Non-Stationary modeling, testing and forecasting

(Updated Spring 2021)

CHAOYI CHEN
Institute of MNB, Corvinus University of Budapest

Empirical Financial Econometrics

Outlines

- Deterministic and Stochastic Trends ⏩ Jump [Online Lecture + Self-study]
- Unit Root theory * ⏩ Jump [Self-study if you have interest :)]
- Brief Introduction to Unit Root Tests ⏩ Jump [Online Lecture + Self-study]
- Cointegration * ⏩ Jump [Self-study if you have interest :)]

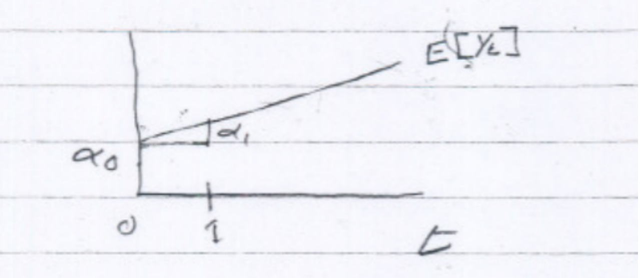
Deterministic and Stochastic Trends

- Trends: stochastic and deterministic
- Detrending
- Stochastic trends
- Random Walk
- Unit Root
- Taking First Differences
- ARIMA Model
- Order of integration terminology
- Differencing to undue integration
- Unit Root + Trend Model
- Forecasting in unit root plus trend model

Trends: stochastic and deterministic

- **Deterministic trends**
- **trends+noise**

$$y_t = \alpha_0 + \alpha_1 t + \varepsilon_t, \quad \varepsilon_t = 0 \sim WN(0, \sigma^2)$$



Trends: stochastic and deterministic

- **Trend + ARMA(Trend-stationary)**

$$
\begin{aligned}
y_t &= \alpha_0 + \alpha_1 t + u_t \\
u_t &= \sum_{i=1}^{p} a_i u_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t
\end{aligned} \tag{1}
$$

- Two components

$$
\begin{aligned}
y_t &= \text{deterministic trend} + \text{Stationary component} && (2)\\
&= (\alpha_0 + \alpha_1 t) + u_t && (3)
\end{aligned}
$$

- For GDP: Long run growth trend $(\alpha_0 + \alpha_1 t)$ + business cycle (u_t)
- u_t can also viewed as (infeasible) detrended version of y_t

Detrending

We can estimate model (67) jointly or following these three steps:

1. Run regression:

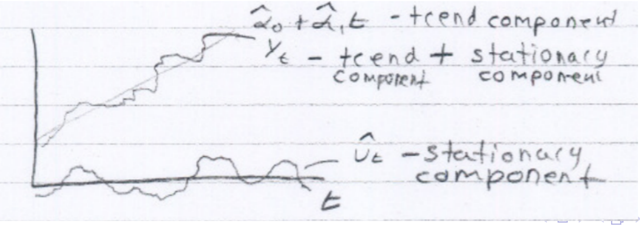
$$y_t = \hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{u}_t$$

2. Define the (feasible) detrended y_t as:

$$\hat{u}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t$$

3. Fit an ARMA(p,q) model to $\hat{u}_t$:

$$\hat{u}_t = \sum_{i=1}^{p} a_1 \hat{u}_{t-i} + \sum_{t=1}^{q} b_1 \varepsilon_{t-i} + \varepsilon_t$$



- **Forecasting:**

$$y_{t+1} = \alpha_0 + \alpha_1(t+1) + u_{t+1} \tag{4}$$

$$E_t y_{t+1} = \alpha_0 + \alpha_1(t+1) + E_t u_{t+1} \tag{5}$$

$$\hat{y}_{t+1|t} = \underbrace{\hat{\alpha}_0 + \hat{\alpha}_1(t+1)}_{\text{trend forecast}} + \underbrace{\hat{u}_{t+1|t}}_{\substack{\text{ARMA}\\ \text{forecast}}} \tag{6}$$

$$\hat{u}_{t+1}|t = \sum_{i=1}^{p} \hat{a}_i \hat{u}_{t-i} + \sum_{i=1}^{q} \hat{b}_i \hat{\varepsilon}_{t-i} \tag{7}$$

Detrending

- **Remarks:** Some macro models are designed to describe growth and some to describe business recycles. The former would use the trend component and the latter would use the stationary component.

Stochastic trends

- **Random Walk Model:**

$$
\begin{aligned}
y_t &= y_{t-1} + \varepsilon_t && (8)\\
\varepsilon_t &\sim WN(0, \sigma^2) \qquad (E_{t-1}\varepsilon_t = 0) && (9)\\
y_0 &= 0 && (10)
\end{aligned}
$$

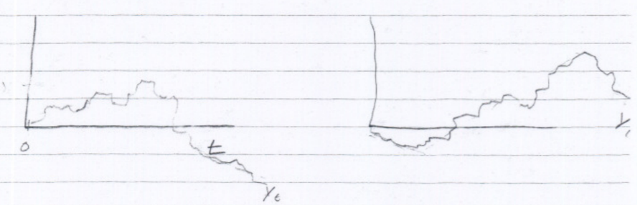
- **Forecasts:**

$$E_t y_{t+1} = y_t + E_t \varepsilon_{t+1} = y_t \qquad \text{A "no-change" forecast}$$

- **Tendency to wander off/stochastic trends**

Random Walk

- **What might a random walk look like?**



- Tendency to trend off, but direction of trend is random and the trending behavior shifts randomly over time.
- No, two realizations of y_t will wander off on quite the same way.
- **examples**
  1. Random walk through a field after too many drinks
  2. Stock price under risk neutral/ rational expectations

Unit Root

- **Unit root model (Unit root non-stationary)**

$$
\begin{aligned}
y_t &= y_{t-1} + u_t; \\
u_t &\sim \text{covariance stationary} \\
E[u_t] &= 0
\end{aligned}
$$

Unit Root

- **Unit root model (Unit root non-stationary)**

$$\begin{aligned} y_t &= y_{t-1} + u_t; \\ u_t &\sim \text{covariance stationary} \\ E[u_t] &= 0 \end{aligned} \tag{11}$$

- **Random Walk:**
  1. $\varepsilon_t \sim WN$ with $E_t \varepsilon_{t+1} = 0$
  2. Change in y_t is unforecastable ε_t
  3. Best forecast of tomorrow (y_{t+1}) is today (y_t)

Unit Root

- **Unit root model (Unit root non-stationary)**

$$\begin{aligned} y_t &= y_{t-1} + u_t; \\ u_t &\sim \text{covariance stationary} \\ E[u_t] &= 0 \end{aligned} \quad (11)$$

- **Random Walk:**
  1. $\varepsilon_t \sim WN$ with $E_t\varepsilon_{t+1} = 0$
  2. Change in y_t is unforecastable ε_t
  3. Best forecast of tomorrow (y_{t+1}) is today (y_t)
- **Unit Root:**
  1. $u_t \sim$ covariance's stationary and mean zero
  2. u_t may be forecastable (e.g. ARMA)
  3. Forecast for y_{t+1} more complicated

Unit Root

- **Unit root model (Unit root non-stationary)**

$$\begin{aligned} y_t &= y_{t-1} + u_t; \\ u_t &\sim \text{covariance stationary} \\ E[u_t] &= 0 \end{aligned} \quad (11)$$

- **Random Walk:**
  1. $\varepsilon_t \sim WN$ with $E_t \varepsilon_{t+1} = 0$
  2. Change in y_t is unforecastable ε_t
  3. Best forecast of tomorrow (y_{t+1}) is today (y_t)
- **Unit Root:**
  1. $u_t \sim$ covariance's stationary and mean zero
  2. u_t may be forecastable (e.g. ARMA)
  3. Forecast for y_{t+1} more complicated
- So the random walk is special case of unit root.
- Both exhibit the type of wandering behavior

Taking First Differences

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

- Bring y_{t-1} to the left-hand-side to solve for the **first difference** of y_t:

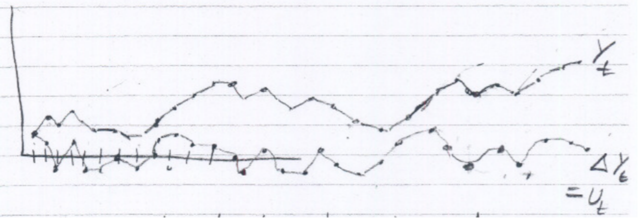
$$\Delta y_t = y_t - y_{t-1} = u_t \sim \text{covariance stationary}$$

In special case of Random Walk:

$$\Delta y_t = y_t - y_{t-1} = \varepsilon_t \sim WN(0, \sigma^2)$$

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

- **Remarks:**
- y_t Unit root $\Rightarrow \Delta y_t$ Stationary
- Difference is the way we "detrend" a stochastic trend.
- Differences in a picture



- Can interpret stochastic trends as long-run growth (or contraction) due to say, random, technology shocks.

