Cointegration with infinite variance noise

Keith Knight

Department of Statistics
University of Toronto

Joint work with Mahinda Samarakoon

Research supported by the Natural Sciences and Engineering Research Council of Canada.

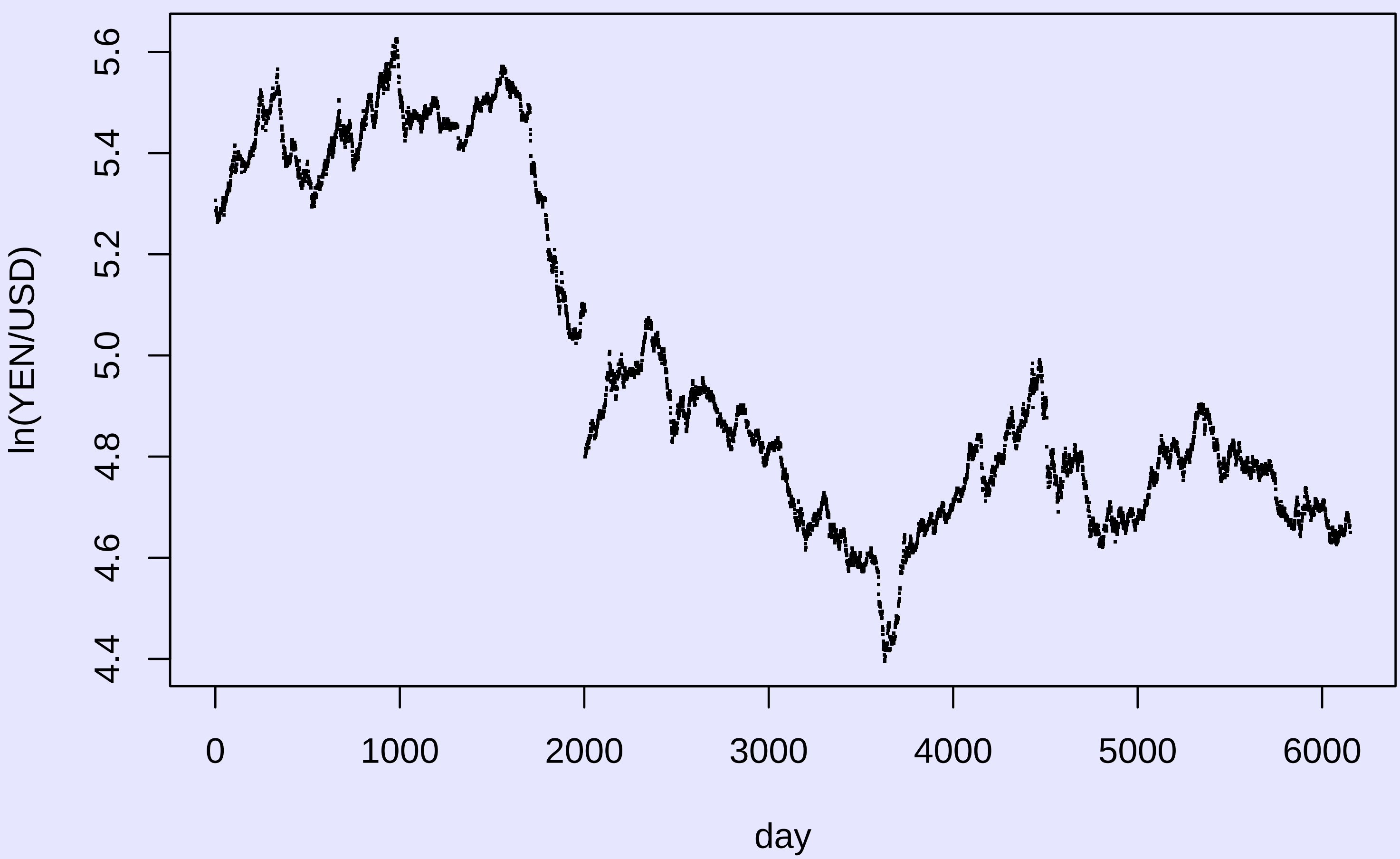
Outline

1. Introduction
   - Heavy tails
   - Cointegration
2. Asymptotics
   - Convergence to stable laws and processes
   - Asymptotics for M-estimators under cointegration
   - Operator stable distributions
3. Final commments

1. INTRODUCTION

Time series analysis with heavy tails

- Mandelbrot (1963, 1967) and Fama (1965) observed that distributions of stock returns are often heavy tailed with possibly infinite variance.
- Since that time, there has been extensive work on examining the plausibility of the infinite variance model.
- Philosophical/modeling question: Are variances infinite or finite with stochastic heteroscedasticity?



