Supplemental Material for
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. 3009

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

http://cowles.yale.edu/

Online Supplement to

"*Kernel-Based Inference in Time-Varying Coefficient Cointegrating Regressions*"

Degui Li[1], Peter C. B. Phillips[2] and Jiti Gao[3]

In Appendix B, we prove some supplementary results which have been used to prove the main results in Appendix A of the main document; and in Appendix C, we give some extensive simulation studies to evaluate the finite sample properties of the proposed methods in relation to the asymptotic theory.

Appendix B: Proofs of the supplemental results

This appendix provides proofs of some supplemental results used in Appendix A.

LEMMA A. *If $B_2(\cdot)$ is full rank vector Brownian motion,*

$$\begin{aligned}
\mathbf{\Lambda}_{22}(z_0) &= B_2(z_0)'B_2(z_0), \\
\mathbf{\Lambda}_{23}(z_0) &= \sqrt{2}\mathbf{\Lambda}_{22}^{1/2}(z_0)\left[\int_{-1}^{1} B_2^*\big(\frac{z+1}{2}\big)'K(z)dz\right] q_2^{\perp}(z_0), \\
\mathbf{\Lambda}_{33}(z_0) &= 2q_2^{\perp}(z_0)'\left[\int_{-1}^{1} B_2^*\big(\frac{z+1}{2}\big)B_2^*\big(\frac{z+1}{2}\big)'K(z)\,dz\right] q_2^{\perp}(z_0),
\end{aligned}$$

where $0 < z_0 < 1$, $B_2^*(\cdot)$ is an independent copy of $B_2(\cdot)$, and $q_2(z_0)$ and $q_2^{\perp}(z_0)$ are defined as in Section 2.2 of the main document, then

$$\mathbf{\Lambda}_2(z_0) = \left[\begin{array}{cc} \mathbf{\Lambda}_{22}(z_0) & \mathbf{\Lambda}_{23}(z_0) \\ \mathbf{\Lambda}_{23}(z_0)' & \mathbf{\Lambda}_{33}(z_0) \end{array}\right] > 0 \quad a.s.$$

PROOF OF LEMMA A. Since $\mathbf{\Lambda}_{22}(z_0) > 0$ *a.s.*, it is sufficient to show that

$$\mathbf{\Lambda}_{33}(z_0) - \mathbf{\Lambda}_{23}(z_0)'\mathbf{\Lambda}_{23}(z_0)/\mathbf{\Lambda}_{22}(z_0) > 0 \quad a.s.$$

Observe that

$$\begin{aligned}
& \mathbf{\Lambda}_{33}(z_0) - \mathbf{\Lambda}_{23}(z_0)'\mathbf{\Lambda}_{23}(z_0)/\mathbf{\Lambda}_{22}(z_0) \\
= \; & 2q_2^{\perp}(z_0)'\left[\int_{-1}^{1} B_2^*\big(\frac{z+1}{2}\big)B_2^*\big(\frac{z+1}{2}\big)'K(z)dz\right] q_2^{\perp}(z_0) - 2\left[\int_{-1}^{1} B_2^*\big(\frac{z+1}{2}\big)'q_2^{\perp}(z_0)K(z)dz\right]^2 \\
= \; & 2\int_{-1}^{1}[A(z)-A]^2\,K(z)dz,
\end{aligned}$$

[1]University of York
[2]Yale University, University of Auckland, Southampton University, and Singapore Management University
[3]Monash University

where $A(z) = q_2^{\perp}(z_0)' B_2^*\left(\frac{z+1}{2}\right)$ and $A = \int_{-1}^{1} A(z)K(z)dz$. The result follows because $K(z) > 0$ and therefore $\int_{-1}^{1} [A(z) - A]^2 K(z)dz = 0$ implies $A(z) - A = 0$ almost everywhere, which holds with probability zero because $B_2(\cdot)$ and $B_2^*(\cdot)$ are full rank vector Brownian motions. □

PROOF OF (A.5). We use a truncation technique to prove this result. Define

$$\widetilde{e}_{t1} = \left(\mathbf{O}_{d_0\times 1}, \mathbf{I}_{d_0}, \mathbf{O}_{d_0\times d_1}\right) \cdot \left(\sum_{j=0}^{m_T} \mathbf{\Phi}_j \varepsilon_{t-j}\right), \quad \overline{e}_{t1} = \left(\mathbf{O}_{d_0\times 1}, \mathbf{I}_{d_0}, \mathbf{O}_{d_0\times d_1}\right) \cdot \left(\sum_{j=m_T+1}^{\infty} \mathbf{\Phi}_j \varepsilon_{t-j}\right)$$

where $\mathbf{O}_{k\times j}$ is a $k \times j$ null matrix, $\mathbf{I}_k$ is a $k \times k$ identity matrix, $m_T \to \infty$ and $m_T = o(Th)$ as $T \to \infty$. Observe that