Differencing with Lag Operator Notation

$$
\begin{aligned}
Ly_t &= y_{t-1} && (12)\\
(1-L)y_t &= y_t - Ly_t = y_t - y_{t-1} = \Delta y_t && (13)\\
(1-L)^0 y_t &= (1)y_t = y_t && (14)\\
(1-L)^2 y_t &= (1-L)(1-L)y_t && (15)\\
&= (1-L)\Delta y_t && (16)\\
&= \Delta^2 y_t \qquad \text{double difference} && (17)
\end{aligned}
$$

$$\Delta^d = (1-L)^d y_t = y_t \qquad \text{differenced d-times}$$

ARIMA Model

- **The ARIMA(p,d,q) is an ARMA(p,q) on the d^{th} difference:**

$$(1-L)^d y_t = \sum_{i=1}^{p} a_i (1-L)^d y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t \qquad (18)$$

where

$$\begin{aligned} \varepsilon_t &\sim WN(0, \sigma^2) \qquad (19) \\ (1-L)^d y_t &\sim \text{stationary for } d = 0, 1, 2 \end{aligned}$$

ARIMA Model

- **The ARIMA(p,d,q) is an ARMA(p,q) on the d^{th} difference:**

$$(1-L)^d y_t = \sum_{i=1}^{p} a_i (1-L)^d y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t \tag{18}$$

where

$$\varepsilon_t \sim WN(0, \sigma^2) \tag{19}$$

$$(1-L)^d y_t \sim \text{stationary for } d = 0, 1, 2 \tag{20}$$

- **In other words:**

$$y_t \sim ARIMA(p, d, q) \quad \Leftrightarrow \quad (1-L)^d y_t \sim ARMA(p, q)$$

ARIMA Model

- **The ARIMA(p,d,q) is an ARMA(p,q) on the d^{th} difference:**

$$(1-L)^d y_t = \sum_{i=1}^{p} a_i (1-L)^d y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t \tag{18}$$

where

$$\varepsilon_t \sim WN(0, \sigma^2) \tag{19}$$

$$(1-L)^d y_t \sim \text{stationary for } d = 0, 1, 2 \tag{20}$$

- **In other words:**

$$y_t \sim ARIMA(p, d, q) \iff (1-L)^d y_t \sim ARMA(p, q) \tag{21}$$

$$y_t \sim ARIMA(p, 1, q) \iff \Delta y_t \sim ARMA(p, q)$$

ARIMA Model

- **The ARIMA(p,d,q) is an ARMA(p,q) on the d^{th} difference:**

$$(1-L)^d y_t = \sum_{i=1}^{p} a_i (1-L)^d y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t \qquad (18)$$

where

$$\varepsilon_t \sim WN(0, \sigma^2) \qquad (19)$$

$$(1-L)^d y_t \sim \text{stationary for } d = 0, 1, 2 \qquad (20)$$

- **In other words:**

$$y_t \sim ARIMA(p,d,q) \Leftrightarrow (1-L)^d y_t \sim ARMA(p,q) \qquad (21)$$

$$y_t \sim ARIMA(p,1,q) \Leftrightarrow \Delta y_t \sim ARMA(p,q) \qquad (22)$$

$$y_t \sim ARMA(p,0,q) \Leftrightarrow y_t \sim ARMA(p,q) \qquad (23)$$

ARIMA Model

- **An ARIMA(p,1,q) is a type of unit root**
- **To see why:** Define

$$
\begin{aligned}
u_t &= (1-L)y_t = y_t - y_{t-1} && (24)\\
&= \text{stationary ARMA(p,q)} && (25)
\end{aligned}
$$

$$
\begin{aligned}
y_t &= y_t - y_{t-1} + y_{t-1} && (26)\\
&= \underbrace{\Delta y_t}_{u_t} + y_{t-1}
\end{aligned}
$$

ARIMA Model

- **An ARIMA(p,1,q) is a type of unit root**
- **To see why:** Define

$$\begin{aligned} u_t &= (1-L)y_t = y_t - y_{t-1} && (24)\\ &= \text{stationary ARMA(p,q)} && (25) \end{aligned}$$

$$\begin{aligned} y_t &= y_t - y_{t-1} + y_{t-1} && (26)\\ &= \underbrace{\Delta y_t}_{u_t} + y_{t-1} && (27) \end{aligned}$$

- **Another way to write ARIMA(p,q)**

$$\begin{aligned} y_t &= y_{t-1} + u_t && (28)\\ u_t &= \text{stationary ARMA(p,q)} && (29) \end{aligned}$$

Estimating an ARIMA(p,1,q) Model

To Estimate an ARIMA(p,1,q)

- **Step 1:** Difference y_t to get $u_t = \Delta y_t =$ stationary $ARMA(p, q)$
- **Step 2:** Estimate an ARMA(p,q) using the differenced data.

$$\Delta y_t = \sum_{i=1}^{p} a_i \Delta y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t$$

Forecasting with ARIMA(p,1,q) model

In theory:

$$y_{t+1} = y_t + u_{t+1} \tag{30}$$

$$E_t y_{t+1} = y_t + E_t u_{t+1} \tag{31}$$

$$E_t u_{t+1} = E_t \Delta y_{t+1} = \sum_{i=1}^{p} a_i \Delta y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} \tag{32}$$

$$E_t y_{t+1} = y_t + \sum_{i=1}^{p} a_i \Delta y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} \tag{33}$$

In practice:

$$\hat{y}_{t+1|t} = y_t + \widehat{\Delta y}_{t+1|t} \tag{34}$$

$$= y_t + \sum_{i=1}^{p} \hat{a}_i \Delta y_{t-i} + \sum_{i=1}^{q} \hat{b}_i \hat{\varepsilon}_{t-i} \tag{35}$$

Basic idea:

1. Difference y_t to make it stationary
2. Use ARMA(p,q) to forecast its change (Δy_{t+1})
3. Your forecast of the level (y_t) is just last periods value (y_{t-1}) plus the predicted change.

Order of integration terminology

- Unit root referred to as integrated of order 1 or $I(1)$:
  Why? unit roots

$$
\begin{aligned}
y_t &= y_{t-1} + u_t && (36)\\
y_0 &= 0 && (37)\\
u_t &\sim \text{stationary} && (38)
\end{aligned}
$$

- **Now iterate backwards:**

$$
\begin{aligned}
y_t &= y_{t-1} + u_t && (39)\\
&= [y_{t-2} + u_{t-1}] + u_t && (40)\\
&= y_{t-3} + u_{t-2} + u_{t-1} + u_t && (41)\\
&= y_1 + u_2 + u_3 + u_4 + \ldots + u_t && (42)\\
&= y_0 + u_1 + u_2 + u_3 + \ldots + u_t && (43)\\
&= u_1 + u_2 + u_3 + \ldots + u_t && (44)
\end{aligned}
$$

Order of integration terminology

- So

$$y_t = u_1 + u_2 + u_3 + \ldots + u_t$$

  is a sum of the u_t's
- So maybe it should have been called "summed of order 1".
- But sums approximate integrals, so we say **integrated of order 1** or **I(1)**

Order of integration terminology

- So

$$y_t = u_1 + u_2 + u_3 + \ldots + u_t$$

  is a sum of the u_t's
- So maybe it should have been called "summed of order 1".
- But sums approximate integrals, so we say **integrated of order 1** or **I(1)**
- u_t is integrated of order zero or I(0) — no summation to get u_t

Order of integration terminology (continued)

- If a process is a sum of I(0) processes, we say that it is I(1)
- The Random Walk, ARIMA(p,1,q), and more general unit root processes are all I(1)
- If a process is the sum of an I(1) process we say that it is I(2)
  e.g if $y_t \sim$ unit root and $x_t = x_{t-1} + y_t, \quad x_0 = 0$, then

$$x_t = y_1 + y_2 + ... + y_t = I(2) \tag{45}$$

Differencing to undue integration

- When we difference a process of order I(d) we get an I(d-1) process:

Differencing to undue integration

- When we difference a process of order I(d) we get an I(d-1) process:
- **Difference an I(2) to get an I(1) process:** In (71) $x_t \sim I(2)$ and

$$x_t = x_{t-1} + y_t, \quad y_t \sim I(1)$$

Then solve for y_t:

$$y_t = x_t - x_{t-1} = \Delta x_t$$

Therefore

$$\Delta x_t \sim I(1)$$

Differencing to undue integration

- When we difference a process of order I(d) we get an I(d-1) process:
- **Difference an I(2) to get an I(1) process:** In (71) $x_t \sim I(2)$ and

$$x_t = x_{t-1} + y_t, \quad y_t \sim I(1)$$

Then solve for y_t:

$$y_t = x_t - x_{t-1} = \Delta x_t$$

Therefore

$$\Delta x_t \sim I(1)$$

- **Difference an I(1) process, to get an I(0) process:**

$$y_t = y_{t-1} + u_t, \quad u_t \sim I(0)$$

$$\Delta y_t = y_t - y_{t-1} = u_t \sim I(0)$$

Unit Root + Trend Model

- Some time series data both unit root behavior and a linear trend
- Think about stock prices:

Unit Root + Trend Model

- Some time series data both unit root behavior and a linear trend
- Think about stock prices:

  i) In the short-run it's argued that stock prices are approximately random walks, and they certainly look like a unit root when plotted.

  ii) On the other hand, over long periods, the level of all the major stock indices have been drifting upwards.This may be due to inflation, economic growth, and technological improvements.

  iii) So while over medium horizons stocks may trend up for a while and then down for a while, over the very long term stock prices have been dominated by an upward trend.

Unit Root + Trend Model

- Also think about GDP
  - Real business cycle theory argues that many permanent changes in GDP due to technology shocks and these are essentially random leading to unit root behavior.
  - On the other hand, while technological progress is random and sporadic, we call it progress because more often than not technology progresses forward rather than moving backward. And this tendency should cause an upward trend.
  - Other factors such as capital accumulation and population growth may also generate an upward trend.

Unit Root + Trend Model

- To model this type of data, we need to allow for both wandering behavior (unit roots) and linear trends:

$$y_t = \text{trend component} + \text{unit root component}$$

$$y_t = \alpha_0 + \alpha_1 t + v_t \quad (46)$$

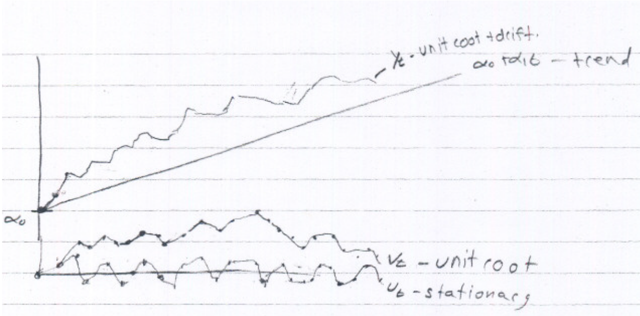
$$v_t = v_{t-1} + u_t \quad (47)$$

where

$$u_t \sim \text{stationary} \qquad \text{e.g:stationary ARMA(p,q) without intercept}$$

$$E[u_t] = 0 \quad (48)$$

Unit Root + Trend Model

- 

Unit Root + Trend Model

- **A trend in the level is an intercept in the difference**
- Subtract y_{t-1} from y_t to take a first difference

$$y_t = \alpha_0 + \alpha_1 t + v_t \tag{49}$$

$$-[\ y_{t-1} = \alpha_0 + \alpha_1 (t-1) + v_{t-1}\ ] \tag{50}$$

$$\Delta y_t = y_t - y_{t-1} = 0 + \alpha_1 (1) + \underbrace{v_t - v_{t-1}}_{u_t} \tag{51}$$

$$\Delta y_t = \alpha_1 + u_t \tag{52}$$

Unit Root + Trend Model

- So,e.g if $u_t = a_1 u_{t-1} + \varepsilon_t, \qquad |a_1| < 1$

$$
\begin{aligned}
\Delta y_t &= \alpha_1 + u_t \\
\Delta y_t &= \alpha_1 + a_1 \underbrace{u_{t-1}}_{\Delta y_{t-1} - \alpha_1} + \varepsilon_t \\
\Delta y_t &= \alpha_1 + a_1(\Delta y_{t-1} - \alpha_1) + \varepsilon_t \\
&= \underbrace{\alpha_1(1 - a_1)}_{\text{set } a_0 = \alpha_1(1-a_1)} + a_1 \Delta y_{t-1} + \varepsilon_t \\
\Delta y_t &= a_0 + a_1 \Delta y_{t-1} + \varepsilon_t \\
a_0 &= \alpha_1(1 - a_1) \qquad (53)
\end{aligned}
$$

Remarks

- The trend $(\alpha_0 + \alpha_1 t)$ in y_t adds an intercept $\alpha_1(1 - a_1)$ in Δy_t.

Remarks

- The trend $(\alpha_0 + \alpha_1 t)$ in y_t adds an intercept $\alpha_1(1 - a_1)$ in Δy_t.
- If y_t has no trend $(\alpha_1 = 0)$ then Δy_t has no intercept $(a_0 = \alpha_1(1 - a_1) = 0)$

Remarks

- The trend $(\alpha_0 + \alpha_1 t)$ in y_t adds an intercept $\alpha_1(1 - a_1)$ in Δy_t.
- If y_t has no trend $(\alpha_1 = 0)$ then Δy_t has no intercept $(a_0 = \alpha_1(1 - a_1) = 0)$
- Likewise if Δy_t has no intercept then y_t has no trend.

Remarks

- The trend $(\alpha_0 + \alpha_1 t)$ in y_t adds an intercept $\alpha_1(1 - a_1)$ in Δy_t.
- If y_t has no trend $(\alpha_1 = 0)$ then Δy_t has no intercept $(a_0 = \alpha_1(1 - a_1) = 0)$
- Likewise if Δy_t has no intercept then y_t has no trend.
- The expectation of the first difference (or change)

$$E\Delta y_t = \frac{a_0}{1 - a_1} = \frac{\alpha_1(1 - a_1)}{1 - a_1} = \alpha_1$$

  is given by the slope of the trend.
- That's intuitive: The greater the average increase in y_t $(E\Delta y_t)$, the faster y_t trend up.

Unit Root + Trend Model

- **Modeling Unit Root+ Trend with/ ARIMA(p,1,q)+ intercept**

$$\underbrace{(1-L)y_t}_{\Delta y_t} = a_0 + \sum_{i=1}^{p} a_i \underbrace{(1-L)y_{t-i}}_{\Delta y_{t-i}} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t$$

$$y_0 = 0; \quad \varepsilon_t \sim WN(0,\sigma^2); \quad (1-L)y_t \text{ stationary} \tag{54}$$

This implies that y_t is a unit root +trend.

Unit Root + Trend Model

- To gain intuition, let's look at the simplest ARIMA(p,1,q)+intercept model possible- the ARIMA(0,1,0)+ intercept model

$$\Delta y_t = a_0 + \varepsilon_t \quad y_0 = 0 \quad \varepsilon_t \sim WN(0, \sigma^2)$$

This is an ARIMA(0,1,0)+ intercept

$$\begin{aligned} y_t &= y_{t-1} + (y_t - y_{t-1}) && (55) \\ &= y_{t-1} + \Delta y_t && (56) \\ y_t &= y_{t-1} + a_0 + \varepsilon_t && (57) \end{aligned}$$

Unit Root + Trend Model

$$
\begin{aligned}
y_1 &= \underbrace{y_0}_{0} + a_0 + \varepsilon_1 && (58)\\
&= a_0 + \varepsilon_1 && (59)
\end{aligned}
$$

$$
\begin{aligned}
y_2 &= \underbrace{y_1}_{a_0+\varepsilon_1} + a_0 + \varepsilon_2 && (60)\\
&= (a_0 + \varepsilon_1) + a_0 + \varepsilon_2 && (61)\\
&= 2a_0 + \varepsilon_1 + \varepsilon_2 && (62)\\
&\dots && (63)\\
y_t &= \underbrace{ta_0}_{\text{trend}} + \underbrace{\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t}_{v_t} && (64)
\end{aligned}
$$

Unit Root + Trend Model

- Note that $v_t = \underbrace{(\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_{t-1})}_{v_{t-1}} + \varepsilon_t$

  $v_t = v_{t-1} + \varepsilon_t$ is a random walk
- So this is a random walk plus trend model

Forecasting in unit root plus trend model

- To forecast unit plus drift using 75:

$$
\begin{aligned}
y_{t+1} &= y_t + \Delta y_{t+1} && (65)\\
\widehat{y}_{t+1|t} &= y_t + \widehat{\Delta y}_{t+1|t} && (66)
\end{aligned}
$$

- So, no different than before, except now include intercept.