Yen per US dollar exchange rate: Dec 1978 to Apr 2005

- A partial list of research in this area includes:
  - Stationary time series: Davis & Resnick (1995, 1996), Davis, Knight & Liu (1992), Anderson & Meerschaert (1997).
  - Unit root testing: Chan & Tran (1989), Knight (1989), Phillips (1990), Rachev, Mittnik & Kim (1998), Hasan (2001), Ahn, Fotopoulos & He (2001), Samarakoon & Knight (2006).
  - Cointegration testing: Caner (1998), Paulauskas & Rachev (1998).
  - Applications: Koedijk and Kool (1992), Falk and Wang (2003), Charemza, Hristova & Burridge (2005), Kirman and Teyssière (2005).

- Classical estimation procedures (typically based on the assumption of normally distributed innovations) perform reasonably well under non-normal noise conditions (when used carefully).
  - For integrated and cointegrated processes, least squares convergence rates are equal for finite and infinite variance noise.
- But .... we can improve on least squares, often substantially.
  - Isolated large shocks to a system provide potentially a lot of information on the system dynamics.
  - Potentially faster convergence rates.

Example: AR(1) process with infinite variance errors

- Define $X_t = \phi X_{t-1} + \varepsilon_t$.
- Estimate ϕ by regressing X_t on X_{t-1}.
- If standard asymptotics carry over, we should be able to estimate ϕ so that

$$\widehat{\phi}_n - \phi = O_p\left\{\left(\sum_{t=2}^{n} X_{t-1}^2\right)^{-1/2}\right\}$$

- Thus we should have faster convergence rates for infinite variance $\{\varepsilon_t\}$ since $\sum X_{t-1}^2$ is increasing at a faster rate in this case.
- But ... least squares estimation does not generally produce the fastest possible rate of convergence.

What is cointegration?

- A univariate stochastic process $\{X_t\}$ is **integrated** if it is non-stationary but its first differences $\nabla X_t = X_t - X_{t-1}$ are stationary.
- If $\{X_t\}$ and $\{Y_t\}$ are both integrated then $\{(X_t, Y_t)\}$ are **cointegrated** if $\{X_t + aY_t\}$ is stationary for some a.
- If $\{\boldsymbol{X}_t\}$ is a vector process whose elements are all non-stationary then it is cointegrated if $\{\boldsymbol{a}^\top \boldsymbol{X}_t\}$ is stationary for some $\boldsymbol{a}$ (called a cointegration vector).
- Economic interpretation: individual variables behave like random walks but are collectively in equilibrium.

Testing for cointegration: Two basic approaches

- Find an estimator $\widehat{\boldsymbol{a}}$ of $\boldsymbol{a}$ (for example, using regression) and test if $\{\widehat{\boldsymbol{a}}^\top \boldsymbol{X}_t\}$ is stationary.
  - For example, use a Dickey-Fuller test (or other unit root test) on $\{\widehat{\boldsymbol{a}}^\top \boldsymbol{X}_t\}$.
- Assume a parametric model (for example, VAR) for $\{\boldsymbol{X}_t\}$ and test for cointegration within that model.

- Assume a VAR(k) model for $\{\boldsymbol{X}_t\}$; we will write this in its error correction form

$$\nabla \boldsymbol{X}_t = \Pi \boldsymbol{X}_{t-k} + \Phi_1 \nabla \boldsymbol{X}_{t-1} + \cdots + \Phi_{k-1} \boldsymbol{X}_{t-k+1} + \boldsymbol{\varepsilon}_t.$$

- We will assume that the components of $\{\boldsymbol{\varepsilon}_t\}$ have infinite variance, either
  - in the domain of attraction of a multivariate stable law, or
  - in the domain of attraction of an operator stable law.
- If $\{\nabla X_t\}$ is stationary,
  - $\Pi = 0$ implies that $\{\boldsymbol{X}_t\}$ is integrated but not cointegrated;
  - Π has full rank implies that $\{\boldsymbol{X}_t\}$ is stationary;
  - $\Pi \neq 0$ but less than full rank implies that $\{\boldsymbol{X}_t\}$ is cointegrated.

