Contents

Chapter 10

Unit Roots

10.1 Introduction

In this chapter we introduce some formal tests of the unit root hypothesis. Recall that in the Box-Jenkins analysis, the strategy for handling nonstationarities has been to difference the model until the autocorrelation function decays. This, of course, is not a very satisfactory guide to practitioners and can lead to testing and estimating procedures with unknown tests sizes or statistical properties.

The pioneering work for testing for unit roots is due to Dickey and Fuller and some of the earliest discussion goes back to Fuller (1976). Our intention in these notes is to introduce the topic and to provide some of the more basic results. Since that time the literature has exploded with a variety of topics on nonstationary time series and this research agenda continues today. Much of the advances in this area are due to Peter Phillips of Yale University and there is a fairly readable primer on the topic (Phillips and Xiao,1998).

The area makes a good topic for those interested in doing special programing on a recent test or estimator. STATA is well-equipped to handle the kinds of techniques necessary (although like most things there could be improvement as in the calculating long-run covariance matrices).

The need to have some formal analysis for handling unit roots really arises from the nature of the data. Many aggregate economic time series (consumption, income, interest rates, money, stock prices, etc.) display strong persistence with sizable fluctuations in both mean and variance over time. Classical hypothesis testing is based on the assumption that the first two population moments (unconditional) are constant over time (covariance stationary), and hence unit roots pose a challenge for our classical $\sqrt{T}$ asymptotic procedures. For example, if a series $y_t \sim I(d)$ with $d \geq 1$, then the usual statistical properties of the first and second sample moments (converging to their population counterparts) do not apply. In this case, a different asymptotic distributional theory is required for non-stationary, non-ergodic processes.

10.2 Unit Roots

We begin our discussion of unit roots with a simple example. Consider the time series, y_t, which evolves according to an $AR(1)$ process.

$$y_t = \phi y_{t-1} + u_t, \quad u_t \sim iid(0, \sigma^2) \tag{10.1}$$

The mean of y_t is:

$$E[y_t] = E[\phi y_{t-1} + u_t] = 0 \tag{10.2}$$

and the variance of y_t at time t is:

$$\begin{aligned} \gamma_0 &= E[y_t^2] = E[\phi y_{t-1} + u_t]^2 \\ &= E[\phi^2 y_{t-1}^2 + 2\phi y_{t-1} + u_t^2] \\ &= \phi^2 \gamma_0 + \sigma^2 \end{aligned}$$

with u_t independent of y_{t-1} and $V[y_t]$ constant for all t. Solving (10.3) for γ_0 yields

$$\gamma_0 = \sigma^2/(1 - \phi^2)$$

As before $\phi < 1$ implies that the $AR(1)$ process is stationary. When $\phi = 1.0$, the variance y_t is unbounded, as t gets arbitrarily large. To see this write

$$\begin{aligned} y_t &= y_{t-1} + u_t \\ &= \sum_{t=1}^{t} u_t + y_{0,} \qquad y_o \quad \text{is given} \\ &= S_t + y_0 \end{aligned} \tag{10.3}$$

where $S_t =$ partial sum process, u_t is assumed to be iid and y_t is said to be integrated of order one, $I(1)$. The case where $\phi = 1$ is called a **unit root process** and can be made stationary by taking first differences. On the other hand, **random walks** refer to the special cases of **iid** u_t. When ϕ>1 the $AR(1)$ process is *explosive.*

If after differencing a time series **d** times, the series is stationary, then we say the series is integrated of order d and denote this by I(d). Clearly, a time series that is stationary to begin with, is just I(0). For convenience, we will confine the discussion to I(1) variables.

An integrated process of order 1 has several important properties:

1. $V[y_t] = t\sigma^2$, which can easily be deduced from (10.3).

2. $CORR(y_t, y_{t-j}) = \sqrt{1 - j/t} \to 1$ as $t \to \infty$, for some fixed j, implying that the autocorrelation function does not die out.

$$\begin{aligned} \rho &= \frac{C[y_t, y_{t-j}]}{SD[y_t] \times SD[y_{t-j}]} \\ &= \frac{(t-j)\sigma^2}{\sqrt{t}\sigma^2\sqrt{(t-j)\sigma^2}} \\ &= \sqrt{\frac{(t-j)^2}{t \times (t-j)}} \\ &= \sqrt{\sqrt{1 - j/t}} \end{aligned}$$

3. The spectrum of y_t is

$$f_y(\omega) = \frac{\sigma^2}{2\pi(1 - \cos^2\omega)^2}$$

Recall the spectrum for a stationary $AR(1)$ process (Chapter 8)

$$f_y(\omega) = \frac{\sigma^2}{2\pi}\left(\frac{1}{1 - 2\phi\cos\omega + \phi^2}\right) \tag{10.4}$$

As $\phi \to 1$, $f_y(\omega)$ is proportional to ω^{-2} for small ω, and that the spectrum has a spike (singularity or undefined) at frequency $\omega = 0$. Low frequencies are associated we movements in the long-run we can think of the series as being dominated by long-term stochastic trends

4. $E\,|[(y_{t+k}y_{t-j})]| > 0$. This last property shows that the optimal k step forecast does not converge in probability to zero (the unconditional mean) as $k \to \infty$. The process is said to have long memory and information today is forever useful in forecasting.

Having introduced the concept of integrated series, we now turn to the issue of **testing for unit roots**. Space and energy prevents us from discussing all of the tests for unit roots. Instead we shall focus on some of the best known tests for unit roots. These include the tests by Dickey and Fuller [1979], [1981]; Evans and Savin [1981]; Phillips [1986], and Phillips and Perron [1986]; Sargan and Bhargava [1983]; Bhargava [1986] and some more recent ones. We note the Schwert [1989] paper in which he conducts numerous Monte Carlo exercises to evaluate the finite sample performance of these various tests. We will also be adding some recent refinements to tests for unit roots at the end of the chapter.

10.3 Testing for Unit Roots in AR Models

The main practical results from these rather technical papers are "t-tests" for unit roots in autoregressions (Dickey and Fuller [1979]), likelihood ratio tests (Dickey and

Fuller [1981]), tests based on the Durbin-Watson statistic (Sargan and Bhargava [1983]) and tests for I(1) series (Phillips [1987], Phillips and Durlauf [1986] and Phillips and Perron [1986]) for the non iid case. We will discuss each in turn.

10.4 Technical Background

Consider the simple $AR(1)$ model.

$$y_t = \phi y_{t-1} + u_t, \qquad t = 1, 2, \ldots, T \tag{10.5}$$

where $y_t = y_0$ for $t = 0$ (and is fixed). The OLS estimator for ϕ in (10.5) is

$$\hat{\phi} = \left[\sum_{t=1}^{T} y_{t-1}^2\right]^{-1} \sum_{t=1}^{T} y_t y_{t-1} \tag{10.6}$$

If y_0 is fixed, and $u_t \rightarrow NID(0, \sigma^2)$, then the OLS estimator is the maximum likelihood estimator. In general, however, the distribution of $\hat{\phi}$ depends on the initial conditions and the distribution of the u_t's. The error in the estimator of ϕ is given by

$$(\hat{\phi} - \phi) = \left[\sum_{t=1}^{T} y_{t-1}^2\right]^{-1} \sum_{t=1}^{T} u_t y_{t-1}$$

For the case of $\mid \phi \mid < 1$, we have established that the limiting distribution as (see Chapter 5)

$$\sqrt{T}(\hat{\phi} - \phi) \sim N(0, 1 - \phi^2) \tag{10.7}$$

This is $\sqrt{T}$ (root T) consistent and we may write

$$\hat{\phi} - \phi \;=\; O_p(T^{-\frac{1}{2}}) \tag{10.8}$$

If $\phi \;=\; 1$, we see from (10.7) that the variance of $\sqrt{T}(\hat{\phi} - \phi)$ is 0, indicating the distribution $\sqrt{T}(\hat{\phi} - \phi)$ is *degenerative*.

Indeed, we may show that for $\phi \;=\; 1$ (see Fuller):

$$\hat{\phi} - 1 \;=\; O_P(T^{-1}) \tag{10.9}$$

From (10.9) we see that for the unit root case $\hat{\phi}$ is T consistent and that there is a limiting distribution. Hence convergence for the estimator $\hat{\phi}$ when $\phi \;=\; 1$, is faster (T *vs* $\sqrt{T}$) than $\phi < 1$.

10.4.1 Explosive Case

White (1958) showed that if $y_0 = 0,\ u_t \sim NID(0, \sigma^2)$ and $|\ \phi\ | > 1.0$, then the limiting distribution of

$$(\phi^2 - 1)^{-1}(\hat{\phi} - \phi)$$

is the ratio of two independent normally distributed random variables. Thus, the limiting distribution of $(\phi^2 - 1)^{-1}(\hat{\phi} - \phi)$ is a *Cauchy* distribution.

10.4.2 Three Different Asymptotics

This changing asymptotic structure depending on the value of ϕ, is one of the reasons why this problem is so tough. We get three sets of asymptotic results depending on the value of ϕ (which is unknown).

1. $\phi < 1 \Leftrightarrow$ asymptotic normal $\sqrt{T}$ consistent
2. $\phi = 1$ nonstandard distribution T consistent
3. $\phi > 1$ Cauchy distribution.

Bayesians have persuasively argued that this is an obvious drawback with classical asymptotics that is not encountered if a Bayesian approach to unit roots is adopted (see Koop, 1991).