Unit Root theory

- Brief introduction to unit root asymptotic and tests

- **Standard Brownian Motion:** A standard Brownian Motion is a stochastic(ie. random), continuous time process that satisfies:
  (A.1) $w(0) = 0$
  (A.2) $w(t)$ has continuous realization.
  (A.3) The increment $w(t) - w(s) \sim N(0, t - s)$ for $t \geq s$
  (A.4) Disjoint increments, such as $w(t) - w(s)$ and $w(r) - w(v)$ where $t > s > r > v$ are independent.

Brief introduction to unit root asymptotic and tests

- **Intuition:** Continuous time counter part to random walk with respect to standard normal errors:

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim iidN(0,1) \\ y_0 &= 0 \end{aligned} \tag{67}$$

(A.1) $w(0) = 0 \leftrightarrow y_0 = 0$
(A.3) $w(t) - w(s) \sim N(0, t - s)$

$$\begin{aligned} y_t = y_{t-1} + \varepsilon_t &= y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ = \ldots &= y_{t-s} + \varepsilon_{s+1} + \varepsilon_{s+2} + \ldots + \varepsilon_t \\ var(y_t - y_s) &= var(\varepsilon_{s+1} + \varepsilon_{s+2} + \ldots + \varepsilon_t) \\ &= (t-s)\underbrace{var(\varepsilon_t)}_{1} \\ &= (t-s) \end{aligned}$$

- (A.4) For $t > s > r > v$:
  $w(t) - w(s)$ independent of $w(r) - w(v)$

$$
\begin{aligned}
y_t - y_s &= \varepsilon_{s+1} + \varepsilon_{s+2} + \ldots + \varepsilon_t \qquad (68) \\
y_r - y_v &= \varepsilon_{v+1} + \varepsilon_{v+2} + \ldots + \varepsilon_r \qquad (69)
\end{aligned}
$$

  (68) and (69) share no ε in common. So, all the ε in (68) are independent of those in (69).

- **Intuition:** If you observed $w(t)$ at discrete point in time $t = 0, 1, 2, 3, \ldots$ you would observe a random walk.

Brief introduction to unit root asymptotic and tests

- Define $[s]$ = greatest integer $\leq s$ (ie. round down function)
- Consider, the normal random walk model:

$$\begin{aligned} y_0 &= 0 \\ y_t &= y_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim iidN(0, \sigma^2) \\ y_t &= \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_t \sim N(0, t\sigma^2) \end{aligned} \quad (70)$$

Brief introduction to unit root asymptotic and tests

- Suppose we have data: $y_1, y_2, ..., y_t$
- And we divide y_t by $\sqrt{T}\sigma$

$$\begin{aligned} y_t &\sim N(0, t\sigma^2) \\ \frac{y_t}{\sigma\sqrt{T}} &\sim N(0, \frac{t}{T}) \end{aligned} \tag{71}$$

- Now consider $w(\frac{t}{T})$:

$$\begin{aligned} w(0) &= 0 \\ w(\frac{t}{T}) &= w(\frac{t}{T}) - w(0) \sim N(0, \frac{t}{T} - 0) \\ w(\frac{t}{T}) &\sim w(0, \frac{t}{T}) \end{aligned} \tag{72}$$

Brief introduction to unit root asymptotic and tests

- Putting (71) and (72) together:

$$\frac{y_t}{\sigma\sqrt{T}} \sim w(\frac{t}{\sqrt{T}}) \tag{73}$$

- This provides the intuition for **Donsker's Theorem** which shows that in the random walk model:

$$\begin{aligned} y_0 &= 0 \\ y_t &= y_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim WN(0, \sigma^2) \end{aligned} \tag{74}$$

$$\frac{y_t}{\sigma\sqrt{T}} \underset{\sim}{approx} w(\frac{t}{T}) \tag{75}$$

Brief introduction to unit root asymptotic and tests

- Now, this turns out to have some implications that form the starting point for unit root tests.
- Note that if we square (75), we get:

$$\frac{y_t^2}{\sigma^2 T} \underset{approx}{\sim} w(\frac{t}{T})^2 \tag{76}$$

- Now recall that$\sum_{t=1}^{T} f(\frac{t}{T})\frac{1}{T} \underset{approx}{=} \int_0^1 f(r)dr$.
- And consider

$$\frac{1}{\sigma^2 T^2}\sum_{t=1}^{T} y_t^2 = \sum_{t=1}^{T}(\frac{y_t}{\sigma^2 T})\frac{1}{T} \underset{d}{\sim} \sum_{t=1}^{T} w(\frac{t}{T})^2\frac{1}{T} \underset{approx}{=} \int_0^1 w(r)^2 dr$$

Brief introduction to unit root asymptotic and tests

- So we have

$$\frac{1}{T}\sum_{t=1}^{T} y_t^2 \underset{approx}{\sim} \sigma^2 \int w(r)^2 dr \tag{77}$$

- Next, consider writing

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

- And solve for $y_{t-1}\varepsilon_t$ as:

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

- And sum over t to get

$$\sum_{t=1}^{T} y_{t-1}\varepsilon_t = \frac{1}{2}\sum_{t=1}^{T}(y_t^2 - y_{t-1}^2) - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t^2 \qquad (78)$$

- Now consider the first term on RHS of (78)

$$
\begin{aligned}
\sum_{t=1}^{T}(y_t^2 - y_{t-1}^2) &= (y_1^2 - y_0^2) + (y_2^2 - y_1^2) + \dots \\
&\quad + (y_{T-1}^2 - Y_{T-2}^2) + (y_T^2 - y_{T-1}^2) \\
&\quad \text{and noting the cancellations} \\
&= y_T^2 - y_0^2 \\
&= y_T^2 \qquad (\text{noting } y_0 = 0)
\end{aligned}
$$

- So plugging this back into (78)

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

- Now divide by T:

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

- And note

$$\frac{1}{T}\sum \varepsilon_t^2 \underset{approx}{=} \sigma^2 = var(\varepsilon_t) \tag{79}$$

$$\frac{y_T^2}{T} \underset{approx}{\sim} \sigma^2 w(1)^2 (\text{by}(76)) \tag{80}$$

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \underset{approx}{\sim} \frac{1}{2}\sigma^2 (w(1)^2 - 1) \tag{81}$$

Brief introduction to unit root asymptotic and tests

- So far, this probably just seems like a bunch of aimless algebra but (77) and (81) give us the tools to analyze autoregression with respect to unit roots.
- Let y_t be given as in (74) consider estimating the auto regression:

$$y_t = a_1 y_{t-1} + \varepsilon_t \tag{82}$$

- We know the true $a_1 = 1$ to match (74) (ie. a unitroot)
- But how does the estimate $\hat{a}_1$ behave?

$$\hat{a}_1 = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

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

Brief introduction to unit root asymptotic and tests

-

$$\hat{a}_1 = \frac{\sum_{t=1}^{T} y_{t-1}(y_{t-1}+\varepsilon_t)}{\sum_{t=1}^{T} y_{t-1}^2} \quad (83)$$

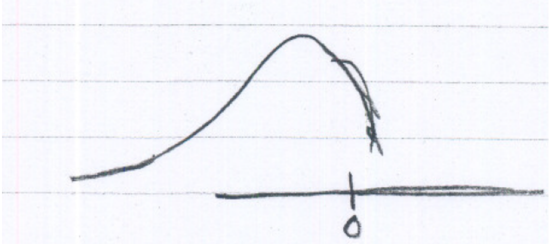
$$= 1 - \frac{\sum_{t=1}^{T} y_{t-1}(y_{t-1}+\varepsilon_t)}{\sum_{t=1}^{T} y_{t-1}^2} \quad (84)$$

