The Unit Root Land

Jesús Gonzalo

Universidad Carlos III de Madrid

Econometrics III (Master in Economic Analysis)

Material used and recommended

- Theodore Caplow,Louis Hicks,Ben J. Wattenberg (2001)
  The First Measured Century: An Illustrated Guide to Trends in America, 1900-2000.
- Hamilton (1994)
  Time Series Analysis (Chapters: 15, 16 and 17).
- Fuller (1996)
  Introduction to Statistical time series (Chapter 10).
- Billingsley (1999)
  Convergence of Probability Measure
- Mikosch (2000)
  Elementary Stochastic Calculus (Chapters 1 and 2).
- Phillips
  Lecture Notes

Outline

* Reasons to study UNIT ROOT MODELS
    - Growth
    - The effect of a Shock
    - Trend-Cycle decomposition
    - Forecasting
* Asymptotic Results
    - Brownian Motion
    - Functional Central Limite Theorem
* Testing for a Unit Root
    - Dickey-Fuller test: How to do it properly.
    - Other tests
    - Problems of testing for UR.

Growth

To represent growth, we consider mainly two models:

- Trend stationary (TS)

$$X_t = \alpha + \beta t + \epsilon_t$$

  where ϵ_t is $I(0)$ (A process Y_t is $I(d)$ if after $\triangle^d$ becomes stationary in the sense of having a Wold decomposition).

- Difference stationary (DS)

$$X_t = 1X_{t-1} + \beta + \epsilon_t$$

  where ϵ_t is $I(0)$. Note that X_t is $I(1)$

$$\Delta X_t = \beta + \epsilon_t$$

  is $I(0)$. Moreover, with some algebra we can write

$$X_t = X_0 + \beta t + \sum_{j=1}^{t} \epsilon_j$$

  where if $\epsilon_t \sim WN$, then X_t is a random walk with drift β.

The Effect of a Shock

We want to analyse the effect of a instantaneous perturbation at time t of ϵ_t in the outcome $t+h$ periods ahead, X_{t+h}:
$IRF = E(X_{t+h} \mid I_{t-1}, \epsilon_t = \delta) - E(X_{t+h} \mid I_{t-1}, \epsilon_t = 0)$.

$$\text{If } \begin{cases} \dfrac{\partial X_{t+h}}{\partial \epsilon_t} = 0 \text{ as } h \to \infty \text{ the shock is transitory.} \\ \dfrac{\partial X_{t+h}}{\partial \epsilon_t} \neq 0 \text{ as } h \to \infty \text{ the shock is permanent.} \end{cases}$$

Example 1 (AR)

Consider $X_t = \rho X_{t-1} + \epsilon_t$ with $|\rho| < 1$. Inverting into MA we obtain

$$X_t = \sum_{j=0}^{\infty} \rho^j \epsilon_{t-j} = \epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} + \cdots$$

Therefore

$$X_{t+h} = \epsilon_{t+h} + \rho \epsilon_{t+h-1} + \cdots + \rho^h \epsilon_t \Rightarrow \frac{\partial X_{t+h}}{\partial \epsilon_t} = \rho^h \to 0 \text{ as } h \to \infty$$

the shock is transitory.

Example 2 (Random Walk)

Consider $X_t = 1X_{t-1} + \epsilon_t$. As before

$$X_{t+h} = \epsilon_{t+h} + \epsilon_{t+h-1} + \cdots + \epsilon_t \Rightarrow \frac{\partial X_{t+h}}{\partial \epsilon_t} = 1 \not\to 0 \text{ as } h \to \infty$$

the shock is permanent.

Example 3 (I(1) process)

Consider $\Delta X_t = C(L)\epsilon_t = C(1)\epsilon_t + (1-L)\tilde{C}(L)\epsilon_t$ or, equally,
$X_t = (1-L)^{-1}C(1)\epsilon_t + \tilde{C}(L)\epsilon_t$ Hence,

$$X_{t+h} = C(1)\frac{\epsilon_{t+h}}{1-L} + \widetilde{C}(L)\epsilon_{t+h}$$

Therefore, under certain assumptions

$$\frac{\partial X_{t+h}}{\partial \epsilon_t} = C(1) \neq 0 \text{ if } X_t \text{ has a unit root}$$

Example 4

Think on $X_t = a + b_t + u_t$ where $u_t = \Psi(L)\epsilon_t$ is stationary.

Trend-Cycle Decomposition

Consider the difference stationary model

$$\Delta X_t = \mu + C(L)\epsilon_t$$
$$\Delta X_t = \mu + C(1)\epsilon_t + (1 - L)\widetilde{C}(L)\epsilon_t$$
$$X_t = \mu t + C(1)\frac{\epsilon_t}{\Delta} + \widetilde{C}(L)\epsilon_t$$

We denote:

- μt: Determinist trend,
- $C(1)\epsilon_t\Delta^{-1}$: Stochastic Trend (or better Permanent Component),
- $\widetilde{C}(L)\epsilon_t$: Cycle (or better Transitory Component).

Remark

Remember the conditions on the polynomial $C(L)$ in order for $\widetilde{C}(L)$ to behave correctly.

Consider the trend stationary model

$$X_t = a + bt + C(L)\epsilon_t$$

with $\sum_{j=1}^{\infty} |C_j| < \infty$. Then

$$X_{t+h} = a + b(t+h) + \epsilon_{t+h} + c_1\epsilon_{t+h-1} + \cdots + c_h\epsilon_t + c_{h+1}\epsilon_{t-1} + \cdots$$

where taking conditional expectations with respect to the information at time t, $\mathcal{I}_t$ we obtain

$$\mathbb{E}\left[X_{t+h}|\mathcal{I}_t\right] = a + b(t+h) + C_h\epsilon_t + C_{h+1}\epsilon_{t-1} + \cdots$$

Note that

$$\mathbb{E}\left[X_{t+h} - \mathbb{E}[X_{t+h}|\mathcal{I}_t]\right]^2 \to c \text{ as } h \to \infty$$

where c is a constant.