10.5 Limiting Distribution of $\hat{\phi}$ when $\phi = 1$ (iid)

Consider the equation with $\phi = 1$ and the errors are **iid**:

$$\begin{aligned} y_t &= \phi y_{t-1} + u_t \quad u_t \sim iid \\ \triangle y_t &= (\phi - 1)y_{t-1} + u_t \end{aligned} \tag{10.10}$$

White [1958] showed that the limiting distribution of $T(\hat{\phi} - 1)$(with $\phi = 1$) can be represented as the ratio of two integrals defined on the Wiener process.

$$T(\hat{\phi} - 1) = \frac{T\sum_{t=1}^{T} \Delta y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \Rightarrow \frac{W(1)^2 - \sigma^2/\sigma^2}{2\int_0^1 W(r)^2 dr} \tag{10.11}$$

Here $W(r)$ is standard Brownian Motion or a Wiener process defined on the interval 0 to 1. Dickey and Fuller [1979] make use of the distribution of ratios of quadratic forms to simulate equation 10.11 by approximating the infinite series by a finite truncation.

10.5.1 Brownian Motion

This discussion is taken from Hamilton. Consider the following process for y_t

$$y_t = y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim iid\ N(0,1) \text{ and } y_o = 0 \tag{10.12}$$

We may write this as the partial sum process as before

$$y_t = S_t = \varepsilon_1 + \ldots + \varepsilon_t \tag{10.13}$$

From 10.12 and 10.13 we may write

$$y_t \sim N(0,t) \tag{10.14}$$

For $s > t$ we have

$$y_s - y_t = \underbrace{\varepsilon_{t+1} + \varepsilon_{t+2} + \ldots + \varepsilon_s}_{(s-t) \text{ terms}} \tag{10.15}$$

Implying

$$y_s - y_t \sim N(0, s-t) \tag{10.16}$$

Given our assumptions $y_s - y_t$ is independent of the change between dates r and q $(y_r - y_q)$ for any dates: $t < s < r < q$ (no overlap). Increments are independent!

Continuous Time Development

Consider the change between y_{t-1} and $y_t \Rightarrow \varepsilon_t \sim N(0,1)$. Suppose we view this change as comprising two additive changes (dated at the midpoint). That is

$$\varepsilon_t = e_{1t} + e_{2t} \tag{10.17}$$

where we associate e_{1t} with the change between y_{t-1} and $y_{t-(1/2)}$. Therefore, we can write:

$$y_{t-(1/2)} - y_{t-1} = e_{1t} \text{ and } y_t - y_{t-(1/2)} = e_{2t} \tag{10.18}$$

Sampled at integer dates $t = 1, 2\ldots$ the processes (10.18) will have the same properties as (10.12), since

$$y_t - y_{t-1} = e_{1t} + e_{2t} \tag{10.19}$$

We do not need to restrict t to be an integer$\Rightarrow$ $\{t+(1/2)\}_{t=0}^{\infty}$.

Therefore, for both integer and noninteger 10.16 holds. Furthermore we can partition the interval t and $t-1$ into more that 2 separate decompositions.

Consider dividing the interval into N separate subperiods:

$$y_t - y_{t-1} = e_{1t} + e_{2t} \ldots + e_{Nt} \quad e_{it} \sim N(0, \frac{1}{N}) \tag{10.20}$$

As $N \to \infty$ we have a *continuous time process.* This process is called **Brownian Motion** or **Wiener Process**. Keep in mind that t is any non-negative real number.

10.5.2 Definition of a Wiener Process

A **Wiener process** $W\ =\ \{W(t)\ t \geq 0\}$ starting from $W(0)\ =\ w\ =\ 0$ is a real-valued Gaussian process such that

1. W has independent increments.

$$W(s+t) - W(s) \sim N(0, \sigma^2 t) \text{ where } \sigma^2 \text{ is a constant}$$

2. If $\sigma^2\ =\ 1$, then we call it **standard Wiener process** and we may write $X(t) \sim N(0,\ t)$.

Clearly Wiener processes are nonstationary. When we write $W(t)$ or $X(t)$ with parentheses, we our denoting the continuous time object. When we write y_t this is the discrete time variable (subscripts). Note that the realization of Brownian motion is a continuous function of t . The process is non-diffentiable using standard calculus since the change at time t may be completely different for $t+\Delta$ for arbitrarily small Δ.

10.5.3 Intuition Behind the Distribution

In equation (10.11), we defined the interval of the Wiener process from 0 to 1. This is accomplished by defining the step-function $X_T(r)$ over the interval r $(0 \leq r \leq 1)$:

$$\begin{aligned} X_T(r) = \ & T^{-1}S_{[Tr]}\ =\ T^{-1}S_{j-1}, & (10.21)\\ & \underbrace{\frac{(j-1)}{T} \leq r \leq j/T \quad j\ =\ 1\ldots T}_{0\ldots 1} & (10.22) \end{aligned}$$

where S is the *partial sums* introduced in 10.3

$$S_t = u_0 + u_1 + \cdots + u_t$$

and [ ] represents the largest integer function. We are breaking the sample T over an integer share Tr

Let as write $X_T(r)$ (we are writing all the fractions r can take on over the interval T). $X_T(r)$ is a step function in r:

$$X_T(r) = \begin{cases} 0 & 0 \leq r < \frac{1}{T} \\ \frac{u_1}{T} & \frac{1}{T} \leq r < \frac{2}{T} \\ \frac{u_1+u_2}{T} & \frac{2}{T} \leq r < \frac{3}{T} \\ \vdots & \vdots \\ \frac{u_1+u_2+\ldots+u_T}{T} & \text{for } r = 1 \end{cases} \tag{10.23}$$

and can be written as a series of sample means over fraction $[Tr]$ of the observations T

$$\begin{aligned} \sqrt{T}\,X_T(r) &= T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} u_t \\ &= \frac{[Tr]^{\frac{1}{2}}}{T^{\frac{1}{2}}} \times \frac{1}{[Tr]^{\frac{1}{2}}} \sum_{t=1}^{[Tr]} u_t \end{aligned}$$

We have a standard central limit result that

$$\frac{1}{[Tr]^{\frac{1}{2}}} \sum_{t=1}^{[Tr]} u_t \Rightarrow N(0,\sigma^2) \tag{10.24}$$

and $\frac{[Tr]^{\frac{1}{2}}}{T^{\frac{1}{2}}} \rightarrow \sqrt{r}$. Putting this and 10.24 we have

$$\sqrt{T}\,X_T(r) \Rightarrow N(0, r\sigma^2) \tag{10.25}$$

This is the distribution of the function at date r.

10.5.4 Functional Central Limit Theorem

Consider the behaviour of the sample mean based on observations $[Tr_1]$ through to $[Tr_2]$. It follows from (10.25) that

$$\sqrt{T}\,[X_T(r_2) - X_T(r_1)] \Rightarrow N(0, \sigma^2(r_2 - r_1)) \tag{10.26}$$

and is independent of any r such that $r < r_1$. Therefore the sequence of stochastic functions $\frac{\sqrt{T}X_T(\cdot)}{\sigma}$ is distributed as a standard Brownian motion

$$\frac{\sqrt{T}X_T\,(\cdot)}{\sigma} \Rightarrow W\,(\cdot) \tag{10.27}$$

This is known as a **functional central limit theorem**. It is a function because of r. As an example, consider if $r = 1$

$$X_T(1) = \frac{1}{T}\sum_{t=1}^{T} u_t \tag{10.28}$$

Therefore when $r = 1$, the conventional central limit theorem obtains as a special case:

$$\frac{\sqrt{T}X_T(1)}{\sigma} = \frac{1}{\sigma\sqrt{T}}\sum_{t=1}^{T} u_t \Rightarrow W(1) \sim N(0,1) \tag{10.29}$$

To recap we have:

1. Brownian motion
2. Functional Central Limit Theorem
3. (now we add) Continuous Mapping Theorem (functions of Brownian motion)

10.5.5 Continuous Mapping Theorem

If S is a continuous function like

$$S_T(r) = \left[\sqrt{T}X_T(r)\right]^2 \tag{10.30}$$

We know that $\sqrt{T}X_T(\cdot) \Rightarrow \sigma W(\cdot)$. The continuous mapping theorem implies

$$S_T(\cdot) \Rightarrow \sigma^2 W(\cdot)^2 \tag{10.31}$$

That is, if the value of $W(r)$ from a realization of Brownian motion at every date r is squared, and multiplied by σ^2, this results in a continuous time process that would follow the same probability law as (10.30), for T sufficiently large.

The continuous mapping theorem allows us to obtain results for various functions of Brownian motion such as (10.11).

