Ph.D. Seminar Series in Advanced Mathematical Methods in Economics and Finance

UNIT ROOT DISTRIBUTION THEORY I

Roderick McCrorie

School of Economics and Finance
University of St Andrews
23 April 2009

- One important aspect of a time series is that it is one realization of a multidimensional random variable

- In this context, an assumption of second-order stationarity is convenient because it facilitates inference through laws of large numbers and central limit theorems in a classical way.

- Another aspect of the second-order stationarity assumption is that it permits a Wold decomposition whereby a time series can be represented as the sum of a linearly regular part involving an infinite-order weighted average of white noise, and a part that is perfectly linearly predictable [e.g. Hannan, 1970, *Multiple Time Series* (John Wiley), p. 137].

- This offered some justification for the then emerging Box-Jenkins methods, where ARIMA models selected on the basis of the data could be viewed as approximations to the regular part in the Wold decomposition

- These models produced satisfactory representations of many observed economic time series, at least for the purpose of prediction, but as the models were selected purely on the basis of the data, they lacked theoretical justification, as if they emerged from a “black box”

- The essential problem faced by the econometrics profession in the late 1960s and 1970s was that structural econometric models, embodying restrictions from economic theory, were often outperformed by the black box models.

- There was also the problem of the proliferation of data-based models for a given time series that were incompatible with each other from the point of view of economic theory

- In the background, there was a debate about whether or not observed aggregate time series, which were manifestly non-stationary, should have random walk or other trend components removed prior to estimation given the loss of long run information that such transformations imply.

- The introduction of the concepts of unit roots and co-integration in the context of non-stationary time series helped to resolve some of these issues.

- Two important papers, which occurred one after the other in an issue in *Econometrica* in 1987 are:-

- Engle, R. F. & C.W.J. Granger (1987) Co-integration and error correction: representation, estimation and testing. *Econometrica* 55, 251-276.

- Phillips, P.C.B. (1987) Time series regression with a unit root. Econometrica 55, 277-301.

- Formative influences include work by A.W. Phillips on continuous time dynamic disequilibrium trade cycle and cyclical growth models; a paper by Sargan (1964) on wages and prices presented at a Colston society conference; the paper by Davidson, Hendry, Srba and Yeo (*Econ. J.*, 1978), universally known as DHSY, on the consumption function; and papers by Granger (*J. Economet.*, 1981) and Dickey and Fuller (*JASA*, 1979; *Econometrica*, 1981)

- Stationary AR(1) model:

$$y_t = \rho y_{t-1} + u_t, \quad |\rho| < 1, \quad y_0 = 0, \quad u_t \sim \text{i.i.d.} N(0, \sigma^2). \quad (1)$$

- Ordinary least squares (OLS) estimator:

$$\widehat{\rho}_T = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} = \rho + \frac{\sum_{t=1}^{T} y_{t-1} u_t}{\sum_{t=1}^{T} y_{t-1}^2} \quad (2)$$

$$\longrightarrow \quad \sqrt{T}\,(\widehat{\rho}_T - \rho) = \frac{T^{-1/2} \sum_{t=1}^{T} y_{t-1} u_t}{T^{-1} \sum_{t=1}^{T} y_{t-1}^2}. \quad (3)$$

- It can be shown that, as $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{p} \gamma_0 = \frac{\sigma^2}{1-\rho^2}, \quad \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}u_t \xrightarrow{d} N(0, \sigma^2\gamma_0),$$

$$\longrightarrow \quad \sqrt{T}\,(\widehat{\rho}_T - \rho) \xrightarrow{d} N(0, \sigma^2/\gamma_0) = N\left(0, (1-\rho^2)\right). \quad (4)$$

- When $\rho = 1$ (4) implies that

$$\sqrt{T}\,(\widehat{\rho}_T - 1) \xrightarrow{p} 0.$$

- The distribution is degenerate (its variance, $1-\rho^2$, is zero).
- Not useful for making inferences when $\rho = 1$.

- The nonstationary AR(1) model (or random walk) is defined by:

$$y_t = \rho y_{t-1} + u_t, \quad \rho = 1, \quad y_0 = 0, \quad u_t \sim \text{i.i.d.} N(0, \sigma^2). \tag{5}$$

- By backward substitution it can be shown that

$$y_t = \sum_{j=1}^{t} u_j, \tag{6}$$

and hence $\text{var}(y_t) = t\sigma^2 \to \infty$ as $t \to \infty$.