(Noting the cancellation)

$$(\underbrace{\hat{a}_1}_{estimate} - \underbrace{1}_{\text{true}}) = \frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

$$T(\hat{a}_1 - 1) = \frac{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2}$$

- $$T(\hat{a}_1 - 1) \underset{\text{approx}}{\sim} \frac{\frac{1}{2}\sigma^2(w(1)^2 - 1)}{\sigma^2 \int w(r)^2 dr} = DF \tag{85}$$
- The RHS of (86) describes a probability distribution known as **Dickey-Fuller** or **unit root** distribution.
- It looks something like:



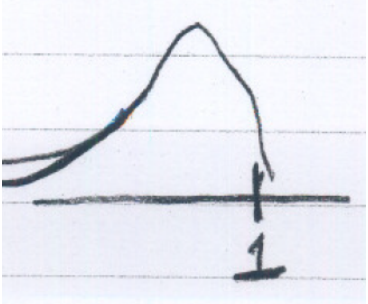
Brief introduction to unit root asymptotic and tests

- **Intuition for non-standard distribution**

$$w(1) \sim N(0,1) \Rightarrow \qquad w(1)^2 \sim \chi_1^2$$

$$\text{Also} \int w(r)^2 dr \text{is a random variable}$$

- Which means that the sampling distribution of $hata_1$ looks like:



- Thus, the behavior $\hat{a}_1$ when $a_1 = 1$(unit root) differs substantially from the usual behavior of regression coefficients in 2 key ways:
- **(A)** The Dick-Fuller Distribution is **not** the usual normal distribution and it is **not** even **symmetric**. And seen in the figure the probability of $\hat{a}_1 < 1$ is much greater than the prob of $\hat{a}_1 > 1$
- In practice, we rarely observe $\hat{a}_1 \geq 1$. So the face that $\hat{a}_1 < 1$ does not ensure that y_t is stationary. This needs to be tested.

- **(B)** $T(\hat{a}_1 - 1) =$ random variable, so $\hat{a}_1$ collapses to 1 at rate T,which is faster than the usual $\sqrt{T}$ rate. This is known as the **super-consistency property**.

Brief Introduction to Unit Root Tests

- Brief introduction to unit root tests
- The Dickey-Fuller unit root test
- Tests Based on the Standard t-statistic but nonstandardcritical values
- In practice: Add at least an intercept
- We can also add a linear trend
- The Augmented Dickey Fuller (ADF) Test
- Variants of the DF test(ADF test)
- Unit Root in AR(2) model and relation to ADF test

Brief introduction to unit root tests

- Suppose we OLS to estimate the AR(1) model without intercept

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

Brief introduction to unit root tests

- Suppose we OLS to estimate the AR(1) model without intercept

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

- When the true process for y_t is a random walk ($a_1 = 1$):

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

- And then we ask, what is the distribution of $\hat{a}_1$?

Brief introduction to unit root tests

- Suppose we OLS to estimate the AR(1) model without intercept

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

- When the true process for y_t is a random walk ($a_1 = 1$):

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

- And then we ask, what is the distribution of $\hat{a}_1$?
- Is it normally distributed?

Brief introduction to unit root tests

- Suppose we OLS to estimate the AR(1) model without intercept

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

- When the true process for y_t is a random walk ($a_1 = 1$):

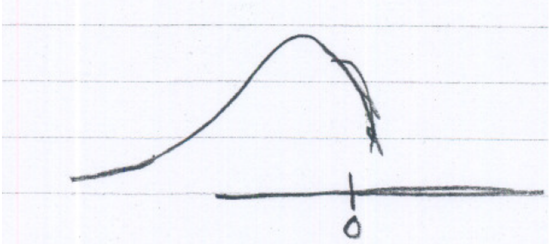
$$y_t = y_{t-1} + \varepsilon_t$$

- And then we ask, what is the distribution of $\hat{a}_1$?
- Is it normally distributed?
- Turns out the answer is: NO

- In the notes called "unit root theory" we derived:

$$T(\hat{a}_1 - 1) \underset{\text{approx}}{\sim} \frac{\frac{1}{2}\sigma^2(w(1)^2 - 1)}{\sigma^2 \int w(r)^2 dr} = DF \tag{86}$$

- The RHS of (86) describes a probability distribution known as **Dickey-Fuller** or **unit root** distribution.
- It looks something like:



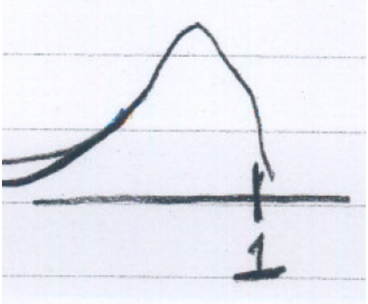
Brief introduction to unit root tests

- **Intuition for non-standard distribution**

$$w(1) \sim N(0,1) \Rightarrow \qquad w(1)^2 \sim \chi_1^2$$

$$\text{Also} \int w(r)^2 dr \text{ is a random variable}$$

- Which means that the sampling distribution of $\hat{a}_1$ looks like:



Thus, the behavior $\hat{a}_1$ when $a_1 = 1$ (unit root) differs substantially from the usual behavior of regression coefficients in 2 key ways:

1. The Dick-Fuller Distribution is **not** the usual normal distribution and it is **not** even **symmetric**. And seen in the figure the probability of $\hat{a}_1 < 1$ is much greater than the prob of $\hat{a}_1 > 1$

   In practice, we rarely observe $\hat{a}_1 \geq 1$. So the face that $\hat{a}_1 < 1$ does not ensure that y_t is stationary. This needs to be tested.

Thus, the behavior $\hat{a}_1$ when $a_1 = 1$ (unit root) differs substantially from the usual behavior of regression coefficients in 2 key ways:

1. The Dick-Fuller Distribution is **not** the usual normal distribution and it is **not** even **symmetric**. And seen in the figure the probability of $\hat{a}_1 < 1$ is much greater than the prob of $\hat{a}_1 > 1$

   In practice, we rarely observe $\hat{a}_1 \geq 1$. So the face that $\hat{a}_1 < 1$ does not ensure that y_t is stationary. This needs to be tested.

2. $T(\hat{a}_1 - 1) =$ random variable, so $\hat{a}_1$ collapses to 1 at rate T,which is faster than the usual $\sqrt{T}$ rate. This is known as the **super-consistency property**.

The Dickey-Fuller unit root test

$$y_t = a_0 y_{t-1} + \varepsilon_t$$

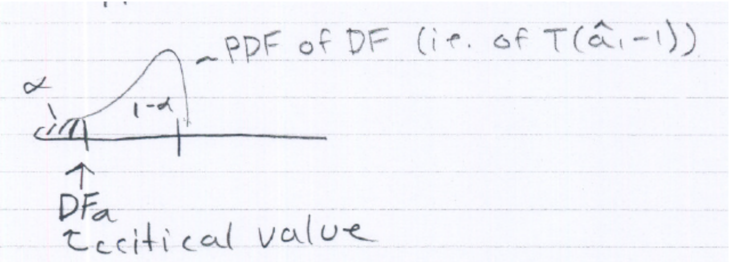
$$\begin{array}{lll} H_o: & a_1 = 1 & \text{unit root} \\ H_A: & a_1 \leq 1 & \text{stationary} \end{array}$$

Test strategy:

$$T(\hat{a}_1 - 1) \underset{H_o \; approx}{\sim} DF$$

(If H_o holds, $\hat{a}_1$ follows DF distribution)

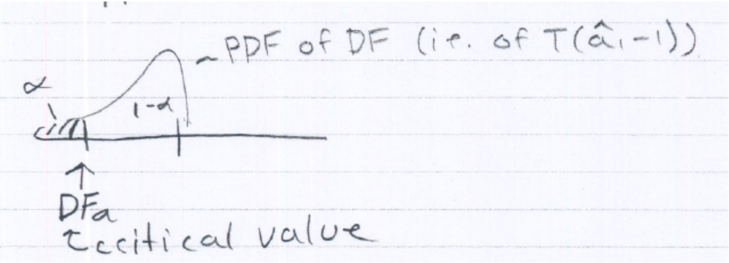
- The PDF of the test statistics looks like:



Reject if $T(\hat{a}_1 - 1) < DF_\alpha$

The Dickey-Fuller unit root test

- The PDF of the test statistics looks like:



Reject if $T(\hat{a}_1 - 1) < DF_\alpha$

- Just like there are probability tables for the normal, t and F distribution. So too is there a table (actually several) for the DF distribution.