$$\begin{aligned}
&\sum_{t=1}^{T} \left\{e_{t1}e_{t1}' - \mathbb{E}[e_{t1}e_{t1}']\right\} K_{th}(z_0) \\
= &\sum_{t=1}^{T} \left\{\widetilde{e}_{t1}\widetilde{e}_{t1}' - \mathbb{E}[\widetilde{e}_{t1}\widetilde{e}_{t1}']\right\} K_{th}(z_0) + \sum_{t=1}^{T} \overline{e}_{t1}\widetilde{e}_{t1}' K_{th}(z_0) + \sum_{t=1}^{T} \widetilde{e}_{t1}\overline{e}_{t1}' K_{th}(z_0) \\
&+ \sum_{t=1}^{T} \overline{e}_{t1}\overline{e}_{t1}' K_{th}(z_0) + \sum_{t=1}^{T} \left\{\mathbb{E}[\widetilde{e}_{t1}\widetilde{e}_{t1}'] - \mathbb{E}[e_{t1}e_{t1}']\right\} K_{th}(z_0) \\
\equiv &\sum_{t=1}^{T} \left\{\widetilde{e}_{t1}\widetilde{e}_{t1}' - \mathbb{E}[\widetilde{e}_{t1}\widetilde{e}_{t1}']\right\} K_{th}(z_0) + \sum_{k=1}^{4} \mathbf{\Xi}_{Tk}. \qquad \text{(B.1)}
\end{aligned}$$

As $m_T \to \infty$ and $\sum_{j=0}^{\infty} j\|\mathbf{\Phi}_j\| < \infty$ in Assumption 1(i), it is easy to show that as $T \to \infty$,

$$\mathbb{E}\left[\|\overline{e}_{t1}\|\right] = o(1), \quad \left\|\mathbb{E}\left[\widetilde{e}_{t1}\widetilde{e}_{t1}'\right] - \mathbb{E}[e_{t1}e_{t1}']\right\| = o(1). \qquad \text{(B.2)}$$

We also note that the kernel weights $K_{th}(z_0)$ are non-zero only for $T(z_0 - h) \le t \le T(z_0 + h)$ by Assumption 3(i), which together with (B.2) and the Cauchy-Schwarz inequality, we can prove that

$$\left\|\mathbf{\Xi}_{Tk}\right\| = o_P(Th), \quad k = 1, 2, 3, 4. \qquad \text{(B.3)}$$

On the other hand, note that $\{\widetilde{e}_{t1}\widetilde{e}_{t1}'\}$ can be seen as a stationary sequence of m_T-dependent random variables. Hence, by a standard argument, we may prove that

$$\mathbb{E}\left[\left(\sum_{t=1}^{T} \left\{\widetilde{e}_{t1}\widetilde{e}_{t1}' - \mathbb{E}[\widetilde{e}_{t1}\widetilde{e}_{t1}']\right\} K_{th}(z_0)\right)^2\right] = O(m_T Th), \qquad \text{(B.4)}$$

which indicates that

$$\sum_{t=1}^{T} \left\{\widetilde{e}_{t1}\widetilde{e}_{t1}' - \mathbb{E}[\widetilde{e}_{t1}\widetilde{e}_{t1}']\right\} K_{th}(z_0) = O_P\left(\sqrt{m_T Th}\right) = o_P(Th). \qquad \text{(B.5)}$$

Combining (B.1), (B.3) and (B.5), completes the proof of (A.5). □

PROOF OF (A.8). By the weak convergence (2.5) in the main document, we have for $t(z) = \delta(z_0) + \lfloor 2Thz \rfloor$ with $0 \le z \le 1$,

$$\frac{1}{\sqrt{2Th}}\left[X_{t2} - X_{\delta(z_0)2}\right] = \frac{1}{\sqrt{2Th}} \sum_{t=\delta(z_0)}^{\delta(z_0)+\lfloor 2Thz \rfloor} e_{t2} \Rightarrow B_2(z).$$

Then, using Assumption 3 and Theorem 3.1 in Ibragimov and Phillips (2008), we have

$$\begin{aligned}
&\frac{1}{Th}\sum_{t=1}^{T} X_{t1}\left[X_{t2}-X_{\delta(z_0)2}\right]' K_{th}(z_0) \\
= \;& 2\left[\sum_{t=1}^{T}\frac{X_{t1}}{\sqrt{2Th}}\frac{X'_{t2}-X'_{\delta(z_0)2}}{\sqrt{2Th}}K\big(\frac{t-Tz_0}{Th}\big)\right] \\
= \;& 2\left[\sum_{t=\delta(z_0)}^{\delta(z_0)+2Th}\frac{X_{t1}}{\sqrt{2Th}}\frac{X'_{t2}-X'_{\delta(z_0)2}}{\sqrt{2Th}}K\big(\frac{t-Tz_0}{Th}\big)\right] \\
\sim \;& 2\left[\sum_{t(z)=\delta(z_0)}^{\delta(z_0)+2Th}\frac{X_{t(z)1}}{\sqrt{2Th}}\frac{X'_{t(z)2}-X'_{\delta(z_0)2}}{\sqrt{2Th}}K\big(\frac{t(z)-Tz_0}{Th}\big)\right] \quad \big(t=t(z)=\delta(z_0)+\lfloor 2Thz\rfloor\big) \\
\sim \;& 2\left[\int_0^1 K\left(2z-1\right)dB_1(z)[B_2(z)]'dz+\mathbf{\Gamma}_{12}\int_0^1 K\left(2z-1\right)dz\right] \\
\Rightarrow \;& 2\left[\int_{-1}^1 K\left(z\right)dB_1\big(\frac{z+1}{2}\big)\big[B_2\big(\frac{z+1}{2}\big)\big]'+\frac{1}{2}\mathbf{\Gamma}_{12}\int_{-1}^1 K(z)dz\right],
\end{aligned}$$

which completes the proof of (A.8) by noting that $\int_{-1}^1 K(z)dz=1$. □

Appendix C: Numerical illustration via simulation

This appendix reports Monte-Carlo simulations designed to examine the finite sample performance of the proposed methods and illustrate the kernel estimation and inferential limit theory given in Sections 2–5 of the main document. In addition, we also report two figures relevant to the empirics in Section 6 of the main document.

