Low-Frequency Robust Cointegration Testing

Ulrich K. Müller   Mark M. Watson
Princeton University

May 21, 2008

Motivation

- Central idea of cointegration: There exists a linear combination that reduces the persistence of time series.

- Inference about cointegrating vector?

- Standard approach of Engle and Granger (1987), Johansen (1988), Phillips and Hansen (1990), Stock and Watson (1993), etc.:
  Common stochastic trend is I(1), error correction term is I(0).

- Two sources of fragility of standard approach

  1. Implied reduction of persistence from I(1) to I(0) is implausible

  2. Assumes exact I(1) properties of stochastic trend

The I(0) / I(1) Persistence Dichotomy

- Up to a scaling factor, the asymptotic properties of an I(0) process and i.i.d. data are identical, in the sense that both satisfy a Functional Central Limit Theorem. In the same sense, an I(1) process is just like a random walk.

- The associated extreme reduction of persistence is implausible (or generates poor approximations) for macroeconomic applications.

- Our solution: Adopt the low-frequency transformation approach of Müller and Watson (2007).

Low Frequency Transformations

- Let $\Psi_j(s) = \sqrt{2}\cos(j\pi s)$. For a scalar sequence $\{a_t\}_{t=1}^T$, define

$$A_{Tj} = \int_0^1 \Psi_j(s) a_{\lfloor sT \rfloor + 1} ds \simeq T^{-1} \sum_{t=1}^{T} \Psi_j(t/T) a_t$$

  and $A_T = (A_{T1}, \cdots, A_{Tq})'$.

- Claim: The q numbers A_T summarize variability of $\{a_t\}_{t=1}^T$ for frequencies lower than $q\pi/T$.

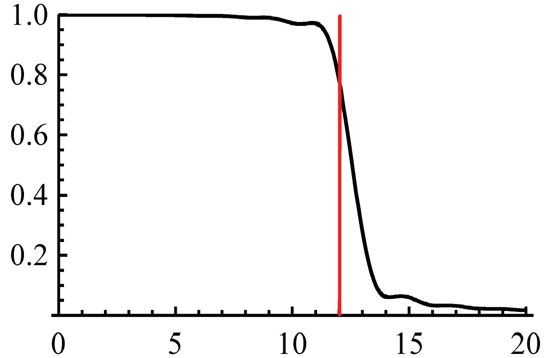
- Consider R^2 from regression of generic periodic series $a_t = \sin(\pi r t/T + \phi)$ on $\Psi_j(t/T)$, $j = 1, \cdots, q$. For T large, this R^2 is well approximated by R^2 of continuous time regression of

$$\sin(\pi r s + \phi), \qquad s \in [0, 1]$$

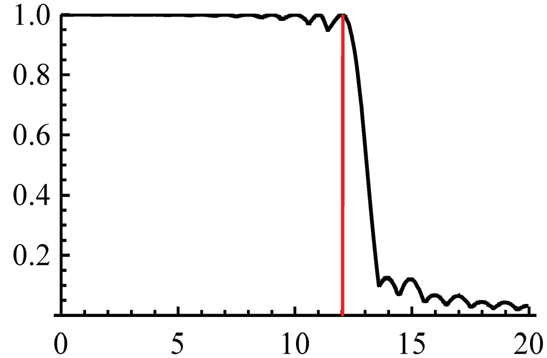
  on $\Psi(s) = (\Psi_1(s), \cdots, \Psi_q(s))'$.

R^2 as Function of r for $q = 12$

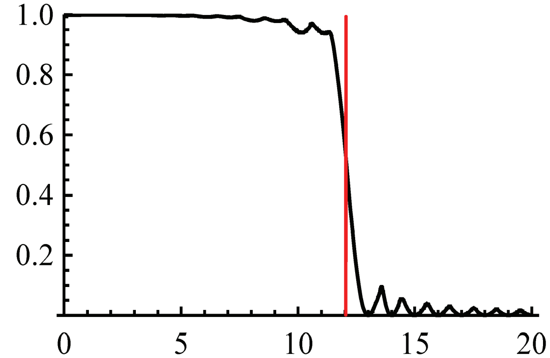
average over ϕ



maximum over ϕ



minimum over ϕ



- Ideally, $R^2 = 1$ for $r \leq 12$ and $R^2 = 0$ for $r > 12$ for all ϕ.

