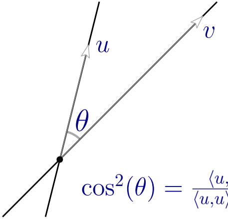
Cointegration, S&P, and Random Matrices

Vadim Gorin
(joint work with Anna Bykhovskaya)

Anna is looking for a job this year.   Like the talk? $\Rightarrow$ Hire her!

Warm up I: angles and correlations

Geometry: A unique invariant (under orthogonal transformations) of a pair of lines through the origin is the **angle.**



$$\cos^2(\theta) = \frac{\langle u,v\rangle^2}{\langle u,u\rangle\langle v,v\rangle}$$

Probability: Dependence of (mean 0) random variables ξ and η is measured by the (squared) **correlation coefficient**.

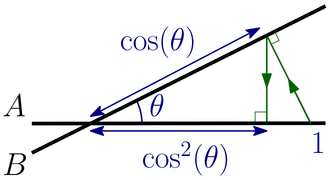
$$\frac{(\mathbb{E}\xi\eta)^2}{\mathbb{E}\xi^2\mathbb{E}\eta^2}$$

Statistics: Dependence of (mean 0) data sequences is measured by the (squared) **sample correlation coefficient**

$$\frac{\left(\sum_{t=1}^{T}\xi_t\eta_t\right)^2}{\sum_{t=1}^{T}\xi_t^2\sum_{t=1}^{T}\eta_t^2}$$

Warm up II: canonical angles and canonical correlations

Geometry: Invariants of a pair of linear subspaces A and B are **canonical angles:** their squared cosines are eigenvalues of $P_A P_B$ — product of orthogonal projections on spaces A and B.



The squared cosines have **variational meaning**. The largest one is

$$\max_{u \in A,\, v \in B} \frac{\langle u, v \rangle^2}{\langle u, u \rangle \langle v, v \rangle}$$

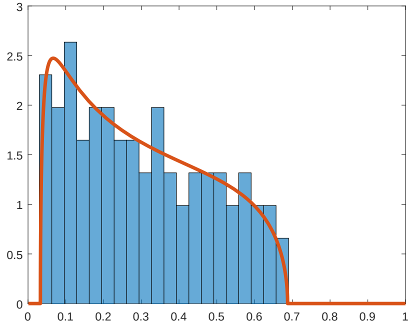
Statistics: Squared **sample canonical correlations** of $N \times T$ datasets A and B: eigenvalues of $P_A P_B$ — product of projectors on (rows of A) and (rows of B) in T–dimensional space.

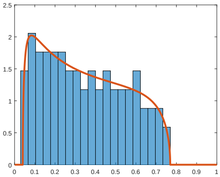
What famous laws are fitted by orange curves?

Histogram:
$X_t = \ln(\text{S\&P 100 stocks})$

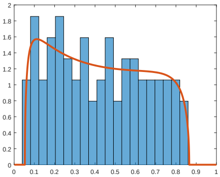
Weekly data 2010-2020.

Squared sample canonical correlations of X_{t-1} and $\Delta X_t := X_t - X_{t-1}$.

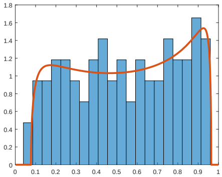




$\perp\ \Delta X_{t-1}$



$\perp\ \Delta X_{t-1}, \Delta X_{t-2}$



$\perp\ \Delta X_{t-1}, \Delta X_{t-2}, \Delta X_{t-3}$

[Figures from (Bykhovskaya-Gorin 2021)]

S&P 100 stocks fit Wachter distribution

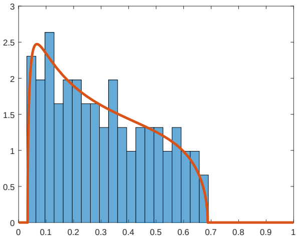
Theorem (Wachter, 1980)

