Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: http://www.tandfonline.com/loi/lecr20

Functional-coefficient cointegration models in the presence of deterministic trends

Masayuki Hirukawa & Mari Sakudo

To cite this article: Masayuki Hirukawa & Mari Sakudo (2015): Functional-coefficient cointegration models in the presence of deterministic trends, Econometric Reviews

To link to this article: http://dx.doi.org/10.1080/07474938.2015.1092845

Accepted author version posted online: 21 Oct 2015.
Published online: 21 Oct 2015.

ECONOMETRIC REVIEWS
http://dx.doi.org/10.1080/07474938.2015.1092845

Taylor & Francis
Taylor & Francis Group

Functional-coefficient cointegration models in the presence of deterministic trends

Masayuki Hirukawa[a] and Mari Sakudo[b,c,d]

[a]Faculty of Economics, Setsunan University, Neyagawa, Osaka, Japan; [b]Development Bank of Japan, Tokyo, Japan; [c]Waseda University, Tokyo, Japan; [d]Japan Economic Research Institute, Tokyo, Japan

ABSTRACT
In this article, we extend the functional-coefficient cointegration model (FCCM) to the cases in which nonstationary regressors contain both stochastic and deterministic trends. A nondegenerate distributional theory on the local linear (LL) regression smoother of the FCCM is explored. It is demonstrated that even when integrated regressors are endogenous, the limiting distribution is the same as if they were exogenous. Finite-sample performance of the LL estimator is investigated via Monte Carlo simulations in comparison with an alternative estimation method. As an application of the FCCM, electricity demand analysis in Illinois is considered.

KEYWORDS
Cointegration; deterministic trend; endogenous regressor; kernel smoothing; local linear regression smoothing; piecewise local linear regression principle

JEL CLASSIFICATION
C13; C14; C22

1. Introduction

Since the seminal work by Engle and Granger (1987), cointegration models have provided an appealing framework for characterizing long-term equilibrium relationships among economic variables. However, empirical works often report weak evidence of cointegration. To account for this phenomenon, many authors have made a variety of attempts to patch up cointegrating regressions. One such attempt is to model the exact quantitative relationships among economic variables as gradually varying, rather than constant, over a long time horizon. In particular, Xiao (2009) incorporated a time-varying nature into the cointegrating regression model by assuming that the cointegrating vector is an unknown smooth function of another stationary variable. As a result, the model can be viewed as a variant of functional-coefficient models or varying-coefficient models (VCMs). In this sense, Xiao's (2009) model is called the functional-coefficient cointegration model (FCCM).

This article extends Xiao's (2009) FCCM to the cases in which nonstationary regressors contain both stochastic and deterministic trends, and establishes asymptotic theories on estimation and inference in this class of FCCMs. Extending the FCCM in this direction is useful for the following reasons. First, many macroeconomic variables that are commonly described as I(1) (e.g., income, output, consumption, price level, and money stock) are actually best regarded as "I(1) with drift" (West, 1988; Hansen, 1992a,b) or thought to be generated by "random walks" using innovations with nonzero means (Granger, 2012). Second, Hansen (1992a) demonstrates that when nonstationary regressors that enter constant-coefficient cointegration models appear to possess both stochastic and deterministic trends, it is better, from the viewpoint of precision in estimation, to estimate the models without detrending the regressors. There are also empirical applications in which cointegrating regressions are estimated without detrending: examples include Engel's law (Ogaki, 1992), money demand (Stock and Watson, 1993), and intertemporal elasticity of substitution of nondurable consumption (Ogaki and Park, 1997), to name a few.

Because the FCCM studied in this article can be viewed as a varying-coefficient version of (or a "time-varying" analog to) Hansen's (1992a,b) models, we naturally attempt to establish distributional theories

CONTACT Masayuki Hirukawa hirukawa@econ.setsunan.ac.jp Faculty of Economics, Setsunan University, 17-8 Ikeda Nakamachi, Neyagawa, Osaka 572-8508, Japan.

on estimation and inference of the model without detrending. This article adopts kernel smoothing to estimate functional-coefficients in the FCCM consistently. Particular attention is paid to local linear (LL) regression smoothing. As is the case with linear regressions with deterministic trends, the LL estimator has a degenerate joint limiting distribution due to multiple convergence rates of coefficient estimates. Then, a nondegenerate distributional theory for LL estimation is explored. The nondegenerate limiting distribution is shown to be mixed-normal. It is worth emphasizing that the limit theory remains unchanged, regardless of whether regressors are exogenous or endogenous. In other words, the second-order effect does not arise, unlike the least-squares estimation for constant-coefficient cointegrating regressions with endogenous regressors.

We also demonstrate that the convergence rate of the estimator of the cointegrating vector depends crucially on the model specification. For instance, when the limit process of regressors contains at least one stochastic trend, LL estimators of the cointegrating vector and coefficients on deterministic trends have the same $Th^{1/2}$ convergence rate, where T and h are the sample size and bandwidth, respectively. The super-consistent nonparametric rate has been already uncovered in the literature; see Juhl (2005), Cai et al. (2009) and Xiao (2009). In contrast, when the limit process consists only of deterministic components, the convergence rate of the LL estimator for the cointegrating vector attains no slower than $T^{3/2}h^{1/2}$. To implement LL estimation, inspired by Ruppert et al. (1995), we propose a solve-the-equation plug-in bandwidth choice rule. This type of implementation method is developed for the first time in the literature on VCMs, to the best of our knowledge.

This article contributes to the literature on VCMs in two respects. First, while recently research directions on VCMs have shifted toward those with nonstationary regressors, the assumption that the regressors contain only stochastic trends has been maintained so far; see Juhl (2005), Cai et al. (2009) and Xiao (2009), for instance. This article can be classified as a complement to these earlier works in the sense that it studies estimation and inference of VCMs including I(1) regressors with trends. Second, several authors investigate consistency of local constant (LC) estimation for VCMs that includes deterministic trends as regressors. For example, Liang and Li (2012) demonstrate that LC estimation for VCMs with stationary regressors and a linear trend is inconsistent, and Li and Li (2013) establish that LC estimation for VCMs with unit-root nonstationary regressors and a linear trend turns out to be consistent. It can be shown that LC estimation of the FCCM considered in this article is still consistent.

The remainder of the article is organized as follows. Section 2 describes the FCCM that can incorporate deterministic trends. In Section 3, a nondegenerate distributional theory on LL estimation is developed. Section 4 delivers component-wise convergence rates of the LL estimator. Based on the results, a plug-in bandwidth choice method is proposed. Section 5 studies hypothesis testing. Particular focuses are on testing the null of constant coefficients and that of no trends in the cointegrating regression. In Section 6, finite-sample performance of the LL estimator is examined via Monte Carlo simulations. Section 7 applies the FCCM for electricity demand analysis in the State of Illinois. Section 8 concludes. Appendices provide all proofs and a brief description of the piecewise local linear regression (PLLR) principle, which is proposed by Banerjee and Pitarakis (2012, 2014) as an alternative to kernel estimation. Additionally, an online supplement that summarizes simulation results is made available on the first author's webpage.

This article adopts the following notational conventions: the symbol "$\Rightarrow$" signifies weak convergence; "$\overset{d}{=}$" is equality in distribution; $\lfloor \cdot \rfloor$ denotes the integer part; $BM(\Omega)$ is the vector Brownian motion with covariance matrix Ω; $\otimes$ is used to represent the tensor (or Kronecker) product; $\mathbf{1}(\cdot)$ is an indicator function; $\mu_{ij}(f)$ denotes the integral $\mu_{ij}(f) = \int t^i f^j(u)\,du$; and $g^{(i)}(x)$ is the ith-order derivative of $g(x)$, i.e., $g^{(i)}(x) = d^i g(x)/dx^i$. For two random variables X and Y, "$X \perp\!\!\!\perp Y$" reads that X is stochastically independent of Y. The L^p-norm for a random matrix $X = (x_{ij})$ is defined as $\|X\|_p = E\left(\sum_{i,j}|x_{ij}|^p\right)^{1/p}$, where $\|X\| \equiv \|X\|_2$ for notational simplicity. Lastly, the expression '$X_T \sim Y_T$' is used whenever $X_T/Y_T \to 1$ as $T \to \infty$.

2. The FCCM when regressors are I(1) with trends

The FCCM considered in this article largely follows the one in Hansen (1992b). We shall be working on a (d_2+2)-dimensional time series $\left(y_t, x_{2t}', z_t\right)' \in \mathbb{R} \times \mathbb{R}^{d_2} \times \mathbb{R}$, where the scalar variable z_t is assumed to be stationary. Let the random variable y_t be generated by a cointegrating regression

$$y_t = x_t'\beta\left(z_t\right) + u_{1t} := x_{1t}'\beta_1\left(z_t\right) + x_{2t}'\beta_2\left(z_t\right) + u_{1t}, \tag{1}$$

where the regressor $x_t = \left(x_{1t}', x_{2t}'\right)'$ is determined by

$$\begin{aligned} x_{1t} &= k_{1t}, \\ x_{2t} &= \Pi_1 k_{1t} + \Pi_2 k_{2t} + S_{2t}, \\ \Delta S_{2t} &= u_{2t}, \end{aligned}$$

and $u_t := \left(u_{1t}, u_{2t}'\right)' \in \mathbb{R} \times \mathbb{R}^{d_2}$ is a zero-mean stationary process, the statistical property of which is described in Assumption 1 below. The system is initialized at time 0, and the initial value y_0 may be any random number.

We now present the definition of $k_t = \left(k_{1t}', k_{2t}'\right)'$. It is an m-dimensional vector of powers of time index t, and it can be further partitioned as

$$k_t := \left(t^{p_1}, \ldots, t^{p_m}\right)' = \left(\left(t^{p_1}, \ldots, t^{p_{m_1}}\right)', \left(t^{p_{(m_1+1)}}, \ldots, t^{p_m}\right)'\right)' := \left(k_{1t}', k_{2t}'\right)'$$

so that $\dim\left(k_{1t}\right) = m_1$ and $\dim\left(k_{2t}\right) = m - m_1 := m_2$. The exponents p_j, $j = 1, \ldots, m$ are assumed to be *known* integers that satisfy $0 \leq p_1 < \cdots < p_m$. Observe that when $p_1 = 0$, the levels regression (1) contains an intercept term. We assume that whenever k_t contains a constant, it is an element of k_{1t} and thus enters (1). In addition, the vectors of "trends" k_{1t} and k_{2t} (here we loosely speak of an intercept as a trend term) can be viewed as those of included and excluded trends, respectively, in the sense that while the former directly enters the cointegrating regression, the latter governs the behavior of the integrated regressor x_{2t} but is not included in the regression. Furthermore, following Hansen (1992a), we can consider two special cases of the regression model (1), namely, the unrestricted FCCM (FCCM-U) and the restricted FCCM (FCCM-R). These can be specified by setting $m_2 = 0$ (i.e., FCCM with all trends included) and $m_1 = 0$ (i.e., FCCM with all trends excluded), respectively.

For notational convenience, we write $d_1 = m_1$ and $d = d_1 + d_2$ hereinafter. We may use d_1 and m_1 interchangeably from statement to statement. Accordingly, the coefficient matrices in x_{2t} are $\Pi_1 \in \mathbb{R}^{d_2 \times d_1}$ and $\Pi_2 \in \mathbb{R}^{d_2 \times m_2}$.

The FCCM (1) is motivated because in many macroeconomic applications the integrated regressor x_{2t} can be described more suitably as "I(1) with drift", as stated in West (1988) and Hansen (1992a,b). Even if there is no deterministic trend in the system, it is a common practice to include an intercept term in the cointegrating regression. In most applications, m tends to be small. Typical choices of k_t would therefore include 1, t, $(1, t)$, $\left(t, t^2\right)$, and $\left(1, t, t^2\right)$.

A notable difference of our system from the one in Hansen (1992b) is that while the integrated regressor x_{2t} includes deterministic trends, the cointegrating vector in our system is not constant but a functional controlled by a stationary variable z_t. Observe that z_t plays a similar role to the threshold variable in threshold autoregressive (TAR) models or the transition variable in smooth transition autoregressive (STAR) models. From the viewpoint that the coefficient vector $\beta\left(z_t\right)$ is assumed to vary smoothly with z_t as in STAR, rather than in an abrupt manner as in TAR, we may also refer to z_t as the transition variable. Following the convention in the TAR and STAR literature and avoiding the curse of dimensionality, we assume that the transition variable is a scalar. An extension to a vector z_t is straightforward, but for simplicity we do not pursue this. In the closely related literature, Cai et al. (2009) and Xiao (2009) consider a similar system to ours. However, in each of these two articles integrated regressors are assumed to be free of deterministic trends (i.e., $d_1 = 0$ and $\Pi_1 = \Pi_2 = 0$ are maintained).

3. Estimation theory

3.1. LL estimation

To estimate the functional-coefficient $\beta(\cdot)$ in (1) consistently, we adopt kernel smoothing. In particular, LL estimation is known to possess appealing properties such as high statistical efficiency in an asymptotic minimax sense, design-adaptation, and automatic boundary correction. For a given design point z, T observations $\left\{\left(y_t, x_{2t}', z_t\right)'\right\}_{t=1}^T$, the kernel function $K(\cdot)$, and the bandwidth parameter h, the LL estimator of $\beta(z)$ is defined as $\hat{\beta}(z)$ in

$$\begin{aligned}
&\begin{bmatrix} \hat{\beta}(z) \\ \hat{\beta}^{(1)}(z) \end{bmatrix} \\
&= \arg\min_{(\theta_0,\theta_1)} \sum_{t=1}^T \left\{y_t - \theta_0' x_t - (z_t - z)\,\theta_1' x_t\right\}^2 K\left(\frac{z_t - z}{h}\right) \\
&= \left\{\sum_{t=1}^T \begin{bmatrix} x_t \\ (z_t - z)\, x_t \end{bmatrix} \begin{bmatrix} x_t' & (z_t - z)\, x_t' \end{bmatrix} K\left(\frac{z_t - z}{h}\right)\right\}^{-1} \sum_{t=1}^T \begin{bmatrix} x_t \\ (z_t - z)\, x_t \end{bmatrix} y_t K\left(\frac{z_t - z}{h}\right).
\end{aligned}$$

In fact, the LL estimator $\hat{\beta}(z)$ admits the concise expression

$$\hat{\beta}(z) = \left\{S_0(z) - S_1(z)\, S_2(z)^{-1} S_1(z)\right\}^{-1} \left\{T_0(z) - S_1(z)\, S_2(z)^{-1} T_1(z)\right\}, \tag{2}$$

where

$$S_i(z) := \sum_{t=1}^T x_t x_t' (z_t - z)^i K\left(\frac{z_t - z}{h}\right) \quad \text{and} \quad T_i(z) := \sum_{t=1}^T x_t y_t (z_t - z)^i K\left(\frac{z_t - z}{h}\right)$$

for $i \geq 0$.