Low-Frequency Transformations

- For data spanning 50 years, below business cycle frequency information (=periods greater than 8 years) is summarized by $q = 12$ weighted averages of the original data.

- Base inference on cointegrating vector on these q weighted averages only, under asymptotics with q fixed as $T \to \infty$. Idea: Make asymptotic approximation relevant for samples where low-frequency information is scarce.

- Take standard I(0) asymptotics seriously only for these $q = 12$ weighted averages

  $\Rightarrow$ I(0) error correction term is assumed to behave like i.i.d. data *only with respect to below-business-cycle frequency variability.*

  $\Rightarrow$ same with model for stochastic trend

Relationship to Standard I(0) Asymptotics

- Under usual asymptotics, the I(0) Error Correction Term $\{z_t\}$ satisfies

$$T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} z_t \Rightarrow \sigma W_z(\cdot) \tag{1}$$

- By the Continuous Mapping Theorem, from (1), for any $n \in \mathbb{N}$

$$\left\{ \int_0^1 \Psi_j(s) z_{[sT]+1} ds \right\}_{j=1}^n \Rightarrow \left\{ \sigma \int_0^1 \Psi_j(s) dW_z(s) \right\}_{j=1}^n \sim \{Z_j\}_{j=1}^n$$

where $Z_j \sim$i.i.d.$\mathcal{N}(0, \sigma^2)$, since with $\Psi_j(s) = \sqrt{2}\cos(j\pi s)$, $\int_0^1 \Psi_i(s)\Psi_j(s) ds = \mathbf{1}[i = j]$.

- Usual asymptotics imply that the error correction term has the same properties as i.i.d. data for all frequencies $|\omega| \leq n\pi/T$, with n arbitrary. Our approach: take this approximation seriously *for below business cycle frequencies only*, by choosing the appropriate $n = q$.

Nature of Stochastic Trend

- As demonstrated by Elliott (1998), standard cointegration inference potentially highly misleading when common stochastic trend is local-to-unity, rather than exactly I(1).

- Valid inference for local-to-unity stochastic trends complicated, because local-to-unity nuisance parameter cannot be estimated consistently: Cavanagh, Elliott and Stock (1995), Campbell and Yogo (2006), Stock and Watson (1996), Jansson and Moreira (2006).

- Still assumes one fairly specific one-parameter model for stochastic trend (and so does Fractional Cointegration).

- Wright (2000): Conduct inference on cointegrating vector by testing whether hypothesized error correction term is I(0). No dependence on precise nature of stochastic trend, but efficiency?

This Paper

- Inference about cointegrating space using low-frequency transformations.

- Stochastic trend specification in terms of flexible Gaussian limiting process $\Rightarrow$ numerous nuisance parameters that cannot be consistently estimated.

- Main challenge: Efficient test in the presence of nuisance parameter under the null hypothesis?

  - Derive upper bounds on power for all tests that control size (cf. Andrews, Moreira and Stock 2007).

  - Low-dimensional nuisance parameter: Generic method to numerically approximate least upper bound and almost efficient test.

  - High-dimensional nuisance parameter: Determine low upper bound and compare to low-frequency version of Wright's (2000) test.

Main Finding

- Focus on tests that maximize power against classical I(1) alternative, but that are restricted to control size for flexible trend specification ("Robust Cointegration Testing").

- For one cointegrating vector, low-frequency version of Wright's (2000) test almost achieves the power bound.

  $\Rightarrow$ In absence of a priori knowledge about nature of persistence, this simple test thus yields robust and almost efficient inference about value of cointegrating vector.

  $\Rightarrow$ Confidence sets can be obtained by inverting the test.

Plan of Talk

1. Introduction

2. Time Series Model and Invariance

3. Low-Frequency Transformation

4. Tests based on Putative Error Correction Term

5. Power bounds on Tests of all Data

6. Results

7. Generalization to Multivariate Systems

8. Conclusion

Canonical Bivariate Cointegrated System

- Canonical representation of bivariate observations (y_t, x_t) in terms of the error correction term z_t and stochastic trend v_t

$$\begin{aligned} y_t &= \Gamma_{yz} z_t + \Gamma_{yv} v_t \\ x_t &= \Gamma_{xz} z_t + \Gamma_{xv} v_t \end{aligned}$$

- Test of null hypothesis whether y_t is the error correction term, i.e.

$$H_0 : \Gamma_{yv} = 0$$

- Error correction term assumed to satisfy

$$T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} z_t \Rightarrow W_z(s)$$

for standard Wiener process W_z.