Deriving the limiting distribution of the OLS estimator in this context is a non-standard statistical problem.

HOWEVER, it is still possible to derive the distribution of the numerator of the test statistic by standard methods.

5. (a) (i) Since $u_t \sim$ i.i.d. $N(0, \sigma^2)$,

$$y_t = u_t + u_{t-1} + \dots + u_1 \sim N(0, \sigma^2 t). \qquad \text{—— (1)}$$

Note also that

$$y_t^2 = (y_{t-1} + u_t)^2 = y_{t-1}^2 + 2y_{t-1}u_t + u_t^2,$$

and so

$$y_{t-1}u_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - u_t^2).$$

Hence

$$\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}(y_T^2 - y_0^2) - \frac{1}{2}\sum_{t=1}^{T} u_t^2.$$

Since $y_0 = 0$, we have

$$\left(\frac{1}{T}\right)\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}\left(\frac{1}{T}\right)y_T^2 - \frac{1}{2}\left(\frac{1}{T}\right)\sum_{t=1}^{T} u_t^2$$

$$\left(\frac{1}{\sigma^2 T}\right)\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}\left(\frac{y_T}{\sigma\sqrt{T}}\right)^2 - \frac{1}{2\sigma^2}\left(\frac{1}{T}\right)\sum_{t=1}^{T} u_t^2$$

—— (2)

(1) implies that $\left(\frac{y_T}{\sigma\sqrt{T}}\right) \sim N(0, 1)$

and so $\left(\frac{y_T}{\sigma\sqrt{T}}\right)^2 \sim \chi^2(1)$ —— (3)

Also, $\sum_{t=1}^{T} u_t^2$ is the sum of T i.i.d. r.v.'s each with <u>mean</u> σ^2 and so, by a law of large numbers,

$$\frac{1}{T} \sum_{t=1}^{T} u_t^2 \xrightarrow{p} \sigma^2 \qquad \text{———} (4)$$

(3) and (4) imply

$$\left(\frac{1}{\sigma^2 T}\right) \sum_{t=1}^{T} y_{t-1} u_t \xrightarrow{D} \frac{1}{2}(X-1),$$

where $X \sim \chi^2(1)$.

(ii) Since $y_t \sim N(0, \sigma^2 t)$, $y_{t-1} \sim N(0, \sigma^2(t-1))$, and so $E(y_{t-1}^2) = \sigma^2(t-1)$.

Now

$$E\left[\sum_{t=1}^{T} y_{t-1}^2\right] = \sigma^2 \sum_{t=1}^{T} (t-1)$$

$$= \sigma^2 \frac{1}{2} T(T+1) - \sigma^2 T$$

$$= \sigma^2 \frac{1}{2} T(T-1).$$

(iii) What the result in part (ii) suggests is that if $\sum_{t=1}^{T} y_{t-1}^2$ is to have a convergent distribution, we must divide by T^2.

To derive the limiting distribution of the statistic

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

we resort to functional central limit theory. The idea is to characterize the limiting distribution of random elements that live in *function spaces* like $C[0, 1]$ or $D[0, 1]$. We consider these because of the memory in the process: we need to consider its <u>whole trajectory</u> and not just its endpoint.

References

Phillips, P.C.B. (1987) Time series regression with a unit root, *Econometrica* 55, 277-301.

Phillips, P.C.B. (1986) Understanding spurious regressions in Econometrics, *Journal of Econometrics* 33, 322-340.

Phillips, P.C.B. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473-495.

Billingsley, P. (1968) *Convergence of Probability Measures*, Wiley.

Davidson, J. (1994) *Stochastic Limit Theory*, OUP.

Hamilton, J.D. (1994) *Time Series Analysis*, Princeton.

Tanaka, K. (1996) *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, Wiley.

LECTURE 7 — FUNCTIONAL CLT's

General background

We are concerned with the weak convergence of stochastic processes defined on a function space.

<u>Definition</u> A <u>metric space</u> is a pair (X, ρ) where X is a set and $\rho : X \times X \longrightarrow \mathbb{R}^+ = \{\alpha . \alpha \in \mathbb{R}, \alpha \geq 0\}$ is a fn such th

(i) $\rho(x,y) = \rho(y,x) \quad \forall x, y \in X$

(ii) $\rho(x,y) + \rho(y,z) \geq \rho(x,z) \quad \forall x, y, z \in X$

(iii) $\rho(x,y) = 0 \iff x = y$.