10.6 Distribution of the Unit Root

With these results for Brownian motion, we can formally derive the limiting distribution of the least squares estimate of ϕ under the unit root given by equation (10.11)

$$T(\hat{\phi} - 1) = \frac{T\sum_{t=1}^{T} \Delta y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$$

We have two parts to consider. The numerator:

$$T\sum_{t=1}^{T}\Delta y_t y_{t-1} = \sum_{t=1}^{T} y_{t-1}u_t \tag{10.32}$$

and the denominator

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

First we need some preliminaries

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

$$y_t^2 = (y_{t-1} + u_t)^2 = y_{t-1}^2 + [2y_{t-1}u_t] + u_t^2 \tag{10.33}$$

Rearranging $[2y_{t-1}u_t]$ and doing some recursive substitution yields

$$\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}\left\{y_T^2 - y_0^2\right\} - \frac{1}{2}\sum_{t=1}^{T} u_t^2 \tag{10.34}$$

In both the numerator and denominator of (10.11), there is the *sum of squares of a random walk*. We have seen from the continuous mapping example the convergence results:

$$S_T(r) = T\left[X_T(r)\right]^2 \tag{10.35}$$

which we have seen we can write

$$S_T(r) = \begin{cases} 0 & 0 \le r < \frac{1}{T} \\ \frac{y_1^2}{T} & \frac{1}{T} \le r < \frac{2}{T} \\ \frac{y_2^2}{T} & \frac{2}{T} \le r < \frac{3}{T} \\ \vdots & \vdots \\ \frac{y_T^2}{T} & \text{for } r = 1 \end{cases} \tag{10.36}$$

Notice that this step function produces random rectangles and if we intergrate this step function:

$$\int_0^1 S_T(r)\,dr = \frac{y_1^2}{T^2} + \frac{y_2^2}{T^2} + \ldots + \frac{y_T^2}{T^2} \tag{10.37}$$

From the continuous mapping theorem we know each of these components $S_T(\cdot) \Rightarrow \sigma^2 W(\cdot)^2$ and therefore

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 [W(r)]^2\,dr \tag{10.38}$$

Now given $y_0 = 0$, multiplying (10.34) by T^{-1}gives

$$T^{-1}\sum_{t=1}^{T} y_{t-1}u_t = \frac{1}{2}\frac{1}{T}y_T^2 - \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T} u_t^2 \tag{10.39}$$

which from (10.36) can be written

$$T^{-1}\sum_{t=1}^{T} y_{t-1}u_t = \underbrace{\frac{1}{2}S_T(1)}_{\frac{1}{2}\sigma^2 W(1)^2} - \underbrace{\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T} u_t^2}_{\sigma^2} \tag{10.40}$$

The last term in (10.40) is $\frac{1}{T}\sum\limits_{t=1}^{T} u_t^2 \Rightarrow \sigma^2$ by the law of large numbers and $S_T(1) \Rightarrow \sigma^2 W(1)^2$. Putting it all together means

$$T^{-1}\sum_{t=1}^{T} y_{t-1}u_t \Rightarrow \frac{1}{2}\sigma^2 W(1)^2 - \frac{1}{2}\sigma^2 \tag{10.41}$$

Finally we have:

$$T(\hat{\phi}-1) = \frac{T^{-1}\sum_{t=1}^{T}\Delta y_t u_{t-1}}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2} \Rightarrow \frac{W(1)^2 - 1}{2\int_0^1 W(r)^2 dr} \tag{10.42}$$

as the σ^2 cancels out and the limiting distribution from (10.11) is obtained.

Of course the pure random walk equation (10.10) is very restrictive. We would like to allow for serially correlated and possibly heteroskedastic errors. We turn to the Beveridge-Nelson Decomposition which illustrates the serial correlation case.

10.7 A Slight Digression: Beveridge-Nelson Decomposition

Let u_t be a stationary process (well defined spectrum at all frequencies) such that

$$u_t = \Psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \Psi_i \varepsilon_{t-j} \tag{10.43}$$

where

$$
\begin{aligned}
E\left[\varepsilon_t\right] &= 0 \\
E\left[\varepsilon_t \varepsilon_\tau\right] &= \begin{cases} \sigma^2 \text{ for } t=\tau \\ 0 \text{ otherwise} \end{cases} \\
\sum_{j=0}^{\infty} & \mid \quad j\ \Psi_j \mid < \ \infty
\end{aligned}
$$

Then the partial sum of u_t

$$
S_t = u_1 + u_2 + \cdots + u_t = \Psi(1)(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) + \eta_t - \eta_0 \tag{10.44}
$$

where

$$
\begin{aligned}
\Psi(1) &\equiv \sum_{j=0}^{\infty} \Psi_j \\
\eta_t &= \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j} \\
\alpha_j &= -(\Psi_{j+1} + \Psi_{j+2} + \cdots) \\
\sum_{j=0}^{\infty} &= j \mid \alpha_j \mid < \ \infty
\end{aligned}
$$

The final condition in (10.44) is slightly stronger than absolute summability but is satisfied by stationary $ARMA$ processes.

Notice this equation is composed of three parts:

1. Random Walk: $\Psi(1)(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t)$
2. Initial Conditions: η_0
3. Stationary Process: η_t

This relationship is called the *Beveridge-Nelson Decomposition*.

10.7.1 Implications for a Unit Root Process

Any $I(1)$ process y_t whose first difference is

$$
\triangle y_t = u_t
$$

then

$$
\begin{aligned}
y_t &= u_1 + u_2 + \cdots + u_t + y_0 && (10.45) \\
&= \Psi(1)(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) + \eta_t - \eta_0 + y_0 && (10.46)
\end{aligned}
$$

Since η_t is stationary if (10.45) is divided by $\sqrt{T}$, only the first term (the non-stationary bit) will contribute to the limiting distribution (all other terms go to 0):

10.7.2 Implication of Decomposition for Serially Correlated Processes

Suppose

$$X_T(r) = \frac{1}{T}\sum_{t=1}^{[Tr]} u_t \tag{10.47}$$

where u_t satisfies (10.43) and (10.44) (we also need a finite fourth moment), then

$$\sqrt{T}X_T(r) \to \sigma\Psi(1)\cdot W(r) \tag{10.48}$$

This gives some insight on how limiting distributions can be derived with serially correlated errors in $I(1)$ processes. Notice that

$$\sigma\Psi(1) = (\ 2\pi f(0))^{\frac{1}{2}} \tag{10.49}$$

where $f(0)$ is the spectral density matrix of the filter Ψ at frequency 0.

We see that the distribution depends on the unknown Ψ, (nuisance parameters) and how we remove these is what distinguishes some of the tests. For those that know the augmented Dickey fuller tests (described below), this motivates the correction terms.

10.8 Testing for Unit Roots (*iid* errors)

Under the assumption of **iid errors**,

$$y_t = y_{t-1} + \epsilon_t$$

Dickey [1976] and Fuller [1976] conducted Monte Carlo experiments to compute the critical values of $T(\hat{\phi}-1)$ from (10.11) and the associated "t" statistic of $\hat{\phi}$. They found the distribution is skewed to the left and has many more large negative values relative to the Student$-t$ distribution.

- Notice that the limiting distribution is obtained by multiplying $(\hat{\phi}-1)$ by the sample size T. Fuller [1976, p.371-373] (see also Savin and Evans, 1981) computes critical values for

$$T(\hat{\phi}-1) \tag{10.50}$$

and $\hat{\tau}$ where,

$$\hat{\tau} = \frac{\hat{\phi}-1}{\sum \hat{s}_{t-1}^2} \tag{10.51}$$

and $\hat{s}$ is an unbiased estimator of the variance under the iid assumption. Under the null, this statistic is $O_p(1)$ and has a non-degenerate distribution (called the *Dickey-Fuller* $t-$ *distribution*).

- The statistic $\hat{\tau}$ is just the "t" statistic (which is *not* distributed as t or normal) for $\phi - 1$ in the following regression

$$\Delta y_t = (\phi - 1)y_{t-1} + u_t \tag{10.52}$$

- For the null of $\phi = 1$ against a two sided alternative, Dickey and Fuller [1979] showed that the statistic $\hat{\tau}$ is a monotone function of the likelihood ratio (LR) test. Fuller [1976] (p.371 and p.373) presents modified "t" statistics for the case where the $AR(1)$ process has a constant or a time trend and a constant.

- Likelihood Ratio test statistics for computing hypothesis of unit roots with time trends and constants are discussed in Dickey and Fuller [1981]. Evans and Savin [1981] have obtained analytic results on the exact small sample distribution of $\hat{\phi}$ showing that it depends on y_0/σ. This indicates that the distribution of $\hat{\phi}$ is going to depend on the unknown parameter y_0 which starts the process and the nuisance parameter σ which is also unknown but can be estimated.

- Nuisance parameters are more acute in the non iid case and hence obtaining the limiting distributions of $\hat{\phi}$ and $\hat{\tau}$ are more problematic.

10.8.1 Three Forms of the Dickey-Fuller Test

Consider the following three variations on an $AR(1)$ process. Throughout this particular section we assume that u_t is iid. Hamilton derives each of these limiting distributions for those interested in pursuing the more technical aspects.

The key feature here to understand that the limiting distribution of the tests are different depending upon whether the model has a constant and/or time trend.

$AR(1)$ **process with no constant or time trend:**

$$\Delta y_t = (\phi - 1)y_{t-1} + u_t \tag{10.53}$$

$$H_0 : \phi = 1.0 \quad H_1 : \phi < 1.0$$

To test the null hypothesis of a unit root against the alternative of $\phi < 1$, we estimate 10.53 by OLS and obtain the t statistic for the estimated coefficient $(\hat{\phi} - 1)$.

- From the distribution, documented in Fuller and many other places, the critical "t" value for $\phi = 1.0$ at the 5% level of significance is -1.96 (compared to -1.65 for a standard normal). A large negative t statistic from OLS estimation of 10.53 is evidence against the null hypothesis of a unit root in the series y_t.

AR(1) process with a constant but no time trend:

$$\Delta y_t = c + (\phi - 1)y_{t-1} + u_t \tag{10.54}$$
$$H_0 : \phi = 1.0 \quad H_1 : \phi < 1.0$$

To test the null hypothesis of a unit root, estimate 10.54 by OLS and obtain the t statistic for the estimated coefficient $(\hat{\phi} - 1)$.

Also note under the null hypothesis $c = 0$. From Fuller [1976], $p373$, the critical "t" value for $\phi = 1.0$ at the 5% level of significance is -2.89. Notice that the critical value is much larger for the case where there is a constant (reflecting the added uncertainty) .

AR(1) process with a constant and a time trend

$$\Delta y_t = c + b\,t + (\phi - 1)y_{t-1} + u_t \tag{10.55}$$
$$H_0 : \phi = 1.0 \quad H_1 : \phi < 1.0$$

Under the null it is assumed that $c = b = 0$.To test the null hypothesis of a unit root, estimate 10.55 by OLS and obtain the t statistic for the estimated coefficient $(\hat{\phi} - 1)$. From Fuller the critical "t" value for $\phi = 1.0$ at the 5% level of significance is -3.43. This is larger than either the standard or demeaned case above.

