Lecture Notes on Unit Roots/Cointegration

John C. Chao

Econ 721 Lecture Notes

September 26, 2022

Unit Root Processes - Introduction

- Another approach to modeling trending behavior in economic time series is to use what are called integrated processes, which must be differenced in order to induce stationarity
- **Definition:** A series $\{X_t\}$ with no deterministic component which has a stationary invertible *ARMA* representation after differencing d times is said to be an integrated process of order d, denoted by

$$X_t \sim I(d).$$

  See Engle and Granger (1987).
- We will focus our discussion here primarily on $I(1)$ processes (and also on $I(0)$ process) because they seem to be the most relevant for applications in economics.

- Consider an $AR(1)$ process

$$Y_t = \rho Y_{t-1} + u_t,\ t = 1, ..., n$$

Under the assumptions that $|\rho| < 1$ and $\{u_t\} \equiv i.i.d.\left(0, \sigma^2\right)$, the asymptotic distribution of the OLS estimator $\widehat{\rho}_n$ of the autoregressive parameter ρ is given by

$$\sqrt{n}\left(\widehat{\rho}_n - \rho\right) \xrightarrow{d} N\left(0, 1 - \rho^2\right) \text{ as } n \to \infty$$

- This seems to suggest that for $\rho = 1$

$$\sqrt{n}\left(\widehat{\rho}_n - 1\right) \xrightarrow{p} 0 \text{ as } n \to \infty$$

so that the rate of convergence of the OLS estimator $\widehat{\rho}_n$ is faster than $\sqrt{n}$ when the true parameter $\rho = 1$. It turns out that this intuition is correct.

Functional Central Limit Theorem (FCLT)

- To analyze the limiting behavior of $\widehat{\rho}_n$ in the unit root case, we need to employ functional central limit theorems which are limit results on partial sums considered as random elements on certain functional spaces. The two functional spaces of interest are
  1. $C[0,1]$ - space of real-valued continuous functions on the $[0,1]$ interval.
  2. $D[0,1]$ - space of real-valued functions on $[0,1]$ which are right continuous and have left limits (or CADLAG functions from the French acronym *continue à droite, limites à gauche*).
- Clearly, $C[0,1] \subset D[0,1]$.
- We want to give these spaces a certain structure that makes them as close as possible to $(\mathbb{R}, d_e)$. This is achieved by endowing them with metrics that makes them **complete** and **separable**.

Functional Central Limit Theorem (FCLT)

- **Cauchy sequence:** A sequence $\{x_n\}$ of points in a metric space $(\mathbb{M}, d)$ is a Cauchy sequence if for all $\varepsilon > 0$, $\exists$ N_ε such that

$$d(x_n, x_m) < \varepsilon \text{ whenever } n, m \geq N_\varepsilon.$$

- A metric space $(\mathbb{M}, d)$ is **complete** if it contains all of its limit points (i.e., the limits of all Cauchy sequences).
- A subset A of a metric space $\mathbb{M}$ is said to be dense in $\mathbb{M}$ if each point in $\mathbb{M}$ can be "well-approximated" by points in A. Formally, A is dense in $\mathbb{M}$ if for each element $m \in \mathbb{M}$ and each $\varepsilon > 0$, $\exists a \in A$ such that $d(m, a) < \varepsilon$.
- A metric space $(\mathbb{M}, d)$ is separable if it contains a countable dense subset (i.e., it is well approximated by some countable subset). Hence, a space is not separable if it contains a noncountable discrete (points separated) subset. Separability is important because if it does not hold then not all the Borel sets of the space are measurable.

Functional Central Limit Theorem (FCLT)

- It can be shown that $C[0,1]$ is complete and separable metric space when endowed with the uniform metric

$$d_u(f,g) = \sup_{t\in[0,1]} |f(t) - g(t)|, \; f,g \in C[0,1].$$

- However, $D[0,1]$ is not a separable metric space under the uniform metric d_u. This can be seen from the following example: consider the set of functions

$$f_\theta(t) = \begin{cases} 0 & t < \theta \\ 1 & t \geq \theta \end{cases} \quad \theta \in [0,1).$$

Note that the set of functions $\{f_\theta(t) : \theta \in [0,1)\}$ is uncountable, but

$$d_u(f_\theta, f_{\theta'}) = 1 \; \forall \theta \neq \theta',$$

so that the elements of this set are all a discrete distance apart. Hence, $(D[0,1], d_u)$ is not separable.

Functional Central Limit Theorem (FCLT)

- Of interest to us is the (standardized) partial sum process

$$X_n(r) = \frac{1}{\sqrt{n}} \sum_{j=1}^{[nr]} u_j = \frac{1}{\sqrt{n}} S_{[nr]}$$

where $[nr]$ denotes the integer part of nr (i.e., the largest integer $\leq nr$), $r \in [0, 1]$, and $\{u_j\} \equiv I(0)$. Note that, for all finite n, the realizations of $X_n(r)$ are not continuous but are elements of $D[0, 1]$.

- **Example:** $n = 3$

$$X_3(r) = \begin{cases} 0 & \text{for } 0 \leq r < 1/3 \\ u_1/\sqrt{3} & \text{for } 1/3 \leq r < 2/3 \\ (u_1 + u_2)/\sqrt{3} & \text{for } 2/3 \leq r < 1 \\ (u_1 + u_2 + u_3)/\sqrt{3} & \text{for } r = 1 \end{cases}$$

Functional Central Limit Theorem (FCLT)

- Although realizations of $X_n(r)$ are not elements of $D[0,1]$, $X_n(r)$ can be approximated by a random element of $C[0,1]$ via the interpolation

$$X_n^*(r) = \frac{1}{\sqrt{n}} S_{[nr]} + \frac{nr - [nr]}{\sqrt{n}} u_{[nr]+1} \in C[0,1]$$

Here, the jumps in $(1/\sqrt{n})\, S_{[nr]}$ are eliminated by line segments that connect the partial sums at each $r = k/n$ for $k = 0, 1, ..., n$. Note that for $(k-1)/n \leq r < k/n$,

$$0 \leq nr - [nr] < 1$$

so that

$$\frac{nr - [nr]}{\sqrt{n}} = O\left(\frac{1}{\sqrt{n}}\right) \text{ uniformly in } r \in \left[\frac{k-1}{n}, \frac{k}{n}\right).$$

It follows that the asymptotic behavior of $X_n^*(r)$ is the same as that of $X_n(r)$.

- We now state a functional central limit theorem for partial sums of i.i.d. sequences
  Theorem 1: Suppose that $\{u_j\} \equiv i.i.d.\left(0, \sigma^2\right)$, $0 < \sigma^2 < \infty$; then,

$$\frac{X_n(r)}{\sigma} = \frac{1}{\sqrt{n}\sigma} \sum_{j=1}^{[nr]} u_j \Longrightarrow W(r) \equiv BM(1).$$

- Here, $W(r)$ denotes the Wiener process, or standard Brownian motion, on $C[0,1]$. Recall that $W(r)$ is completely defined by its properties
  (i) $W(0) = 0$;
  (ii) $W(r) \equiv N(0, r)$;
  (iii) $W(s)$ is independent of $W(r) - W(s)$ for $0 \leq s < r \leq 1$;
  (iv) $W(r)$ has continuous sample path with probability one.

Donsker's Theorem

- **Remark:** Note that the Donsker's theorem is obtained under the same assumptions as the Lindeberg-Lévy central limit theorem. In fact, the former contains the latter as a special case since setting $r = 1$, we have

$$\frac{X_n(1)}{\sigma} = \frac{1}{\sqrt{n}\sigma} \sum_{j=1}^{n} u_j \Longrightarrow W(1) \equiv N(0,1).$$

Continuous Mapping Theorem

- To analyze estimators and test statistics associated with unit root models, we need results that not only give us the limiting behavior of partial sums but also that of continuous functional of partial sums.
- **Theorem 2:** Let $h(\cdot)$ be any continuous functional on $D[0,1]$. If

$$X_n(r) \Longrightarrow B(r) \text{ on } D[0,1],$$

where $B(r) \equiv \sigma W(r) \equiv BM(\sigma^2)$. Then,

$$h(X_n(r)) \Longrightarrow h(B(r)).$$

- **Example:** Suppose that $X_n(r) \Longrightarrow B(r)$; then,

$$h(X_n(r)) = \int_0^1 X_n(r)\, dr \Longrightarrow \int_0^1 B(r)\, dr = h(B(r))$$

since the integral here is a continuous functional.

Asymptotics for Integrated Processes with i.i.d. Innovations

- Consider the simple $I(1)$ process

$$Y_t = Y_{t-1} + u_t,\ t = 1, ..., n;$$

To illustrate some of the basic ideas behind the asymptotics of unit root processes, we will first show that

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} Y_t \Longrightarrow \int_0^1 B(r)\, dr \text{ as } n \to \infty.$$

- **Remark:** Note that, unlike a law of large numbers result for stationary, weakly dependent processes; here, we have to divide by $n^{3/2}$ instead of n. Even so, this "average"

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} Y_t$$

does not stabilize to some population mean as $n \to \infty$ but instead goes to some random limit.

Asymptotics for Integrated Processes with i.i.d. Innovations

- To proceed, write

$$
\begin{aligned}
Y_t &= \sum_{j=1}^{t} u_j + Y_0 = S_t + Y_0 \equiv I(1), \\
X_n(r) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{[nr]} u_j = \frac{1}{\sqrt{n}} S_{[nr]} \in D[0,1]
\end{aligned}
$$

- Note that

$$
X_n(r) = \begin{cases} 0 & \text{for } 0 \leq r < 1/n \\ u_1/\sqrt{n} & \text{for } 1/n \leq r < 2/n \\ (u_1 + u_2)/\sqrt{n} & \text{for } 2/n \leq r < 3/n \\ \vdots & \vdots \\ (u_1 + u_2 + \cdots + u_n)/\sqrt{n} & \text{for } r = 1 \end{cases}
$$

Asymptotics for Integrated Processes with i.i.d. Innovations

- Given that the realizations of $X_n(r)$ is a step function, it is apparent that

$$\int_{(t-1)/n}^{t/n} X_n(r)\,dr = \frac{1}{n}\sum_{j=1}^{t-1} u_j/\sqrt{n} = \frac{1}{n^{3/2}} S_{t-1}$$

- Now, define $S_0 = 0$, and note that

$$\begin{aligned}
\sum_{t=1}^{n} Y_t &= \sum_{t=1}^{n} \left[S_{t-1} + u_t + Y_0\right] \\
&= n^{3/2} \sum_{t=1}^{n} \left[\frac{1}{n^{3/2}} S_{t-1}\right] + \sum_{t=1}^{n} u_t + nY_0 \\
&= n^{3/2} \sum_{t=1}^{n} \left[\int_{(t-1)/n}^{t/n} X_n(r)\,dr\right] + \sum_{t=1}^{n} u_t + nY_0 \\
&= n^{3/2} \int_0^1 X_n(r)\,dr + \sum_{t=1}^{n} u_t + nY_0
\end{aligned}$$

Asymptotics for Integrated Processes with i.i.d. Innovations

- Hence, by the Donsker's Theorem and the continuous mapping theorem, we have conditional on Y_0,

$$\begin{aligned} \frac{1}{n^{3/2}} \sum_{t=1}^{n} Y_t &= \int_0^1 X_n(r)\,dr + \frac{1}{n^{3/2}} \sum_{t=1}^{n} u_t + \frac{1}{\sqrt{n}} Y_0 \\ &= \int_0^1 X_n(r)\,dr + o_p(1) \\ &\Longrightarrow \int_0^1 B(r)\,dr \quad \text{as } n \to \infty. \end{aligned}$$

  where $B(r) \equiv \sigma W(r) \equiv BM(\sigma^2)$
- Similarly, one can show that

$$\frac{1}{n^{3/2}} \sum_{t=1}^{[nr]} Y_t \Longrightarrow \int_0^r B(s)\,ds \quad \text{as } n \to \infty.$$

Asymptotic Distribution of the OLS Estimator

- Consider estimating by OLS the coefficient ρ of the $AR(1)$ model

$$Y_t = \rho Y_{t-1} + u_t,$$

where the true value $\rho_0 = 1$ and where $\{u_t\} \equiv i.i.d.\left(0, \sigma^2\right)$, with $0 < \sigma^2 < \infty$.