3.2. Regularity conditions

To describe the convergence properties on $\hat{\beta}(z)$, we make the following assumptions.

Assumption 1. The random sequence $v_t := \left(u_t', z_t\right)'$ is a strictly stationary, strong mixing process with the mixing coefficient $\alpha(k)$ of size $-\delta\gamma/(\delta - \gamma)$ and $\|v_t\|_\delta < \infty$ for some $\delta > \gamma > 2$. Also, u_t satisfies $E(u_t) = 0$, and its long-run variance $\Omega = \sum_{j=-\infty}^{\infty} E\left(u_t u_{t-j}'\right) > 0$. Moreover, v_t satisfies one of the following: (i) $z_t (:= \varsigma_{t-j}$ for some $j \geq 1)$ is predetermined, u_{1t} is independent and identically distributed (*iid*), and $u_{1t} \perp\!\!\!\perp \mathcal{F}_{t-1}$, where $\mathcal{F}_t = \sigma(v_t, v_{t-1}, \ldots)$ is the smallest σ-field containing current and past $v_t := \left(u_t', \varsigma_t\right)'$; (ii) $u_{1t} \perp\!\!\!\perp \left(u_{2s}', z_s\right)', \forall t, s$; or (iii) $z_t \perp\!\!\!\perp u_s, \forall t, s$.

Assumption 2. The kernel function $K(\cdot)$ is a symmetric, continuous probability density function with support $[-1, 1]$.

Assumption 3. The nonnegative sequence of bandwidth $h = h_T$ satisfies $h \to 0$ and $Th \to \infty$ as $T \to \infty$.

Assumption 4. $f(z)$ (the marginal density of z_t) and $f_s(z_0, z_s)$ (the joint density of (z_0, z_s) for $s \geq 1$) satisfy $\sup_z f(z) < \infty$ and $\sup_{z_0, z_s, s} f_s(z_0, z_s) < \infty$. In addition, $f(z)$ is continuously differentiable with a uniformly bounded derivative, and $f(z) > 0$ for a given design point z.

Assumption 5. $\beta(z)$ is twice continuously differentiable for all $z \in \mathbb{R}$.

Assumption 6. Each column of Π_1 contains a nonzero element, and rank $(\Pi_2) = m_2 \leq d_2$.

Assumption 1 is very similar to the one in the literature on unit-root TAR models (e.g., Caner and Hansen, 2001, Assumption 2) or threshold cointegration models (e.g., Gonzalo and Pitarakis, 2006, Assumptions A1–A3 and A5). It follows from this assumption that the elements of S_{2t} are not mutually cointegrated. The mixing condition is the same as the one in Assumption 1 of Hansen (1992c), which in turn establishes the following functional central limit theorem (FCLT) for the partial sum process $S_t = \left(S_{1t}, S_{2t}'\right)' := \sum_{s=1}^{t} u_s$

$$\frac{1}{\sqrt{T}} S_{\lfloor Tr \rfloor} = \frac{1}{\sqrt{T}} \left[\begin{array}{c} S_{1\lfloor Tr \rfloor} \\ S_{2\lfloor Tr \rfloor} \end{array} \right] \Rightarrow \left[\begin{array}{c} B_{S_1}(r) \\ B_{S_2}(r) \end{array} \right] := B_S(r) \overset{d}{=} BM(\Omega), \quad r \in [0,1],$$

where Ω can be partitioned as

$$\Omega = \left[\begin{array}{cc} \omega_{11} & \Omega_{21}' \\ \Omega_{21} & \Omega_{22} \end{array} \right].$$

Furthermore, condition (i) in Assumption 1 is commonly imposed in the aforementioned articles as well as in Juhl (2005). Condition (ii) is taken from Xiao (2009, Assumption A). These conditions allow for standard choices of the transition variable in STAR such as $z_t = x_{2,t-1,l} - x_{2,t-1-j,l}$ or $z_t = \Delta x_{2,t-j,l}$ for some $j \geq 1$, where $x_{2,t,l}$, the lth element of x_{2t}, is assumed to have up to a linear trend. Also note that while $\Omega_{21} = 0$ under condition (ii), conditions (i) and (iii) each allow for endogeneity in the cointegrating regression (1) so that $\Omega_{21} \neq 0$ in general. However, unlike least-squares estimation of constant-coefficient cointegrating regressions, local averaging does not cause the so-called second-order effect; see Remark 1 below for discussion. In their functional-coefficient model with integrated regressors, Sun et al. (2011) do not even assume the independence of z_t and u_s, and derive a nonstandard convergence result. Because our aim is to develop a mixed-normal limit theorem for standard inference, the independence assumption is maintained throughout.

Assumptions 2–5 are standard in the literature on kernel regression. The assumption of the compact support in Assumption 2 is made solely for brevity of the exposition. It can be relaxed to allow for kernels with support on the entire real line, at the expense of lengthier proofs. Lastly, the condition on Π_1 in Assumption 6 restates that the integrated regressor x_{2t} contains a full set of included trends k_{1t}. While Theorem 1 below holds even when $\Pi_1 = 0$ and nonetheless the regression contains deterministic trends, the condition plays a key role in determining the convergence rate of each component of $\hat{\beta}(z)$; see Section 4 for details. The rank condition in Assumption 6 is also required for the asymptotic result, as in Hansen (1992b). The assumption also implies that the number of excluded trends is small relative to the number of integrated regressors.

3.3. A nondegenerate distributional theory

To deliver the distributional theory on $\hat{\beta}(z)$, we need to derive the multivariate invariance principle including the kernel-weighted partial sum process. Following Xiao (2009), define

$$\underline{K}_t(z) := K\left(\frac{z_t - z}{h}\right) - E\left\{K\left(\frac{z_t - z}{h}\right)\right\}.$$

For $\underline{K}_t(\cdot)$, let the kernel-weighted partial sum process be $U_t(z) := \sum_{s=1}^{t} \underline{K}_s(z) u_{1s}$. Then, we have the following multivariate invariance principle for $\left(S_t', U_t(z)\right)'$.

Lemma 1. *If Assumptions 1–4 hold, then*

$$\left[\begin{array}{c} \frac{1}{\sqrt{T}} S_{\lfloor Tr \rfloor} \\ \frac{1}{\sqrt{Th}} U_{\lfloor Tr \rfloor}(z) \end{array} \right] \Rightarrow \left[\begin{array}{c} B_S(r) \\ B_{U(z)}(r) \end{array} \right] \overset{d}{=} BM\left\{\Omega_U(z)\right\}, \quad r \in [0,1],$$

where $\left(B_S(r)', B_{U(z)}(r)\right)'$ *is a* (d_2+2)*-dimensional Brownian motion with covariance matrix*

$$\Omega_U(z) = \begin{bmatrix} \Omega & 0 \\ 0 & \mu_{02}(K)\,\sigma_{11} f(z) \end{bmatrix},$$

and $\sigma_{11} = E\left(u_{1t}^2\right)$.

It is well known that when the regressors have deterministic trends in a constant-coefficient cointegrating regression, or when we run a linear regression with deterministic trends as a part of regressors, the joint limiting distribution of the least-squares estimators becomes degenerate. Our FCCM is not free of this issue, either. Therefore, we adopt a key trick employed in Hansen (1992a,b) to establish a nondegenerate asymptotic result in LL estimation. Let $D_T := \text{diag}\left\{T^{p_1}, \ldots, T^{p_m}\right\}$. Then, as shown in Hansen (1992a, p. 90),

$$D_T^{-1} k_{\lfloor Tr \rfloor} \to k(r) = \left(r^{p_1}, \ldots, r^{p_m}\right)' \tag{3}$$

uniformly over $r \in [0,1]$, where $0^0 \equiv 1$ by convention. Again, for notational convenience, D_T and $k(r)$ are partitioned as $D_T := \text{diag}\{D_{1T}, D_{2T}\} = \text{diag}\left\{\text{diag}\left\{T^{p_1}, \ldots, T^{p_{m_1}}\right\}, \text{diag}\left\{T^{p_{(m_1+1)}}, \ldots, T^{p_m}\right\}\right\}$, and $k(r) := \left(k_1(r)', k_2(r)'\right)' = \left(\left(r^{p_1}, \ldots, r^{p_{m_1}}\right)', \left(r^{p_{(m_1+1)}}, \ldots, r^{p_m}\right)'\right)'$.

We consider the linear transformation of the partial sum process $x_{\lfloor Tr \rfloor}$ using a $d \times d$ standardizing matrix

$$\Gamma_T = \begin{bmatrix} D_{1T} & 0 \\ \Pi_1 D_{1T} & \Gamma_{2T}^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} D_{1T}^{-1} & 0 \\ -\Gamma_{2T}\Pi_1 & \Gamma_{2T} \end{bmatrix},$$

where

$$\Gamma_{2T} = \begin{bmatrix} D_{2T}^{-1}\left(\Pi_2'\Pi_2\right)^{-1}\Pi_2' \\ T^{-1/2}\left(\Pi_2^{*\prime}\Omega_{22}\Pi_2^*\right)^{-1/2}\Pi_2^{*\prime} \end{bmatrix},$$

and the matrix $\Pi_2^* \in \mathbb{R}^{d_2 \times (d_2 - m_2)}$ spans the null space of Π_2. Observe that Π_2^* annihilates k_{2t}, the vector of excluded deterministic trends, from x_{2t}. Then, it is straightforward to see that

$$\Gamma_T x_{\lfloor Tr \rfloor} \Rightarrow J(r) := \begin{bmatrix} k_1(r) \\ k_2(r) \\ W_{d_2-m_2}(r) \end{bmatrix}, \quad r \in [0,1], \tag{4}$$

where $W_{d_2-m_2}(r) := \left(\Pi_2^{*\prime}\Omega_{22}\Pi_2^*\right)^{-1/2}\Pi_2^{*\prime}B_2(r) \overset{d}{=} BM\left(I_{d_2-m_2}\right)$.

Additionally, by Lemma A2 of Phillips and Hansen (1990), $J(r)$ is a full-ranked process in the sense that $\int_0^1 J(r)J(r)'\,dr > 0$ almost surely. This fact and the weak convergence result (4) jointly demonstrate that Γ_T is an appropriate weighting matrix for x_t. We can also see from (4) that the d-dimensional process x_t is asymptotically dominated by the d_1-dimensional trend process $k_1(r)$, the m_2-dimensional trend process $k_2(r)$, and the $(d_2 - m_2)$-dimensional stochastic trend $W_{d_2-m_2}(r)$.

We now present the following nondegenerate distributional theory on the LL estimator of $\beta(\cdot)$ for a given design point z.

Theorem 1. *If Assumptions 1–6 hold, then*

$$\sqrt{Th}\,\Gamma_T'^{-1}\left\{\hat{\beta}(z) - \beta(z) - \frac{1}{2}\mu_{21}(K)\,\beta^{(2)}(z)\,h^2 + o_p\left(h^2\right)\right\} \Rightarrow MN\left(0, \Sigma(z)\right), \tag{5}$$

where

$$\Sigma(z) := \frac{\mu_{02}(K)\,\sigma_{11}}{f(z)}\left\{\int_0^1 J(r)\,J(r)'\,dr\right\}^{-1}.$$

Remark 1. The limiting distribution remains unchanged, regardless of whether integrated regressors are exogenous or endogenous in (1). The invariance of the limit theory in the presence of endogeneity (condition (iii) in Assumption 1) is attributed to the fact that smoothing is made on the range of the weakly dependent process z_t. Weak dependence implies that when we pick the observations z_t that are close to the design point z and take a local average over the range, the selected observations are not necessarily close to each other in time and thus likely to behave as if they were independent. As a result, together with the independence between z_t and the error process u_t, the second-order effect does not arise even when the regressors are endogenous. Accordingly, there is no need for a second-order correction in the estimation of our FCCM, unlike the least-squares estimation for constant-coefficient cointegrating regressions with endogenous regressors.

In the related literature, Sun et al. (2011) allow z_t and u_t to be correlated, at the expense of nonstandard asymptotic results. Our aim is to demonstrate that under some regularity conditions, it is possible to obtain a normal limit theorem without a second-order bias correction even when integrated regressors are endogenous. Furthermore, it is not hard to see that replacing LL with LC estimation in our FCCM does not lose consistency, unlike the results in Liang and Li (2012), who study VCMs with stationary regressors and a linear trend. Extra variability of I(1) components in nonstationary regressors restores consistency, as explained in Li and Li (2013), who establish consistency for LC estimation of VCMs with unit-root nonstationary regressors and a linear trend.

