Non-Stationary Series.

There are three important types of time series which one is likely to find in financial econometrics:

Stationary.

A (weakly) stationary time series has a constant mean, a constant variance and and the covariance is independent of time. Stationarity is essential for standard econometric theory. Without it we cannot obtain consistent estimators.

A quick way of telling if a process is stationary is to plot the series against time. If the graph crosses the mean of the sample many times, chances are that the variable is stationary, otherwise that is an indication of persistent trends away from the mean of the series.

Trend Stationary

A trend stationary variable is a variable whose mean grows around a fixed trend. This provides a classical way of describing an economic time series which grows at a constant rate. A trend-stationary series tends to evolve around a steady, upward sloping curve without big swings away from that curve. Detrending the series will give a stationary process. For simplicity assume that the following process.

$$y_t = \alpha + \mu t + \varepsilon_t \ where \ \varepsilon_t \tilde{} N(0, \sigma^2)$$

Notice that the mean of this process varies with time but the variance is constant.

$$E(y_t) = \alpha + \mu t$$

$$V(y_t) = E(\alpha + \mu t + \varepsilon_t - (\alpha + \mu t))^2 = \sigma^2$$

Notice that if you define a new variable, say $y_t^*, y_t^* = y_t - (\alpha + \mu t)$ then y_t is stationary.

Unit roots

An autoregressive process of order p, AR(p), has a unit root if the polynomial in L, $(1 - \phi_1 L - ...\phi_p L^p)$ has a root equal to one.

A Random Walk.

The simplest example of a process with a unit root is a random walk, i.e.,

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

where ε_t is i.i.d. with zero mean and constant variance.

Lagging the process one period we can write

$$y_{t-1} = y_{t-2} + \varepsilon_{t-1}$$

Substituting back in equation (1) we get

$$y_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

Repeating this procedure we can easily show that

$$y_t = y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t$$

Mean of a RW process

If y_0 is fixed the mean is constant over time

$$E(y_t) = E(y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t) = y_0$$

The variance of y_t , "conditional" on knowing y_0 , can be computed as

$$\begin{aligned} V(y_t) &= V(y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t) \\ &= V(\varepsilon_1) + V(\varepsilon_2) + ... + V(\varepsilon_{t-1}) + V(\varepsilon_t) = t\sigma^2 \end{aligned}$$

As we move further into the future this expression becomes infinite. We conclude that the variance of a unit root process is infinite.

A unit root process will only cross the mean of the sample very infrequently, and the process will experience long positive and negative strays away from the sample mean.

A process that has a unit root is also called **integrated of order one**, denoted as I(1). By contrast a stationary process is an **integrated of order zero** process, denoted as I(0).

$$\text{ARIMA}(p, d, q)$$

A common practice using the Box and Jenkins methodology was just to take first differences of the series and analyze de differenced process. We can then define an ARIMA(p, d, q) process as

$(1 - \phi_1 L - ...\phi_p L^p)(1 - L)^d y_t = (1 + \phi_1 L - ... + \phi_q L^q)\varepsilon_t$

How do we test for a unit root?

In this section we will show that the standard t-test cannot be applied for a process with unit root. The standard testing procedure that we use for stationary series yields a degenerate distribution. We also find the distribution

under these circumstances but it turns out that is not a t-distribution and that is biased to the left.

Consider the following model

$$y_t = \alpha + y_{t-1} + \varepsilon_t \tag{2}$$

where ε_t "is assumed" to be N(0,σ^2). It can easily be shown that asymptotically

$$\sqrt{T}(\hat{\alpha}_T - \alpha) \ \underrightarrow{L} \ N(0, (1-\alpha^2))$$

If we where able to use this distribution for $\alpha = 1$ then

$$\sqrt{T}(\hat{\alpha}_T - \alpha) \ \underrightarrow{P} \ 0$$

Even though this is valid, it is not very useful for hypothesis testing.

To obtain a non-degenerate asymptotic distribution for $\hat{\alpha}_T$ in the unit root case, it turns out that we have to multiply by T and not by the square root of T. Then the unit root coefficient converges at a faster rate T than for the stationary case.

To get a better sense of why scaling by T is necessary when the true value of α is unity consider

$$\hat{\alpha}_T - \alpha = \frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

then

$$T(\hat{\alpha}_T - \alpha) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}$$

Now, under the null that $\alpha = 1$, y_t can be written as

$$\begin{aligned} y_t &= y_0 + \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t \\ &= \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t \qquad if\ we\ assume\ y_0 = 0. \end{aligned}$$

Then under the null that $\alpha = 1$, y_t ~N$(0, \sigma^2 t)$. To find the distribution of the *numerator* we need to do some easy but tedious algebra. We start by noting that under the null,

$$y_t^2 = (y_{t-1} + \varepsilon_t)^2 = y_{t-1}^2 + 2y_{t-1}\varepsilon_t + \varepsilon_t^2$$

and rearranging terms we obtain

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

Then, the sum which appears in the *numerator* can be expressed as

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

Then, recalling that $y_0 = 0$ and multiplying by (T^{-1}) we obtain the expression of the *numerator* as the sum of two terms