- The OLS estimator in this case is given by

$$\widehat{\rho}_n = \frac{\sum_{t=2}^n Y_{t-1} Y_t}{\sum_{t=2}^n Y_{t-1}^2}$$

- By the usual regression algebra, we can write the deviation of $\widehat{\rho}_n$ from the true value as

$$\widehat{\rho}_n - 1 = \frac{\sum_{t=2}^n Y_{t-1} u_t}{\sum_{t=2}^n Y_{t-1}^2}$$

Asymptotic Distribution of the OLS Estimator

- It turns out that the rate of convergence of $\widehat{\rho}_n$ in this case is n, so that, upon appropriate standardization, we obtain

$$n\left(\widehat{\rho}_n - 1\right) = \frac{n^{-1}\sum_{t=2}^{n} Y_{t-1}u_t}{n^{-2}\sum_{t=2}^{n} Y_{t-1}^2}$$

- We first examine the limiting behavior of the denominator on the right-hand side of the expression above. Note that

$$X_n^2(r) = \begin{cases} 0 & \text{for } 0 \leq r < 1/n \\ u_1^2/n & \text{for } 1/n \leq r < 2/n \\ \left(u_1 + u_2\right)^2/n & \text{for } 2/n \leq r < 3/n \\ \vdots & \vdots \\ \left(u_1 + u_2 + \cdots + u_n\right)^2/n & \text{for } r = 1 \end{cases}$$

Asymptotic Distribution of the OLS Estimator

- Given that the realizations of $X_n^2(r)$ is also a step function, we can write
$$\int_{(t-1)/n}^{t/n} X_n^2(r)\,dr = \frac{1}{n}\left(\sum_{j=1}^{t-1} u_j/\sqrt{n}\right)^2 = \frac{1}{n^2}S_{t-1}^2$$
- Next, write

$$\begin{aligned}
\sum_{t=1}^{n} Y_{t-1}^2 &= \sum_{t=1}^{n}\left[S_{t-1}+Y_0\right]^2 \\
&= \sum_{t=1}^{n}\left[S_{t-1}^2 + 2S_{t-1}Y_0 + Y_0^2\right] \\
&= n^2\sum_{t=1}^{n}\left[\frac{1}{n^2}S_{t-1}^2\right] + 2Y_0\sum_{t=1}^{n}S_{t-1} + nY_0^2 \\
&= n^2\sum_{t=1}^{n}\left[\int_{(t-1)/n}^{t/n} X_n^2(r)\,dr\right] + 2Y_0\sum_{t=1}^{n}S_{t-1} + nY_0^2
\end{aligned}$$

- so that

$$\sum_{t=1}^{n} Y_{t-1}^2 = n^2 \sum_{t=1}^{n} \left[\int_{(t-1)/n}^{t/n} X_n^2(r)\, dr \right] + 2Y_0 \sum_{t=1}^{n} S_{t-1} + nY_0^2$$
$$= n^2 \int_0^1 X_n^2(r)\, dr + 2Y_0 \sum_{t=1}^{n} S_{t-1} + nY_0^2$$

- Dividing by n^2 and conditioning on Y_0, we obtain

$$\frac{1}{n^2} \sum_{t=1}^{n} Y_{t-1}^2 = \int_0^1 X_n^2(r)\, dr + 2Y_0 \frac{1}{n^2} \sum_{t=1}^{n} S_{t-1} + \frac{Y_0^2}{n}$$
$$= \int_0^1 X_n^2(r)\, dr + o_p(1)$$
$$\Longrightarrow \int_0^1 [B(r)]^2\, dr = \sigma^2 \int_0^1 [W(r)]^2\, dr \quad \text{as } n \to \infty.$$

Asymptotic Distribution of the OLS Estimator

- Next, to get a handle on the numerator on the right-hand side of the expression

$$n\left(\widehat{\rho}_n - 1\right) = \frac{n^{-1}\sum_{t=2}^{n} Y_{t-1}u_t}{n^{-2}\sum_{t=2}^{n} Y_{t-1}^2}$$

we first write

$$Y_{t-1} = \sum_{j=1}^{t-1} u_j + Y_0 = S_{t-1} + Y_0$$

- Note that, conditional on Y_0,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}u_t &= \frac{1}{n}\sum_{t=1}^{n} S_{t-1}u_t + Y_0\frac{1}{n}\sum_{t=1}^{n} u_t \\ &= \frac{1}{n}\sum_{t=1}^{n} S_{t-1}u_t + o_p(1)\end{aligned}$$

Asymptotic Distribution of the OLS Estimator

- Moreover, note that

$$
\begin{aligned}
S_n^2 &= \left(\sum_{j=1}^{n} u_j \right)^2 \\
&= \sum_{j=1}^{n} u_j^2 + 2 \sum_{i=2}^{n} \sum_{j=1}^{i-1} u_j u_i \\
&= \sum_{j=1}^{n} u_j^2 + 2 \sum_{i=2}^{n} S_{i-1} u_i
\end{aligned}
$$

- This, in turn, implies that

$$
\frac{1}{n} \sum_{i=2}^{n} S_{i-1} u_i = \frac{1}{2} \left(\frac{1}{n} S_n^2 - \frac{1}{n} \sum_{j=1}^{n} u_j^2 \right)
$$

- It follows that

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}u_t &= \frac{1}{n}\sum_{t=1}^{n} S_{t-1}u_t + o_p(1) \\
&= \frac{1}{2}\left(\frac{1}{n}S_n^2 - \frac{1}{n}\sum_{t=1}^{n} u_t^2\right) + o_p(1) \\
&= \frac{1}{2}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t\right)^2 - \frac{1}{2}\frac{1}{n}\sum_{t=1}^{n} u_t^2 + o_p(1) \\
&= \frac{1}{2}\left(X_n(1)\right)^2 - \frac{1}{2}\frac{1}{n}\sum_{t=1}^{n} u_t^2 + o_p(1) \\
\Longrightarrow & \ \frac{1}{2}\left(B(1)^2 - \sigma^2\right) = \frac{1}{2}\sigma^2\left(\chi^2(1) - 1\right)
\end{aligned}$$

where $B(1) = \sigma W(1) \equiv N(0, \sigma^2)$ and $\chi^2(1)$ denotes a Chi-square random variable with one degree of freedom.

- Putting the pieces together and invoking the continuous mapping theorem, we have that

$$\begin{aligned} n\left(\widehat{\rho}_n - 1\right) &= \frac{n^{-1}\sum_{t=2}^{n} Y_{t-1}u_t}{n^{-2}\sum_{t=2}^{n} Y_{t-1}^2} \\ &\Longrightarrow \frac{(1/2)\,\sigma^2\left(\chi^2(1) - 1\right)}{\sigma^2\int_0^1 [W(r)]^2\, dr} \equiv \frac{(1/2)\left(\chi^2(1) - 1\right)}{\int_0^1 [W(r)]^2\, dr}. \end{aligned}$$

Asymptotic Distribution of the OLS Estimator

- Note that the asymptotic distribution of $n\left(\widehat{\rho}_n - 1\right)$ is nonstandard but nuisance parameter free, so that, at least in the case with i.i.d. innovations, $n\left(\widehat{\rho}_n - 1\right)$ itself can be used as a statistic for testing the null hypothesis

$$H_0 : \rho_0 = 1$$

- Observe also that the probability that a $\chi^2\left(1\right)$ random variable is less than one is 0.68. since

$$\int_0^1 \left[W\left(r\right)\right]^2 dr > 0 \ a.s.,$$

the probability of $n\left(\widehat{\rho}_n - 1\right)$ being negative approaches 0.68 as n approaches infinity. Hence, in contrast to the stationary or stable case, the limiting distribution of $n\left(\widehat{\rho}_n - 1\right)$ in this case is skewed to the left.

Extension to Integrated Processes with Serially Correlated Innovations

- We consider now an extension of the Donsker Theorem to cases with serially correlated innovations based on the approach of Phillips and Solo (1992). In particular, we want to establish a FCLT for partial sums of a general linear process
- Some Notations:
  1. Let L be a lag operator, so that $L\varepsilon_t = \varepsilon_{t-1}$ and, more generally, $L^j\varepsilon_t = \varepsilon_{t-j}$.
  2. Define
$$C(L) = \sum_{j=0}^{\infty} c_j L^j.$$

Extension to Integrated Processes with Serially Correlated Innovations

- Consider the linear process

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$$

- **Assumption LP:**
  (i) $\{\varepsilon_t\} \equiv i.i.d.\left(0, \sigma^2\right)$, $0 < \sigma^2 < \infty$
  (ii) $\sum_{j=0}^{\infty} \sqrt{j}\left|c_j\right| < \infty$

- **Remark:** Note that the condition $\sum_{j=0}^{\infty} \sqrt{j}\left|c_j\right| < \infty$ is stronger than absolute summability, i.e., $\sum_{j=0}^{\infty} \left|c_j\right| < \infty$, so that Assumption LP(ii) above requires faster decay in the coefficient c_j as $j \to \infty$.

Extension to Integrated Processes with Serially Correlated Innovations

- **Theorem** (Phillips and Solo, *Annals of Statistics*, 1992)**:** Under Assumption LP,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t \Longrightarrow B(r) \equiv BM\left(\omega^2\right),$$

where

$$\omega^2 = \sigma^2 C(1)^2 = \sigma^2 \left(\sum_{j=0}^{\infty} c_j \right)^2$$

is the long-run variance.

Extension to Integrated Processes with Serially Correlated Innovations

- **Rough Outline of Proof:** By the Beveridge-Nelson (BN) decomposition, we have that

$$C(L) = C(1) - \widetilde{C}(L)(1 - L)$$

where

$$\widetilde{C}(L) = \sum_{j=0}^{\infty} \widetilde{c}_j L^j \text{ and } \widetilde{c}_j = \sum_{s=j+1}^{\infty} c_s$$

We will give a more explicit derivation of the BN decomposition later, but note first that, by making use of this decomposition, we can write

$$\begin{aligned} u_t &= C(L)\varepsilon_t \\ &= C(1)\varepsilon_t - \widetilde{C}(L)(\varepsilon_t - \varepsilon_{t-1}) \\ &= C(1)\varepsilon_t - (\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_{t-1}) \end{aligned}$$

where $\widetilde{\varepsilon}_t = \widetilde{C}(L)\varepsilon_t$.

Extension to Integrated Processes with Serially Correlated Innovations

- **Rough Outline of Proof (con't):** It follows that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} u_t &= C(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t - \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\left(\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_{t-1}\right) \\
&= C(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t - \frac{1}{\sqrt{n}}\left(\widetilde{\varepsilon}_{[nr]} - \widetilde{\varepsilon}_0\right) \\
&\quad \left(\text{since } \sum_{t=1}^{[nr]}\left(\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_{t-1}\right) \text{ is a telescoping sum}\right)
\end{aligned}$$

Next, note that, by the Donsker Theorem (i.e., FCLT for *i.i.d.* sequence),

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t \Longrightarrow \sigma W(r) \equiv BM\left(\sigma^2\right).$$

Extension to Integrated Processes with Serially Correlated Innovations

- **Rough Outline of Proof (con't):** Application of the continuous mapping theorem yields

$$C(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t \Longrightarrow \sigma C(1)W(r) \equiv BM(\omega^2).$$

Moreover, we can show that

$$\sup_{r\in[0,1]}\left|\frac{\widetilde{\varepsilon}_{[nr]}-\widetilde{\varepsilon}_0}{\sqrt{n}}\right| \xrightarrow{p} 0 \text{ as } n\to\infty$$

from which it follows that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}u_t = C(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t - \frac{1}{\sqrt{n}}\left(\widetilde{\varepsilon}_{[nr]}-\widetilde{\varepsilon}_0\right) \Longrightarrow BM(\omega^2).$$

Extension to Integrated Processes with Serially Correlated Innovations

- **Beveridge-Nelson Decomposition:** An explicit derivation of the BN decomposition can be given as follows. Let $\widetilde{c}_j = \sum_{s=j+1}^{\infty} c_s$ as previously defined and write

$$
\begin{aligned}
& C(L) \\
= \; & \sum_{j=0}^{\infty} c_j L^j \\
= \; & \sum_{j=0}^{\infty} c_j - \sum_{j=1}^{\infty} c_j + \left(\sum_{j=1}^{\infty} c_j - \sum_{j=2}^{\infty} c_j \right) L + \left(\sum_{j=2}^{\infty} c_j - \sum_{j=3}^{\infty} c_j \right) L^2 + \cdots \\
= \; & \sum_{j=0}^{\infty} c_j - \sum_{j=1}^{\infty} c_j (1 - L) - \sum_{j=2}^{\infty} c_j L (1 - L) - \sum_{j=3}^{\infty} c_j L^2 (1 - L) - \cdots \\
= \; & \sum_{j=0}^{\infty} c_j - \sum_{j=0}^{\infty} \widetilde{c}_j L^j (1 - L) = C(1) - \widetilde{C}(L)(1 - L)
\end{aligned}
$$

Extension to Integrated Processes with Serially Correlated Innovations

- **Lemma:** Let $\widetilde{C}(L) = \sum_{j=0}^{\infty} \widetilde{c}_j L^j$ and $\widetilde{c}_j = \sum_{s=j+1}^{\infty} c_s$. Then,

  (a)
  $$\sum_{j=0}^{\infty} \widetilde{c}_j^2 < \infty \text{ if } \sum_{j=0}^{\infty} \sqrt{j} \left| c_j \right| < \infty$$

  (b)
  $$\sum_{j=0}^{\infty} \left| \widetilde{c}_j \right| < \infty \text{ if } \sum_{j=0}^{\infty} j \left| c_j \right| < \infty.$$

Phillips-Perron Unit Root Test

- Consider the time series model

$$Y_t = \alpha + \rho Y_{t-1} + u_t,$$

where $\{u_t\}$ follows a linear process, i.e.,

$$u_t = C(L)\,\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$$

- **Assumptions:**
  (i) $\{\varepsilon_t\} \equiv i.i.d.\,(0, \sigma^2)$, where $0 < \sigma^2 < \infty$
  (ii) $\sum_{j=0}^{\infty} \sqrt{j}\,|c_j| < \infty$

Phillips-Perron Unit Root Test

- Suppose we wish to test the null hypothesis

$$H_0 : \alpha = 0, \; \rho = 1$$

versus the alternative hypothesis

$$H_1 : \alpha > 0, \; |\rho| < 1$$

- **Remark:** Note that, for the model studied here, i.e.,

$$Y_t = \alpha + \rho Y_{t-1} + u_t,$$

the regressor Y_{t-1} is correlated with the error u_t since u_t is serially correlated. Ordinarily, in models involving $I(0)$ variables, the OLS estimator of the regresssion coefficients will be inconsistent in this case, and we will be looking to estimate this model by some instrumental variable (IV) methods. However, we will see that, in the $I(1)$ case, OLS will still be consistent, and this fact was exploited by the Phillips-Perron approach to this problem.

Phillips-Perron Unit Root Test

- The Phillips-Perron approach in this setting is to proceed by first analyzing the limiting distribution of the OLS estimator of ρ under H_0. As we will see, the OLS estimator of ρ will still be consistent under the unit root null hypothesis but will have an asymptotic distribution that is not nuisance parameter free. Hence, the Phillips-Perron appoarch involves modifying the usual test statistics so that the resulting test procedure will be asymptotically similar, or nuisance parameter free.
- To consolidate notations a bit, write

$$\begin{aligned} Y_t &= \alpha + \rho Y_{t-1} + u_t \\ &= X_t' \beta + u_t, \end{aligned}$$

where $X_t = (1, Y_{t-1})'$ and $\beta = (\alpha, \rho)'$. Also, let $\widehat{\beta}_n = (\widehat{\alpha}_n, \widehat{\rho}_n)'$, where $\widehat{\alpha}_n$ and $\widehat{\rho}_n$ are the OLS estimators of α and ρ, respectively, and let $\beta_0 = (\alpha_0, \rho_0)' = (0, 1)'$ be the value of α and ρ under H_0.