Rewriting the auto-regression in DF form

- In unit root testing it is common to rewrite (86) as follows:

$$
\begin{aligned}
y_t &= a_1 y_{t-1} + \varepsilon_t \\
\underbrace{y_t - y_{t-1}}_{\Delta y_t} &= \underbrace{(a_1 - 1)}_{\gamma_1} y_{t-1} + \varepsilon_t
\end{aligned}
$$

$$
\begin{aligned}
\Delta y_t &= \gamma_1 y_{t-1} + \varepsilon_t \qquad (87) \\
H_o : \gamma_1 &= 0 \Leftrightarrow a_1 = 1 \\
H_A : \gamma_1 &< \Leftrightarrow a_1 < 1 \qquad (88)
\end{aligned}
$$

Then

$$T\hat{\gamma}_1 = T(\hat{a}_1 - 1)$$

So we can base our test on $T\hat{\gamma}_1$

Tests Based on the Standard t-statistic but nonstandard critical values

- **Test based on the standard t statistics for a test of $\gamma_1 = 0$**
  1. t-stat is calculated in the usual way and it just standard t-stat
  2. But it does **Not have a t-distribution**. It always like the picture we attached before.

- **Tests incorporating as intercept:**
- We first considered no intercept because it was the simplest case to derive the theory for[1]
- But, in practice omitting the intercept is usually too restrictive
- For example, what if the series is stationary around a mean that is different from zero
- A more realistic specification is

$$\Delta y_t = \alpha_0 + \gamma_1 y_{t-1} + \varepsilon_t \quad (89)$$

$$H_0: \gamma_1 = 0$$
$$H_A: \gamma_1 < 0$$

[1] See the companion notes in the file "unit_root_theory.pdf"

In practice: Add at least an intercept (continued)

- Tests can again be based on either $T\hat{\gamma}_1$ or the standard t-stat
- These again have non-standard DF type distributions
- But they are bit different, so that different critical values are required.

We can also add a linear trend

$$\Delta y_t = \alpha_0 + \alpha_1 t + \gamma_1 y_{t-1} + \varepsilon_t \qquad (90)$$

$$\begin{aligned} H_0: \gamma_1 &= 0 \\ H_A: \gamma_1 &< 0 \end{aligned}$$

- Under the alternative hypothesis y_t is trend stationary rather than stationary. Recall

  Trend Stationary = Linear Trend + Mean zero stationary component

The Augmented Dickey Fuller (ADF) Test

- **The Augmented DF(ADF) test**
- Note DF test assumes $\varepsilon \sim WN$
- Appropriate for random walk, but not for general unit root.
- Usually we have to add more lags of y_t or Δy_t before it's safe to assume $\varepsilon \sim WN$.
- This is exactly what the Augmented DF test does.

The Augmented Dickey Fuller (ADF) Test

- The ADF test is based on:

$$\Delta y_t = \alpha_0 + \alpha_1 t + \gamma_1 y_{t-1} + \sum_{i=1}^{p} a_i \Delta y_{t-i} + \varepsilon_t$$

$$\begin{aligned} H_o: &\quad \gamma_1 = 0 \\ H_A: &\quad \gamma_1 \leq 0 \end{aligned}$$

- The Δy_{t-i} act like a strainer that filters out the persistence in Δy_t so that $\varepsilon_t \sim WN$
- The term $\alpha_1 t$ may dropped if no linear trend.
- The critical values are the same as in the DF test and don't depend on lag length p.

Variants of the DF test(ADF test)

- **When to use DF vs ADF**

$$H_o: \quad \text{Random walk} \rightarrow \text{use DF}$$
$$H_o: \quad \text{General unit root} \rightarrow \text{use ADF}$$

- If in doubt, use the ADF
- A more recent test called ADF-GLS offers power improvements and is likely the best choice. We cover this in a separate handout.

Unit Root in AR(2) model and relation to ADF test

- How do we define/test for a unit root in the AR(2) model?:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t \tag{91}$$

Unit Root in AR(2) model and relation to ADF test

- How do we define/test for a unit root in the AR(2) model?:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t \qquad (91)$$

- Add/subtract $a_2 y_{t-1}$ to this AR(2) model:

$$\begin{aligned} y_t &= a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_2 y_{t-1} - a_2 y_{t-1} + \varepsilon_t && (92) \\ &= a_0 + (a_1 + a_2) y_{t-1} - a_2 (y_{t-1} - y_{t-2}) + \varepsilon_t && (93) \\ &= a_0 + (a_1 + a_2) y_{t-1} - a_2 \Delta y_{t-1} + \varepsilon_t && (94) \end{aligned}$$

$$\Delta y_t = y_t - y_{t-1} = a_0 + (a_1 + a_2 - 1) y_{t-1} - a_2 \Delta y_{t-1} + \varepsilon_t$$

Unit Root in AR(2) model and relation to ADF test

- How do we define/test for a unit root in the AR(2) model?:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t \tag{91}$$

- Add/subtract $a_2 y_{t-1}$ to this AR(2) model:

$$\begin{aligned} y_t &= a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_2 y_{t-1} - a_2 y_{t-1} + \varepsilon_t & (92) \\ &= a_0 + (a_1 + a_2) y_{t-1} - a_2 (y_{t-1} - y_{t-2}) + \varepsilon_t & (93) \\ &= a_0 + (a_1 + a_2) y_{t-1} - a_2 \Delta y_{t-1} + \varepsilon_t & (94) \end{aligned}$$

$$\Delta y_t = y_t - y_{t-1} = a_0 + (a_1 + a_2 - 1) y_{t-1} - a_2 \Delta y_{t-1} + \varepsilon_t \tag{95}$$

- Equation (95) has ADF form.
- y_t will be mean-reverting (stationary) when $a_1 + a_2 < 1$ and has a unit root when $a_1 + a_2 = 1$

Cointegration

- Co-trending and cointegration
- Cointegration regression
- Testing for cointegration
- Spurious Regression
- Cointegration and error correction
- Vector error correction mode

Co-trending and cointegration

- **Co-trending**: Two variables, say y_{1t} and y_{2t} share a common trend.
- **Cointegration**: y_{1t} and y_{2t} share a common *stochastic* trend:

$$y_{1t} \sim I(1) \; (unitroot)$$
$$y_{2t} \sim I(1).$$

  But some linear combination, e.g. $y_{1t} - y_{2t}$, is stationary about the mean.
- **Analogy**:
  - Random walk; A dunk man walk
  - Conintegration: Two drunk men walking and talking

Cointegration: a general form

- Let

$$\tilde{\beta} = \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{bmatrix} \; and \; y_t = \begin{bmatrix} \tilde{y}_{1t} \\ \tilde{y}_{2t} \end{bmatrix}.$$

A more general linear combination takes the form:

$$\tilde{\beta}' y_t = [\tilde{\beta}_1, \tilde{\beta}_2] \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \tilde{\beta} y_{1t} + \tilde{\beta}_2 y_{2t}.$$

- Let $\tilde{U}_t$ be a stationary process, e.g a stationary ARMA, with mean zero.
- Then, if

$$\tilde{\beta}' y_t = \tilde{\beta}_0 + \tilde{U}_{0t}$$
$$\tilde{U}_{0t} \sim I(0) \; (stationary$$
$$y_{1t}) \sim I(1) \; \& \; y_{2t} \sim I(1),$$

we say that $y_t = (y_{1t}, y_{2t})'$ is cointegrated with cointegrating vector $\tilde{\beta}$.

Cointegration regression

- Note that linear combination of stationary random variables are stationary.
- So, if $\tilde{\beta}_1 \neq 0$, we have

$$\frac{1}{\tilde{\beta}_1}[\tilde{\beta}' y_t] = \frac{1}{\tilde{\beta}_1}[\tilde{\beta}_0 + \tilde{U}_{0t}] = \frac{\tilde{\beta}_0}{\tilde{\beta}_1} + \frac{1}{\tilde{\beta}_1}\tilde{U}_{0t}$$

is also stationary.

$$\Longrightarrow \frac{1}{\tilde{\beta}_1}[\tilde{\beta}] = \frac{1}{\tilde{\beta}_1}\begin{bmatrix}\tilde{\beta}_1 \\ \tilde{\beta}_2\end{bmatrix} = \begin{bmatrix}1 \\ \frac{\tilde{\beta}_2}{\tilde{\beta}_1}\end{bmatrix}$$

is also a conintegrating vector.

