Series Estimation of Cointegrated System with Time Varying Coefficients

Yu Bai*, Jiti Gao, Hsein Kew

November 2025

Faculty of Finance, City University of Macau.

Cointegration

→ The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003 was divided equally between Robert F. Engle III "for methods of analyzing economic time series with time-varying volatility (ARCH)" and Clive W.J. Granger "for methods of analyzing economic time series with common trends (cointegration)"

Cointegration

$$y_t = \mathbf{X}_t^\top \boldsymbol{\beta} + u_{0t},$$
$$\mathbf{X}_t = \mathbf{X}_{t-1} + \mathbf{u}_{xt}, \quad t = 1, 2, \cdots, n,$$

where

→ $\mathbf{X}_t$: $K \times 1$-dimensional $I(1)$ vector;

→ $\mathbf{u}_t = \left(u_{0t}, \mathbf{u}_{xt}^\top\right)^\top$: $(K+1) \times 1$-dimensional vector of stationary process.

The theory of OLS estimator $\hat{\boldsymbol{\beta}}_n = \left(\sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t^\top\right)^{-1} \left(\sum_{t=1}^n \mathbf{X}_t \mathbf{y}_t\right)$ is well understood in the literature. See, for instance, Stock, 1987, Phillips and Hansen, 1990, and many references therein.

Cointegration with time-varying coefficients

→ What if $\boldsymbol{\beta}$ varies over time?

→ To remain agnostic on the forms of time variation, we may consider the following kernel estimator for any $u \in (0,1)$:

$$\hat{\boldsymbol{\beta}}_n(u) = \left(\sum_{t=1}^{n} K\left(\frac{t-nu}{nh}\right) \mathbf{X}_t \mathbf{X}_t^\top \right)^{-1} \left(\sum_{t=1}^{n} K\left(\frac{t-nu}{nh}\right) \mathbf{X}_t \mathbf{y}_t \right),$$

where $K(\cdot)$ is some kernel function, and $h := h_n$ is a bandwidth parameter.

→ Just a rolling window estimator if $K(x) = \frac{1}{2}\mathbb{1}_{\{-1 \le x \le 1\}}$.

Cointegration with time-varying coefficients

→ As shown in Phillips, Li, and Gao, 2017, under certain regularity conditions, as $nh \to \infty$

$$\frac{1}{n^2 h} \sum_{t=1}^{n} K\left(\frac{t-nu}{nh}\right) \mathbf{X}_t \mathbf{X}_t^\top \xrightarrow{d} u \mathbf{B}_u(\cdot) \mathbf{B}_u^\top(\cdot),$$

where $\mathbf{B}_u(\cdot)$ is a K-dimensional Brownian motion with covariance matrix $\cdot$.

→ $\mathbf{B}_u(\cdot)\mathbf{B}_u^\top(\cdot)$: a Wishart variate with 1 degree of freedom.

→ So, unless $K = 1$, the signal matrix $\frac{1}{nh} \sum_{t=1}^{n} K\left(\frac{t-nu}{nh}\right) \mathbf{X}_t \mathbf{X}_t^\top$ becomes asymptotically singular.

Cointegration with time-varying coefficients

→ To overcome this technical difficulty, Phillips, Li, and Gao, 2017 introduce a novel rotational decomposition approach.

→ Let $\mathbf{b}_n \equiv b_{nu} = \frac{1}{\sqrt{n}}\mathbf{X}_{\lfloor (u-h)n \rfloor}$ and $\mathbf{q}_n = \frac{\mathbf{b}_n}{\|\mathbf{b}_n\|}$. Define

$$\mathbf{Q}_n = \left(\mathbf{q}_n, \mathbf{q}_n^{\perp}\right), \quad \mathbf{D}_n = \text{diag}\{n\sqrt{h}, (nh)\mathbf{I}_{K-1}\}.$$

→ Under certain regularity conditions, Phillips, Li, and Gao, 2017 establish the asymptotic distribution on

$$\mathbf{D}_n \mathbf{Q}_n^{\top} \left(\hat{\boldsymbol{\beta}}_n(u) - \boldsymbol{\beta}(u)\right).$$

→ This is nice, but ...

What we do?

- → We propose a series estimation method for cointegrated system with time varying coefficients.
- → We develop a test for structure change in cointegrated system.