Phillips-Perron Unit Root Test

- Using these notations, we have, by the usual regression algebra, that under H_0

$$\begin{aligned}
\widehat{\beta}_n - \beta_0 &= \begin{pmatrix} \widehat{\alpha}_n \\ \widehat{\rho}_n - 1 \end{pmatrix} \\
&= \begin{pmatrix} n & \sum_{t=1}^{n} Y_{t-1} \\ \sum_{t=1}^{n} Y_{t-1} & \sum_{t=1}^{n} Y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{n} u_t \\ \sum_{t=1}^{n} Y_{t-1} u_t \end{pmatrix} \\
&= \left(\sum_{t=1}^{n} X_t X_t' \right)^{-1} \sum_{t=1}^{n} X_t u_t
\end{aligned}$$

- It turns out that the proper standardization in this case is to premultiply $\widehat{\beta}_n - \beta_0$ by the diagonal matrix

$$D_n = \begin{pmatrix} \sqrt{n} & 0 \\ 0 & n \end{pmatrix}$$

Phillips-Perron Unit Root Test

- Hence,

$$
\begin{aligned}
& \begin{pmatrix} \sqrt{n}\widehat{\alpha}_n \\ n\left(\widehat{\rho}_n - 1\right) \end{pmatrix} \\
= \quad & D_n \left(\widehat{\beta}_n - \beta_0 \right) \\
= \quad & \left(D_n^{-1} \sum_{t=1}^{n} X_t X_t' D_n^{-1} \right)^{-1} D_n^{-1} \sum_{t=1}^{n} X_t u_t \\
= \quad & \begin{pmatrix} 1 & n^{-3/2} \sum_{t=1}^{n} Y_{t-1} \\ n^{-3/2} \sum_{t=1}^{n} Y_{t-1} & n^{-2} \sum_{t=1}^{n} Y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} n^{-1/2} \sum_{t=1}^{n} u_t \\ n^{-1} \sum_{t=1}^{n} Y_{t-1} u_t \end{pmatrix}
\end{aligned}
$$

Phillips-Perron Unit Root Test

- The results of Phillips (1987) and Phillips and Perron (1988) show that the following convergence results hold jointly

$$\begin{aligned}
\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{t-1} &\Longrightarrow \omega\int_0^1 W(r)\,dr, \\
\frac{1}{n^{2}}\sum_{t=1}^{n} Y_{t-1}^2 &\Longrightarrow \omega^2\int_0^1 [W(r)]^2\,dr \\
\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t &\Longrightarrow \omega W(1) \equiv N\left(0,\omega^2\right), \\
\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}u_t &\Longrightarrow \frac{1}{2}\left[\omega^2\left(W(1)\right)^2-\gamma_0\right] \equiv \frac{1}{2}\left[\omega^2\chi^2(1)-\gamma_0\right],
\end{aligned}$$

where $\omega^2 = \sigma^2 C(1)^2 = \sigma^2\left(\sum_{j=0}^{\infty} c_j\right)^2$ (long-run variance) and $\gamma_0 = E\left[u_t^2\right] = \sigma^2\sum_{j=0}^{\infty} c_j^2$.

Phillips-Perron Unit Root Test

- By the continuous mapping theorem

$$\begin{pmatrix} \sqrt{n}\widehat{\alpha}_n \\ n\left(\widehat{\rho}_n - 1\right) \end{pmatrix} \implies \begin{pmatrix} 1 & \omega \int_0^1 W(r)\, dr \\ \omega \int_0^1 W(r)\, dr & \omega^2 \int_0^1 [W(r)]^2\, dr \end{pmatrix}^{-1} \times \begin{pmatrix} \omega W(1) \\ \frac{1}{2}\left[\omega^2 \chi^2(1) - \gamma_0\right] \end{pmatrix}$$

as $n \longrightarrow \infty$

Phillips-Perron Unit Root Test

- It follows by elementary calculations that, as $n \to \infty$,

$$
\begin{aligned}
n\left(\widehat{\rho}_n - 1\right) &\Longrightarrow \frac{\frac{1}{2}\left[\omega^2 \chi^2(1) - \gamma_0\right] - \omega^2 W(1) \int_0^1 W(r)\, dr}{\omega^2 \left\{ \int_0^1 [W(r)]^2\, dr - \left(\int_0^1 W(r)\, dr \right)^2 \right\}} \\
&= \frac{\frac{1}{2}\left[\chi^2(1) - \gamma_0/\omega^2\right] - W(1) \int_0^1 W(r)\, dr}{\int_0^1 [W(r)]^2\, dr - \left(\int_0^1 W(r)\, dr \right)^2}
\end{aligned}
$$

- or

$$n\left(\widehat{\rho}_n - 1\right) \implies \frac{\frac{1}{2}\left[\chi^2(1) - 1\right] - W(1)\int_0^1 W(r)\,dr}{\int_0^1 [W(r)]^2\,dr - \left(\int_0^1 W(r)\,dr\right)^2} + \frac{\frac{1}{2}\left(\omega^2 - \gamma_0\right)/\omega^2}{\int_0^1 [W(r)]^2\,dr - \left(\int_0^1 W(r)\,dr\right)^2}$$

- **Remark 1:** Note that in the special case where $C(L) = 1$, we have $u_t = \varepsilon_t$, so that $\{u_t\} \equiv i.i.d.\left(0, \sigma^2\right)$. Moreover, in this case,

$$\omega^2 = \sigma^2 C(1)^2 = \sigma^2 \text{ and } \gamma_0 = E\left[u_t^2\right] = \sigma^2.$$

Hence, from previous results, we know that, in this case,

$$n\left(\widehat{\rho}_n - 1\right) \Longrightarrow \frac{\frac{1}{2}\left[\chi^2(1) - 1\right] - W(1)\int_0^1 W(r)\,dr}{\int_0^1 [W(r)]^2\,dr - \left(\int_0^1 W(r)\,dr\right)^2}$$

- **Remark 1(con't):** It follows that the second term

$$\frac{\frac{1}{2}\left(\omega^2-\gamma_0\right)/\omega^2}{\int_0^1\left[W\left(r\right)\right]^2dr-\left(\int_0^1 W\left(r\right)dr\right)^2}$$

  is a second-order bias due to the serial correlation in the process $\{u_t\}$. As mentioned before, unlike in the $I\left(0\right)$ case, $\widehat{\rho}_n$ is consistent even in the presence of serial correlation in u_t, and the effect of this serial correlation shows up only as a second-order bias.

- **Remark 2:** Note also that, unlike the case with $i.i.d.$ innovations, $n\left(\widehat{\rho}_n-1\right)$ cannot be used directly as a statistic for testing the unit root null hypothesis since its asymptotic distribution now involves the nuisance parameters ω^2 and γ_0, whose true values are in general unknown.

- **Remark 3:** We can estimate $\gamma_0 = E\left[u_t^2\right]$ consistently using the estimator

$$s_n^2 = \frac{1}{n-2}\sum_{t=1}^{n}\widehat{u}_t^2$$

where $\widehat{u}_t = Y_t - \widehat{\alpha}_n - \widehat{\rho}_n Y_{t-1}$ is the OLS residual.

- **Estimation of Long-Run Variance:** To consistently estimate ω^2, note first that

$$\begin{aligned}\omega^2 &= \sigma^2 C(1)^2 \\ &= \sigma^2 \left(\sum_{j=0}^{\infty} c_j \right)^2 \\ &= E\left[u_t^2\right] + 2 \sum_{j=1}^{\infty} E\left[u_t u_{t-j}\right] \\ &= \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j\end{aligned}$$

- **Remark:** Note that ω^2 depends on an infinite number of unknown parameters, i.e., $\gamma_0, \gamma_1, \gamma_2, \ldots$. Realistically, with finite data, we cannot hope to estimate an infinite number of unknown parameters. However, we could pursue a strategy where we try to estimate a finite number of γ_j's and consider an estimation framework where we allow the dimension of the parameter space to increase as sample size increases.

- **Newey-West Estimator:** We can estimate ω^2 using an estimator proposed by Newey and West (1987), which in the special case we are dealing with here has the form

$$\widehat{\omega}^2 = \widehat{\gamma}_0 + 2\sum_{j=1}^{q(n)} \left[1 - \frac{j}{q(n)+1}\right] \widehat{\gamma}_j,$$

where

$$\widehat{\gamma}_j = \frac{1}{n}\sum_{t=j+1}^{n} \widehat{u}_t \widehat{u}_{t-j} \text{ for } j = 0, 1, ..., q(n).$$

- For the unit root testing problem, Phillips (1987), under some additional conditions, show that

$$\widehat{\omega}^2 \xrightarrow{p} \omega^2 \text{ if } q(n) \to \infty \text{ as } n \to \infty \text{ such that } \frac{q(n)}{n^{1/4}} \to 0.$$

Phillips-Perron Unit Root Test

- **Remark:** One might think that a natural estimator for ω^2 is

$$\widetilde{\omega}^2 = \widehat{\gamma}_0 + 2 \sum_{j=1}^{q(n)} \widehat{\gamma}_j,$$

However, it turns out that while $\widetilde{\omega}^2$ will be consistent under some conditions, it is not necessarily non-negative in finite sample. On the other hand, the Newey-West estimator $\widehat{\omega}^2$ is guaranteed to be non-negative as we will show below.

- **Non-negativity of $\widehat{\omega}^2$:** Also, define

$$\widehat{U}_{t,q} = \begin{pmatrix} \widehat{u}_t \\ \widehat{u}_{t-1} \\ \vdots \\ \widehat{u}_{t-q(n)+1} \\ \widehat{u}_{t-q(n)} \end{pmatrix},$$

where $\widehat{u}_j = 0$ for $j = n+1, ..., n+q(n)$ and $j = -q(n)+1, ..., -1, 0$.

- Also, define

$$\widehat{\Gamma} = \frac{1}{n} \sum_{t=1}^{n+q(n)} \widehat{U}_{t,q} \widehat{U}'_{t,q}$$

- **Non-negativity of $\widehat{\omega}^2$:** It is easily check that, under the above definition for $\widehat{\Gamma}$, we have

$$\widehat{\Gamma} = \begin{pmatrix} \widehat{\gamma}_0 & \widehat{\gamma}_1 & \cdots & \widehat{\gamma}_{q(n)-1} & \widehat{\gamma}_{q(n)} \\ \widehat{\gamma}_1 & \widehat{\gamma}_0 & \ddots & & \widehat{\gamma}_{q(n)-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \widehat{\gamma}_{q(n)-1} & & \ddots & \ddots & \widehat{\gamma}_1 \\ \widehat{\gamma}_{q(n)} & \widehat{\gamma}_{q(n)-1} & \cdots & \widehat{\gamma}_1 & \widehat{\gamma}_0 \end{pmatrix}.$$

where

$$\widehat{\gamma}_j = \frac{1}{n} \sum_{t=j+1}^{n} \widehat{u}_t \widehat{u}_{t-j} \text{ for } j = 0, 1, ..., q(n).$$

- **Non-negativity of $\widehat{\omega}^2$:** Moreover, note that $\widehat{\Gamma} = n^{-1} \sum_{t=1}^{n+q(n)} \widehat{U}_{t,q} \widehat{U}_{t,q}'$ is positive semidefinite. Hence, let ι be an $(q(n)+1) \times 1$ vector of ones, i.e., $\iota = (1, 1, ..., 1)'$, and we have that

$$
\begin{aligned}
0 &\leq \frac{\iota' \widehat{\Gamma} \iota}{q(n)+1} \\
&= \frac{1}{q(n)+1} \left\{ (q(n)+1) \widehat{\gamma}_0 + 2q(n) \widehat{\gamma}_1 + \cdots + 2\widehat{\gamma}_{q(n)} \right\} \\
&= \frac{1}{q(n)+1} \left\{ (q(n)+1) \widehat{\gamma}_0 + 2 \sum_{j=1}^{q(n)} [(q(n)+1) - j] \widehat{\gamma}_j \right\} \\
&= \widehat{\gamma}_0 + 2 \sum_{j=1}^{q(n)} \left[1 - \frac{j}{q(n)+1} \right] \widehat{\gamma}_j \\
&= \widehat{\omega}^2
\end{aligned}
$$

- **Modified Test Statistic for Testing the Unit Root Null Hypothesis:**

  First, define

$$
\begin{aligned}
s_n^2 &= \frac{1}{n-2}\sum_{t=1}^{n}\widehat{u}_t^2,\ \widehat{\sigma}_{\widehat{\rho}}^2 = \frac{s_n^2}{\sum_{t=2}^{n}\left(Y_{t-1}-\overline{Y}_{-1}\right)^2}, \\
\widehat{\omega}^2 &= \widehat{\gamma}_0 + 2\sum_{j=1}^{q(n)}\left[1-\frac{j}{q(n)+1}\right]\widehat{\gamma}_j, \text{ and } T_n = \frac{\widehat{\rho}_n - 1}{\widehat{\sigma}_{\widehat{\rho}}}
\end{aligned}
$$

where

$$
\widehat{u}_t = Y_t - \widehat{\alpha}_n - \widehat{\rho}_n Y_{t-1} \text{ and } \overline{Y}_{-1} = \frac{1}{n-1}\sum_{t=2}^{n} Y_{t-1}.
$$

Phillips-Perron Unit Root Test

- **Modified Test Statistic for Testing the Unit Root Null Hypothesis (con't):** Phillips and Perron (1988) proposed the following modified test statistics, which are designed to remove the effect of the second-order bias in the asymptotic distribution of the OLS estimator

$$\begin{aligned} Z_\rho &= n\left(\widehat{\rho}_n - 1\right) - \frac{1}{2}\left(\frac{n^2\widehat{\sigma}^2_{\widehat{\rho}}}{s_n^2}\right)\left(\widehat{\omega}^2 - s_n^2\right) \\ Z_t &= \left(\frac{s_n}{\widehat{\omega}}\right) T_n - \frac{1}{2}\left(\frac{n\widehat{\sigma}_{\widehat{\rho}}}{s_n}\right)\left(\frac{\widehat{\omega}^2 - s_n^2}{\widehat{\omega}}\right) \end{aligned}$$

- **Modified Test Statistic for Testing the Unit Root Null Hypothesis (con't):** In addition, Phillips and Perron (1988) showed that, under $H_0 : \rho_0 = 1$,

$$Z_\rho \Longrightarrow \frac{\frac{1}{2}\left[\chi^2(1) - 1\right] - W(1)\int_0^1 W(r)\,dr}{\int_0^1 [W(r)]^2\,dr - \left(\int_0^1 W(r)\,dr\right)^2}$$

$$Z_t \Longrightarrow \frac{\frac{1}{2}\left[\chi^2(1) - 1\right] - W(1)\int_0^1 W(r)\,dr}{\sqrt{\int_0^1 [W(r)]^2\,dr - \left(\int_0^1 W(r)\,dr\right)^2}}$$

- **Remark:** Note that, unlike the asymptotic distribution of $n\left(\widehat{\rho}_n - 1\right)$, these two modified test statistics have asymptotic null distributions that are free of nuisance parameters.