ρ is called a <u>metric</u> on X and abstracts the four basic properties of Euclidean distance: the distance between distinct points is strictly positive; the distance from a point to itself is zero; distance is symmetric; and the triangle inequality holds.

<u>Example</u> Let $X = C([0,1])$, the space of cts real-valued functio on $[0,1]$. Write $\rho_\infty(f,g) = \sup_{0 \leq \alpha \leq 1} |f(\alpha) - g(\alpha)|$ for $f, g \in X$. Then ρ_∞ is a metric on X.

~~Note that~~ $\cancel{\sup_{\alpha \in [0,1]} |f(\alpha) - g(\alpha)| \leq \sup_{\alpha \in [0,1]} |f(\alpha)}$

7. (a) ρ_∞ takes values in $[0, \infty[$ because $\{|f(\alpha) - g(\alpha)| : \alpha \in [0,1]\}$ is always a non-empty set, bounded below by 0, and also bounded above because f, g and $f-g$ are cts functions and $[0,1]$ is a closed bounded interval.

Now

(i) $\rho_\infty(f,f) = \sup_{\alpha\in[0,1]} 0 = 0$;

(ii) $$\rho_\infty(f,g) = \sup_{\alpha\in[0,1]} |f(\alpha) - g(\alpha)|$$
$$= \sup_{\alpha\in[0,1]} |g(\alpha) - f(\alpha)|$$
$$= \rho_\infty(g,f);$$

(iii) if $\alpha \in [0,1]$ then
$$|f(\alpha) - h(\alpha)| \leq |f(\alpha) - g(\alpha)| + |g(\alpha) - h(\alpha)|$$
$$\leq \rho_\infty(f,g) + \rho_\infty(g,h),$$
for all $\alpha \in [0,1]$, so $\rho_\infty(f,h) \leq \rho_\infty(f,g) + \rho_\infty(g$

(iv) if $\rho_\infty(f,g) = 0$ then $|f(\alpha) - g(\alpha)| \leq 0$ for every $\alpha \in [0,1]$

7. (a) (iv) cont. so $f(\alpha) = g(\alpha)$ for every $\alpha \in [0,1]$ and $f = g$.

Thus ρ_∞ is a metric. 4

(b) Let (X, ρ) be a metric space.

(i) A sequence $\langle x_n \rangle_{n \in \mathbb{N}}$ in X is convergent to $x \in X$ if $\rho(x_n, x) \to 0$ as $n \to \infty$;

(ii) A sequence $\langle x_n \rangle_{n \in \mathbb{N}}$ in X is a Cauchy sequence if $\rho(x_n, x_m) \to 0$ as $m, n \to \infty$;

(iii) X is complete if every Cauchy sequence in X is convergent.

(iv) (X, ρ) is separable if either $X = \emptyset$ or there is a sequence $\langle x_n \rangle_{n \in \mathbb{N}}$ in X such that $X = \overline{\{x_n : n \in \mathbb{N}\}}$ (i.e. it contains a countable dense subset.)

3

(c) When discussing convergence in distribution, completeness is important to avoid probability mass escaping from the space as $n \to \infty$. Separability is useful because if it does not hold not all the Borel sets of the space are measurable. Separability also ensures that weak convergence on product spaces is equivalent to weak convergence on the component spaces. Two theorems are especially useful :-

1. If X is complete and separable, then each probability measure on $(X, \mathcal{B}(X))$ is tight.

2. A family Π of probability measures on X is tight if for every $\varepsilon > 0$ there exists a compact set K s.t. $P(K) > 1 - \varepsilon$ for all P in Π. Prohorov's Theorem :-

(a) If Π is tight, then it is relatively compact.
(b) If X is complete and separable, if Π is relatively compact, then it is tight.

3

NB The following notation conforms with Tanaka (1996) and differs from Phillips (1987) and Hamilton (1994).

We construct a function that measures a scaled ($1/\sqrt{T}$) partial sum of the errors u_j up to a <u>certain fraction</u> r of the total sample T:

(i) $X_T(t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{[Tt]} u_j$, where $[Tt]$ is the integer part of Tt, $t \in [0, 1]$,

$$= \frac{1}{\sqrt{T}} \sum_{i=1}^{j} u_j \text{ for } \frac{j-1}{T} \leq t \leq \frac{j}{T},$$

Clearly $X_T(t) \in D[0, 1]$.