These examples illustrate the wide appeal of Dickey-Fuller tests. In all cases, the Dickey-Fuller test comes down to estimating an OLS regression. If the resulting t statistic on $(\hat{\phi} - 1)$ is a large negative value then we may reject the null hypothesis of a unit root in the data.

10.9 Unit Root Tests with AR(p) Errors

10.9.1 Order of Process is Known

While it is unlikely that the order of the serial correlation is ever known in practice, it is useful to study this first. The results for the simple $AR(1)$ process can be generalized to any $AR(p)$ process provided p is finite. An $AR(p)$ process, with no constant and no time trend can be written as

$$y_t = \phi y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + u_t \tag{10.56}$$

Assuming $u_t \sim N(0, \sigma^2)$, Dickey and Fuller have shown that a test for $\phi = 1$ in 10.56 has the same limiting distribution under H_o as a test for $\hat{\phi}$ in a pure $AR(1)$ process.

- Essentially, we add lags of Δy_{t-i} to reduce u_t to white noise thereby leaving a test statistic that is identical to the one with iid errors. We can see the addition of the Δy_{t-i} as a removal of the nuisance terms identified in (10.48). Note that we are assuming that p, the order of the AR process is **known**. The test is generally applied and is equivalent to 10.57 below.

- This result extends in an obvious fashion (as in 10.54 and 10.55) to models with a constant and/or time trends with the comparison to the appropriate distribution.

10.9.2 Testing unit roots with arbitrary AR structure (p is unknown)-ADF Tests

At this stage we consider p to be finite but for proofs infinite serial correlation is permitted under some memory restrictions. In order to test whether a series y_t is I(1), Dickey and Fuller advocate the following test regression.

$$\Delta y_t = \beta y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + u_t \tag{10.57}$$

where p is **chosen** large enough to ensure the residuals $\hat{\varepsilon}_t$ are empirically white noise and $\beta = (\phi - 1)$. The order p is data dependent and we can choose p

1. Start with a large p, say $p = T^{\frac{1}{4}}$ and then test downward using standard normal test theory (F or t tests) (STATA produces appropriate standard errors–see $dfuller$) .

2. Use model selection criteria, along the same lines as outlined in Chapter 6. Kim (1998, Econometrica 66, pp. 359-380) has suggested the usual BIC (SC) criteria:

$$SC(k) \;=\; ln\, \tilde{\sigma}_k^2 + \frac{k\; lnT}{T} \tag{10.58}$$

is not the appropriate for nonstationary data and presents the alternative:

$$SC(k)^{ns} = (T-k)\ln(\tilde{\sigma}_k^2) + 2\ln(T^k)/T \tag{10.59}$$

10.9.3 Comments and Notes on ADF Tests

1. Visual inspection of the autocorrelation function is useful and should be plotted. The null hypothesis is $H_o : y_t {\sim} I(1)$. A valid test statistic is a test on β obtained by OLS estimation of (10.57). The null hypothesis is rejected if the "t-test" has a negative and significant value.

2. If $\hat{\beta}$ is positive and significant, it means we are rejecting the hypothesis of a unit root, with an indication that the root is outside the unit circle (explosive). Since explosive behaviour is rare in economic time series it is often the case that one-sided tests for unit roots are constructed against the alternative hypothesis of a root inside the unit circle.

3. It is important to recognize that the test statistic *does not* have a t distribution under the null hypothesis of unit root. Critical values for the Dickey-Fuller test are found in Fuller [1976]. If $p = 0$ in (10.57), some authors refer to the t test on $\hat{\beta}$ as a **Dickey-Fuller (DF) test**. For $p > 0$ the t test on $\hat{\beta}$ is called as an **Augmented Dickey-Fuller test (ADF)**. If one is unable to reject the null hypothesis, then estimate 10.57 in first differences. Empirically we are finding that many macroeconomic time series are I(1) or at most I(2).

10.9.4 Calculation of Power

In addition to the three cases considered ($\phi < 1, \phi = 1$ and $\phi > 1$) there is also a branch of the unit roots literature that deals with **roots "near unity"**. This also unifies the asymptotics (local to unity) of the stationary, unit root and explosive case. The issue here is essentially one of **power**.

- In such cases, the standard asymptotic theory still applies if $\phi < 1$, but is likely to provide a poor approximation close to the unit root in small samples.
- The notion of roots near unity can be understood by replacing ϕ with $\phi_T = 1 - \epsilon/T$ where ϵ is an arbitrary small constant. Distribution theory can then be worked out using $1 - \epsilon/T$ in place of ϕ. Unit roots and near unity are discussed in Phillips [1987], Phillips and Perron [1986] and Park and Phillips [1986, 1987].

10.9.5 Non Zero Constants and Trends

One potential point of confusion is that if (10.57) contains a constant term or a time trend, then different critical values (which are also found in Fuller [1976] p.371) are needed. While the time trend and constant are included in the regression, they are developed **under the null** (part of the maintained hypothesis) that the coefficient for the constant is zero.

To understand this, consider :

$$y_t = c + \phi y_{t-1} + u_t \qquad u_t \sim iid(0, \sigma_u^2), \quad c \neq 0$$

This may be expressed in the following form:

$$y_t = ct + y_o + \sum_{i=1}^{t} u_t \qquad (t = 1, 2, \dots T)$$

from which may obtain the expected sampling variability:

- $$E\left[\sum_{t=1}^{T} y_t^2\right] = c^2 \frac{T(T-1)(2T+1)}{6} + y_0 c \frac{T(T+1)}{2} + y_o T + \sigma_u^2 \frac{T(T+1)}{2}$$

- This suggests that the deterministic trend will dominate the unit root for large T as the first component is $O(T^3)$ and the last component is only $O(T^2)$. Therefore, in the long-run, the series would look like a linear trend. A tendency that becomes stronger as c is large relative to σ_u^2. This result also implies that the limiting distribution t-statistic for $(\hat{\phi} - 1)$ in 10.54 (with $c \neq 0$) has a standard normal distribution and not Dickey-Fuller distribution (see West, 1989).

- The normality result holds regardless of whether a unit root is present or not. Importantly, with just a constant in the unit root test equation the test is inconsistent (power against a fixed alternative does not go to one as the sample size increases) against the trend stationary alternative.

- In order to gain power in this direction, a time trend should be included in the regression. The guideline is to add 1 higher power in t than is maintained under H_o.

- Also, this asymptotic normality result is a very poor guide for the small sample distribution unless c is large relative to σ_u^2. Monte Carlo work by Hylleberg and Mizon [1989] question the practical finite sample implications of this finding. In view of this last observation, it is common to see authors first run their data on a constant and a time trend (demeaning and detrending the data) and then to apply the Dickey-Fuller type tests with comparisons against the appropriate critical values. These tests appear to have reasonable small sample properties.

10.10 More on Unit Root and Trend Stationary Process

In a paper in the *Journal of Applied Econometrics,* Peter Phillips (1998) discusses what he calls "reasonable spurious regressions"and the discussion anticipates some material on cointegration in Chapter 11.

- The concept of spurious regression usually arises in the context of a regression of unrelated $I(1)$ series. These are also called "nonsensical correlations".

Suppose the true DGP is

$$y_t = y_{t-1} + u_t \tag{10.60}$$

The researcher fits a linear trend

$$y_t = \widehat{b}\, t + \widehat{u_t} \tag{10.61}$$

Equation 10.61 is a spurious nonsensical correlation since (using results in Dulauf and Phillips, 1988 *Econometrica*) we have

1. $\widehat{b} \to 0$
2. $t_b = O(\sqrt{T}) \Rightarrow$ "t" statsitics are diverging

- Thus even though $\widehat{b}$ is converging to its true value of 0, it is always significant. Our conclusion from such regression results is that (10.61) is misspecified.

Now consider the true DGP to be

$$y_t = b\,t + u_t, \quad u_t \text{ is stationary} \tag{10.62}$$

and the researcher fits

$$y_t = \widehat{a}\,y_{t-1} + \widehat{u_t} \tag{10.63}$$

The following results are obtained

1. $\widehat{a} \to 1$
2. $T(\widehat{a} - 1)$ converges to a constant

- The model does not have unit power against a stationary alternative (power is not going to one when the null is false). This again motivates the inclusion of a time trend in 10.63 since in this case, $\widehat{a} \to 0$.

10.11 Sargan and Bhargava –Durbin Watson Tests

Because the limiting distribution of the DF test is not invariant to either a constant or a time trend, Sargan and Bhargava [1983] have proposed a test for unit roots based on the Durbin-Watson (or Von Neuman ratio) statistic. The Sargan and Bhargava test is essentially a test to see whether the Durbin-Watson (DW) is significantly different from zero (under the null hypothesis). The critical values for this DW based test are found in Sargan and Bhargava [1983].

Based on the work by Anderson [1958] and Durbin and Watson [1952], Sargan and Bhargava have shown that a test for unit roots can be obtained using the distribution of the Von-Neumann ratio.

$$\frac{\sum_{t=1}^{T}(y_t - \phi y_{t-1})}{\sum_{t=1}^{T}(y_t - \bar{y})} \tag{10.64}$$

This ratio is now used in many tests that do permit serially correlated errors (see Elliot et all below).