Related literature

- → Kernel-based estimation: Phillips, Li, and Gao, 2017, Li, Phillips, and Gao, 2020
- → Kernel-based instability test, single equation framework: Cai, Yunfei Wang, and Yonggang Wang, 2015
- → Series estimation, single-equation framework: Park and Hahn, 1999
- → Time-varying VECM: Bierens and Martins, 2010

Model

Phillips, 1991:

$$\underbrace{\mathbf{y}_t}_{d\times 1} = \mathbf{A}_t \underbrace{\mathbf{X}_t}_{K\times 1} + \mathbf{u}_{0t}, \tag{1}$$

$$\mathbf{X}_t = \mathbf{X}_{t-1} + \mathbf{u}_{xt}, \quad t = 1, 2, \cdots, n, \tag{2}$$

where

→ $\mathbf{A}_t := \mathbf{A}(t/n)$: an $d \times K$ matrix of coefficients

→ $(\mathbf{X}_t)_t$: initialized at some $\mathbf{X}_0 = O_p(1)$, with $\mathbb{E}\,\|\mathbf{X}_0\|^8 < \infty$

Model

→ $\mathbf{u}_t = (\mathbf{u}_{0t}^\top, \mathbf{u}_{xt}^\top)^\top$ are determined according to the linear process

$$\mathbf{u}_t = \mathbf{\Phi}(\mathcal{L})\boldsymbol{\varepsilon}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \boldsymbol{\varepsilon}_{t-j}, \tag{3}$$

where $\mathbf{\Phi}(\mathcal{L}) = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \mathcal{L}^j$. ▸ details

Assumption 1

(i) $(\boldsymbol{\varepsilon}_t)_t$ is an $i.i.d.$ $(d+K)$-dimensional random vector with $\mathbb{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$, $\mathbf{\Sigma}_\varepsilon := \mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') > 0$, and $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^p < \infty$, for some $p > 8$. (ii) The coefficient matrices in (3) satisfy $\sum_{j=0}^{\infty} j^{\gamma_0} \|\mathbf{\Phi}_j\| < \infty$, for some $\gamma_0 > 2$.

Model

→ Our objective: to construct an estimator for the time-varying cointegrating functional coefficients $\mathbf{A}(t/n)$

→ Let $\mathbf{A}^{(i,j)}(t/n)$ be the (i, j)th elements in $\mathbf{A}(t/n)$, where $i = 1, 2, \cdots, d$, $j = 1, 2, \cdots, K$.

Assumption 2

$\mathbf{A}^{(i,j)}(\cdot)$ is a squared integrable real function:
$\mathbf{A}^{(i,j)}(\cdot) \in L^2(0,1) = \left\{a(r) : \int_0^1 a^2(r)dr < \infty\right\}$ and is $q+1$ times continuously differentiable on $[0,1]$, where $q > \frac{2p}{p-2}$ and p is given in Assumption 1(i).

Series estimator and its asymptotic properties

→ Under Assumption 2, $\mathbf{A}(t/n)$ admits an orthogonal expansion

$$\mathbf{A}(t/n) = \sum_{i=0}^{\infty} \mathbf{B}_i \psi_i(t/n),$$

where $\mathbf{B}_i$ is of dimension $d \times K$.

→ $(\psi_i(t/n))_i$ are Chebyshev time polynomials, which are defined by

$$\psi_0(t/n) = 1, \quad \psi_i(t/n) = \sqrt{2} \cos\left(i\pi t/n\right), \tag{4}$$

where $i = 1, 2, 3, \cdots, m-1$.

- first used in Bierens, 1997, see also Bierens and Martins, 2010

Series estimator and its asymptotic properties

→ Define a truncated series

$$\mathbf{A}_m(t/n) = \sum_{i=0}^{m-1} \mathbf{B}_i^{(m)} \psi_i(t/n) = \mathbf{B}^{(m)} \left(\boldsymbol{\psi}_{(m)}(t/n) \otimes \mathbf{I}_K \right), \tag{5}$$

where $\boldsymbol{\psi}_{(m)}(t/n) = (\psi_0(t/n), \psi_1(t/n), \cdots, \psi_{m-1}(t/n))^\intercal$.

→ It can be shown that

$$\left\| \text{vec}\left(\mathbf{A}(r) - \mathbf{A}_m(r) \right) \right\| = O\left(\frac{1}{m^q} \right), \tag{6}$$

for any $r \in [0, 1]$.

- $\mathbf{A}(t/n)$ can be well approximated by (5), and the approximation error goes to zero as $m \to \infty$.

→ Approximated model:

$$\mathbf{y}_t = \sum_{i=0}^{m-1} \mathbf{B}_i \psi_i(t/n) \mathbf{X}_t + \mathbf{u}_{0t}^{(m)}, \tag{7}$$

where $\mathbf{u}_{0t}^{(m)} = \mathbf{u}_{0t} + \left(\mathbf{A}_t - \mathbf{A}_m(t/n)\right) \mathbf{X}_t$.

→ Define

$$\underset{mK\times 1}{\mathbf{X}_t^{(m)}} = (\mathbf{X}_t^{\intercal}, \psi_1(t/n)\mathbf{X}_t^{\intercal}, \cdots, \psi_{m-1}(t/n)\mathbf{X}_t^{\intercal})^{\intercal}$$

$$\underset{d\times mK}{\mathbf{B}^{(m)}} = (\mathbf{B}_0, \mathbf{B}_1, \cdots, \mathbf{B}_{m-1}).$$

We can write (7) more conveniently as

$$\mathbf{y}_t = \mathbf{B}^{(m)}\mathbf{X}_t^{(m)} + \mathbf{u}_{0t}^{(m)}. \tag{8}$$