Augmented Dickey-Fuller Test for Unit Root

- Dickey and Fuller (1979) approaches the problem of testing for a unit root by by specifing an $AR\left(p\right)$ model

$$\left(1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p\right) Y_t=\phi\left(L\right) Y_t=\varepsilon_t$$

or

$$Y_t=\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+\cdots+\phi_p Y_{t-p}+\varepsilon_t$$

where $\{\varepsilon_t\} \equiv i.i.d.\left(0,\sigma^2\right)$.

- Define

$$\begin{aligned}
\rho &= \phi_1+\cdots+\phi_p, \\
\zeta_j &= -\left(\phi_{j+1}+\cdots+\phi_p\right) \text{ for } j=1,...,p-1
\end{aligned}$$

Augmented Dickey-Fuller Test for Unit Root

- It follows by essentially the Beveridge-Nelson decomposition for finite-order lag polynomial that we can write

$$
\begin{aligned}
& 1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p \\
= \; & 1-\left[\left(\phi_1+\phi_2+\cdots+\phi_p\right)-\left(\phi_2+\cdots+\phi_p\right)\right] L \\
& -\left[-\left(\phi_3+\cdots+\phi_p\right)+\left(\phi_2+\cdots+\phi_p\right)\right] L^2 \\
& \cdots-\left[-\phi_p+\left(\phi_{p-1}+\phi_p\right)\right] L^{p-1}-\phi_p L^p \\
= \; & 1-\left[\rho+\zeta_1\right] L-\left[\zeta_2-\zeta_1\right] L^2-\cdots-\left[\zeta_{p-1}-\zeta_{p-2}\right] L^{p-1} \\
& +\zeta_{p-1} L^p \\
= \; & 1-\rho L-\zeta_1 L(1-L)-\cdots-\zeta_{p-1} L^{p-1}(1-L) \\
= \; & (1-\rho L)-\left(\zeta_1 L+\cdots+\zeta_{p-1} L^{p-1}\right)(1-L)
\end{aligned}
$$

- Given this decomposition, we can rewrite the $AR(p)$ model as

$$\left[(1-\rho L)-\left(\zeta_1 L+\cdots+\zeta_{p-1}L^{p-1}\right)(1-L)\right]Y_t=\varepsilon_t$$

or

$$Y_t=\rho Y_{t-1}+\zeta_1\Delta Y_{t-1}+\cdots+\zeta_{p-1}\Delta Y_{t-p+1}+\varepsilon_t$$

- To see why this transformation of the $AR(p)$ model yields a useful representation for unit root testing, suppose that the p^{th} order polynomial equation

$$1-\phi_1 z-\phi_2 z^2-\cdots-\phi_p z^p=0$$

contains a single unit root, and all other roots are outside the unit circle. Then, by the fact that $z=1$ is a root, we have

$$1-\phi_1-\phi_2-\cdots-\phi_p=0$$

Augmented Dickey-Fuller Test for Unit Root

- or

$$1 = \phi_1 + \cdots + \phi_p = \rho$$

Hence, by rewriting the model in this way, we have transformed a null hypothesis about the root of a p^{th}-degree polynomial to a hypothesis which imposes a simple restriction on a single parameter ρ. The latter is obviously much easier to test.

- Moreover, under the null hypothesis that $\rho = 1$, we have

$$\begin{aligned} & 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \\ = \; & (1 - z) - \left(\zeta_1 z + \cdots \cdot + \zeta_{p-1} z^{p-1} \right) (1 - z) \\ = \; & \left(1 - \zeta_1 z - \cdots \cdot - \zeta_{p-1} z^{p-1} \right) (1 - z) \, , \end{aligned}$$

so that it must be true that all roots of the equation

$$1 - \zeta_1 z - \cdots \cdot - \zeta_{p-1} z^{p-1} = 0$$

are outside the unit circle.

Augmented Dickey-Fuller Test for Unit Root

- Furthermore, we can write

$$\left(1 - \zeta_1 L - \cdots - \zeta_{p-1} L^{p-1}\right)(1 - L) Y_t = \varepsilon_t$$

or

$$\left(1 - \zeta_1 L - \cdots - \zeta_{p-1} L^{p-1}\right) \Delta Y_t = \varepsilon_t$$

- Given that all roots of the polynomial equation

$$1 - \zeta_1 z - \cdots - \zeta_{p-1} z^{p-1} = 0$$

are outside the unit circle, we can invert the lag polynomial to obtain the moving-average representation

$$\begin{aligned} \Delta Y_t &= \left(1 - \zeta_1 L - \cdots - \zeta_{p-1} L^{p-1}\right)^{-1} \varepsilon_t \\ &= \psi(L) \varepsilon_t \\ &= u_t \end{aligned}$$

Augmented Dickey-Fuller Test for Unit Root

- Recall that the Phillips-Perron setup considered the equation

$$Y_t = \rho Y_{t-1} + u_t$$

where

$$u_t = C(L)\,\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$$

Hence, under $H_0 : \rho = 1$, we have

$$Y_t = Y_{t-1} + u_t$$

or

$$\Delta Y_t = u_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$$

Augmented Dickey-Fuller Test for Unit Root

- Comparing the Dickey-Fuller setup with the Phillips-Perron setup, we see that, under H_0,
  - (i) the Dickey-Fuller model has the moving-average representation

$$\begin{aligned} \Delta Y_t &= \psi(L)\,\varepsilon_t \\ &= \left(1 - \zeta_1 L - \cdots - \zeta_{p-1} L^{p-1}\right)^{-1} \varepsilon_t \end{aligned}$$

    which depends on a finite set of unknown parameters $\zeta_1, \ldots, \zeta_{p-1}$, whereas
  - (ii) the Phillips-Perron model has the moving-average representation

$$\Delta Y_t = u_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$$

    which can potentially depend on an infinite number of unknown parameters $c_0, c_1, c_2, \ldots$.
- It is in the sense that one may consider the Phillips-Perron framework for unit root testing to be more general.

Augmented Dickey-Fuller Test for Unit Root

- Now, consider again testing the null hypothesis

$$H_0 : \alpha = 0, \ \rho = 1$$

versus

$$H_1 : \alpha > 0, \ \ |\rho| < 1$$

- To implement the augmented Dickey-Fuller test in this case, one would first estimate the parameters of the regression

$$Y_t = \alpha + \rho Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \cdots + \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

by OLS.

- It turns out that the OLS estimator $\widehat{\rho}_n$, obtained from running the regression above, has the following large sample property under H_0

$$n\left(\widehat{\rho}_n - 1\right) \Longrightarrow \psi\left(1\right)\left\{\frac{\frac{1}{2}\left[\chi^2\left(1\right) - 1\right] - W\left(1\right)\int_0^1 W\left(r\right) dr}{\int_0^1 \left[W\left(r\right)\right]^2 dr - \left(\int_0^1 W\left(r\right) dr\right)^2}\right\}$$

where

$$\psi\left(1\right) = 1 - \zeta_1 - \cdots - \zeta_{p-1}$$

Augmented Dickey-Fuller Test for Unit Root

- It follows, by the Cramér Convergence Theorem, that under H_0

$$\frac{n\left(\widehat{\rho}_n - 1\right)}{1 - \widehat{\zeta}_1 - \cdots - \widehat{\zeta}_{p-1}} \Longrightarrow \left\{ \frac{\frac{1}{2}\left[\chi^2(1) - 1\right] - W(1)\int_0^1 W(r)\,dr}{\int_0^1 \left[W(r)\right]^2 dr - \left(\int_0^1 W(r)\,dr\right)^2} \right\}$$

where $\widehat{\zeta}_1, ..., \widehat{\zeta}_{p-1}$ denote OLS estimators of the parameters $\zeta_1, ..., \zeta_{p-1}$.

- Moreover, the t-statistic, under H_0, has the following asymptotic distribution

$$T_n \Longrightarrow \frac{\frac{1}{2}\left[\chi^2(1) - 1\right] - W(1)\int_0^1 W(r)\,dr}{\sqrt{\int_0^1 [W(r)]^2\,dr - \left(\int_0^1 W(r)\,dr\right)^2}}$$

- Hence, both the T_n statistic and the statistic

$$\frac{n\left(\widehat{\rho}_n - 1\right)}{1 - \widehat{\zeta}_1 - \cdots - \widehat{\zeta}_{p-1}}$$

have asymptotic null distributions that are free of nuisance parameter. Hence, both statistics can be used to implement a test of the unit root null hypothesis.

Spurious Regression

- Consider the $m \times 1$ vector $I(1)$ process

$$Y_t = Y_{t-1} + u_t, \;\; t = 1, ..., n,$$

where $\{u_t\}$ follows a general linear process, i.e.,

$$u_t = \Psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}.$$

- **Assumptions:**

(i) $\{\varepsilon_t\} \equiv i.i.d.\,(0, \Sigma_\varepsilon)$, where $\Sigma_\varepsilon > 0$ (i.e., Σ_ε is positive definite);

(ii) $\max_{1 \leq k \leq m} E\left(\varepsilon_{kt}^4\right) < \infty$;

(iii) $\sum_{j=1}^{\infty} j \left\| \Psi_j \right\| < \infty$;

(iv) $\Psi(1)$ is nonsingular (as we will see, this is an assumption of the absence of cointegration)

- **Multivariate Partial Sum Process:** A multivariate extension of the partial sum process can be defined as

$$X_n\left(r\right)=\frac{1}{\sqrt{n}}\sum_{j=1}^{[nr]}u_j\in D\left[0,1\right]^m$$

where $D\left[0,1\right]^m$ is a product space of m copies of $D\left[0,1\right]$.
- One can show that the vector linear process

$$u_t=\Psi\left(L\right)\varepsilon_t=\sum_{j=0}^{\infty}\Psi_j\varepsilon_{t-j}$$

under Assumptions (i)-(iv) satisfies a multivariate FCLT, so that, as $n\longrightarrow\infty$,

$$X_n\left(r\right)\Longrightarrow B\left(r\right)\equiv BM\left(\Omega\right)$$

where

$$\Omega=\Psi\left(1\right)\Sigma_\varepsilon\Psi\left(1\right)'$$

is the long-run covariance matrix.

Spurious Regression

- **Remark:** Observe that, under Assumptions (i) and (iv), $\Omega > 0$.
- To consider the phenomenon known as "spurious regression", partition the random vector Y_t as follows

$$Y_t = \begin{bmatrix} \underset{1\times 1}{y_{1t}} \\ \underset{g\times 1}{Y_{2t}} \end{bmatrix} \equiv I(1)$$

where $g = m - 1$. Also, partition $B(r)$ and Ω conformably with $Y_t = \begin{pmatrix} y_{1t} & Y_{2t}' \end{pmatrix}'$ as

$$B(r) = \begin{bmatrix} \underset{1\times 1}{B_1(r)} \\ \underset{g\times 1}{B_2(r)} \end{bmatrix}, \Omega = \begin{pmatrix} \underset{1\times 1}{\omega_{11}} & \underset{1\times g}{\omega_{12}} \\ \underset{g\times 1}{\omega_{21}} & \underset{g\times g}{\Omega_{22}} \end{pmatrix}$$

Spurious Regression

- Consider the least squares regression

$$y_{1t} = \widehat{\beta}_n' Y_{2t} + \widehat{v}_t$$

- Note that, if $\omega_{21} = 0$, then the variables y_{1t} and Y_{2t} have stochastic trends but are unrelated in the long run. Indeed, in the prototypical Granger-Newbold spurious regression setup examined in Granger and Newbold (1974), $\{u_t\} \equiv i.i.d.N\left(0, I_m\right)$ are unrelated at any frequency. Hence, in these cases, one might expect that $\widehat{\beta}_n$ will converge in probability to a zero vector. However, that turns out not to be the case. In fact, even if $\omega_{21} \neq 0$, the relationship between the series is not in general strong enough to permit $\widehat{\beta}_n$ to converge to any constant value.

- More precisely, it can be shown that

$$\begin{aligned}\widehat{\beta}_n &= \left(\frac{1}{n^2}\sum_{t=1}^{n} Y_{2t}Y_{2t}'\right)^{-1}\left(\frac{1}{n^2}\sum_{t=1}^{n} Y_{2t}y_{1t}\right) \\ &\Longrightarrow \left(\int_0^1 B_2(r)\,B_2(r)'\,dr\right)^{-1}\int_0^1 B_2(r)\,B_1(r)\,dr.\end{aligned}$$

- In addition, let

$$s^2 = \frac{1}{n} \sum_{t=1}^{n} \widehat{v}_t^2,$$

and it can be shown that

$$\frac{s^2}{n} \implies \int_0^1 [B_1(r)]^2 dr - \left(\int_0^1 B_1(r) B_2(r)' dr \left(\int_0^1 B_2(r) B_2(r)' dr \right)^{-1} \times \int_0^1 B_2(r) B_1(r) dr \right)$$

Spurious Regression

- Thus, s^2 diverges as $n \nearrow \infty$; more precisely,

$$s^2 = O_p\left(n\right)$$

Intuitively, this is because $\widehat{v}_t \equiv I\left(1\right)$ in this case.