Remark 2. We also find similarity of Theorem 1 to Theorem 1 of Hansen (1992a). For a fair comparison, we concentrate on the case in which all the regressors in the constant-coefficient cointegrating regression are exogenous. Theorem 1 of Hansen (1992a) demonstrates that the least-squares estimators of unrestricted and restricted regressions, after being standardized suitably in the same manner as (4), are asymptotically mixed-normal with covariance matrices $\omega_{11}\left\{\int_0^1 J_U(r)\,J_U(r)'\,dr\right\}^{-1}$ and $\omega_{11}\left\{\int_0^1 J_R(r)\,J_R(r)'\,dr\right\}^{-1}$, respectively, for some processes $J_U(r)$ and $J_R(r)$ that contain both deterministic and stochastic trends. Intuitively, because of kernel smoothing, our result can be obtained by slowing down the expansion rate from $\sqrt{T}$ in Theorem 1 of Hansen (1992a) to $\sqrt{Th}$, including the $O(h^2)$ leading bias term, and replacing ω_{11} (the long-run variance of u_{1t}) with $\{\mu_{02}(K)/f(z)\}$-times its instantaneous variance σ_{11}.

4. Component-wise convergence results

4.1. Bias-variance trade-off

Theorem 1 resolves the issue of degeneration in the limiting distribution due to multiple convergence rates of the components of the LL estimator $\hat{\beta}(z)$ by considering the limiting behavior of linear combinations of the components. However, it is often desirable to know the convergence rate of each component, especially for implementation purposes of the nonparametric estimators, as described shortly.

Here we consider the mean integrated squared error (MISE) of $\hat{\beta}(z)$. Because $\hat{\beta}(z)$ is multidimensional in general, for a d-dimensional symmetric positive definite weighting matrix $W(z)$, the MISE of

$\hat{\beta}(z)$ is defined as

$$MISE\left\{\hat{\beta}(z)\right\} = \int E\left[\left\{\hat{\beta}(z) - \beta(z)\right\}' W(z)\left\{\hat{\beta}(z) - \beta(z)\right\}\right] dz. \tag{6}$$

For simplicity, we choose $W(z) = \psi(z) I_d$ for some nonnegative scalar weighting function $\psi(\cdot)$. Then, (6) reduces to

$$\begin{aligned} MISE\left\{\hat{\beta}(z)\right\} &= \int \left\|\hat{\beta}(z) - \beta(z)\right\|^2 \psi(z)\, dz \\ &= \int \left\|Bias\left\{\hat{\beta}(z)\right\}\right\|^2 \psi(z)\, dz + \int \operatorname{tr}\left[Var\left\{\hat{\beta}(z)\right\}\right] \psi(z)\, dz, \end{aligned} \tag{7}$$

where $\psi(\cdot)$ is assumed to ensure finiteness of integrals.

The integrated squared bias (first) and integrated variance (second) terms in (7) can now be approximated. It immediately follows from Theorem 1 that the integrated squared bias term can be approximated by

$$\int \left\|Bias\left\{\hat{\beta}(z)\right\}\right\|^2 \psi(z)\, dz \sim \frac{h^4}{4}\{\mu_{21}(K)\}^2 \int \left\|\beta^{(2)}(z)\right\|^2 \psi(z)\, dz = O\left(h^4\right). \tag{8}$$

To approximate the integrated variance term, we need to examine orders of magnitude in the diagonal elements of $Var\left\{\hat{\beta}(z)\right\} \sim (Th)^{-1}\Gamma_T' \Sigma(z)\Gamma_T$. A simple calculation yields the order of magnitude in the variance of each component of the LL estimator $\hat{\beta}(\cdot)$.

Table 1 displays that the order of magnitude in each variance term depends on whether $d_2 > m_2$ (i.e., at least one stochastic trend remains in the limit process $J(r)$) or $d_2 = m_2$ (i.e., $J(r)$ consists only of deterministic components). In particular, the expression O_p (rather than O) when $d_2 > m_2$ reflects that because $\int_0^1 J(r) J(r)' dr$ is stochastic, $\Sigma(z) = O_p(1)$. In contrast, when $d_2 = m_2$, deterministic trends asymptotically dominate in the transformed process $\Gamma_T x_t$, and as a consequence, $\Sigma(z) = O(1)$. In addition, multiple convergence rates appear in general. A single convergence rate applies to the estimator of each coefficient only if (1) satisfies either $d_2 > m_2, p_1 > 0$ (no intercept) or $d_2 = m_2, m_1 = 0$ (a special case of the FCCM-R). Table 1 also reports that when $d_2 > m_2$, the LL estimator of the cointegrating vector, as well as the one of the coefficient on each trend, attains a super-consistent nonparametric rate of $Th^{1/2}$. This convergence rate has been already uncovered in the related literature; see Juhl (2005), Cai et al. (2009), and Xiao (2009), for instance. In contrast, when $d_2 = m_2$, the LL estimator of the cointegrating vector reaches a further super-consistent nonparametric rate of $T^{p_{(m_1+1)}+1/2}h^{1/2}$, which depends on the slowest element of the excluded trend vector k_{2t} and is no slower than $T^{3/2}h^{1/2}$.

Based on (8) and Table 1, Table 2 reports the MISE-optimal bandwidth and optimal MISE of each component of $\hat{\beta}(\cdot)$. The implementation method for LL estimation below is built on this result. As suggested there, LL estimation in general requires multi-step (or component-wise) smoothing with different bandwidths employed for different components.

Table 1. Order of magnitude in $Var\left\{\hat{\beta}_k(\cdot)\right\}, k = 1, \ldots, d$.

Regressor	$d_2 > m_2$	$d_2 = m_2$
Intercept (if any)	$O_p\left\{(Th)^{-1}\right\}$	$O\left\{(Th)^{-1}\right\}$
Trend (if any)	$O_p\left\{\left(T^2h\right)^{-1}\right\}$	$O\left\{\left(T^{2p_k+1}h\right)^{-1}\right\}$, $k = 1, \ldots, m_1$
Integrated Regressor	$O_p\left\{\left(T^2h\right)^{-1}\right\}$	$O\left\{\left(T^{2p_{(m_1+1)}+1}h\right)^{-1}\right\}$

Table 2. MISE-optimal bandwidth and optimal MISE of $\hat{\beta}_k(\cdot)$, $k = 1, \ldots, d$.

Regressor	$d_2 > m_2$		$d_2 = m_2$	
	h^*	$MISE^*$	h^*	$MISE^*$
Intercept (if any)	$O_p\left(T^{-1/5}\right)$	$O_p\left(T^{-4/5}\right)$	$O\left(T^{-1/5}\right)$	$O\left(T^{-4/5}\right)$
Trend (if any)	$O_p\left(T^{-2/5}\right)$	$O_p\left(T^{-8/5}\right)$	$O\left\{T^{-(2p_k+1)/5}\right\}$ for $k = 1, \ldots, m_1$	$O\left\{T^{-4(2p_k+1)/5}\right\}$
Integrated Regressor	$O_p\left(T^{-2/5}\right)$	$O_p\left(T^{-8/5}\right)$	$O\left\{T^{-\left(2p_{(m_1+1)}+1\right)/5}\right\}$	$O\left\{T^{-4\left(2p_{(m_1+1)}+1\right)/5}\right\}$

4.2. An expositional example: West's (1988) regression with a single "random walk with drift" regressor

Before introducing an implementation method, it is worth illustrating how component-wise convergence results differ depending on model specifications. West (1988) investigates the linear regression without detrending when a single regressor obeys a random walk with drift. In our context, an equivalent model can be obtained by setting $m = 2$, $m_1 = 1$, $d_2 = m_2 = 1$, and $\left(p_1, p_2\right) = (0, 1)$ in the levels regression (1) so that

$$y_t = \beta_0(z_t) + \beta_2(z_t)\, x_{2t} + u_{1t}, \tag{9}$$

where $x_{1t} = k_{1t} = 1$, $x_{2t} = \pi_0 + \pi_1 t + S_{2t}$ $(\pi_0, \pi_1 \neq 0)$, and $\Delta S_{2t} = u_{2t}$. Note that (9) is not the FCCM-R in that an intercept term is included in the regression. Now, by Theorem 1,

$$\sqrt{Th}\begin{bmatrix} 1 & \pi_0 \\ 0 & \pi_1 T \end{bmatrix}\left\{\hat{\beta}(z) - \beta(z) - \frac{1}{2}\mu_{21}(K)\,\beta^{(2)}(z)\,h^2 + o_p\left(h^2\right)\right\} \Rightarrow N\left(0, \Sigma(z)\right),$$

where $\beta(z) = \left(\beta_0(z), \beta_2(z)\right)'$ and

$$\Sigma(z) = \frac{\mu_{02}(K)\,\sigma_{11}}{f(z)}\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}^{-1} = \frac{\mu_{02}(K)\,\sigma_{11}}{f(z)}\begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix}.$$

Note that the limiting distribution is normal, not mixed-normal, because the asymptotic variance is free of random components. The intercept estimator $\hat{\beta}_0(z)$ has a usual nonparametric convergence rate of $\sqrt{Th}$, whereas the estimator of the cointegrating vector $\hat{\beta}_2(z)$ has the following asymptotic distribution:

$$T^{3/2}h^{1/2}\left\{\hat{\beta}_2(z) - \beta_2(z) - \frac{1}{2}\mu_{21}(K)\,\beta_2^{(2)}(z)\,h^2 + o_p\left(h^2\right)\right\} \Rightarrow N\left(0, \frac{12\mu_{02}(K)\,\sigma_{11}}{\pi_1^2 f(z)}\right).$$

Observe that the convergence rate of $\hat{\beta}_2(z)$ is unusually rapid $T^{3/2}h^{1/2}$, which is analogous to the result in West (1988).

Including a linear trend t in (9), we can reformulate the regression to the FCCM-U

$$y_t = \beta_0(z_t) + \beta_1(z_t)\, t + \beta_2(z_t)\, x_{2t} + u_{1t}. \tag{10}$$

Because $m_1 = 2$, $m_2 = 0$, and $d_2 = 1$ in this model, it follows from Theorem 1 that

$$\sqrt{Th}\begin{bmatrix} 1 & 0 & \pi_0 \\ 0 & T & \pi_1 T \\ 0 & 0 & \Omega_{22}^{1/2}\sqrt{T} \end{bmatrix}\left\{\hat{\beta}(z) - \beta(z) - \frac{1}{2}\mu_{21}(K)\,\beta^{(2)}(z)\,h^2 + o_p\left(h^2\right)\right\} \Rightarrow MN\left(0, \Sigma(z)\right),$$

where $\beta(z) = \left(\beta_0(z), \beta_1(z), \beta_2(z)\right)'$ and

$$\Sigma(z) = \frac{\mu_{02}(K)\,\sigma_{11}}{f(z)}\begin{bmatrix} 1 & 1/2 & \int_0^1 W(r)\,dr \\ 1/2 & 1/3 & \int_0^1 rW(r)\,dr \\ \int_0^1 W(r)\,dr & \int_0^1 rW(r)\,dr & \int_0^1 W(r)^2\,dr \end{bmatrix}^{-1}.$$

The result implies that while $\hat{\beta}_0(z)$ is asymptotically $\sqrt{Th}$-mixed normal, both $\hat{\beta}_1(z)$ and $\hat{\beta}_2(z)$ are asymptotically $T\sqrt{h}$-mixed normal. In particular, the asymptotic distribution of $\hat{\beta}_2(z)$ reduces to

$$Th^{1/2}\left\{\hat{\beta}_2(z)-\beta_2(z)-\frac{1}{2}\mu_{21}(K)\beta_2^{(2)}(z)h^2+o_p\left(h^2\right)\right\}\Rightarrow MN\left(0,\frac{\mu_{02}(K)\sigma_{11}}{\Omega_{22}f(z)}\left\{\int_0^1 W^\tau(r)^2\,dr\right\}^{-1}\right),$$

where $W^\tau(r):=W(r)-(4-6r)\int_0^1 W(s)\,ds-(12r-6)\int_0^1 sW(s)\,ds$ is the demeaned and detrended Brownian motion.

4.3. A solve-the-equation plug-in bandwidth choice method

The bandwidth choice is always an important practical question in kernel smoothing. Before proceeding, it is worth noting that except for the cases with $d_2>m_2, p_1>0$, and $d_2=m_2, m_1=0$, we must make a multistep LL smoothing for a full estimation of the levels regression (1). For this purpose, we could extend a two-step smoothing as in Cai et al. (2009). In the first step, we should compute the LL estimator with the fastest convergence rate. Table 2 suggests that the estimator of the cointegrating vector attains the fastest rate, regardless of whether $d_2>m_2$ or $d_2=m_2$. Once the LL estimate of the cointegrating vector is obtained, we subtract the estimated part from y_t and then compute the LL estimator with the second fastest convergence rate using the residual, and so on. A detailed discussion is found in Cai et al. (2009, Section 2.4).

The plug-in approach considered here closely follows the idea in Ruppert et al. (1995, Section 5). Our bandwidth choice rule can be viewed as a variant of the solve-the-equation plug-in (SP) approach, which is classified as a second-generation bandwidth selector in Jones et al. (1996). The SP rule originates from kernel density estimation (Park and Marron, 1990; Sheather and Jones, 1991), and it is also applied to the long-run variance estimation (Hirukawa, 2010). A noticeable difference in our approach is that while a nonparametric estimate of the second-order derivative of the quantity of interest (either density or conditional expectation) plays a key role in the aforementioned articles, our approach fits a polynomial model to the derivative, as explained shortly. Furthermore, the SP rule is readily applicable to the implementation of Xiao's (2009) FCCM with a minor modification.