Series estimator and its asymptotic properties

→ $\mathbf{B}^{(m)}$ can thus be estimated by LS (least squares):

$$\hat{\mathbf{B}}_n^{(m)} = \left[\sum_{t=1}^{n} \mathbf{y}_t \left(\mathbf{X}_t^{(m)}\right)^\intercal\right] \left[\sum_{t=1}^{n} \mathbf{X}_t^{(m)} \left(\mathbf{X}_t^{(m)}\right)^\intercal\right]^{-1}.$$

→ For a given $u \in [0,1]$, we have the series estimator

$$\hat{\mathbf{A}}_n^{(m)}(u) = \hat{\mathbf{B}}_n^{(m)} \mathbf{T}_{(m)}(u), \tag{9}$$

where $\mathbf{T}_{(m)}(u) = \boldsymbol{\psi}_{(m)}(u) \otimes \mathbf{I}_K$.

Assumption 3

$m = \lfloor c \cdot n^\kappa \rfloor$ is a sequence of integers for some $c > 0$, where $\frac{1}{q} < \kappa < \frac{1}{2} - \frac{1}{p}$.

Theorem

Suppose that Assumptions 1-3 are satisfied. For any given $u \in [0,1]$, we have, as $n \to \infty$

$$n \cdot \mathit{vec}\left(\hat{\mathbf{A}}_n^{(m)}(u) - \mathbf{A}(u)\right) = \left(\mathbf{T}_{(m)}^{\intercal}(u) \otimes \mathbf{I}_d\right)\left(\boldsymbol{\Lambda}_{(m),xx}^{-1} \otimes \mathbf{I}_d\right)\boldsymbol{\Gamma}_{(m),ux} + o_p(1), \qquad (10)$$

where

$$\boldsymbol{\Lambda}_{(m),xx} = \int_0^1 \left(\boldsymbol{\psi}_{(m)}(r)\boldsymbol{\psi}_{(m)}^{\intercal}(r) \otimes \left(\mathbf{B}_r(\boldsymbol{\Omega}_{xx})\mathbf{B}_r^{\intercal}(\boldsymbol{\Omega}_{xx})\right)\right) dr,$$

$$\boldsymbol{\Gamma}_{(m),ux} = \int_0^1 \left(\left(\boldsymbol{\psi}_{(m)}(r) \otimes \mathbf{B}_r(\boldsymbol{\Omega}_{xx})\right) \otimes \mathbf{I}_d\right) d\mathbf{B}_r(\boldsymbol{\Omega}_{00}) + \begin{bmatrix} \mathit{vec}\left(\boldsymbol{\Lambda}_{0x} + \boldsymbol{\Sigma}_{0x}\right) \\ \mathbf{0}_{dK(m-1)\times 1} \end{bmatrix}.$$

→ Define $\hat{\mathbb{V}}_{n(m)}(u) = \left\{ \mathbf{T}_{(m)}^{\intercal}(u) \left(\frac{1}{n^2} \mathbf{\Lambda}_{n(m),xx} \right)^{-1} \mathbf{T}_{(m)}(u) \right\} \otimes \hat{\mathbf{\Omega}}_{n,00}$, where $\hat{\mathbf{\Omega}}_{n,00}$ is a consistent estimator of $\mathbf{\Omega}_{00}$.

Corollary

Suppose that conditions in Theorem 1 are satisfied and $\mathbf{\Lambda}_{0x} = \mathbf{0}_{dK}$, $\mathbf{\Sigma}_{0x} = \mathbf{0}_{dK}$. We then have

$$n \left(\hat{\mathbb{V}}_{n(m)}(u) \right)^{-1/2} vec \left(\hat{\mathbf{A}}_n^{(m)}(u) - \mathbf{A}(u) \right) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}_{dK \times 1}, \mathbf{I}_{dK} \right), \tag{11}$$

for any fixed $u \in [0, 1]$.

Series estimator and its asymptotic properties

- $\to$ $\left\|\hat{\mathbb{V}}_{n(m)}(u)\right\| = O(1)m$, the order involved in the asymptotic normality is $O_p(1)\,\dfrac{n}{\sqrt{m}}$.
- $\to$ The convergence rate is comparable with the kernel estimate in the literature.
- $\to$ As shown in Phillips, Li, and Gao, 2017, the *type I* super convergence rate is given by $nh^{1/2}$, where h is the bandwidth parameter.
- $\to$ Thinking of m^{-1} as equivalent to the bandwidth h, the convergence rate in the two situations are quite comparable.

Introduction

Model and estimation theory

Testing for structural changes

Monte Carlo study

An empirical application

Conclusion

Construction of the test statistic

→ The null hypothesis is

$$\mathbb{H}_0 : \; \mathbf{A}_t = \mathbf{A}, \text{ for some constant coefficient matrix } \mathbf{A} \text{ and for all } t. \tag{12}$$

The alternative hypothesis is $\mathbb{H}_A : \; \mathbb{H}_0$ is false.