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

To find the distribution of this expression we divide each side by σ^2 which yields the following result

$$(\sigma^2 T)^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t = (1/2)(\frac{y_T}{\sigma\sqrt{T}})^2 - \sum_{t=1}^{T}(\frac{1}{2\sigma^2 T})\varepsilon_t^2$$

Consider the first term of this expression. Since we have shown above that y_t ~N$(0, \sigma^2 t)$, standardizing we obtain

$$(y_T/\sigma\sqrt{T})\tilde{}N(0,1),$$

and then squaring this expression we find that the first term of the numerator is distributed Chi-square

$$(y_T/\sigma\sqrt{T})^2\tilde{}\chi^2(1).$$

It can be shown using the law of large numbers that the second term converges in probability to σ^2, i.e.

$$(1/T)\sum_{t=1}^{T}\varepsilon_t^2 \quad \underrightarrow{P} \quad \sigma^2, \qquad or \qquad (1/\sigma^2 T)\sum_{t=1}^{T}\varepsilon_t^2 \quad \underrightarrow{P} \quad 1$$

If we put both results together we can see that the numerator converges to

$$(\sigma^2 T)^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \qquad \underrightarrow{L} \qquad (1/2)(X-1) \quad where \quad X\tilde{}\chi^2(1).$$

It can also be shown using the law of large numbers that the the *denominator* converges in probability to

$$E(T^{-2}\sum_{t=1}^{T} y_{t-1}^2).$$

Now,as y_{t-1} ~N$(0, \sigma^2(t-1))$, then $E(y_{t-1}^2) = \sigma^2(t-1)$. Therefore the expected value of the denominator can be written as

$$E(T^{-2}\sum_{t=1}^{T} y_{t-1}^2) = T^{-2}\sigma^2\sum_{t=1}^{T}(t-1) = \sigma^2 T^{-2}(T-1)T/2$$

Then if we multiply $(\hat{\alpha}_T - \alpha)$ by T instead than by $\sqrt{T}$, we obtain a non-degenerate asymptotic distribution, but this distribution is not gaussian.

A convenient way of finding the distributions of processes with unit roots is using continuous time stochastic processes defined below.

Brownian Motion

To show what is a BM we may start first considering a simple RW without drift.

$$y_t = y_{t-1} + \varepsilon_t, \quad \varepsilon_t \tilde{} N(0,1)$$

or

$$y_t = \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t \quad and \quad y_t \tilde{} N(0,t)$$

Now consider the change between t and s

$$y_s - y_t = \varepsilon_t + \varepsilon_{t+1} + ... + \varepsilon_s$$

is $\tilde{} N(0, s-t)$ and is independent of changes between dates r and q whenever $t < s < r < q$.

Consider now the change between y_t and y_{t-1} with $\varepsilon_t \tilde{} N(0,1)$, and suppose we view ε_t as the sum of two independent Gaussian variables.

$$\varepsilon_t = e_{1t} + e_{2t} \quad with \quad e_{it} \tilde{} N(0, 1/2)$$

We may associate e_{1t} with the change between y_{t-1} and y_t at some interim point, say $y_{t-1/2}$, such that

$$y_{t-1/2} - y_{t-1} = e_{1t}$$

and

$$y_t - y_{t-1/2} = e_{2t}$$

Sampled at an integer $y_t - y_{t-1}$ ~N(0,1), but we can consider n possible divisions as above such that,

$$y_t - y_{t-1} = e_{1t} + e_{2t} + ... + e_{Nt}$$

where e_{it} ~iid N(0,1/N) .

The limit when $N \to \infty$ is a continuous process called a Standard Brownian Motion. The value this process takes at date t is denoted $W(t)$. A realization of a continuous time process can be viewed as a stochastic function, denoted $W(.)$ where $W : t \in [0, \infty) \to R$.

Definition of a SBM

$W(.)$ is a continuous time process,associating each date $t \in [0,1]$ with the scalar $W(t)$ such that

(*a*) $W(0) = 0$

(*b*) for any dates $0 < t_1 < t_2 ... < t_k < 1$,

$W(t_2) - W(t_1), ..., W(t_k) - W(t_{k-1})$.
are independent multivariate gaussian with $W(t_s) - W(t_t)$~N(0,$s - t$)

(*c*) $W(t)$ is continuous with probability 1.

The functional Central Limit Theorem

Recall the simplest version of the Central limit Theorem.

