# Unit Roots and Co-integration

*Topic 3: Unit Roots*

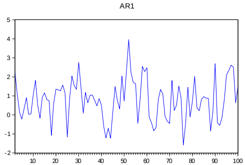
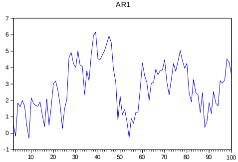
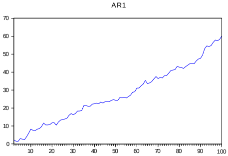
by Vanessa Berenguer-Rico

University of Oxford

January 2014

# The Unit Root Land

- Let $\varepsilon_t \sim i.i.d.\,(0,1)$, $x_0 = 0$ and consider the following DGPs:

| $x_t = 0.5 + 0.3x_{t-1} + \varepsilon_t;$ | $x_t = 0.5 + 0.7x_{t-1} + \varepsilon_t$ | $x_t = 0.5 + x_{t-1} + \varepsilon_t$ |
|---|---|---|
|  |  |  |

# The Unit Root Land

- Recall:

$$x_t = \beta + x_{t-1} + u_t; \quad u_t \sim i.i.d.\left(0, \sigma^2\right)$$

$$x_t = x_0 + \beta t + \sum_{j=1}^{t} u_j$$

- Stochastic Properties

$$E\left[x_t\right] = x_0 + \beta t$$

$$V\left[x_t\right] = \sigma^2 t$$

$$Cov\left[x_t, x_s\right] = \min\left\{t, s\right\}\sigma^2$$

Some Properties of Unit Root processes:

- The rate of growth of a unit root process is stationary:

$$x_t = \beta + x_{t-1} + u_t \quad vs \quad \Delta x_t = \beta + u_t$$

- Shocks have a permanent effect on the future of the series
- The forecast error $(x_{t+h} - E\left[x_{t+h}|I_t\right])$ is unbounded if the forecast horizon tends to infinity
- Standard inference does not hold

# The Unit Root Land

Consider the OLS estimation of the AR(1) process,

$$y_t = \rho y_{t-1} + u_t,$$

where $u_t \sim i.i.d.N\left(0, \sigma^2\right)$ and $y_0 = 0$. The OLS estimate of $\rho$ is given by

$$\hat{\rho}_T = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2} = \rho + \frac{\sum_{t=1}^{T} y_{t-1} u_t}{\sum_{t=1}^{T} y_{t-1}^2}.$$

# The Unit Root Land

- If $|\rho| < 1$, then the LLN and the CLT can be applied to obtain the asymptotic distribution of the OLS estimator of $\rho$. See Hamilton (p. 215)
- LLN:

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{p} E\left[y_{t-1}^2\right] = \frac{\sigma^2}{1-\rho^2}.$$

- CLT:

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T} y_{t-1}u_t \xrightarrow{d} N\left(0, \lim_{T\to\infty} E\left[y_{t-1}^2 u_t^2\right]\right) = N\left(0, \frac{\sigma^4}{1-\rho^2}\right)$$

- Therefore,

$$T^{1/2}\left(\hat{\rho}_T - \rho\right) \xrightarrow{d} N\left(0, 1-\rho^2\right)$$

# The Unit Root Land

- Then, if $\rho = 1$, the distribution collapses to a point mass at zero; that is, $T^{1/2}(\hat{\rho}_T - \rho) \xrightarrow{p} 0$. Obviously, this is not very helpful for hypothesis testing

- To obtain a non-degenerate asymptotic distribution for $\hat{\rho}_T$ in the unit root case, it turns out that we have to multiply $\hat{\rho}_T$ by $T$ rather than by $T^{1/2}$

- Thus, the unit root coefficient converges at a faster rate $(T)$ than a coefficient for a stationary regression $\left(T^{1/2}\right)$, but at a slower rate than the coefficient on a time trend $\left(T^{3/2}\right)$

- Moreover, the asymptotic distribution when $\rho = 1$ is not standard. It can be described in terms of functionals of **Brownian motions**

# Brownian Motion

## Definition

A Standard Brownian motion $W(.)$ is a continuous-time stochastic process, associating each date $r \in [0,1]$ with the scalar $W(r)$ such that:

(a) $W(0) = 0$

(b) For any dates $0 \leq r_1 < r_2 < ... < r_k \leq 1$, the changes $[W(r_2) - W(r_1)], [W(r_3) - W(r_2)], ..., [W(r_k) - W(r_{k-1})]$ are independent Gaussian with $[W(s) - W(r)] \sim N(0, s - r)$

(c) For a given realization, $W(r)$ is continuous in $r$ with probability 1

# The Functional Central Limit Theorem

- The CLT establishes convergence of random variables, the FCLT establishes conditions for convergence of random functions

- Let $\varepsilon_t$ be an *i.i.d.* $(0, \sigma^2)$ sequence

- The CLT considers

$$T^{1/2}\bar{\varepsilon}_T = T^{1/2}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t$$

- The FCLT considers

$$T^{1/2}X_T(r) = T^{1/2}\frac{1}{T}\sum_{t=1}^{[Tr]}\varepsilon_t$$

# The Functional Central Limit Theorem

- Consider an estimator of the sample mean that only considers the $r$th fraction of the observations, $r \in [0, 1]$, that is

$$X_T\left(r\right) = \frac{1}{T}\sum_{t=1}^{[Tr]} \varepsilon_t,$$

  where $[Tr]$ denotes the integer part of $Tr$

- Then, for any given realization, $X_T\left(r\right)$ is a step **function** in $r$:

$$X_T\left(r\right) = \begin{cases} 0 & 0 \leq r < 1/T \\ \varepsilon_1/T & 1/T \leq r < 2/T \\ \left(\varepsilon_1 + \varepsilon_2\right)/T & 2/T \leq r < 3/T \\ \vdots & \vdots \\ \left(\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_T\right)/T & r = 1 \end{cases}$$