→ We initiate our test statistic as a L_2-type test statistic:

$$\int_0^1 \left[\text{vec}\left(\hat{\mathbf{A}}_n^{(m)}(u) - \hat{\mathbf{A}}_n\right)\right]^{\intercal} \left[\text{vec}\left(\hat{\mathbf{A}}_n^{(m)}(u) - \hat{\mathbf{A}}_n\right)\right] du, \tag{13}$$

where $\hat{\mathbf{A}}_n$ is the OLS estimator for (1) under $\mathbb{H}_0$:

$$\hat{\mathbf{A}}_n = \left[\sum_{t=1}^{n} \mathbf{y}_t \mathbf{X}_t^{\intercal}\right] \left[\sum_{t=1}^{n} \mathbf{X}_t \mathbf{X}_t^{\intercal}\right]^{-1}. \tag{14}$$

Construction of the test statistic

→ To avoid the random denominator issue, we modify the test statistic as follows.

→ First, modify $\hat{\mathbf{B}}_n^{(m)}$ with $\hat{\mathbf{B}}_n^{(m)} \left[\sum_{t=1}^n \mathbf{X}_t^{(m)} \left(\mathbf{X}_t^{(m)}\right)^\intercal\right]$ to construct a modified version of $\hat{\mathbf{A}}_n^{(m)}(u)$:

$$\tilde{\mathbf{A}}_n^{(m)}(u) := \left[\sum_{t=1}^n \mathbf{y}_t \left(\mathbf{X}_t^{(m)}\right)^\intercal\right] \left(\boldsymbol{\psi}_{(m)}(u) \otimes \mathbf{I}_K\right).$$

→ Analogously, we may also have a similar version for $\hat{\mathbf{A}}_n$:

$$\tilde{\mathbf{A}}_n := \hat{\mathbf{A}}_n \left[\sum_{t=1}^n \mathbf{X}_t \left(\mathbf{X}_t^{(m)}\right)^\intercal\right] \left(\boldsymbol{\psi}_{(m)}(u) \otimes \mathbf{I}_K\right).$$

Construction of the test statistic

→ Instead of comparing the distance between $\hat{\mathbf{A}}_n^{(m)}(u)$ and $\hat{\mathbf{A}}_n$, we measure the distance between $\tilde{\mathbf{A}}_n^{(m)}(u)$ and $\tilde{\mathbf{A}}_n$:

$$\int_0^1 \left[\text{vec}\left(\tilde{\mathbf{A}}_n^{(m)}(u) - \tilde{\mathbf{A}}_n\right)\right]^\intercal \left[\text{vec}\left(\tilde{\mathbf{A}}_n^{(m)}(u) - \tilde{\mathbf{A}}_n\right)\right] du$$

$$= \left(\sum_{t=1}^{n} \left(\mathbf{X}_t^{(m)} \otimes \mathbf{I}_d\right) \hat{\mathbf{u}}_{0t}\right)^\intercal \left(\int_0^1 \left(\left(\boldsymbol{\psi}_{(m)}(u) \otimes \mathbf{I}_K\right) \otimes \mathbf{I}_d\right) \left(\left(\boldsymbol{\psi}_{(m)}^\intercal(u) \otimes \mathbf{I}_K\right) \otimes \mathbf{I}_d\right) du\right)$$

$$\times \left(\sum_{t=1}^{n} \left(\mathbf{X}_t^{(m)} \otimes \mathbf{I}_d\right) \hat{\mathbf{u}}_{0t}\right),$$

where $\hat{\mathbf{u}}_{0t} = \mathbf{y}_t - \hat{\mathbf{A}}_n \mathbf{X}_t$.

Construction of the test statistic

→ In view of the orthogonality of the basis, we can simplify the above expression, leading to the following test statistic

$$\hat{L}_n := \sum_{t=1}^{n}\sum_{s=1}^{n} \hat{\mathbf{u}}_{0t}^{\intercal} \left(\mathbf{X}_t^{(m)} \otimes \mathbf{I}_d\right)^{\intercal} \left(\mathbf{X}_s^{(m)} \otimes \mathbf{I}_d\right) \hat{\mathbf{u}}_{0s}. \tag{15}$$

→ We then have

$$\begin{aligned}\hat{L}_n &= \sum_{t=1}^{n} \hat{\mathbf{u}}_{0t}^{\intercal} \left(\mathbf{X}_t^{(m)} \otimes \mathbf{I}_d\right)^{\intercal} \left(\mathbf{X}_t^{(m)} \otimes \mathbf{I}_d\right) \hat{\mathbf{u}}_{0t} + 2\sum_{t=2}^{n}\sum_{s=1}^{t-1} \hat{\mathbf{u}}_{0t}^{\intercal} \left(\mathbf{X}_t^{(m)} \otimes \mathbf{I}_d\right)^{\intercal} \left(\mathbf{X}_s^{(m)} \otimes \mathbf{I}_d\right) \hat{\mathbf{u}}_{0s} \\ &:= \hat{L}_{an} + \hat{L}_{bn}.\end{aligned}$$

Asymptotic distribution of the test statistic

→ In the case when $\mathbf{X}_t$ is stationary (Gao, Tong, and Wolff, 2002), $\hat{L}_{an}$ determines the asymptotic mean, while standardized version of $\hat{L}_{bn}$ determines the asymptotic distribution.