Consider N–particle **Jacobi ensemble**: $1 > x_1 > \cdots > x_N > 0$

$$\sim \prod_{i<j}(x_i - x_j)\prod_{i=1}^{N}(x_i)^{\frac{N}{2}(\mathfrak{p}-1)}(1-x_i)^{\frac{N}{2}(\mathfrak{q}-1)}$$

Set $\lambda_\pm = \frac{1}{(\mathfrak{p}+\mathfrak{q})^2}(\sqrt{\mathfrak{p}(\mathfrak{p}+\mathfrak{q}-1)} \pm \sqrt{\mathfrak{q}})^2$. *Then as* $N \to \infty$:

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{x_i} \longrightarrow \frac{\mathfrak{p}+\mathfrak{q}}{2\pi}\cdot\frac{\sqrt{(x-\lambda_-)(\lambda_+-x)}}{x(1-x)}\mathbf{1}_{[\lambda_-,\lambda_+]}\,dx$$



Weekly S&P stocks fit with:

$\mathfrak{p} = 2$,

$\mathfrak{q} = \frac{\text{number of weeks}}{\text{number of stocks}} - 1.$

S&P 100 stocks fit Wachter distribution

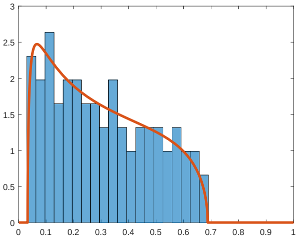
Theorem (Wachter, 1980)

Consider N–particle **Jacobi ensemble***:* $1 > x_1 > \cdots > x_N > 0$

$$\sim \prod_{i<j}(x_i - x_j) \prod_{i=1}^{N} (x_i)^{\frac{N}{2}(\mathfrak{p}-1)} (1 - x_i)^{\frac{N}{2}(\mathfrak{q}-1)}$$

Set $\lambda_\pm = \frac{1}{(\mathfrak{p}+\mathfrak{q})^2}\left(\sqrt{\mathfrak{p}(\mathfrak{p}+\mathfrak{q}-1)} \pm \sqrt{\mathfrak{q}}\right)^2.$ *Then as* $N \to \infty$*:*

$$\frac{1}{N}\sum_{i=1}^{N} \delta_{x_i} \longrightarrow \frac{\mathfrak{p}+\mathfrak{q}}{2\pi} \cdot \frac{\sqrt{(x-\lambda_-)(\lambda_+ - x)}}{x(1-x)} \mathbf{1}_{[\lambda_-,\lambda_+]}\, dx$$



Weekly S&P stocks fit with:

$\mathfrak{p} = 2,$

$\mathfrak{q} = \frac{\text{number of weeks}}{\text{number of stocks}} - 1.$

What does Standard & Poor's has to do with Jacobi?

Basic modeling of a time series

$$\boxed{X_t = X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T}$$

X_t : $N \times 1$	observed vector
ε_t : i.i.d. $\mathcal{N}(0, \Lambda)$	unobserved random innovations
Λ : $N \times N$	unknown covariance matrix
μ : $N \times 1$	unknown intercept vector

$\hookrightarrow$ **high-dimensional random walk**

Basic modeling of a time series

$$\boxed{X_t = X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T}$$

X_t : $N \times 1$	observed vector
ε_t : i.i.d. $\mathcal{N}(0, \Lambda)$	unobserved random innovations
Λ : $N \times N$	unknown covariance matrix
μ : $N \times 1$	unknown intercept vector

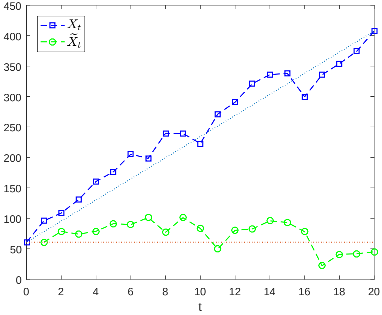
$\hookrightarrow$ **high-dimensional random walk**

[Bykhovskaya–Gorin 2021] **Jacobi ensemble** is hiding already here!