Example C.1. (i) Let $d=3$ and consider the data generating process:

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

where $X_t=(X_{t1},X_{t2}')'$, $X_{t1}=e_{t1}$, $\Delta X_{t2}=(e_{t2},e_{t3})'$, the process $e_t=(e_{t1},e_{t2},e_{t3})'$ is independently generated from a three-dimensional normal distribution with mean zero and variance matrix:

$$\mathbf{\Omega}_1=\begin{pmatrix}1&\rho&\rho\\ \rho&1&\rho\\ \rho&\rho&1\end{pmatrix}\text{ with }\rho=0,\ 0.2\text{ or }0.4,$$

the error term e_{t0} is independently generated from $\mathsf{N}(0,0.5^2)$ and is independent of e_t, the coefficient functions $\boldsymbol{\beta}_t=\boldsymbol{\beta}(t/T)=(\beta_{t1},\beta_{t2},\beta_{t3})'$ are defined as

$$\beta_{t1}=\beta_1(t/T)=1+t/T,\quad \beta_{t2}=\beta_2(t/T)=\sqrt{1+t/T},\quad \beta_{t3}=\beta_3(t/T)=1+(t/T)^2. \tag{C.2}$$

When $\rho=0.2$ or 0.4, the stationary component X_{t1} is correlated with the nonstationary component X_{t2}. We use the local kernel estimation technique $\widehat{\boldsymbol{\beta}}(z)=\left[\widehat{\beta}_1(z),\widehat{\beta}_2(z),\widehat{\beta}_3(z)\right]'$ to estimate the

Table 1: Standard errors of kernel estimates in three different directions for setting (i)

	z	$\rho = 0$			$\rho = 0.2$			$\rho = 0.4$		
T		200	600	1000	200	600	1000	200	600	1000
	0.25	0.1554	0.1248	0.1134	0.1592	0.1234	0.1083	0.1675	0.1246	0.1147
$g^1_{T1}(z)$	0.50	0.1571	0.1322	0.1183	0.1681	0.1331	0.1201	0.1626	0.1290	0.1187
	0.75	0.1822	0.1423	0.1189	0.1792	0.1443	0.1247	0.1839	0.1390	0.1228
	0.25	0.0376	0.0189	0.0132	0.0430	0.0179	0.0147	0.0402	0.0191	0.0133
$g^1_{T2}(z)$	0.50	0.0245	0.0120	0.0095	0.0284	0.0142	0.0102	0.0269	0.0147	0.0102
	0.75	0.0206	0.0101	0.0074	0.0234	0.0114	0.0079	0.0270	0.0120	0.0081
	0.25	0.1294	0.0888	0.0763	0.1324	0.0984	0.0767	0.1590	0.1150	0.0947
$g^1_{T3}(z)$	0.50	0.1158	0.0819	0.0689	0.1179	0.0743	0.0727	0.1246	0.0900	0.0706
	0.75	0.1638	0.1067	0.0865	0.1523	0.1113	0.0837	0.1664	0.1152	0.0864

coefficient functions $\boldsymbol{\beta}(z)$, where the kernel function is chosen as the Epanechnikov kernel $K(u) = \frac{3}{4}(1-u^2)I(-1 \leq u \leq 1)$.

This simulation design serves to illustrate the kernel estimation theory in Section 2. From the above scheme, the global rotation matrix is $\mathbf{H} = \mathbf{I}_3$ and the local rotation matrix is $\overline{\mathbf{Q}}_T(z) = \mathsf{diag}\{1, \mathbf{Q}_{T2}(z)\}$, where $\mathbf{Q}_{T2}(z) = \left[q_{T2}(z), q^{\perp}_{T2}(z)\right]$ with

$$q_{T2}(z) = X_{\lfloor T(z-h)\rfloor,2}/\|X_{\lfloor T(z-h)\rfloor,2}\| =: [p_1(z), p_2(z)]' \text{ and } q^{\perp}_{T2}(z) = [p_2(z), -p_1(z)]'.$$

From the definition of the rotation matrices, the kernel estimates in the three different directions (as discussed in Section 2) can be expressed as

$$\widehat{\beta}_1(z), \quad p_1(z)\widehat{\beta}_2(z) + p_2(z)\widehat{\beta}_3(z), \quad p_2(z)\widehat{\beta}_2(z) - p_1(z)\widehat{\beta}_3(z),$$

respectively. Their corresponding rates of convergence are given in (2.14)–(2.16) of the main document, respectively. In the simulation, we consider three interior points in the kernel estimation ($z = 0.25, 0.5, 0.75$). For the point-wise kernel estimates at these points, we calculate the respective estimated standard errors over 1000 replications of the following transformed and centered quantities:

$$\begin{aligned}
g^1_{T1}(z) &= \widehat{\beta}_1(z) - \beta_1(z), \\
g^1_{T2}(z) &= p_1(z)\left[\widehat{\beta}_2(z) - \beta_2(z)\right] + p_2(z)\left[\widehat{\beta}_3(z) - \beta_3(z)\right], \\
g^1_{T3}(z) &= p_2(z)\left[\widehat{\beta}_2(z) - \beta_2(z)\right] - p_1(z)\left[\widehat{\beta}_3(z) - \beta_3(z)\right].
\end{aligned}$$