# The Functional Central Limit Theorem

The simplest FCLT is known as Donsker's theorem (Donsker, 1951)

**Theorem**

*Let $\varepsilon_t$ be a sequence of i.i.d. random variables with mean zero. If $\sigma^2 \equiv var(\varepsilon_t) < \infty$, $\sigma^2 \neq 0$, then*

$$T^{1/2} X_T(r) / \sigma \xrightarrow{d} W(r)$$

Billingsley (1968)!

# The Functional Central Limit Theorem

$$\underline{S_T(r) = T^{1/2} X_T(r) / \sigma \xrightarrow{d} W(r) = S(r) \text{ if:}}$$

**(a)** For any finite collection of $k$ particular dates, $0 \leq r_1 \leq r_2 < ... < r_k \leq 1$,

$$(S_T(r_1), S_T(r_2), ..., S_T(r_k)) \xrightarrow{d} (S(r_1), S(r_2), ..., S(r_k)).$$

That is, the finite-dimensional distributions of $S_T(r)$ converge to those of $S(r)$.

**(b)** For each $\varepsilon > 0$,

$$P\left(\sup_{|r_1 - r_2| < \delta} |S_T(r_1) - S_T(r_2)| > \varepsilon\right) \longrightarrow 0,$$

uniformly in $T$ as $\delta \to 0$.

**(c)** $P(|S_T(0)| > \lambda) \to 0$ uniformly in $T$ as $\lambda \to \infty$.

# The Continuous Mapping Theorem

- The Continuous Mapping Theorem, CMT, states that if $Y_T(.) \xrightarrow{d} Y(.)$ and $g$ is a continuos functional, then $g(Y_T(.)) \xrightarrow{d} g(Y(.))$
- Example: $S_T(r) = T^{1/2} X_T(r) \xrightarrow{d} \sigma W(r)$
- Example: $S_T^2(r) = \left[T^{1/2} X_T(r)\right]^2 \xrightarrow{d} \sigma^2 [W(r)]^2$
- Example: $\int_0^1 S_T(r)\,dr = \int_0^1 T^{1/2} X_T(r)\,dr \xrightarrow{d} \sigma \int_0^1 W(r)\,dr$
- Example: $\int_0^1 S_T^2(r)\,dr = \int_0^1 \left[T^{1/2} X_T(r)\right]^2 dr \xrightarrow{d} \sigma^2 \int_0^1 [W(r)]^2\,dr$

# Application to Unit Root Processes

- Consider the random walk

$$y_t = y_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim i.i.d.\left(0, \sigma^2\right)$, and $y_0 = 0$, so that

$$y_t = \sum_{j=1}^{t} \varepsilon_j$$

- Then, one can construct the stochastic function $X_T\left(r\right)$ as follows

$$X_T\left(r\right) \begin{cases} 0 & 0 \leq r < 1/T \\ y_1/T = \varepsilon_1/T & 1/T \leq r < 2/T \\ y_2/T = \left(\varepsilon_1 + \varepsilon_2\right)/T & 2/T \leq r < 3/T \\ \vdots & \vdots \\ y_T/T = \left(\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_T\right)/T & r = 1 \end{cases}$$

# Application to Unit Root Processes

- Notice that

$$\int_0^1 X_T(r)\,dr = \frac{y_1}{T^2} + \frac{y_2}{T^2} + \ldots + \frac{y_T}{T^2} = \frac{1}{T^2}\sum_{t=1}^{T} y_t$$

- Hence,

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_t = \int_0^1 T^{1/2} X_T(r)\,dr \xrightarrow{d} \sigma \int_0^1 W(r)\,dr$$

- Similarly,

$$\frac{1}{T^2}\sum_{t=1}^{T} y_t^2 = \int_0^1 \left[T^{1/2} X_T(r)\right]^2 dr \xrightarrow{d} \sigma^2 \int_0^1 [W(r)]^2\,dr$$

# Application to Unit Root Processes

- Recall, if
$$y_t = \rho y_{t-1} + \varepsilon_t,$$
where $\varepsilon_t \sim i.i.d.\left(0, \sigma^2\right)$, then the OLS estimator of $\rho$ is
$$\hat{\rho}_T = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2} = \rho + \frac{\sum_{t=1}^{T} y_{t-1} \varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

- If $\rho = 1$
$$\left(\hat{\rho}_T - 1\right) = \frac{\sum_{t=1}^{T} y_{t-1} \varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}.$$

# Application to Unit Root Processes

- For the numerator, notice that $y_t^2 = y_{t-1}^2 + \varepsilon_t^2 + 2y_{t-1}\varepsilon_t$. Hence,

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t &= \frac{1}{2}\left(\frac{1}{T}\sum_{t=1}^{T}\left(y_t^2 - y_{t-1}^2\right) - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\right) \\
&= \frac{1}{2}\left(\frac{1}{T}y_T^2 - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\right) \\
&\xrightarrow{d} \frac{1}{2}\sigma^2\left(W^2(1) - 1\right)
\end{aligned}$$

- And for the denominator

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 [W(r)]^2\, dr$$

# Application to Unit Root Processes

- Therefore,

$$T\left(\hat{\rho}_T - 1\right) = \frac{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{d} \frac{\left(W^2\left(1\right) - 1\right)}{2\int_0^1 \left[W\left(r\right)\right]^2 dr}$$

- **Remark 1**: The OLS estimator converges at a rate $T$: Super-consistent!