- Granger representation of $\{\boldsymbol{X}_t\}$:

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + A\left\{B^\top(I - \Phi_1 - \cdots - \Phi_{k-1})A\right\}^{-1} B^\top \sum_{u=1}^{t} \boldsymbol{\varepsilon}_t + \boldsymbol{\zeta}_t$$

  where

  - $\{\boldsymbol{\zeta}_t\}$ is stationary;
  - $B^\top\Pi = \Pi A = 0$ for maximal rank matrices A and B.

- $\{\boldsymbol{X}_t\}$ looks like a random walk in $r = \text{rank}(A) = \text{rank}(B)$ dimensions.
  - Π full rank: $A = B = 0$, $\boldsymbol{X}_t = \boldsymbol{X}_0 + \boldsymbol{\zeta}_t$.
  - $\Pi = 0$: $A = B = I$,

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + (I - \Phi_1 - \cdots - \Phi_{k-1})^{-1} \sum_{u=1}^{t} \boldsymbol{\varepsilon}_t + \boldsymbol{\zeta}_t$$

- Define the cointegration space of $\{\boldsymbol{X}_t\}$:

$$\mathcal{C} = \{\boldsymbol{a} : \{\boldsymbol{a}^\top \boldsymbol{X}_t\} \text{ is stationary}\}$$

  $\mathcal{C}$ is simply the row space of Π.

- Cointegration rank is determined essentially by finding good lower rank approximations to an unconstrained (and typically full rank) estimator of Π.
  - Start by testing $H_0 : \Pi = 0$.

- Finite variance errors: look at canonical correlations between $\{\nabla \boldsymbol{X}_t\}$ and $\{\boldsymbol{X}_{t-k}\}$, adjusted for $\boldsymbol{X}_{t-1}, \cdots, \boldsymbol{X}_{t-k+1}$.
  - Johansen (1988, 1991, ...) develops asymptotic distribution theory.

- We will consider component-by-component M-estimators of the parameters in the model.
- Define Y_t to be an arbitrary component of ∇X_t.
- Our M-estimators minimize

$$\sum_{t=k+1}^{n} \rho(Y_t - \boldsymbol{X}_{t-k}^{\top}\boldsymbol{\pi} - \nabla X_{t-1}^{\top}\boldsymbol{\phi}_1 - \cdots - \nabla X_{t-k+1}^{\top}\boldsymbol{\phi}_{k-1})$$

  over some appropriate space where ρ is a convex function increasing slower than x^2.
- These estimators can be "stacked" to give estimators of Π, $\Phi_1, \cdots, \Phi_{k-1}$.

2. ASYMPTOTICS

Stable laws and processes

- Assume that the innovations $\{\boldsymbol{\varepsilon}_t\}$ to lie in the domain of attraction of a multivariate stable law with index $\alpha \in (0,2)$.
- This means that

$$P\left(\|\boldsymbol{\varepsilon}_t\| > x\right) = x^{-\alpha} L(x)$$

where L is a slowly varying function, and for unit vectors $\boldsymbol{a}$,

$$\lim_{x\to\infty} \frac{P\left(\|\boldsymbol{\varepsilon}_t\| > x, \boldsymbol{\varepsilon}_t/\|\boldsymbol{\varepsilon}_t\| \in A\right)}{P\left(\|\boldsymbol{\varepsilon}_t\| > x\right)} = \nu(A)$$

for some measure A.

- Note that this assumption is quite restrictive — it implies the same tail index in every direction.

- Under these assumptions, we have

$$a_n^{-1} \sum_{t=1}^{n} (\boldsymbol{\varepsilon}_t - \boldsymbol{b}_n) \xrightarrow{d} \boldsymbol{S}_\alpha$$

  where $\boldsymbol{S}_\alpha$ is an α-stable random vector.

- $a_n = n^{1/\alpha} L^*(n)$ where L^* is another slowly varying function.

- We will assume in this talk that $\boldsymbol{b}_n = \mathbf{0}$ (i.e. no drift).
  - When $\alpha > 1$, this means $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$.
  - When $\alpha < 1$, we can always set $\boldsymbol{b}_n = \mathbf{0}$.

- Define the two partial sum processes

$$\boldsymbol{S}_n(u) = a_n^{-1} \sum_{t=1}^{\lfloor nu \rfloor} \boldsymbol{\varepsilon}_t$$

and

$$W_n(u) = n^{-1/2} \sum_{t=1}^{\lfloor nu \rfloor} \phi(\boldsymbol{\varepsilon}_t).$$

where $E[\phi(\boldsymbol{\varepsilon}_t)] = 0$ and $E[\phi^2(\boldsymbol{\varepsilon}_t)] < \infty$.