The SP rule is built on the MISE-optimal bandwidth for the LL estimator of interest. Because the same convergence rates of the estimators of the coefficients on trends and the cointegrating vector apply when $d_2>m_2$, we can obtain these LL estimators using a single bandwidth. In light of this, define the selector matrix S as

$$S:=\begin{cases}\begin{bmatrix}0_{(d-1)\times 1} & I_{d-1}\end{bmatrix}' & \text{if } d_2>m_2,\ p_1=0\\ I_d & \text{if } d_2>m_2,\ p_1>0\\ \begin{bmatrix}0_{d_2\times d_1} & I_{d_2}\end{bmatrix}' & \text{if } d_2=m_2\end{cases}.$$

For this S and a suitably chosen exponent α, it can be shown that as $T\to\infty$, $T^\alpha\Gamma_T S$ converges to a nonzero constant matrix S^* (say), as in Lemma 2 in Section 5. Specifically, put

$$\alpha=\begin{cases}1/2 & \text{if } d_2>m_2\\ p_{(m_1+1)} & \text{if } d_2=m_2\end{cases},$$

and write

$$\Sigma_{S'\beta}=S^{*\prime}\left\{\int_0^1 J(r)J(r)'\,dr\right\}^{-1}S^*.$$

The next proposition delivers an approximation to the MISE of the LL estimator for selected coefficients $S'\hat{\beta}(z)$ and the MISE-optimal bandwidth. This is a direct outcome from Theorem 1, (7), (8), and the results in Table 2.

Proposition 1. *If Assumptions 1–6 hold, then MISE* $\left\{S'\hat{\beta}(z)\right\}$ *can be approximated by*

$$MISE\left\{S'\hat{\beta}(z)\right\} = \frac{h^4}{4}\{\mu_{21}(K)\}^2 \int \left\|S'\beta^{(2)}(z)\right\|^2 \psi(z)\,dz + \frac{\mu_{02}(K)\,\sigma_{11}}{T^{2\alpha+1}h}\int \operatorname{tr}\left(\Sigma_{S'\beta}\right)\frac{\psi(z)}{f(z)}dz + o\left(h^4 + \frac{1}{T^{2\alpha+1}h}\right),$$

The MISE-optimal bandwidth is

$$h^* = \left[\frac{\mu_{02}(K)\,\sigma_{11}\int \operatorname{tr}\left(\Sigma_{S'\beta}\right)\left\{\psi(z)/f(z)\right\}dz}{\{\mu_{21}(K)\}^2\int \left\|S'\beta^{(2)}(z)\right\|^2\psi(z)\,dz}\right]^{1/5} T^{-(2\alpha+1)/5}. \tag{11}$$

The right-hand side of (11) contains three unknown quantities, namely, σ_{11}, $\int \left\|S'\beta^{(2)}(z)\right\|^2 \psi(z)\,dz$ and $\int \operatorname{tr}\left(\Sigma_{S'\beta}\right)\left\{\psi(z)/f(z)\right\}dz$. We "estimate" (or find proxies of) these quantities in the following manner.

On σ_{11}. Suppose that h^* is known. Then, we have the functional-coefficient estimates $\left\{\hat{\beta}(z_t;h^*)\right\}_{t=1}^{T}$, where $\hat{\beta}(\cdot;h)$ signifies the LL estimate of $\beta(\cdot)$ using the bandwidth h. A natural estimator of σ_{11} is $\hat{\sigma}_{11}^* = \hat{\sigma}_{11}(h^*)$, where

$$\hat{\sigma}_{11} = \hat{\sigma}_{11}(h) := \frac{1}{T}\sum_{t=1}^{T}\{\hat{u}_{1t}(h)\}^2 - \left\{\overline{\hat{u}_1(h)}\right\}^2, \tag{12}$$

$$\hat{u}_{1t}(h) = y_t - x_t'\hat{\beta}(z_t;h), \text{ and}$$

$$\overline{\hat{u}_1(h)} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{1t}(h).$$

On $\int \left\|S'\beta^{(2)}(z)\right\|^2 \psi(z)\,dz$. Following Ruppert et al. (1995, p. 1259), put $\psi(z) = f(z)\,\mathbf{1}(z \in [a,b])$ for a prespecified compact interval $[a,b]$ $(-\infty < a < b < \infty)$. Then, a natural estimator of $\int \left\|S'\beta^{(2)}(z)\right\|^2 \mathbf{1}(z \in [a,b])\,f(z)\,dz$ is

$$\sum_i \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}_t\left(\hat{\beta}_{it}^{*(2)}\right)^2 = \begin{cases} \sum_{i=2}^{d}\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}_t\left(\hat{\beta}_{it}^{*(2)}\right)^2 & \text{if } d_2 > m_2, p_1 = 0 \\ \sum_{i=1}^{d}\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}_t\left(\hat{\beta}_{it}^{*(2)}\right)^2 & \text{if } d_2 > m_2, p_1 > 0 \\ \sum_{i=d_1+1}^{d}\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}_t\left(\hat{\beta}_{it}^{*(2)}\right)^2 & \text{if } d_2 = m_2 \end{cases}, \tag{13}$$

where $\mathbf{1}_t = \mathbf{1}(z_t \in [a,b])$, and $\hat{\beta}_{it}^{*(2)} = \hat{\beta}_i^{(2)}(z_t;h^*)$ is an estimate (or a proxy) of $\beta_i^{(2)}(z_t)$. We can obtain $\left\{\left\{\hat{\beta}_{it}^{(2)}\right\}_{i=1}^{d}\right\}_{t=1}^{T}$ in the following manner. For each i, we regress the LL estimate $\hat{\beta}_{it}^* = \hat{\beta}_i(z_t;h^*)$ on a pth order polynomial in z_t, i.e., $\hat{\beta}_{it}^* = \sum_{j=0}^{p}\delta_{ij}z_t^j + e_t$ for some prespecified $p\ (\geq 2)$. Let $\hat{\delta}_{ij}$ be the least-squares estimate of δ_{ij}. Then, $\hat{\beta}_{it}^{*(2)}$ is given by

$$\hat{\beta}_{it}^{*(2)} = \hat{\beta}_i^{(2)}\left(z_t;h^*\right) = \sum_{j=2}^{p} j\left(j-1\right)\hat{\delta}_{ij}z_t^{j-2}.$$

On $\int \operatorname{tr}\left(\Sigma_{S'\beta}\right)\left\{\psi(z)/f(z)\right\}dz$. Setting $\psi(z) = f(z)\,\mathbf{1}(z \in [a,b])$ gives $\int \operatorname{tr}\left(\Sigma_{S'\beta}\right)\left\{\psi(z)/f(z)\right\}dz = \operatorname{tr}\left(\Sigma_{S'\beta}\right)(b-a)$. Because it is not hard to see that $\Sigma_{S'\beta}$ may be estimated by $T^{2\alpha+1}S'\left(\sum_{t=1}^{T}x_t x_t'\right)^{-1}S$, a proxy of $\int \mathrm{tr}\left(\Sigma_{S'\beta}\right) dz$ can be obtained as

$$\mathrm{tr}\left\{T^{2\alpha+1}S'\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} S\right\}(b-a). \tag{14}$$

Substituting $\hat{\sigma}_{11}^{*}$, (13), and (14) into (11) establishes the fixed-point problem

$$h^{*} = \left[\frac{\mu_{02}(K)\,\hat{\sigma}_{11}^{*}\mathrm{tr}\left\{T^{2\alpha+1}S'\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} S\right\}(b-a)}{\{\mu_{21}(K)\}^2 \sum_i (1/T)\sum_{t=1}^{T} \mathbf{1}_t \left(\hat{\beta}_{it}^{*(2)}\right)^2}\right]^{1/5} T^{-(2\alpha+1)/5}. \tag{15}$$

Solving this equation numerically for h^{*} yields the SP bandwidth $\hat{h}_{SP}$.

A few issues remain in implementing fixed-point equation (15). First, we must choose the compact interval $[a,b]$ and the order of polynomial p. Our preliminary Monte Carlo studies indicate that $p=4$ works well. On the other hand, the interval $[a,b]$ should depend on the range of the transition variable z_t. If the support of its marginal density is known *a priori* to be compact (as in PLLR estimation), the interval should match the support. Additionally, it is often the case that z_t takes the form of a percentage change (e.g., z_t is defined as the difference of a log-transformed integrated regressor). In this situation, $[a,b]=[0,1]$ or $[-1,1]$ may be a reasonable choice. Second, the right-hand side of (15) is a highly nonlinear function of h^{*}, and there may be multiple roots. In case of multiple roots, we follow the suggestion in Park and Marron (1990) and define $\hat{h}_{SP}$ as the largest root that solves (15).

5. Hypothesis testing

Below we provide a brief discussion on hypothesis testing. Hypothesis tests of interest include testing the null of parameter constancy and testing the null of no trends in the cointegrating regression. The basic idea for each testing closely follows Xiao (2009) and Cai and Xiao (2012).

5.1. Testing for parameter constancy

We first study the testing problem for the null of constant coefficients in (1), i.e.,

$$H_0 : \beta(z) = \beta.$$

To test H_0 against H_1 : at least one element of $\beta(z)$ is varying, we may follow the idea of Sun et al. (2008) and construct an L^2-type test statistic. However, their test statistic is built on a somewhat restrictive assumption on the regression error u_{1t}, which does not admit our Assumption 1(iii). Although it could be possible to relax the assumption, an extension in this direction is beyond the scope of this article.

Instead, we adopt the approach taken by Xiao (2009) and Cai and Xiao (2012). The test statistic is the maximum of Wald statistics using $\hat{\beta}(\cdot)$ on q distinct design points $\{z_i\}_{i=1}^{q}$. Observe that under H_0, the regression model (1) collapses to a usual constant-coefficient cointegrating regression

$$y_t = x_t'\beta + u_{1t}. \tag{16}$$

Because the value of β is left unspecified, the Wald statistic using the estimator $\hat{\beta}(z_i)$ requires a consistent estimator of β in (16). What matters is the convergence rate of the estimator, not its efficiency, and thus we consider the ordinary least squares (OLS) estimator of β. Given the OLS estimator $\hat{\beta}_{OLS}$ and the set of design points $\{z_i\}_{i=1}^{q}$, we consider the Wald statistic

$$W_1(z_i) = \left\{\hat{\beta}(z_i) - \hat{\beta}_{OLS}\right\}' \left[\mu_{02}(K)\,\hat{\sigma}_{11}\left\{\sum_{t=1}^{T} x_t x_t' K\left(\frac{z_t - z_i}{h}\right)\right\}^{-1}\right]^{-1} \left\{\hat{\beta}(z_i) - \hat{\beta}_{OLS}\right\},$$

where $\hat{\sigma}_{11}$ is defined in (12).

Because Theorem 1 of Hansen (1992b) implies that $\sqrt{T}\Gamma_T'^{-1}\left(\hat{\beta}_{OLS}-\beta\right)=O_p(1)$, we have

$$\sqrt{Th}\Gamma_T'^{-1}\left\{\hat{\beta}(z)-\hat{\beta}_{OLS}\right\}=\sqrt{Th}\Gamma_T'^{-1}\left\{\hat{\beta}(z)-\beta\right\}+O_p\left(\sqrt{h}\right)\Rightarrow MN\left(0,\Sigma(z)\right).$$

A standard argument on kernel smoothing then establishes that $\{W_1(z_i)\}_{i=1}^{q}$ is a set of mutually asymptotically independent χ^2 random variables. Define the test statistic as $T_{1q}:=\max_{1\le i\le q}W_1(z_i)$. The next theorem refers to the distributional theory on T_{1q}.

Theorem 2. *If Assumptions 1–6 hold, then $T_{1q}\Rightarrow\max_{1\le i\le q}\chi_i^2(d)$ under H_0, where $\chi_1^2(d),\ldots,\chi_q^2(d)$ are independent χ^2 random variables with d degrees of freedom.*

We can reject H_0 if T_{1q} takes a very large value. Because its limiting distribution is free of nuisance parameters, it is easy to tabulate the critical values. The critical values for the distribution of the maximum of q independent χ^2 random variables with d degrees of freedom can be obtained by solving the nonlinear equation $\{F_d(x)\}^q=1-\alpha$ numerically for x, where $F_d(x)=\Pr\left\{\chi^2(d)\le x\right\}$ is the cdf of χ^2 distribution with d degrees of freedom, and α is the nominal size of this test. The remaining question is how to choose the design points $\{z_i\}_{i=1}^{q}$. As in Cai and Xiao (2012, Example 4), we may consider a set of equally spaced grid points over a prespecified compact interval.

5.2. Testing for no trends in the cointegrating regression

Restricted coefficient estimation is more efficient than unrestricted estimation, as indicated in Section 4.2, for instance. Accordingly, we next consider the test of an exclusion restriction

$$H_0:R'\beta(z)=0$$

against $H_1:R'\beta(z)\neq 0$ in (1). Specifically, the selector matrix R takes the form of

$$R=\begin{bmatrix}R_1\\0\end{bmatrix}\in\begin{cases}\mathbb{R}^{d\times(d_1-1)} & \text{if } p_1=0\\ \mathbb{R}^{d\times d_1} & \text{if } p_1>0\end{cases},$$

where

$$R_1=\begin{cases}\begin{bmatrix}0 & I_{d_1-1}\end{bmatrix}' & \text{if } p_1=0\\ I_{d_1} & \text{if } p_1>0\end{cases}.$$

Observe that R selects all coefficients on trends but an intercept in the levels regression (1). The reason why the exclusion restriction is not imposed on an intercept is that in many applications an intercept is introduced in cointegrating regressions, whether there may be trends as extra regressors or not.

As in the previous section, the test statistic takes the form of the maximum of Wald statistics. To deliver the distributional theory of this test statistic, we first strengthen Assumptions 3 and 6 as follows.

Assumption 3'. The nonnegative sequence of bandwidth $h=h_T$ satisfies $h\to 0$, $Th\to\infty$, and $T^{2p_m+1}h^5\to 0$ as $T\to\infty$.

Assumption 6'. Each column of Π_1 contains a nonzero element, $\text{rank}(\Pi_2)=m_2$, and $d_2\ge m$.