- **Remark 2**: The asymptotic distribution is not standard

# Application to Unit Root Processes: A useful Lemma

## Lemma (iid)

*Suppose that* $y_t = y_{t-1} + \varepsilon_t$, *where* $y_0 = 0$ *and* $\varepsilon_t \sim i.i.d.\left(0, \sigma^2\right)$. *Then,*

(a) $T^{-1/2} \sum_{t=1}^{T} \varepsilon_t \xrightarrow{d} \sigma W(1)$

(b) $T^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{d} (1/2)\,\sigma^2 \left\{ [W(1)]^2 - 1 \right\}$

(c) $T^{-3/2} \sum_{t=1}^{T} t\varepsilon_t \xrightarrow{d} \sigma W(1) - \sigma \int_0^1 W(r)\, dr$

(d) $T^{-3/2} \sum_{t=1}^{T} y_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r)\, dr$

(e) $T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 [W(r)]^2\, dr$

(f) $T^{-5/2} \sum_{t=1}^{T} t y_{t-1} \xrightarrow{d} \sigma \int_0^1 r W(r)\, dr$

(g) $T^{-3} \sum_{t=1}^{T} t y_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 r [W(r)]^2\, dr$

(h) $T^{-(v+1)} \sum_{t=1}^{T} t^v \xrightarrow{d} 1/(v+1)$ *for* $v = 0, 1, \ldots$

# Allowing for Serial Correlation in the error term

- The assumption $u_t = \varepsilon_t \sim i.i.d.$ will be typically violated for many economic time series

- What if

$$y_t = y_{t-1} + u_t,$$

with

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

and

$$\varepsilon_t \sim i.i.d.\left(0, \sigma^2\right)?$$

# Allowing for Serial Correlation in the error term

Phillips and Solo (1992): **Beveridge-Nelson decomposition**

## Lemma (BN)

*Let* $\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$. *Then*

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

*where* $\tilde{\Psi}(L) = \sum_{j=0}^{\infty} \tilde{\psi}_j L^j$, $\tilde{\psi}_j = \sum_{k=j+1}^{\infty} \psi_k$. *If* $p \geq 1$, *then*

$$\sum_{j=1}^{\infty} j^p \left|\psi_j\right|^p < \infty \Longrightarrow \sum_{j=0}^{\infty} \left|\tilde{\psi}_j\right|^p < \infty \text{ and } \left|\Psi(1)\right| < \infty.$$

*If* $p < 1$, *then*

$$\sum_{j=1}^{\infty} j \left|\psi_j\right|^p < \infty \Longrightarrow \sum_{j=0}^{\infty} \left|\tilde{\psi}_j\right|^p < \infty.$$

# Allowing for Serial Correlation in the error term

## Lemma (BN′)

*Let*

$$u_t = \Psi(L)\,\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \varepsilon_t \sim i.i.d.\left(0, \sigma^2\right),$$

*and*

$$\sum_{j=0}^{\infty} j \left|\psi_j\right| < \infty.$$

*Then*

$$u_t = \Psi(1)\,\varepsilon_t + \eta_t - \eta_{t-1} \text{ and } \sum_{j=1}^{t} u_j = \Psi(1) \sum_{j=1}^{t} \varepsilon_j + \eta_t - \eta_0,$$

$$\eta_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j},\ \alpha_j = -\left(\psi_{j+1} + \psi_{j+2} + \psi_{j+3} + \ldots\right) \text{ and } \sum_{j=0}^{\infty} \left|\alpha_j\right| < \infty.$$

# Allowing for Serial Correlation in the error term

**Remarks**:

- Condition $\sum_{j=0}^{\infty} j \left|\psi_j\right| < \infty$ is satisfied by any stationary ARMA process

- If $y_t = y_{t-1} + u_t$, then

$$y_t = \sum_{j=1}^{t} u_j + y_0 = \Psi\left(1\right) \sum_{j=1}^{t} \varepsilon_j + \eta_t - \eta_0 + y_0$$

- Notice that $\eta_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$ is a stationary process

- Hence,

$$\frac{1}{t} \sum_{j=1}^{t} u_j = \Psi\left(1\right) \frac{1}{t} \sum_{j=1}^{t} \varepsilon_j + \frac{1}{t} \eta_t - \frac{1}{t} \eta_0 = \Psi\left(1\right) \frac{1}{t} \sum_{j=1}^{t} \varepsilon_j + o_p\left(1\right)$$

# Allowing for Serial Correlation in the error term

**Remarks**:

- Let

$$X_T(r) \equiv \frac{1}{T}\sum_{t=1}^{[Tr]} u_t,$$

where $u_t$ satisfies conditions of Lemma BN′ with $E\left[\varepsilon_t^4\right] < \infty$. Then,

$$\sqrt{T}X_T(r) \xrightarrow{d} \sigma\psi(1)\,W(r)\,.$$

- Therefore, for $r = 1$,

$$\sqrt{T}X_T(1) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t \xrightarrow{d} \sigma\psi(1)\,W(1)\,,$$

and since $W(1) \sim N(0,1)$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t \xrightarrow{d} N\left(0, \sigma^2\left[\psi(1)\right]^2\right).$$