We observe from Table 1 that all estimated standard errors decrease as the sample size T increases. Broadly speaking, the standard errors $g^1_{T2}(\cdot)$ are much smaller than those of $g^1_{T3}(\cdot)$,

and the standard errors of g^1_{T1} (in the stationary direction) have the largest values. These results generally support the limit theory and relative convergence rates obtained in Section 2 where $g^1_{T1}(z) = O_P(\frac{1}{\sqrt{Th}})$, $g^1_{T2}(z) = O_P(\frac{1}{T\sqrt{h}})$, and $g^1_{T3}(z) = O_P(\frac{1}{Th})$ for fixed $z \in (0,1)$. Also, we observe that correlation between the stationary component X_{t1} and the nonstationary component X_{t2} does not noticeably impact the finite sample estimation results.

(ii) We next let $d = 2$ and consider mixed stochastic and deterministic trends in generating the nonstationary regressors. Define

$$Y_t = \boldsymbol{\beta}_t' X_t + e_{t0}, \quad t = 1, \cdots, T, \tag{C.3}$$

where $X_t = X_{t-1} + \boldsymbol{\mu} + u_t$, $\boldsymbol{\mu} = (0.1, 0.2)'$, the initial value $X_0 = \mathbf{0}_2$, u_t is independently generated as bivariate normal with mean zero and

$$\boldsymbol{\Omega}_2 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \text{ with } \rho = 0,\ 0.2 \text{ or } 0.5,$$

the error term e_{t0} is defined as in setting (i), and the coefficient functions $\boldsymbol{\beta}_t = \boldsymbol{\beta}(t/T) = (\beta_{t1}, \beta_{t2})'$ are defined by

$$\beta_{t1} = \beta_1(t/T) = 1 + t/T, \quad \beta_{t2} = \beta_2(t/T) = 1 + (t/T)^2. \tag{C.4}$$

The aim of this design is to assess the asymptotic theory of Section 3.1 when the nonstationary regressors have a mixture of stochastic and deterministic trends. From the data generating scheme, we have the local rotation matrix $\widetilde{\mathbf{Q}}_T(z) = \left[\widetilde{q}_T(z), \widetilde{q}_T^{\perp}(z)\right]$ with $\widetilde{q}_T(z) = X_{\lfloor T(z-h)\rfloor}/\|X_{\lfloor T(z-h)\rfloor}\| =: [\widetilde{p}_1(z), \widetilde{p}_2(z)]'$ and $\widetilde{q}_T^{\perp}(z) = [\widetilde{p}_2(z), -\widetilde{p}_1(z)]'$. Let $\widehat{\boldsymbol{\beta}}(z) = \left[\widehat{\beta}_1(z), \widehat{\beta}_2(z)\right]'$ be the kernel regression estimator of $\boldsymbol{\beta}(z)$. The kernel estimates in the two different directions (as discussed in Section 3.1) are

$$\widetilde{p}_1(z)\widehat{\beta}_1(z) + \widetilde{p}_2(z)\widehat{\beta}_2(z) \ \text{ and } \ \widetilde{p}_2(z)\widehat{\beta}_1(z) - \widetilde{p}_1(z)\widehat{\beta}_2(z),$$

respectively. They have convergence rates given in (3.10) and (3.11), respectively. We use the same interior points ($z = 0.25, 0.5, 0.75$) as in setting (i) and calculate the estimated standard errors over 1000 replications of the following quantities:

$$\begin{aligned} g^2_{T1}(z) &= \widetilde{p}_1(z)\left[\widehat{\beta}_1(z) - \beta_1(z)\right] + \widetilde{p}_2(z)\left[\widehat{\beta}_2(z) - \beta_2(z)\right], \\ g^2_{T2}(z) &= \widetilde{p}_2(z)\left[\widehat{\beta}_1(z) - \beta_1(z)\right] - \widetilde{p}_1(z)\left[\widehat{\beta}_2(z) - \beta_2(z)\right]. \end{aligned}$$

From Table 2, it is clear that the values in both directions approach zero as the sample size increases, and $g^2_{T1}(z)$ evidently converges at a rate faster than $g^2_{T2}(z)$, as indicated in Section 3.1 where $g^2_{T1}(z) = O_P(\frac{1}{T\sqrt{Th}} + h^\gamma)$ and $g^2_{T2}(z) = O_P(\frac{1}{Th} + h^\gamma)$ for fixed $z \in (0,1)$. As in setting (i), Table 2 shows that correlation within the nonstationary regressors X_t has little impact on the results.