The additional condition "$T^{2p_m+1}h^5\to 0$" in Assumption 3' makes the leading bias term asymptotically negligible by undersmoothing. The FCCM-U, for instance, satisfies $p_m=1$ so that the bandwidth for LL should be $h=O\left(T^{-\alpha}\right)$, $\alpha\in(3/5,1)$. Moreover, $d_2\ge m$ in Assumption 6' means that the number of integrated regressors is at least as many as the number of trends. The condition $d_2\ge m$ establishes the lemma below, which determines the degrees of freedom in the limiting null distribution of the test statistic.

Lemma 2. *If Assumption 6' holds, then*

$$R^* := \lim_{T\to\infty} \sqrt{T}\Gamma_T R \tag{17}$$

is of the same rank as R.

For the set of design points $\{z_i\}_{i=1}^q$, again the Wald statistic

$$W_2(z_i) = \left\{R'\hat{\beta}(z_i)\right\}'\left[\mu_{02}(K)\hat{\sigma}_{11}R'\left\{\sum_{t=1}^T x_t x_t' K\left(\frac{z_t - z_i}{h}\right)\right\}^{-1} R\right]^{-1}\left\{R'\hat{\beta}(z_i)\right\}$$

is utilized, where $\hat{\sigma}_{11}$ is again given in (12). Define the test statistic as $T_{2q} := \max_{1\le i\le q} W_2(z_i)$. The next theorem refers to the distributional theory on T_{2q}. For the implementation of this test (e.g., obtaining critical values and choosing design points), the discussion in the previous section directly applies.

Theorem 3. *If Assumptions 1, 2, 3', 4, 5, and 6' hold, then*

$$T_{2q} \Rightarrow \begin{cases} \max_{1\le i\le q}\chi_i^2(d_1-1) & \text{if } p_1 = 0 \\ \max_{1\le i\le q}\chi_i^2(d_1) & \text{if } p_1 > 0 \end{cases}$$

under H_0, where $\chi_1^2(s), \ldots, \chi_q^2(s)$ are independent χ^2 random variables with s degrees of freedom.

Remark 3. Each of the two tests above is built on the first-order asymptotic theory described in Theorem 1. Asymptotic tests have the advantage of freedom of nuisance parameters in the limiting distributions. In finite samples, however, we cannot ignore the effect of the second-order bias due to endogenous regressors and/or serial dependence of the transition variable z_t. As a result, relying simply on first-order approximations to the distributions of test statistics may not provide a satisfactory solution to inference based on the levels regression (1). As an alternative, we could consider bootstrap-based tests, which may achieve some finite-sample improvement.

Moreover, in some applications, each test might lead to different conclusions across different choices of the number of design points q or different sets of design points $\{z_i\}_{i=1}^q$. This is due to the fact that the distributional theory on which Theorems 2 and 3 are grounded is pointwise in z. Then, a referee suggested that we consider the Wald statistic $W_1(z)$ (or $W_2(z)$) as a process over the whole range of z and conduct inference in a similar manner to the sup-Wald test for structural breaks. We find this extension appealing, and it would be worthwhile to investigate the properties of test statistics based on the Wald process as a future research topic.

6. Finite-sample performance

6.1. Setup

Our primary focus is on the precision of LL estimators of functional coefficients in finite samples. Recently Banerjee and Pitarakis (2012, 2014) have applied PLLR estimation (Banerjee, 2007) for estimating the FCCM with a single pure I(1) regressor and reported that PLLR estimators often outperform kernel estimators in finite samples. From this viewpoint, our Monte Carlo study compares finite-sample performance of LL and PLLR. Additionally, in Appendix B we develop the distributional theory on PLLR estimation when multiple I(1) regressors with trends enter the FCCM.

We generate the data $\left\{(y_t, x_{2t}, z_t)\right\}_{t=1}^T \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}$ from two regression specifications (9) and (10) with the initial value $S_{20} = 0$ and $\pi_0 = \pi_1 = 0.3$. The error process $u_t = (u_{1t}, u_{2t})'$ obeys the VAR(1) model

$$u_t = \Phi u_{t-1} + \epsilon_t,\ \Phi = \begin{bmatrix} \phi_{11} & 0 \\ 0 & 0 \end{bmatrix}, \quad \phi_{11} \in \{0.0, \pm 0.4, \pm 0.8\},$$

where

$$\epsilon_t = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \overset{iid}{\sim} N(0, V), \quad V = \begin{bmatrix} 1 & \sigma_{21} \\ \sigma_{21} & 1 \end{bmatrix}, \quad \sigma_{21} \in \{0.0, \pm 0.4, \pm 0.8\}.$$

A maintained assumption on PLLR is that the marginal density of the transition variable z_t is compact. In order for z_t to have a marginal density with compact support, we first generate z_t^* from the AR(1) model $z_t^* = \rho z_{t-1}^* + w_t,\ \rho \in \{0.0, \pm 0.4, \pm 0.8\}$, where $w_t \overset{iid}{\sim} N(0, 2)$ is independent of ϵ_t. Then, z_t^* is transformed to $z_t = 2\Phi\left(z_t^*\right) - 1$, where $\Phi(\cdot)$ is the cdf of $N(0, 1)$.[1] Observe that $z_t \in [-1, 1]$. The functional coefficients are set equal to $\beta_0(z) = \beta_1(z) = \beta_2(z) = \beta(z)$, where $\beta(z)$ takes the following functional forms. The following functional forms B-D are taken from Banerjee and Pitarakis (2012):

$$\begin{aligned} &\text{A: } \beta(z) = 1. \\ &\text{B: } \beta(z) = 0.3 - 0.5 \exp\left(-1.25 z^2\right). \\ &\text{C: } \beta(z) = 0.5 / \left\{1 + \exp(-4z)\right\} - 0.75. \\ &\text{D: } \beta(z) = 0.25 \exp\left(-z^2\right). \end{aligned}$$

For each combination of the regression and the functional coefficient, 1,000 Monte Carlo samples with the sample size $T = 100$ or 250 are simulated. For each Monte Carlo sample, the functional-coefficient (or the cointegrating vector) $\beta_2(z)$ is estimated by LL and PLLR. LL estimation employs the Epanechnikov kernel $K(u) = (3/4)\left(1 - u^2\right)\mathbf{1}(|u| \le 1)$. The bandwidth is chosen via the solve-the-equation plug-in method in Section 4. This bandwidth is denoted by SP, and the LL estimator using the bandwidth is referred to as LL-SP. To compute the bandwidth, we set the compact interval and the order of the polynomial equal to $[a, b] = [-1, 1]$ and $p = 4$ so that the interval matches the support of the marginal density of z_t. Often a very simple bandwidth formula is applied in the literature (e.g., Juhl, 2005; Xiao, 2009). In light of this, a "rule-of-thumb" bandwidth is also employed. Specifically, $\hat{h}_{ROT,R} = 2\hat{\sigma}_z T^{-3/5}$ for (9) and $\hat{h}_{ROT,U} = 2\hat{\sigma}_z T^{-2/5}$ for (10) are considered, where $\hat{\sigma}_z$ is the sample standard deviation of observations $\{z_t\}_{t=1}^T$. Each bandwidth is denoted by ROT, and the LL estimator using the bandwidth is referred to as LL-ROT. Our aim is to compare the gain in accuracy from the SP bandwidth with the one from the ROT bandwidth. It is not hard to see that for PLLR estimation, the MISE-optimal bin length for $\tilde{\beta}_2(z)$ for $d_2 > m_2$ is $\ell^* = O\left(T^{-2/3}\right)$, and thus we set the number of bins equal to $N = 10\left\lfloor T^{2/3}/10 \right\rfloor$ so that $\ell = O\left(T^{-2/3}\right)$ is preserved.[2] This rule roughly mimics the bin numbers chosen in Banerjee and Pitarakis (2012, 2014). In our setup, $N = 20$ (or $\ell = 0.100$) for $T = 100$ and $N = 30$ (or $\ell = 0.067$) for $T = 250$.

As in Banerjee and Pitarakis (2012, 2014), midpoints of the bins are used as the design points for both LL and PLLR. For a nonparametric estimator $\bar{\beta}_2(z)$, we employ the root mean squared error (RMSE)

$$RMSE\left\{\bar{\beta}_2(z)\right\} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left\{\bar{\beta}_2(z_j) - \beta_2(z_j)\right\}^2}$$

as its performance measure, where the design points are $\{z_j\}_{j=1}^N = \left\{\left(H_j^L + H_j^U\right)/2\right\}_{j=1}^N$. Medians of RMSEs over 1,000 replications are taken and reported for performance evaluation.

[1] It appears that the dependent structure of z_t^* is well transmitted to z_t. Averages of the first-order sample autocorrelations of z_t over 1,000 replications are typically $-0.74, -0.37, -0.00, 0.36, 0.73$ for $\rho = -0.8, -0.4, 0.0, 0.4, 0.8$, respectively.

[2] It may not be a good idea to apply the MISE-optimal bin length for $\tilde{\beta}_2(z)$ for $d_2 = m_2$ $\left(\ell^* = O\left(T^{-1}\right)\right)$ to the regression (9). If we put $N = T$ and z_t is uniformly distributed, then the number of observations falling into each bin will be only one on average.

6.2. Results

Simulation results can be found on the first author's webpage (http://www.setsunan.ac.jp/∼hirukawa/). For each combination of $\beta(z)$, the sample size, and the estimation method, the estimate from the restricted model (9) tends to yield a smaller RMSE than the one from the unrestricted model (10). This reflects a slower convergence rate and a stochastic component in the asymptotic variance of the estimator for the latter. A closer examination also reveals that as ρ or ϕ_{11} moves away from zero, the performance measure becomes larger, often drastically. This may suggest necessity for an alternative approximation to the estimator in the presence of a persistent regression error or transition variable.

Results also indicate that in most cases, LL-SP performs better than LL-ROT and PLLR. Improvement in the performance measure by LL-SP is often substantial. To be more precise, LL-SP performs best for all restricted models and for most unrestricted models. Even though it is outperformed for $T = 100$, it performs best for $T = 250$ except in only a few examples. In other words, given typical numbers of observations for cointegration analysis, the SP bandwidth is expected to work reasonably well.

In contrast, the performance of PLLR is typically as good as that of LL-ROT at best, as opposed to the results provided in Banerjee and Pitarakis (2012, 2014). Additionally, PLLR exhibits substantial variability, and its instability is not resolved even after the sample size is increased. We may attribute the contradictory PLLR results between Banerjee and Pitarakis (2012, 2014) and us to the following two aspects. First, they estimate Xiao's (2009) version of the FCCM using a higher-order kernel and/or an *ad hoc* bandwidth including ROT and the reciprocal of the number of bins. These may be disadvantages to the kernel estimator because nonparametric estimators employing higher-order kernels tend to yield unstable estimates in finite samples and finite-sample behavior of kernel estimators is also affected considerably by bandwidth selectors. Second, they consider much larger sample sizes (e.g., $T = 500, 1{,}000, 2{,}000$). PLLR appears to work well in such a data-rich environment, whereas our Monte Carlo design adopts typical sample sizes for cointegration analysis using low-frequency data. It could be the case that PLLR fits well with medium- to high-frequency data.

7. An application: Estimating electricity demand functions in the state of Illinois

We now apply the FCCM to electricity demand analysis in Illinois. Our primary focus is how price or income elasticity of electricity demand varies with the level of temperature. For this purpose, we specify the FCCM to be estimated as

$$\log q_t = \beta_0(z_t) + \beta_1(z_t)\,t + \beta_2(z_t)\log p_t + \beta_3(z_t)\log y_t + u_{1t}, \tag{18}$$

where, with a slight abuse of notation, q_t and p_t are the consumption and price of electricity, y_t is income or output, and z_t is temperature. We work with 288 monthly observations from January 1990 to December 2013. For each of the end-use sectors in Illinois — namely, the residential, commercial, and industrial sectors — electricity consumption and price data are obtained from the US Energy Information Administration. Electricity prices are then converted into relative prices by dividing them by the consumer price index of energy in the Chicago–Gary–Kenosha area. Real disposable personal income and industrial production index are chosen as measures of income (for the residential sector) and output (for the commercial and industrial sectors), respectively.[3] These are taken from the Federal Reserve Economic Data at the Federal Reserve Bank of St. Louis. Finally, CustomWeather provides monthly average temperature in Illinois. Table 3 reports the results of augmented Dickey-Fuller (ADF) unit root tests and Kwiatkowski et al. (1992) (KPSS) stationarity tests. We can find evidence of stationarity in z_t. The ADF test fails to reject the null of a unit root after detrending for most of the remaining variables; the only exception is $\log p_t$ for the commercial sector. Moreover, results from the KPSS tests

[3]Because there is no statewide monthly income or output available, we employ nationwide income and output measures by assuming that income and output levels in Illinois are constant proportions of the corresponding nationwide ones over the entire sample period.

Table 3. Results of unit root and stationarity tests.

Variable		ADF		KPSS	
		demeaned	detrended	level stationary	trend stationary
$\log q_t$	Residential	−1.578	−3.423	2.306∗	0.287∗
	Commercial	−1.606	−1.145	2.197∗	0.436∗
	Industrial	−2.128	−2.424	1.135∗	0.100
$\log p_t$	Residential	−0.586	−2.977	2.191∗	0.209∗
	Commercial	−0.510	−3.763∗	2.274∗	0.181∗
	Industrial	−1.022	−3.139	2.055∗	0.109
$\log y_t$	Disposable Income	−2.263	−0.041	2.473∗	0.527∗
	Industrial Production	−2.167	−1.956	1.897∗	0.510∗
z_t	Temperature	−3.118∗	−3.138	0.131	0.058
Critical Value (5%)		−2.879	−3.429	0.463	0.146

The ADF tests are based on regression with 12 lags. The Bartlett kernel and the Newey and West (1994) automatic bandwidth are employed for the long-run variance estimate in the KPSS test. "∗" indicates significance at the 5% level.

Table 4. Results of specification tests.