# Allowing for Serial Correlation in the error term

"**Lemma iid**" can be easily generalized by using **BN**

## Lemma (LP)

*Let* $u_t = \Psi(L)\varepsilon_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$ *where* $\sum_{j=0}^{\infty} j\left|\psi_j\right| < \infty$, $\varepsilon_t \sim i.i.d.\left(0,\sigma^2\right)$ *and* $E\left[\varepsilon_t^4\right] < \infty$*. Define*

$$\gamma_j \equiv E\left(u_t u_{t-j}\right) = \sigma^2 \sum_{s=0}^{\infty} \psi_s \psi_{s+j} \quad \text{for } j = 0, 1, 2, ...$$

$$\lambda \equiv \sigma \sum_{j=0}^{\infty} \psi_j = \sigma\psi(1)$$

$$y_t \equiv \sum_{j=1}^{t} u_j \quad \text{for } t = 1, 2, ..., T$$

*with* $y_0 = 0$*. Then,*

# Allowing for Serial Correlation in the error term

**Lemma (LP)**

(a) $T^{-1/2}\sum_{t=1}^{T}u_t \xrightarrow{d} \lambda W(1)$

(b) $T^{-1/2}\sum_{t=1}^{T}u_{t-j}\varepsilon_t \xrightarrow{d} N\left(0,\sigma^2\gamma_0\right)$ *for* $j=1,2,...$

(c) $T^{-1}\sum_{t=1}^{T}u_t u_{t-j} \xrightarrow{p} \gamma_j$ *for* $j=0,1,2,...$

(d) $T^{-1}\sum_{t=1}^{T}y_{t-1}\varepsilon_t \xrightarrow{d} (1/2)\,\sigma\lambda\left\{[W(1)]^2-1\right\}$

(e) $T^{-1}\sum_{t=1}^{T}y_{t-1}u_{t-j} \xrightarrow{d}$

$$\begin{cases} (1/2)\left\{\lambda^2[W(1)]^2-\gamma_0\right\} & \text{for } j=0 \\ (1/2)\left\{\lambda^2[W(1)]^2-\gamma_0\right\}+\sum_{s=0}^{j-1}\gamma_s & \text{for } j=1,2,... \end{cases}$$

# Allowing for Serial Correlation in the error term

## Lemma (LP)

(f) $T^{-3/2}\sum_{t=1}^{T} y_{t-1} \xrightarrow{d} \lambda \int_0^1 W(r)\,dr$

(g) $T^{-3/2}\sum_{t=1}^{T} tu_{t-j} \xrightarrow{d} \lambda \left\{W(1) - \int_0^1 W(r)\,dr\right\}$ *for* $j = 0, 1, 2, \ldots$

(h) $T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \lambda^2 \int_0^1 [W(r)]^2\,dr$

(i) $T^{-5/2}\sum_{t=1}^{T} ty_{t-1} \xrightarrow{d} \lambda \int_0^1 rW(r)\,dr$

(j) $T^{-3}\sum_{t=1}^{T} ty_{t-1}^2 \xrightarrow{d} \lambda^2 \int_0^1 r[W(r)]^2\,dr$

(k) $T^{-(v+1)}\sum_{t=1}^{T} t^v \xrightarrow{d} 1/(v+1)$ *for* $v = 0, 1, \ldots$

# Allowing for Serial Correlation in the error term

**Remarks**:

- Again, there are simpler ways to describe individual results. For example:

(i) $\lambda W(1)$ in (a) is a $N\left(0, \lambda^2\right)$ distribution

(ii) $(1/2)\,\sigma\lambda\left\{[W(1)]^2 - 1\right\}$ in (d) is $(1/2)\,\sigma\lambda\left\{\chi^2(1) - 1\right\}$

(ii) $\lambda\int_0^1 W(r)\,dr$ and $\lambda\left\{W(1) - \int_0^1 W(r)\,dr\right\}$ in (f) and (g), respectively, are both $N\left(0, \lambda^2/3\right)$ distributions

- Lemmas idd and IP can be used to construct unit root tests

# Testing for Unit Roots

- Unit Root tests are hypotheses testing procedures whose objective is to determine whether a process contains a unit root
- Traditional unit root tests are designed for **testing the null hypothesis of a unit root versus the hypothesis of trend-stationary** (although there are many other types)
- Among this group, one the most popular test for unit roots is the pioneer Dickey-Fuller (DF) test
- The literature with regard to this area is very large and we only give here a very brief outline to the **DF test**
- For an overview on unit root testing: Phillips and Xiao (1998)

It is important to bear in mind that in sharp contrast to standard inference

- The asymptotic distribution of the DF statistic is not standard
- Whether the true model contains or not deterministic components and whether the regression model contains or not deterministic components changes the asymptotic distribution and then, different tables of critical values should be used in each case

# The Dickey-Fuller test

- The Hypotheses

$$\begin{cases} H_o : y_t \sim I(1) \\ H_a : y_t \sim I(0) \end{cases}$$

- The Auxiliary Regression

$$y_t = \rho y_{t-1} + u_t,$$

and hence

$$\begin{cases} H_o : y_t \sim I(1) \equiv \rho = 1 \\ H_a : y_t \sim I(0) \equiv \rho < 1 \end{cases}$$

- Equivalently,

$$\Delta y_t = \theta y_{t-1} + u_t,$$

where $\theta = (\rho - 1)$ and hence

$$\begin{cases} H_o : y_t \sim I(1) \equiv \theta = 0 \\ H_a : y_t \sim I(0) \equiv \theta < 0 \end{cases}$$