Flexible Stochastic Trend

- Stochastic trend is assumed to satisfy

$$T^{-1/2} v_{\lfloor sT \rfloor} \Rightarrow \int_{-\infty}^{s} g(s,t) dW_v(t)$$

and R is correlation between standard Wiener processes W_v and W_z.

- We focus on 4 cases:

1. Standard I(1) case: $g(s,t) = \mathbf{1}[t \geq 0]$
2. Local-to-unity case: $g(s,t) = \mathbf{1}[t \geq 0] e^{-c(s-t)}$, $c \geq 0$
3. General stationary case: $g(s,t) = g^S(s-t)$
4. Unrestricted case

Invariance

- Recall

$$
\begin{aligned}
y_t &= \Gamma_{yz} z_t + \Gamma_{yv} v_t \\
x_t &= \Gamma_{xz} z_t + \Gamma_{xv} v_t
\end{aligned}
$$

- Impose invariance to transformations

$$\{y_t, x_t\}_{t=1}^{T} \to \{A_{yy} y_t, A_{xx} x_t + A_{xy} y_t\}_{t=1}^{T}$$

for $A_{yy}, A_{xx} \neq 0$.

- No loss in generality in setting $\Gamma_{yz} = \Gamma_{xv} = 1$ and $\Gamma_{xz} = 0$, so that

$$
\begin{aligned}
y_t &= z_t + \Gamma_{yv} v_t \\
x_t &= v_t
\end{aligned}
$$

Testing Problem and Local Alternatives

- Local alternatives of the form $\Gamma_{yv} = T^{-1}B$ for B fixed, so that

$$T^{-1/2}\sum_{t=1}^{\lfloor sT\rfloor} y_t \Rightarrow W_z(s) + B\int_0^s\int_{-\infty}^u g(u,t)dW_v(t)du$$

- Null hypothesis

$$H_0 : B = 0, \quad g(s,t) \in \mathcal{G}_0$$

- Alternative hypothesis

$$H_a : B = B_1, \quad g(s,t) = \mathbf{1}[t \geq 0]$$

$\Rightarrow$ focus on power against traditional I(1) alternative

Low Frequency Transformations I

- Recall that the low frequency transformation of $\{a_t\}_{t=1}^{T}$ are the q numbers

$$A_T = \int_0^1 \Psi(s) a_{\lfloor sT \rfloor + 1} ds$$

where $\Psi(s) = (\sqrt{2}\cos(\pi s), \sqrt{2}\cos(2\pi s), \cdots, \sqrt{2}\cos(q\pi s))'$.

- By assumed limiting behavior of z_t and v_t,

$$\begin{bmatrix} T^{1/2} Z_T \\ T^{-1/2} V_T \end{bmatrix} \Rightarrow \begin{bmatrix} Z \\ V \end{bmatrix} \sim \mathcal{N}(0, \Sigma_{(Z,V)}) \text{ and } \Sigma_{(Z,V)} = \begin{bmatrix} I_q & \Sigma_{ZV} \\ \Sigma_{VZ} & \Sigma_{VV} \end{bmatrix}$$

since $Z = \int_0^1 \Psi(s) dW_z(s)$, so that $E[ZZ'] = \int_0^1 \Psi(s)\Psi(s)' ds = I_q$.

The elements of Σ_{ZV} and Σ_{VV} can be explicitly computed, but depend on a parameter θ, describing the stochastic trend $g(s,t)$ and R, the correlation between W_z and W_v.

Low Frequency Transformations II

- Under local alternatives with $\Gamma_{yv} = T^{-1}B$, $Y_T = Z_T + BT^{-1}V_T$

$$\begin{bmatrix} T^{1/2}Y_T \\ T^{-1/2}X_T \end{bmatrix} \Rightarrow \begin{bmatrix} Y \\ X \end{bmatrix} = \begin{bmatrix} Z + BV \\ V \end{bmatrix} \sim \mathcal{N}\left(0, \Sigma_{(Y,X)}\right) \qquad (2)$$

where

$$\Sigma_{(Y,X)} = \begin{bmatrix} I_q & BI_q \\ 0 & I_q \end{bmatrix} \begin{bmatrix} I_q & \Sigma_{ZV} \\ \Sigma_{VZ} & \Sigma_{VV} \end{bmatrix} \begin{bmatrix} I_q & BI_q \\ 0 & I_q \end{bmatrix}'$$

- Consider tests of $H_0 : B = 0$ based on data $\{y_t, x_t\}_{t=1}^T$ that control asymptotic size whenever the weak convergence (2) holds with $B = 0$.