Search for Jacobi

$$X_t = X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \Sigma)$$

Step 1. De-trend the data and define $\tilde{X}_t = X_{t-1} - \frac{t-1}{T}(X_T - X_0)$.



Search for Jacobi

$$X_t = X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \Sigma)$$

Step 1. **De-trend** the data and define $\tilde{X}_t = X_{t-1} - \frac{t-1}{T}(X_T - X_0)$.
Step 2. **De-mean** the data and define

$$R_{0t} = \tilde{X}_t - \frac{1}{T}\sum_{\tau=1}^{T}\tilde{X}_\tau, \quad R_{1t} = (X_t - X_{t-1}) - \frac{1}{T}\sum_{\tau=1}^{T}(X_\tau - X_{\tau-1}).$$

Search for Jacobi

$$\boxed{X_t = X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \Sigma)}$$

Step 1. **De-trend** the data and define $\tilde{X}_t = X_{t-1} - \frac{t-1}{T}(X_T - X_0)$.
Step 2. **De-mean** the data and define

$$R_{0t} = \tilde{X}_t - \frac{1}{T}\sum_{\tau=1}^{T} \tilde{X}_\tau, \quad R_{1t} = (X_t - X_{t-1}) - \frac{1}{T}\sum_{\tau=1}^{T}(X_\tau - X_{\tau-1}).$$

Step 3. Define $S_{ij} = \sum_{t=1}^{T} R_{it} R_{jt}^*, \quad i, j = 0, 1$, and
solve $\det\left(S_{10} S_{00}^{-1} S_{01} - \lambda S_{11}\right) = 0$;
$\hookrightarrow \lambda_1 \geq \ldots \geq \lambda_N =$ eigenvalues of $S_{10} S_{00}^{-1} S_{01} S_{11}^{-1}$.

Summary: These are **squared sample canonical correlations** of X_{t-1} and ΔX_t, after removing constants.

Search for Jacobi

$$\boxed{X_t = X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \Lambda)}$$

$\lambda_1 \geq \ldots \geq \lambda_N$ are **squared sample canonical correlations** of X_{t-1} and ΔX_t, after removing constants.

Theorem (Bykhovskaya-G. 2021)

Suppose $2 + C^{-1} < \frac{T}{N} < C$. *One can couple* $\lambda_1 \geq \cdots \geq \lambda_N$ *and* $x_1 \geq \cdots \geq x_N$ *of the* ***Jacobi ensemble*** $\mathbf{J}(N; \frac{N}{2}, \frac{T-2N}{2})$, *so that for each* $\epsilon > 0$

$$\boxed{\lim_{T,N\to\infty} \text{Prob}\left(\max_{1\leq i\leq N} |\lambda_i - x_i| < \frac{1}{N^{1-\epsilon}}\right) = 1.}$$

$$\mathbf{J}(N; \tfrac{N}{2}, \tfrac{T-2N}{2}) \sim \prod_{i<j}(x_i - x_j) \prod_{i=1}^{N} (x_i)^{\frac{N}{2}-1}(1 - x_i)^{\frac{T-2N}{2}-1}.$$

Wachter distribution for random walks

$$X_t = X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \dots, T, \qquad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \Lambda)$$

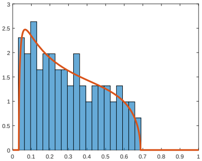
$\lambda_1 \geq \ldots \geq \lambda_N$ are **squared sample canonical correlations** of X_{t-1} and ΔX_t, after removing constants.