Now consider the difference stationary model

$$X_t = \mu + X_{t-1} + C(L)\epsilon_t$$

Then

$$\begin{aligned} X_{t+h} = h\mu + X_t + \epsilon_{t+h} + (1 + c_1)\epsilon_{t+h-1} + \cdots + \\ + \cdots + (1 + c_1 + \cdots + c_{h-1})\epsilon_{t+1} + \cdots \end{aligned}$$

where, again, taking conditional expectations with respect to $\mathcal{I}_t$ we obtain

$$\mathbb{E}\left[X_{t+h}|\mathcal{I}_t\right] = h\mu + X_t + [1 + c_1 + \cdots + c_h]\epsilon_t + [\cdots]\epsilon_{t-1}$$

Note that now

$$\mathbb{E}\left[X_{t+h} - \mathbb{E}[X_{t+h}|\mathcal{I}_t]\right]^2 \to \infty \text{ as } h \to \infty$$

Asymptotic results

Consider the AR(1) model

$$X_t = \rho X_{t-1} + \epsilon_t$$

where $\epsilon_t \sim iid\ N(0, \sigma^2)$. We want to test whether $\rho = 1$. Recall that

- If $|\rho| < 1$ we know

$$\sqrt{T}(\hat{\rho}_T - \rho) \overset{d}{\sim} N(0, 1 - \rho^2)$$

$$\sqrt{T}(\hat{\rho}_T - 1) \xrightarrow{p} 0$$

- But if $\rho = 1$ what do we have?

$$\hat{\rho}_T - 1 = \frac{\sum_1^T X_{t-1}\epsilon_t}{\sum_1^T X_{t-1}^2}$$

Asymptotic results

Consider first the term in the numerator

$$\frac{1}{T}\sum_{t=1}^{T} X_{t-1}\epsilon_t$$

Note that if $\rho = 1$, from our model we have

$$X_t^2 = (X_{t-1} + \epsilon_t)^2 = X_{t-1}^2 + 2X_{t-1}\epsilon_t + \epsilon_t^2$$

where rearranging terms we can obtain

$$X_{t-1}\epsilon_t = \frac{1}{2}\left[X_t^2 - X_{t-1}^2 - \epsilon_t^2\right]$$

Summing over t and dividing by $\sigma^2 T$ gives

$$\frac{1}{\sigma^2 T}\sum_{t=1}^{T} X_{t-1}\epsilon_t = \frac{1}{2}\left(\frac{1}{\sigma\sqrt{T}}X_T\right)^2 - \frac{1}{2\sigma^2 T}\sum_{t=1}^{T}\epsilon_t^2$$

Asymptotic results

Since

$$\frac{1}{\sigma\sqrt{T}}X_T \sim N(0,1), \quad then(\frac{1}{\sigma\sqrt{T}}X_T^2 \overset{d}{\sim} \chi_1^2 \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \xrightarrow{p} \sigma^2$$

Then

$$\frac{1}{\sigma^2 T}\sum_{t=1}^{T}X_{t-1}\epsilon_t \overset{d}{\sim} \frac{1}{2}(\chi_1^2 - 1)$$

Now consider the term in the denominator

$$\frac{1}{T^2}\sum_{t=1}^{T}X_{t-1}^2$$

Note that

$$X_{t-1} \overset{d}{\sim} N(0,\sigma^2(t-1))$$

Asymptotic results

and

$$\mathbb{E}\left(\sum_{t=1}^{T} X_{t-1}^2\right) = \mathbb{E}\left[\sum_{t=1}^{T-1}\left(\sum_{i=1}^{t}\epsilon_i\right)^2\right] = \frac{1}{2}T(T-1)\sigma^2$$

$$\mathbb{V}\left(\sum_{1}^{T} X_{t-1}^2\right) = \frac{1}{3}T(T-1)(T^2-T+1)\sigma^4$$

Therefore

$$\frac{\sum_{t=1}^{T} X_{t-1}^2}{.5T(T-1)\sigma^2}\begin{cases}\mathbb{E}[\cdot] = 1\\ \mathbb{V}[\cdot] = \frac{4(T^2-T+1)}{3(T^2-T)} \to \frac{4}{3} \text{ as } t \to \infty\end{cases}$$

Then we have that

$$\frac{1}{T^2}\sum_{t=1}^{T} X_{t-1}^2$$

does not converge to a constant in mean square; BUT to a random variable.

Brownian Motion

Definition 1 (Brownian Motion)

B is a stochastic process $[B(t) : 0 \leq t < \infty]$ on a probabilistic space $(\Omega, \mathcal{F}, \mathcal{P})$ with properties:

1. $B(0, \omega) = 0$, $\forall \omega \in \Omega$
2. $B(., \omega)$ is continuous for each $\omega \in \Omega$
3. For $0 < t_1 < t_2 < .... < t_{n-1} < t_n$, the increments $B(t_1), B(t_2) - B(t_1), ..., B(t_n) - B(t_{n-1})$ are independent and normally distributed, with mean 0 and variances $t_1, t_2 - t_1, ..., t_n - t_{n-1}$

Definition 2

A Brownian motion is a gaussian process with $\mathbb{E}(B(t)) = 0$ and $\mathbb{C}(B(s), B(t)) = s \wedge t$.

For a graphical view see the applets on my web.

Let $\{u_t\}_{t=0}^{\infty}$ be a stochastic process such that:

1. $\mathbb{E}(u_t) = 0$, for all t,
2. $\sup \mathbb{E}|u_t|^{\beta} < \infty$, for some $\beta > 2$,
3. $\sigma^2 = \lim \mathbb{E}[T^{-1}S_T^2]$ exists and $\sigma^2 > 0$, where
$$S_T = \sum_{i=1}^{T} u_i$$
and σ^2 can also be written as $\sigma^2 = \sigma_u^2 + 2\lambda$ with $\sigma_u^2 = \mathbb{E}(u_i^2)$ and
$$\lambda = \sum_{j=2}^{\infty} \mathbb{E}(u_i, u_j)$$
4. u_t is strongly-mixing with mixing coefficients α_m such that
$$\sum_{1}^{\infty} \alpha_m^{1-2/\beta} < \infty$$

Let's consider

$$X_T = \frac{1}{\sqrt{T}} S_T = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} u_i$$

Note that

$$\frac{1}{\sqrt{T}} S_T \overset{d}{\sim} N(0, \sigma^2)$$

and

$$\frac{1}{\sqrt{[T/2]}} \sum_{i=1}^{[T/2]} u_i \overset{d}{\sim} N(0, \sigma^2)$$

where $[T/2]$ denotes the largest integer that is less or equal to $T/2$.

