t=1}^{T}X_tX_t'K(\frac{t-Tz_k}{Th})\right]^{+}\mathbf{R}'=\sigma_e^2\mathbf{R}\mathbf{H}_1\mathbf{\Lambda}_{11}^{+}\mathbf{H}_1'\mathbf{R}'+o_P(1) \tag{A.29}$$

over $z_1,\cdots,z_m$. In view of (5.5), the limit matrix in (A.29) is of rank r and does not rely on z_k. By (5.1) and (5.6), and noting that $Th^{1+2\gamma}=o(1)$, it follows immediately that

$$\begin{aligned}\sqrt{Th}\mathbf{R}\left[\widehat{\boldsymbol{\beta}}(z_k)-\widehat{\boldsymbol{\beta}}\right] &= \sqrt{Th}\mathbf{R}\mathbf{H}_1\mathbf{H}_1'\left[\widehat{\boldsymbol{\beta}}(z_k)-\boldsymbol{\beta}_0\right]+o_P(1)\\ &\Rightarrow \mathbf{R}\mathbf{H}_1\boldsymbol{\xi}=_d\mathsf{N}\left(\mathbf{0}_r,\nu_0\sigma_e^2\mathbf{R}\mathbf{H}_1\mathbf{\Lambda}_{11}^{+}\mathbf{H}_1'\mathbf{R}'\right)\end{aligned} \tag{A.30}$$

under $\mathcal{H}_0$, $k = 1, \cdots, m$. By (A.29) and (A.30), we may further show that under $\mathcal{H}_0$,

$$W_T(z_k) \Rightarrow \chi_r^2(k), \quad k = 1, \cdots, m, \tag{A.31}$$

i.e., the point-wise Wald test statistic $W_T(z_k)$ defined in (5.4) is central chi-square with r degrees of freedom in the limit. As $z_1 < z_2 < \cdots < z_m$ are an equidistant grid points in $(0, 1)$ and m is a fixed positive integer, we must have $z_k - z_{k-1} > 2h$, which together with the arguments in the proof of Theorem 1, indicates that the components $\{W_T(z_k) : k = 1, \cdots, m\}$ are asymptotically independent. This fact, together with (A.31), leads to the conclusion that the limit distributions $\chi_r^2(1), \cdots, \chi_r^2(m)$ are independent, and consequently

$$W_T \Rightarrow \chi_{mr}^2 \tag{A.32}$$

under the null hypothesis $\mathcal{H}_0$.

Note that the FM-kernel estimator defined in (4.8) only makes the bias corrections in the direction $\mathbf{H}_2$, which ensures that the asymptotic distributions given in (5.1) and (A.30) continue to hold for the FM-kernel estimator $\widehat{\boldsymbol{\beta}}_{\#}(\cdot)$. As $W_T^{\#}$ is constructed in the same manner as W_T but using FM-kernel estimates of the time-varying coefficients, the same arguments can show that the limit distribution in (A.32) continues to apply, which completes the proof of Theorem 5. □

Proof of Theorem 6. We first analyze the asymptotic form of the matrix

$$\mathbf{R}(z_k)\left[\frac{1}{Th}\sum_{t=1}^{T} X_t X_t' K\left(\frac{t - Tz_k}{Th}\right)\right]^{+}\mathbf{R}(z_k)'$$

to extract the limit distribution for $W_T^{\diamond}(z_k)$. From (5.11), we may rewrite $\mathbf{R}(z)'$ as

$$\mathbf{R}(z)' = \left[\mathbf{H}_1\mathbf{S}_* + (\mathbf{O}_{d\times r_0}, \mathbf{H}_2\overline{\mathbf{Q}}_{T2}(z)\mathbf{S}_{h2}) \ \vdots\ \mathbf{H}_2\overline{\mathbf{Q}}_{T2}(z)\mathbf{S}_2\right], \tag{A.33}$$