Cointegration regression cont.

- In other words,

$$\beta' y_t = [1, -\beta_1] \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \frac{1}{\tilde{\beta}_1} [\tilde{\beta}_1, \tilde{\beta}_2] \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \frac{1}{\tilde{\beta}_1} [\tilde{\beta}_0 + \tilde{U}_{0t}].$$

Define $\beta_0 = \frac{\tilde{\beta}_0}{\tilde{\beta}_1}$ and $U_{0,t} = \frac{1}{\tilde{\beta}_1} \tilde{U}_{0t}$ also stationary,

$$\Longrightarrow \beta' y_t = [1, -\beta_1] \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = y_{1t} - \beta_1 y_{2t} = \beta_0 + U_{0t}, \tag{96}$$

$$\left.\begin{aligned} \Longrightarrow y_{1t} &= \beta_0 + \beta_1 y_{2t} + U_{0,t} \\ U_{0,t} &\sim \text{ stationary} \\ y_{2t} &\sim I(1) \end{aligned}\right\} \text{Cointegrating Regression.}$$

- To make the cointegration regression more complete

$$\left.\begin{aligned} y_{1t} &= \beta_0 + \beta_1 y_{2t} + U_{0,t} \\ y_{2t} &= y_{2t-1} + U_{2t}, y_0 = 0 \\ U_{0,t} \;\&\; U_{2,t} &\text{ stationary} \end{aligned}\right\} \tag{97}$$

Intepretation of cointegration regression

- Capture long-run comovements of y_{1t}, y_{2t}
- Short-run dynamics require specification of the models (e.g ARMA(p.q)) for U_{0t}.
- Note that U_{0t} is neither i.id nor WN. So it does not have the usual interpretation of a regression error in the traditional sense.
- As a result, y_{1t} & y_{2t} do not have the usual interpretation as dependent and independent variable.
- A few results, which I leave as homework for you to demonstrate underscore this. They are:
  Homework: Let (97) hold
  - Show that for $\beta_1 \neq 0$ (97) may be rewritten as cointegrating in which Y_{2t} is the LHS variable & y_{1t} is the RHS variable.
  - SHow that (97) may be written as a conintegrating regression of y_{1t} on Y_{2t-1},
  - SHow that for $\beta_1 \neq 0$ (97) can also be written as a cointegrating regression of y_{2t} on y_{1t-1}.

Estimation of cointegration regression

- Estimation by OLS has some problems Properties of $\hat{\beta}_{1,OLS}$ in (97)

$$= T(\hat{\beta}_{1,OLS} - \beta) = \frac{T^{-1}\sum_{t=1}^{T} y_{2t}U_{0t}}{T^{-2}\sum_{t=1}^{T} y_{2t}^2} \xrightarrow{d} \begin{cases} \text{General Nonstandard} \\ \text{Distribution} \\ \Downarrow \\ \text{Not the same as the unit root} \\ \text{distribution but some what like it.} \end{cases}$$

- Consequences:
  - Super-consistency [good]
  - Non-standard distribution [problem]:
    - second order bias
    - standard error not right
    - confidence interval not right
    - t & F tests not right

Correction to OLS

- Various corrections to OLS are available to ensure that the corrected estimator, say $\hat{\beta}_{1,c}$ satisfies:

$$T(\hat{\beta}_{1,c} - \beta) \overset{approx}{\sim} \text{Normal}.$$

- These corrections would take some time to explain in detail, but are avaialble in software sych as Eviews, Matlab, R. Perhaps, the most common are:
  - Fully Modified least Square, (FM-OLS)
  - Dynamic Least Square (Leads/Lags), (DOLS)
- The DOLS is referred to as a lags & leads estimator because it adds lags & leads of Δy_{2t} into the regression to strain out the impatc of the cointegrating residual $U_{0,t}$ that correlated with $\Delta y_{2t \pm j}$,

$$y_{1t} = \beta_0 + \beta_1 y_{2t} + \sum_{j=-p}^{p} \delta_j \Delta y_{2t-j} + V_{0,t}, \ (DOLS).$$

- Then robust (HAC) standard errors can be used to construct inference on β_1 (see Stock Watson for futher details)

Testing for cointegration

- The key assumptions underlying (97) are:
  1. y_{1t} and y_{2t} have unit roots.
  2. $U_{0,t}$ is stationary $\iff$ $(\text{y}_{1t}, y_{2,t})$ is cointegrated.
- Before estimating & using cointegrating regression, both assumptions must be tested.
- Assumption (1) shoudl first be tested using the unit root tests dicussed previously.
- If we rejrect unit roots, (1) fails & we don't employ cointegration methods.
- If we fail to reject unit roots in y_{1t} & y_{2t}, we proceed to test Assumption (2) using a cointegreting test as discussed in the next slide.
- Only if both A.1 and A.2 are verified, we can safely proceed to use cointegration methodology.

Cointegration test when β_1 is known

- Key insight:
  - (y_{1t}, y_{2t}) cointegrated with cointegrating vector (1, -β_1)
  - $\Longleftrightarrow \beta_0 + U_{0,t} = y_{1t} - \beta_1 y_{2t} \sim$ stationary.
- Method:
  - Step 1. Form $V_{0,t} = \beta_0 + U_{0,t} = y_{1t} - \beta_1 y_{2t}$.
  - Step 2. Test $U_{0,t}$, for a unit root in $V_{0,t}$ including an intercept , but no trend.
  - Step 3. Use the result to determine cointegration:
    1. Fail to reject unit root $\Longrightarrow$ suggests $U_{0,t}$ may not be stationary $\Longrightarrow$ Fail to reject no cointegration $\Longrightarrow$ Do not use cointegration methods.
    2. Reject the unit root $\Longrightarrow$ $U_{0,t}$ stationary $\Longrightarrow$ reject null hypothesis of no cointegration $\Longrightarrow$ cointegration $\Longrightarrow$ use cointegration methods.
- Formal Hypothesis
- H_0: (y_{1t}, y_{2t}) not cointegrated with conitegrating vector (1, -β_1) $\Longleftrightarrow$ $U_{0,t}$ has a unit root.
- H_A: (y_{1t}, y_{2t}) cointegrated with conitegrating vector (1, -β_1) $\Longleftrightarrow$ $U_{0,t}$ is stationary.

Cointegration test when β_1 is not known

- $$\left.\begin{array}{l} \text{H}_0\text{: } (y_{1t}, y_{2t}) \text{ not cointegrated} \\ \iff y_{1t} - \beta_1 y_{2t} \text{ has unit root for any } \beta_1 \end{array}\right\} \text{No-cointegration (Null)}$$
  $$\left.\begin{array}{l} \text{H}_A\text{: } (y_{1t}, y_{2t}) \text{ cointegrated} \\ \iff y_{1t} - \beta_1 y_{2t} \text{ is stationary for some } \beta_1 \end{array}\right\} \text{Yes-cointegration (Alt)}$$
- Key Points
  * Still test $y_{1t} - \beta_1 y_{2t}$ for unit root
    - $y_{1t} - \beta_1 y_{2t} \sim \text{I}(1) \implies$ no conitegration.
    - $y_{1t} - \beta_1 y_{2t} \sim \text{I}(0) \implies$ conitegration.
  * But now, must estinmate $\hat{\beta}_1$ first. So essentially conduct unit root test on $y_{1t} - \hat{\beta}_1 y_{2t}$.
  * However, estimating $\hat{\beta}_1$ changes the distribution of the unit root tests on $y_{1t} - \hat{\beta}_1 y_{2t}$.
  * Therefore, alternative critical values must be employed.
- The Engle-Granger- ADF Test (EG-ADF) is a cointegration test that follows exactly the procedure discussed above, providing teh correct critical values. It is available on programs such as Eviews, Matlab, R. See the Stock & Watson textbook for futher details.