If ε_t~iid with mean zero and variance σ^2, then the sample mean $\tilde{\varepsilon}_t = T^{-1}\sum_{t=1}^{T}\varepsilon_t$ and the central limit theorem states that

$$\sqrt{T}\tilde{\varepsilon}_T \quad \underrightarrow{L} \quad N(0,\sigma^2)$$

Consider now an estimator based on the following principle: Given a sample size T, we calculate the mean of the first half of the sample and throw out the rest of the observations.

$$\widetilde{\varepsilon}_{[T/2]^*} = ([T/2]^*)^{-1}\sum_{t=1}^{[T/2]}\varepsilon_t$$

where $[T/2]^*$ is the larger integer $\geq$ than $T/2$, i.e.

$$[T/2]^* = T/2 \; for \; T \; even$$

$$[T/2]^* = (T-1)/2 \; for \; T \; odd.$$

This will also satisfy

$$\sqrt{[T/2]^*}\tilde{\varepsilon}_{[T/2]^*} \quad \underrightarrow{L} \quad N(0,\sigma^2)$$

Moreover the estimator will be independent of an estimator that uses only the second half of the sample.

More generally we can construct a Variable $X_T(r)$ from the sample mean of the first r^{th} fraction of observations, where $r \in [0,1]$ defined by

$$X_T(r) = \frac{1}{T}\sum_{t=1}^{Tr^*}\varepsilon_t$$

For any given realization $X_T(r)$ is a step function in r, with

$$X_T(r) = \begin{cases} 0 & \text{for } 0 \leq r < (1/T) \\ \varepsilon_1/T & \text{for } (1/T) \leq r < (2/T) \\ (\varepsilon_1+\varepsilon_2)/T & \text{for } (2/T) \leq r < (3/T) \\ (\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4+...+\varepsilon_T)/T & \text{for } r = 1 \end{cases}$$

then

$$\sqrt{T}X_T(r) = (1/\sqrt{T})\sum_{t=1}^{Tr^*}\varepsilon_t = (\sqrt{Tr^*}/\sqrt{T})(1/\sqrt{Tr^*})\sum_{t=1}^{Tr^*}\varepsilon_t$$

but

$$(1/\sqrt{Tr^*})\sum_{t=1}^{Tr^*}\varepsilon_t \quad \underrightarrow{L} \quad N(0,\sigma^2) \quad while \quad (\sqrt{Tr^*}/\sqrt{T}) \to \sqrt{r}$$

Hence the asymptotic distribution of $\sqrt{T}X(r)$ is that of $\sqrt{r}$ times a $N(0,\sigma^2)$ or $\sqrt{T}X_T(r) \underrightarrow{L} N(0,\sigma^2 r)$.

Clearly this implies that

$$\frac{\sqrt{T}X_T(r)}{\sigma} \quad \underrightarrow{L} \quad N(0,r).$$

Notice that this is evaluated at a given value r (NB $\frac{\sqrt{T}X_T(r)}{\sigma}$ is a random variable).

We could consider the behaviour of the sample mean based on Tr_1^* through Tr_2^* for $r_2 > r_1$ and conclude that

$$\frac{\sqrt{T}(X_T(r_2) - X_T(r_1))}{\sigma} \quad \underrightarrow{L} \quad N(0, r_2 - r_1)$$

Then a sequence of stochastic functions $\frac{\sqrt{T}(X_T(.)}{\sigma}|_{T=1}^{\infty}$ has an asymptotic probability law that is described by a standard brownian motion.

$$\frac{\sqrt{T}(X_T(.))}{\sigma} \quad \underrightarrow{L} \quad W(.) \ (this\ is\ a\ random\ function)$$

This function evaluated at $r = 1$ is just the sample mean, $X_T(1) = T^{-1}\sum_{t=1}^{T}\varepsilon_t$, then when $r = 1$ the CLT is a special case of this function, that is $\frac{\sqrt{T}(X_T(1))}{\sigma} \quad \underrightarrow{L} \quad W(1) = N(0,1)$

Convergence in Functions

Let $S(.)$ represent a continuous time stochastic process with $S(r)$ representing its value at some date r for $r \in [0,1]$. Suppose further that for any given realization $S(.)$ is a continuous function of r with probability 1, being $\{S_T(.)\}|_{T=1}^{\infty}$ a sequence of such continuous functions,

We say that

$S_T(.) \underrightarrow{L} S(.)$ whenever the following holds:

a) For any finite collection of k particular dates $0 < r_1 < r_2 .... < r_k \leq 1$, the sequence of k dimensional random vectors $y_T \xrightarrow{L} y$ where

$$y_T \equiv \begin{bmatrix} S_T(r_1) \\ S_T(r_k) \end{bmatrix} \quad and \; y \equiv \begin{bmatrix} S(r_1) \\ S(r_k) \end{bmatrix}$$

b) For each $\varepsilon > 0$ the probability that $S_T(r_1)$ differs from $S_T(r_2)$ for any dates r_1 and r_2 within a distance δ of each other goes to zero uniformly in T as $\delta \to 0$..

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

This definition applies to sequences of continuous functions though $X_T(r)$ is a discontinuous step function. Fortunately the discontinuities occur at countable points.