- Under the null $\phi = 1$, where $\bar{y}$ is the sample mean, this test is uniformly most powerful against a one sided stationary alternative. Hence a major advantage of the test is its invariance to both a time trend and a constant term in the DGP.
- However, the test **lacks generality** in the sense that it is only powerful in discriminating between a simple random walk and the stationary $AR(1)$ process.
- Sargan and Bhargava propose running the following regression

$$y_t = c + u_t$$

and test the null hypothesis that the $\hat{u}_t$ are $AR(1)$ against a stationary $AR(1)$ process.

$$\begin{aligned} H_0 &: \quad u_t = u_{t-1} + \epsilon_t, \epsilon_t \sim IN(0, \sigma^2) \\ H_1 &: \quad u_t = \phi u_{t-1} + \epsilon_t, \epsilon_t \sim IN(0, \sigma^2), \quad \phi < 1 \end{aligned}$$

where $\hat{u}_t' s$ replace u_t.

- The Durbin-Watson statistic will approach zero if the residuals are non-stationary. At the 5% level of significance, the critical Durbin-Watson statistic is .386
- One problem with the DW based test (which is called the CRDW test) is that the power of the test becomes very low as ϕ approaches 1. This is also true for the DF tests, but the loss in power is not as great. Some small sample Monte Carlo work by Granger and Engle [1985], led them to prefer the DF and ADF tests over the DW test.

10.12 Testing for Unit Roots in ARMA Models

- The discussion above shows that the Dickey-Fuller tests in the $AR(1)$ model can, in a straightforward manner, be extended to test for a unit root in the $AR(p)$ process for any finite order p. This is done by augmenting lagged values of the first differences in the test regression.
- Matters become somewhat more complicated when the process is ARMA(p, q) because the corresponding AR process is of an infinite order (see Chapter 3).
- The residuals in **any finite** order augmented Dickey-Fuller test will not be orthogonal to the regressors. In response to this problem, there are at least two strategies that can be taken.

1. Approximation of the infinite AR process by a finite order Dickey-Fuller test regression such as Said and Dickey [1984].
2. Using the robust method of Phillips [1987].

10.12.1 Said and Dickey Approach for ARMA(p, q)

Said and Dickey [1984] generalize the Dickey-Fuller tests to the ARMA(p, q) model by approximating the model as a finite order AR process. OLS is used to estimate the model and the resulting test statistics are shown to have the same limiting distribution as the original Dickey-Fuller statistics reported in Fuller [1976]. Consider the following case

$$\begin{aligned} H_0 &: ARIMA(p,1,q) \\ H_1 &: ARMA(p+1,q) \end{aligned}$$

Suppose that

$$(1-\phi L)y_t = u_t \qquad \gamma(L)u_t = \theta(L)\epsilon_t \tag{10.65}$$

where the polynomials, $\gamma(L)$ and $\theta(L)$ are lag operators. We assume that $\epsilon_t \sim \text{iid}(0, \sigma^2)$ and u_t is stationary.

The polynomials $\gamma(L)$ and $\theta(L)$ are assumed of finite order. If $\phi < 1$, the process $\{y_t\}$ is a stationary ARMA$(p+1, q)$ except for transitory start up effects. But if $\phi = 1$, then the process $\{y_t\}$ is ARIMA$(p, 1, q)$. Under the null hypothesis of a unit root we obtain, $\Delta y_t = u_t$. A test for a unit root for (10.65) can be obtained by OLS estimation of

$$\Delta y_t = (\phi - 1)y_{t-1} - \sum_{j=1}^{m} \alpha_j \Delta y_{t-j} + \epsilon_t \tag{10.66}$$

where m is chosen to be $O(T^{\frac{1}{3}})$. That is, m is increasing in T at **most** at rate $O(T^{\frac{1}{3}})$. Hence $m = T^{\frac{1}{4}}$ rounded to the nearest integer is fine for the theory.

- Monte Carlo evidence has not found much support for mechanical rules for choosing m and that better finite sample test properties are obtained with data dependent methods (model selection criteria or testing of the $\alpha' s$ as we discussed in the finite but unknown p case)

- STATA again can be programmed to handle this

- The OLS estimates from (10.65) of ϕ, γ and θ are all consistent under the null. However, the distribution of $T(\hat{\phi} - 1)$ depends on the nuisance parameters, that is, the $\gamma_i' s$ and the θ_j's.

- The t statistic for $(\phi - 1) = 0$ for (10.66) does not depend on unknown parameters and asymptotically has the original Dickey-Fuller t statistic distribution.

- Qualitatively, the same results can be obtained for a model with a constant and time trend.

10.12.2 Phillips-Perron Approach

This approach, which makes use of a functional central limit theory to obtain robust corrections for infinite dimensional nuisance parameters (see Phillips [1987] and Phillips and Perron [1986]). The problem of infinite nuisance parameters comes about in ARMA representations because any MA process can be written as an infinite AR process.

- Phillips [1987] shows that the Dickey-Fuller tests are affected by autocorrelation in the errors. He develops **a modified Dickey-Fuller test** statistic that has the same asymptotic distribution tabulated by Dickey and Fuller (essentially the statistic is adjusted to take care of the serial correlations in the errors)

- These Phillips-Perron tests allow for autocorrelation and conditional heteroskedasticity. Basically, the Phillips and Perron tests are just adjusted Dickey-Fuller tests where the adjustments involve the autocovariances of the errors from a ARIMA(1,0,0) model.

- The estimation of the impact of the nuisance parameters is nonparametric (hence the unit root tests are often referred to as non parametric unit root tests but this is still a bit of a misnomer)

- The derivation of these Phillips and Perron tests make use of functional central limit theory which we discussed earlier. Here we just present some of the results.

- Consider an ARIMA(1,0,0) process with a constant.

$$y_t = \delta + \phi y_{t-1} + u_t \quad t = 1, 2, \ldots, T, u_t \sim (0, \sigma_t^2) \tag{10.67}$$

  In (10.67) the disturbances, u_t, may be **autocorrelated** and **heteroskedastic** (both conditional and some unconditional). As before there are restrictions on the amount of dependence and the nature of the heteroskedasticity.

- As previously discussed, for the iid case, Dickey and Fuller have derived the distribution for $T(\hat{\phi} - 1)$ and $\hat{\tau} = (\hat{\phi} - 1)/s(\hat{\phi})$ where $s(\hat{\phi})$ is the standard deviation of the estimate of $\hat{\phi}$ calculated by least squares under the iid for u_t.

- Phillips proposes the modified test statistic (called the Z_α tests)

$$Z_\alpha = T(\hat{\phi} - 1) - .5(s_{Tm}^2 - s_u^2)(T\sum_{t=2}^{T}(y_{t-1} - \bar{y}_{t-1})^2)^{-1} \tag{10.68}$$

  with the last term "adjusts" for possible serial correlation and heteroskedasticity. Define

$$s_u^2 \;=\; (T-1)^{-1}\sum_{t=2}^{T}\hat{u}_t^2 \tag{10.69}$$

and the sample variance s_{Tm}^2 (called the long-run variance)

$$s_{Tm}^2 = T^{-1}\sum_{t=2}^{T}\hat{u}_t^2 + 2T^{-1}\sum_{j=1}^{m} w_{jm}\sum_{t=j+}^{T}\hat{u}_t\,\widehat{u_{t-j}} \tag{10.70}$$

and the weights $w_{jm} = (1 - j/(m+1))$ ensure the estimate of the variances s_{Tm}^2 is positive.

- Recall these weights from spectral estimation where (10.70) is employed in the tent weighting scheme.
- This kind of estimation is discussed in Newey and West [1987]. Again, m should increase in T and we may set $m = T^{\frac{1}{4}}$ and take the nearest integer. Recently Newey and West (1994, *Review of Economic Studies*) have suggested a data dependent choice for m (which we are planning to program)

Issues

- If u_t is iid, the second terms in (10.70) are (asymptotically) zero, the second term in (10.68) is zero (since $s_u^2 = s_{Tm}^2$ asymptotically) and the $Z_\alpha = T(\hat{\phi}-1)$ which is the same as Dickey-Fuller.
- Phillips also modifies the regression t statistics $\hat{\tau}$ (called the Z_t tests).

$$Z_t = \hat{\tau}(s_u/s_{Tm}) - .5(s_{Tm}^2 - s_u^2)(s_{Tm}^2 T\sum_{t=2}^{T}(y_{t-1} - \bar{y}_{t-1})^2)^{-1/2} \tag{10.71}$$

where s_{Tm}^2 is given by (10.70). The Z_t under H_o has the same limiting distribution as the $t-$distribution of Dickey-Fuller distribution given earlier. Phillips and Perron [1986] develop adjusted Dickey-Fuller tests for the case where the **alternative** is a stationary ARIMA$(p,0,q)$ process around a deterministic trend.

- If instead of (10.23) we have a model which includes a constant and a time trend, so the ARIMA(1,0,0) is

$$\begin{aligned} y_t &= \delta + \beta\, t + \phi y_{t-1} + u_t \\ t &= 1, 2, \ldots, T, u_t \sim (0, \sigma_t^2) \end{aligned} \tag{10.72}$$

Phillips and Perron [1986] develop the following adjustments to the basic Dickey-Fuller tests $T(\hat{\phi} - 1)$ and $\hat{\tau} = (\hat{\phi} - 1)/s(\hat{\phi})$ in (10.68) and(10.71) respectively.

$$Z_\alpha = T(\hat{\phi} - 1) - (s_{Tm}^2 - s_u^2)(T^6/24D_{XX}) \tag{10.73}$$

and

$$Z_t = \hat{\tau}(s_u/s_{Tm}) - (s_{Tm}^2 - s_u^2)T^3(s_{Tm}4(3D_{XX})^{1/2})^{-1} \tag{10.74}$$

where D_{XX} is the determinant of the regressor cross product matrix in 10.72. The statistics Z_α and Z_t should have the asymptotic distribution tabulated by Dickey and Fuller for ϕ and τ (with a trend and intercept) even when the regression errors in 10.72 are autocorrelated and heteroskedastic.