- An elegant theory can be constructed using $D[0, 1]$ which contains jumps but not isolated points but it is not separable under the uniform metric discussed above, meaning there are “too many” sets on which to define a probability space.

- Skorokhod (1956) showed that the space can be rendered separable using what is now called the “Skorokhod metric” (his $J1$ metric), which allows functions to be compared “sideways” as well as vertically:

  for $x, y \in D[0,1]$,

$$d_S = \inf_{\lambda \in \Lambda} \left\{ \varepsilon > 0 : \sup_t \left| \lambda(t) - t \right| \leq \varepsilon, \sup_t \left| x(t) - y(\lambda(t)) \right| \leq \varepsilon \right\},$$

  where Λ denotes the set of all increasing functions $\lambda : [0,1] \rightarrow [0,1]$.

- This is the crucial element in terms of what is needed to define the Borel σ-algebra; however the metric space $(D[0,1], d_S)$ is not complete and Phillips used a modification introduced by Billingsley (1968) that preserves the same topology:

$$d_B = \inf_{\lambda \in \Lambda'} \left\{ \varepsilon > 0 : \|\lambda\| \leq \varepsilon, \sup_t |x(t) - y(\lambda(t))| \leq \varepsilon \right\},$$

where

$$\|\lambda\| = \sup_{t \neq s} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right| ,$$

and Λ' denotes the set of all increasing functions such that $\|\lambda\| \leq \infty$.

(ii) continuous version of the partial sum process

- It is possible to use $C[0, 1]$ endowed with the uniform metric even although most of the functions of interest are not continuous but this involves an awkward construction that requires extra terms defined to make the relevant partial sum process continuous be shown to be asymptotically negligible

$$X_T(t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{[Tt]} u_j + T(t - \frac{j}{T}) \frac{1}{\sqrt{T}} u_j \in C[0,1].$$

The asymptotic behaviour of the partial sum process in (ii) is the same as the that of (i).

Both (i) and (ii) set up the following two-stage approach:

- In the first stage, we consider a real-valued stochastic process $(x_t, t \in \mathbb{N})$ such that $n^{-1/2} x_{[rn]} \Rightarrow \sigma W(r)$, $\sigma > 0$, where $[rn]$, $r \in [0,1]$, denotes the integer part of rn, "$\Rightarrow$" denotes weak convergence in $D[0, 1]$ as described above, and W represents standard Brownian motion on [0,1]. The approach is set up such that one of a variety of such FCLT's could be employed [Donsker, Erdos and Kac, McLeish, Hernndorff].

e.g. Donsker's theorem: u_j i.i.d.$(0, \sigma^2)$ and $E(u_j^4) < \infty$,

then $X_T(t) = \dfrac{1}{\sigma\sqrt{T}} \sum_{j=1}^{[Tt]} u_j \Rightarrow W(r)$, standard BM on $C[0, 1]$.

- In the second stage, the continuous mapping theorem is applied, which preserves weak convergence under continuous mappings from the original metric space $C[0, 1]$ or $D[0, 1]$ into Euclidean space, e.g. $\int_0^1 X_T(t)dt$ or $\sup_{t\in[0,1]} X_T(t)$.

5. (b) Noting that $y_j = w_1 + \dots + w_j$, we define for $\frac{j-1}{T} \le t \le \frac{j}{T}$,

$$X_T(t) = \frac{1}{\sqrt{T}} \sum_{i=1}^{j} w_j + T\left(t - \frac{j}{T}\right) \frac{1}{\sqrt{T}} w_j$$

so that $X_T\left(\frac{j}{T}\right) = \frac{y_j}{\sqrt{T}}$, and $\frac{X_T}{\sigma} \Rightarrow BM(1)$

by Donsker's theorem

Consider

$$T(\hat{\rho} - 1) = \frac{\frac{1}{T} \sum_{j=2}^{T} y_{j-1} (y_j - y_{j-1})}{\frac{1}{T^2} \sum_{j=2}^{T} y_{j-1}^2}$$

$$= \frac{U_T}{V_T},$$

where $U_T = \frac{1}{T} \sum_{j=2}^{T} y_{j-1} (y_j - y_{j-1})$

$$= \frac{1}{2} X_T^2(1) - \frac{1}{2T} \sum_{j=1}^{T} \varepsilon_j^2$$

$$V_T = \frac{1}{T^2} \sum_{j=2}^{T} y_{j-1}^2 = \frac{1}{T} \sum_{j=1}^{T} X_T^2\left(\frac{j}{T}\right) - \frac{1}{T^2} y_T^2.$$

Define /.