Sector	Cointegration	Parameter constancy	No trends
Residential	1.35	160.48∗	7.78
Commercial	−0.30	291.20∗	68.45∗
Industrial	−1.23	87.61∗	7.17
Critical Value (5%)	1.65	16.37	9.10

The Bartlett kernel and the Newey and West (1994) automatic bandwidth are employed for the long-run variance estimate in the cointegration test by Xiao (2009). Critical values at the 5% level for nulls of parameter constancy, and no trends are calculated from the distributions of $\max_{1\le i\le 20} \chi_i^2$ (4) and $\max_{1\le i\le 20} \chi_i^2$ (1), respectively. "∗" indicates significance at the 5% level.

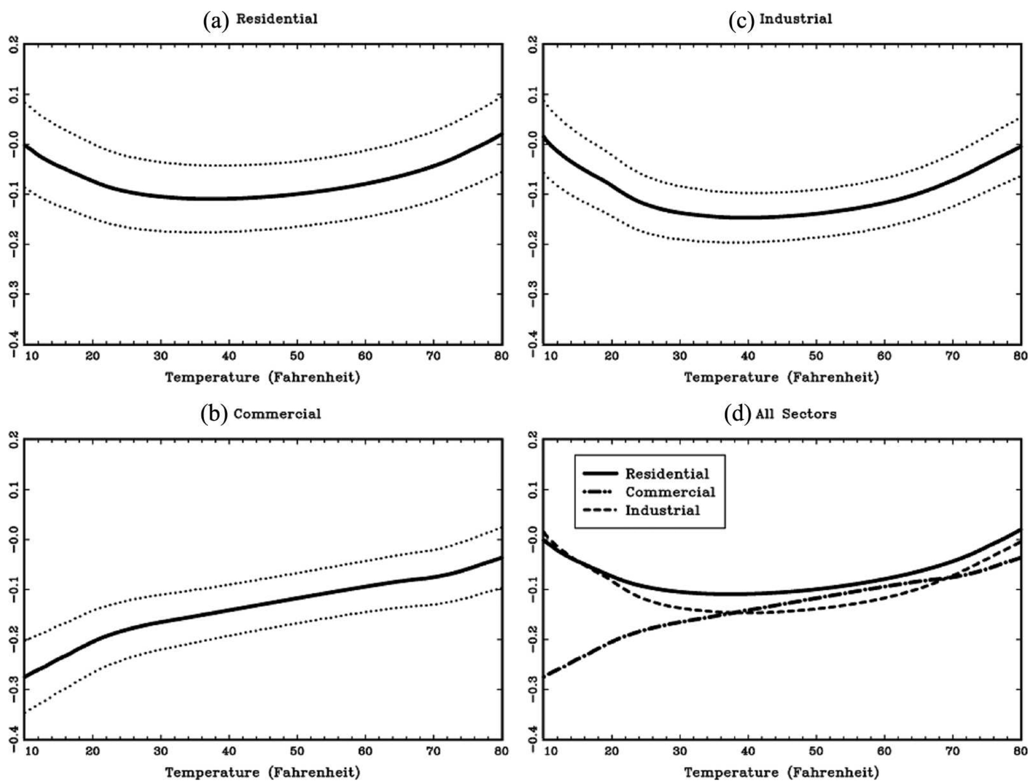

Figure 1. Estimated price elasticity of electricity demand against temperature $\hat{\beta}_2(z)$.

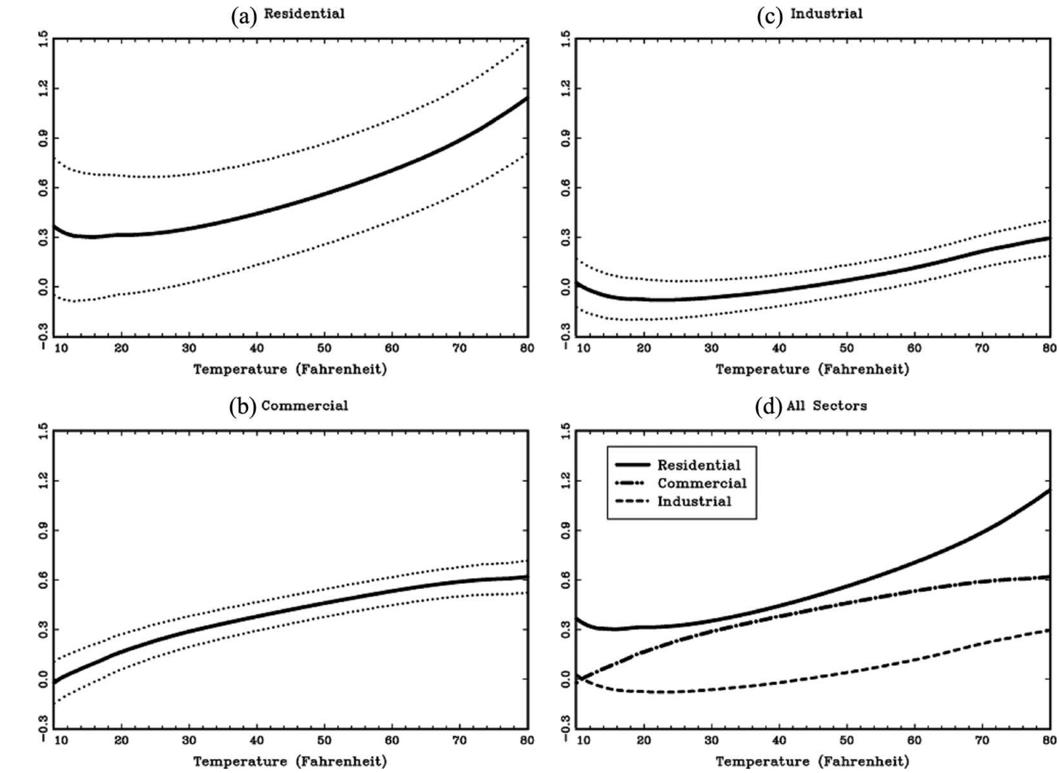

Figure 2. Estimated income elasticity of electricity demand against temperature $\hat{\beta}_3(z)$.

confirm that these variables are likely to be I(1) with a linear trend, except $\log q_t$ and $\log p_t$ for the industrial sector. All in all, the data set appears to fit with the framework of the FCCM.

LL estimation of (18) employs the Epanechnikov kernel and the SP bandwidth. Plots of LL estimates $\hat{\beta}_2(z)$ and $\hat{\beta}_3(z)$ and their 95% confidence bands for each of three sectors are presented in Figures 1 and 2. Additionally, results from the tests for cointegration, parameter constancy, and no trends are reported in Table 4, where the cointegration test is a direct application of the one proposed in Xiao (2009) and equi-spaced 20 design points (i.e., $q = 20$) are chosen for the second and third tests. Xiao's (2009) test does not reject the null of cointegration for any sector. As the figures indicate, the null of parameter constancy is strongly rejected for all sectors. On the other hand, there is evidence of the linear trend term in (18) for the commercial sector alone.

Figure 1 indicates that price elasticity for each sector is very small in magnitude, demonstrating that electricity is a necessity. The elasticity for the residential and industrial sectors is U-shaped. This comes from the fact that price sensitivity tends to increase at the temperature at which there is little demand for heating or cooling. In contrast, the price elasticity for the commercial sector is upward sloping. A rationale could be that at low temperature commercial customers are willing to switch to other heating apparatuses, such as ones fueled by gas, when finding the electricity price high. Furthermore, Figure 2 indicates that income elasticity is upward sloping for each sector. It may be the case that because high-income people tend to live in larger houses or high-production firms are likely to use larger buildings, their electricity demand for cooling becomes higher on hot summer days.

8. Conclusion

This paper has extended Xiao's (2009) FCCM to the cases in which nonstationary regressors have both stochastic and deterministic trends. LL estimation is employed to estimate the FCCM consistently, and a nondegenerate distributional theory for the LL estimator is explored. It is demonstrated that endogeneity in integrated regressors does not cause a second-order effect. In addition, the convergence rate of the nonparametric estimator of each coefficient is shown to depend on the model specification. In any case, the convergence rate on the estimator of the cointegrating vector becomes no slower than $Th^{1/2}$. We also study hypothesis tests for the null of constant coefficients and for the null of no trends in the regression. A solve-the-equation plug-in bandwidth choice rule is proposed as an implementation method for LL estimation, and its better finite-sample property is confirmed by Monte Carlo simulations. Finally, the FCCM is applied for estimating electricity demand functions in Illinois, and changes in price and income elasticity of electricity demand due to temperature are investigated.

Appendix A: Technical proofs

All the proofs are based on condition (iii) in Assumption 1. Proofs for all other cases are quite similar and thus omitted.

A.1. Proof of Lemma 1

The set of assumptions allows us to apply the FCLT (see, for example, Hansen, 1992c), and thus our remaining task is to specify the long-run variance matrix $\Omega_U(z)$. To do so, we only need to show the following two statements:

$$Var\left\{\frac{1}{\sqrt{Th}}U_T(z)\right\} = \mu_{02}(K)\,\sigma_{11}f(z) + o(1), \tag{A1}$$

$$Cov\left\{\frac{1}{\sqrt{T}}S_T, \frac{1}{\sqrt{Th}}U_T(z)\right\} = o(1). \tag{A2}$$

Proof of (A1). Observe that

$$\begin{aligned} Var\left\{\frac{1}{\sqrt{Th}}U_T(z)\right\} &= \frac{1}{h}\sum_{k=-(T-1)}^{T-1}\left(1-\frac{|k|}{T}\right)E\left\{\underline{K}_t(z)\,u_{1t}\underline{K}_{t-k}(z)\,u_{1t-k}\right\} \\ &= E\left\{\frac{1}{h}\underline{K}_t^2(z)\,u_{1t}^2\right\} + \frac{2}{h}\sum_{k=1}^{T-1}\left(1-\frac{|k|}{T}\right)E\left\{\underline{K}_t(z)\,u_{1t}\underline{K}_{t-k}(z)\,u_{1t-k}\right\} \\ &= A_1 + 2A_2 \text{ (say)}. \end{aligned}$$

A_1 can be approximated by

$$\begin{aligned} A_1 &= \left[E\left\{\frac{1}{h}K^2\left(\frac{z_t-z}{h}\right)\right\} - hE^2\left\{\frac{1}{h}K\left(\frac{z_t-z}{h}\right)\right\}\right]E\left(u_{1t}^2\right) \\ &= \left[\mu_{02}(K)\left\{f(z)+o(1)\right\}+O(h)\right]\sigma_{11} \\ &\to \mu_{02}(K)\,f(z)\,\sigma_{11}. \end{aligned}$$

Hence, (A1) is established if $A_2 = o(1)$. Now, $|A_2|$ is bounded by

$$\begin{aligned} \sum_{k=1}^{\infty}\frac{1}{h}\left|E\left\{\underline{K}_t(z)\,u_{1t}\underline{K}_{t-k}(z)\,u_{1t-k}\right\}\right| &= \left(\sum_{k=1}^{d_T}+\sum_{k=d_T+1}^{\infty}\right)\frac{1}{h}\left|E\left\{\underline{K}_t(z)\,u_{1t}\underline{K}_{t-k}(z)\,u_{1t-k}\right\}\right| \\ &= A_{21} + A_{22} \text{ (say)}, \end{aligned}$$

where the increasing sequence d_T is specified shortly. We evaluate A_{22} first. By Davydov's lemma (see, for example, Corollary A.2 of Hall and Heyde, 1980) and the stationarity of (z_t, u_{1t}),

$$\left|E\left\{\underline{K}_t(z)\,u_{1t}\underline{K}_{t-k}(z)\,u_{1t-k}\right\}\right| \le 8\left(\left\|\underline{K}_t(z)\,u_{1t}\right\|_\delta\right)^2\alpha(k)^{1-2/\delta}. \tag{A3}$$

Because $\|u_{1t}\|_\delta < \infty$ and $E\left|\underline{K}_t(z)\right|^\delta \leq O(h)$ by C_r-inequality, we have

$$\left\|\underline{K}_t(z)\,u_{1t}\right\|_\delta = \left\|\underline{K}_t(z)\right\|_\delta \|u_{1t}\|_\delta \leq O\left(h^{1/\delta}\right). \tag{A4}$$

The size of the mixing coefficient also implies that

$$\alpha(k) \leq ck^{-\{\delta\gamma/(\delta-\gamma)+\epsilon\}} \tag{A5}$$

for some $\epsilon > 0$. Substituting (A4)–(A5) into (A3) yields

$$\frac{1}{h}\left|E\left\{\underline{K}_t(z)\,u_{1t}\underline{K}_{t-k}(z)\,u_{1t-k}\right\}\right| \leq ch^{-(1-2/\delta)}k^{-\{\delta\gamma/(\delta-\gamma)+\epsilon\}(1-2/\delta)}.$$

For such ϵ, define $d_T := \lfloor h^{-a}\rfloor$ for some $a \in (a_1, 1)$, where $a_1 = \{\delta\gamma/(\delta-\gamma)+\epsilon-1/(1-2/\delta)\}^{-1}$ and $0 < a_1 < 1$ is ensured by $\delta > \gamma > 2$. Then,

$$A_{22} \leq ch^{-(1-2/\delta)}\sum_{k=d_T+1}^{\infty} k^{-\{\delta\gamma/(\delta-\gamma)+\epsilon\}(1-2/\delta)}.$$

It follows from $\delta > \gamma > 2$ that $\{\delta\gamma/(\delta-\gamma)+\epsilon\}(1-2/\delta) > 2$. As a result,

$$\sum_{k=d_T+1}^{\infty} k^{-\{\delta\gamma/(\delta-\gamma)+\epsilon\}(1-2/\delta)} \leq \int_{d_T}^{\infty} x^{-\{\delta\gamma/(\delta-\gamma)+\epsilon\}(1-2/\delta)}dx = cd_T^{-(1-2/\delta)/a_1} = O\left\{h^{(a/a_1)(1-2/\delta)}\right\},$$

and thus $A_{22} \leq O\left\{h^{(a/a_1-1)(1-2/\delta)}\right\} \to 0$.