Table 2: Standard errors of kernel estimates in two different directions for setting (ii)

	z	$\rho = 0$			$\rho = 0.2$			$\rho = 0.5$		
n		200	600	1000	200	600	1000	200	600	1000
	0.25	0.0509	0.0091	0.0047	0.0568	0.0110	0.0052	0.0529	0.0125	0.0053
$g^2_{T1}(z)$	0.50	0.0270	0.0045	0.0026	0.0237	0.0046	0.0026	0.0244	0.0054	0.0027
	0.75	0.0136	0.0034	0.0020	0.0167	0.0033	0.0021	0.0165	0.0035	0.0020
	0.25	0.3628	0.2612	0.2495	0.4173	0.2912	0.2650	0.4803	0.3501	0.2925
$g^2_{T2}(z)$	0.50	0.4046	0.3460	0.3186	0.4456	0.3602	0.3258	0.5399	0.4396	0.4333
	0.75	0.5627	0.4743	0.4673	0.5719	0.5444	0.5221	0.7106	0.6611	0.6197

(iii) As in setting (i), we let $d = 3$ and consider the data generating process (C.1) with $X_t = (X_{t1}, X'_{t2})'$, where

$$\Delta X_{t1} = e_{t1} \quad \text{and} \quad X_{t2} = \begin{pmatrix} 0.05 \\ 0.15 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.3 \\ 0 & 1 \end{pmatrix} \mathbf{X}_{t-1,2} + \begin{pmatrix} e_{t2} \\ e_{t3} \end{pmatrix}.$$

It can be shown that X_{t2} is cointegrated with a linear deterministic trend. The processes $e_t = (e_{t1}, e_{t2}, e_{t3})'$ and e_{t0} are defined as in setting (i), and the coefficient functions $\boldsymbol{\beta}_t = \boldsymbol{\beta}(t/T) = (\beta_{t1}, \beta_{t2}, \beta_{t3})'$ are defined in the same way as before.

This design is intended to assess the limit theory in Section 3.2 where the nonstationary regressors are cointegrated with deterministic trends. The global rotation matrix is $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$, where

$$\mathbf{H}_1 = (0, 5/\sqrt{34}, -3/\sqrt{34})' \quad \text{and} \quad \mathbf{H}_2 = \begin{pmatrix} 0 & 3/\sqrt{34} & 5/\sqrt{34} \\ 1 & 0 & 0 \end{pmatrix}'.$$

The local rotation matrix $\overline{\mathbf{Q}}_T(z)$ is the same as that in setting (i) with $\mathbf{H}'_2 X_t$ replacing X_{t2}. The kernel estimates in three different directions (as discussed in Section 3.2) are $\mathbf{H}'_1\widehat{\beta}(z)$, $q_{T2}(z)'\mathbf{H}'_2\widehat{\beta}(z)$, and $q^{\perp}_{T2}(z)'\mathbf{H}'_2\widehat{\beta}(z)$, respectively. As in settings (i) and (ii), we consider three interior points: $z = 0.25, 0.5, 0.75$, and then calculate the standard errors over 1000 replications of the following quantities:

$$\begin{aligned}
g^3_{T1}(z) &= 5/\sqrt{34}\left(\widehat{\beta}_2(z) - \beta_2(z)\right) - 3/\sqrt{34}\left(\widehat{\beta}_3(z) - \beta_3(z)\right), \\
g^3_{T2}(z) &= p_1(z)\left[3/\sqrt{34}\left(\widehat{\beta}_2(z) - \beta_2(z)\right) + 5/\sqrt{34}\left(\widehat{\beta}_3(z) - \beta_3(z)\right)\right] + p_2(z)\left[\widehat{\beta}_1(z) - \beta_1(z)\right], \\
g^3_{3T}(z) &= p_2(z)\left[3/\sqrt{34}\left(\widehat{\beta}_2(z) - \beta_2(z)\right) + 5/\sqrt{34}\left(\widehat{\beta}_3(z) - \beta_3(z)\right)\right] - p_1(z)\left[\widehat{\beta}_1(z) - \beta_1(z)\right].
\end{aligned}$$

From Table 3, all estimated standard errors decrease as the sample size increases. Broadly speaking, the values of $g^3_{T2}(z)$ converge to zero much faster than $g^3_{T1}(z)$ and $g^3_{T3}(z)$, and $g^3_{T1}(z)$ looks to have the largest value, supporting the limit results in Section 3.2. In addition, the presence

Table 3: Standard errors of kernel estimates in three different directions for setting (iii)

	z	$\rho = 0$			$\rho = 0.2$			$\rho = 0.5$		
n		200	600	1000	200	600	1000	200	600	1000
	0.25	0.3452	0.2706	0.2401	0.4086	0.3085	0.2739	0.4813	0.3704	0.3264
$g^3_{T1}(z)$	0.50	0.3672	0.2993	0.2665	0.4107	0.3288	0.2977	0.5561	0.3925	0.3416
	0.75	0.4130	0.3910	0.3541	0.4755	0.4171	0.3699	0.6056	0.4907	0.4578
	0.25	0.0518	0.0158	0.0084	0.0555	0.0135	0.0068	0.0522	0.0141	0.0062
$g^3_{T2}(z)$	0.50	0.0243	0.0063	0.0035	0.0244	0.0075	0.0034	0.0273	0.0061	0.0033
	0.75	0.0191	0.0052	0.0026	0.0173	0.0047	0.0027	0.0212	0.0047	0.0027
	0.25	0.1629	0.0945	0.0613	0.1853	0.0853	0.0718	0.2071	0.0993	0.0682
$g^3_{T3}(z)$	0.50	0.2933	0.2324	0.2204	0.3012	0.2419	0.2189	0.3607	0.2602	0.2314
	0.75	0.3999	0.3573	0.3385	0.3965	0.3660	0.3605	0.4570	0.4000	0.3904

of the correlation between the stationary component X_{t1} and the nonstationary component X_{t2} still has no impact on the kernel estimation results.