Corollary 1 (Bykhovskaya-G. 2021)

$$\lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i} = \frac{\mathfrak{p}+\mathfrak{q}}{2\pi} \cdot \frac{\sqrt{(x-\lambda_-)(\lambda_+ - x)}}{x(1-x)} \mathbf{1}_{[\lambda_-,\lambda_+]}\, dx,$$

$$\lambda_\pm = \tfrac{1}{(\mathfrak{p}+\mathfrak{q})^2}\left(\sqrt{\mathfrak{p}(\mathfrak{p}+\mathfrak{q}-1)} \pm \sqrt{\mathfrak{q}}\right)^2, \quad \mathfrak{p} = 2, \quad \mathfrak{q} = \frac{T}{N} - 1.$$

- Exact match with S&P 100 data.



- Another approach in [Onatski-Wang-2018]

Airy$_1$ point process for random walks

$$X_t = X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \dots, T, \qquad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \Lambda)$$

$\lambda_1 \geq \ldots \geq \lambda_N$ are **squared sample canonical correlations** of X_{t-1} and ΔX_t, after removing constants.

Corollary 2 (Bykhovskaya-G. 2021)

Set $\lambda_\pm = \frac{1}{(\mathfrak{p}+\mathfrak{q})^2}\left(\sqrt{\mathfrak{p}(\mathfrak{p}+\mathfrak{q}-1)} \pm \sqrt{\mathfrak{q}}\right)^2$, $\mathfrak{p} = 2$, $\mathfrak{q} = \frac{T}{N} - 1$. *Then*

$$\lim_{N,T\to\infty} \sqrt[3]{\lambda_+ - \lambda_-}\left(\frac{\mathfrak{p}+\mathfrak{q}}{2\lambda_+(1-\lambda_+)}\right)^{2/3} N^{2/3}\left[\lambda_i - \lambda_+\right]_{i\geq 1} = \left[\mathfrak{a}_i\right]_{i\geq 1},$$

where $[\mathfrak{a}_i]_{i=1}^{\infty}$ *is the Airy*$_1$ *point process.*

- Important for developing statistical **cointegration tests.**
- Similar result for $\beta = 2$ with complex ε_t and Airy$_2$.

Classical appearance of the Jacobi ensemble

Theorem (Hotelling; Fisher; Hsu 1936-1939)

Suppose:

- $X = (N \times T)$ *Gaussian matrix with i.i.d. mean* 0 *columns;*
- $Y = (K \times T)$ *Gaussian matrix with i.i.d. mean* 0 *columns;*
- $N \leq K$, $N + K \leq T$ *and* X **and** Y **are independent**.

Then squared sample canonical correlations of X *and* Y *are*

$$\mathbf{J}\left(N; \tfrac{K-N+1}{2}, \tfrac{T-N-K+1}{2}\right) \sim \prod_{i<j}(x_i - x_j) \prod_{i=1}^{N} (x_i)^{\frac{K-N-1}{2}} (1-x_i)^{\frac{T-N-K-1}{2}}.$$

Classical appearance of the Jacobi ensemble

Theorem (Hotelling; Fisher; Hsu 1936-1939)

Suppose:

- $X = (N \times T)$ *Gaussian matrix with i.i.d. mean 0 columns;*
- $Y = (K \times T)$ *Gaussian matrix with i.i.d. mean 0 columns;*
- $N \le K$, $N + K \le T$ *and* X **and** Y **are independent**.

Then squared sample canonical correlations of X *and* Y *are*

$$\mathbf{J}\left(N; \tfrac{K-N+1}{2}, \tfrac{T-N-K+1}{2}\right) \sim \prod_{i<j}(x_i - x_j) \prod_{i=1}^{N} (x_i)^{\frac{K-N-1}{2}} (1 - x_i)^{\frac{T-N-K-1}{2}}.$$

Key differences:

1. We deal with **(maximally) dependent** X_{t-1} and ΔX_t.
2. We **approximate** by Jacobi ensemble instead of exact match.