Define the continuous fn $h_3(x) = (h_{31}(x), h_{32}(x))$ for $x \in C$, where

$$h_{31}(x) = \frac{1}{2} x^2(1), \quad h_{32}(x) = \int_0^1 x^2(t)\,dt$$

Then

$$U_T = h_{31}(X_T) - \frac{1}{2T} \sum_{j=1}^{T} \varepsilon_j^2$$

$$V_T = h_{32}(X_T) + R_{1T} - \frac{1}{T^2} y_T^2,$$

where $R_{1T} = \frac{1}{T} \sum_{j=1}^{T} X_T^2\left(\frac{j}{T}\right) - \int_0^1 X_T^2(t)\,dt$

$$= \sum_{j=1}^{T} \int_{(j-1)/T}^{j/T} \left[X_T^2\left(\frac{j}{T}\right) - X_T^2(t) \right] dt$$

We show that $R_{1T} \xrightarrow{p} 0$. We have

$$\left| X_T^2\left(\frac{j}{T}\right) - X_T^2(t) \right| \le 2 \sup_{0 \le t \le 1} |X_T(t)| \max_{1 \le j \le T} \frac{|\varepsilon_j|}{\sqrt{T}}$$

Now

$$P\left(\max_{1 \le j \le T} \frac{|\varepsilon_j|}{\sqrt{T}} \le \delta \right) = \prod_{j=1}^{T} P\left(\frac{|\varepsilon_j|}{\sqrt{T}} \le \delta \right)$$

$$= \left\{ 1 - P\left(\frac{|\varepsilon_1|}{\sqrt{T}} > \delta \right) \right\}^T$$

$$\ge \left[1 - \frac{1}{\delta^2 T} E\left\{ \varepsilon_1^2 I\left(\frac{|\varepsilon_1|}{\sqrt{T}} > \delta \right) \right\} \right]$$

and so

$$\max_{1 \le j \le T} \frac{|\varepsilon_j|}{\sqrt{T}} \xrightarrow{p} 0.$$

Hence

$$\sup_{0 \le t \le 1} |X_T(t)| \max_{1 \le j \le T} \frac{|\varepsilon_j|}{\sqrt{T}} \xrightarrow{p} 0, \text{ and } R_{1T} \xrightarrow{p} 0.$$

Applying Donsker's theorem and the continuous mapping theorem

$$\left(\frac{U_T}{\sigma^2}, \frac{V_T}{\sigma^2} \right) \Rightarrow \left(h_{31}(W) - \tfrac{1}{2}, h_{32}(W) \right)$$

Hence

$$T(\hat{\rho} - 1) \Rightarrow \frac{h_{31}(W) - \frac{1}{2}}{h_{32}(W)} \equiv \frac{\frac{1}{2}(W^2(1) - 1)}{\int_0^1 W^2(t)\,dt}$$

Note that

$$\frac{\hat{\rho} - 1}{\hat{\sigma} / \sqrt{\sum_{j=2}^{T} y_{j-1}^2}} = \frac{U_T / V_T}{\hat{\sigma} / \sqrt{V_T}} = \frac{U_T}{\hat{\sigma} \sqrt{V_T}}$$

Now

$$\sigma^2 = \frac{1}{T-1} \sum_{j=2}^{T} \left(y_j - y_{j-1} - (\hat{\rho} - 1) y_{j-1} \right)^2$$

$$= \frac{1}{T-1} \Big[\sum_{j=2}^{T} \varepsilon_j^2 - 2(\hat{\rho} - 1) \sum_{j=2}^{T} y_{j-1} \varepsilon_j$$

$$+ (\hat{\rho} - 1)^2 \sum y_{j-1}^2 \Big]$$

5. (b) cont.

