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.

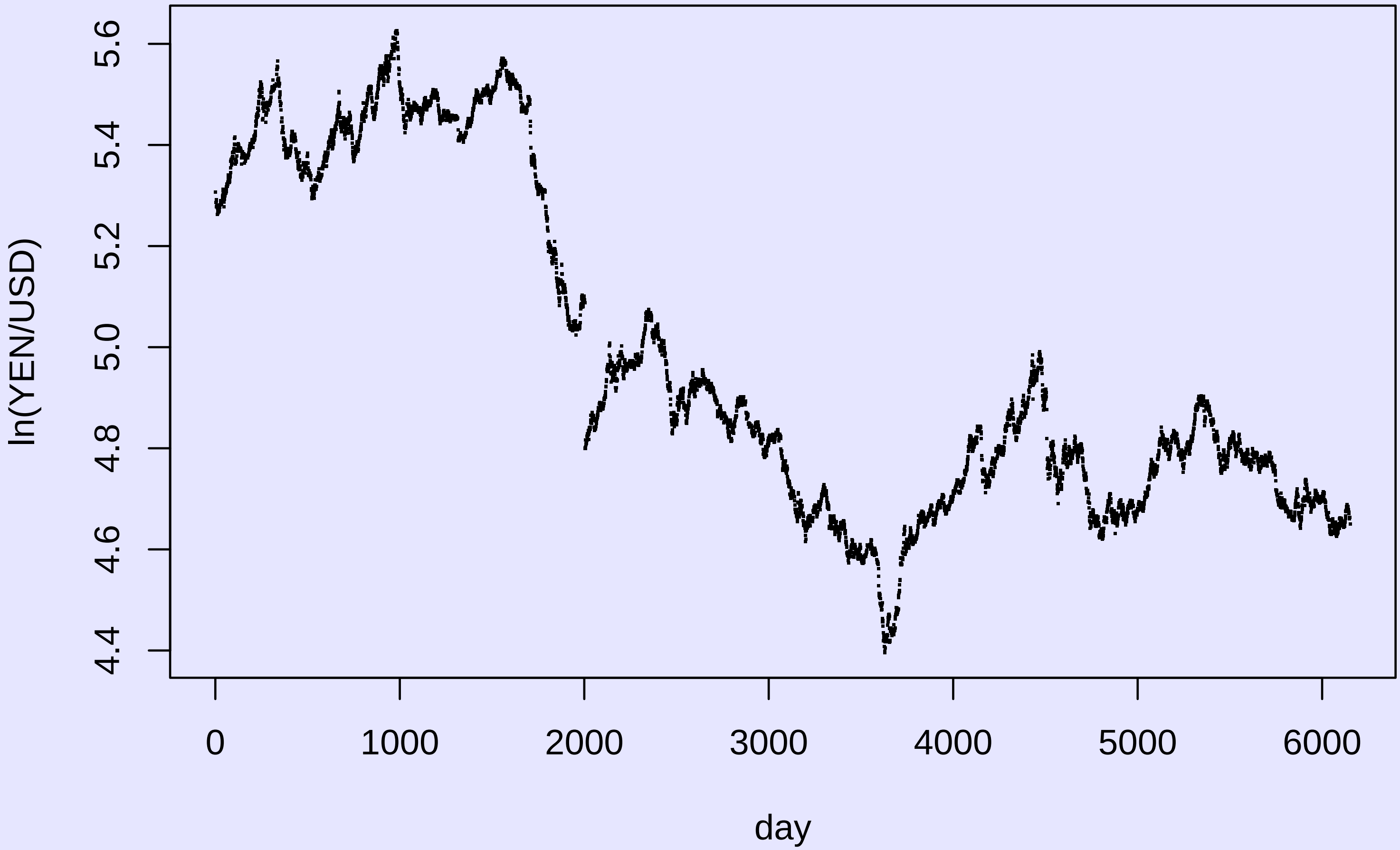
Outline

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