- All 3 forms of the test can be calculated in STATA using the command *pperon* with options specifying no constant, constant, constant trend and the lag length m in the Newey-West long-run covariance term.
- Evans and Savin [1984] have shown that without correction, the τ statistics depend on the unknown intercept δ in 10.67. They demonstrate that in a model that includes both a constant and a time trend, the inclusion of the time trend makes the distribution of $\hat{\phi}$ independent of δ even if the coefficient of the constant is zero. See the earlier discussion on ADF tests.

Recent Monte Carlo evidence by Schwert [1989] indicates the tests developed by Phillips [1987] and Phillips and Perron [1986] seem more sensitive to model misspecification than the higher order AR approximation suggested by Said and Dickey [1984]. The ADF tests are easy to calculate and have good finite sample properties. It is not surprising that they are the most widely used tests for nonstationarity.

10.13 Increasing the Power of Unit Root Tests: $DFGLS$

Graham Elliott, Thomas Rothenberg, and James Stock wrote *Efficient Tests for an Autoregressive Unit Root*, Econometrica (64(4) July, 1996, pages 813-836). We demonstrate in program unit.do.

10.13.1 The Basic Approach

Suppose the data is generated using the following data generating process:

$$y_t = d_t + u_t$$

$$u_t = \alpha u_{t-1} + v_t \quad v_t \text{ is stationary} \tag{10.75}$$

Here d_t contains the deterministic components of the model (typically includes a constant and/or time trend). The test for a unit root has as its null hypothesis

$$H_0 : \alpha = 1.$$

Now the trick here is the alternative. Consider a specific alternative say $\bar{\alpha} < 1$

$$H_0 : \alpha = \bar{\alpha} < 1$$

- Elliott et al compute the asymptotic bound on the power of unit root tests, and propose a test that comes closer to attaining that bound than most other tests.
- The idea is to find out what values of $\alpha < 1$ have maximal power call it $\bar{\alpha}$. It can be shown that there is a range for $\bar{\alpha}$ that gets very near the maximal power boundary (as we shall see $\bar{\alpha}$ depends on T and whether there is a trend and constant)
- This test does not require knowledge of the nuisance parameters (the parameters in d_t, the initial value of the error u_0, or the distribution of v_t), so is easy to implement.
- It is shown, that if there is no deterministic component ($d_t = 0$ for all t), the standard Dickey-Fuller test comes close to the power bound; if there is a deterministic component (as will usually be the case), the power of the test can be considerably improved by using the proposed testing procedure.

10.13.2 The Test

In order for the test to be valid, there are some conditions on u_0 and $\{v_t\}$ that must be satisfied:

1. variance of u_0 bounded in the neighbourhood around 1.
2. $\{v_t\}$ is stationary, ergodic (unchanging over time), with finite autocovariances: $\gamma(k) = Ev_t v_{t-k}$ where: (a) $\omega^2 = \sum_{k=-\infty}^{\infty} \gamma(k)$ is finite and non-zero;

   (b) $T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} v_t \Rightarrow \omega W_0(s)$, where W is a Wiener process; $0 < s < 1$; $[sT]$ is the largest integer less than or equal to sT.

- These assumptions are standard in this literature, and are satisfied by all stationary and invertible ARMA processes.
- For this testing procedure, the alternative hypothesis is that α takes on a specific alternative value; $H_1 : \alpha = \bar{\alpha}$, where $\bar{\alpha}$ is chosen optimally to maximize the power of the test. By comparison, in most unit root tests, the alternative hypothesis takes the form $H_1 : \alpha < 1$.
- To construct the test statistic we use (following notation on STATA)

$$
\begin{aligned}
\tilde{y}_1 &= y_1 \\
\tilde{y}_t &= y_t - \bar{\alpha} y_{t-1} \quad t = 2, \dots T \\
x_1 &= 1 \\
x_t &= 1 - \bar{\alpha} \quad t = 2, \dots T \\
z_1 &= 1 \\
z_t &= t - \bar{\alpha}(t-1) \quad t = 2, \dots T
\end{aligned}
$$

Estimate by OLS the following (notice that this is like a quasi-difference generalized least sqaures procedure and hence its name):

$$\tilde{y}_t = \delta_0 x_t + \delta_1 z_t + error$$

and denote the estimates respectively as $\hat{\delta}_0$ and $\hat{\delta}_1$ Obtain the residual based on the untransformed variables

$$y_t^* = y_t - (\hat{\delta}_0 \; + \hat{\delta}_1 \, t)$$

Do a standard ADF test using these residuals:

$$\Delta y_t^* = \alpha + \beta y_{t-1}^* + \sum_{j=1}^{p} \gamma_i \Delta y_{t-j}^* + \epsilon_t$$

- The t-test on β has a Dickey Fuller distribution under the null
- The calculation of the test depends on the choice of $\bar{\alpha}$ which in turn depends on whether there is a trend or no trend ($\delta_1 = 0$) in the model
- Clearly there will be as many tests as there are $\bar{\alpha}$, and this needs to be decided. Fortunately, there is a range of choices for $\bar{\alpha}$, that appears to work very well regardless of the true value under the alternative.
- Let π be the asymptotic power of the test. The power curves are not sensitive to $\bar{\alpha}(\pi)$ for $0.25 \leq \pi \leq 0.95$, so $\bar{\alpha}$ can be set at any level on a range with no more than a minimal impact on the results. The authors suggest

  $$\bar{\alpha} = 1 + \frac{\bar{c}}{T}$$

  with $\bar{c} = -7$ when d_t contains just a constant, and $\bar{c} = -13.5$ when it contains a constant and time trend. With these choices, the limiting power functions of the P_T tests are within 0.01 of the power bound for a test size ϵ satisfying $0.01 \leq \epsilon \leq 0.10$.
- To make the test we need to select p and STATA with the command $dfgls$ first calculates a maximum p,Monte Carlo evidence favours:

  $$p_{\max} = int[12(T+1)/100]^{\frac{1}{4}}$$

  and then applying:

1. Ng-Peron sequential $t - statistic$: Test downward sequentially the highest a_p starting with $p_{\max}$ at some significance level (STATA default is 10%) until a a rejection and choose that value for p
2. Schwarz criterion (choose p to minimize)

   $$SIC = \ln(\frac{1}{T - p_{\max}}) \sum_{t=p_{\max}+1}^{T} \hat{\eta}_t + (p+1)\frac{\ln(T - p_{max})}{T - p_{\max}}$$

3. NG-Peron modified Akaike information criteria (see STATA manual)

10.14 Increasing the Power of Unit Root Tests: Additional Regressors

- Another attempt at increasing the power of the unit root tests has been suggested by Bruce E. Hansen ("Rethinking the Univariate Approach to Unit Root testing: Using Covariates to Increase Power", in *Econometric Theory 1995*).
- We briefly discuss the framework behind his approach. As of writing this procedure has not been programmed in STATA.
- The procedure has not really been picked up in the applied literature owing to its poor finite sample performance under the null hypothesis
- However, it does lead nicely into the systems approach of Johansen examined in Chapter 11

10.15 The Basic Argument

Unit root tests have low power partly due to the convention of testing for them in the context of univariate time series, ignoring relevant information in multivariate data sets.

- If we are testing for a unit root in series y_t, assume that there are **related** time-series x_t. Assume further that x_t is I(1). Consider,

$$\Delta y_t = \gamma y_{t-1} + u_t \tag{10.76}$$

  where for convenience of exposition we assume u_t is iid $(0, \sigma_u^2)$ and that Δx_t has mean zero.

- By testing $H_0 : \gamma = 0$, say following Phillips (1987), we test for a unit root, using the OLS t-statistic for γ.
- Consider the following decomposition:

$$e_t = u_t - \Delta x_t b \tag{10.77}$$

  where we rewrite the Δy_t process as:

$$\Delta y_t = \gamma y_{t-1} + \underbrace{\Delta x_t' b + e_t}_{u_t} \tag{10.78}$$

  Essentially we have decomposed u_t into two bits:

  1. New error e_t

2. Explained part $\Delta x_t' b$

- The test of $H_0 : \gamma = 0$, in (10.78) should yield a test with higher power since

$$\sigma_e^2 = \sigma_u^2 - \frac{\sigma_{xu}^2}{\sigma_x^2} \tag{10.79}$$

is typically smaller than in the usual unit root test (σ_u^2).