- $\boldsymbol{S}_n$ and W_n converge weakly to independent processes (Resnick and Greenwood, 1979):
  - $\boldsymbol{S}_n \xrightarrow{d} \boldsymbol{S}_\alpha$, a stable process;
  - $W_n \xrightarrow{d} W$, a Brownian motion.

Asymptotics for M-estimation

- Asymptotic distribution theory for estimators of Π combines the techniques used in
  - Davis *et al.* (1992) for stationary AR processes,
  - Knight (1989, 1991) for the unit root AR(1) process, and
  - Samarakoon & Knight (2006) for general unit root tests.
- The asymptotics depend on whether we do unconstrained minimization or minimize over $\pi \in \mathcal{C}^{\perp}$.
  - unconstrained: point process (i.e. non-standard) asymptotics.
  - constrained: more classical asymptotics involving a stable process and a Brownian motion.

What are the regularity conditions?

- $\{\boldsymbol{\varepsilon}_t\}$ are in the domain of attraction of a stable law with index $\alpha \in (0,2)$ with $\boldsymbol{b}_n = 0$;

- ρ is a convex function with derivatives $\psi = \rho'$ and $\psi' = \rho''$ satisfying

$$\begin{aligned} |\psi(x+y) - \psi(x)| &\leq K_1 |y|^{\delta_1} \quad \text{and} \\ |\psi'(x+y) - \psi'(x)| &\leq K_2 |y|^{\delta_2} \end{aligned}$$

  where $\delta_1 > \max\{2(\alpha-1)/\alpha, 0\}$, $\delta_2 > 0$, and K_1, K_2 are positive constants;

- $E[\psi(\varepsilon_{ti})] = 0$, $E[\psi^2(\varepsilon_{ti})] < \infty$, and $0 < E[\psi'(\varepsilon_{ti})] < \infty$ where $\boldsymbol{\varepsilon}_t = (\varepsilon_{t1}, \cdots, \varepsilon_{tp})^\top$.

Results: Focus on estimation of Π with rows constrained to $\mathcal{C}^{\perp}$.

- If we minimize over $\pi \in \mathcal{C}^{\perp}$ then

  $$n^{1/2} a_n A^\top \widehat{\Pi}_n^\top \xrightarrow{d} \left(\int_0^1 A^\top \boldsymbol{S}_\alpha(s) \boldsymbol{S}_\alpha^\top(s) A \, ds\right)^{-1} \left(\int_0^1 A^\top \boldsymbol{S}_\alpha(s) \, d\boldsymbol{W}^\top(s)\right) \Gamma^{-1}$$

  where

  - columns of A are an orthonormal basis for $\mathcal{C}^{\perp}$;
  - $\boldsymbol{W}$ is a zero-mean Gaussian process with

    $$E[\boldsymbol{W}(s_1)\boldsymbol{W}^\top(s_2)] = \min(s_1, s_2)\Sigma, \;\; \Sigma = \Big(\mathrm{Cov}[\psi(\varepsilon_{ti}), \psi(\varepsilon_{tj})]\Big);$$

  - $\Gamma = \mathrm{diag}(E[\psi'(\varepsilon_{t1})], \cdots, E[\psi'(\varepsilon_{tp})])$.

- Faster convergence than LS: $O_p(n^{-1/2} a_n^{-1})$ vs $O_p(n^{-1})$.

- Given $\widehat{\Gamma}_n$ and $\widehat{\Sigma}_n$ consistent estimators of Γ and Σ then

$$\mathcal{T}_n = \Upsilon_n^\top \left(\widehat{\Pi}_n A\right)^\top \left(\widehat{\Gamma}_n \widehat{\Sigma}_n^{-1} \widehat{\Gamma}_n\right) \left(\widehat{\Pi}_n A\right) \Upsilon_n \xrightarrow{d} \mathcal{W}_r(p, I),$$

  a standard Wishart distribution with $r = \dim(\mathcal{C}^\perp)$ where

$$\Upsilon_n \Upsilon_n^\top = A^\top \left(\sum_{t=k+1}^{n} \boldsymbol{X}_{t-k} \boldsymbol{X}_{t-k}^\top \right) A.$$

- To test $H_0 : \mathcal{C} = \mathcal{C}_0$, use test statistics based on the eigenvalues of $\mathcal{T}_n$ whose asymptotic distribution theory is relatively straightforward.