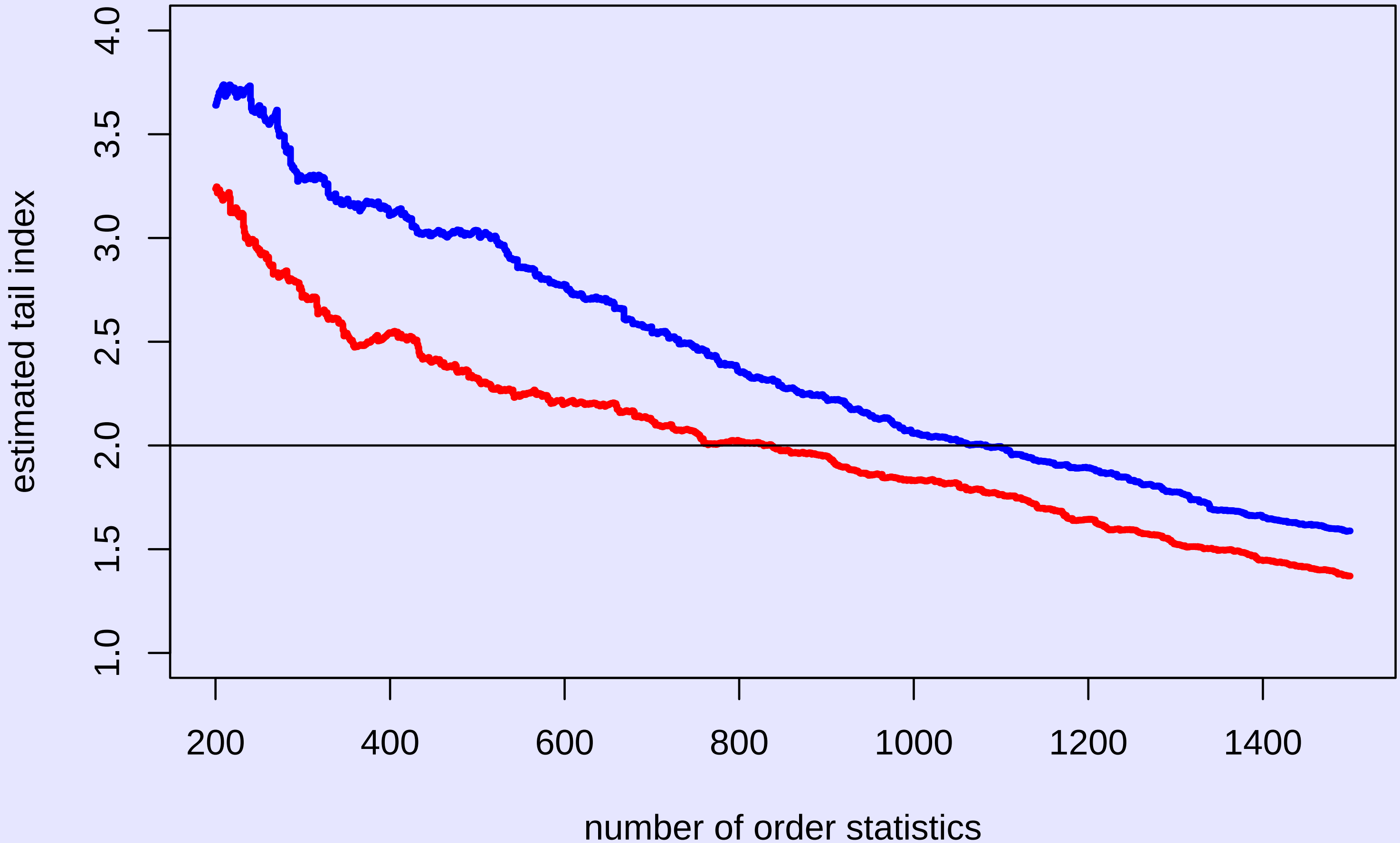
1. INTRODUCTION

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



Hill estimates for returns: lower tail = red, upper tail = blue.