(iv) In the above three settings, the error e_{t0} is independent of the nonstationary regressors, excluding the existence of endogeneity. We now relax this restriction when generating the simulated data. As shown in Section 2.3, the correlation between e_{t0} and the stationary regressors often leads to estimation inconsistency. We now consider the setting (i) with the only difference that e_{t0} and $(e_{t2}, e_{t3})'$ are jointly determined by a three-dimensional normal distribution with mean zero and the covariance matrix:

$$\mathbf{\Omega}_3 = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix},$$

and e_{t1} is independent of $(e_{t0}, e_{t2}, e_{t3})'$.

We compare finite sample performance between the conventional kernel estimates and the FM kernel estimates of Section 4. We use the same interior points ($z = 0.25, 0.5, 0.75$) and calculate averages of the following quantities over 1000 replications:

$$\begin{aligned} g^{\#}_{T1}(z) &= \widehat{\beta}^{\#}_1(z) - \beta_1(z), \\ g^{\#}_{T2}(z) &= p_1(z)\left[\widehat{\beta}^{\#}_2(z) - \beta_2(z)\right] + p_2(z)\left[\widehat{\beta}^{\#}_3(z) - \beta_3(z)\right], \\ g^{\#}_{T3}(z) &= p_2(z)\left[\widehat{\beta}^{\#}_2(z) - \beta_2(z)\right] - p_1(z)\left[\widehat{\beta}^{\#}_3(z) - \beta_3(z)\right], \end{aligned}$$

Table 4: Averages of $g^{\#}_{Ti}(z)$ and $g^{4}_{Ti}(z)$ over 1000 replications for setting (iv)

	z	FM-kernel				z	Kernel		
T		200	600	1000			200	600	1000
	0.25	0.01368	0.00401	0.00002		0.25	0.01229	0.00198	-0.00009
$g^{\#}_{T1}(z)$	0.50	0.00687	0.00477	0.00443	$g^{4}_{T1}(z)$	0.50	0.00803	0.00550	0.00429
	0.75	-0.01006	-0.00884	0.00536		0.75	-0.01451	-0.00938	0.00507
	0.25	0.00188	0.00073	0.00065		0.25	-0.00053	0.00037	0.00085
$g^{\#}_{T2}(z)$	0.50	0.00042	0.00070	0.00056	$g^{4}_{T2}(z)$	0.50	0.00024	0.00062	-0.00060
	0.75	0.00044	0.00033	0.00037		0.75	-0.00051	0.00036	-0.00044
	0.25	-0.00573	-0.00343	-0.00184		0.25	-0.00526	-0.00523	0.00298
$g^{\#}_{T3}(z)$	0.50	-0.00345	-0.00198	0.00050	$g^{4}_{T3}(z)$	0.50	-0.00132	-0.00323	0.00173
	0.75	0.00116	0.00302	0.00051		0.75	-0.00180	0.00358	0.00092

where $\widehat{\boldsymbol{\beta}}_{\#}(z) = \left[\widehat{\beta}^{\#}_1(z), \widehat{\beta}^{\#}_2(z), \widehat{\beta}^{\#}_3(z)\right]'$ is defined in (4.8). The estimate of the long run covariance matrix is constructed under $l_T = \lfloor T^{1/4} \rfloor$, $\tau = 1/4$, and $k(x) = I(-1 \leq x \leq 1)$. Accordingly, we define $g^{4}_{Ti}(z)$ in a similar way as $g^{\#}_{Ti}(z)$, with $\widehat{\beta}^{\#}_i(z)$ being replaced by the conventional kernel estimate $\widehat{\beta}_i(z)$ for $i = 1, 2, 3$.

The simulation results are reported in Table 4 for both the FM and conventional kernel methods. The methods exhibit similar finite sample performance in the three directions. The results show that the FM-kernel method improves estimation well when $T = 1000$, but has similar performance to the conventional kernel estimation when the sample size is small or moderate. Broadly speaking, the simulation shows that reasonably large samples are needed to realize gains from the use of FM kernel methods in the present setting. These findings generally support the limit theory in Section 4.

Example C.2. We use the data generating process

$$Y_t = \boldsymbol{\beta}_t' X_t + e_{t0}, \quad \boldsymbol{\beta}_t \equiv \boldsymbol{\beta}_0 = (2, 2, 4)', \quad t = 1, \cdots, T, \tag{C.5}$$

where $X_t = (X_{t1}, X_{t2})'$, $X_{t1} = (e_{t1}, e_{t2})'$, $\Delta X_{t2} = e_{t3}$, $e_t = (e_{t1}, e_{t2}, e_{t3})'$ and e_{t0} are defined as in setting (i) with $\rho = 0$ or 0.2. This simulation is designed to examine the size performance of the nonparametric test statistics proposed in Section 5 of the main document.