  Müller (2007): Test with highest asymptotic power is simply best test in limiting problem with (Y, X) observed, evaluated at sample analogues (Y_T, X_T).

Tests based on Y Only

- Problem is to test $H_0 : B = 0$ in
$$\begin{bmatrix} Y \\ X \end{bmatrix} = \begin{bmatrix} Z + BV \\ V \end{bmatrix}$$
and distribution of V depends on nuisance parameters.

- In analogy to Wright (2000), denote by JW the (scale invariant) point optimal test that ignores X, i.e. the best test of
$$H_0 : Y = Z \quad \text{against} \quad H_1 : Y = Z + b_1 V_1$$
with $b_1 = 10$, where V_1 is the limit of the low-frequency transformation of a classical I(1) stochastic trend.

- The test is similar, since it does not depend on X, so that its distribution under the null does not depend on any nuisance parameters. But it is *ad hoc*, since it ignores the potentially valuable information contained in X.

Power Bounds for Tests Using (Y, X)

Four steps:

1. Density of a Maximal Invariant

2. Parameterization of $\boldsymbol{\Sigma}_{(Y,X)}$

3. General Result on Power Bounds

4. Implementation

Density of a Maximal Invariant

- Recall that

$$\begin{bmatrix} Y \\ X \end{bmatrix} = \begin{bmatrix} Z + BV \\ V \end{bmatrix} \sim \mathcal{N}\left(0, \Sigma_{(Y,X)}\right)$$

and impose invariance to transformations

$$\begin{bmatrix} Y \\ X \end{bmatrix} \rightarrow \begin{bmatrix} A_{yy}Y \\ A_{xx}X + A_{xy}Y \end{bmatrix}$$

- All invariant tests can be written as functions of a maximal invariant.

- **Lemma**: The density of a maximal invariant $Q = h(Y, X)$ is given by a known function f_Q, which depends on $\Sigma_{(Y,X)}$.

Parameterization of $\Sigma_{(Y,X)}$ under Alternative

- Recall

$$\begin{bmatrix} Y \\ X \end{bmatrix} = \begin{bmatrix} Z + BV \\ V \end{bmatrix} \sim \mathcal{N}\left(0, \Sigma_{(Y,X)}\right)$$

where

$$\Sigma_{(Y,X)} = \begin{bmatrix} I_q & BI_q \\ 0 & I_q \end{bmatrix} \begin{bmatrix} I_q & \Sigma_{ZV} \\ \Sigma_{VZ} & \Sigma_{VV} \end{bmatrix} \begin{bmatrix} I_q & BI_q \\ 0 & I_q \end{bmatrix}'$$

- Focus on alternatives with traditional I(1) stochastic trend

$\Rightarrow$ determines Σ_{VV} and Σ_{ZV} up to $R = E[W_z(1)W_v(1)]$

- Trace out power envelope as a function of $B = B_1$ and $R = R_1$

Parameterization of $\Sigma_{(Y,X)}$ under Null

- Under null hypothesis:

$$\Sigma_{(Y,X)} = \Sigma_{(Z,V)} = \begin{bmatrix} I_q & \Sigma_{ZV} \\ \Sigma_{VZ} & \Sigma_{VV} \end{bmatrix}$$

$\Rightarrow$ denote by $\theta \in \Theta$ the nuisance parameter that describes $\Sigma_{(Z,V)}$ for different stochastic trend models

- Unrestricted model:

Lemma: With $g(s,t)$ unrestricted, there are no constraints (other than positive definiteness of $\Sigma_{(Z,V)}$) on Σ_{VZ} and Σ_{VV}.

$\Rightarrow \theta$ is of dimension $q^2 + q(q+1)/2$

Parameterization of $\Sigma_{(Y,X)}$ under Null, ctd.

- Stationary model where $g(s,t) = g^S(s-t)$:

  convenient step function parameterization with 40 steps

- Local to unity model:

  θ consists of mean reversion parameter c, and R

- I(1) model:

  $\theta = R$

General Result about Power Bounds

- Face a hypothesis testing problem of the form

$$H_0 \quad : \quad \text{the density of } U \text{ is } f_\theta(u),\ \theta \in \Theta$$
$$H_1 \quad : \quad \text{the density of } U \text{ is } h$$

  Construction of an efficient test φ^*?