We now turn to A_{21}. Notice that $\left|E\left\{\underline{K}_t(z)\,u_{1t}\underline{K}_{t-k}(z)\,u_{1t-k}\right\}\right| = \left|E\left\{\underline{K}_t(z)\,\underline{K}_{t-k}(z)\right\}\right|\left|E\left(u_{1t}u_{1t-k}\right)\right|$, where $\left|E\left(u_{1t}u_{1t-k}\right)\right| < \infty$ by Cauchy–Schwarz inequality and the stationarity of u_{1t}. Moreover,

$$\left|E\left\{\underline{K}_t(z)\,\underline{K}_{t-k}(z)\right\}\right| \leq E\left|K\left(\frac{z_t-z}{h}\right)K\left(\frac{z_{t-k}-z}{h}\right)\right| + E\left|K\left(\frac{z_t-z}{h}\right)\right|E\left|K\left(\frac{z_{t-k}-z}{h}\right)\right|.$$

Uniform boundedness of the joint density for (z_t, z_{t-k}) in Assumption 4 implies that the first term is $O\left(h^2\right)$, whereas the second term is also $O\left(h^2\right)$. In conclusion, $\left|E\left\{\underline{K}_t(z)\,u_{1t}\underline{K}_{t-k}(z)\,u_{1t-k}\right\}\right| \leq O\left(h^2\right)$. Therefore, $A_{21} \leq O(d_T h) = O\left(h^{1-a}\right) \to 0$ is also established.

Proof of (A2). The left-hand side of (A2) reduces to

$$\frac{1}{\sqrt{h}}\sum_{k=-(T-1)}^{T-1}\left(1-\frac{|k|}{T}\right)E\left\{u_t\underline{K}_{t-k}(z)\,u_{1t-k}\right\}.$$

For $k = 0$, $\left\|E\left\{\underline{K}_t(z)\,u_tu_{1t}\right\}/\sqrt{h}\right\| \leq \left\{E\left|\underline{K}_t(z)\right|/\sqrt{h}\right\}E\left\|u_tu_{1t}\right\| = O\left(\sqrt{h}\right) \to 0$, whereas absolute convergence of the sum to zero for $|k| \geq 1$ can be shown in a similar manner to the proof of (A1). □

A.2. Proof of Theorem 1

This proof requires the following lemma.

Lemma A1. *If Assumptions 1–4 and 6 hold, then*

$$\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\Gamma_T x_t K\left(\frac{z_t-z}{h}\right)u_{1t} \Rightarrow \int_0^1 J(r)\,dB_{U(z)}(r),$$

where $B_{U(z)}(r)$ is independent of $J(r)$.

A.2.1. Proof of Lemma A1.
Notice that

$$
\begin{aligned}
&\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\Gamma_T x_t K\left(\frac{z_t-z}{h}\right)u_{1t}\\
&=\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\Gamma_T x_t \underline{K}_t(z)\,u_{1t}+\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\Gamma_T x_t E\left\{K\left(\frac{z_t-z}{h}\right)\right\}u_{1t}\\
&=B_1+B_2\ \ \text{(say)}.
\end{aligned}
$$

By Theorem 4.1 of Hansen (1992c) and Lemma 1, we have

$$
\begin{aligned}
B_1 &= \begin{bmatrix} \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} D_T^{-1}k_t\underline{K}_t(z)\,u_{1t}+o_p(1)\\ \frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\left(\Pi_2^{*\prime}\Omega_{22}\Pi_2^{*}\right)^{-1/2}\Pi_2^{*\prime}\left(S_{2t}/\sqrt{T}\right)\underline{K}_t(z)\,u_{1t}\end{bmatrix}\\
&\Rightarrow \begin{bmatrix}\int_0^1 k(r)\,dB_{U(z)}(r)\\ \left(\Pi_2^{*\prime}\Omega_{22}\Pi_2^{*}\right)^{-1/2}\Pi_2^{*\prime}\int_0^1 B_{S_2}(r)\,dB_{U(z)}(r)\end{bmatrix}\\
&=\int_0^1 J(r)\,dB_{U(z)}(r),
\end{aligned}
$$

where $J(r)$ and $B_{U(z)}(r)$ are uncorrelated and thus independent due to Gaussianity. On the other hand,

$$
B_2=\sqrt{h}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Gamma_T x_t E\left\{\frac{1}{h}K\left(\frac{z_t-z}{h}\right)\right\}u_{1t}=\sqrt{h}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Gamma_T x_t u_{1t}\right)\{f(z)+o(1)\}.
$$

It follows from Theorem 3(c) of Hansen (1992b) that

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Gamma_T x_t u_{1t}\Rightarrow\int_0^1 J(r)\,dB_{S_1}(r)+\begin{bmatrix}0\\ \Lambda_{21}^{*}\end{bmatrix},
$$

where $\Lambda_{21}^{*}:=\left(\Pi_2^{*\prime}\Omega_{22}\Pi_2^{*}\right)^{-1/2}\Pi_2^{*\prime}\Lambda_{21}$, and $\Lambda_{21}:=\sum_{j=0}^{\infty}E\left(u_{2t}u_{1t+j}\right)$. Therefore, $B_2=\sqrt{h}O_p(1)\,O(1)=O_p\left(\sqrt{h}\right)\xrightarrow{p}0$, and thus the lemma is established. □

A.2.2. Proof of Theorem 1.
The numerator of $\hat{\beta}(z)$ can be rewritten as

$$
\begin{aligned}
&T_0(z)-S_1(z)\,S_2(z)^{-1}T_1(z)\\
&\quad=\left\{\sum_{t=1}^{T}x_t x_t'\beta(z_t)\,K\left(\frac{z_t-z}{h}\right)-S_1(z)\,S_2(z)^{-1}\sum_{t=1}^{T}x_t x_t'\beta(z_t)\,(z_t-z)\,K\left(\frac{z_t-z}{h}\right)\right\}\\
&\qquad+\left\{\sum_{t=1}^{T}x_t u_{1t}K\left(\frac{z_t-z}{h}\right)-S_1(z)\,S_2(z)^{-1}\sum_{t=1}^{T}x_t u_{1t}\,(z_t-z)\,K\left(\frac{z_t-z}{h}\right)\right\}\\
&\quad=G_1(z)+G_2(z)\ \ \text{(say)}.
\end{aligned}
\tag{A6}
$$

Using a second-order Taylor expansion of $\beta(z_t)$ around $z_t=z$, we have

$$
\begin{aligned}
G_1(z)&=\left\{S_0(z)-S_1(z)\,S_2(z)^{-1}S_1(z)\right\}\beta(z)\\
&\quad+\frac{1}{2}\left\{S_2(z)-S_1(z)\,S_2(z)^{-1}S_3(z)\right\}\beta^{(2)}(z)+R(z),
\end{aligned}
\tag{A7}
$$

where

$$R(z) = \sum_{t=1}^{T} x_t x_t' \xi_t(z) K\left(\frac{z_t - z}{h}\right) - S_1(z) S_2(z)^{-1} \sum_{t=1}^{T} x_t x_t' \xi_t(z)(z_t - z) K\left(\frac{z_t - z}{h}\right),$$

$$\xi_t(z) = \beta(z_t) - \left\{\beta(z) + \beta^{(1)}(z)(z_t - z) + \frac{1}{2}\beta^{(2)}(z)(z_t - z)^2\right\}.$$

Substituting (A6) and (A7) into (2) yields

$$\begin{aligned}
&\sqrt{Th}\Big[\Gamma_T'^{-1}\left\{\hat{\beta}(z) - \beta(z)\right\} \\
&\quad - \frac{1}{2}\Gamma_T'^{-1}\left\{S_0(z) - S_1(z) S_2(z)^{-1} S_1(z)\right\}^{-1}\left\{S_2(z) - S_1(z) S_2(z)^{-1} S_3(z)\right\}\beta^{(2)}(z) \\
&\quad - \Gamma_T'^{-1}\left\{S_0(z) - S_1(z) S_2(z)^{-1} S_1(z)\right\}^{-1} R(z)\Big] \\
&= \Gamma_T'^{-1}\left\{S_0(z) - S_1(z) S_2(z)^{-1} S_1(z)\right\}^{-1}\sqrt{Th}G_2(z). \qquad \text{(A8)}
\end{aligned}$$

Notice that

$$\begin{aligned}
&\Gamma_T'^{-1}\left\{S_0(z) - S_1(z) S_2(z)^{-1} S_1(z)\right\}^{-1}\left\{S_2(z) - S_1(z) S_2(z)^{-1} S_3(z)\right\} \\
&= \left[\frac{1}{Th}\Gamma_T\left\{S_0(z) - S_1(z) S_2(z)^{-1} S_1(z)\right\}\Gamma_T'\right]^{-1}\left[\frac{1}{Th}\Gamma_T\left\{S_2(z) - S_1(z) S_2(z)^{-1} S_3(z)\right\}\Gamma_T'\right]\Gamma_T'^{-1} \\
&= G_3(z)^{-1} G_4(z)\Gamma_T'^{-1} \text{ (say).} \qquad \text{(A9)}
\end{aligned}$$

Now,

$$G_3(z) = \frac{1}{Th}\Gamma_T S_0(z)\Gamma_T' - \left\{\frac{1}{Th^2}\Gamma_T S_1(z)\Gamma_T'\right\}\left\{\frac{1}{Th^3}\Gamma_T S_2(z)\Gamma_T'\right\}^{-1}\left\{\frac{1}{Th^2}\Gamma_T S_1(z)\Gamma_T'\right\}.$$

In particular,

$$\frac{1}{Th}\Gamma_T S_0(z)\Gamma_T' = \frac{1}{T}\sum_{t=1}^{T}\Gamma_T x_t(\Gamma_T x_t)' E\left\{\frac{1}{h}K\left(\frac{z_t - z}{h}\right)\right\} + \frac{1}{Th}\sum_{t=1}^{T}\Gamma_T x_t(\Gamma_T x_t)'\underline{K}_t(z).$$

It can be shown in a similar manner to the proof of Lemma A.3 of Sun et al. (2011) that $\left(1/\sqrt{Th}\right)\sum_{t=1}^{T}\Gamma_T x_t(\Gamma_T x_t)'\underline{K}_t(z) = O_p(1)$. Therefore,

$$\begin{aligned}
\frac{1}{Th}\Gamma_T S_0(z)\Gamma_T' &= \left\{\int_0^1 J(r) J(r)' dr + o_p(1)\right\}\{f(z) + o(1)\} + O_p\left\{(Th)^{-1/2}\right\} \\
&= f(z)\int_0^1 J(r) J(r)' dr + o_p(1).
\end{aligned}$$

A similar argument establishes that

$$\frac{1}{Th^3}\Gamma_T S_2(z)\Gamma_T' = \mu_{21}(K) f(z)\int_0^1 J(r) J(r)' dr + o_p(1).$$

Symmetry of $K(\cdot)$ implies that $\left(T^{1/2}h^{3/2}\right)^{-1}\Gamma_T S_1(z)\Gamma_T' = O_p(1)$. Hence,

$$G_3(z) = f(z)\int_0^1 J(r) J(r)' dr + o_p(1). \qquad \text{(A10)}$$

Similarly, it can be shown that $\left(T^{1/2}h^{7/2}\right)^{-1}\Gamma_T S_3(z)\Gamma_T' = O_p(1)$, and thus

$$\begin{aligned} h^{-2}G_4(z) &= \frac{1}{Th^3}\Gamma_T S_2(z)\Gamma_T' - \left\{\frac{1}{Th^2}\Gamma_T S_1(z)\Gamma_T'\right\}\left\{\frac{1}{Th^3}\Gamma_T S_2(z)\Gamma_T'\right\}^{-1}\left\{\frac{1}{Th^4}\Gamma_T S_3(z)\Gamma_T'\right\} \\ &= \mu_{21}(K) f(z)\int_0^1 J(r)J(r)'\,dr + o_p(1). \end{aligned} \tag{A11}$$

Furthermore, it is not hard to see that

$$\Gamma_T'^{-1}\left\{S_0(z) - S_1(z)S_2(z)^{-1}S_1(z)\right\}^{-1}R(z) = G_3(z)^{-1}\left\{\frac{1}{Th}\Gamma_T R(z)\Gamma_T'\right\}\Gamma_T'^{-1}, \tag{A12}$$

where

$$\frac{1}{Th}\Gamma_T R(z)\Gamma_T' = o_p\left(h^2\right). \tag{A13}$$

It follows from (A9), (A11), (A12), and (A13) that the left-hand side of (A8) reduces to

$$\sqrt{Th}\Gamma_T'^{-1}\left\{\hat{\beta}(z) - \beta(z) - \frac{1}{2}\mu_{21}(K)\beta^{(2)}(z)h^2 + o_p\left(h^2\right)\right\}. \tag{A14}$$

On the other hand, the right-hand side of (A8) becomes $G_3(z)^{-1}\left(1/\sqrt{Th}\right)\Gamma_T G_2(z)$. Notice that

$$\frac{1}{\sqrt{Th}}\Gamma_T G_2(z) = \frac{1}{\sqrt{Th}}\sum_{t=1}^T \Gamma_T x_t K\left(\frac{z_t - z}{h}\right)u_{1t} - \frac{1}{\sqrt{Th}}\sum_{t=1}^T \Gamma_T x_t(z_t - z)K\left(\frac{z_t - z}{h}\right)u_{1t},$$

where the second term is at most $O_p(h)$. Then, by Lemma A1,

$$\frac{1}{\sqrt{Th}}\Gamma_T G_2(z) = \int_0^1 J(r)\,dB_{U(z)}(r) + o_p(1).$$

Therefore,

$$\begin{aligned} G_3(z)^{-1}\frac{1}{\sqrt{Th}}\Gamma_T G_2(z) &\Rightarrow \left\{f(z)\int_0^1 J(r)J(r)'\,dr\right\}^{-1}\int_0^1 J(r)\,dB_{U(z)}(r) \\ &\stackrel{d}{=} MN\left(0, \frac{\mu_{02}(K)\sigma_{11}}{f(z)}\left\{\int_0^1 J(r)J(r)'\,dr\right\}^{-1}\right). \end{aligned} \tag{A15}$$

Finally, combining (A14) and (A15) with (A8) establishes the distributional theory. □

A.3. Proof of Theorem 2.