→ As we show here, when $\mathbf{X}_t$ follows a unit root process, after a suitable standardization, the leading term is $\hat{L}_{an}$, while $\hat{L}_{bn}$ becomes asymptotically negligible.

Asymptotic distribution of the test statistic

Theorem

Suppose that Assumptions 1-4 are satisfied. Under $\mathbb{H}_0$*, we have, as* $n \to \infty$

$$\frac{1}{n^2 m\,tr(\boldsymbol{\Sigma}_{00})}\hat{L}_n \xrightarrow{d} \int_0^1 \left\|\mathbf{B}_r(\boldsymbol{\Omega}_{xx})\right\|^2 dr.$$

→ The critical values are obtained by the Dependent Wild Bootstrap (DWB). ▸ details

Introduction

Model and estimation theory

Testing for structural changes

Monte Carlo study

An empirical application

Conclusion

DGPs

$$\begin{aligned} y_{1t} &= a_{11}(u)\, x_{1,t} + a_{12}(u)\, x_{2,t} + u_{01,t} \\ y_{2t} &= a_{21}(u)\, x_{1,t} + a_{22}(u)\, x_{2,t} + u_{02,t} \end{aligned}$$

where

- $\rightarrow$ $x_{i,t}$: $x_{i,t} = x_{i,t-1} + u_{xi,t}$, $u_{xi,t} = \rho u_{xi,t-1} + \varepsilon_{xi,t}$
- $\rightarrow$ $u_{0i,t}$: $u_{0i,t} = \rho_1 u_{0i,t-1} + \varepsilon_{0i,t}$
- $\rightarrow$ $\boldsymbol{\varepsilon}_t = \left(\varepsilon_{01,t}, \varepsilon_{02,t}, \varepsilon_{x1,t}, \varepsilon_{x2,t}\right)^\top \sim \mathcal{N}\left(\mathbf{0}, (1-\lambda)\mathbf{I} + \lambda\mathbf{J}\right)$
- $\rightarrow$ $a_{i,j}(\tau)$:

$$\mathbf{A}_t := \mathbf{A}(u) = \begin{bmatrix} a_{11}(u) & a_{12}(u) \\ a_{12}(u) & a_{22}(u) \end{bmatrix} = \begin{bmatrix} 1+u & 1+u+u^2 \\ u+u^2 & u \end{bmatrix}$$

Simulation designs

→ Sample size: $n = 100, 200, 400$

→ # of replications: $M = 100000$

→ Performance measure: mean squared error (MSE)

$$MSE\left(\hat{a}_{ij}\right) = \frac{1}{n}\sum_{t=1}^{n}\left(\hat{a}_{ij}\left(\frac{t}{n}\right) - a_{ij}\left(\frac{t}{n}\right)\right)^2$$

for $i, j = 1, 2$ and their standard deviations (SD).

→ m selected by the Generalised Cross Validation (GCV) details

n	MSE $\hat{a}_{11}$	$\hat{a}_{12}$	$\hat{a}_{21}$	$\hat{a}_{22}$
$\lambda = \rho_1 = \rho_2 = 0$				
100	0.0300	0.0277	0.0277	0.0300
200	0.0131	0.0115	0.0115	0.0130
400	0.0059	0.0050	0.0050	0.0059
$\lambda = 0.5, \rho_1 = \rho_2 = 0$				
100	0.0673	0.0648	0.0648	0.0673
200	0.0302	0.0286	0.0286	0.0301
400	0.0131	0.0123	0.0123	0.0131
$\lambda = \rho_1 = \rho_2 = 0.5$				
100	0.2236	0.2207	0.2209	0.2235
200	0.0660	0.0647	0.0648	0.0659
400	0.0175	0.0168	0.0168	0.0175

n	SD $\hat{a}_{11}$	$\hat{a}_{12}$	$\hat{a}_{21}$	$\hat{a}_{22}$
$\lambda = \rho_1 = \rho_2 = 0$				
100	0.0398	0.0338	0.0337	0.0399
200	0.0155	0.0122	0.0123	0.0154
400	0.0064	0.0047	0.0047	0.0064
$\lambda = 0.5, \rho_1 = \rho_2 = 0$				
100	0.0889	0.0839	0.0837	0.0886
200	0.0344	0.0320	0.0320	0.0344
400	0.0131	0.0118	0.0118	0.0131
$\lambda = \rho_1 = \rho_2 = 0.5$				
100	0.2070	0.2038	0.2035	0.2064
200	0.0530	0.0512	0.0512	0.0529
400	0.0139	0.0128	0.0128	0.0139

Empirical size and power of the test

→ To investigate the size of the test (Case 1), we set

$$a_{11}(\tau) = a_{12}(\tau) = a_{21}(\tau) = a_{22}(\tau) = 1.$$

→ To investigate the power of the test, we consider the following two cases:

(A) Structural Break

$$y_{1t} = \begin{cases} 0.8x_{1,t} + 0.8x_{2,t} + u_{01,t} & \text{if } t \leq 0.3T \\ 1x_{1,t} + 1x_{2,t} + u_{01,t} & \text{otherwise} \end{cases}$$

and similarly for y_{2t}.