Since $\hat{e} - 1 = O_p(T^{-1})$ and

$$\sum_{j=2}^{T} y_{j-1}\varepsilon_j = O_p(T), \quad \sum_{j=2}^{T} y_{j-1}^2 = O_p(T^2),$$

$\hat{\sigma}^2 \xrightarrow{p} \sigma^2$. Noting that $V_T \Rightarrow \sigma^2 \int_0^1 W^2(t)\,dt$,

we have

$$t_{e=1} \Rightarrow \frac{\frac{1}{2}(W^2(1) - 1)}{\sqrt{\int_0^1 W^2(t)\,dt}}.$$

- We have derived the asymptotic distribution of the OLS estimator in the AR(1) model with a unit root, given by

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

- Can we use $\widehat{\rho}_T$ to test whether $\rho = 1$ in the AR(1) model?
- Can we test for the presence of a unit root in more general ARMA models?

- Consider estimating, by OLS, the model

$$y_t = \rho y_{t-1} + u_t, \quad y_0 = 0, \quad u_t \sim \text{i.i.d.}\, N(0, \sigma^2). \tag{1}$$

- Aim to test $H_0 : \rho = 1$ against $H_1 : \rho < 1$.
- Note: explosive alternatives ($\rho > 1$) are excluded under H_1.
- Note: the test is one-sided and we can use

$$T\left(\widehat{\rho}_T - 1\right) \Rightarrow \frac{(1/2)[W(1)^2 - 1]}{\int_0^1 W(r)^2 dr} \text{ under } H_0. \tag{2}$$

- Critical values in Table B.5 (case 1) of Hamilton (1994).
- Note: the test is one-sided so we need significantly negative values of $T\left(\widehat{\rho}_T - 1\right)$ in order to reject H_0.

An alternative is the Dickey-Fuller t-test, the statistic being

$$t_T = \frac{(\widehat{\rho}_T - 1)}{\widehat{\sigma}_{\widehat{\rho}_T}}, \tag{3}$$

$$\widehat{\sigma}_{\widehat{\rho}_T} = \sqrt{\frac{s_T^2}{\sum_{t=1}^{T} y_{t-1}^2}}, \tag{4}$$

$$s_T^2 = \frac{1}{T-1} \sum_{t=1}^{T} \left(y_t - \widehat{\rho}_T y_{t-1}\right)^2 . \tag{5}$$

- Under $H_0 : \rho = 1$, t_T does not have a limiting $N(0,1)$ distribution as it would in the stationary case.
- From (3) and (4) we have

$$t_T = \frac{T(\widehat{\rho}_T - 1)\left[T^{-2}\sum_{t=1}^{T} y_{t-1}^2\right]^{1/2}}{s_T} \tag{6}$$

$$= \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}u_t}{\left[T^{-2}\sum_{t=1}^{T} y_{t-1}^2\right]^{1/2} s_T}. \tag{7}$$

- The consistency of $\widehat{\rho}_T$ ensures that $s_T^2 \xrightarrow{p} \sigma^2$.
- Using earlier results we find that, as $T \to \infty$,

$$t_T \Rightarrow \frac{(\sigma^2/2)[W(1)^2 - 1]}{\left[\sigma^2 \int_0^1 W(r)^2 dr\right]^{1/2} \sigma} = \frac{(1/2)[W(1)^2 - 1]}{\left[\int_0^1 W(r)^2 dr\right]^{1/2}}. \qquad (8)$$

- Critical values in Table B.6 (case 1) of Hamilton (1994).
- Need significantly negative values of t_T in order to reject the null hypothesis $H_0 : \rho = 1$.

- Consider the ARMA(p,q) process

$$\widetilde{\phi}(L)y_t = \theta(L)\epsilon_t, \quad \epsilon_t \sim \text{i.i.d.}(0, \sigma^2). \tag{9}$$

- If a unit root exists,

$$\widetilde{\phi}(z) = (1 - z)\phi(z),$$

  where $\phi(z)$ is a polynomial of order $p - 1$.

- Hence

$$(1 - L)y_t = \frac{\theta(L)}{\phi(L)}\epsilon_t \quad \text{or} \quad y_t = y_{t-1} + u_t$$

  where $u_t = \dfrac{\theta(L)}{\phi(L)}\epsilon_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$.

- We could therefore consider

$$y_t = \rho y_{t-1} + u_t, \quad u_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \tag{10}$$

  and test $H_0 : \rho = 1$ against $H_1 : \rho < 1$ using $\widehat{\rho}_T$ as before.
- But the serial correlation in u_t will affect the results.
- A useful tool is the *Beveridge-Nelson decomposition*.