where $\mathbf{S}_* = (\mathbf{S}_1, \mathbf{S}_{h1})$. As in the proof of Theorem 4.5 in Phillips (1995), we may neglect the component submatrix $(\mathbf{O}_{d\times r_0}, \mathbf{H}_2\overline{\mathbf{Q}}_{T2}(z)\mathbf{S}_{h2})$ in the following analysis because its associated term is of negligible asymptotic order. Then, by (A.33), we have

$$\mathbf{R}(z_k)\left[\frac{1}{Th}\sum_{t=1}^{T} X_t X_t' K\left(\frac{t - Tz_k}{Th}\right)\right]^{+}\mathbf{R}(z_k)'$$
$$= \begin{pmatrix} \mathbf{S}_*'\mathbf{H}_1'\left[\frac{1}{Th}\sum_{t=1}^{T} X_t X_t' K\left(\frac{t-Tz_k}{Th}\right)\right]^{+}\mathbf{H}_1\mathbf{S}_* & \mathbf{S}_*'\mathbf{H}_1'\left[\frac{1}{Th}\sum_{t=1}^{T} X_t X_t' K\left(\frac{t-Tz_k}{Th}\right)\right]^{+}\mathbf{H}_2\mathbf{S}_2 \\ \mathbf{S}_2'\mathbf{H}_2'\left[\frac{1}{Th}\sum_{t=1}^{T} X_t X_t' K\left(\frac{t-Tz_k}{Th}\right)\right]^{+}\mathbf{H}_1\mathbf{S}_* & \mathbf{S}_2'\mathbf{H}_2'\left[\frac{1}{Th}\sum_{t=1}^{T} X_t X_t' K\left(\frac{t-Tz_k}{Th}\right)\right]^{+}\mathbf{H}_2\mathbf{S}_2 \end{pmatrix}.$$

Noting that

$$\left[\frac{1}{Th}\sum_{t=1}^{T}X_tX_t'K(\frac{t-Tz_k}{Th})\right]^{+}=Th\mathbf{H}\overline{\mathbf{Q}}_T(z_k)\overline{\mathbf{D}}_T^{-1}\left[\overline{\mathbf{D}}_T^{-1}\overline{\mathbf{Q}}_T(z_k)'\mathbf{H}'\mathbf{\Lambda}_T(z_k)\mathbf{H}\overline{\mathbf{Q}}_T(z_k)\overline{\mathbf{D}}_T^{-1}\right]^{+}\mathbf{D}_T^{-1}\overline{\mathbf{Q}}_T(z_k)'\mathbf{H}',$$

and letting $\overline{\mathbf{D}}_T^{\diamond}=\mathsf{diag}\left\{\mathbf{I}_{d_0},\sqrt{T},(\sqrt{Th})\mathbf{I}_{d_1-1}\right\}$, as in the proof of Proposition 1, we have

$$\overline{\mathbf{D}}_T^{\diamond}\mathbf{R}(z_k)\left[\frac{1}{Th}\sum_{t=1}^{T}X_tX_t'K(\frac{t-Tz_k}{Th})\right]^{+}\mathbf{R}(z_k)'\overline{\mathbf{D}}_T^{\diamond}=\begin{pmatrix}\mathbf{S}_*'\mathbf{\Lambda}_{11}^{+}\mathbf{S}_* & \mathbf{O}_{r_1\times r_2}\\ \mathbf{O}_{r_2\times r_1} & \mathbf{S}_2'\mathbf{\Lambda}_{T2}^{+}(z_k)\mathbf{S}_2\end{pmatrix}, \tag{A.34}$$

where the definition of $\mathbf{\Lambda}_{T2}(\cdot)$ can be found in the proof of Proposition 1. Lemma B.3 in Phillips, Li and Gao (2017) further shows that $\mathbf{\Lambda}_{T2}(z)$ is asymptotically non-singular for any $z\in[h,1-h]$.