Now consider the following partial sum process

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

where $r \in [0,1]$. For any given realization, $X_T(r)$ is a step function in r, with

$$X_T(r) = \begin{cases} 0 & \text{for} \quad 0 \leq r < 1/T \\ \frac{u_1}{\sqrt{T}} & \text{for} \quad 1/T \leq r < 2/T \\ \frac{u_1 + u_2}{\sqrt{T}} & \text{for} \quad 2/T \leq r < 3/T \\ \cdots & \cdots \\ \frac{u_1 + u_2 ... + u_T}{\sqrt{T}} & \text{for} \quad r = 1 \end{cases}$$

Note that

- For $T = 1$

$$X_1(r) = \frac{1}{\sqrt{1}} \sum_{t=1}^{[1r]} u_t = \frac{1}{\sqrt{1}} u_0 = 0$$

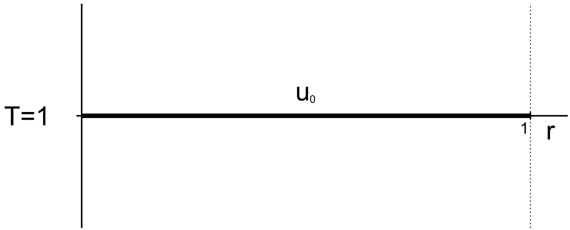


Figure: T=1

- For $T = 2$

$$X_2(r) = \frac{1}{\sqrt{2}} \sum_{t=1}^{[2r]} u_t = \begin{cases} \frac{1}{\sqrt{2}} u_0 & \text{if} \quad 0 \leq r < \frac{1}{2} \\ \frac{1}{\sqrt{2}} (u_0 + u_1) & \text{if} \quad \frac{1}{2} \leq r < 1 \end{cases}$$

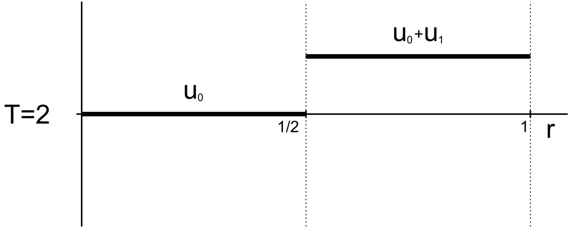


Figure: T=2

- For $T = 3$

$$X_3(r) = \frac{1}{\sqrt{3}} \sum_{t=1}^{[3r]} u_t = \begin{cases} \frac{1}{\sqrt{3}} u_0 & \text{if} \quad 0 \leq r < \frac{1}{3} \\ \frac{1}{\sqrt{3}} (u_0 + u_1) & \text{if} \quad \frac{1}{3} \leq r < \frac{2}{3} \\ \frac{1}{\sqrt{3}} (u_0 + u_1 + u_2) & \text{if} \quad \frac{2}{3} \leq r < 1 \end{cases}$$

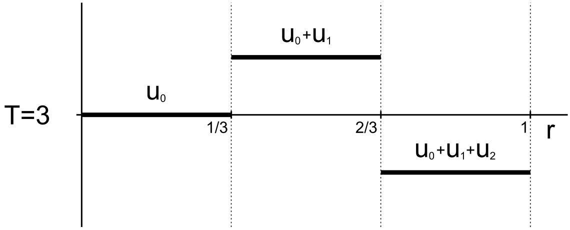


Figure: T=3

Then

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

where by a CLT we have that

$$\frac{1}{\sqrt{[Tr]}} \sum_{t=1}^{[Tr]} u_t \overset{d}{\sim} N(0, \sigma^2)$$

and

$$\frac{\sqrt{[Tr]}}{\sqrt{T}} \to \sqrt{r}$$

Therefore

$$X_T(r) \overset{d}{\sim} N(0, r\sigma^2) \quad \text{and} \quad \frac{X_T(r)}{\sigma} \overset{d}{\sim} N(0, r)$$

Now, consider $(r_2 > r_1)$. We have that

$$\begin{aligned}\mathbb{V}\left[X_T(r_2) - X_T(r_1)\right] &= \mathbb{V}\left[X_T(r_2)\right] + \mathbb{V}\left[X_T(r_1)\right] - 2\mathbb{C}\left[X_T(r_2), X_T(r_1)\right] \\ &= \sigma^2(r_2 + r_1 - 2r_1) = \sigma^2(r_2 - r_1)\end{aligned}$$

Hence

$$\frac{X_T(r_2) - X_T(r_1)}{\sigma} \overset{d}{\sim} N(0, r_2 - r_1)$$

Then by a Functional Central Limit Theorem (See Appendix for weak convergence that will be represented by $\Rightarrow$, fidi convergence + tighness), the process

$$\left\{\frac{X_T(\cdot)}{\sigma}\right\}_{t=1}^{\infty}$$

has an asymptotic probability law that is described by the Wiener process $W(\cdot)$

$$\frac{X_T(\cdot)}{\sigma} \Rightarrow W(\cdot) \quad \text{or} \quad X_T(\cdot) \Rightarrow \sigma^2 W(\cdot)$$

Application to Unit Root processes

Consider the unit root process

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

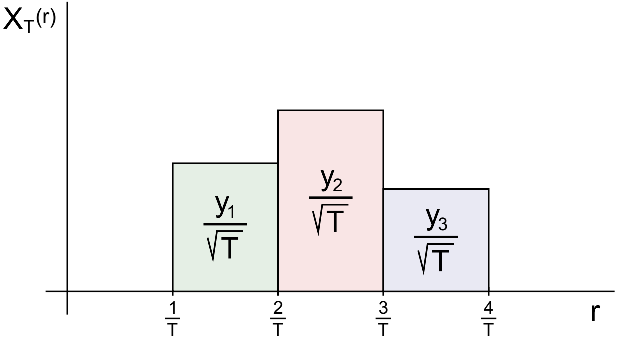
Which can also be expressed as

$$y_t = u_1 + u_2 + ... + u_t$$

Then we can define the partial sum process

$$X_T(r) = \begin{cases} 0 & \text{for} \quad 0 \leq r < \frac{1}{T} \\ \frac{y_1}{\sqrt{T}} = \frac{u_1}{\sqrt{T}} & \text{for} \quad \frac{1}{T} \leq r < \frac{2}{T} \\ \frac{y_2}{\sqrt{T}} = \frac{u_1+u_2}{\sqrt{T}} & \text{for} \quad \frac{2}{T} \leq r < \frac{4}{T} \\ \dots & \\ \frac{y_T}{\sqrt{T}} = \frac{u_1+\cdots+u_T}{\sqrt{T}} & \text{for} \quad r = 1 \end{cases}$$