Convergence for a sequence of Random Functions.

It will be helpful to extend the earlier definition of convergence in probability to sequences of random functions.

Let $\{S_T(.)\}|_{T=1}^{\infty}$ and $\{V_T(.)\}|_{T=1}^{\infty}$ denote sequences of random continuous functions with

$$S_T : r \in [0,1] \to R, \quad V_T : r \in [0,1] \to R.$$

Let the scalar Y_T represent the largest amount by which $S_T(r)$ differs from $V_T(r)$.

$$Y_T = Sup|S_T(r) - V_T(r)|$$

Then $\{Y_T\}|_{T=1}^{\infty}$ is a sequence of random variables and we could talk about its probability limit. If the sequence converges in prob to 0, then we can say that

$$S_T(r) \quad \xrightarrow{P} \quad V_T(r) \quad or \quad Sup|S_T(r) - V_T(r)| \quad \xrightarrow{P} \quad 0.$$

This can be generalized to sequences of functions. For example, if $\{S_T(.)\}|_{T=1}^{\infty}$ and $\{V_T(.)\}|_{T=1}^{\infty}$ are sequences of continuous functions with $V_T(.) \xrightarrow{P} S_T(.)$ and $S_T(.) \xrightarrow{L} S(.)$, for $S(.)$ a continuous function, then $V_T(.) \xrightarrow{L} S(.)$

Continuous Mapping Theorem

If $g(.)$ is a continuous functional, which could associate a real variable y with the stochastic function $S(.)$, then the theorem states that if $S_T(.) \xrightarrow{L} S(.)$ and $g(.)$ is a continuous functional, then $g(S_T(.)) \xrightarrow{L} g(S(.))$

(Examples of continuous functionals)
$y = \int_0^1 S(r)dr$ or $y = \int_0^1 [S(r)]^2 dr$

Example
Given $\frac{\sqrt{T}X_T(.)}{\sigma} \underrightarrow{L} W(.)$, we can simply use the theorem to get

$$\sqrt{T}X_T(.) \quad \underrightarrow{L} \quad \sigma W(.)$$

or

$$[\sqrt{T}X_T(.)]^2 \quad \underrightarrow{L} \quad [\sigma W(.)]^2$$

Applications to Unit Root Processes

Example: a Random Walk.

$$y_t = y_{t-1} + \varepsilon_t \qquad \varepsilon_t \tilde{} iid(0, \sigma^2)$$

or

$$y_t = \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1} + \varepsilon_t$$

this can be used to express the stochastic function

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

Notice that the t^{th} rectangle has width $1/T$ and height y_{t-1}/T. The total area is then

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

Multiplying both sides by $\sqrt{T}$

$$\int_0^1 \sqrt{T}X_T(r)dr = T^{-3/2} \sum_{t=1}^{T} y_{t-1}$$

But we know from the continuous mapping theorem that as $T \to \infty$

$$\int_0^1 \sqrt{T}X_T(r)dr \quad \underrightarrow{L} \quad \int_0^1 [\sigma W(r)]dr$$

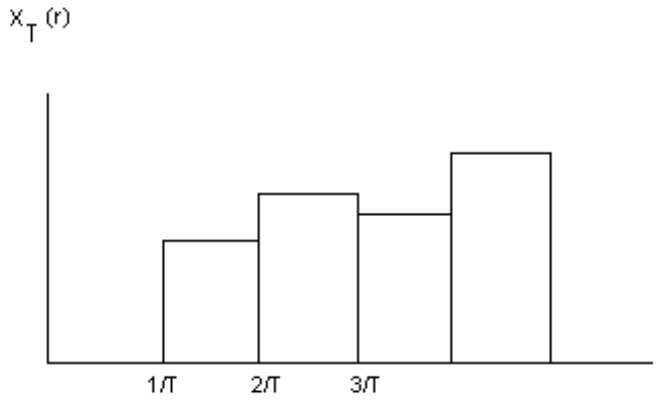

Figure 1:

(since $[\sqrt{T}X_T(.)] \quad \underrightarrow{L} \quad [\sigma W(.)]$), implying

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1} \quad \underrightarrow{L} \quad \int_0^1 [\sigma W(r)]dr$$

Consider now the following autoregressive process

$$y_t = \rho y_{t-1} + \varepsilon_t \qquad \varepsilon_t \tilde{} iid(0, \sigma^2)$$

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

then and assume that we want to test the null

$H_0)$ $\rho = 1$.