(B) Smooth Structural Changes

$$y_{1t} = F(u)(1 + 0.5x_{1,t} + 0.5x_{2,t}) + u_{01,t}$$

where $u = t/T$ and $F(u) = 1.5 - 0.8\exp\left(-1.1(u-0.5)^2\right)$ and similarly for y_{2t}.

Empirical size and power of the test

- $\rightarrow$ Sample size: $n = 100, 200, 400$
- $\rightarrow$ # of replications: $M = 100000$
- $\rightarrow$ nominal size: 5%
- $\rightarrow$ $m = \lfloor c \cdot n^{\kappa} \rfloor$, $c = 1$, $\kappa = 0.15, 0.25, 0.3$

Table: Rejection frequencies for the test

	κ			κ		
n	0.15	0.25	0.30	0.15	0.25	0.30
	Case 1					
100	0.073	0.073	0.067			
200	0.055	0.069	0.060			
400	0.050	0.056	0.051			
	Cases A			**Cases B**		
100	0.394	0.314	0.303	0.311	0.392	0.355
200	0.732	0.699	0.707	0.504	0.524	0.582
400	0.917	0.893	0.923	0.798	0.841	0.808

Introduction

Model and estimation theory

Testing for structural changes

Monte Carlo study

An empirical application

Conclusion

Bond risk premia

Cochrane and Piazzesi, 2005:

$$rx_{t+1}^{(n)} = \delta_0^{(n)} + \delta_1^{(n)} y_t^{(1)} + \delta_2^{(n)} f_t^{(2)} + \cdots + \delta_5^{(n)} f_t^{(5)} + u_{t+1}^{(n)}$$

where

→ $n = 2, 3, 4, 5$

→ $y_t^{(n)}$: log-yield of an n-year bond

→ $f_t^{(n)}$: forward rate with maturity n

→ $rx_{t+1}^{(n)}$: excess return of an n-year bond

Bond risk premia

- → When $\delta_1^{(n)} = \cdots = \delta_5^{(n)} = 0$, $rx_{t+1}^{(n)}$ are not predictable and are equal to a constant $\delta_0^{(n)}$.
- → This is consistent with the expectation hypothesis (EH) of the term structure of interest rates.
- → We ask the following two questions:
  - (i) Any evidence of time-varying predictability?
  - (ii) Does modeling time variation improve forecast accuracy?
- → Data: June 1961- December 2024, obtained from Cynthia Wu's website

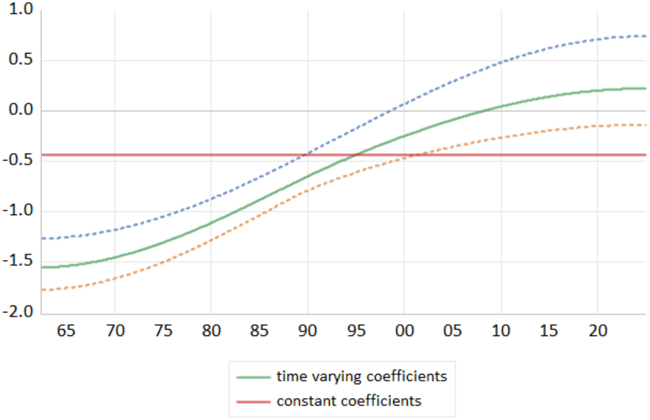

Figure: Plot for $\hat{\delta}_{1,t}^{(2)}$ for forecasting $rx_{t+1}^{(2)}$ using $y_t^{(1)}$, with 95% confidence intervals; m selected by GCV.

▸ construction of CI

Out-of-sample(OOS) forecasting performance

→ Evaluation period: November 2020 - December 2024

→ benchmark model: EH

→ Performance measure: OOS R^2(Campbell and Thompson, 2008):

$$R^2_{oos} = 1 - \frac{\sum_{t=1}^{R} e^{(n)2}_{t,TV}}{\sum_{t=1}^{R} e^{(n)2}_{t,EH}},$$

where

- $e_{t,TV}(n)$: forecast error from the model with time-varying coefficients
- $e_{t,EH}(n)$: forecast error from the EH benchmark model

Out-of-sample(OOS) forecasting performance

	$n = 2$	$n = 3$	$n = 4$	$n = 5$
R^2_{oos}	0.192	0.168	0.159	0.160
DM-test	-2.142	-2.506	-2.776	-2.894
p-value	0.032	0.012	0.006	0.004

$\rightarrow$ DM-test: Diebold and Mariano, 1995 test for equal forecast accuracy

Conclusion

- → We propose a series estimation method for cointegrated system with time varying coefficients and establish its asymptotic properties.
- → We develop a test for structural changes in cointegrated system, which does not require prior information about the alternative.
- → Monte Carlo simulations show that both the estimator and test have satisfactory finite sample performance.
- → An empirical application on bond risk premia further demonstrate the usefulness of the method.