Application to Unit Root processes



$$\int_0^1 X_T(r)dr = \frac{y_1}{T^{\frac{3}{2}}} + \frac{y_2}{T^{\frac{3}{2}}} + \dots + \frac{y_{T-1}}{T^{\frac{3}{2}}}$$

$$= T^{-\frac{3}{2}} \sum_{t=1}^{T} y_{t-1}$$

Application to Unit Root processes

By the Continuous Mapping Theorem we have that

$$\int_0^1 X_T(r)dr = T^{-\frac{3}{2}} \sum_{t=1}^{T} y_{t-1} \Rightarrow \sigma \int_0^1 W(r)dr$$

where

$$\int_0^1 W(r)dr$$

is the Integrated Brownian Motion.

$$\mathbb{E}\int_0^t W(s)ds = \int_0^t \mathbb{E}W(s)ds = 0$$

For $s \leq t$

$$\begin{aligned}
\mathbb{C}\left[\int_0^s W(y)dy \int_0^t W(u)du\right] &= \mathbb{E}\int_0^s \int_0^t W(y)W(u)dydu \\
&= \int_0^s \int_0^t \mathbb{E}W(y)W(u)dydu \\
&= \int_0^s \int_0^t min\{y, u\}dydu \\
&= \int_0^s \left(\int_0^u ydy + \int_u^t udu\right) du \\
&= s^2\left(\frac{t}{2} - \frac{s}{6}\right)
\end{aligned}$$

By the previous result, if $t = s = 1$, then

$$\mathbb{C}\left[\int_0^s W(y)dy \int_0^t W(u)du\right] = \frac{1}{3}$$

and then

$$\int_0^1 W(r)dr \equiv N\left(0, \frac{1}{3}\right).$$

Application to Unit Root processes

Other averages

$$
\begin{aligned}
T^{-\frac{3}{2}} \sum_{t=1}^{T} y_{t-1} &= T^{-\frac{3}{2}} \left[y_0 + y_1 + \cdots + y_{T-1} + y_T \right] \\
&= T^{-\frac{3}{2}} \left[u_1 + (u_1 + u_2) + \cdots + (u_1 + u_2 + \cdots + u_{T-1}) \right] \\
&= T^{-\frac{3}{2}} \left[(T-1)\, u_1 + \cdots + [T - (T-1)]\, u_{T-1} \right] \\
&= T^{-\frac{3}{2}} \sum_{t=1}^{T} (T - t)\, u_t \\
&= T^{-\frac{1}{2}} \sum_{t=1}^{T} u_t - T^{-\frac{3}{2}} \sum_{t=1}^{T} t u_t .
\end{aligned}
$$

Therefore

$$T^{-\frac{3}{2}}\sum_{t=1}^{T} tu_t = T^{-\frac{1}{2}}\sum_{t=1}^{T} u_t - T^{-\frac{3}{2}}\sum_{t=1}^{T} y_{t-1}$$

$$\Rightarrow \sigma W(1) - \sigma \int_0^1 W(r)dr \equiv N\left(0, \frac{\sigma^2}{3}\right)$$

because

$$\mathbb{C}\left[\int_0^t W(r)dr, W(t)\right] = \frac{t^2}{2}.$$

Consider now

$$\sum_{t=1}^{T} y_{t-1}^2$$

Define

$$S_T(r) = [X_T(r)]^2 \,,$$

then

$$S_T(r) = \begin{cases} 0 & \text{for } 0 \leq r < \frac{1}{T} \\ \frac{y_1^2}{T} & \text{for } \frac{1}{T} \leq r < \frac{2}{T} \\ \frac{y_2^2}{T} & \text{for } \frac{2}{T} \leq r < \frac{3}{T} \\ \ldots & \\ \frac{y_T^2}{T} & \text{for } r = 1 \end{cases}$$

If follows that

$$\int_0^1 (X_T(r))^2\, dr = \frac{y_1^2}{T^2} + \frac{y_2^2}{T^2} + ... + \frac{y_{T-1}^2}{T^2}$$
$$\Rightarrow \sigma^2 \int_0^1 (W(r))^2\, dr$$

Hence

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

$$T^{-\frac{5}{2}}\sum_{t=1}^{T} t y_{t-1} = T^{-\frac{3}{2}}\sum_{t=1}^{T}\underbrace{\left(\frac{t}{T}\right)}_{r} y_{t-1} \Rightarrow \sigma\int_0^1 rW(r)dr$$

$$T^{-3}\sum_{t=1}^{T} t y_{t-1}^2 = T^{-2}\sum_{t=1}^{T}\left(\frac{t}{T}\right) y_{t-1}^2 \Rightarrow \sigma^2\int_0^1 r\left(W(r)\right)^2 dr$$

Application to Unit Root processes

Other useful results

An important sum where we can not apply a FCLT and then the continuous mapping theorem as with the previous ones is

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T} y_{t-1}u_t &= \frac{1}{2T}y_T^2 - \frac{1}{2T}\left(u_1^2 + u_2^2 + ... + u_T^2\right) \\
&= \left(\frac{1}{2}\right) X_T^2(1) - \left(\frac{1}{2}\right)\left(\frac{1}{T}\right)\left(u_1^2 + u_2^2 + ... + u_T^2\right)
\end{aligned}
$$

Since $X_T^2(1) \Rightarrow \sigma^2 \left[W(1)\right]^2$ and by LLN $\frac{1}{T}\left(u_1^2 + u_2^2 + ... + u_T^2\right) \to \sigma^2$ then

$$
T^{-1}\sum_{t=1}^{T} y_{t-1}u_t \Rightarrow \frac{1}{2}\sigma^2\left[\left[W(1)\right]^2 - 1\right].
$$

Notice that when u_t is not $i.i.d$, then $\sigma^2 = \sigma_u^2 + \lambda$.

