# Cointegration, Stationarity and Error Correction Models.

## STATIONARITY

Wold's decomposition theorem states that a stationary time series process with no deterministic components has an infinite moving average (MA) representation. This in turn can be represented by a finite autoregressive moving average (ARMA) process. A variable $y_t$ is said to be integrated of order d [$y_t \sim I(d)$] if it has a stationary, invertible, non deterministic, ARMA representation after differencing d times.

This is typical of the sort of statement you see in modern econometrics. What does it mean? Let us turn to a more intuitive definition of **stationarity**, i.e. its mean, variance and covariances are independent of time. This is called weak or covariance stationarity. The **order of integration** of a series is ascertained by the application of a set of tests, commonly known as tests for **"unit roots"**

More formally weak or covariance stationary requires for $y_t$:

$E(y_t) = \mu$
$E[(y_t - \mu)^2] = \sigma^2$
$E[(y_t - \mu)(y_{t-j} - \mu)] = \sigma_{oj}$

One way for testing for this is to look at a plot of the correlelogram:
http://www.stat.wvu.edu/~fdesta/seminar/node12.html

# Correlogram

A useful aid in interprating a set of autocorrelation coefficents is a graph called a correlogram in which is plotted against the lag($k$); where is the autocorrelation coefficent at lag($k$). A correlogram can be used to get a general understanding on the following aspects of our time series:

- A random series: if a time series is completely random then for Large ($N$), will be approximately zero for all non-zero values of ($k$).
- Short-term correlation: stationary series often exhibit short-term correlation charcterized by a fairly large value of followed by (2) or (3) more coefficents which, while significantly greater than zero, tend to get successively smaller.
- Non-stationary series:if a time series contains a trend, then the values of will not come to zero except for very large values of the lag.
- seasonal fluctuations:

We will not go down this road but instead look at some of the plethora of tests which have appeared in recent years, the most common of which is the Dickey Fuller test: Consider the model

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

It is intuitive that if y is stationary then $|\delta|<1.0$ If $|\delta|=1.0$ then:

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

and continuous substitution makes y a function of all past error terms. The series is not stationary, the variance of $y_t = t\sigma^2$ . In testing for unit roots we are essentially testing $\rho=1$ against the alternative $\delta<1$. In the case where $\delta=1$ we say y has a unit root, but if we difference the series once, then $\Delta y_t = \varepsilon_t$ will be stationary because of assumption $\varepsilon_t$ is stationary. 1(a) is called a random walk. This is an example of a difference stationary process (DSP), that is y is not stationary but some difference of y is. Many time series are DSPs.

We do not test for stationarity by estimating (1) and testing $\delta<1.0$. Instead we subtract $y_{t-1}$ from both sides of(1):

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

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

it is this we estimate and test for $(\delta-1)<0$. From now on we refer to this as:

$\Delta y_t = \rho y_{t-1} + \varepsilon_t$ (3)

We cannot use the t tables to test $\delta <0$ because Dickey and Fuller (DF) show that under the null hypothesis that $\delta =1$, the least squares estimate of $\delta$ is not distributed around unity but less than unity. This also impacts on (3), see more later on in the lecture on this bias. Thus DF have constructed alternative tables.

The augmented Dickey Fuller test is when we add lagged dependent variables to the rhs of (3)

$\Delta y_t = \rho y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \ldots\ldots\ldots\ldots + \varepsilon_t$

but it is still focused on testing $\rho<0$

# COINTEGRATION

For a group of variables to be cointegrated it is required that:

i) they should all be I(d) i.e. integrated to the same order, often I(1)
ii) a linear combination of the two series should exist which is integrated to a lower order (generally I(0))

The intuition behind this is that unless the variables are integrated to the same order then in general the equation does not make sense:

$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t}$

where Y and $X_1 \sim I(0)$ and $X_2 \sim I(2)$ then we have different properties left and right hand side and that does not make sense. An equilibrium relationship implies that whereas the series may have different trends or cyclical components, then movements in one are mirrored by combination of movements in the other. For example, in the case where all are I(1) but a linear combination: $y_t - \beta_0 - \beta_1 X_{1t} - \beta_2 X_{2t}$ is I(0)

**TESTING FOR COINTEGRATION**

For example take:

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t \qquad (4)$$

Estimate this (e.g. by OLS) then take the residuals and apply the Dickey Fuller test to those residuals:

$$\Delta\varepsilon_t = \rho\varepsilon_{t-1} + \upsilon_t \qquad (5)$$

A test for cointegration is a test that $\varepsilon_t \sim I(0)$. The critical values are different from the statndard Dickey Fuller tests because $\varepsilon_t$ is an estimate. Critical values for this test are given in Engle and Granger (1989) and McKinnon, 1991).