- Moreover, since

$$\left(\sum_{t=1}^{n} Y_{2t} Y_{2t}'\right)^{-1} = O_p\left(\frac{1}{n^2}\right)$$

It follows that

$$\widehat{\sigma}^2_{\widehat{\beta}_i} = s^2 \left[\left(\sum_{t=1}^{n} Y_{2t} Y_{2t}'\right)^{-1}\right]_{ii} = O_p\left(\frac{1}{n}\right)$$

Spurious Regression

- Hence, in this case,

$$\widehat{\sigma}^2_{\widehat{\beta}_i} = s^2 \left[\left(\sum_{t=1}^{n} Y_{2t} Y_{2t}' \right)^{-1} \right]_{ii} \xrightarrow{p} 0 \text{ as } n \to \infty$$

- An important consequence of this is that

$$T_n = \frac{\widehat{\beta}_i}{\widehat{\sigma}_{\widehat{\beta}_i}} \text{ diverges with probability approaching one.}$$

- The above result holds even if $\omega_{21} = 0$ and $\{u_t\} \equiv i.i.d.N(0, I_m)$ as in the original Granger-Newbold spurious regression setup. Hence, even if y_{1t} and Y_{2t} are unrelated at any frequency, testing the null hypothesis $H_0 : \beta_i = 0$ using the usual t-statistic might nevertheless give a false sense of regression significance.

- Consider an $m \times 1$ vector time series Y_t such that

$$Y_t \equiv I(1).$$

Y_t is said to be cointegrated if there exists at least one nonzero (and nonrandom) $m \times 1$ vector a such that

$$a' Y_t \equiv I(0).$$

In this case, a is called a cointegrating vector.

- **Remark:** For simplicity, consider first the case where $m = 2$. Intuitively, cointegration in this bivariate system means that the two components of the random vector $Y_t = (Y_{1t}, Y_{2t})'$ share a common stochastic trend. By taking linear combination with respect to a cointegrating vector, this common stochastic trend is elminated or annihilated, so that

$$a_1 Y_{1t} + a_2 Y_{2t} \equiv I(0)$$

Cointegration

- **Remark (con't):** In fact, if $a_1 \neq 0$, i.e., the cointegration is nontrivial, then we can normalize by dividing through by a_1 to obtain

$$Y_{1t} + \frac{a_2}{a_1} Y_{2t} = u_t \equiv I(0)$$

or

$$Y_{1t} = \beta Y_{2t} + u_t$$

where

$$\beta = -\frac{a_2}{a_1}$$

Hence, with appropriate normalization, cointegration relationship can be reframed as a regression relationship. However, note that, unlike spurious regression, here $u_t \equiv I(0)$.

Cointegration

- If $m > 2$, then there may be more than one (linearly independent) cointegrating vectors. Hence, we define the cointegrating rank as the number of linearly independent cointegrating vectors in a multivariate system.
- More precisely, suppose that

$$A = (a_1, ..., a_r)$$

is an $m \times r$ matrix with rank $r < m$ such that

$$A' Y_t \equiv I(0),$$

and suppose further that

$$\underline{a}' Y_t \equiv I(1)$$

for any other $m \times 1$ vector $\underline{a}$ that is linearly independent of the columns of A, then we say that the cointegrating rank is r.

- Note that the cointegrating vectors $(a_1, .., a_r)$ are not unique since if

$$A' Y_t \equiv I(0);$$

then,

$$b' A' Y_t \equiv I(0),$$

for any nonzero $r \times 1$ vector b. Hence, $d = Ab$ is also a cointegrating vector.

- Note that in the case where the cointegrating rank $= m$; then, Y_t must itself be $I(0)$ as well, since in this case

$$A' Y_t = u_t \equiv I(0)$$

and A is an $m \times m$ nonsingular matrix, implying that

$$Y_t = (A')^{-1} u_t \equiv I(0).$$

- An interpretation for cointegration may be given as follows. Suppose that

$$A'Y_t = 0$$

describes some long-run “equilibrium" relationship. Then,

$$A'Y_t = u_t \equiv I\left(0\right)$$

measures the “equilibrium error" or the extent that the system is out of equilibrium. We would then expect this error to be $I\left(0\right)$.

- **Restrictions Implied by Cointegration:** Consider the $m \times 1$ vector process

$$\Delta Y_t = \mu + \Psi(L)\varepsilon_t$$

where $\Delta = 1 - L$ is the first difference operator and where

$$\begin{aligned} \{\varepsilon_t\} &\equiv i.i.d.(0, \Omega), \ \ \Omega > 0 \text{ and} \\ \Psi(L) &= \sum_{j=0}^{\infty} \Psi_j L^j \text{ with } \Psi_0 = I_m. \end{aligned}$$

We want to show that, for $A' Y_t$ to be $I(0)$, it must be true that

$$\text{(i) } A'\Psi(1) = 0 \text{ and (ii) } A'\mu = 0.$$

- To proceed, note first that, by the Beveridge-Nelson decomposition, we have

$$\Psi(L) = \Psi(1) - \widetilde{\Psi}(L)(1 - L),$$

where

$$\widetilde{\Psi}(L) = \sum_{j=0}^{\infty} \widetilde{\Psi}_j L^j \text{ with } \widetilde{\Psi}_j = \sum_{s=j+1}^{\infty} \Psi_s.$$

We assume that $\sum_{j=0}^{\infty} \sqrt{j} \, \|\Psi_j\| < \infty$ which implies that $\sum_{j=0}^{\infty} \left\| \widetilde{\Psi}_j \right\|^2 < \infty$.

- Given this setup, we can write

$$\begin{aligned}\Delta Y_j &= \mu + \Psi(L)\varepsilon_j \\ &= \mu + \Psi(1)\varepsilon_j - \widetilde{\Psi}(L)(1-L)\varepsilon_j \\ &= \mu + \Psi(1)\varepsilon_j - \widetilde{\Psi}(L)(\varepsilon_j - \varepsilon_{j-1}) \\ &= \mu + \Psi(1)\varepsilon_j - (\widetilde{\varepsilon}_j - \widetilde{\varepsilon}_{j-1})\end{aligned}$$

where $\widetilde{\varepsilon}_j = \widetilde{\Psi}(L)\varepsilon_j$.

- Summing both sides of the above equation from $j = 1$ to t, we get

$$\sum_{j=1}^{t} \Delta Y_j = \sum_{j=1}^{t} \mu + \Psi(1) \sum_{j=1}^{t} \varepsilon_j - \sum_{j=1}^{t} (\widetilde{\varepsilon}_j - \widetilde{\varepsilon}_{j-1})$$

or

$$Y_t - Y_0 = \mu t + \Psi(1) \sum_{j=1}^{t} \varepsilon_j - (\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_0)$$

- or

$$Y_t = Y_0 + \mu t + \Psi(1) \sum_{j=1}^{t} \varepsilon_j - (\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_0)$$

From the above equation, it is apparent that the nonstationary components of Y_t come from
(i) the linear trend term: μt and
(ii) the stochastic trend term: $\Psi(1) \sum_{j=1}^{t} \varepsilon_j$
Premultiplying the above equation by A', we get

$$A'Y_t = A'Y_0 + A'\mu t + A'\Psi(1) \sum_{j=1}^{t} \varepsilon_j - A'(\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_0)$$

from which it is apparent that for $A'Y_t$ to be $I(0)$, it must be that

$$\begin{aligned} A'\Psi(1) &= 0 \quad \text{(stochastic cointegration restriction)} \\ A'\mu &= 0 \quad \text{(deterministic cointegration restriction)} \end{aligned}$$

- **Remark:** Note also that

$$A'\Psi(1) = 0$$

implies that

$$|\Psi(z)| = 0 \text{ when } z = 1$$

i.e., $z = 1$ is a root of the determinantal equation $|\Psi(z)| = 0$. Hence, $\Psi(z)$ is noninvertible. This, in turn, implies that a cointegrated system cannot be represented by a finite order vector autoregression in the first differenced data ΔY_t.

- **Additional Restrictions Implied by Cointegration:** Note that, although a VAR in first differences is not compatible with a cointegrated system, a VAR in levels could be. Now, suppose that vector moving average process

$$\Delta Y_t = \mu + \Psi(L)\varepsilon_t$$

has a VAR representation in levels, i.e.,

$$Y_t = \alpha + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t$$

More succinctly, we can write

$$\Phi(L) Y_t = \alpha + \varepsilon_t$$

where $\Phi(L) = I_m - \Phi_1 L - \cdots - \Phi_p L^p$.

- **Additional Restrictions Implied by Cointegration (con't):**
  We want to show that if the cointegrating rank equals r, then, it is possible to write

$$\Phi(1) = -BA' \ \text{ (i.e., } \Phi(1) \text{ is of reduced rank)}$$

where $\Phi(1) = I_m - \Phi_1 - \cdots - \Phi_p$, where A $(m \times r)$ is the cointegrating matrix and where B $(m \times r)$ is called the loading matrix.

- To show this, we first multiply both sides of

$$\Delta Y_t = \mu + \Psi\left(L\right)\varepsilon_t$$

by $\Phi\left(L\right)$ to obtain

$$\left(1-L\right)\Phi\left(L\right)Y_t = \Phi\left(1\right)\mu + \Phi\left(L\right)\Psi\left(L\right)\varepsilon_t$$

Now, substituting the right-hand side of the equation $\Phi\left(L\right)Y_t = \alpha + \varepsilon_t$ into the equation above, we get

$$\left(1-L\right)\left(\alpha + \varepsilon_t\right) = \Phi\left(1\right)\mu + \Phi\left(L\right)\Psi\left(L\right)\varepsilon_t$$

or

$$\left(1-L\right)\varepsilon_t = \Phi\left(1\right)\mu + \Phi\left(L\right)\Psi\left(L\right)\varepsilon_t$$

since $\left(1-L\right)\alpha = 0$.

- Since the above equation must hold for all realization of ε_t, this suggests that

$$\Phi(1)\mu = 0$$

and

$$(1-z) I_m = \Phi(z)\Psi(z)$$

for all $z = e^{i\omega}$ with $-\pi \le \omega \le \pi$. In particular, for $\omega = 0$ or $z = 1$, we have

$$\Phi(1)\Psi(1) = 0$$

Let ϕ_i' $(1 \times m)$ be the i^{th} row of $\Phi(1)$; then,

$$\begin{aligned} \phi_i'\mu &= 0, \\ \phi_i'\Psi(1) &= 0. \end{aligned}$$

so that ϕ_i is a cointegrating vector.

- Let the cointegrating rank equals r and let $a_1, ..., a_r$ form a basis for the space of cointegrating vectors, then there exists a $r \times 1$ vector b_i^* such that

$$\begin{aligned} \phi_i &= (a_1, .., a_r)\, b_i^* \\ &= \underset{m \times r}{A} \underset{r \times 1}{b_i^*} . \end{aligned}$$

- Doing this for all rows of $\Phi(1)$, we have

$$\Phi(1)' = (\phi_1, ..., \phi_m) = A\,(b_1^*, ..., b_m^*) = AB^{*\prime}$$

or

$$\Phi(1) = B^* A' = -BA',$$

where $B = -B^*$.

Vector Error-Correction Model (VECM)

- Consider the $VAR(p)$ model

$$Y_t = \alpha + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t$$

where $\{\varepsilon_t\} \equiv i.i.d.\,(0, \Omega)$ with $\Omega > 0$.

- Using a Beveridge-Nelson type decomposition of the matrix polynomial $\Phi(z) = I_m - \Phi_1 z - \cdots - \Phi_p z^p$, we can rewrite this VAR model as

$$Y_t = \alpha + HY_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t,$$

where

$$\begin{aligned} H &= \Phi_1 + \cdots + \Phi_p \\ \Gamma_i &= \left\{ \begin{array}{cc} -\Phi_p & \text{for } i = p-1 \\ -\left[\Phi_{i+1} + \cdots + \Phi_p\right] & \text{for } i = 1, ..., p-2 \end{array} \right\} \end{aligned}$$

Vector Error-Correction Model (VECM)

- Subtracting Y_{t-1} from both sides of the equation

$$Y_t = \alpha + HY_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t,$$

we further obtain

$$\Delta Y_t = \alpha + \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t,$$

where

$$\begin{aligned} \Pi &= H - I_m \\ &= (\Phi_1 + \cdots + \Phi_p) - I_m \\ &= -\Phi(1) \end{aligned}$$

- If there are r linearly independent cointegrating relations (i.e., if the cointegrating rank $= r$), it follows from our previous result that

$$\Pi = -\Phi(1) = BA'.$$

Vector Error-Correction Model (VECM)

- Imposing this reduced rank restriction, and we have arrived at the vector error-correction representation

$$\Delta Y_t = \alpha + BA' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t,$$

- **Remark (i):** In the case where A is known, all regressors on the right-hand side of the vector error-correction model given above are $I(0)$.
- **Remark (ii):** Note also that, in the error-correction representation given above, the change in Y_t (i.e., ΔY_t) depends not only on its lagged values but also on the magnitude of the "equilibrium error" $A' Y_{t-1}$.

Vector Error-Correction Model (VECM)

- **Remark (iii):** In the absence of additional restrictions (or normalization), B and A are not separately identified since, for any nonsingular $r \times r$ matrix F, we have

$$BA' = (BF)\left(F^{-1}A'\right) = \overline{B}\overline{A}'$$

where $\overline{B} = BF$ and $\overline{A} = AF'^{-1}$. That is (B, A) is observationally equivalent to $\left(\overline{B}, \overline{A}\right)$ in the sense that they would give rise to the same value of the likelihood function.