# The Dickey-Fuller test

- Three possible specifications to consider deterministic components:
- **No Deterministic Components**:

$$(i) \quad \Delta y_t = \theta y_{t-1} + u_t$$

- **Constant Term**:

$$(ii) \quad \Delta y_t = \alpha + \theta y_{t-1} + u_t$$

- **Linear Trend**:

$$(iii) \quad \Delta y_t = \alpha + \beta t + \theta y_{t-1} + u_t$$

# The Dickey-Fuller test

- Which specification to use in practice?
- One should use an auxiliary regression that is plausible under both $H_o$ and $H_1$
- A graphical simple device:
- If the data looks trended, then $(iii)$ would offer a plausible specification under both hypothesis
- Otherwise, $(ii)$ is recommended

# The Dickey-Fuller test

- Auxiliary Regression

$$\Delta y_t = f(t) + \theta y_{t-1} + u_t,$$

where $\theta = (\rho - 1)$

- Two scenarios:

(a) $u_t$ uncorrelated: DF test

(b) $u_t$ correlated: Augmented DF (ADF) test, Phillips-Perron (1988)

# The Dickey-Fuller test

Consider the following case

$$y_t = y_{t-1} + u_t \text{ where } u_t = \varepsilon_t \sim i.i.d. \left(0, \sigma^2\right)$$

- Auxiliary Regression

$$\Delta y_t = \theta y_{t-1} + u_t,$$

where $\theta = (\rho - 1)$

- Test statistic under the $H_o : \theta = 0$ is

$$t_\theta = \frac{\hat{\theta}_T}{\hat{\sigma}_{\hat{\theta}_T}} = \frac{\frac{1}{T} \sum_{t=1}^{T} y_{t-1} \varepsilon_t}{\left(\frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^2\right)^{1/2} s_T},$$

where, given that $\hat{\theta}_T = \hat{\rho}_T - 1$,

$$s_T^2 = \frac{1}{(T-1)} \sum_{t=1}^{T} (y_t - \hat{\rho}_T y_{t-1})^2$$

# The Dickey-Fuller test

Recall

$$y_t = y_{t-1} + u_t \text{ where } u_t = \varepsilon_t \sim i.i.d. \left(0, \sigma^2\right),$$

Therefore,
(i)

$$s_T^2 \xrightarrow{p} \sigma^2$$

(ii)

$$\frac{1}{T} \sum_{t=1}^{T} y_{t-1} \varepsilon_t \xrightarrow{d} (1/2)\, \sigma^2 \left\{ [W(1)]^2 - 1 \right\}$$

(iii)

$$\frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 [W(r)]^2\, dr$$

# The Dickey-Fuller test

- Hence,

$$t_{\theta} \xrightarrow{d} \frac{(1/2)\left\{[W(1)]^2 - 1\right\}}{\left(\int_0^1 [W(r)]^2 \, dr\right)^{1/2}}$$

- **Remark**: This distribution is not standard and therefore, it has to be tabulated. Tables of critical values can be found in the Appendix of most time series books. Check DFtest.prg!

- **Remark**: Whether the true model contains or not deterministic components and whether the regression model contains or not deterministic components changes the asymptotic distribution and then, different tables of critical values should be use in each case

# The Dickey-Fuller test

- The assumption $u_t = \varepsilon_t \sim i.i.d.$ will be typically violated for many economic time series

- To relax the previous assumption one could model $u_t$ as a stationary process that admits a Wold decomposition $u_t = \psi(L)\,\varepsilon_t$ with $\varepsilon_t \sim i.i.d.$

- It is important to see that in this case the distribution of the DF test will be different

- We will consider briefly both the **ADF** test and the **Phillips-Perron** correction

# The Augmented Dickey-Fuller test

- The ADF test proposes a parametric correction to address the presence of autocorrelation in the error

- It is based on the following auxiliary regression

$$\Delta y_t = f\left(t\right) + \theta y_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta y_{t-j} + \varepsilon_t$$

- Under $H_o : \theta = 0$ the test based on the corresponding t-statistic has the same asymptotic distribution as in the non-autocorrelated case

- As before, the consideration of deterministic components will change the distribution

# The Augmented Dickey-Fuller test

- ADF auxiliary regression

$$\Delta y_t = f(t) + \theta y_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta y_{t-j} + \varepsilon_t$$

- In practice, the order $p$ is unknown and perhaps, it is infinite. Said and Dickey (1984) showed that as long as $p$ goes to infinity sufficiently slowly relative to $T$, $p = T^{1/3}$, then the OLS t-test of $\theta = 0$ can be carried out using the DF critical values

- In practice, information criteria are often use to select the order of the polynomial lags of $\Delta y_t$ (General to Specific)

# Phillips-Perron approach

- Phillips and Perron (1988) generalized the DF test, remember

$$\Delta y_t = f(t) + \theta y_{t-1} + u_t,$$

  to the case when $u_t$ is serially correlated and possibly heteroskedastic as well

- For now we will assume that the true process is

$$\Delta y_t = u_t = \psi(L)\,\varepsilon_t,$$

  where $\psi(L)$ and $\varepsilon_t$ satisfy the condition of Lemma LP

# Phillips-Perron approach

- Recall, if
$$y_t = \alpha + \rho y_{t-1} + u_t,$$
with $|\rho| < 1$ and $u_t$ is autocorrelated, then the OLS estimator, $\hat{\rho}_T$, of $\rho$ is not consistent.