Notice that $W_1(z_i)$ can be rewritten as

$$\begin{aligned} &\left[\sqrt{Th}\Gamma_T'^{-1}\left\{\hat{\beta}(z_i) - \hat{\beta}_{OLS}\right\}\right]'\left[\mu_{02}(K)\hat{\sigma}_{11}\left\{\frac{1}{Th}\sum_{t=1}^T(\Gamma_T x_t)(\Gamma_T x_t)' K\left(\frac{z_t - z}{h}\right)\right\}^{-1}\right]^{-1} \\ &\times\left[\sqrt{Th}\Gamma_T'^{-1}\left\{\hat{\beta}(z_i) - \hat{\beta}_{OLS}\right\}\right]. \end{aligned}$$

It can be shown that $\hat{\sigma}_{11} \xrightarrow{p} \sigma_{11}$, and thus

$$\mu_{02}(K)\hat{\sigma}_{11}\left\{\frac{1}{Th}\sum_{t=1}^T(\Gamma_T x_t)(\Gamma_T x_t)' K\left(\frac{z_t - z}{h}\right)\right\}^{-1} \Rightarrow \frac{\mu_{02}(K)\sigma_{11}}{f(z)}\left\{\int_0^1 J(r)J(r)'\,dr\right\}^{-1} = \Sigma(z).$$

Therefore, $W_1(z_i) \Rightarrow \chi^2(d)$. Moreover, $W_1(z_1), \ldots, W_1(z_q)$ are asymptotically independent χ^2 random variables, and thus the stated result immediately follows. □

A.4. Proof of Lemma 2.

A straightforward calculation yields

$$R^* = \begin{bmatrix} 0 \\ R_1^* \end{bmatrix} \in \begin{cases} \mathbb{R}^{d\times(d_1-1)} & \text{if } p_1 = 0 \\ \mathbb{R}^{d\times d_1} & \text{if } p_1 > 0 \end{cases},$$

where

$$R_1^* := -\left(\Pi_2^{*\prime}\Omega_{22}\Pi_2^*\right)^{-1/2}\Pi_2^{*\prime}\Pi_1 R_1 \in \begin{cases} \mathbb{R}^{(d_2-m_2)\times(d_1-1)} & \text{if } p_1 = 0 \\ \mathbb{R}^{(d_2-m_2)\times d_1} & \text{if } p_1 > 0 \end{cases}.$$

Invoking $d_1 = m_1$ and using $d_2 \geq m$, we have $d_2 - m_2 \geq m_1 = d_1 > d_1 - 1$. Therefore, R^* is of full column rank regardless of p_1, and thus

$$\text{rank}\left(R^*\right) = \text{rank}\left(R\right) = \begin{cases} d_1 - 1 & \text{if } p_1 = 0 \\ d_1 & \text{if } p_1 > 0 \end{cases}.$$

□

A.5. Proof of Theorem 3.

Under H_0,

$$T\sqrt{h}R'\hat{\beta}(z) = T\sqrt{h}R'\left\{\hat{\beta}(z) - \beta(z)\right\} = \left(\sqrt{T}\Gamma_T R\right)'\left[\sqrt{Th}\Gamma_T'^{-1}\left\{\hat{\beta}(z) - \beta(z)\right\}\right]. \tag{A16}$$

By Assumption 3' and $\Gamma_T'^{-1} = O\left(T^{p_m}\right)$, Theorem 1 can be now rewritten as

$$\sqrt{Th}\Gamma_T'^{-1}\left\{\hat{\beta}(z) - \beta(z)\right\} \Rightarrow \left\{f(z)\int_0^1 J(r)J(r)'dr\right\}^{-1}\int_0^1 J(r)\,dB_{U(z)}(r). \tag{A17}$$

Substituting (17) and (A17) into (A16) yields

$$T\sqrt{h}R'\hat{\beta}(z) \Rightarrow R^{*\prime}\left\{f(z)\int_0^1 J(r)J(r)'dr\right\}^{-1}\int_0^1 J(r)\,dB_{U(z)}(r) \overset{d}{=} MN\left(0, R^{*\prime}\Sigma(z)R^*\right). \tag{A18}$$

On the other hand,

$$\begin{aligned} &T^2h\left[\mu_{02}(K)\hat{\sigma}_{11}R'\left\{\sum_{t=1}^T x_t x_t' K\left(\frac{z_t - z_i}{h}\right)\right\}^{-1}R\right] \\ &= \left(\sqrt{T}\Gamma_T R\right)'\left[\mu_{02}(K)\hat{\sigma}_{11}\left\{\frac{1}{Th}\sum_{t=1}^T\left(\Gamma_T x_t\right)\left(\Gamma_T x_t\right)'K\left(\frac{z_t - z_i}{h}\right)\right\}^{-1}\right]\left(\sqrt{T}\Gamma_T R\right) \\ &\Rightarrow R^{*\prime}\left[\frac{\mu_{02}(K)\sigma_{11}}{f(z)}\left\{\int_0^1 J(r)J(r)'dr\right\}^{-1}\right]R^* \\ &= R^{*\prime}\Sigma(z)R^*. \end{aligned} \tag{A19}$$

Combining (A18) and (A19) with Lemma 2 gives

$$W_2(z_i) = \left\{R'\hat{\beta}(z_i)\right\}' \left[\mu_{02}(K)\hat{\sigma}_{11}R'\left\{\sum_{t=1}^{T} x_t x_t' K\left(\frac{z_t - z}{h}\right)\right\}^{-1} R\right]^{-1} \left\{R'\hat{\beta}(z_i)\right\}$$
$$\Rightarrow \begin{cases} \chi^2(d_1 - 1) & \text{if } p_1 = 0 \\ \chi^2(d_1) & \text{if } p_1 > 0 \end{cases}.$$

Because $W_2(z_1), \ldots, W_2(z_q)$ are asymptotically independent χ^2 random variables, the stated result is immediately established. □

Appendix B: PLLR estimation

B.1. The estimator

As an alternative to LL, Banerjee and Pitarakis (2012, 2014) advocate PLLR to estimate $\beta(\cdot)$ in (1) consistently. To implement PLLR, assume that the support of $f(z)$ is a compact interval $H = [H^L, H^U]$, $-\infty < H^L < H^U < \infty$. Then, H is partitioned into N disjoint bins $\{H_j\}_{j=1}^N$ with an equal length ℓ, i.e., $H_j = \left(H_j^L, H_j^U\right] = \left(H^L + (j-1)\ell, H^L + j\ell\right]$. Using the observations $\left\{\left(y_t, x_{2t}', z_t\right)'\right\}_{t=1}^T$ for all $z_t \in H_j$, the levels regression (1) is estimated as a linear one of y_t on x_t by OLS. The OLS estimator constitutes the PLLR estimator of $\beta(z)$ for a design point z as long as z also falls into H_j. Formally, the PLLR estimator $\tilde{\beta}(z)$ is defined as

$$\tilde{\beta}(z) = \arg\min_{\theta} \sum_{j=1}^{N}\sum_{t=1}^{T}\left(y_t - x_t'\theta\right)^2 \mathbf{1}_j(z_t)\mathbf{1}_j(z) = \sum_{j=1}^{N}\left\{\sum_{t=1}^{T} x_t x_t' \mathbf{1}_j(z_t)\right\}^{-1} \sum_{t=1}^{T} x_t y_t \mathbf{1}_j(z_t)\mathbf{1}_j(z),$$

where $\mathbf{1}_j(\cdot) = \mathbf{1}\left(\cdot \in H_j\right)$.

B.2. Distributional theory

To derive the asymptotic properties of $\tilde{\beta}(z)$, Assumptions 3 and 4 are modified as follows.

Assumption 3". The nonnegative sequence of bin length $\ell = \ell_T$ satisfies $\ell \to 0$ and $T\ell \to \infty$ as $T \to \infty$.

Assumption 4". H, the support of $f(z)$, is compact, and $f(z)$ is first-order Lipschitz continuous over H. In addition, $\sup_{z_0, z_s, s} f_s(z_0, z_s) < \infty$ and $f(z) > 0$ for a given design point z.

Assumption 3" suggests that the bin length ℓ in PLLR plays a similar role to the bandwidth in kernel smoothing. Differentiability of $f(z)$ in Assumption 4 is relaxed to Lipschitz continuity in Assumption 4", which suffices for approximating the leading bias term of the PLLR estimator. Boundedness of f from above (or existence of $\max f(z)$, to be precise) is ensured by compactness of H and continuity of f.

The nondegenerate distributional theory on $\tilde{\beta}(z)$ is provided below. The proof is similar to the one of Theorem 1 and thus omitted.

Theorem B1. *If Assumptions 1, 3", 4", 5, and 6 hold, then*

$$\sqrt{T\ell}\,\Gamma_T'^{-1}\left\{\tilde{\beta}(z) - \beta(z) - \mathcal{B}(z)\ell + o_p(\ell)\right\} \Rightarrow MN(0, \mathcal{V}(z)),$$

where

$$\mathcal{B}(z) := \sum_{j=1}^{N}\left(\frac{H_j^U - z}{\ell} - \frac{1}{2}\right)\beta^{(1)}(z)\,\mathbf{1}_j(z)\,,\ \mathcal{V}(z) := \frac{\sigma_{11}}{f(z)}\left\{\int_0^1 J(r)\,J(r)'\,dr\right\}^{-1},$$

and $\left(H_j^U - z\right)/\ell \in [0, 1]$.

Remark B1. Banerjee and Pitarakis (2012, 2014) allow z_t and u_t to be correlated. As a consequence, they derive only the convergence rate of $\tilde{\beta}(z)$. In contrast, we again attain a mixed-normal limit without a second-order bias correction in the presence of endogenous integrated regressors. Moreover, the leading bias term is $O(\ell)$ rather than $O\left(\ell^2\right)$. This comes from the fact that the "kernel" $\mathbf{1}_j(\cdot)$ is asymmetric with respect to the design point z unless it is the midpoint of the bin.

Acknowledgments

For their constructive comments and suggestions, we would like to thank the Associate Editor, two anonymous referees, Eiji Kurozumi, Morten Nielsen, Peter Phillips, and the participants of, respectively, the New Zealand Econometric Study Group Meeting 2011, the Recent Development in Statistics, Empirical Finance and Econometrics Symposium 2011 at Kyoto University, the Asian Meeting of the Econometric Society 2012, and a seminar at Osaka University. The views expressed herein and those of the authors do not necessarily reflect the views of the Development Bank of Japan.

Funding

The first author gratefully acknowledges financial support from Japan Society of the Promotion of Science (grant number 23530259).

References

Banerjee, A. N. (2007). A method of estimating the average derivative. *Journal of Econometrics* 136:65–88.

Banerjee, A. N., Pitarakis, J.-Y. (2012). Functional cointegration: definition and nonparametric estimation. MPRA Paper No. 38846.

Banerjee, A. N., Pitarakis, J.-Y. (2014). Functional cointegration: definition and nonparametric estimation. *Studies in Nonlinear Dynamics & Econometrics* 18:507–520.

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

Cai, Z., Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics* 167:413–425.

Caner, M., Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica* 69:1555–1596.

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

Gonzalo, J., Pitarakis, J.-Y. (2006). Threshold effects in cointegrating relationships. *Oxford Bulletin of Economics and Statistics* 68:813–833.

Granger, C. W. J. (2012). Useful conclusions from surprising results. *Journal of Econometrics* 169:142–146.

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

Hansen, B. E. (1992a). Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends. *Journal of Econometrics* 53:87–121.

Hansen, B. E. (1992b). Tests for parameter instability in regressions with I(1) processes. *Journal of Business & Economic Statistics* 10:45–59.

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

Hirukawa, M. (2010). A two-stage plug-in bandwidth selection and its implementation for covariance estimation. *Econometric Theory* 26:710–743.

Jones, M. C., Marron, J. S., Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* 91:401–407.

Juhl, T. (2005). Functional-coefficient models under unit root behaviour. *Econometrics Journal* 8:197–213.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54:159–178.

Li, K., Li, W. (2013). Estimation of functional coefficient models with time trend and integrated regressors. *Economics Letters* 119:89–93.

Liang, Z., Li, Q. (2012). Functional Coefficient Regression Models with Time Trend. *Journal of Econometrics* 170:15–31.

Newey, W. K., West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61:631–653.

Ogaki, M. (1992). Engel's law and cointegration. *Journal of Political Economy* 100:1027–1046.

Ogaki, M., Park, J. Y. (1997). A cointegration approach to estimating preference parameters. *Journal of Econometrics* 82:107–134.

Park, B. U., Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association* 85:66–72.

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

Ruppert, D., Sheather, S. J., Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association* 90:1257–1270.

Sheather, S. J., Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B.* 53:683–690.

Stock, J. H., Watson, M. W. (1993). A Simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61:783–820.

Sun, Y., Cai, Z., Li, Q. (2008). Consistent nonparametric test on parametric smooth coefficient model with nonstationary data. Working Paper, University of North Carolina at Charlotte.

Sun, Y., Hsiao, C., Li, Q. (2011). Measuring correlations of integrated but not cointegrated variables: a semiparametric approach. *Journal of Econometrics* 164:252–267.

West, K. D. (1988). Asymptotic normality, when regressors have a unit root. *Econometrica* 56:1397–1417.

Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics* 152:81–92.