On the other hand, by (5.15) and using Assumption 1* and Theorem 4, we have

$$\sqrt{Th}\overline{\mathbf{D}}_T^{\diamond}\mathbf{R}(z_k)\left[\widehat{\boldsymbol{\beta}}_{\#}(z_k)-\widehat{\boldsymbol{\beta}}\right]\Rightarrow\boldsymbol{\xi}_{\#}(z_k)=\left[\boldsymbol{\xi}_1^{\#}(z_k)',\boldsymbol{\xi}_2^{\#}(z_k)'\right]' \tag{A.35}$$

under the null hypothesis $\mathcal{H}_0^{\diamond}$, where

$$\begin{aligned}\boldsymbol{\xi}_1^{\#}(z_k) &=_d \mathsf{N}\left(\mathbf{0}_{r_1},\sigma_e^2\mathbf{S}_*'\mathbf{\Lambda}_{11}^{+}\mathbf{S}_*\right),\\ \left[\mathbf{S}_2'\mathbf{\Lambda}_{T2}^{+}(z_k)\mathbf{S}_2\right]^{-1/2}\boldsymbol{\xi}_2^{\#}(z_k) &\Rightarrow \mathsf{N}\left(\mathbf{0}_{r_2},\omega_{0|2}\mathbf{I}_{r_2}\right),\quad \omega_{0|2}=\omega-\mathbf{\Omega}_{02}\mathbf{\Omega}_{22}^{+}\mathbf{\Omega}_{20},\end{aligned}$$

and $\boldsymbol{\xi}_1^{\#}(z_k)$ is independent of $\boldsymbol{\xi}_2^{\#}(z_k)$ according to Remark 2(b). Hence, using (A.34) and (A.35), we may show that

$$W_T^{\#}(z_k)\Rightarrow\chi_{r_1}^2+\frac{\omega_{0|2}}{\sigma_e^2}\chi_{r_2,*}^2, \tag{A.36}$$

where $\chi_{r_1}^2$ and $\chi_{r_2,*}^2$ are two independent chi-square distributions with degrees of freedom r_1 and r_2, respectively. Furthermore, as the limit variates $\left\{\boldsymbol{\xi}_{\#}(z_k):k=1,\cdots,m\right\}$, are mutually independent when $z_k-z_{k-1}\geq 2h$, we finally obtain the following limit distribution in (5.16). □

References

Baek, Y., Cho, J. and Phillips, P. C. B. (2015). Testing linearity using power transforms of regressors. *Journal of Econometrics* **187**, 376–384.

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

Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* **16**, 367–401.

Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* **80**, 1157–1183.

Cheng, X. and Phillips, P. C. B. (2009). Semiparametric cointegrating rank selection. *Econometrics Journal* **12**, 83–104.

Engle, R. and Granger, C. W. J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* **55**, 251–276.

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

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

Giraitis, L., Kapetanios, G. and Yates, T. (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics* **179**, 46–65.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59**, 1551–1580.

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

Li, D., Phillips, P. C. B. and Gao, J. (2016). Uniform consistency of nonstationary kernel-weighted sample covariances for nonparametric regression. *Econometric Theory* **32**, 355–385.

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

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

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

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

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

Phillips, P. C. B. (1988). Multiple regression with integrated time series. *Contemporary Mathematics* **80**, 79–105.

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. (1996). Econometric model determination. *Econometrica* **64**, 763–812.

Phillips, P. C. B. (2007). Regression with Slowly Varying Regressors and Nonlinear Trends. *Econometric Theory* **23**, 557–614.

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

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

Phillips, P. C. B., Li, D. and Gao, J. (2017). Estimating smooth structural changes in cointegration models. *Journal of Econometrics* **196**, 180–195.

Phillips, P. C. B. and Ouliaris S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica* **58**, 165–193.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* **75**, 335–346.

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

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

Toda, H. Y. and Phillips, P. C. B. (1993). Vector autoregressions and causality. *Econometrica* **61**, 1367–1393.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *Annals of Statistics* **40**, 2601–2633.

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

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

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

Zhang, T. and Wu, W. B. (2012). Inference of time varying regression models. *Annals of Statistics* **40**, 1376–1402.

Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society Series B* **72**, 513–531.