- However, if $\rho = 1$, the rate $T$ of convergence of $\hat{\rho}_T$ turns out to ensure that $\hat{\rho}_T \xrightarrow{p} 1$ even when $u_t$ is serially correlated

- Phillips and Perron (1988) proposed estimating
$$y_t = f(t) + \rho y_{t-1} + u_t,$$
by OLS even when $u_t$ is serially correlated and then modifying the DF statistic to take account of the serial correlation

# Phillips-Perron approach

<u>*How to take account of the serial correlation?*</u>

**Example**: Consider the DF auxiliary regression

$$y_t = \alpha + \rho y_{t-1} + u_t,$$

when the true model is

$$y_t = y_{t-1} + u_t,$$

with $u_t$ satisfying the condition of Lemma LP. Then,

$$\left(\frac{\gamma_0}{\lambda^2}\right)^{1/2} t_\rho - \left\{\frac{1}{2}\left(\lambda^2 - \gamma_0\right)/\lambda\right\} \times \left\{T\hat{\sigma}_{\hat{\rho}_T} \div s_T\right\} \xrightarrow{d} D,$$

where $s_T^2 = (T-2)^{-1} \sum_{t=1}^{T} \left(y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1}\right)^2$ and $D$ is distributed as the DF statistic that assumes $u_t = \varepsilon_t \sim i.i.d.$ and uses the auxiliary regression

$$y_t = \alpha + \rho y_{t-1} + u_t.$$

# Phillips-Perron approach

The statistic

$$\left(\frac{\gamma_0}{\lambda^2}\right)^{1/2} t_\rho - \left\{\frac{1}{2}\left(\lambda^2 - \gamma_0\right)/\lambda\right\} \times \left\{T\hat{\sigma}_{\hat{\rho}_T} \div s_T\right\},$$

requires knowledge of the population parameters $\gamma_0$ and $\lambda^2$. Remember

$$\gamma_0 \equiv E\left(u_t^2\right) = \sigma^2 \sum_{s=0}^{\infty} \psi_s^2,$$

and

$$\lambda^2 \equiv [\sigma\psi(1)]^2 = \left[\sigma \sum_{j=0}^{\infty} \psi_j\right]^2 = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j,$$

where $\gamma_j$ is the $j$th autocovariance of $u_t$, are the short and long run variances of $u_t$, respectively.

# Phillips-Perron approach

Although $\gamma_0$ and $\lambda^2$ are unknown, they are easy to estimate consistently. Phillips-Perron (1988) used

$$\hat{\gamma}_0 = (T-2)^{-1} \sum_{t=1}^{T} \hat{u}_t^2 = s_T^2,$$

and the Newey-West estimator

$$\hat{\lambda}^2 = \hat{\gamma}_0 + \sum_{j=1}^{q} \left[1 - j/\left(q+1\right)\right] \hat{\gamma}_j,$$

where

$$\hat{\gamma}_j = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j},$$

and $\hat{u}_t = y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1}$.

# Phillips-Perron approach

**Remarks**:

- The Newey-West estimator $\hat{\lambda}^2$ can provide a consistent estimate of $\lambda^2$ for an $MA(\infty)$ process, provided that $q$ grows sufficiently slowly relative to $T$
- Phillips (1987) established such consistency assuming that $q_T \to \infty$ and $q_T / T^{1/4} \to 0$; for example $q_T = AT^{1/5}$ satisfies this requirement
- This is an asymptotic result and does not tell us exactly how $q$ should be chosen in practice

# Nelson and Plosser

- Nelson and Plosser (1982): “**Trends and Random Walks in Macroeconomic Time Series**”
- *“This paper investigates whether macroeconomic time series are better characterized as stationary fluctuations around a deterministic trend or as non-stationary processes that have no tendency to return to a deterministic path.”*
- *“Using long historical time series for the U.S. we are unable to reject the hypothesis that these series are non-stationary stochastic processes with no tendency to return to a trend line.”*

# Nelson and Plosser

**The data set**:

- The U.S. historical time series include measures of output, spending, money, prices, and interest rates
- The data are annual with starting dates varying from 1860 to 1909 and ending in 1970 in all cases
- All series except the bond yield are transformed to natural logs
- Remember: An extended version of this data set, NelsonPlosserData.wf1, available

# Nelson and Plosser

Table 2

Sample autocorrelations of the natural logs of annual data.[a]