Appendix Slides

More on $\mathbf{u}_t$

→ Under Assumption1, $\mathbf{u}_t$ has covariance matrix $\boldsymbol{\Sigma} = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_j \boldsymbol{\Sigma}_\varepsilon \boldsymbol{\Phi}_j'$.

→ By functional central limit theory (FCLT) for linear process (cf. Phillips and Solo, 1992), we have

$$n^{-1/2} \sum_{t=1}^{n} u_t \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega}\right), \tag{16}$$

with covariance matrix $\boldsymbol{\Omega} = \boldsymbol{\Phi}(1)\boldsymbol{\Sigma}_\varepsilon \boldsymbol{\Phi}(1)' > 0$ and $\boldsymbol{\Phi}(1) = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_j$.

→ Partition:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{0x} \\ \boldsymbol{\Sigma}_{x0} & \boldsymbol{\Sigma}_{xx} \end{bmatrix}, \quad \boldsymbol{\Phi}(1) = \begin{bmatrix} \boldsymbol{\Phi}_0(1) \\ \boldsymbol{\Phi}_x(1) \end{bmatrix}, \quad \boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{00} & \boldsymbol{\Omega}_{0x} \\ \boldsymbol{\Omega}_{x0} & \boldsymbol{\Omega}_{xx} \end{bmatrix}, \tag{17}$$

→ One-sided long-run covariance matrix

$$\boldsymbol{\Lambda} := \begin{bmatrix} \boldsymbol{\Lambda}_{00} & \boldsymbol{\Lambda}_{0x} \\ \boldsymbol{\Lambda}_{x0} & \boldsymbol{\Lambda}_{xx} \end{bmatrix} = \sum_{j=1}^{\infty} \mathbb{E}\left(\mathbf{u}_t \mathbf{u}'_{t-j}\right).$$

Selection of m

→ To select m, we use the Generalised Cross Validation (GCV)

$$\hat{m}_{GCV} = \arg\min_{2 \leq m \leq M} \left(1 - \frac{mK}{n}\right)^{-2} \frac{1}{nd} \sum_{t=1}^{n} \sum_{j=1}^{d} \left(\hat{u}_{tj}^{(m)}\right)^2$$

where $\hat{u}_{tj}^{(m)}$ is the j-element of $\hat{\mathbf{u}}_{0t}^{(m)} = \mathbf{y}_t - \hat{\mathbf{B}}^{(m)} \mathbf{X}_t^{(m)}$.

→ In the presence of serial correlation in the residuals, Shao, 2010 suggest the Dependent Wild Bootstrap (DWB) .

→ Let ζ_t denote the dependent weights (multipliers) ζ_t that mimic serial dependence, and it satifies $E\left(\zeta_t\right) = 0$ and $E\left(\zeta_t^2\right) = 1$.

→ Define $n \times n$ covariance matrix

$$\sum_{ij} = k\left(\frac{|i-j|}{l_n}\right), \quad 1 \leq i, j \leq n$$

where $k\left(.\right)$ is a kernel function and l_n is the bandwidth satisfying $l_n \to \infty$ with $l_n / n \to 0$.

→ We draw $\zeta = \left(\zeta_1, ..., \zeta_n\right)^{\mathrm{T}} \sim N\left(0, \Sigma\right)$, use the Quadratic Spectral (QS) kernel and $l_n = 1.3221 \times T^{1/5}$ (Andrews, 1991).

→ Step 1: Generate

$$\mathbf{u}_t^{(0)*} = \zeta_t \mathbf{u}_t^{(0)}$$

for $t = 1, 2, ..., n$ and

$$\mathbf{u}_{xt}^* = \zeta_t \mathbf{u}_{xt}$$

where $\mathbf{u}_{xt} = \mathbf{X}_t - \mathbf{X}_{t-1}$ for $t = 2, 3, ..., n$.

→ Step 2: Rebuild the bootstrap regressor as

$$\begin{aligned} \mathbf{X}_t^* &= \mathbf{X}_{t-1}^* + \mathbf{u}_{xt}^* \text{ for } t = 2, 3, ..., n \\ \mathbf{X}_1^* &= \mathbf{X}_1 \end{aligned}$$

and rebuild the bootstrap regressand as

$$\mathbf{y}_t^* = \hat{\mathbf{A}}_n \mathbf{X}_t^* + \hat{\mathbf{u}}_{0t}^*$$

→ Step 3: Compute $\hat{L}^{(m)*}$ by replacing $\hat{\mathbf{u}}_t$ with $\hat{\mathbf{u}}_t^*$ and $\mathbf{X}_t^{(m)}$ with $\mathbf{X}_t^{(m)*}$ in $\hat{L}^{(m)}$.