Consider the global null hypothesis:

$$\mathcal{H}_0 : \ \mathbf{R}\left(\boldsymbol{\beta}_t - \boldsymbol{\beta}_0\right) = \mathbf{0}_2, \quad \mathbf{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

which is equivalent to testing $\beta_{t1} = \beta_{t2} \equiv 2$. In this case, it is easy to verify the rank condition (5.5) with $r = 2$. So we can use the generalized Wald test statistic constructed in (5.4) and (5.7), and examine the size performance in finite samples.

As in (5.10), we also construct the following local null hypothesis:

$$\mathcal{H}_0^{\diamond}: \ \mathbf{R}(z)\left(\boldsymbol{\beta}_t - \boldsymbol{\beta}_0\right) = \mathbf{0}_2, \ \ \mathbf{R}(z) = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & q_T(z) \end{pmatrix}, \ \ q_T(z) = X_{\lfloor T(z-h)\rfloor,2}/|X_{\lfloor T(z-h)\rfloor,2}|.$$

In this case, the rank condition (5.5) fails, so we adopt the generalized Wald test statistic defined in (5.13) and (5.14) and examine its size performance.

A bootstrap scheme is introduced below to generate the critical values l_α^* where $\alpha = 1\%$, 5%, and 10%, as it is well known that bootstrap critical values usually outperform the asymptotic critical values in finite samples. The number of Monte-Carlo replication is $M = 1000$ and that of the bootstrap procedure is $M_b = 250$.

1. Generate simulated data $\{(X_t^{(i)}, Y_t^{(i)}), 1 \le t \le T\}$, for $i = 1, \cdots, M$. Let $\bar{Y}_t = \frac{1}{M}\sum_{i=1}^M Y_t^{(i)}$ and $\bar{X}_t = \frac{1}{M}\sum_{i=1}^M X_t^{(i)}$. Given a specific bandwidth h, estimate β_t by the kernel local level regression technique $\widehat{\beta}_t$ based on $\{(\bar{X}_t, \bar{Y}_t), 1 \le t \le T\ \}$. Let $\{z_k\}_{k=1}^m$ be an equidistant grid of points between $z_1 = h$ and $z_m = 1 - h$ with the number of the grid points $m = \lfloor \frac{1-2h}{2h} \rfloor$, satisfying the distance assumption required for asymptotically independent components $\{W_T(z_k) : k = 1, \cdots, m\}$.

2. Let $\widehat{e}_t = \bar{Y}_t - \bar{X}_t'\widehat{\beta}_t$. Generate the bootstrap residuals $\{e_t^*\}$ by a regression bootstrap method of the form $e_t^* = \widehat{e}_t \eta_t^*$, where $\{\eta_t^*, 1 \le t \le T\}$ is a sequence of i.i.d random variables drawn from a two-point distribution of the form

$$\mathsf{P}\left(\eta_1^* = \frac{-(\sqrt{5}-1)}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}} \ \text{ and } \ \mathsf{P}\left(\eta_1^* = \frac{\sqrt{5}+1}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}. \tag{C.6}$$

3. Generate $Y_t^* = \bar{X}_t'\widehat{\beta}_t + e_t^*$. Use the dataset $\{(\bar{X}_t, Y_t^*), 1 \le t \le T\}$ to re-estimate β_t using the bandwidth h_{cv} obtained from the first bootstrap sample by cross-validation, and denote the resulting estimate by $\widehat{\beta}_t^*$. Construct the equation error variance estimate $\hat{\sigma}_e^{*2} = \frac{1}{\lfloor(1-\tau)T\rfloor - \lfloor \tau T\rfloor}\sum_{t=\lfloor \tau T\rfloor+1}^{\lfloor(1-\tau)T\rfloor} \widehat{e}_t^{*2}$ with $\tau = 1/4$ and $\widehat{e}_t^* = Y_t^* - \bar{X}_t'\widehat{\beta}_t^*$. Compute W_T^* as a corresponding version of W_T with $\widehat{\beta}_t$, $\widehat{\sigma}_e^2$, and $\{(X_t, Y_t), 1 \le t \le T\}$ being replaced by $\widehat{\beta}_t^*$, $\widehat{\sigma}_e^{*2}$, and $\{(\bar{X}_t, Y_t^*), 1 \le t \le T\}$ in the expression of W_T.

4. Repeat steps 2 and 3 $M_b = 250$ times and thereby produce M_b versions of W_T^*, signified by W_{Tj}^* for $j = 1, \cdots, M_b$. Use the M_b values of W_{Tj}^* to construct empirical bootstrap distribution. The bootstrap distribution of W_T^* given $\mathcal{S}_T = \{(\bar{X}_t, \bar{Y}_t), 1 \le t \le T\}$ is defined by $\mathsf{P}^*(W_T^* < x) = \mathsf{P}(W_T^* < x|\mathcal{S}_T)$. Let l_α^* be the quantile such that $\mathsf{P}^*(W_T^* \ge l_\alpha^*) = \alpha$ and then l_α^* is used to approximate the critical value l_α.