- If $\sigma_{xu}^2 = 0 \Longrightarrow \Delta x_t'$ cannot explain any of the error u_t

10.15.1 The Distribution for this Test Statistic

- Unfortunately the test becomes a bit of a mess in the serially correlated error case
- Using the notation of Hansen (no longer assuming iid errors), there are three forms of the test considered:

$$a(L)\Delta y_t = \gamma y_{t-1} + b(L)'(\underbrace{\Delta x_t - \mu_{\Delta x}}_{\text{demean } \Delta x_t}) + e_t \tag{10.80}$$

$$a(L)\Delta y_t = \mu + \gamma y_{t-1} + b(L)'(\Delta x_t - \mu_{\Delta x}) + e_t \tag{10.81}$$

$$a(L)\Delta y_t = \mu + \theta t + \gamma y_{t-1} + b(L)'(\Delta x_t - \mu_{\Delta x}) + e_t \tag{10.82}$$

where we define

$$v_t = b(L)'(\Delta x_t - \mu_{\Delta x}) + e_t \tag{10.83}$$

- We note that $b(L)$ may include both leads and lags on x_t. Hansen suggests that in many applications only lag values of x_t will appear.
- There is little practical guidance given in choosing the order of $a(L)$ or $b(L)$ and much work needs to be done before this is routinely applied (the lack of this work suggests the flaws in the procedure)
- Hansen goes on to demonstrate that the asymptotic distribution of such test statistics are not Dickey-Fuller (as we might expect), but rather a combination of the Dickey-Fuller and Normal distributions. In particular, the form of this under the null hypothesis of $\gamma = 0$ (for equation (3)-no mean) is:

$$t(\hat{\gamma}) \to \rho \frac{\int_0^1 W_1 dW_1}{(\int_0^1 W_1^2)^{1/2}} + (1-\rho^2)^{1/2} N(0,1) \tag{10.84}$$

where:

$$\rho^2 = \frac{\sigma_{ve}^2}{\sigma_v^2 \sigma_e^2} \tag{10.85}$$

- Thus the distribution depends upon the unknown ρ (which needs to be estimated to determine the appropriate distribution and hence the appropriate critical value). Note also that when $\rho = 1$ ($\triangle x_t$ is not explaining anything and $v_t = e_t$), this collapses to the ordinary Dickey-Fuller test.

- To apply this test a consistent estimator of ρ is required. This is accomplished by using an estimate of the long-run covariance matrix:

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_v^2 & \hat{\sigma}_{ve} \\ \hat{\sigma}_{ve} & \hat{\sigma}_e^2 \end{bmatrix} = \sum_{k=-M}^{M} w(k/M) \frac{1}{T} \sum_t \hat{\eta}_{t-k} \hat{\eta}_t' \tag{10.86}$$

where:

$$\hat{\eta}_t = \begin{bmatrix} \hat{v}_t & \hat{e}_t \end{bmatrix}^T \tag{10.87}$$

and $w(.)$ is any kernel weight function which produces positive semi-definite covariance matrices, and M is a bandwidth.

- Critical values are computed using Monte-Carlo simulations, for differing values of ρ, and these illustrate higher power than using standard Dicky-Fuller tables, particularly at low values of ρ.

- However, one problem with this test is that in finite samples, there is **more size distortion** than in the standard ADF test, with this test sometimes over and sometimes under rejecting.

10.16 Unit Roots and Seasonality (Optional)

We skip this discussion since it is quite difficult to understand without studying Chapter 8 on the frequency distribution.

The tests of a unit root we have discussed so far are only applicable to time series which are integrated at a non-seasonal frequency. Such a root corresponds to a zero-frequency peak in the spectrum and describes the long memory properties of the series. However, many economic time series exhibit pronounced seasonality, so that they exhibit long memory properties which correspond to peaks at **seasonal frequencies** in the spectrum as well.

- To motivate the discussion of integration in the context of seasonality, we introduce the following definition: Let $A(L)$ have a root with modulus one at frequency λ and also let $B(L)$ have all roots at seasonal frequencies as well as zero frequency.

10.16.1 Definition of Seasonally Integrated

A time series y_t with no deterministic component is **seasonally integrated** of order d at frequency λ, denoted as $y_t \sim I_\lambda(d)$, if λ is the smallest integer for which the following process:

$$A(L)^d B(L) y_t = C(L)\epsilon_t \tag{10.88}$$

has the following properties:

1. $C(e^{i\omega})C(e^{-i\omega})^T$ is finite for all ω and $\{\epsilon_t\}$ is a sequence of white-noise processes
2. $A(L) \Rightarrow$ has a singularity at frequency λ

   $$f_a(\omega) = \sum_{j=0}^{\infty} a_j e^{-i\omega j} \;\; = 0, \text{ for } \omega = \lambda$$

   (actually this is the inverse but we will refer to it as a transfer function for this section).
3. Notice that the typical unit root process studied earlier also shared this singularity property at the 0 frequency

   $$(1 - L) \Rightarrow f_a(\omega) = 1 - e^{-i\omega} = 0 \text{ if } \omega = 0$$

4. $B(L) \Rightarrow$ has no singularity at any frequency

   $$f_b(\omega) = \sum_{j=0}^{\infty} b_j e^{-i\omega j} \;\; \neq 0, \quad \forall\, \omega$$

5. The standard definition of an I(d) process can be subsumed in the definition above for $\lambda = 0$ and $B(L) = 1$. And for $d = 0$ with $B(L) = 1$, the series y_t is an I(0) process.

The definition above allows us to conclude that if $y_t \sim I_\lambda(1)$ with $B(L) = 1$, then

1. $\lim Var\,[y_t] \to \infty$
2. y_t has long memory so that the innovation has a permanent impact on the seasonal pattern of y_t.

   (a) The spectrum of y_t is given by:

   $$f(\omega) = \frac{\overline{D}}{(\omega - \lambda)} 2 \tag{10.89}$$

for $\omega \to \lambda$, so that it has an infinite peak at frequency λ.

3. y_t is asymptotically orthogonal to processes which have unit roots at other frequencies.

Now for simplicity we will consider the value $d = 0$ and $d = 1$ in this note. The following discussion shows how we may **nests** tests of unit roots at the 0 frequency **and** tests for unit roots at other (seasonal) frequencies as well.

10.16.2 Seasonal Unit Roots and their Cycles

We will take quarterly periodicity as a benchmark case for discussion (there are similar versions of the tests for monthly data which are more complicated):

$$(1 - L^4)y_t = \epsilon_t \tag{10.90}$$

which can be rewritten as:

$$\begin{aligned}(1 - L^4)y_t &= (1 - L)(1 + L + L^2 + L^3)y_t = (1 - L)w_{1t} \\ &= (1 + L)(1 - L + L^2 - L^3)y_t = (1 + L)w_{2t} \\ &= (1 + L^2)(1 - L^2)y_t = (1 + L^2)w_{3t}\end{aligned}$$

where the $w_{it}, (i = 1, 2, 3)$ are defined from above. Essentially we have rewritten $(1 - L^4)y_t$ in three alternative ways since

$$(1 - L^4) = (1 - L)(1 + L)(1 + L^2) \tag{10.91}$$

The $(1 - L^4)y_t$ has four roots with modulus one:

1. one at zero frequency
2. one at two cycles per year
3. two complex pairs at one cycle per year.

Let $2\pi = 1$ quarter. If there is a seasonal unit root then $(1 - L^4)y_t$ is stationary. Using the decomposition in 10.91

1. $(1 - L)$ in the frequency domain implies a transfer function (its inverse from Chapter 8)

$$\begin{aligned}f_a(\omega) &= 1 - e^{-i\omega} \\ &= 1 - \cos(\omega) + i\sin(\omega)\end{aligned}$$

If $\omega = 0 \Rightarrow f_a(0) = 0$ and that $\frac{0}{2\Pi} = 0$ cycles per year
2. $(1 + L)$ in the frequency domain implies a transfer function

$$\begin{aligned} f_a(\omega) &= 1 + e^{-i\omega} \\ &= 1 + \cos(\omega) + i\sin(\omega) \end{aligned}$$

If $\omega = \Pi \Rightarrow f_a(\Pi) = 0$ and that $\frac{\Pi}{2\Pi} = \frac{1}{2}$ cycle per quarter or 2 cycles per year.
3. $(1 + L^2)$ in the frequency domain implies a transfer function

$$\begin{aligned} f_a(\omega) &= 1 + e^{-i2\omega} \\ &= 1 + \cos(2\omega) - i\sin(\omega) \end{aligned}$$

There are 2 solutions $(i, -i)$

1. $\omega = \frac{\Pi}{2} \Rightarrow f_a(\frac{\Pi}{2}) = 0 \Rightarrow \frac{\frac{\Pi}{2}}{2\Pi} = \frac{1}{4}$ cycle per quarter (1 per year)
2. $\omega = \frac{3\Pi}{2} \Rightarrow f_a(\frac{3\Pi}{2}) = 0 \Rightarrow \frac{\frac{3\Pi}{2}}{2\Pi} = \frac{3}{4}$cycle per quarter.

These two solutions are indistinguishable and is known as the *aliasing problem.* The idea is that the higher frequency (faster) movements get "folded" or aliased into the lower frequency (slower).

Hence using the above definition, y_t in (10.91), can be denoted as $y_t \sim I_\lambda(1)$ for $\lambda = 0, 1/2$, and $1/4$.

Based on this factorization, Hylleberg, Engle, Granger and Yeo (1990) and Otto and Wirjanto (1988) derive tests for unit roots at some or all seasonal frequencies and/or zero frequency. Hylleberg et al use a somewhat complicated function in approximation theory to obtain the test regression from which the test-statistics can be computed. The following discussion which is asymptotically equivalent and is quite simple based on the Lagrange-Multiplier (LM) principle which can be computed via an artificial regression.

To motivate the seasonal unit-root tests, we assume that the model under the null hypothesis is given by:

$$(1 - L^4)(1 - \alpha_1 L - \ldots - \alpha_p L^p)y_t = \epsilon_t \tag{10.92}$$

that is, we set

$$A(L)^d = (1 - L^4)$$

$$B(L) = (1 - \alpha_1 L - \ldots \alpha_p L^p),$$

and $C(L) = 1$ in the earlier definition. The test regression takes the following form:

$$\begin{aligned}\hat{\epsilon}_t &= \pi_1 L w_{1t} + \pi_2(-L) w_{2t} \\ &+ \pi_3(L^2) w_{3t} + \pi_4(-L) w_{4t} \\ &+ \sum_{i=1}^{p} \mu_i L \hat{\varepsilon}_t + \text{ residuals}\end{aligned}$$

where $L^i w_{1t} = w_{1t-i}$, etc. or, using the LM principle,

$$\begin{aligned}\hat{\epsilon}_t &= \pi_1 L w_{1t}^* + \pi_2(-L) w_{2t}^* \ + \pi_3(-L^2) w_{3t}^* + \pi_4(-L) w_{4t}^* \\ &+ \sum_{i=1}^{p} \phi_i (1 - L^4) y_{t-1-i} + \text{ residuals}\end{aligned} \tag{10.93}$$

where: $Lw_{1t}^* = \hat{\alpha}(L) L w_{1t}, (-L) w_{2t}^* = \hat{\alpha}(L)(-L) w_{2t}$, and so on, $\hat{\epsilon}_t$ is the estimated residuals from model (10.92). Note that although test-statistics computed from these two artificial regressions are asymptotically equivalent, the artificial regression (10.93) would be less susceptible to misspecification than the artificial regression (10.93).