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

Then the distribution of the denominator can be easily obtained since

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \quad \underrightarrow{L} \quad \int_0^1 [\sigma W(r)]^2 dr$$

becuase

$$[\sqrt{T}X_T(.)] \underrightarrow{L} [\sigma W(.)],$$

$$[\sqrt{T}X_T(.)]^2 \underrightarrow{L} [\sigma W(.)]^2,$$

$$\int_0^1 [\sqrt{T}X_T(r)]^2 dr \quad \underrightarrow{L} \quad \int_0^1 [\sigma W(r)]^2 dr,$$

and

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 = \int_0^1 [\sqrt{T}X_T(r)]^2 dr$$

$$(\text{since}(X_T(r))^2 = \begin{cases} 0 & \text{for } 0 \le r < (1/T) \\ (y_1/T)^2 & \text{for } (1/T) \le r < (2/T) \\ (y_2/T)^2 & \text{for } (2/T) \le r < (3/T) \\ . & . \\ (y_T/T)^2 & \text{for } r = 1 \end{cases} \text{ and } \int_0^1 [\sqrt{T}X_T(r)]^2 dr =$$

$$T\left((y_1)^2/T^3 + (y_2)^2/T^3 + ... + (y_{T-1})^2/T^3\right))$$

Also

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \quad \underrightarrow{L} \quad (1/2)\sigma^2[W(1)^2 - 1]$$

(Recall that $T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t = (1/2T)y_T^2 - \sum_{t=1}^{T}(1/2T)\varepsilon_t^2$ (when $y_0 = 0$)) then

$$\boxed{T(\hat{\rho}_T - 1) = \frac{(1/2)\sigma^2[W(1)^2 - 1]}{\int_0^1 [\sigma W(r)]^2 dr}}$$

Recall that $W(1)^2$ is chi square with one degree of freedom. The prob that a Chi-Square is less than 1 is .68 and since the denominator is positive, the prob that $\hat{\rho}_T - 1$ is negative approaches 0.68 as T tends to infinity. In other words in two thirds of the samples generated by a RW, the estimate will be less than the true value of unity or negative values will be twice as likely than positive values.

In practice critical values for the random variable $T(\hat{\rho}_T - 1)$ are found by calculating the exact small-sample distribution assuming the innovations are gaussian, generally by Monte Carlo.

$\hat{\rho}_T$ is a super consistent estimator

It follows that $\hat{\rho}_T$ is a super consistent estimate of the true value.

$$\sqrt{T}(\hat{\rho}_T - 1) = \frac{T^{-3/2}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}$$

but the numerator converges to $T^{-\frac{1}{2}}(1/2)\sigma^2[W(1)^2 - 1]$. The chi-square function has finite variance then the variance of the numerator is of order $(1/T)$, meaning that the numerator converges in probability to zero, hence $\sqrt{T}(\hat{\rho}_T - 1) \xrightarrow{P} 0$

Testing for the existence of unit roots

Consider the following model

$$y_t = \mu + \beta t + \alpha y_{t-1} + \varepsilon_t \tag{3}$$

where ε_t "is assumed" to be N(0,σ^2)

We want to test the Hypothesis of the existence of a unit root therefore we set the following null and alternative hypothesis.

H_0) $\alpha = 1$ (unit root)

H_1) $\alpha < 1$ (Integrated of order zero)

The obvious estimator of is the OLS estimator, $\hat{\beta}$. The problem is that under the null hypothesis there is considerable evidence of the non - adequacy of the asymptotic (approximate in large samples) distribution. Therefore

Equation (3) can be reparameterizied as

$$\Delta y_t = \mu + (\alpha - 1)y_{t-1} + \beta t + \varepsilon_t$$

or

$$\Delta y_t = \mu + \lambda y_{t-1} + \beta t + \varepsilon_t \tag{4}$$

For this expression the relevant hypothesis should be written as

$H_0)\lambda = 0$ (unit root)

$H_1)\lambda < 0$ (Integrated of order zero)

Fuller (1976) tabulated, using Monte Carlo methods, critical values for alternative cases, for example for a sample size of 100 the 5 % critical values are

$\mu = 0,\ \beta = 0$	-2.24
$\mu \neq 0,\ \beta = 0$	-3.17
$\mu \neq 0,\ \beta \neq 0$	-3.73

Therefore the method simply consist to check the t-statistic of $\hat{\lambda}$ against the critical values of Fuller (1976). Notice that the critical values depend on

i) the sample size

ii) whether you include a constant and/or a time trend.

This procedure is only valid when there is no evidence of serial correlation in the residuals, $\hat{\varepsilon}_t$. To see if this condition is satisfied you should look at the diagnostic tests for serial correlation in the regression. If there is serial correlation you should need to include additional lags, say $\Delta y_{t-1}, \Delta y_{t-2}$ etc to equation (4) until the serial correlation of the residuals disappears, that is

Augmented Dickey Fuller

$$\Delta y_t = \mu + \lambda y_{t-1} + \beta t + \alpha_1 \Delta y_{t-1} + ... + \alpha_k \Delta y_{t-k} + \varepsilon_t$$

In this case we chose to augment the regression with k lags. This is usually denoted as ADF(k). To choose the order of augmentation of the DF regression several procedures have been proposed in the literature. Some of these consist in:

(*i*) *choosing k as a function of the number of observations as in Schwert (1989)*

$k = INT(12(T/100)^{1/12})$

(*ii*) *information based rules such as AIC and BIC.*

(*iii*) *Sequential rules*

General to specific seems to be preferable to the other methods.