| Series | Period | Sample autocorrelations $T$ | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_5$ | $r_6$ |
|---|---|---|---|---|---|---|---|---|
| Random walk[b] | | 100 | 0.95 | 0.90 | 0.85 | 0.81 | 0.76 | 0.70 |
| Time aggregated[b] random walk | | 100 | 0.96 | 0.91 | 0.86 | 0.82 | 0.77 | 0.73 |
| Real GNP | 1909–1970 | 62 | 0.95 | 0.90 | 0.84 | 0.79 | 0.74 | 0.69 |
| Nominal GNP | 1909–1970 | 62 | 0.95 | 0.89 | 0.83 | 0.77 | 0.72 | 0.67 |
| Real per capita GNP | 1909–1970 | 62 | 0.95 | 0.88 | 0.81 | 0.75 | 0.70 | 0.65 |
| Industrial production | 1860–1970 | 111 | 0.97 | 0.94 | 0.90 | 0.87 | 0.84 | 0.81 |
| Employment | 1890–1970 | 81 | 0.96 | 0.91 | 0.86 | 0.81 | 0.76 | 0.71 |
| Unemployment rate | 1890–1970 | 81 | 0.75 | 0.47 | 0.32 | 0.17 | 0.04 | −0.01 |
| GNP deflator | 1889–1970 | 82 | 0.96 | 0.93 | 0.89 | 0.84 | 0.80 | 0.76 |
| Consumer prices | 1860–1970 | 111 | 0.96 | 0.92 | 0.87 | 0.84 | 0.81 | 0.77 |
| Wages | 1900–1970 | 71 | 0.96 | 0.91 | 0.86 | 0.82 | 0.77 | 0.73 |
| Real wages | 1900–1970 | 71 | 0.96 | 0.92 | 0.88 | 0.84 | 0.80 | 0.75 |
| Money stock | 1889–1970 | 82 | 0.96 | 0.92 | 0.89 | 0.85 | 0.81 | 0.77 |
| Velocity | 1869–1970 | 102 | 0.96 | 0.92 | 0.88 | 0.85 | 0.81 | 0.79 |
| Bond yield | 1900–1970 | 71 | 0.84 | 0.72 | 0.60 | 0.52 | 0.46 | 0.40 |
| Common stock prices | 1871–1970 | 100 | 0.96 | 0.90 | 0.85 | 0.79 | 0.75 | 0.71 |

[a]The natural logs of all the data are used except for the bond yield. $T$ is the sample size and $r_i$ is the $i$th order autocorrelation coefficient. The large sample standard error under the null hypothesis of no autocorrelation is $T^{-\frac{1}{2}}$ or roughly 0.11 for series of the length considered here.

[b]Computed by the authors from the approximation due to Wichern (1973).

# Nelson and Plosser

Table 3

Sample autocorrelations of the first difference of the natural logs of annual data.[a]

| Series | Period | $T$ | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_5$ | $r_6$ | $s(r)$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | Sample autocorrelations | | | | | | |
| Time aggregated random walk[b] | | | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| Real GNP | 1909–1970 | 62 | 0.34 | 0.04 | −0.18 | −0.23 | −0.19 | 0.01 | 0.13 |
| Nominal GNP | 1909–1970 | 62 | 0.44 | 0.08 | −0.12 | −0.24 | −0.07 | 0.15 | 0.13 |
| Real per capita GNP | 1909–1970 | 62 | 0.33 | 0.04 | −0.17 | −0.21 | −0.18 | 0.02 | 0.13 |
| Industrial production | 1860–1970 | 111 | 0.03 | −0.11 | −0.00 | −0.11 | −0.28 | 0.05 | 0.09 |
| Employment | 1890–1970 | 81 | 0.32 | −0.05 | −0.08 | −0.17 | −0.20 | 0.01 | 0.11 |
| Unemployment rate | 1890–1970 | 81 | 0.09 | −0.29 | 0.03 | −0.03 | −0.19 | 0.01 | 0.11 |
| GNP deflator | 1889–1970 | 82 | 0.43 | 0.20 | 0.07 | −0.06 | 0.03 | 0.02 | 0.11 |
| Consumer prices | 1860–1970 | 111 | 0.58 | 0.16 | 0.02 | −0.00 | 0.05 | 0.03 | 0.09 |
| Wages | 1900–1970 | 71 | 0.46 | 0.10 | −0.03 | −0.09 | −0.09 | 0.08 | 0.12 |
| Real wages | 1900–1970 | 71 | 0.19 | −0.03 | −0.07 | −0.11 | −0.18 | −0.15 | 0.12 |
| Money stock | 1889–1970 | 82 | 0.62 | 0.30 | 0.13 | −0.01 | −0.07 | −0.04 | 0.11 |
| Velocity | 1869–1970 | 102 | 0.11 | −0.04 | −0.16 | −0.15 | −0.11 | 0.11 | 0.10 |
| Bond yield | 1900–1970 | 71 | 0.18 | 0.31 | 0.15 | 0.04 | 0.06 | 0.05 | 0.12 |
| Common stock prices | 1871–1970 | 100 | 0.22 | −0.13 | −0.08 | −0.18 | −0.23 | 0.02 | 0.10 |

[a]The first differences of the natural logs of all the data are used except for the bond yield. $T$ is the sample size and $r_i$ is the estimated $i$th order autocorrelation coefficient. The large sample standard error for $r$ is given by $s(r)$ under the null hypothesis of no autocorrelation.

[b]Theoretical autocorrelations as the number of aggregated observations becomes large; result due to Working (1960).