Strategy of the proof

$$X_t = X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \Lambda)$$

$\lambda_1 \geq \ldots \geq \lambda_N$ are **squared sample canonical correlations** of X_{t-1} and ΔX_t, after removing constants.

Theorem (Bykhovskaya-G. 2021)

One can couple $\lambda_1 \geq \cdots \geq \lambda_N$ *and* $x_1 \geq \cdots \geq x_N$ *of the* **Jacobi ensemble** $\mathbf{J}(N; \frac{N}{2}, \frac{T-2N}{2})$, *so that*

$$\lim_{T,N\to\infty} \mathrm{Prob}\left(\max_{1\leq i\leq N} |\lambda_i - x_i| < \frac{1}{N^{1-\epsilon}} \right) = 1.$$

Proof: Step 1. Linear algebra + rotational symmetries.

$$(\lambda_1, \ldots, \lambda_N) \stackrel{d}{=} \text{ eigenvalues of } [\tilde{U}]_{NN}\left([\tilde{U}^* \tilde{U}]_{NN}\right)^{-1}[\tilde{U}^*]_{NN}$$

$$\tilde{U} = (I_{T-1} - oFo^*)^{-1}, \quad o \sim \text{ Haar-random in } SO(T-1).$$

$F \sim$ deterministic with eigenvalues evenly spaced on the unit circle

Strategy of the proof

Theorem

$$\lim_{T,N\to\infty} \text{Prob}\left(\max_{1\le i\le N} |\lambda_i - x_i| < \frac{1}{N^{1-\epsilon}}\right) = 1.$$

Proof: Step 1. Linear algebra + rotational symmetries.

$$(\lambda_1, \ldots, \lambda_N) \overset{d}{=} \text{ eigenvalues of } [\tilde{U}]_{NN}\left([\tilde{U}^*\tilde{U}]_{NN}\right)^{-1}[\tilde{U}^*]_{NN}$$

$$\tilde{U} = (I_{T-1} - oFo^*)^{-1}, \quad o \sim \text{ Haar-random in } SO(T-1).$$

$F \sim$ deterministic with eigenvalues evenly spaced on the unit circle

Step 2. A new matrix integral leading to the **Jacobi ensemble**.

$$(x_1, \ldots, x_N) \overset{d}{=} \text{ eigenvalues of } [U]_{NN}\left([U^*U]_{NN}\right)^{-1}[U^*]_{NN}$$

$$U = (I_{T-1} - Q)^{-1}, \quad Q \sim \text{ Haar-random in } SO(T-1).$$

Strategy of the proof

Theorem

$$\lim_{T,N\to\infty} \mathrm{Prob}\left(\max_{1\le i\le N} |\lambda_i - x_i| < \frac{1}{N^{1-\epsilon}}\right) = 1.$$

Proof: Step 1. Linear algebra + rotational symmetries.

$$(\lambda_1, \dots, \lambda_N) \overset{d}{=} \text{ eigenvalues of } [\tilde{U}]_{NN}\left([\tilde{U}^* \tilde{U}]_{NN}\right)^{-1}[\tilde{U}^*]_{NN}$$

$$\tilde{U} = (I_{T-1} - oFo^*)^{-1}, \quad o \sim \text{ Haar-random in } SO(T-1).$$

$F \sim$ deterministic with eigenvalues evenly spaced on the unit circle

Step 2. A *new* matrix integral leading to the **Jacobi ensemble**.

$$(x_1, \dots, x_N) \overset{d}{=} \text{ eigenvalues of } [U]_{NN}\left([U^* U]_{NN}\right)^{-1}[U^*]_{NN}$$

$$U = (I_{T-1} - Q)^{-1}, \quad Q \sim \text{ Haar-random in } SO(T-1).$$

Step 3. Rigidity for eigenvalues of Q: $\quad Q \overset{d}{=} oFo^* +$ small error.

Analogy: **canonical** vs **grand canonical** ensembles.