Small sample properties of the Dickey Fuller tests.

The power (the ability to reject the null Hypothesis) of the Dickey Fuller Tests is notoriously weak. Thus it can be difficult to reject the null of a unit root test even if the true series is stationary. That is, most of the time an ADF test will not reject the null of unit root even if the true model is an autoregressive model with an autoregressive coefficient of say .8 (that is an Integrated of order zero series). In addition the ADF tests often come up with conflicting results depending in the order of the lag structure. A good practice is to start with a quite general model and delete the non significant lags of Δy_t.

Phillips-Perron-type tests for unit roots

The ADF test includes additional lagged terms to account for the fact that the DGP might be more complicated than an AR(1). An alternative approach is that suggested by Phillips(1987) and Perron (1988). They make a non-parametric correction to the standard deviation which provides a consistent estimator of the variance.

They use

$$S_{Tl}^2 = T^{-1}\sum_{t=1}^{T}(\varepsilon_t^2) + 2T^{-1}\sum_{t=1}^{l}\sum_{t=j+1}^{T}\varepsilon_t\varepsilon_{t-j}$$

$$S_{\varepsilon}^2 = T^{-1}\sum_{t=1}^{T}(\varepsilon_t^2)$$

where $_l$ is the lag truncation parameter used to ensure that the autocorrelation of the residuals is fully captured.

An asymptotically valid test $\phi = 1$, for

$$\Delta y_t = \mu + (\phi - 1)y_{t-1} + \varepsilon_t \qquad \varepsilon_t \tilde{} iiD(0, \sigma^2)$$

when the underlying DGP is not necessarily an AR(1) process, is given by the Phillips Z-test.

$$Z(\tau_\mu) = (S_\varepsilon / S_{Tl})\tau_\mu - (1/2)(S_{Tl}^2 - S_\varepsilon^2)[S_{Tl}[T^2\sum_{t=2}^{T}(y_t - \overline{y})^2]^{.5}]^{-1}$$

where τ_μ is the t-statistic associated with testing the null hypothesis $\rho = 1$. The critical values for this test statistic are the same as those used for the same case in the fuller table. Monte Carlo work suggests that the Phillips-type test has poor size properties (tendency to over reject when is true) when the underlying DGP has large negative MA components.

Why is this Important?

The study of econometric models with non-stationary data has been one of the most important concerns of econometricians in the last 20 years. Therefore the topic is very vast and we just will *mention* some of the most important issues.

Spurious Regressions.

Granger and Newbold (1974) have shown that, using I(1), you can obtain an apparently significant regression (say with a high R^2) even if the regressor and the dependent variable are independent. He generated independent random walks regress one against the other and obtain very high R^2 for this equation. They conclude that this result is spurious. As a rule of thumb whenever you

obtain a very high R^2 and a very low DW you should suspect that the result are spurious.

Regressing Series that are integrated of the same order

Another important issue is that whenever you try to explain a variable, say y_t , by another variable, say x_t , you should check that these variables are integrated of the same order to obtain meaningful results.

Usual problems with ADF tests and possible solutions

Structural breaks and unit roots

We have already mentioned that the ADF test has very low power to reject the null hypothesis of the existence of a unit root. Therefore is very difficult using this type of test to distinguish between an autoregressive process with root, say .95, and a unit root process. This is particularly true if there is a structural break in the mean of the series. Perron (1988) have shown that an I(0) process with a structural break in the mean will be difficult to distinguish from a I(1) process. If we know were the break takes place, the natural thing to do, is to partial out the break by using dummy variables and test for unit roots once the break has been partialled out.

A possible solution to try to identify these breaks is to perform the ADF test recursively and to compute recursive t-statistics.

Perron (1989) showed that if a series is stationary around a deterministic time trend which has undergone a permanent shift sometime during the period under consideration, failure to take account of this change in the slope will be mistaken by the usual ADF unit root test as a persistent innovation to a stochastic (non-stationary) trend. That is, a unit root test which does not take into account a break in the series will have very low power. There is a similar loss in power if there is a shift in the intercept.

If the breaks in the series are known then it is relatively simple to adjust the ADF test by including dummy variables to ensure there are as many deterministic regressors as there are deterministic components in the DGP.

However is unlikely that we will know the break then we can proceed by using the critical values provided by Banjeree, Lumsdaine and Stock (1992).