- **Remark (iv):** In practice, a normalization that is often used to achieve identification is to set

$$\underset{r \times m}{A'} = \begin{bmatrix} \underset{r \times r}{I_r} & \underset{r \times (m-r)}{-\Gamma} \end{bmatrix}$$

Partition Y_t conformably, we get

$$\underset{m \times 1}{Y_t} = \begin{bmatrix} \underset{r \times 1}{Y_{1t}} \\ \underset{(m-r) \times 1}{Y_{2t}} \end{bmatrix}$$

- **Remark (iv) (con't):** It follows that, under this normalization, we have

$$A'Y_t = \left[\begin{array}{cc} I_r & -\Gamma \end{array} \right] \left[\begin{array}{c} Y_{1t} \\ Y_{2t} \end{array} \right] = Y_{1t} - \Gamma Y_{2t} = u_t \equiv I(0)$$

or

$$Y_{1t} = \Gamma Y_{2t} + u_t \text{ (a multivariate regression representation)}$$

This is known as Phillips' Triangular Representation (cf. Phillips, 1990).

- **Remark (v):** In the case where the cointegrating matrix A is not known, it can be shown that the usual estimators of A (OLS or ML) will be super-consistent in the sense that the convergence rate will be n instead of $\sqrt{n}$, provided that A can be identified by the normalization restriction discussed earlier. Hence, one can envision a two-step procedure where one first gets an estimate of A, say $\widehat{A} = \begin{bmatrix} I_r & -\widehat{\Gamma} \end{bmatrix}$, by running the regression

$$Y_{1t} = \Gamma Y_{2t} + u_t$$

and then plug this estimate into the VECM specification and then estimate the remaining parameters by running the second-stage regression

$$\Delta Y_t = \alpha + B\widehat{A}' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \widehat{\varepsilon}_t,$$

Vector Error-Correction Model (VECM)

- Rewriting the vector error-correction model

$$\Delta Y_t = \alpha + BA' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t,$$

  as

$$\Delta Y_t - \Gamma_1 \Delta Y_{t-1} - \cdots - \Gamma_{p-1} \Delta Y_{t-p+1} = \alpha + BA' Y_{t-1} + \varepsilon_t$$

  and taking expectation on both sides of the above equation, we get

$$\left(I_m - \Gamma_1 - \cdots - \Gamma_{p-1} \right) \mu = \alpha - B\mu_1^*$$

  where $\mu_1^* = -E\left[A' Y_{t-1}\right]$ and $\mu = E\left[\Delta Y_t\right]$. (Recall that $\Delta Y_t = \mu + \Psi\left(L\right) \varepsilon_t$).
- Since the roots of

$$\left| I_m - \Gamma_1 z - \cdots - \Gamma_{p-1} z^{p-1} \right| = 0$$

  are all outside the unit circle, i.e., $z = 1$ is not a root of the determinantal equation given above, it follows that $\left(I_m - \Gamma_1 - \cdots - \Gamma_{p-1} \right)$ is invertible.

- Hence, we can write

$$\mu = \left(I_m - \Gamma_1 - \cdots - \Gamma_{p-1}\right)^{-1}\left(\alpha - B\mu_1^*\right)$$

where $\mu_1^* = -E\left[A'Y_{t-1}\right]$ and $\mu = E\left[\Delta Y_t\right]$. (Recall that $\Delta Y_t = \mu + \Psi\left(L\right)\varepsilon_t$). It follows that, for this system to have no drift in any of the variables (i.e., $\mu = 0$), we would have to impose the restriction

$$\alpha = B\mu_1^*.$$

Otherwise, there are potentially $m - r$ separate time trends in Y_t.

Testing the Null Hypothesis of No Cointegration

- Suppose that $\underset{m\times 1}{Y_t} \equiv I(1)$ but suppose that economic theory suggests that the possible existence of a particular cointegrating vector $\underset{m\times 1}{a}$. (Note that, here, a is a known vector) In this case, the null hypothesis of no cointegration can be tested in a straightforward manner as follows:

(i) Construct $u_t = a'Y_t$.

(ii) Test the null hypothesis that $u_t \equiv I(1)$ using either the Phillips-Perron test of the augmented Dickey-Fuller test.

(iii) If the null hypothesis that $u_t \equiv I(1)$ is rejected; then, there is evidence in favor of cointegration. Otherwise, one finds evidence for an absence of cointegration.

- **Example - Testing Purchasing Power Parity (PPP):** This theory asserts that apart from transportation cost, goods should be sold for the same price in two countries. To be more specific, let

  P_t - index of price level in U.S. (in dollars per good)

  P_t^*- index of price level in the U.K. (in pounds per good)

  S_t - rate of exchange between the two currencies (in dollars per pou

  Under PPP, we would have

  $$P_t = S_t P_t^*$$

  so that, upon taking a logarithmic transformation on both sides of the above equation, we get

  $$\ln P_t = \ln S_t + \ln P_t^*.$$

Testing the Null Hypothesis of No Cointegration

- **Example - Testing Purchasing Power Parity (con't):** However, since in practice various reasons (such as errors in measuring prices, transportation costs, and differences in quality) prevent PPP from holding exactly in every time period t, a weaker but empiricallly more plausible version of PPP may be formulated as

$$\ln P_t - \ln S_t - \ln P_t^* = u_t \equiv I(0),$$

i.e., $\ln P_t$, $\ln S_t$, and $\ln P_t^*$ are cointegrated with cointegrating vector $a = (1, -1, -1)'$. Given data on P_t, S_t, and P_t^*; the sequence $\{u_t\}$ is an observed time series, so one can test the null hypothesis that $u_t \equiv I(1)$ (i.e., the weak form of PPP does not hold) versus the alternative hypothesis that $u_t \equiv I(0)$ (i.e., the weak form of PPP does hold) using either the Phillips-Perron test or the augmented Dickey-Fuller test.

Testing the Null Hypothesis of No Cointegration

- **Case 2- Cointegrating Vector Must Be Estimated:** Consider now the case where the true value of the cointegrating vector is unknown and must be estimated. More precisely, let $\{Y_t\}$ denote an $m \times 1$ vector time series such that $Y_t \equiv I(1)$. Partition Y_t as follows

$$Y_t = \begin{pmatrix} \underset{1\times 1}{y_{1t}} \\ \underset{g\times 1}{Y_{2t}} \end{pmatrix}$$

and consider the time series regression

$$y_{1t} = \beta' Y_{2t} + v_t$$

Note that we would expect that

$$\begin{aligned} v_t &\equiv I(0) \text{ if } Y_t \text{ is cointegrated} \\ v_t &\equiv I(1) \text{ if } Y_t \text{ is } \textbf{not} \text{ cointegrated} \end{aligned}$$

- Hence, it seems that we can design a test for cointegration based on estimating the regression

$$y_{1t} = \beta' Y_{2t} + v_t$$

and then testing the residual process

$$\widehat{v}_t = y_{1t} - \widehat{\beta}_n' Y_{2t}$$

for the presence of a unit root using either the Phillips-Perron test or the augmented Dickey-Fuller test. Here, $\widehat{\beta}_n$ denotes the OLS estimator of β in the regression above. One complication with this test strategy is that, under the null hypothesis, $v_t \equiv I(1)$ (i.e., Y_t is not cointegrated), so that we have a spurious regression situation. As we have discussed previously, in this case, $\widehat{\beta}_n$ is not a consistent estimtor of any population quantity and $\widehat{v}_t$ is also not a residual in the ordinary sense.

Residual Based Tests of Cointegration

- It turns out, however, that the Phillips-Perron test and the augmented Dickey-Fuller test can still be applied in this situation but the asymptotic critical values are different from those used in the usual applications of these tests.
- To implement the Phillips-Perron test, we estimate the regression

$$\widehat{v}_t = \rho \widehat{v}_{t-1} + e_t, \ t = 2, ..., n;$$

from which we obtain the OLS estimator

$$\widehat{\rho}_n = \frac{\sum_{t=2}^{n} \widehat{v}_{t-1} \widehat{v}_t}{\sum_{t=2}^{n} \widehat{v}_{t-1}^2}$$

- The null hypothesis that $\rho = 1$ can now be tested using the statistics

$$Z_\rho = (n-1)\left(\widehat{\rho}_n - 1\right) - \frac{1}{2}\left\{\frac{(n-1)^2 \widehat{\sigma}^2_{\widehat{\rho}_n}}{s_n^2}\right\}\left(\widehat{\omega}^2 - s_n^2\right),$$

$$Z_t = \left(\frac{s_n}{\widehat{\omega}}\right) T_n - \frac{1}{2}\left\{\frac{(n-1)\widehat{\sigma}_{\widehat{\rho}_n}}{s_n}\right\}\left[\frac{\widehat{\omega}^2 - s_n^2}{\widehat{\omega}}\right]$$

where

$$s_n^2 = \frac{1}{n-1}\sum_{t=2}^{n} \widehat{e}_t^2 \quad \text{(OLS estimate of the variance of } e_t\text{)}$$

$$\widehat{\sigma}^2_{\widehat{\rho}_n} = \frac{s_n^2}{\sum_{t=2}^{n} \widehat{v}_{t-1}^2}$$

(the usual OLS formula for the sample variance of $\widehat{\rho}_n$)

Residual Based Tests of Cointegration

- $$\begin{aligned} \widehat{\omega}^2 &= \widehat{\gamma}_0 + 2 \sum_{j=1}^{q(n)} \left[1 - \frac{j}{q(n) + 1} \right] \widehat{\gamma}_j \quad \text{(Newey-West estimator)} \\ T_n &= \frac{\widehat{\rho}_n - 1}{\widehat{\sigma}_{\widehat{\rho}_n}} \end{aligned}$$

  where $\widehat{e}_t = \widehat{v}_t - \widehat{\rho}_n \widehat{v}_{t-1}$ and where

  $$\widehat{\gamma}_j = \frac{1}{n-1} \sum_{t=j+2}^{n} \widehat{e}_t \widehat{e}_{t-j} \text{ for } j = 0, 1, ..., q(n).$$

- Phillips and Ouliaris (1990) show that if $q(n) \to \infty$ as $n \to \infty$ such that $q(n)/n \to 0$, then under some additional conditions

$$Z_\rho \Longrightarrow \int_0^1 R(r)\, dR(r) \text{ and } Z_t \Longrightarrow \int_0^1 R(r)\, dS(r)$$

where

$$\begin{aligned} R(r) &= \frac{Q(r)}{\left[\int_0^1 [Q(r)]^2 dr\right]^{1/2}},\ S(r) = \frac{Q(r)}{(\kappa'\kappa)^{1/2}}, \\ Q(r) &= W_1(r) \\ &\quad - \left[\left(\int_0^1 W_1(r) W_2(r)' dr\right)\left(\int_0^1 W_2(r) W_2(r)' dr\right)^{-1} \right. \\ &\qquad \left. \times W_2(r)\right] \end{aligned}$$

Residual Based Tests of Cointegration

- $$\underset{1\times g}{\kappa'} = \left(1, -\left(\int_0^1 W_1(r) W_2(r)' dr\right)\left(\int_0^1 W_2(r) W_2(r)' dr\right)^{-1}\right)$$
  $$W(r) = \left(\begin{array}{cc} \underset{1\times 1}{W_1(r)} & \underset{1\times g}{W_2(r)'} \end{array}\right)' \equiv BM(I_m).$$

- To implement the augmented Dickey-Fuller test, we estimate the regression

$$\widehat{v}_t = \rho \widehat{v}_{t-1} + \gamma_1 \Delta \widehat{v}_{t-1} + \cdots + \gamma_{p-1} \Delta \widehat{v}_{t-p+1} + \eta_t$$

Residual Based Tests of Cointegration

- The null hypothesis $H_0 : \rho = 1$ can be tested using the t-statistic

$$T_n = \frac{\widehat{\rho}_n - 1}{\sqrt{s_n^2 \left(\widehat{v}_{-1}' M_X \widehat{v}_{-1}\right)^{-1}}}$$

where

$$s_n^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} \left(\widehat{v}_t - \widehat{\rho}_n \widehat{v}_{t-1} - \widehat{\gamma}_1 \Delta \widehat{v}_{t-1} - \cdots - \widehat{\gamma}_{p-1} \Delta \widehat{v}_{t-p+1}\right)^2$$

$$X = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}, \text{ with } x_t = \left(\Delta \widehat{v}_{t-1}, \ldots., \Delta \widehat{v}_{t-p+1}\right)'.$$

Residual Based Tests of Cointegration

- Phillips and Ouliaris (1990) show that

$$T_n \Longrightarrow \int_0^1 R(r)\, dS(r)$$

if $p \to \infty$ as $n \to \infty$ such that

$$p = o\left(n^{1/3}\right).$$

- Consider the m-variate error-correction model

$$\Delta Y_t = \Gamma_0 Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

where

$$\{\varepsilon_t\} \equiv i.i.d.N\left(0, \Omega\right) \text{ with } \Omega > 0.$$

Following the approach of Johansen (1988), we will derive a likelihood ratio test for the null hypothesis that the cointegrating rank $= r$, in which case we have the reduced rank restriction

$$\Gamma_0 = BA'$$

where A and B are $m \times r$ matrices of full column rank $r < m$. As noted previously, A is the cointegrating matrix whereas B is called the loading matrix.