Strategy of the proof

Theorem

$$\lim_{T,N\to\infty} \mathrm{Prob}\left(\max_{1\le i\le N} |\lambda_i - x_i| < \frac{1}{N^{1-\epsilon}}\right) = 1.$$

Proof: Step 1. Linear algebra + rotational symmetries.

$$(\lambda_1,\dots,\lambda_N) \overset{d}{=} \text{ eigenvalues of } [\tilde{U}]_{NN}\left([\tilde{U}^*\tilde{U}]_{NN}\right)^{-1}[\tilde{U}^*]_{NN}$$

$$\tilde{U} = (I_{T-1} - oFo^*)^{-1}, \quad o \sim \text{ Haar-random in } SO(T-1).$$

$F \sim$ deterministic with eigenvalues evenly spaced on the unit circle

Step 2. A *new* matrix integral leading to the **Jacobi ensemble**.

$$(x_1,\dots,x_N) \overset{d}{=} \text{ eigenvalues of } [U]_{NN}\left([U^*U]_{NN}\right)^{-1}[U^*]_{NN}$$

$$U = (I_{T-1} - Q)^{-1}, \quad Q \sim \text{ Haar-random in } SO(T-1).$$

Step 3. Rigidity for eigenvalues of Q: $\quad Q \overset{d}{=} oFo^* + $ small error.

Step 4. Show that small error *passes through* inversions. ∎

Cointegration

Why do we care about canonical correlations of X_t and ΔX_t?

Cointegration

Why do we care about canonical correlations of X_t and ΔX_t?

A more general time-series model:

$$\boxed{X_t = X_{t-1} + \Pi X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d.}}$$

$N \times N$ matrix Π allows for richer temporal behavior.

- $\Pi = 0$: coordinates "grow" like random walks.
- $\Pi = -I_N$: coordinates are stationary in time (i.i.d.).

Cointegration

Why do we care about canonical correlations of X_t and ΔX_t?

A more general time-series model:

$$\boxed{X_t = X_{t-1} + \Pi X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \dots, T, \qquad \varepsilon_t \sim \text{i.i.d.}}$$

$N \times N$ matrix Π allows for richer temporal behavior.

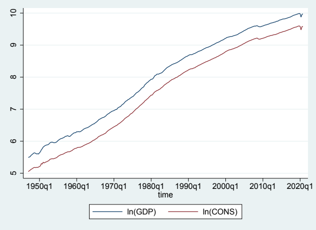
- $\Pi = 0$: coordinates "grow" like random walks.
- $\Pi = -I_N$: coordinates are stationary in time (i.i.d.).
- Π of rank r $<\approx>$ there are r **cointegrating relations** — stationary linear combinations of non-stationary coordinates.

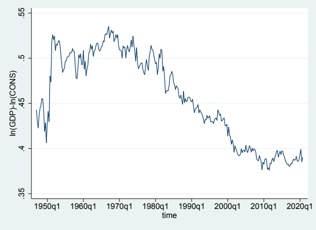
[need a technical condition to get rid of a "faster than random walk" growth cases]

Cointegration is an important topic in economics.

- Π of rank r $<\approx>$ there are r **cointegrating relations** — stationary linear combinations of non-stationary coordinates.

Example:
ln(GDP) and ln(consumption) are cointegrated.





The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003

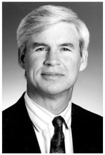


Photo from the Nobel Foundation archive.
Robert F. Engle III
Prize share: 1/2

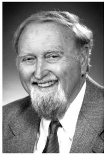


Photo from the Nobel Foundation archive.
Clive W.J. Granger
Prize share: 1/2

TITLE	CITED BY	YEAR
Co-integration and error correction: representation, estimation, and testing RF Engle, CWJ Granger Econometrica: journal of the Econometric Society, 251-276	44357	1987