# Nelson and Plosser

**Table 4**

**Sample autocorrelations of the deviations from the time trend.[a]**

| Series | Period | $T$ | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_5$ | $r_6$ |
|---|---|---|---|---|---|---|---|---|
| | | | Sample autocorrelations | | | | | |
| Detrended random walk[b] | | 61 | 0.85 | 0.71 | 0.58 | 0.47 | 0.36 | 0.27 |
| | | 101 | 0.91 | 0.82 | 0.74 | 0.66 | 0.58 | 0.51 |
| Real GNP | 1909–1970 | 62 | 0.87 | 0.66 | 0.46 | 0.26 | 0.19 | 0.07 |
| Nominal GNP | 1909–1970 | 62 | 0.93 | 0.79 | 0.65 | 0.52 | 0.43 | 0.05 |
| Real per capita GNP | 1909–1970 | 62 | 0.87 | 0.65 | 0.43 | 0.24 | 0.11 | 0.04 |
| Industrial production | 1860–1970 | 111 | 0.84 | 0.67 | 0.53 | 0.40 | 0.30 | 0.28 |
| Employment | 1890–1970 | 81 | 0.89 | 0.71 | 0.55 | 0.39 | 0.25 | 0.17 |
| Unemployment rate | 1890–1970 | 81 | 0.75 | 0.46 | 0.30 | 0.15 | 0.03 | −0.01 |
| GNP deflator | 1889–1970 | 82 | 0.92 | 0.81 | 0.67 | 0.54 | 0.42 | 0.30 |
| Consumer prices | 1860–1970 | 111 | 0.97 | 0.91 | 0.84 | 0.78 | 0.71 | 0.63 |
| Wages | 1900–1970 | 71 | 0.93 | 0.81 | 0.67 | 0.54 | 0.42 | 0.31 |
| Real wages | 1900–1970 | 71 | 0.87 | 0.69 | 0.52 | 0.38 | 0.26 | 0.19 |
| Money stock | 1889–1970 | 82 | 0.95 | 0.83 | 0.69 | 0.53 | 0.37 | 0.21 |
| Velocity | 1869–1970 | 102 | 0.91 | 0.81 | 0.72 | 0.65 | 0.59 | 0.56 |
| Bond yield | 1900–1970 | 71 | 0.85 | 0.73 | 0.62 | 0.55 | 0.49 | 0.43 |
| Common stock prices | 1871–1970 | 100 | 0.90 | 0.76 | 0.64 | 0.53 | 0.46 | 0.43 |

[a]The data are residuals from linear least squares regression of the logs of the series (except the bond yield) on time. See footnote for table 3.

[b]Approximate expected sample autocorrelations based on Nelson and Kang (1981).

# Nelson and Plosser

**Table 5**

Tests for autoregressive unit roots[a]

$$z_t = \hat{\mu} + \hat{\gamma}t + \hat{\rho}_1 z_{t-1} + \hat{\rho}_2(z_{t-1} - z_{t-2}) + \cdots + \hat{\rho}_k(z_{t-k+1} - z_{t-k}) + \hat{\mu}_t.$$

| Series | $T$ | $k$ | $\hat{\mu}$ | $t(\hat{\mu})$ | $\hat{\gamma}$ | $t(\hat{\gamma})$ | $\hat{\rho}_1$ | $\tau(\hat{\rho}_1)$ | $s(\hat{u})$ | $r_1$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Real GNP | 62 | 2 | 0.819 | 3.03 | 0.006 | 3.03 | 0.825 | −2.99 | 0.058 | −0.02 |
| Nominal GNP | 62 | 2 | 1.06 | 2.37 | 0.006 | 2.34 | 0.899 | −2.32 | 0.087 | 0.03 |
| Real per capita GNP | 62 | 2 | 1.28 | 3.05 | 0.004 | 3.01 | 0.818 | −3.04 | 0.059 | −0.02 |
| Industrial production | 111 | 6 | 0.103 | 4.32 | 0.007 | 2.44 | 0.835 | −2.53 | 0.097 | 0.03 |
| Employment | 81 | 3 | 1.42 | 2.68 | 0.002 | 2.54 | 0.861 | −2.66 | 0.035 | 0.10 |
| Unemployment rate | 81 | 4 | 0.513 | 2.81 | −0.000 | −0.23 | 0.706 | −3.55* | 0.407 | 0.02 |
| GNP deflator | 82 | 2 | 0.260 | 2.55 | 0.002 | 2.65 | 0.915 | −2.52 | 0.046 | −0.03 |
| Consumer prices | 111 | 4 | 0.090 | 1.76 | 0.001 | 2.84 | 0.986 | −1.97 | 0.042 | −0.06 |
| Wages | 71 | 3 | 0.566 | 2.30 | 0.004 | 2.30 | 0.910 | −2.09 | 0.060 | 0.00 |
| Real wages | 71 | 2 | 0.487 | 3.10 | 0.004 | 3.14 | 0.831 | −3.04 | 0.034 | −0.01 |
| Money stock | 82 | 2 | 0.133 | 3.52 | 0.005 | 3.03 | 0.916 | −3.08 | 0.047 | 0.03 |
| Velocity | 102 | 1 | 0.052 | 0.99 | −0.000 | −0.65 | 0.941 | −1.66 | 0.067 | 0.11 |
| Interest rate | 71 | 3 | −0.186 | −0.95 | 0.003 | 1.75 | 1.03 | 0.686 | 0.283 | −0.02 |
| Common stock prices | 100 | 3 | 0.481 | 2.02 | 0.003 | 2.37 | 0.913 | −2.05 | 0.158 | 0.20 |

[a] $z_t$ represents the natural logs of annual data except for the bond yield. $t(\hat{\mu})$ and $t(\hat{\gamma})$ are the ratios of the OLS estimates of $\mu$ and $\gamma$ to their respective standard errors. $\tau(\hat{\rho}_1)$ is the ratio of $\hat{\rho}_1 - 1$ to its standard error. $s(\hat{u})$ is the standard error of the regression and $r_1$ is the first-order autocorrelation coefficient of the residuals. The values of $\tau(\hat{\rho}_1)$ denoted by an (*) are smaller than the 0.05 one tail critical value of the distribution of $\tau(\hat{\rho}_1)$ and similarly for $\hat{\rho}_1$. It should also be noted that $t(\hat{\mu})$ and $t(\hat{\gamma})$ are not distributed as normal random variables.