- **Conditional Log-likelihood Function:** Conditional on the initial observations $\left(Y_{-p+1}, Y_{-p+2}, ..., Y_0\right)$, we can write the log-likelihood function for the vector error-correction model as

$$\begin{aligned} & \ln L\left(\Omega, \Gamma_0, \Gamma_1, ..., \Gamma_{p-1}\right) \\ = \; & -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln\left|\Omega\right| \\ & -\frac{1}{2}\sum_{t=1}^{n}\left\{\left(\Delta Y_t - \Gamma_0 Y_{t-1} - \Gamma_1 \Delta Y_{t-1} - \cdots - \Gamma_{p-1}\Delta Y_{t-p+1}\right)'\Omega^{-1}\right. \\ & \qquad \left. \times\left(\Delta Y_t - \Gamma_0 Y_{t-1} - \Gamma_1 \Delta Y_{t-1} - \cdots - \Gamma_{p-1}\Delta Y_{t-p+1}\right)\right\} \end{aligned}$$

- Johansen (1988) maximizes this log-likelihood function in steps.
- **Step 1:**

(a) Regress ΔY_t on $\Delta Y_{t-1}, ..., \Delta Y_{t-p+1}$ by OLS to obtain the $m \times 1$ vector of residuals $\widehat{u}_t$ from the estimated regression

$$\Delta Y_t = \widehat{\Pi}_1 \Delta Y_{t-1} + \cdots + \widehat{\Pi}_{p-1} \Delta Y_{t-p+1} + \widehat{u}_t$$

(b) Also, regress Y_{t-1} on $\Delta Y_{t-1}, ..., \Delta Y_{t-p+1}$ by OLS to obtain the $m \times 1$ vector of residuals $\widehat{v}_t$ from the estimated regression

$$Y_{t-1} = \widehat{\Theta}_1 \Delta Y_{t-1} + \cdots + \widehat{\Theta}_{p-1} \Delta Y_{t-p+1} + \widehat{v}_t$$

(c) Concentrate the log-likelihood function with respect to $\Gamma_1, ..., \Gamma_{p-1}$ to obtain

$$\begin{aligned} & \ln L\left(\Omega, \Gamma_0\right) \\ = \; & -\frac{nm}{2} \ln 2\pi - \frac{n}{2} \ln |\Omega| - \frac{1}{2} \sum_{t=1}^{n} \left(\widehat{u}_t - \Gamma_0 \widehat{v}_t\right)' \Omega^{-1} \left(\widehat{u}_t - \Gamma_0 \widehat{v}_t\right) \end{aligned}$$

- **Step 1:**

(c) Moreover, under the null hypothesis

$$H_0 : \text{ cointegrating rank } = r$$

we have $\Gamma_0 = BA'$, so that we can further write the concentrated log-likelihood function as

$$\begin{aligned} & \ln L\left(\Omega, A, B\right) \\ = \quad & -\frac{nm}{2} \ln 2\pi - \frac{n}{2} \ln \left|\Omega\right| - \frac{1}{2} \sum_{t=1}^{n} \left(\widehat{u}_t - BA'\widehat{v}_t\right)' \Omega^{-1} \left(\widehat{u}_t - BA'\widehat{v}_t\right). \end{aligned}$$

- Before discussing step 2, we first define some notations which will be useful. Let

$$\underset{m\times 1}{Z_{0t}} = \Delta Y_t,\ \underset{m\times 1}{Z_{1t}} = Y_{t-1}, \text{ and } \underset{m(p-1)\times 1}{Z_{2t}} = \left(\Delta Y_{t-1}', ..., \Delta Y_{t-p+1}'\right)'$$

- Also, define

$$M_{ij} = \frac{1}{n}\sum_{t=1}^{n} Z_{it}Z_{jt}' \ \ i,j = 0,1,2$$

- Note that, based on these notations, we have

$$\begin{aligned} \widehat{u}_t &= Z_{0t} - M_{02}M_{22}^{-1}Z_{2t}, \\ \widehat{v}_t &= Z_{1t} - M_{12}M_{22}^{-1}Z_{2t}. \end{aligned}$$

- Further define

$$
\begin{aligned}
& S_{uu} \\
= \; & \frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t\widehat{u}_t' \\
= \; & \frac{1}{n}\sum_{t=1}^{n}\left(Z_{0t}-M_{02}M_{22}^{-1}Z_{2t}\right)\left(Z_{0t}-M_{02}M_{22}^{-1}Z_{2t}\right)' \\
= \; & \frac{1}{n}\sum_{t=1}^{n}Z_{0t}Z_{0t}'-M_{02}M_{22}^{-1}\frac{1}{n}\sum_{t=1}^{n}Z_{2t}Z_{0t}'-\frac{1}{n}\sum_{t=1}^{n}Z_{0t}Z_{2t}'M_{22}^{-1}M_{20} \\
& +M_{02}M_{22}^{-1}\frac{1}{n}\sum_{t=1}^{n}Z_{2t}Z_{2t}'M_{22}^{-1}M_{20} \\
= \; & M_{00}-M_{02}M_{22}^{-1}M_{20}
\end{aligned}
$$

- Similarly, we have

$$S_{uv} = \frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t\widehat{v}_t' = M_{01} - M_{02}M_{22}^{-1}M_{21},$$
$$S_{vv} = \frac{1}{n}\sum_{t=1}^{n}\widehat{v}_t\widehat{v}_t' = M_{11} - M_{12}M_{22}^{-1}M_{21}.$$

- **Step 2:** To further concentrate the function

$$\begin{aligned} & \ln L(\Omega, A, B) \\ = \; & -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln|\Omega| - \frac{1}{2}\sum_{t=1}^{n}\left(\widehat{u}_t - BA'\widehat{v}_t\right)'\Omega^{-1}\left(\widehat{u}_t - BA'\widehat{v}_t\right). \end{aligned}$$

we will, for a fixed A, maximize $\ln L(\Omega, A, B)$ with respect to B and Ω by regressing $\widehat{u}_t$ on $A'\widehat{v}_t$ to obtain

$$\begin{aligned} \widehat{B}(A) &= \frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t\widehat{v}_t'A\left(A'\left[\frac{1}{n}\sum_{t=1}^{n}\widehat{v}_t\widehat{v}_t'\right]A\right)^{-1} \\ &= S_{uv}A\left(A'S_{vv}A\right)^{-1} \end{aligned}$$

- and

$$
\begin{aligned}
& \widehat{\Omega}(A) \\
= \; & \frac{1}{n} \sum_{t=1}^{n} \left(\widehat{u}_t - \widehat{B}(A) A' \widehat{v}_t \right) \left(\widehat{u}_t - \widehat{B}(A) A' \widehat{v}_t \right)' \\
= \; & \frac{1}{n} \sum_{t=1}^{n} \widehat{u}_t \widehat{u}_t' - \widehat{B}(A) A' \frac{1}{n} \sum_{t=1}^{n} \widehat{v}_t \widehat{u}_t' - \frac{1}{n} \sum_{t=1}^{n} \widehat{u}_t \widehat{v}_t' A \widehat{B}(A)' \\
& + \widehat{B}(A) A' \frac{1}{n} \sum_{t=1}^{n} \widehat{v}_t \widehat{v}_t' A \widehat{B}(A)' \\
= \; & S_{uu} - 2 S_{uv} A \left(A' S_{vv} A \right)^{-1} A' S_{vu} \\
& + S_{uv} A \left(A' S_{vv} A \right)^{-1} \left(A' S_{vv} A \right) \left(A' S_{vv} A \right)^{-1} A' S_{vu} \\
= \; & S_{uu} - S_{uv} A \left(A' S_{vv} A \right)^{-1} A' S_{vu} \\
= \; & S_{uu} - S_{uv} A \left(A' S_{vv} A \right)^{-1} \left(A' S_{vv} A \right) \left(A' S_{vv} A \right)^{-1} A' S_{vu} \\
= \; & S_{uu} - \widehat{B}(A) \left(A' S_{vv} A \right) \widehat{B}(A)' .
\end{aligned}
$$

- Substituting $\widehat{B}(A)$ and $\widehat{\Omega}(A)$ into

$$
\begin{aligned}
& \ln L(\Omega, A, B) \\
= \ & -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln|\Omega| - \frac{1}{2}\sum_{t=1}^{n}\left(\widehat{u}_t - BA'\widehat{v}_t\right)'\Omega^{-1}\left(\widehat{u}_t - BA'\widehat{v}_t\right),
\end{aligned}
$$

we obtain

$$
\begin{aligned}
& \ln L(A) \\
= \ & -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln\left|\widehat{\Omega}(A)\right| \\
& -\frac{1}{2}\sum_{t=1}^{n}\left(\widehat{u}_t - \widehat{B}(A)A'\widehat{v}_t\right)'\widehat{\Omega}(A)^{-1}\left(\widehat{u}_t - \widehat{B}(A)A'\widehat{v}_t\right) \\
= \ & -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln\left|\widehat{\Omega}(A)\right| \\
& -\frac{1}{2}\sum_{t=1}^{n} tr\left\{\widehat{\Omega}(A)^{-1}\left(\widehat{u}_t - \widehat{B}(A)A'\widehat{v}_t\right)\left(\widehat{u}_t - \widehat{B}(A)A'\widehat{v}_t\right)'\right\}
\end{aligned}
$$

Johansen's Maximum Likelihood Procedure

- or

$$
\begin{aligned}
& \ln L(A) \\
= & -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln\left|\widehat{\Omega}(A)\right| \\
& -\frac{n}{2}tr\left\{\widehat{\Omega}(A)^{-1}\sum_{t=1}^{n}\frac{\left(\widehat{u}_t - \widehat{B}(A)A'\widehat{v}_t\right)\left(\widehat{u}_t - \widehat{B}(A)A'\widehat{v}_t\right)'}{n}\right\} \\
= & -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln\left|\widehat{\Omega}(A)\right| - \frac{n}{2}tr\left\{\widehat{\Omega}(A)^{-1}\widehat{\Omega}(A)\right\} \\
= & -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln\left|\widehat{\Omega}(A)\right| - \frac{n}{2}tr\left\{I_m\right\} \\
= & -\frac{nm}{2}\ln 2\pi - \frac{n}{2}\ln\left|\widehat{\Omega}(A)\right| - \frac{nm}{2} \\
= & C - \frac{n}{2}\ln\left|S_{uu} - S_{uv}A\left(A'S_{vv}A\right)^{-1}A'S_{vu}\right|
\end{aligned}
$$

where $C = -\frac{nm}{2}(\ln 2\pi + 1)$.

Johansen's Maximum Likelihood Procedure

- **Step 3:** From the concentrated log-likelihood function

$$\ln L(A) = C - \frac{n}{2} \ln \left| S_{uu} - S_{uv} A \left(A' S_{vv} A \right)^{-1} A' S_{vu} \right|$$

it is clear that maximizing $\ln L(A)$ with respect to A is the same as minimizing

$$\ln \left| S_{uu} - S_{uv} A \left(A' S_{vv} A \right)^{-1} A' S_{vu} \right|$$

with respect to A. Moreover, note that from a standard result for determinants of partitioned matrices, i.e.,

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = |A_{11}| \left| A_{22} - A_{21} A_{11}^{-1} A_{12} \right| = |A_{22}| \left| A_{11} - A_{12} A_{22}^{-1} A_{21} \right|,$$

- **Step 3 (con't):** we have that

$$\begin{vmatrix} S_{uu} & S_{uv}A \\ A'S_{vu} & A'S_{vv}A \end{vmatrix} = |S_{uu}| \left|A'S_{vv}A - A'S_{vu}S_{uu}^{-1}S_{uv}A\right| = |A'S_{vv}A| \left|S_{uu} - S_{uv}A\left(A'S_{vv}A\right)^{-1}A'S_{vu}\right|$$

from which it follows that

$$\left|S_{uu} - S_{uv}A\left(A'S_{vv}A\right)^{-1}A'S_{vu}\right| = \frac{|S_{uu}| \left|A'\left(S_{vv} - S_{vu}S_{uu}^{-1}S_{uv}\right)A\right|}{|A'S_{vv}A|}$$

- It follows that maximizing $\ln L(A)$ with respect to A is equivalent to minmizing the objective function

$$Q(A) = \frac{\left|A'\left(S_{vv} - S_{vu}S_{uu}^{-1}S_{uv}\right)A\right|}{|A'S_{vv}A|}$$

with respect to A.

Johansen's Maximum Likelihood Procedure

- To do so, we make use of the following lemma
  Lemma: Let M be an $m \times m$ symmetric and positive semidefinite matrix and let N be an $m \times m$ symmetric and positive definite matrix. The function
  $$Q(X) = \frac{|X'MX|}{|X'NX|}$$
  is maximized (alternatlvely, minimized) among all $m \times r$ matrices (with $r < m$) by $\widehat{X} = (\widehat{x}_1, .., \widehat{x}_r)$ $\left(\text{alternatively, } \widehat{X} = (\widehat{x}_{m-r+1}, .., \widehat{x}_m)\right)$ and the maximal (alternatively, minimal) value is
  $$\prod_{i=1}^{r} \lambda_i \left(\text{alternatively, } \prod_{i=m-r+1}^{m} \lambda_i\right)$$
  where λ_i and $\widehat{x}_i$ $(i = 1, ..., m)$ are solutions of the generalized eignevalue problem
  $$M\widehat{x}_i = \lambda_i N\widehat{x}_i.$$

Johansen's Maximum Likelihood Procedure

- Here, $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$ can be obtained as the roots of the determinantal equation

$$|\lambda N - M| = 0$$

where $\widehat{x}_1, ..., \widehat{x}_r$ are the associated eigenvectors.

- **Remark:** We can also choose $\widehat{X}$ times any nonsingular $r \times r$ matrix as the maximizing (alternatively, minimizing) argument. To see this, let

$$\widehat{X} = \arg\min Q\left(X\right) = \arg\min \frac{|X'MX|}{|X'NX|}$$

and let B is any nonsingular $r \times r$ matrix. Let

$$\widetilde{X} = \widehat{X}B.$$

- **Remark (con't):** Then,

$$\begin{aligned} Q\left(\widetilde{X}\right) &= \frac{\left|\widetilde{X}'M\widetilde{X}\right|}{\left|\widetilde{X}'N\widetilde{X}\right|} = \frac{\left|B'\widehat{X}'M\widehat{X}B\right|}{\left|B'\widehat{X}'N\widehat{X}B\right|} \\ &= \frac{|B|\left|\widehat{X}'M\widehat{X}\right||B|}{|B|\left|\widehat{X}'N\widehat{X}\right||B|} = \frac{\left|\widehat{X}'M\widehat{X}\right|}{\left|\widehat{X}'N\widehat{X}\right|} \\ &= Q\left(\widehat{X}\right) \end{aligned}$$

so that $\widetilde{X}$ also minimizes $Q(X)$.

Johansen's Maximum Likelihood Procedure

- Now, in light of the lemma above, to minimize the objective function

$$Q\left(A\right)=\frac{\left|A'\left(S_{vv}-S_{vu}S_{uu}^{-1}S_{uv}\right)A\right|}{\left|A'S_{vv}A\right|},$$

  we need to solve the (generalized) eigenvalue problem

$$\left|\rho S_{vv}-\left(S_{vv}-S_{vu}S_{uu}^{-1}S_{uv}\right)\right|=0$$

  for the r smallest roots.