TITLE	CITED BY	YEAR
Statistical analysis of cointegration vectors S Johansen Journal of economic dynamics and control 12 (2-3), 231-254	25794	1988
Maximum likelihood estimation and inference on cointegration—with appucations to the demand for money S Johansen, K Juselius Oxford Bulletin of Economics and statistics 52 (2), 169-210	18848	1990
Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models S Johansen Econometrica: journal of the Econometric Society, 1551-1580	14219	1991

Cointegration tests

$$\boxed{X_t = X_{t-1} + \Pi X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d.}}$$

- $\Pi = 0$: coordinates "grow" like random walks.
- $\Pi = -I_N$: coordinates are stationary in time (i.i.d.).
- Π of rank r $<\approx>$ there are r **cointegrating relations**

How to statistically **test the rank of Π?**

Cointegration tests

$$X_t = X_{t-1} + \Pi X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d.}$$

- $\Pi = 0$: coordinates "grow" like random walks.
- $\Pi = -I_N$: coordinates are stationary in time (i.i.d.).
- Π of rank r: $\approx$ there are r **cointegrating relations**

Theorem (Anderson 1951; Johansen 1988)

The (Gaussian) **maximum likelihood ratio test** *of*

$$H_0: \quad \Pi = 0 \qquad vs \qquad H_1: \quad rank(\Pi) \leq r$$

is based on the value of the statistic

$$\mathbf{LR} = -\sum_{i=1}^{r} \ln(1 - \lambda_i).$$

$\lambda_1 \geq \cdots \geq \lambda_N \approx$ *squared sample canonical correlations of* X_{t-1} *and* ΔX_t

Cointegration tests

$$X_t = X_{t-1} + \Pi X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d.}$$

- $\Pi = 0$: coordinates "grow" like random walks.
- $\Pi = -I_N$: coordinates are stationary in time (i.i.d.).
- Π of rank r: $\approx$ there are r **cointegrating relations**

Theorem (Anderson 1951; Johansen 1988)

The (Gaussian) **maximum likelihood ratio test** *of*

$$H_0 : \quad \Pi = 0 \qquad vs \qquad H_1 : \quad rank(\Pi) \leq r$$

is based on the value of the statistic

$$\mathbf{LR} = -\sum_{i=1}^{r} \ln(1 - \lambda_i).$$

$\lambda_1 \geq \cdots \geq \lambda_N \approx$ *squared sample canonical correlations of* X_{t-1} *and* ΔX_t

We **reject** H_0 if **LR** is atypically large ($>$ 95% percentile).

Cointegration tests

$$\boxed{X_t = X_{t-1} + \mathit{\Pi} X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d.}}$$

Need: asymptotics of $\mathbf{LR} = -\sum_{i=1}^{r} \ln(1 - \lambda_i)$ under $H_0 : \mathit{\Pi} = 0$.

$\lambda_1 \geq \cdots \geq \lambda_N \approx$ squared sample canonical correlations of X_{t-1} and $\mathit{\Delta} X_t$

Cointegration tests

$$X_t = X_{t-1} + \Pi X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d.}$$

Need: asymptotics of $\mathbf{LR} = -\sum_{i=1}^{r} \ln(1 - \lambda_i)$ under $H_0 : \Pi = 0$.

$\lambda_1 \geq \cdots \geq \lambda_N \approx$ squared sample canonical correlations of X_{t-1} and ΔX_t

Classical results. (Johansen 1988, 1991) Limit theorems based on **fixed** N and **large** $T \longrightarrow$ integral functionals of Brownian motions.

- Widely used.
- Perform badly for intermediate N.
- [Onatski-Wang 2018]: explanation of bad performance based on **joint** $N, T \to \infty$, $T/N \to c$ asymptotics.