To obtain the previous result note that on the one hand we have

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{T}\left[u_1u_2 + (u_1+u_2)\,u_3 + ... + (u_1 + ... + u_{T-1})\,u_T\right]$$

On the other hand

$$\begin{aligned}
\frac{1}{T}y_T^2 &= \frac{1}{T}\left(u_1 + u_2 + ... + u_T\right)^2 \\
&= \frac{1}{T}\left[\left(u_1^2 + ... + u_T^2\right) + 2\,(u_1u_2) + \cdots \right. \\
&\qquad \left. \cdots + 2\,(u_1u_T) + ... + 2\,(u_{T-1}u_T)\right] \\
&= \frac{1}{T}\left(u_1^2 + u_2^2 + ... + u_T^2\right) + \frac{2}{T}\sum_{t=1}^{T} y_{t-1}u_t
\end{aligned}$$

Re-arranging we obtain the desired result.

Consider

$$X_t = X_{t-1} + u_t$$

where now $u_t \sim iid$ with $\mathbb{E}[u_t] = 0$ and $\mathbb{V}[u_t] = \sigma^2$.
Then we have

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} u_t \Rightarrow \sigma W(1)$$

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} X_{t-1} u_t \Rightarrow \frac{1}{2}\sigma^2 \left[[W(1)]^2 - 1 \right] = \int_0^1 W dW$$

$$T^{-\frac{3}{2}} \sum_{t=1}^{T} t u_t \Rightarrow \sigma W(1) - \sigma \int_0^1 W(r) dr \equiv N\left(0, \frac{\sigma^2}{3}\right)$$

$$T^{-\frac{3}{2}} \sum_{t=1}^{T} X_{t-1} \Rightarrow \sigma \int_0^1 W(r)dr$$

$$T^{-2} \sum_{t=1}^{T} X_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 (W(r))^2 \, dr$$

$$T^{-\frac{5}{2}} \sum_{t=1}^{T} tX_{t-1} \Rightarrow \sigma \int_0^1 rW(r)dr$$

$$T^{-3} \sum_{t=1}^{T} tX_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 r\,(W(r))^2 \, dr$$

$$T^{-(r+1)} \sum_{t=1}^{T} t^r \xrightarrow{p} \frac{1}{r+1} \text{ for } r = 0, 1, \dots$$

Unit Root Distribution Tables

Figure: $T(\hat{\alpha} - 1)$

Table 8.5.1. Empirical cumulative distribution of $n(\hat{\rho} - 1)$ for $\rho = 1$

Sample Size	Probability of a Smaller Value							
n	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
				$\hat{\rho}$				
25	−11.9	−9.3	−7.3	−5.3	1.01	1.40	1.79	2.28
50	−12.9	−9.9	−7.7	−5.5	0.97	1.35	1.70	2.16
100	−13.3	−10.2	−7.9	−5.6	0.95	1.31	1.65	2.09
250	−13.6	−10.3	−8.0	−5.7	0.93	1.28	1.62	2.04
500	−13.7	−10.4	−8.0	−5.7	0.93	1.28	1.61	2.04
∞	−13.8	−10.5	−8.1	−5.7	0.93	1.28	1.60	2.03
				$\hat{\rho}_\mu$				
25	−17.2	−14.6	−12.5	−10.2	−0.76	0.01	0.65	1.40
50	−18.9	−15.7	−13.3	−10.7	−0.81	−0.07	0.53	1.22
100	−19.8	−16.3	−13.7	−11.0	−0.83	−0.10	0.47	1.14
250	−20.3	−16.6	−14.0	−11.2	−0.84	−0.12	0.43	1.09
500	−20.5	−16.8	−14.0	−11.2	−0.84	−0.13	0.42	1.06
∞	−20.7	−16.9	−14.1	−11.3	−0.85	−0.13	0.41	1.04
				$\hat{\rho}_\tau$				
25	−22.5	−19.9	−17.9	−15.6	−3.66	−2.51	−1.53	−0.43
50	−25.7	−22.4	−19.8	−16.8	−3.71	−2.60	−1.66	−0.65
100	−27.4	−23.6	−20.7	−17.5	−3.74	−2.62	−1.73	−0.75
250	−28.4	−24.4	−21.3	−18.0	−3.75	−2.64	−1.78	−0.82
500	−28.9	−24.8	−21.5	−18.1	−3.76	−2.65	−1.78	−0.84
∞	−29.5	−25.1	−21.8	−18.3	−3.77	−2.66	−1.79	−0.87

NOTE. This table was constructed by David A. Dickey using the Monte Carlo method. Details are given in Dickey (1975). Standard errors of the estimates vary, but most are less than 0.15 for entries in the left half of the table and less than 0.03 for entries in the right half of the table.

Source: Fuller (1976).

Figure: $\hat{\tau}$

Table 8.5.2. Empirical cumulative distribution of $\hat{\tau}$ for $\rho = 1$

Sample Size	Probability of a Smaller Value							
n	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
				$\hat{\tau}$				
25	−2.66	−2.26	−1.95	−1.60	0.92	1.33	1.70	2.16
50	−2.62	−2.25	−1.95	−1.61	0.91	1.31	1.66	2.08
100	−2.60	−2.24	−1.95	−1.61	0.90	1.29	1.64	2.03
250	−2.58	−2.23	−1.95	−1.62	0.89	1.29	1.63	2.01
500	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	2.00
∞	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	2.00
				$\hat{\tau}_\mu$				
25	−3.75	−3.33	−3.00	−2.63	−0.37	0.00	0.34	0.72
50	−3.58	−3.22	−2.93	−2.60	−0.40	−0.03	0.29	0.66
100	−3.51	−3.17	−2.89	−2.58	−0.42	−0.05	0.26	0.63
250	−3.46	−3.14	−2.88	−2.57	−0.42	−0.06	0.24	0.62
500	−3.44	−3.13	−2.87	−2.57	−0.43	−0.07	0.24	0.61
∞	−3.43	−3.12	−2.86	−2.57	−0.44	−0.07	0.23	0.60
				$\hat{\tau}_\tau$				
25	−4.38	−3.95	−3.60	−3.24	−1.14	−0.80	−0.50	−0.15
50	−4.15	−3.80	−3.50	−3.18	−1.19	−0.87	−0.58	−0.24
100	−4.04	−3.73	−3.45	−3.15	−1.22	−0.90	−0.62	−0.28
250	−3.99	−3.69	−3.43	−3.13	−1.23	−0.92	−0.64	−0.31
500	−3.98	−3.68	−3.42	−3.13	−1.24	−0.93	−0.65	−0.32
∞	−3.96	−3.66	−3.41	−3.12	−1.25	−0.94	−0.66	−0.33

This table was constructed by David A. Dickey using the Monte Carlo method. Details are given in Dickey (1975). Standard errors of the estimates vary, but most are less than 0.02.