- Moreover, set $\lambda=1-\rho$ and note that solving the above eigenvalue problem is the same as solving the slightly modified eigenvalue problem

$$\left|\lambda S_{vv}-S_{vu}S_{uu}^{-1}S_{uv}\right|=0$$

  for the r largest eigenvalues $\widehat{\lambda}_1, ..., \widehat{\lambda}_r$ and the associated eigenvectors $\widehat{x}_1, ..., \widehat{x}_r$ which satisfy the equation

$$\widehat{\lambda}_i S_{vv}\widehat{x}_i=S_{vu}S_{uu}^{-1}S_{uv}\widehat{x}_i.$$

Johansen's Maximum Likelihood Procedure

- Next, we impose that normalization

$$\widehat{x}_j' S_{vv} \widehat{x}_i = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

so that

$$\widehat{X}' S_{vv} \widehat{X} = I_r$$

- From the equation

$$\widehat{\lambda}_i S_{vv} \widehat{x}_i = S_{vu} S_{uu}^{-1} S_{uv} \widehat{x}_i,$$

we also see that

$$\widehat{x}_j' S_{vu} S_{uu}^{-1} S_{uv} \widehat{x}_i = \begin{cases} \widehat{\lambda}_i & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

- Hence,

$$\widehat{X}' S_{vu} S_{uu}^{-1} S_{uv} \widehat{X} = \widehat{\Lambda}_r = diag\left(\widehat{\lambda}_1, ..., \widehat{\lambda}_r\right)$$

- Let $\widehat{A}_r = \widehat{X}$ and note that

$$\begin{aligned} & \ln L\left(\widehat{A}_r\right) \\ = & -\frac{nm}{2}\left(\ln 2\pi + 1\right) - \frac{n}{2}\ln\left[\frac{|S_{uu}|\left|\widehat{A}_r'\left(S_{vv} - S_{vu}S_{uu}^{-1}S_{uv}\right)\widehat{A}_r\right|}{\left|\widehat{A}_r' S_{vv}\widehat{A}_r\right|}\right] \end{aligned}$$

- so that

$$
\begin{aligned}
& \ln L\left(\widehat{A}_r\right) \\
= & -\frac{nm}{2}\left(\ln 2\pi + 1\right) - \frac{n}{2}\ln\left|S_{uu}\right| \\
& -\frac{n}{2}\ln\left[\frac{\left|\widehat{A}_r'\left(S_{vv} - S_{vu}S_{uu}^{-1}S_{uv}\right)\widehat{A}_r\right|}{\left|\widehat{A}_r' S_{vv}\widehat{A}_r\right|}\right] \\
= & -\frac{nm}{2}\left(\ln 2\pi + 1\right) - \frac{n}{2}\ln\left|S_{uu}\right| - \frac{n}{2}\ln\left[\prod_{i=1}^{r}\left(1 - \widehat{\lambda}_i\right)\right]
\end{aligned}
$$

- It further follows that

$$L\left(\widehat{A}_r\right) = C'\left[\prod_{i=1}^{r}\left(1 - \widehat{\lambda}_i\right)\right]^{-n/2},$$

where

$$C' = (2\pi)^{-nm/2}\left|S_{uu}\right|^{-n/2}\exp\left\{-\frac{nm}{2}\right\}.$$

- Suppose we wish to test

$$H_0 : \text{ cointegrating rank } = r$$

versus

$$H_1 : \text{ cointegrating rank } > r$$

- The likelihood ratio test statistic for testing the above null hypothesis is given by

$$LR = \frac{L\left(\widehat{A}_r\right)}{L\left(\widehat{A}_m\right)} = \frac{C'\left[\prod_{i=1}^{r}\left(1-\widehat{\lambda}_i\right)\right]^{-n/2}}{C'\left[\prod_{i=1}^{m}\left(1-\widehat{\lambda}_i\right)\right]^{-n/2}}$$

Johansen's Likelihood Ratio Test

- Simplifying, we have

$$
\begin{aligned}
LR &= \frac{C'\left[\prod_{i=1}^{r}\left(1-\widehat{\lambda}_i\right)\right]^{-n/2}}{C'\left[\prod_{i=1}^{m}\left(1-\widehat{\lambda}_i\right)\right]^{-n/2}} \\
&= \frac{\left[\prod_{i=1}^{m}\left(1-\widehat{\lambda}_i\right)\right]^{n/2}}{\left[\prod_{i=1}^{r}\left(1-\widehat{\lambda}_i\right)\right]^{n/2}} \\
&= \left[\prod_{i=r+1}^{m}\left(1-\widehat{\lambda}_i\right)\right]^{n/2}
\end{aligned}
$$

Johansen's Likelihood Ratio Test

- Johansen (1988) showed that, under H_0,

$$\begin{aligned}
& -2 \ln LR \\
= & -n \sum_{i=r+1}^{m} \ln \left(1 - \widehat{\lambda}_i\right) \\
= & \; n \sum_{i=r+1}^{m} \widehat{\lambda}_i + o_p(1) \\
= & \; tr \left\{ \left[\int_0^1 W(r)\, dW(r)' \right]' \left[\int_0^1 W(r)\, W(r)'\, dr \right] \right. \\
& \left. \times \left[\int_0^1 W(r)\, dW(r)' \right] \right\}
\end{aligned}$$

where $W(r) \equiv BM(I_{m-r})$.

Johansen's Sequential Procedure for Cointegrating Rank Determination

- Johansen (1992) proposes estimating the cointegrating rank using a sequence of likelihood ratio tests

$$
\begin{array}{ccccc}
\text{Test } H_0 : r = 0 & \overset{\text{reject}}{\longrightarrow} & \text{Test } H_0 : r = 1 & \overset{\text{reject}}{\longrightarrow} & \cdots \\
\text{fail to reject } \downarrow & & \text{fail to reject } \downarrow & & \\
\widehat{r} = 0 & & \widehat{r} = 1 & &
\end{array}
$$

$$
\begin{array}{ccccc}
\cdots & \overset{\text{reject}}{\longrightarrow} & \text{Test } H_0 : r = m - 1 & \overset{\text{reject}}{\longrightarrow} & \widehat{r} = m \\
& & \text{fail to reject } \downarrow & & \\
& & \widehat{r} = m - 1 & &
\end{array}
$$

Asymptotic Properties of the Johansen Sequential Procedure

- If all tests are performed using a fixed significance level α; then, results given in Johansen (1992) show that his sequential procedure produces an estimator $\widehat{r}$ of the cointegrating rank with the following asymptotic properties

(i) $\Pr\left(\widehat{r} = r^0\right) \rightarrow 1 - \alpha$ as $n \rightarrow \infty$.

(ii) $\Pr\left(\widehat{r} < r^0\right) \rightarrow 0$ as $n \rightarrow \infty$.

(iii) $\Pr\left(\widehat{r} > r^0\right) \rightarrow \alpha$ as $n \rightarrow \infty$.

Here, r^0 denotes the true cointegrating rank.

Joint Estimation of Cointegrating Rank and VAR Lag Order Using Order Selection Methods

- Consider a family of the m-variate vector error correction models (VECMs)

$$\mathcal{M}_{p,r} : \Delta Y_t = B_r A_r' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

where

$$\begin{aligned} \mathcal{P} &= \{0, 1, ..., \overline{p}\}, \ \mathcal{R} = \{0, 1, ..., m\}, \\ \{\varepsilon_t\} &\equiv i.i.d. N(0, \Omega) \text{ with } \Omega > 0. \end{aligned}$$

- Suppose we impose the à priori normalization

$$\underset{r \times m}{A_r'} = \left[\begin{array}{cc} \underset{r \times r}{I_r} & \underset{r \times (m-r)}{\overline{A}_r'} \end{array} \right],$$

For $\mathcal{M}_{p,r}$,

$$\# \text{ of free parameters } = mr + r(m-r) + m^2(p-1)$$

Joint Estimation of Cointegrating Rank and VAR Lag Order Using Order Selection Methods

- The VECM can be estimated by the maximum likelihood method from which we obtain the residual vector

$$\widehat{\varepsilon}_t \left(p, r \right) = \Delta Y_t - \widehat{B}_r \widehat{A}_r' Y_{t-1} - \widehat{\Gamma}_1 \Delta Y_{t-1} - \cdots - \widehat{\Gamma}_{p-1} \Delta Y_{t-p+1},$$

Here,

$$\widehat{B}_r, \widehat{A}_r = \left[\begin{array}{cc} I_r & \widehat{\overline{A}}_r' \end{array} \right], \widehat{\Gamma}_1, ..., \widehat{\Gamma}_{p-1}$$

denote, respectively, the maximum likelihood estimates of the loading matrix, the cointegrating matrix, and the coefficient matrices of the short-run dynamics.

- Using the residual vectors, we can also define the maximum likelihood estimate of the error covariance matrix Ω, viz

$$\widehat{\Omega} \left(p, r \right) = \frac{1}{n} \sum_{t=1}^{n} \widehat{\varepsilon}_t \left(p, r \right) \widehat{\varepsilon}_t \left(p, r \right)'$$

Joint Estimation of Cointegrating Rank and VAR Lag Order Using Order Selection Methods

- Chao and Phillips (1999) proposed jointly estimating p and r using information criteria. In particular, the paper analyzed the large sample properties of the following information criteria for estimating p and r.

(a) AIC (Akaike Information Criterion)
Criterion:

$$AIC(p, r) = \ln \left| \widehat{\Omega}(p, r) \right| + \left\{ mr + r(m - r) + m^2 (p - 1) \right\} \frac{2}{n}$$

Order Estimates:

$$\left(\widehat{p}_{AIC}, \widehat{r}_{AIC} \right) = \arg \min_{p \in \mathcal{P},\ r \in \mathcal{R}} AIC(p, r)$$

where

$$\mathcal{P} = \{0, 1, ..., \overline{p}\},\ \mathcal{R} = \{0, 1, ..., m\}$$

Joint Estimation of Cointegrating Rank and VAR Lag Order Using Order Selection Methods

(b) BIC (Bayesian Information Criterion - also known as the Schwarz Criterion)
Criterion:

$$BIC(p,r) = \ln \left| \widehat{\Omega}(p,r) \right| + \left\{ mr + r(m-r) + m^2 (p-1) \right\} \frac{\ln n}{n}$$

Order Estimates:

$$(\widehat{p}_{BIC}, \widehat{r}_{BIC}) = \arg \min_{p \in \mathcal{P},\ r \in \mathcal{R}} BIC(p,r)$$

where

$$\mathcal{P} = \{0, 1, \ldots, \overline{p}\},\ \mathcal{R} = \{0, 1, \ldots, m\}$$

(c) PIC (Posterior Information Criterion)
Criterion:

$$
\begin{aligned}
& PIC\left(p, r\right) \\
= \;& \ln\left|\widehat{\Omega}\left(p, r\right)\right| + \frac{m}{n}\ln\left|\widehat{U}'\widehat{U}\right| + \frac{r}{n}\ln\left|Y_{2,-1}'Y_{2,-1}\right| \\
& + \frac{\left(m-r\right)}{n}\ln\left|\widehat{B}_r'\widehat{\Omega}\left(p, r\right)\widehat{B}_r\right| \\
= \;& \ln\left|\widehat{\Omega}\left(p, r\right)\right| + \frac{m}{n}\ln n^{r+m(p-1)} + \frac{r}{n}\ln n^{2(m-r)} + O_p\left(n^{-1}\right) \\
= \;& \ln\left|\widehat{\Omega}\left(p, r\right)\right| + \left\{mr + 2r\left(m-r\right) + m^2\left(p-1\right)\right\}\frac{\ln n}{n} + O_p\left(n^{-1}\right)
\end{aligned}
$$

Order Estimates:

$$
\left(\widehat{p}_{PIC}, \widehat{r}_{PIC}\right) = \arg\min_{p\in\mathcal{P},\; r\in\mathcal{R}} PIC\left(p, r\right)
$$

(c) PIC (Posterior Information Criterion)
Here,

$$\underset{[r+m(p-1)]\times 1}{\widehat{U}_t} = \begin{bmatrix} \underset{r\times 1}{\widehat{A}_r' Y_{t-1}} \\ \Delta Y_{t-1} \\ \vdots \\ \Delta Y_{t-p+1} \end{bmatrix} = \begin{bmatrix} Y_{1,t-1} - \widehat{\overline{A}}_r' Y_{2,t-1} \\ \Delta Y_{t-1} \\ \vdots \\ \Delta Y_{t-p+1} \end{bmatrix},$$

$$\underset{n\times[r+m(p-1)]}{\widehat{U}} = \begin{bmatrix} \widehat{U}_1' \\ \widehat{U}_2' \\ \vdots \\ \widehat{U}_n' \end{bmatrix} \text{ and } \underset{n\times(m-r)}{Y_{2,-1}} = \begin{bmatrix} Y_{2,0}' \\ Y_{2,1}' \\ \vdots \\ Y_{2,n-1}' \end{bmatrix}$$

Joint Estimation of Cointegrating Rank and VAR Lag Order Using Order Selection Methods

- Asymptotic Properties:

(a) **AIC:**

$$\widehat{p}_{AIC} \overset{p}{\nrightarrow} p^0 \text{ and } \widehat{r}_{AIC} \overset{p}{\nrightarrow} r^0 \text{ as } n \longrightarrow \infty$$

(b) **BIC:**

$$\widehat{p}_{BIC} \overset{p}{\longrightarrow} p^0 \text{ and } \widehat{r}_{BIC} \overset{p}{\longrightarrow} r^0 \text{ as } n \longrightarrow \infty$$

(c) **PIC:**

$$\widehat{p}_{PIC} \overset{p}{\longrightarrow} p^0 \text{ and } \widehat{r}_{PIC} \overset{p}{\longrightarrow} r^0 \text{ as } n \longrightarrow \infty$$