Beveridge-Nelson decomposition

Let $u_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ where $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$, $E(\epsilon_t \epsilon_s) = 0$ $(t \neq s)$, and

$$\sum_{j=0}^{\infty} j|\psi_j| < \infty. \tag{11}$$

Then

$$\sum_{j=1}^{t} u_j = \psi(1) \sum_{j=1}^{t} \epsilon_j + \eta_t - \eta_0, \tag{12}$$

where $\eta_t = \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j}$, $\alpha_j = -\sum_{k=j+1}^{\infty} \psi_k$ and $\sum_{j=0}^{\infty} |\alpha_j| < \infty$.

- Thus an I(1) (integrated of order one) process $\Delta y_t = u_t$ can be written

$$y_t = \sum_{j=1}^{t} u_j + y_0 = \underbrace{\psi(1) \sum_{j=1}^{t} \epsilon_j}_{\text{(a)}} + \underbrace{\eta_t}_{\text{(b)}} \underbrace{- \eta_0 + y_0}_{\text{(c)}} . \qquad (13)$$

- Note that: (a) is a random walk, (b) is stationary, and (c) represents initial conditions.

The representation in (13) is useful for establishing the behaviour of

$$\begin{aligned}\sqrt{T}X_T(r) &= T^{-1/2}\sum_{t=1}^{[Tr]} u_t \\ &= \psi(1)T^{-1/2}\sum_{t=1}^{[Tr]} \epsilon_t + T^{-1/2}\left(\eta_{[Tr]} - \eta_0\right) \\ &\Rightarrow \sigma\psi(1)W(r), \qquad (14)\end{aligned}$$

using the earlier convergence results for the partial sums of white noise (provided that $E(\epsilon_t^4) < \infty$ as well) and the fact that $T^{-1/2}(\eta_{[Tr]} - \eta_0) \xrightarrow{p} 0$ as $T \to \infty$.

- From (14) we find that, defining $\lambda = \sigma\psi(1)$,

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 = \int_0^1 \left[\sqrt{T}X_T(r)\right]^2 dr \Rightarrow \lambda^2 \int_0^1 W(r)^2 dr. \tag{15}$$

- Furthermore, recall that

$$T^{-1}\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}\left(\frac{1}{T}y_T^2 - \frac{1}{T}\sum_{t=1}^{T} u_t^2\right);$$

$$\frac{1}{T}y_T^2 = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t\right]^2 \Rightarrow \lambda^2 W(1)^2; \tag{16}$$

$$\frac{1}{T}\sum_{t=1}^{T} u_t^2 \xrightarrow{p} \gamma_0 = E(u_t^2) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2. \tag{17}$$

- Hence we find that

$$T\left(\widehat{\rho}_T - 1\right) \Rightarrow \frac{\frac{1}{2}[\lambda^2 W(1)^2 - \gamma_0]}{\lambda^2 \int_0^1 W(r)^2 dr} = \frac{\frac{1}{2}[W(1)^2 - (\gamma_0/\lambda^2)]}{\int_0^1 W(r)^2 dr}. \tag{18}$$

- (2) is a special case in which $\psi(z) = 1$ and $\gamma_0 = \sigma^2 = \lambda^2$.
- Note that γ_0 and λ^2 represent nuisance parameters in (18) i.e. we can't obtain critical values for the distribution unless they are known.
- There are two main approaches to tackling this problem – one uses a nonparametric correction, while the other is fully parametric.

- We can write (18) in terms of the distribution in (2) (the critical values for which are tabulated) as follows:

$$T\left(\widehat{\rho}_T - 1\right) \Rightarrow \frac{(1/2)[W(1)^2 - 1]}{\displaystyle\int_0^1 W(r)^2 dr} + \frac{(1/2)[1 - (\gamma_0/\lambda^2)]}{\displaystyle\int_0^1 W(r)^2 dr}. \quad (19)$$

- The second component on the RHS of (19) can be written

$$\eta = \frac{(1/2)(\lambda^2 - \gamma_0)}{\lambda^2 \displaystyle\int_0^1 W(r)^2 dr}. \quad (20)$$

- The idea behind the Phillips-Perron test is to find a random variable $\widehat{\eta}_T$ such that $\widehat{\eta}_T \Rightarrow \eta$ and to base inferences on

$$T\left(\widehat{\rho}_T - 1\right) - \widehat{\eta}_T \Rightarrow \frac{(1/2)[W(1)^2 - 1]}{\displaystyle\int_0^1 W(r)^2 dr}.$$

- It is easy to estimate γ_0 using

$$\widehat{\gamma}_0 = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t^2, \tag{21}$$

where $\widehat{u}_t = y_t - \widehat{\rho}_T y_{t-1}$.