→ Step 4: Repeat Steps 1-3 for B bootstrap replications to generate null distribution $\hat{L}^{(m)*(b)}$ for $b = 1, \cdots, B$. Reject H$_0$ if $\hat{L}^{(m)} > c_\alpha$ where c_α is the $(1-\alpha)$-quantile of bootstrap null distribution.

→ Step 1: Generate

$$\hat{\mathbf{u}}_{0t}^{(m)*} = \zeta_t \hat{\mathbf{u}}_{0t}^{(m)}$$

for $t = 1, 2, ..., n$ and

$$\mathbf{u}_{xt}^{*} = \zeta_t \mathbf{u}_{xt}$$

where $\mathbf{u}_{xt} = \mathbf{X}_t - \mathbf{X}_{t-1}$ for $t = 2, 3, ..., n$.

→ Step 2: Rebuild the bootstrap regressor as

$$\begin{aligned} \mathbf{X}_t^* &= \mathbf{X}_{t-1}^* + \mathbf{u}_{xt}^* \text{ for } t = 2, 3, ..., n \\ \mathbf{X}_1^* &= \mathbf{X}_1 \end{aligned}$$

and rebuild the bootstrap regressand as

$$\mathbf{y}_t^* = \hat{\mathbf{A}}_n^{(m)}(\tau_t)\, \mathbf{X}_t^* + \hat{\mathbf{u}}_{0t}^{(m)*}$$

Appendix Slides

Bootstrap Confidence Interval for $\hat{A}_n(\tau_t)$.

→ Step 3: Re-estimate $\hat{\mathbf{A}}_n^{(m)}(\tau_t)^*$ by replacing $\mathbf{y}_t$ with $\mathbf{y}_t^*$ and $\mathbf{X}_t$ with $\mathbf{X}_t^*$ in $\hat{\mathbf{A}}_n^{(m)}(\tau_t)$.

→ Step 4: Repeat Step 1 to Step 3 for $b = 1, ..., B$ to obtain $\left\{\hat{\mathbf{A}}_n^{(m)}(\tau_t)^{*(b)}\right\}_{b=1}^{B}$. A $100(1-\alpha)\%$ confidence interval is the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of $\left\{\hat{\mathbf{A}}_n^{(m)}(\tau_t)^{*(b)}\right\}$; i.e.

$$\left[\hat{\mathbf{A}}_n^{(m)}(\tau_t)^*_{(\alpha/2)} \ , \ \hat{\mathbf{A}}_n^{(m)}(\tau_t)^*_{(1-\alpha/2)}\right]$$

References I

Andrews, Donald WK (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation". *Econometrica: Journal of the Econometric Society*, pp. 817–858.

Bierens, Herman J (1997). "Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the US price level and interest rate". *Journal of Econometrics* 81.1, pp. 29–64.

Bierens, Herman J and Luis F Martins (2010). "Time-varying cointegration". *Econometric Theory* 26.5, pp. 1453–1490.

Cai, Zongwu, Yunfei Wang, and Yonggang Wang (2015). "Testing instability in a predictive regression model with nonstationary regressors". *Econometric Theory* 31.5, pp. 953–980.

Campbell, John Y and Samuel B Thompson (2008). "Predicting excess stock returns out of sample: Can anything beat the historical average?" *The Review of Financial Studies* 21.4, pp. 1509–1531.

Cochrane, John H and Monika Piazzesi (2005). "Bond risk premia". *American economic review* 95.1, pp. 138–160.

References II

Diebold, Francis X and Robert S Mariano (1995). "Comparing predictive accuracy". *Journal of Business & economic statistics* 13.3.

Gao, Jiti, Howell Tong, and Rodney Wolff (2002). "Model specification tests in nonparametric stochastic regression models". *Journal of Multivariate Analysis* 83.2, pp. 324–359.

Li, Degui, Peter CB Phillips, and Jiti Gao (2020). "Kernel-based inference in time-varying coefficient cointegrating regression". *Journal of Econometrics* 215.2, pp. 607–632.

Park, Joon Y and Sang B Hahn (1999). "Cointegrating regressions with time varying coefficients". *Econometric Theory* 15.5, pp. 664–703.

Phillips, Peter CB (1991). "Optimal inference in cointegrated systems". *Econometrica: Journal of the Econometric Society*, pp. 283–306.

Phillips, Peter CB and Bruce E Hansen (1990). "Statistical inference in instrumental variables regression with I (1) processes". *The review of economic studies* 57.1, pp. 99–125.

References III

Phillips, Peter CB, Degui Li, and Jiti Gao (2017). “Estimating smooth structural change in cointegration models”. *Journal of Econometrics* 196.1, pp. 180–195.

Phillips, Peter CB and Victor Solo (1992). “Asymptotics for linear processes”. *The Annals of Statistics*, pp. 971–1001.

Shao, Xiaofeng (2010). “The dependent wild bootstrap”. *Journal of the American Statistical Association* 105.489, pp. 218–235.

Stock, James H (1987). “Asymptotic properties of least squares estimators of cointegrating vectors”. *Econometrica: Journal of the Econometric Society*, pp. 1035–1056.