- In contrast, the "classical" (i.e. finite variance) asymptotic theory is much more complicated.

- No uniformly optimal test statistic based on the eigenvalues of $\mathcal{T}_n$ exists.
- Two natural possibilities: maximum eigenvalue and trace.
- Maximum eigenvalue statistic: suggests a new subspace to be added to $\mathcal{C}_0$.
  - Limiting distribution can be evaluated analytically, albeit painfully (Muirhead, 1982) or via simulation.
- Trace statistic: more of an omnibus test.
  - χ^2 limiting distribution.

Note: This latter asymptotic result does *not* depend on α.

Question: Can we weaken the assumption on $\{\boldsymbol{\varepsilon}_t\}$ so that this result still holds?

- We want to allow projections of $\boldsymbol{\varepsilon}_t$ to have different tail indices.
- Replace normalizing constants $\{a_n\}$ by normalizing matrices $\{\Delta_n\}$.

Solution: Consider domains of attraction of operator stable laws.

Example: $\{X_i\}$, $\{Y_i\}$ i.i.d. sequences with $E(X_i) = 0$, $E(X_i^2) = 1$, $Y_i \sim$ Cauchy.

- Define

$$\mathbf{U}_i = \begin{pmatrix} X_i + Y_i \\ X_i - Y_i \end{pmatrix}.$$

- Elements of $\mathbf{U}_i$ are in the domain of attraction of a Cauchy distribution and

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{U}_i \xrightarrow{d} \begin{pmatrix} Y_0 \\ -Y_0 \end{pmatrix}$$

  where $Y_0 \sim$ Cauchy.

- The limiting distribution is concentrated on a one-dimensional subspace of R^2.

- We get a more interesting limiting distribution by normalizing the partial sum by matrices.
- Define

$$\Delta_n = \begin{pmatrix} n^{1/2} & n \\ n^{1/2} & -n \end{pmatrix}$$

- Then

$$\Delta_n^{-1} \sum_{i=1}^{n} \mathbf{U}_i \xrightarrow{d} \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix}$$

where X_0 and Y_0 are independent, $X_0 \sim \mathcal{N}(0,1)$ and $Y_0 \sim$ Cauchy.

What are operator stable laws?

- Limits of partial sums are operator stable laws P_E, where the index E is a matrix.
- If $\mathbf{U}_1, \cdots, \mathbf{U}_n$ are i.i.d. P_E then for some $\boldsymbol{b}_n$,

$$n^{-E} \sum_{i=1}^{n} \mathbf{U}_i - \boldsymbol{b}_n \sim P_E$$

  where

$$n^{-E} = \exp[-E \ln(n)] = \sum_{k=0}^{\infty} \frac{(-1)^k \ln^k(n) E^k}{k!}.$$

- Canonical form of the characteristic function was given by Sharpe (1969).
- Applications: Meerschaert & Scheffler (2000, 2001).

- The matrix E has eigenvalues $\lambda_1, \cdots, \lambda_p$ with $\mathrm{Re}(\lambda_j) \geq 1/2$.
- $\mathrm{Re}(\lambda_j)$ $(j = 1, \cdots, p)$ play the role of $1/\alpha$:
  - If $\mathrm{Re}(\lambda_j) > 1/2$ for all j then P_E is an infinite variance operator stable law.
  - $\mathrm{Re}(\lambda_j) = 1/2$ corresponds to a Gaussian component that is independent of the infinite variance components.
- P_E must *not* be concentrated on a lower dimensional hyperplane.
  - A lower dimensional projection of an operator stable distribution is not necessarily operator stable.
  - But ... one-dimensional projections have potentially different tail indices.

- An i.i.d. sequence $\{\mathbf{U}_i\}$ is in the domain of attraction of P_E if there exists a sequence of matrices $\{\Delta_n\}$ and vectors $\{\boldsymbol{b}_n\}$ such that

$$\Delta_n^{-1} \sum_{i=1}^{n} \mathbf{U}_i - \boldsymbol{b}_n \xrightarrow{d} P_E.$$

  - $\{\Delta_n\}$ is regularly varying in the following sense:

$$\lim_{n\to\infty} \Delta_{\lfloor sn \rfloor} \Delta_n^{-1} = s^E \quad \text{for each } s > 0.$$