- Note that the spectral density of u_t is

$$s_u(\omega) = \frac{\sigma^2}{2\pi}\left|\psi(e^{-i\omega})\right|^2 \longrightarrow 2\pi s_u(0) = \sigma^2\psi(1)^2 = \lambda^2,$$

$$\text{and} \;\; s_u(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_j e^{-ij\omega} \longrightarrow 2\pi s_u(0) = \sum_{j=-\infty}^{\infty}\gamma_j.$$

- A possible estimator for λ^2 is therefore

$$\widehat{\lambda}^2 = \sum_{j=-q}^{q} b_j\widehat{\gamma}_j = \widehat{\gamma}_0 + 2\sum_{j=1}^{q} b_j\widehat{\gamma}_j, \tag{22}$$

$$\widehat{\gamma}_j = (1/T)\sum_{t=j+1}^{T}\widehat{u}_t\widehat{u}_{t-j}, \;\; b_j = 1 - \frac{j}{(q+1)}.$$

- This is the Newey-West estimator which ensures $\widehat{\lambda}^2 > 0$.

- Note: the lag truncation number q needs to be chosen large enough so that $\widehat{\lambda}^2$ approximates $2\pi s_u(0)$ sufficiently well.
- The Phillips-Perron test statistic is therefore

$$\widehat{\eta}_T = \frac{(1/2)(\widehat{\lambda}^2 - \widehat{\gamma}_0^2)}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2} \Rightarrow \eta \text{ as } T \to \infty. \tag{23}$$

- The nonparametric adjustment for nuisance parameters yields a limiting distribution which is free of nuisance parameters and for which critical values are tabulated.

- Suppose y_t satisfies the AR(p) process

$$\phi(L)y_t = \epsilon_t,$$

where $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$.

- Let

$$\rho = \phi_1 + \phi_2 + \ldots + \phi_p,$$

$$\zeta_j = -[\phi_{j+1} + \phi_{j+2} + \ldots + \phi_p], \quad (j = 1, \ldots, p-1).$$

- Then (see Hamilton p.517) it can be shown that

$$\phi(z) = (1 - \rho z) - \zeta(z)(1 - z), \tag{24}$$

where $\zeta(z) = \zeta_1 z + \zeta_2 z^2 + \ldots + \zeta_{p-1} z^{p-1}$.

- It follows that

$$y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta y_{t-j} + \epsilon_t,$$

or, subtracting y_{t-1} from both sides,

$$\Delta y_t = (\rho - 1) y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta y_{t-j} + \epsilon_t. \tag{25}$$

- This is the ADF($p - 1$) regression, and the test statistic is the usual t-ratio on $\rho - 1$ for testing $H_0 : \rho - 1 = 0$.
- The t-statistic has the same distribution as in (8).

- Alternatively, suppose that Δy_t has the Wold representation
$$\Delta y_t = \psi(L)\epsilon_t.$$
- The ADF approach attempts to approximate the polynomial $\psi(z)$ by $1/[1-\zeta(z)]$, yielding
$$\Delta y_t = \sum_{j=1}^{p-1} \zeta_j \Delta y_{t-j} + \epsilon_t.$$
- Adding y_{t-1} to the RHS enables the test of a unit root to be carried out in the usual way.

- It can be important to incorporate constants and time trends into the PP and ADF regressions:

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

$$\Delta y_t = \alpha + \delta t + (\rho - 1) y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta y_{t-j} + \epsilon_t.$$

- But it is important to bear in mind the implied model for y_t under both the null (H_0) and alternative (H_1) hypotheses.
- Including a constant and/or time trend in the regression affects the limiting distributions of the test statistics and different sets of critical values are needed; these are given in Hamilton (1994).

- To apply the PP and ADF tests it is necessary to choose the lag truncation parameter q in the former and the number of lags p in the latter.
- For the ADF test, it is possible to use information criteria (e.g. Schwarz's or Akaike's or Hannan-Quinn's) to determine p.
- But bear in mind that p needs to be sufficiently large to capture the serial correlation properties in y_t i.e. the residuals need to be approximately white noise for the critical values to be appropriate – standard serial correlation tests can be used to check this.