Dickey-Fuller Tests for Unit Root

- Case (a)
  1. DGP $X_t = X_{t-1} + u_t$
  2. Regression $X_t = \rho X_{t-1} + u_t$
- Case (b)
  1. DGP $X_t = X_{t-1} + u_t$
  2. Regression $X_t = a + \rho X_{t-1} + u_t$
- Case (b')
  1. DGP $X_t = a + X_{t-1} + u_t$
  2. Regression $X_t = a + \rho X_{t-1} + u_t$
- Case (c)
  1. DGP $X_t = X_{t-1} + u_t$
  2. Regression $X_t = a + bt + \rho X_{t-1} + u_t$
- Case (c')
  1. DGP $X_t = a + X_{t-1} + u_t$
  2. Regression $X_t = a + bt + \rho X_{t-1} + u_t$

Dickey-Fuller Tests for Unit Root

Case (a)

DGP is the Random Walk given by

$$X_t = X_{t-1} + u_t$$

Regression:

$$X_t = \rho X_{t-1} + u_t$$

where $u_t \sim iid$ with $\mathbb{E}[u_t] = 0$ and $\mathbb{V}[u_t] = \sigma^2$. The OLS estimate of ρ is given by

$$\hat{\rho}_T = \frac{\sum_{t=1}^{T} X_{t-1} X_t}{\sum_{t=1}^{T} X_{t-1}^2} = \rho + T^{-1} \frac{T^{-1} \sum_{t=1}^{T} X_{t-1} u_t}{T^{-2} \sum_{t=1}^{T} X_{t-1}^2}$$

Therefore

$$T\left(\hat{\rho}_T - \rho\right) = \frac{T^{-1} \sum_{t=1}^{T} X_{t-1} u_t}{T^{-2} \sum_{t=1}^{T} X_{t-1}^2}$$

By previous results we know

$$T^{-1}\sum_{t=1}^{T} X_{t-1}u_t \Rightarrow \frac{1}{T}\sigma^2\left[[W(1)]^2-1\right]$$

and

$$T^{-2}\sum_{t=1}^{T} X_{t-1}^2 \Rightarrow \sigma^2\int_0^1 (W(1))^2 dr$$

Therefore

$$T\left(\hat{\rho}_T-1\right) \Rightarrow \frac{\frac{1}{2}\left[[W(1)]^2-1\right]}{\int_0^1 \left(W(r)\right)^2 dr}$$

Remember that when $|\rho|<1$

$$\sqrt{T}\left(\hat{\rho}-\rho\right) \rightarrow N\left(0,1-\rho^2\right)$$

The Dickey-Fuller statistic is

$$t_{\rho=1} = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} = \frac{\hat{\rho}_T - 1}{\left[\frac{S_T^2}{\sum_{t=1}^{T} X_{t-1}^2}\right]^{1/2}}$$

where

$$S_T^2 = \frac{1}{T-1} \sum_{t=1}^{T} \left(X_t - \hat{\rho}_T X_{t-1}\right)^2 .$$

To find the Asymptotic Distribution of $t_{\rho=1}$ we need to rewrite it as:

$$t_{\rho=1} = T\left(\hat{\rho}_T - 1\right)\left[T^{-2}\sum_{t=1}^{T} X_{t-1}^2\right]^{\frac{1}{2}} \frac{1}{\left[S_T^2\right]^{\frac{1}{2}}}$$

$$= \frac{T^{-1}\sum_{t=1}^{T} X_{t-1}u_t}{\left[T^{-2}\sum_{t=1}^{T} X_{t-1}^2\right]^{\frac{1}{2}}\left[S_T^2\right]^{\frac{1}{2}}}$$

$$\Rightarrow \frac{\frac{1}{2}\sigma^2\left[[W(1)]^2 - 1\right]}{[\sigma^2\int_0^1[W(r)]^2dr]^{\frac{1}{2}}[\sigma^2]^{\frac{1}{2}}} = \frac{\frac{1}{2}\left[[W(1)]^2 - 1\right]}{[\int_0^1[W(r)]^2dr]^{\frac{1}{2}}}$$

Since $S_T^2 \xrightarrow{p} \sigma^2$ by consistency of $\hat{\rho}_T$.

DGP is given by

$$X_t = \alpha + X_{t-1} + u_t$$

Regression:

$$X_t = \alpha + \rho X_{t-1} + u_t$$

where $u_t \sim iid$ with $\mathbb{E}[u_t] = 0$ and $\mathbb{V}[u_t] = \sigma^2$. We can rewrite X_t as

$$X_t = X_0 + \alpha t + (u_1 + u_2 + \ldots + u_t) = X_0 + \alpha t + \xi_t$$

Now note that

$$\frac{1}{T}\sum_{t=1}^{T} X_{t-1} = \frac{1}{T}\sum_{t=1}^{T} \left[X_0 + \alpha\,(t-1) + \xi_{t-1}\right]$$

We have that

$$T^{-1}\sum_{t=1}^{T} X_0 \to X_0$$

Dickey-Fuller Tests for Unit Root

Case (b')

Analogously

$$T^{-2}\sum_{i=1}^{T}\alpha(t-1)\rightarrow\frac{\alpha}{2}$$

and

$$T^{-\frac{3}{2}}\sum_{i=1}^{T}\xi_{t-1}\Rightarrow\sigma\int_0^1 W(r)d$$

Therefore need to divide $\sum_{t=1}^{T}X_{t-1}$ by T^2 to get something that 'makes sense'. Note that doing that we obtain

$$\begin{aligned}\frac{1}{T^2}\sum_{t=1}^{T}X_{t-1} &= \frac{1}{T^2}\sum_{t=1}^{T}\left[X_0+\alpha\left(t-1\right)+\xi_{t-1}\right]\\ &\rightarrow 0+\frac{\alpha}{2}+0=\frac{\alpha}{2}\end{aligned}$$