5. Obtain the size function by

$$\alpha_T^* = \mathsf{P}(W_T \ge l_\alpha^*|\mathcal{H}_0) \ \text{ or } \ \alpha_T^* = \mathsf{P}(W_T \ge l_\alpha^*|\mathcal{H}_0^{\diamond}). \tag{C.7}$$

Table 5: Empirical size performance under the global null hypothesis

T	$\rho = 0$			$\rho = 0.2$		
	1%	5%	10%	1%	5%	10%
200	0.0178	0.0588	0.0936	0.0188	0.0568	0.0922
400	0.0134	0.0520	0.0957	0.0141	0.0527	0.0935
1000	0.0108	0.0510	0.0982	0.0101	0.0504	0.0976

Table 6: Empirical size performance under the local null hypothesis

T	$\rho = 0$			$\rho = 0.2$		
	1%	5%	10%	1%	5%	10%
200	0.0149	0.0523	0.0946	0.0168	0.0584	0.0946
400	0.0118	0.0514	0.0949	0.0148	0.0549	0.0978
1000	0.0105	0.0504	0.0989	0.0100	0.0507	0.0987

The sample sizes considered are $T = 200$, 400, and 1000. Since the size function defined in equation (C.7) is also a function of the significance level α, different bandwidths $\{h_i : i = 1, 2, 3\}$ are used in simulations with respect to $\alpha = 1\%$, 5%, and 10%, respectively. The simulation results of the empirical size for both the global and local null hypothesis tests are summarized in Tables 5 and 6. We observe that almost all the calculated sizes fluctuate around the given significance level. They appear to be oversized at the 1% and 5% significance level and undersized at 10% for both hypothesis tests, but these size distortions decrease as the sample size increases. The presence of the correlation between the stationary component X_{t1} and the nonstationary component X_{t2} again has no noticeable impact on size performance.

Additional Figures for the Empirical Example in Section 6

Figure S.1 shows the estimated residuals from the time-varying coefficient model (6.1) and the linear model (6.3), respectively. 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. Section 6 in the main document introduces a local level kernel estimation method to estimate the possibly nonlinear trend function for $\widehat{v}_t$. Figure S.2 plots the detrended residuals $\widehat{u}_t$ (from the linear consumption function) against the residuals $\widehat{e}_t$ from the time varying coefficient consumption function.

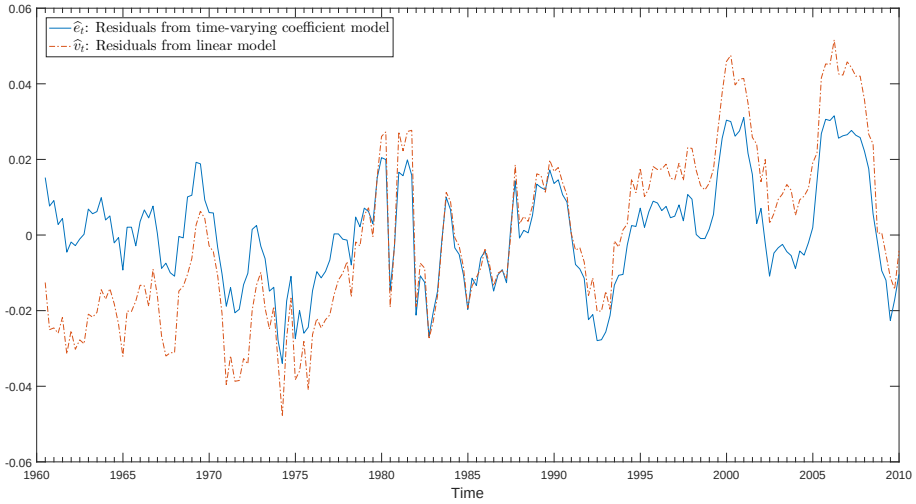

Figure S.1: Estimated residuals from the time-varying coefficient model and the linear model

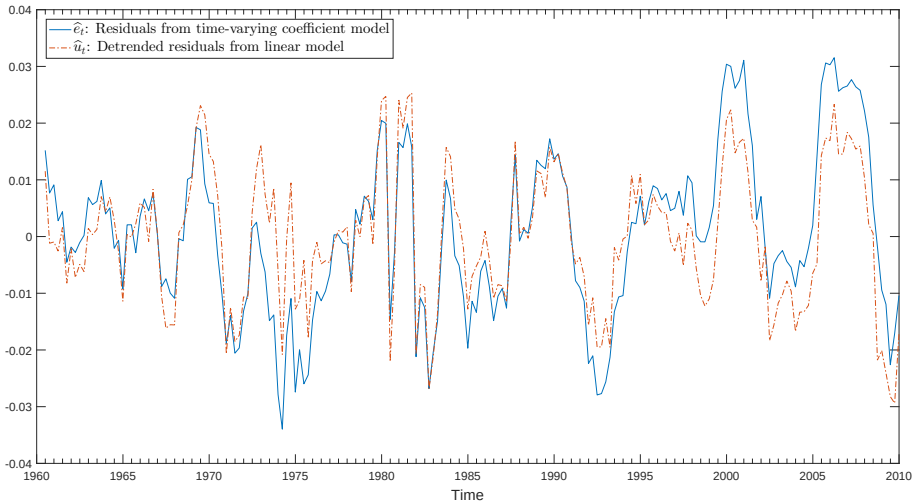

Figure S.2: Estimated residuals from the time-varying coefficient model and detrended residuals from the linear model

References

Ibragimov, R. and Phillips, P. C. B. (2008). Regression asymptotics using martingale convergence methods. *Econometric Theory* **24**, 888–947.