- If there's no Gaussian component then for any set D bounded away from $\mathbf{0}$, we have

$$\lim_{n\to\infty} nP(\Delta_n^{-1}\mathbf{U}_i \in D) = \phi(D)$$

Example: Use Δ_n from earlier example:

$$\Delta_{\lfloor sn \rfloor} \Delta_n^{-1} = \frac{1}{2} \begin{pmatrix} s^{1/2} + s & s^{1/2} - s \\ s^{1/2} - s & s^{1/2} + s \end{pmatrix}$$

- The eigenvalues of $\Delta_{\lfloor sn \rfloor} \Delta_n^{-1}$ are $s^{1/2}$ and s and the eigenvectors are $(1, \pm 1)^\top$ so that

$$\Delta_{\lfloor sn \rfloor} \Delta_n^{-1} = s^E$$

where

$$E = \begin{pmatrix} 3/4 & -1/4 \\ -1/4 & 3/4 \end{pmatrix}$$

has eigenvalues $1/2$ and 1.

Application to cointegration

- Recall Granger representation of $\{\boldsymbol{X}_t\}$:

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + A\left\{B^\top(I - \Phi_1 - \cdots - \Phi_{k-1})A\right\}^{-1} B^\top \sum_{u=1}^{t} \boldsymbol{\varepsilon}_t + \boldsymbol{\zeta}_t$$

  with $\{\boldsymbol{\zeta}_t\}$ stationary.

- Assume that $\{B^\top \boldsymbol{\varepsilon}_t\}$ lie in the domain of attraction of an operator stable distribution:

$$\Lambda_n^{-1} \sum_{t=1}^{n} B^\top \boldsymbol{\varepsilon}_t \xrightarrow{d} \boldsymbol{V} \sim P_E$$

  for some E and some sequence of matrices $\{\Lambda_n\}$.

- Includes earlier assumption on $\{\boldsymbol{\varepsilon}_t\}$ as a special case.

- Look at asymptotic behaviour of $\{\boldsymbol{X}_t\}$ on $\mathcal{C}^{\perp}$.
- Redefine $\boldsymbol{S}_n$ as follows:

$$\begin{aligned}\boldsymbol{S}_n(u) &= \Delta_n^{-1} A^{\top} \boldsymbol{X}_{\lfloor nu \rfloor} \\ &= \Lambda_n^{-1} \sum_{t=1}^{\lfloor nu \rfloor} B^{\top} \boldsymbol{\varepsilon}_t + o_p(1)\end{aligned}$$

  where

$$\Delta_n^{-1} = \Lambda_n^{-1} \left\{ B^{\top}(I - \Phi_1 - \cdots - \Phi_{k-1}) A \right\}$$

- $\boldsymbol{S}_n \xrightarrow{f-d} \boldsymbol{S}_E$, a operator stable Lévy process.

- Under the operator stable assumption plus regularity conditions on ρ, we have

$$n^{1/2}\Delta_n^\top A^\top \widehat{\Pi}_n^\top \xrightarrow{d} \left(\int_0^1 \boldsymbol{S}_E(s)\boldsymbol{S}_E^\top(s)\,ds\right)^{-1}\left(\int_0^1 \boldsymbol{S}_E(s)\,d\boldsymbol{W}^\top(s)\right)\Gamma^{-1}$$

where $\boldsymbol{W}$ is the same Gaussian process as before.

- We also have (as before)

$$\mathcal{T}_n = \Upsilon_n^\top \left(\widehat{\Pi}_n A\right)^\top \left(\widehat{\Gamma}_n \widehat{\Sigma}_n^{-1}\widehat{\Gamma}_n\right)\left(\widehat{\Pi}_n A\right)\Upsilon_n \xrightarrow{d} \mathcal{W}_r(p, I).$$

- Limiting distribution is independent of E, $\{\Delta_n\}$ — we don't need to estimate tail indices!

3. FINAL COMMENTS

- The results can be extended to allow drift and other $I(0)$ terms (including an intercept) in the model.
  - Need only correct for estimation of these additional parameters.
- Asymptotic theory for estimators of $\Phi_1, \cdots, \Phi_{k-1}$ is non-standard — point process asymptotics.
- Open question: Is a "domain of attraction" assumption necessary?
  - Does $\mathcal{T}_n \xrightarrow{d} \mathcal{W}_r(p, I)$ if $\boldsymbol{a}^\top \boldsymbol{\varepsilon}_t$ has infinite variance for all non-zero $\boldsymbol{a}$?
- Extensions to domains of attraction with a Normal component also are possible.