- **Lemma:** Let φ be any size α test of H_0 against H_1. For any probability distribution Λ, let φ_Λ be the Neyman-Pearson level α test of

$$H_\Lambda : \text{the density of } U \text{ is } \int f_\theta(u) d\Lambda(\theta)$$

  against H_1. Then φ_Λ is at least as powerful as φ.

- Proof: Since φ is of size α under H_0, it is also a valid level α test of H_Λ against H_1. But by assumption, φ_Λ is the best level α test in this problem, so its power is at least as high.

Two Uses for Upper Bounds on Power

1. Use numerical methods to estimate φ^*. The power bound tells us that a candidate $\hat{\varphi}^*$ is close enough to being efficient for practical purposes.

   - Paper described an algorithm that numerically determines $\hat{\varphi}^*$ with power within 2.5% of the power bound. Idea is to exploit numerical advantages of distributions for Λ with mass at N points $\{\theta_1, \cdots, \theta_N\}$.

   - Requires verification that $\hat{\varphi}^*$ controls size, which is feasible only when Θ is low dimensional.

     $\Rightarrow$ can only be implemented for I(1) and local-to-unity model

2. Compare power bound to power of an *ad hoc* test that is known to control size under H_0. If the power of the *ad hoc* is close to the bound, then it is close to optimal.

   $\Rightarrow$ Application to JW test. Still requires low bound.

Implementation for High Dimensional Parameters

- Let Σ_1 be the covariance matrix of (Y, X) under the alternative model with $B = B_1$ and $R = R_1$. Denote by $\Sigma_0(\theta, \gamma)$ the covariance matrix of $(\gamma_1 Y, (\gamma_2 X + \gamma_3 Y))$ under the null model, where $\gamma = (\gamma_1, \gamma_2, \gamma_3)$.

- We consider Λ that put all mass on the point θ^*. Since discriminating the null and alternative amounts to discriminating between two zero mean normals with covariance matrices $\Sigma_0(\theta, \gamma)$ and Σ_1, it makes sense to choose θ and γ so that these distributions are close. We thus use the numerical minimizer $\theta = \theta^*$ of the Kullback-Leibler divergence between the distributions $N(0, \Sigma_1)$ and $N(0, \Sigma_0(\theta, \gamma))$ over (θ, γ).

- Resulting power bound entirely driven by restriction of size control for $\theta = \theta^*$, size of Θ irrelevant.

  $\Rightarrow$ model with $\Sigma_0(\theta^*, \gamma)$ is not *empirically* unreasonable—after all, hard to discriminate from model with Σ_1

Power Bounds of 5% Level Tests for $q = 12$

Because of invariance, power only depends on $|R|$, $|B|$, and **sign**$(B \cdot R)$

$\lvert R\rvert$	$B \cdot R$	I(1)	LTU	ST	UNR	JW
			$\lvert B\rvert = 7$			
0	0	*0.50*	*0.50*	0.41	0.36	0.36
0.5	$+$	*0.65*	*0.58*	0.40	0.36	0.36
0.5	$-$	*0.65*	*0.66*	0.54	0.36	0.36
0.9	$+$	*0.94*	*0.65*	0.44	0.36	0.36
0.9	$-$	*0.94*	*0.93*	0.89	0.36	0.36
			$\lvert B\rvert = 14$			
0	0	*0.81*	*0.81*	0.69	0.64	0.63
0.5	$+$	*0.90*	*0.78*	0.67	0.65	0.63
0.5	$-$	*0.90*	*0.88*	0.81	0.64	0.63
0.9	$+$	*1.00*	*0.82*	0.72	0.66	0.63
0.9	$-$	*1.00*	*1.00*	0.99	0.65	0.63

Italic entries denote direct approximations to least upper bounds (are by construction within 2.5% of least upper bound)

Generalization to Multivariate Systems

- Model now

$$
\begin{aligned}
y_t &= \Gamma_{yz} z_t + \Gamma_{yv} v_t \\
x_t &= \Gamma_{xz} z_t + \Gamma_{xv} v_t
\end{aligned}
$$

where $r \times 1$ vector y_t are putative error correction terms, and $k \times 1$ vector x_t contain stochastic trends. $H_0 : \Gamma_{yv} = 0$.