Cointegration tests

$$X_t = X_{t-1} + \Pi X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \dots, T, \qquad \varepsilon_t \sim \text{i.i.d.}$$

$\lambda_1 \geq \cdots \geq \lambda_N \approx$ squared sample canonical correlations of X_{t-1} and ΔX_t

Corollary 3 (Bykhovskaya-G. 2021)

Suppose $2 + C^{-1} < T/N < C$ *and* $\mathbf{H_0}$ **holds**: $\mathbf{\Pi} = \mathbf{0}$. *Then*

$$\frac{\sum_{i=1}^{r} \ln(1-\lambda_i) - r \cdot c_1(N,T)}{N^{-2/3} c_2(N,T)} \xrightarrow[T,N\to\infty]{d} \sum_{i=1}^{r} \mathfrak{a}_i,$$

where $$c_1(N,T) = \ln(1-\lambda_+),$$

$$c_2(N,T) = -\frac{2^{2/3}\lambda_+^{2/3}}{(1-\lambda_+)^{1/3}(\lambda_+ - \lambda_-)^{1/3}} (\mathfrak{p}+\mathfrak{q})^{-2/3} < 0,$$

$$\mathfrak{p} = 2, \quad \mathfrak{q} = \frac{T}{N} - 1, \quad \lambda_\pm = \frac{1}{(\mathfrak{p}+\mathfrak{q})^2}\left[\sqrt{\mathfrak{p}(\mathfrak{p}+\mathfrak{q}-1)} \pm \sqrt{\mathfrak{q}}\right]^2.$$

Finite sample performance of tests

Our Test

	N	$LR_{N,T}$	LR	RALR
	5	6.60	20.75	3.59
	6	5.45	31.66	2.68
$T = 30$	7	4.52	47.44	1.98
	8	3.80	67.42	2.00
	9	3.16	85.00	1.32
	10	2.60	96.69	0.96

Empirical size under no cointegration hypothesis (5% nominal level). DGP: $\Delta X_{it} = \varepsilon_{it}$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $MC = 1{,}000{,}000$ for $LR_{N,T}$ and $MC = 10{,}000$ for LR and RALR.

Critical values for H_0 rejection based on:

- $LR_{N,T}$ — our asymptotic theorem.
- LR — Johansen's asymptotic theorem.
- $RALR$ — empirical correction to LR of [Reinsel-Ahn 1992].

Open problems

Universality:

Conjecture

The Airy_1 asymptotic behavior for largest eigenvalues and tests extends to **non-Gaussian innovations** ε_t*. All we need is the existence of second moments.*

Conjecture

The Airy_1 asymptotic behavior for largest eigenvalues and tests extends to different treatments of constants μ (de-trending/de-meaning steps in our procedure).

Open problems

Universality:

Conjecture

The Airy_1 asymptotic behavior for largest eigenvalues and tests extends to **non-Gaussian innovations** ε_t*. All we need is the existence of second moments.*

Conjecture

The Airy_1 asymptotic behavior for largest eigenvalues and tests extends to different treatments of constants μ (de-trending/de-meaning steps in our procedure).

Integrability:

Question

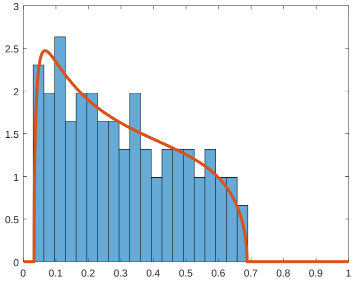
What is the law of $\sum_{i=1}^{r} \mathfrak{a}_i$, where $(\mathfrak{a}_i)_{i=1}^{\infty}$ is the Airy_1 point process?

(For Airy_2 point process this is also unknown.)

Summary

$$X_t = X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \Lambda)$$

1. **Squared sample canonical correlations** of a high-dimensional random walk and its time-increments are closely approximated by the **Jacobi ensemble** $\mathbf{J}(N; \frac{N}{2}, \frac{T-2N}{2})$.
2. Consistent with behavior of logarithms of S&P 100 stocks.



3. Leads to **cointegration tests** with superior performance.
4. **No** cointegration in S&P.