To derive the test-statistics we make use of two facts: (i) unit roots at different frequencies are asymptotically orthogonal; and (ii) a unit root at a given frequency is asymptotically orthogonal to any lags of y_t that are stationary. Thus, a test of a unit root at certain frequency would be invariant to the inclusion of other lags of y_t irrespective of whether they have a unit root or not.

Note also that a deterministic component can be added to the artificial regressions 10.92 and 10.93 where the deterministic component is supposed to capture the purely deterministic processes that may be contained in y_t and is of the following form:

$$c_t = \beta_0 + \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 t \tag{10.94}$$

where D_{it} is the quarterly dummies and t is the time trend.

Test Results (all distributions use nonstandard critical values)

1. To test for a unit root at zero frequency (i.e.$y_t \sim I_0(1)$) simply perform a t-test on $\pi_1 = 0$.

2. To test for a bi-annual cycle root (i.e. $y_t \sim I_{1/2}(1)$) a t-test on $\pi_2 = 0$ is performed.

3. To test for an annual cycle root (i.e. $y_t \sim I_{1/4}(1)$) we can perform either a joint F-test of $\pi_3 = \pi_4 = 0$, or two sequential t-tests of $\pi_4 = 0$ and then $\pi_3 = 0$.

Notes

1. A time series y_t contains no seasonal unit roots if the t-test of $\pi_2 = 0$ and the F-test of $\pi_3 = \pi_4 = 0$ are **both** rejected.

2. The series contains no unit roots and is stationary in a wide sense if each of the t-tests of $\pi_j = 0\,(j = 1, 2)$ are rejected, as well as the F-test of $\pi_3 = \pi_4 = 0$.

Intuition behind Tests

The regressand is essentially y_t after it has been filtered in two ways, by $\hat{\alpha}(L)$ and by the seasonal filter $(1 - L^4)$. If there are unit roots at all frequencies, then applying $(1 - L^4)$ is clearly appropriate and it will remove all four unit roots. In 10.93 each frequency is examined individually.

Consider the first regressor Lw^*_{1t}, we see that y_t is filtered by $\hat{\alpha}(L)$ and also to remove unit roots at frequencies other than zero. If the series does have a unit root at zero frequency, then $(1 - L^4)$ will have removed it from the regressand, but Lw^*_{1t} will still have the zero-frequency unit root. Thus the least-squares estimator of π_1 will be asymptotically distributed as a non-standard unit root distribution. Under the null hypothesis of a zero-frequency unit root and hence the t-test on $\pi_1 = 0$ is the standard test for a unit root, with critical values are tabulated in Fuller (1976).

Similar arguments apply to the other three regressors. Thus the Dickey-Fuller type tests for a unit root at zero frequency can be subsumed in the general tests we derive here as a special case.

The actual asymptotic distribution of the test statistics has been derived by Chan and Wei (1988, Annal of Statistics) and its finite-sample distribution by Hylleberg (1988, unpublished). Critical values for the test statistics were computed from a Monte Carlo experiment using 24,000 replications, based upon the data generating process $(1 - L^4)y_t = \epsilon_t$, where ϵ_t is NID(1,0) and are contained in HEGY (1990).

10.17 Summary

At the present time, the literature on unit roots is still in a state of development. There are a number of tests available for testing non-stationarity in time series but not a great deal known about how sensitive these tests are to various forms of misspecification. For instance, the Dickey-Fuller tests are derived under the assumption of IID$(0, \sigma^2)$ errors. Many time series, especially financial data, display various forms of serial correlation or ARCH behaviour. One would then be interested in the performance of these tests under non-homoskedastic errors. Of course the tests developed by Phillips and Perron are invariant to some forms of serial correlation and conditional heteroskedasticity, but these tests are also very sensitive to other forms of model misspecification (especially regarding the appropriate lag lengths and weights).

Moreover, all of the tests suffer severe power loss when used to test against alternatives with roots near unity. The reason these test work is because the null of a unit root is testable only in the sense that the unit root dominates the asymptotic behaviour of a stationary time series. When a root is very close to unity the tests often fail to reject the null of a unit root. This means that unnecessary transformations for stationarity are conducted.

10.18 Some Additional References

Anderson,T.W. [1942], "Distribution of the Serial Correlation Coefficients," *Annals of Mathematical Statistics*, 13, pp.1-13.

Bhargava, A. [1986], "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 53, pp.369-384.

Cressie, U. [1988], "A Graphical Procedure for Determining Nonstationarity in Time Series," *JASA*,83,404, pp.1108-1116.

Crowder, M.J. [1980], "On the Asymptotic Properties of Least Squares Estimators in Autoregressions," *Annals of Statistics*, pp. 132-146.

Dickey, D. and W. Fuller [1979], "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *JASA*, 74, pp.427-431.

__________________________[1981], "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, pp.1057-1-72.

Durbin,J. and G.S. Watson [1950], "Testing for Serial Correlation in Least Squares Regressions," *Biometrika*, 37, pp. 409-428.

Engle,R.F. and C.W.J. Granger [1987], "Cointegration and Error Correction:Representation and Testing," *Econometrica*, 55, pp.251-276.

Evans,G. and N. Savin [1981], "Testing for Unit Roots," *Econometrica*, 49, pp.753-779.

Evans,G. and N. Savin [1981], "Testing for Unit Roots:2," *Econometrica*, 52, pp.1241-1269.

Fuller, W.A. [1976], *Introduction to Statistical Time Series*, New York:John Wiley and Sons.

Graham Elliott, Thomas Rothenberg, and James Stock [1996] Efficient Tests for an Autoregressive Unit Root, *Econometrica* 64 813-836.

Koop, G. [1991], "Intertemporal Properties of Real Output: A Bayesian Analysis", *Journal of Business and Economic Statistics*, 9, pp. 253-264.

Granger, C.W.J. and R.F. Engle [1985], "Dynamic Model Specification with Equilibrium Constraints: Cointegration and Error Correction," UCSD Working Paper.

Hannan, E.J. and C.C. Heyde [1972], "On Limit Theorems for Quadratic Functions of Discrete Time Series," *Annals of Mathematical Statistics*, 43, pp.2058-2066.

Hansen, Bruce, (1995) Rethinking the Univariate Approach to Unit Root testing: Using Covariates to Increase Power, *Econometric Theory*, 11, 1148-1171.

Harvey, A.C. [1981], *Time Series Models*, Phillip Allan, Oxford.

Hylleberg S., Engle, R.F., Granger, C.W.J. and B.S. Yoo [1990], "Seasonal Integration and Cointegration", *Journal of Econometrics*, 44, pp. 215-238.

Hylleberg, S. and G. E. Mizon, [1989] "A Note on the Distribution of the Least Squares Estimator of a Random Walk with Drift," *Economics Letters*, 29, pp.225-230.

King,R., Plosser,C.I., Stock,J. and M. Watson [1987], "Stochastic Trends and Economic Fluctuations," Working Paper 2229 NBER, Cambridge MA.

Nelson,C.R and C.I. Plosser [1982], "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics* pp.139-142.

Newey, W.K. and K.D. West [1987], "A Simple Positive Semi-definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, pp.703-708.

Phillips, P.C.B. [1987], "Time Series Regressions with Unit Roots," *Econometrica, 55, pp.271-302.*

Phillips, P.C.B. [1988], "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56, pp. 1021-1044.

Phillips, P.C.B. and Zhijic Xiao [1998] "A Primer on Unit Root Testing", *Journal of Economic Surveys*, 12, pp. 423-469.

Phillips, P.C.B. and S.N. Durlauf [1986], "Multiple Time Series Regressions with Integrated Processes," *Review of Economic Studies*, 53, pp.473-496.

Phillips, P.C.B. and P. Perron [1986], "Testing for a Unit Root in Time Series Regressions," University of Montreal D.P. #8633.

Said,E.S. and D.A. Dickey [1984], "Testing for Unit Roots in Autoregressive-moving Average Models of Unknown Order," *Biometrika*, 71, pp.599-607.

Sargan, J.D. and A. Bhargava [1983], "Testing Residuals from Least Squares Regressions for Being Generated by the Gaussian Framework," *Econometrica*, 51, pp.153-74.

Schwert, G.W. [1989], "Testing for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics* , 7, pp.147-159.

Spanos, A. [1986], *Statistical Foundations of Econometric Modelling*, Cambridge University Press.

Stock,J.H. [1987], "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors,"*Econometrica,* 55, pp.1035-1056.

Stock, J.H. and M.W. Watson [1988], "Testing for Common Trends," *JASA*, 85, 404, pp.1097-1107.

West, K. [1988], "Asymptotic Normality, when the Regressors have a Unit Root," *Econometrica*, 56, pp.1397-1417.

White, J.S. [1958], "The Limiting Distribution of The Serial Correlation Coefficients in the Explosive Cases," *Annals of Mathematical Statistics*,29, pp.1188-1197.