**THE ERROR CORRECTION MODEL**

***We cannot estimate (4) directly if the variables are I(1) as the conventional statistical tests, including the t tests, are not valid with non-stationary variables. Hence an alternative approach is the ECM model.***

For example take:

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t \qquad (6)$$

we have established that all variables are I(1) and cointegrate, then an error correction model (ECM) of the form

$$\Delta y_t = \boldsymbol{\delta_1 \Delta X_{1t} + \delta_2 \Delta X_{2t}} + \lambda[y_{t-1} - (\beta_0 + \beta_1 X_{1t-1} + \beta_2 X_{2t-1})] \qquad (7)$$

is a meaningful short-run adjustment equation to estimate as all variables are I(0). The formal proof of the proposition that any cointegrated series will have an error correction representation is given in Engle and Granger, Econometrica, 1987.

$\boldsymbol{\delta_1 \Delta X_{1t} + \delta_2 \Delta X_{2t}}$ can be thought of as short run adjustment factors, neither variables have to be included and other variables, which do not impact on the long run equilibrium equation, can be included. $[y_{t-1} - (\beta_0 + \beta_1 X_{1t-1} + \beta_2 X_{2t-1})]$ is the deviation from equilibrium in the previous period. Thus if $y_{t-1} > (\beta_0 + \beta_1 X_{1t-1} + \beta_2 X_{2t-1})$ we would expect y to fall back towards its equilibrium value. Thus we expect $\lambda<0$. It is the speed of adjustment back to equilibrium. A value of -0.90 means adjustment is rapid and a value of -0.05 that adjustment is very slow.

The pair of equations (6) and (7) can be thought of as providing information on both the short run parameters, **δ's** and λ and the long run equilibrium parameters, the β's.

**Engle Granger Two Step Method..**

We can estimate (7) either directly by non linear least squares or even OLS. Engle Granger propose the two step method.

Step 1. Estimate 6: $y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t}$ call the residual Z

Step 2 Estimate $\Delta y_t = \boldsymbol{\delta_1 \Delta X_{1t} + \delta_2 \Delta X_{2t}} + \lambda Z_{t-1}$.

Note this way we do not get t statistics on the long run equilibrium equation.

## Trend Stationarity

The following is an AR(1) (autoregressive with lag depth or order 1) with a deterministic linear trend term

$$Y_t = \theta Y_{t-1} + \delta + \gamma t + \varepsilon_t \qquad (1)$$

Where $|\theta| < 1$

The moving average (MA) representation of this on past error terms is

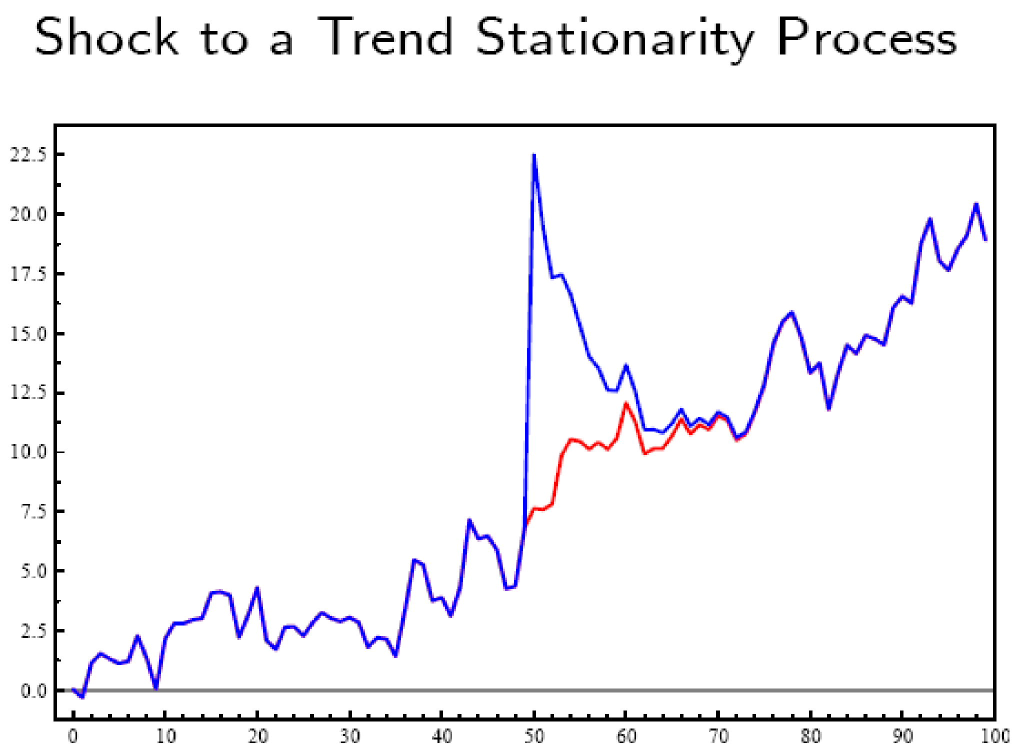
$$Y_t = \theta^t Y_0 + \mu_0 + \mu_1 t + \varepsilon_t + \theta\varepsilon_{t-1} + \theta^2\varepsilon_{t-2} + \theta^3\varepsilon_{t-3} + \ldots.. \qquad (2)$$