Recursive t-statistics

The recursive ADF -statistic is computed using sub samples $t = 1..k\ fork = k_0, ...T$, where k is the start up value and T is the sample size of the full sample. The most general model (with drift and trend) is estimated for each sub sample and the minimum value of $\tau_\tau(k/T)$ across all the sub samples is chosen and compared with the table provided by Banjeree, Lumsdaine and Stock

Rolling ADF tests

This method could also be applied using a (large enough) window to see if there are clear changes in the pattern of a series. This can be done by carrying out the following steps

T	Percentile	τ_τ	Recursive min τ_τ	Rolling min τ_τ
100	.025	-3.73	-4.62	-5.29
	.050	-3.45	-4.33	-5.01
	.100	-3.15	-4.00	-4.71
250	.025	-3.69	-4.42	-5.07
	.050	-3.43	-4.18	-4.85
	.100	-3.13	-3.91	-4.59
500	.025	-3.68	-4.42	-5.00
	.050	-3.42	-4.18	-4.79
	.100	-3.13	-3.88	-4.55

Tests with stationarity as null

The **KPSS** test
Consider the following model.

$$y_t = \delta t + \xi_t + \varepsilon_t$$

where ε_t is a stationary process and ξ_t is a random walk given by

$$\xi_t = \xi_{t-1} + u_t \quad u_t \tilde{} iid(0, \sigma_u^2)$$

The null of stationarity is formulated as
$H_0) \sigma_u^2 = 0$
The test statistic for this hypotesis is given by

$$LM = \frac{\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_e^2}$$

where e_t are the residuals of a regression of y_t on a constant and a time trend, $\hat{\sigma}_e^2$ is the residual variance for this regression and S_t is the partial sum of e_t defined by

$$S_t = \sum_{t=1}^{T} e_t \quad t = 1, 2, .., T.$$

For testing the null of the level stationary instead of trend stationary the test is constructed the same way except that e is obtained as the residual from a regression of y on an intercept only. The test is an upper tail test. When the errors are iid the asymptotic distribution of the test is derived in Nabeya and Tanaka (1988). In other cases need to be conviniently adjusted.

Variance Ratio Tests

Given the low power of the ADF test, the variance ratio test will provide us with another tool to discriminate between stationary and non-stationary series.

Consider y_t and assume that it follows follows a random walk, i.e

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

then by iterative substitution we know that

$$y_t = y_{t-k} + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3} + ... + \varepsilon_{t-k+1}$$

Now if we denote the difference between y_t and y_{t-k} as $\Delta_k y_t$, then

$$\Delta_k y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3} + ... + \varepsilon_{t-k+1}$$

and clearly the variance of $\Delta_k y_t$, is $\sigma^2 \Delta_k y_t = k$, where σ^2 is the variance of ε.

We can define a "variance ratio" function (a function of k) as

$$\lambda_1(k) = \frac{Var(\Delta_k y_t)}{Var(\Delta_1 y_t)} = k.$$

Therefore a plot of λ_1 against k should be an increasing straight line.

Alternatively we may define a new function $\lambda_2(k)$ as $\lambda_2(k) = \lambda_1(k)/k$, then if there is a unit root, $\lambda_2(k)$ tends to one when k tends to infinite.

However if y_t does not contain a unit root then lim $\lambda_2(k)$ when k tends to infinite is equal to zero.

Proof

To see this assume the following AR(1) process

Example AR(1)

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \qquad t = 1, ..., T$$

we have seen that by iterative substitution we can express this process as

$$y_t = \phi_1^k y_{t-k} + \phi_1^{k-1} \varepsilon_{t-(k-1)} + \phi_1^{k-2} \varepsilon_{t-(k-2)} + ... + \phi_1 \varepsilon_{t-1} + \varepsilon_t$$

Then subtracting y_{t-k} in both sides of the equation we get the following expression

$$y_t - y_{t-k} = (\phi_1^k - 1) y_{t-k} + \phi_1^{k-1} \varepsilon_{t-(k-1)} + \phi_1^{k-2} \varepsilon_{t-(k-2)} + ... + \phi_1 \varepsilon_{t-1} + \varepsilon_t$$

then the variance of $y_t - y_{t-k}$, $Var(\Delta_k y_t)$

$$V(y_t - y_{t-k}) = (\phi_1^k - 1)^2 V(y_{t-k}) + V(\sum_{j=0}^{k-1} \phi_1^j \varepsilon_{t-j})$$

Notice that
$V(y_{t-k}) = V(y_t) = (1/(1-\phi_1^2))\sigma^2$
and
$V(\sum_{j=0}^{k-1} \phi_1^j \varepsilon_{t-j}) = ((1-\phi_1^{2k})/(1-\phi_1^2))\sigma^2$
therefore we can write the variance of Δy_{t-k} as

$$V(y_t - y_{t-k}) = (\phi_1^k - 1)^2(1/(1-\phi_1^2))\sigma^2 + ((1-\phi_1^{2k})/(1-\phi_1^2))\sigma^2$$

In the same way we can express for a stationary process the variance of the first difference of y_t, $(y_t - y_{t-1}) = (\phi_1 - 1)y_{t-1} + \varepsilon_t$.