- Invariance

$$
(y_t, x_t) \rightarrow (A_{yy} y_t, A_{xx} x_t + A_{xy} y_t)
$$

where A_{yy} and A_{xx} are nonsingular, so that without loss of generality

$$
\begin{aligned}
y_t &= z_t + \Gamma_{yv} v_t \\
x_t &= v_t.
\end{aligned}
$$

Multivariate ECM and Stochastic Trend

- Error Correction Term:

$$T^{-1/2}\sum_{t=1}^{\lfloor sT\rfloor} z_t \Rightarrow S_z W(s) = W_z(s), \quad \text{where } S_z S_z' = I_r$$

and W is a $(r+k)$ vector standard Wiener process

- Stochastic Trend

$$T^{-1/2}v_{\lfloor sT\rfloor} \Rightarrow \int_{-\infty}^{s} H(s,t)dW(t)$$

where $H : [0,1]^2 \mapsto \mathbb{R}^{k\times(r+k)}$

Restrictions on $H(s,t)$

- $H(s,t) = G(s,t)S_v$ with $S_v S_v' = I_k$, so that

$$T^{-1/2} v_{\lfloor sT \rfloor} \Rightarrow \int_{-\infty}^{s} G(s,t) dW_v(t)$$

  and $R = S_z S_v'$ is correlation between W_z and W_v

- $G(s,t) = \text{diag}(g_1(s,t), \cdots, g_k(s,t))$: k independent common trends

- $g_i(s,t) = g_i^S(s-t)$, $i = 1, \cdots, k$: k independent, asymptotically stationary common trends

- $G(s,t) = \mathbf{1}[t > 0]e^{C(s-t)}$: Local-to-Unity model with $r \times r$ mean reverting matrix C

- $G(s,t) = \mathbf{1}[t > 0]$: I(1) model

Differences to Bivariate Model

- Densities of maximal invariants become messy, no closed form solution in general

  $\Rightarrow$ but: Nice closed form for point-optimal Y–only test against I(1) alternatives as long as $r \leq k$ (low-frequency version of multivariate JW test). Happens to be uniformly most powerful over some values of B and R.

- Power Bounds for $r = 1$ and $k = 2$ also hold for $r = 1$ and $k > 2$

  $\Rightarrow$ Under alternative with I(1) stochastic trends, by invariance, can assume that last $k-2$ stochastic trends are independent of first 2 stochastic trends, and putative error correction term

  $\Rightarrow$ In all null stochastic trend models, can choose identical behavior of last $k - 2$ stochastic trends, so that it becomes optimal to ignore them.

Power Bounds of 5% Level Tests for $r = 1$, $k \geq 2$

Power depends only on $||B||, ||R||$ and $\omega = \text{tr}(B'R)/||B|| \cdot ||R||$

$\|\|R\|\|$	I(1)	LTU	ST	DIAG	UNR/G-model
	$\|\|B\|\| = 7$, Power of JW test: 0.36				
0	0.42	0.42	0.39	0.36	0.36
0.5	0.54 0.47 0.54	0.54 0.48 0.51	0.47 0.42 0.38	0.36 0.37 0.36	0.36 0.36 0.36
0.9	0.92 0.86 0.92	0.92 0.69 0.64	0.82 0.59 0.43	0.36 0.43 0.36	0.36 0.36 0.36
	$\|\|B\|\| = 14$, Power of JW test: 0.62				
0	0.71	0.71	0.67	0.64	0.64
0.5	0.82 0.77 0.83	0.83 0.77 0.76	0.76 0.71 0.66	0.65 0.64 0.64	0.65 0.64 0.64
0.9	0.99 0.98 0.99	1.00 0.85 0.85	0.97 0.84 0.71	0.66 0.74 0.66	0.66 0.65 0.66

Side-by-side entries correspond to $\omega = \{-1, 0, 1\}$

Same calculations for $r = 2$ and $k = 1$ reveals differences up to 14 percentage points between best Y-only test and bound on tests in unrestricted model

Conclusions

- Study of efficient inference on cointegrating vector in cointegrated system. Focus on low-frequency variability.

- Restrict attention to tests that control size for a wide range of stochastic trend specifications.

- Low-frequency version of Wright's (2000) test essentially efficient for $r = 1$ cointegrating vectors.

- Numerical method to identify approximately asymptotic (weighted average) power maximizing test in tightly parametrized stochastic trend model.

  $\Rightarrow$ Method is generic and can be applied to other nonstandard testing problems with nuisance parameter under the null hypothesis