Spurious Regression

- Discoverd by Granger-Newbold (1974) and explained by Phillips (1986, 1998).
- Consider two unit roots processes, y_{1t} & y_{2t}, that are not cointegrated and in fact have no relation whatsoever ▸▸ Graph
- Say we run the regression: $y_{1t} = \hat{\beta}_0 + \hat{\beta}_1 y_{2t}$
- In principle, there is no relation between y_{1t} & y_{2t}, so we ought to find $\beta_1 = 0$ & $R^2 = 0$.
- However, Granger & Newbold (1974) found that
  1. $\hat{\beta}_1$ tends to be significantly different from zero.
  2. The R^2 tends to be large.
- Thus the regression is spurious in the sense that it uncovers a relationship between unrelated series.
- Phillips (1998) provides smome intuition: The stochastic trends in y_{1t} are partially picked up by the stochastic trends in y_{2t} despite the lacks of any causal relation between two.

Spurious Regression & the importance of cointegration and unit root tests

- Spurious regression can provide highly misleading results
- Spurious regression occurs when both your y (y_{1t}) and x (x_{1t}) variables:
  1. have unit roots
  2. are not cointegrated
- So if you run a time series regression of y_{1t} on x_t (or x_{t-1}) without testing for unit roots & cointegration, you risk running a spurious regression.

Cointegration and error correction

- Suppose $y_{1t} = \beta_1 y_{2t} + U_{0,t}$ and $U_{0,t}$ is stationary
- Long-run comovement: y_{1t} & y_{2t} move together over the long-run.
- Now consider modeling Δy_{1t} & Δy_{2t}.
- Suppose Δy_{1t} & Δy_{2t} as autoregressions, e.g.

$$\Delta y_{1t} = a_1(0) + \sum_{i=1}^{p} a_{11}(i)\Delta y_{1t-i} + \varepsilon_{1t}$$

$$\Delta y_{2t} = a_2(0) + \sum_{i=1}^{p} a_{22}(i)\Delta y_{2t-i} + \varepsilon_{2t}.$$

- Then where is the mechanism by which y_{1t} & y_{2t} move together in the long run?

Cointegration and error correction Cont.

- We can add this by as what is known as an error correction term. To simplify, we omit AR terms for now,

$$\begin{aligned}
\Delta y_{1t} &= a_0 + \overbrace{\alpha_1 (y_{1t-1} - \beta_1 y_{2t-1})}^{U_{0t}} + \varepsilon_{1t} \\
\Delta y_{2t} &= b_0 + \underbrace{\alpha_2 (y_{1t-1} - \beta_1 y_{2t-1})}_{U_{0t}} + \varepsilon_{2t}.
\end{aligned}$$

- $y_{1t-1} - \beta_1 y_{2t-1}) = 0 \implies (y_{1t}, y_{2t})$ in long-run equilibrium $\implies$ no adjustment needed
- $y_{1t-1} - \beta_1 y_{2t-1}) \neq 0 \implies$ Out of equilibrium $\implies$ Either Δy_{1t} and/or Δy_{2t} should adjust to restore equilibrium.

Cointegration and error correction: an example

- Assume $\beta_1 > 0$, $\alpha_1 < 0$, $\alpha_2 > 0$.
- Suppose $y_{1,t-1} > \beta_1 y_{2,t-1}$:
  - (A)
    1. $y_{1,t-1} - \beta_1 y_{2,t-1} > 0$
    2. $\alpha_1 < 0$
    3. $\underset{(-)}{\alpha_1} - \underbrace{(y_{1,t-1} - \beta_1 y_{2,t-1})}_{(+)} = (-)$.

    We expect a negative influcence on Δy_{1t} pushing y_{1t} back down to $\beta_1 y_{2t}$ to restore equilibrium.
  - (B)
    1. $y_{1,t-1} - \beta_1 y_{2,t-1} > 0$
    2. $\alpha_2 > 0$
    3. $\underset{(+)}{\alpha_2} - \underbrace{(y_{1,t-1} - \beta_1 y_{2,t-1})}_{(+)} = (+)$.

    We expect a positive influcence on Δy_{2t} pushing $\beta_1 y_{2t}$ back up to the equilibrium value.
  - Terminology: $\beta_1 \equiv$ conintegrating coefficient. $y_{1,t-1} - \beta_1 y_{2,t-1} =$ deviation from long-run equilibrium. $\alpha_2(y_{1,t-1} - \beta_1 y_{2,t-1}) =$ error correction term. $\alpha_1, \alpha_2 =$ speed of adjustment parameters.

Vector error correction model

- In practivce, we combine the error correction term with lags of both Δy_{1t} and Δy_{2t},

$$\Delta y_{1t} = a_1(0) + \alpha_1(y_{1,t-1} - \beta_1 y_{2,t-1}) + \sum_{i=1}^{p} a_{11}(i)\Delta y_{1t-i} + \sum_{i=1}^{p} a_{12}(i)\Delta y_{2t-i} + \varepsilon_{1t} \tag{98}$$

$$\Delta y_{2t} = a_2(0) + \alpha_2(y_{1,t-1} - \beta_1 y_{2,t-1}) + \sum_{i=1}^{p} a_{21}(i)\Delta y_{1t-i} + \sum_{i=1}^{p} a_{22}(i)\Delta y_{22t-i} + \varepsilon_{2t}.$$

- The model above is known as a Vector Error Correction Model (VECM). To see why, we proceed to next slide

Vector error correction model Cont.

- To see why (98) is known as a VECM, we rewrite (98) in a vector (or matrix) form:

$$\underbrace{\begin{bmatrix}\Delta y_{1t}\\ \Delta y_{2t}\end{bmatrix}}_{\Delta y_t} = \underbrace{\begin{bmatrix}a_1(0)\\ a_2(0)\end{bmatrix}}_{A(0)} + \begin{bmatrix}\alpha_1(y_{1,t-1}-\beta_1 y_{2,t-1})\\ \alpha_2(y_{1,t-1}-\beta_1 y_{2,t-1})\end{bmatrix} + \sum_{i=1}^{p} \underbrace{\begin{bmatrix}a_1 1(i) & a_{12}(i)\\ a_{21}(i) & a_{22}(i)\end{bmatrix}}_{A(i)} \underbrace{\begin{bmatrix}\Delta y_{1t-i}\\ \Delta y_{2t-i}\end{bmatrix}}_{\Delta y_{t-i}} + \underbrace{\begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{bmatrix}}_{\varepsilon_t}$$

$$\Longrightarrow \Delta y_t = A(0) + \begin{bmatrix}\alpha_1(y_{1,t-1}-\beta_1 y_{2,t-1})\\ \alpha_2(y_{1,t-1}-\beta_1 y_{2,t-1})\end{bmatrix} + \sum_{i=1}^{p} A(i)\Delta y_{t-i} + \varepsilon_t.$$

- Note that:

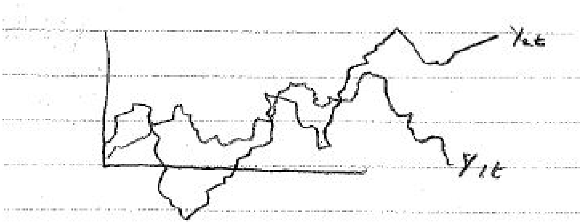
$$\begin{bmatrix}\alpha_1(y_{1,t-1}-\beta_1 y_{2,t-1})\\ \alpha_2(y_{1,t-1}-\beta_1 y_{2,t-1})\end{bmatrix} = \underbrace{\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}}_{\substack{\alpha\\ 2\times 1}} \underbrace{\begin{bmatrix}1 & -\beta_1\end{bmatrix}}_{\substack{\beta'\\ 1\times 2}} \underbrace{\begin{bmatrix}y_{1,t-1}\\ y_{2,t-1}\end{bmatrix}}_{\substack{y_{t-1}\\ 2\times 1}}$$

$$\Longrightarrow \underbrace{\Delta y_t}_{2\times 1} = \underbrace{A(0)}_{2\times 1} + \underbrace{\alpha}_{2\times 1} \underbrace{\beta'}_{1\times 2} \underbrace{y_{t-1}}_{2\times 1} + \sum_{i=1}^{p} \underbrace{A(i)}_{2\times 2} \underbrace{\Delta y_{t-i}}_{2\times 1} + \underbrace{\varepsilon_t}_{2\times 1}.$$

This is the VECM in its matrix form.

VECM estimation

- Estimation:
  - Step 1: estimate $\hat{\beta}$ using cointegrating regression.
  - Step 2: estimate $A(\hat{0})$, $\hat{\alpha}$, $A(\hat{i})$ using a regression with β replaced by $\hat{\beta}$.
- Note that step (89) can be done by two separate regressions or by one vector regression as they will be numerically equivalent.



Back