$$E(Y_t) = \theta^t Y_0 + \mu_0 + \mu_1 t \rightarrow \mu_0 + \mu_1 t \text{ as } t \rightarrow \infty \qquad (3)$$

This has a finite, unchanging variance, but no constant mean (because of $\mu_1 t$). Thus the process is not stationary.
However, the deviation from the mean

$$Y_t = Y_t - E[Y_t] = Y_t - \mu_0 - \mu_1 t \qquad (4)$$

is stationary and hence $Y_t$ is called a **trend stationary process.** Thus the shocks to the process are transitory and the process is 'mean reverting', with the mean $\mu_0 + \mu_1 t$ being the 'attractor'.




## UNIT ROOT PROCESSES

The following is an AR(1) model with a unit root $\theta$=1

$$Y_t = \theta Y_{t-1} + \delta + \varepsilon_t = Y_{t-1} + \delta + \varepsilon_t \qquad (5)$$

It is clearly non-staionary with no specified mean. But the process $\Delta Y_t$ is stationary and we term $Y_t$ **a difference stationary process** (in this case I(1))

$$Y_t = Y_{t-1} + \delta + \varepsilon_t \qquad (6)$$

$$Y_t = Y_{t-2} + \delta + \varepsilon_t + \delta + \varepsilon_{t-1} \text{ (substituting for } Y_{t-1}) \qquad (7)$$

$$Y_t = Y_{t-3} + \delta + \varepsilon_t + \delta + \varepsilon_{t-1} + \delta + \varepsilon_{-2} \text{ (substituting for } Y_{t-2})$$

$$Y_{t-3} + \delta + \delta + \delta + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{-2} \qquad (7)$$

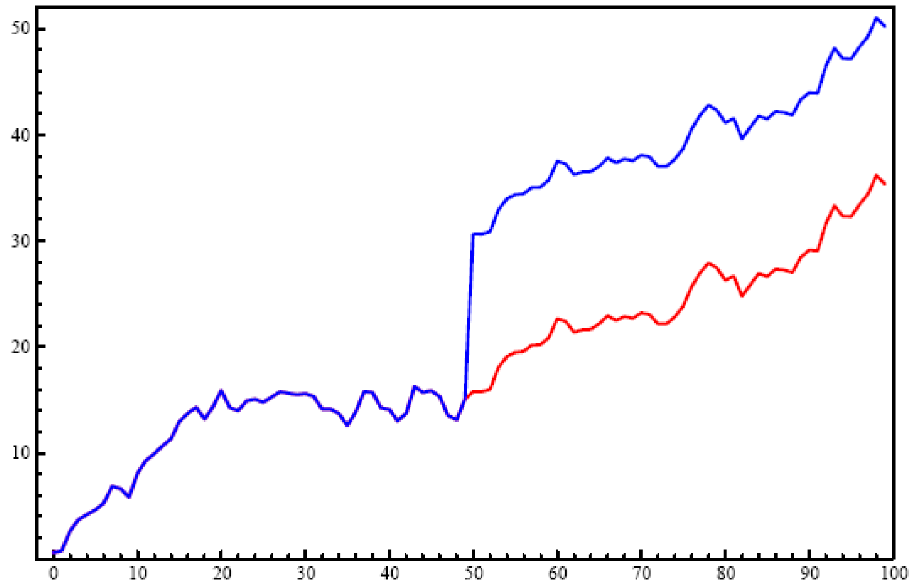
And so on until:

$$Y_t = Y_0 + \delta t + \Sigma\varepsilon_i \qquad (8)$$

$$\text{Hence } E[Y_t] = Y_0 + \delta t \qquad (9)$$

The effect of the initial value $Y_0$ stays in the process. We can also see from (8) that shocks have permanent effects and the figure below shows the impact of a large shock. The process has no attractor.

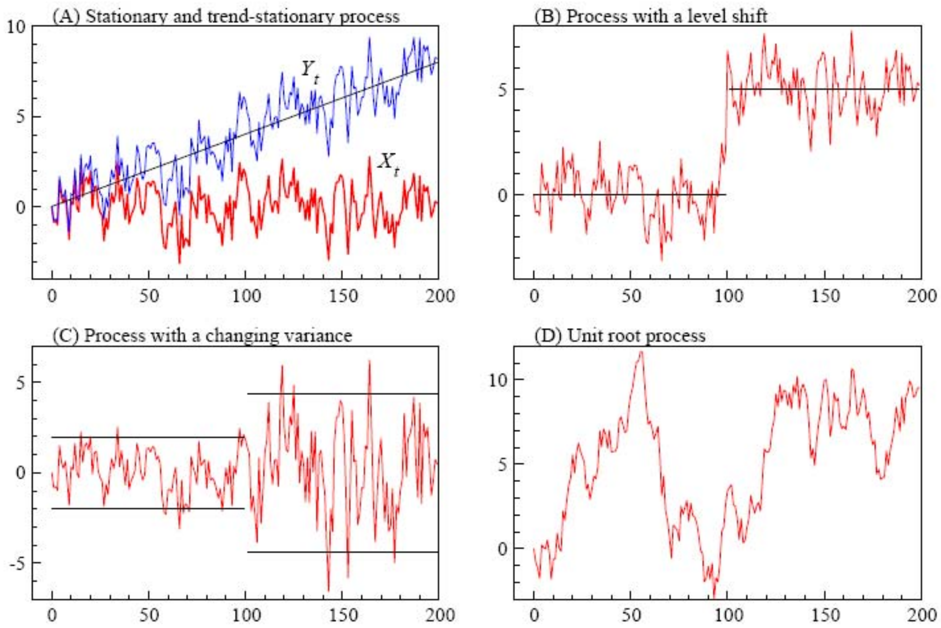
Shock to a Unit Root Process




## DETERMINISTIC TRENDS AND TREND STATIONARITY

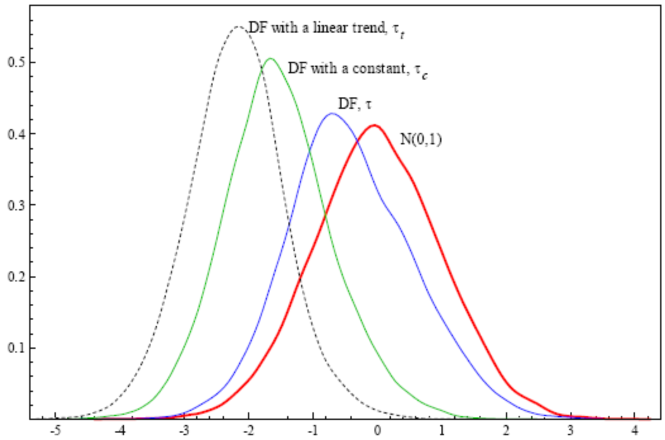
A time series that is stationary around a deterministic trend is called a trend stationary process.

**Examples of Different Processes**




To test for trend stationarity we include a trend term and typically a constant term in the Dickey Fuller/ADF regressions. However this changes the asymptotic distribution relating to the test shifting the distribution to the left, the inclusion of a time trend shifts it still further to the left which is why we need different critical values for the DF and ADF tests depending on whether a constant and/or a time trend is included in the specification.

The DF Distributions




But apart from this the test is the same and relates to the t statistic on $Y_{t-1}$. We are also interested in whether the coefficient on the constant term is significant.