$$V(y_t - y_{t-1}) = (\phi_1 - 1)^2 V(y_{t-1}) + \sigma^2 = (\phi_1 - 1)(1/(1-\phi_1^2))\sigma^2 + \sigma^2$$

then the variance ratio can be written as

$$\lambda_1(k) = \frac{(\phi_1^k - 1)^2(1/(1-\phi_1^2))\sigma^2 + (\ (1-\phi_1^{2k})/(1-\phi_1^2)\ )\sigma^2}{(\phi_1 - 1)^2(1/(1-\phi_1^2))\sigma^2 + \sigma^2}$$

then the limit of $\lambda_1(k)$ when k tends to infinite for a stationary process is

$$lim_{k\to\infty}\lambda_1(k) = \frac{1}{1-\phi_1}$$

which is a constant provided that $\phi_1 \neq 1$.

It should be clear from the previous result that the limit of $\lambda_2(k)$ equals 0 when k tends to infinite.

Trend stationary vs difference stationary processes

A trend stationary variable may be written as

$$y_t = \alpha + \mu t + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3} + ...$$

then y_{t-k} is simply

$$y_{t-k} = \alpha + \mu(t-k) + \varepsilon_{t-k} + \theta_1\varepsilon_{t-1-k} + \theta_2\varepsilon_{t-2-k} + \theta_3\varepsilon_{t-3-k} + ...$$

The k^{th} difference can be obtained simply by subtracting the two above equations

$$\begin{aligned} y_t - y_{t-k} &= \mu k + \varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_{k-1}\varepsilon_{t-(k-1)} + (\theta_k - 1)\varepsilon_{t-k} + \\ &\quad (\theta_{k+1} - \theta_1)\varepsilon_{t-(K-1)} + ... + (\theta_{k+q} - \theta_q)\varepsilon_{t-q} + ... \\ &= \mu k + \sum_{j=0}^{k-1} \theta_j\varepsilon_{t-j} + \sum_{j=0}^{\infty} c(\theta_{k+j} - \theta_j)^2 \end{aligned}$$

Then the variance of $y_t - y_{t-k}$ may be written as

$$V(y_t - y_{t-k}) = \sigma^2 \sum_{j=0}^{k-1} \theta_j^2 + \sigma^2 \sum_{j=0}^{\infty} (\theta_{k+j} - \theta_j)^2$$

From the previous equation we can see that when k tends to infinity, the variance of $\Delta_k y_t$ is equal to

$$V(\Delta_k y_t) = 2\sigma^2 \sum_{j=0}^{k-1} \theta_j^2$$

Now the first difference of a trend stationary process, Δy_t is

$$\begin{aligned} y_t - y_{t-1} &= \mu + \varepsilon_t + (\theta_1 - 1)\varepsilon_{t-1} + ... + (\theta_{k+1} - \theta_k)\varepsilon_{t-(k+1)} \\ &\quad +... + (\theta_{k+q} - \theta_{k+q-1})\varepsilon_{t-q} + ... \end{aligned}$$

Then the variance of the first difference can be written as

$$\begin{aligned} V(y_t - y_{t-1}) &= V(\varepsilon_t + (\theta_1 - 1)\varepsilon_{t-1} + ... + (\theta_{k+1} - \theta_k)\varepsilon_{t-(k+1)} \\ &\quad +... + (\theta_{k+q} - \theta_{k+q-1})\varepsilon_{t-(k+q)} + ...) \\ &= \sigma^2(1 + \sum_{j=0}^{\infty} (\theta_{j+1} - \theta_j)^2) \end{aligned}$$

The variance ratio should be

$$\lambda_1(k) = \frac{Var(\Delta_k y_t)}{Var(\Delta_1 y_t)} = \frac{2\sum_{j=0}^{k-1} \theta_j^2}{(1 + \sum_{j=0}^{\infty} (\theta_{j+1} - \theta_j)^2)}$$

This expression is a constant and might be greater or smaller than one depending on the θ_j values. Therefore can simply distinguish between the two models by simply noting that under the random walk assumption λ_1 increase with k and that under the trend stationary assumption λ_1 tends to a constant. Alternatively we can consider λ_2 $Var(\Delta_k y_t)/k$. and note both, that when the model is a random walk this expression tends to 1 (see proof above), and that this ratio should tend to zero when k tends to infinity since $Var(\Delta_k y_t)$ is constant for the trend stationary model.

Sampling distribution of λ(k) under the Random Walk Hypothesis.

$H_0)$ $\alpha = 1$ *or* $y_t = \mu + y_{t-1} + \varepsilon_t$ ε_t ~IID N(0,)

It can be shown that asymptotically under the null, $\sqrt{Tk}(\hat{\lambda}_2(k) - 1) \overset{a}{\sim} \mathrm{N}(0, 2(k-1))$. Then tests of the null Hypothesis can be carried out on the standardized statistics.