Similarly

$$\sum_{t=1}^{T} X_{t-1}^2 = \underbrace{\sum_{t=1}^{T} X_0^2}_{Op(T)} + \underbrace{\sum_{t=1}^{T} \alpha^2 (t-1)^2}_{Op(T^3)} + \underbrace{\sum_{t=1}^{T} \xi_{t-1}^2}_{Op(T^2)} +$$

$$+ \underbrace{2\sum_{t=1}^{T} X_0 \alpha(t-1)}_{Op(T^2)} + \underbrace{2\sum_{t=1}^{T} X_0 \xi_{t-1}}_{Op(T^{\frac{3}{2}})} + \underbrace{2\sum_{t=1}^{T} \alpha(t-1)\xi_{t-1}}_{Op(T^{\frac{5}{2}})}$$

Then

$$T^{-3} \sum_{t=1}^{T} X_{t-1}^2 \to \frac{\alpha^2}{3}$$

Finally

$$\sum_{t=1}^{T} X_{t-1} u_t = \underbrace{X_0 \sum_{t=1}^{T} u_t}_{Op(T^{\frac{1}{2}})} + \underbrace{\sum_{t=1}^{T} \alpha(t-1)}_{Op(T^{\frac{3}{2}})} + \underbrace{\sum_{t=1}^{T} \xi_{t-1} u_t}_{Op(T)}$$

So

$$T^{-\frac{3}{2}} \sum_{t=1}^{T} X_{t-1} u_t \to T^{-\frac{3}{2}} \sum_{t=1}^{T} \alpha(t-1) u_t$$

Putting everything together we obtain

$$\begin{bmatrix} T^{\frac{1}{2}}(\hat{\alpha}_T - \alpha) \\ T^{\frac{3}{2}}(\hat{\rho}_T - 1) \end{bmatrix} \rightarrow N(0, Q^{-1}\sigma^2)$$

where

$$Q = \begin{bmatrix} 1 & 2^{-1}\alpha \\ 2^{-1}\alpha & 3^{-1}\alpha^2 \end{bmatrix}$$

Then what about $t_{\rho=1}$?

Dickey-Fuller Tests for Unit Root

Case (c) and (c')

DGP is given by

$$X_t = X_{t-1} + u_t$$

Regression:

$$X_t = \alpha + bt + \rho X_{t-1} + u_t$$

where $u_t \sim iid$ with $\mathbb{E}[u_t] = 0$ and $\mathbb{V}[u_t] = \sigma^2$.

$$t_{\rho=1} \rightarrow \text{D.F. distribution that does not depend on } \alpha \text{ or } \sigma$$

THINK ON A STRATEGY TO TEST FOR UNIT ROOTS (use the Eviews applets of my web page)

Let

$$X_t = X_{t-1} + u_t$$

where $u_t = \Psi(L)e_t$ with $e_t \sim iid$. In the model

$$X_t = \rho X_{t-1} + u_t$$

we want to test

$$H_0 : \rho = 1 \qquad H_1 : |\rho| < 1$$

which is equivalent to rewriting the model as

$$\Delta X_t = (\rho - 1)X_{t-1} + u_t$$

and testing

$$H_0 : (\rho - 1) = 0 \qquad H_1 : (\rho - 1) < 0$$

Dickey-Fuller Tests for Unit Root

What if u_t is correlated?

Note that

$$\psi(L)^{-1} = A(L) = A(1) + (1-L)\tilde{A}(L)$$

with $A(0) = 1$. Therefore we have

$$\begin{aligned} \Delta X_t &= (\rho - 1)X_{t-1} + \psi(L)e_t \\ \psi(L)^{-1}\Delta X_t &= (\rho - 1)\psi(L)^{-1}X_{t-1} + e_t \\ \Delta X_t &= (\rho - 1)A(1)X_{t-1} + (\rho - 1)\Delta A(L)X_{t-1} + \\ &\quad + (-A(L) + A(0))\Delta X_t + e_t \end{aligned}$$

Hence we can write

$$\Delta X_t = a^* X_{t-1} + \{\text{LAGS of } \Delta X_t\} + e_t$$

It can be proved that $t_{a*=0}$ has the same Asymptotic Distribution as $t_{\rho=1}$ in case (a).

Consider

$$
\begin{aligned}
X_t &= \rho X_{t-1} + u_t && (1)\\
X_t &= a + \rho X_{t-1} + u_t \\
X_t &= a + b_t + \rho X_{t-1} + u_t \\
X_t &= (\text{polynomial in t}) + \rho X_{t-1} + u_t && (2)
\end{aligned}
$$

Under the H_0 of (1) we have that

$$
t_{\rho=1} = \frac{(\hat{\rho}_{t-1})}{\hat{\sigma}_{\hat{\rho}_t}} \Rightarrow \frac{\int_0^1 W(r)dW(r)}{\left(\int_0^1 W^2(r)\right)^{1/2}}
$$

In general under the H_0 of (2) we have

$$t_{\rho=1} = \frac{(\hat{\rho}_{t-1})}{\hat{\sigma}_{\hat{\rho}_t}} \Rightarrow \frac{\int_0^1 W_D(r)dW(r)}{\left(\int_0^1 W_D^2(r)^2 dr\right)^{1/2}}$$

where

$$W_D(r) = W(r) - \left[\int_0^1 WD'\right]\left[\int_0^1 DD'\right]^{-1} D(r)$$

is the Hilbert projection in $L_2[0,1]$ of W onto the space orthogonal to D, i.e.

$$D(r) = (1, r, ..., r^p)'$$

Phillips-Perron Tests

Consider

$$X_t = \rho X_{t-1} + u_t$$

with $u_t = C(L)\epsilon_t$. We now that

$$T(\hat{\rho} - 1) = \frac{\frac{1}{T}\sum_{t=1}^{T} X_{t-1}u_t}{\frac{1}{T^2}\sum_{t=1}^{T} X_{t-1}^2} \Rightarrow \frac{\left[\int B(r)dB(r) + \lambda\right]}{\left[\int B^2(r)dr\right]}$$

$$t_{\rho=1} = \frac{(\hat{\rho}_T - 1)}{\left(\hat{\sigma}_u^2\left[\sum_1^T X_{t-1}^2\right]^{-1}\right)^{1/2}} \Rightarrow \frac{\int_0^1 B(r)dB(r) + \lambda}{\sigma_u\left[\int_0^1 B^2(r)dr\right]^{1/2}}$$