- 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), Martins (2009), Samarakoon & Knight (2009).
  - Cointegration testing: Caner (1998), Paulauskas & Rachev (1998), Xiao (2009).
  - Applications: Koedijk and Kool (1992), Falk and Wang (2003), Charemza, Hristova & Burridge (2005), Kirman and Teyssière (2005).

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 $\{\boldsymbol{X}_t\}$ is a vector process whose elements are all integrated then it is cointegrated if $\{\boldsymbol{a}^\top \boldsymbol{X}_t\}$ is stationary for some $\boldsymbol{a} \neq 0$ (called a cointegration vector).
- Economic interpretation: individual variables behave like random walks but are collectively in equilibrium.

Example: Long run PPP

$$\text{Is } \ln(\text{exchange rate}_t) - \ln(\text{CPI}_t^{(1)}/\text{CPI}_t^{(2)}) \text{ stationary?}$$

Testing for cointegration: Two basic approaches:

(1) 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.

(2) Assume a parametric model (for example, VAR) for $\{\boldsymbol{X}_t\}$ and test for cointegration within that model.

- We will consider the second approach using a VAR model.
- Assume the VAR(k) model written 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.

- 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.

  – If $\Pi = 0$ then there is no cointegration.

  – If Π has full rank then $\{\boldsymbol{X}_t\}$ is stationary.

- Cointegration rank is determined essentially by finding good lower rank approximations to an unconstrained estimator of Π.

  – Start by testing $H_0 : \Pi = 0$ (no cointegration).

- Johansen (1988, 1991, ...) and others have developed extensive asymptotic distribution theory in the finite variance case.

Question: How to extend to the infinite variance case?

- Finite variance procedures work quite well but the asymptotic theory is complicated (Caner, 1998; Paulauskas & Rachev, 1998).
- For simplicity, 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

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

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

  where

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

- 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 $\{\Lambda_n\}$.

What are operator stable laws?

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

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

  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 $\mathrm{Re}(\lambda_j) \geq 1/2$ — these play the role of $1/\alpha$ where α is the stable index:
  - 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, which 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 $\{\boldsymbol{U}_t\}$ 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

$$\Delta_n^{-1} \sum_{t=1}^{n} \boldsymbol{U}_t - \boldsymbol{b}_n \xrightarrow{d} P_E.$$

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

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

  - For any set D bounded away from $\mathbf{0}$, we have

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

    for some measure ϕ.

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.

- Now look at operator stable limiting distribution.
- 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.
- For this example,

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

  and E has eigenvalues $1/2$ and 1.

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 (2009) for general unit root tests.
- The asymptotics depend on whether we do unconstrained minimization or project onto $\mathcal{C}^{\perp}$.
  - unconstrained — point process (i.e. non-standard) asymptotics.
  - projections — more classical asymptotics involving an operator stable process and a Brownian motion.

What are the regularity conditions?

- $\{B^\top \boldsymbol{\varepsilon}_t\}$ are in the domain of attraction of an operator stable law with index E whose eigenvalues are greater than $1/2$ 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 $\boldsymbol{\varepsilon}_t = (\varepsilon_{t1}, \cdots, \varepsilon_{tp})^\top$.

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

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

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.

Results: Focus on estimation of Π with rows projected to $\mathcal{C}^\perp$; these should be close to 0.

- If $\widehat{\Pi}_n$ is the "stacked" estimator then

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

  where

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

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

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

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

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

  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 $\boldsymbol{T}_n$:
  - maximum eigenvalue — can analytically evaluate the limiting distribution, albeit painfully (Muirhead, 1982) or via simulation;
  - trace — χ^2 limiting distribution.

- Rate of convergence is faster than for LS:
  - For LS estimation, we have $O_p(n^{-1})$ convergence rate;
  - for M-estimation, we have essentially $O_p(n^{-1/2}n^{-\tau}L(n))$ convergence where τ is the smallest of the real components of the eigenvalue of E and $L(n)$ is a slowly varying function.
- Limiting distribution is independent of E and $\{\Delta_n\}$ — thus 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 $\boldsymbol{T}_n \xrightarrow{d} \mathcal{W}_r(p, I)$ 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.