Where $\sigma_u^2 = \mathbb{V}[u_t]$, $B(r)$ is a Brownian Motion with variance $w^2 = \sigma_\epsilon^2 C(1)^2$ and

$$\lambda = \sum_{j=1}^{\infty} \mathbb{E}(u_0 u_j)$$

hence

$$w^2 = \sigma_u^2 + 2\lambda$$

Phillips-Perron Tests

THINK ON WHAT DOES IT HAPPEN when $u_t = \epsilon_t$?

$$z_\alpha = T(\hat{\rho} - 1) - \frac{\hat{\lambda}}{\frac{1}{T^2}\sum_{t=1}^{T} X_{t-1}^2} \Rightarrow \frac{\int_0^1 W(r)dw}{\int W^2(r)dr}$$

$$z_t = \hat{\sigma}_u \hat{w}^{-1} t_{\rho=1} - \hat{\lambda}\left\{\hat{w}\left(T^{-2}\sum_{t=1}^{T} X_{t-1}^2\right)^{1/2}\right\}^{-1} \Rightarrow \frac{\int W(r)dW}{\left(\int W^2(r)dr\right)^{1/2}}$$

where

$$W(r) = \frac{1}{w}B(r).$$

Efficient Unit Root Tests

When there is no trend/intercept in the model, the Dickey Fuller Test is very close to being optimal, i.e. for local alternatives it comes close to having optimal power in relation to the best test computed for the known local alternative using Neyman-Pearson Lemma (this is the so-called power envelope) and is based on the likelihood ratio

$$\frac{L_{H_0}(\rho = 1)}{L_{H_1}(\rho = 1 + \frac{1}{T} \text{ given } c)}$$

However, when there is a trend in the model the Dickey Fuller Tests relies on trend removal by regression and it turns out that efficiency can be gained by 'improving' the trend removal process.

Efficient Unit Root Tests

Consider

$$y_t = \beta' D_t + X_t$$

where $D_t = (1, t, ...)$ are deterministic components and

$$X_t = \rho X_{t-1} + \epsilon_t$$

Under local alternatives

$$\rho = 1 + \frac{c}{T}$$

You could use OLS to estimate the deterministic components. This is what Dickey Fuller does. It makes sense to think that if we find a more efficient method to estimate β, we would get better power.

Efficient Unit Root Tests

Elliot, Rothember and Stock (1996, Econometrica) proposes to use GLS to estimate β under H_1, so for different values of 'c'

$$\widetilde{\beta} = (\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}'\widetilde{y}$$

where

$$\begin{aligned} \widetilde{D}_t &= D_t - \rho D_{t-1} = (1 - \rho L)D_t \\ \widetilde{y}_t &= y_t - \rho y_{t-1} = (1 - \rho L)y_t \end{aligned}$$

so

$$\begin{aligned} \widetilde{y}_t &= \widetilde{D}_t\beta + X_t \\ \widetilde{X}_t &= \widetilde{y}_t - \widetilde{D}_t\widetilde{\beta} \quad \text{(detrended data)} \end{aligned}$$

Now, we make a Unit Root test on $\widetilde{X}_t$. Note that:

- The Asymptotic Distribution of the Dickey Fuller on the GLS detrended data depends on 'c'.
- For $D = (1, t)$, Elliot, Rothember and Stock advise to use $c = 13.5$.

Additional Aspects of Unit Root Tests

It is commonly claimed that Unit Root Tests have serious

- Size problems (Phillips-Perron Tests)
- Power problems (Dickey-Fuller Tests)

My opinion

J.Gonzalo and T.Lee (1996). *No Lack of Relative Power of the DF Tests*. JTSA, 17, 37-47.

Why the null is unit root (non-stationary), instead of $|root| < 1$ (stationarity)? Think on terms of Type I errors.

To solve the size problems of the PP tests Perron and Ng (1996) [1] propose some modifications of the Pihllips Perron Test. Let

$$Mz_\alpha = z_\alpha + \frac{1}{2} T(\hat{\rho}_T - 1)^2$$

Under H_0, $M_{z\alpha} \overset{d}{\equiv} z_\alpha$. Then

- Define
$$MSB = \left(T^{-2} \frac{\sum_{t=1}^{T} X_{t-1}^2}{\hat{w}^2} \right)^{1/2}$$
- It can be checked that
$$z_t = MSB \cdot z_\alpha$$

[1] Perron and Ng (1996): *Useful Modifications to Some Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties*. RES, 63, 435-63.

Additional Aspects of Unit Root Tests

- This suggests a new modified Phillips-Perron Test

$$Mz_t = MSB \cdot Mz_\alpha$$

therefore

$$Mz_t = z_t + \frac{1}{2}\left(\frac{\sum_{t=1}^{T} X_{t-1}^2}{\hat{w}^2}\right)^{1/2} (\hat{\rho}_T - 1)^2$$

Under H_0,

$$Mz_t \overset{d}{\equiv} z_t$$

THINK ON HOW TO IMPLEMENT IT.

Testing Stationarity

We want to test the hypothesis

$$\begin{cases} H_0: & \text{no unit root} \\ H_1: & \text{unit root} \end{cases}$$

Let

$$y_t = \beta' D_t + X_t + v_t$$

with

$$X_t = X_{t-1} + u_t$$

Then the test is given by

$$\begin{cases} H_0 : \sigma_u^2 = 0 \\ H_1 : \sigma_u^2 > 0 \end{cases}$$

Testing Stationarity

Let $\hat{e}_t$ be the residuals from the regression of y_t on D_t and

$$\hat{\sigma}_v^2 = \frac{1}{T}\sum_1^T \hat{e}_t^2$$

Then the LM statistic can be constructed as

$$LM = \frac{\frac{1}{T^2}\sum_1^T \hat{S}_t^2}{\hat{\sigma}_v^2}$$

Where S_t is the partial sum process of the residuals $\sum_{j=1}^t \hat{e}_j^2$. Under H_0

$$LM \Rightarrow \int_0^1 V^2(r)dr$$

where $V_{(r)}$ is a Brownian Bridge $(B(r) - rB(1))$. When v_t is serially correlated $\hat{\sigma}_v^2$ must be substituted by a LRV estimator using $\hat{e}_t^2$.