Time Series Analysis

Regression Analysis of Time Series, Cointegration and Error Correction Mechanisms

Michael Colaresi
colaresi@msu.edu

Outline

Outline

1. Regression Analysis of Time Series
   - M-test
2. Nonstationarity and Cointegration
   - An alternative procedure: cointegration
   - Alternative means for dealing with unit roots

Outline

ADL Specification

General ADL

- It is a simple extension to add (literally) AR and MA components to the ADL model. For example, a general ADL model could be written:

$$y_t = \alpha + \sum_{k=1}^{K} \sum_{n=0}^{N} \beta_{kn} X_{k,t-n} + \frac{\theta(B)}{\phi(B)} \epsilon_t \qquad (1)$$

where

$$\frac{\theta(B)}{\phi(B)} = \frac{1 + \theta_1 B + \ldots + \theta_\infty B^\infty}{1 + \phi_1 B + \ldots + \phi_\infty B^\infty} \qquad (2)$$

Practical ADL

Practical ADL

- However, in practice estimation of MA components is difficult and thus AR components are more common. Thus we see ADL models such as:

$$y_t = \alpha + \sum_{p=1}^{P} \phi_p y_{t-p} + \sum_{k=1}^{K} \Sigma_{n=1}^{N} \beta_{kn} x_{k,t-n} + \epsilon_t \tag{3}$$

  with the errors being white noise, P representing the number of autoregressive parameters, K representing the number of exogenous variables and N representing the number of exogenous variable lags.

Multipliers

Long-run multiplier for variable k

$$\frac{\sum_{n=0}^{N} \beta_{kn}}{(1 - \sum_{p=1}^{P} \phi_p)} \tag{4}$$

This simplifies to

$$\sum_{n=1}^{N} \beta_{kn} \tag{5}$$

if

$$\phi_p = 0, \ \forall \ p. \tag{6}$$

Multipliers

Impact multiplier and Interim multipliers for variable k

These remains the same as before,

Impact: β_{k0}

Interim: change up to m divided by total change, where $m < N$.

Specification

- Probably one of the least discussed issues in journals is that of specification.
- Specification issues in time series are very difficult to address.
  - because unlike with cross-sections, where we are modeling based on an $N \times N$ covariance matrix
  - in time series we have at least K $(T - p) \times (T - p)$ covariance matrices.
- Similar information, more parameters

Information Loss

Deep Thought

All specification involves a loss of information. The goal is, however, to specify models that are reparameterizations of the data generation process (DGP). The model itself may be wrong, but the parameters "left over" are sufficient to avoid invalidating inference on the part of the DGP we recover.
In short, wrong models will have parameters that are functions of the true model.

Information Loss

Deep Thought

All specification involves a loss of information. The goal is, however, to specify models that are reparameterizations of the data generation process (DGP). The model itself may be wrong, but the parameters "left over" are sufficient to avoid invalidating inference on the part of the DGP we recover.
In short, wrong models will have parameters that are functions of the true model.

Intuition

If you have estimate model $\hat{M}_1$ which is an attempt to approximate the DGP model Π, but that model ignores autoregressive terms, heteroskedasticity and endogeneity, what are we left with? What did we learn about Π from $\hat{M}_1$?

Impications

What happens when you have the wrong specification?
To illustrate this point, consider the following model:

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \alpha Y_{t-1} + u_t \tag{7}$$

with long run multiplier

$$\frac{\beta_0 + \beta_1}{1 - \alpha}$$

and

$$X_t = \lambda X_{t-1} + v_t$$

where

$$u_t \sim IN\left(0, \sigma_u^2\right),\; v_t \sim IN\left(0, \sigma_v^2\right),\; E\left[v_t, u_s\right] = 0,\; |\alpha| < 1,\; |\lambda| < 1$$

Implications II

Now assume that (7) is the DGP, the "true model" (Π) What happens if we haphazardly specify alternative models $\hat{M}_i$? Consider the restrictions and implications of the following.

1. Static Regression
2. ARIMA(1,0,0)
3. First difference model
4. Leading indicator model
5. Dead start model

1: Static regression

$$Y_t = \gamma X_t + u_{1t}$$

Restriction on (7): $\beta_1 = \alpha = 0$.

1. **Implications:**
   - Range of "γ":
     $$\beta_0 \leq \gamma \leq \frac{\beta_0 + \beta_1\lambda}{1 - \alpha\lambda}$$
     $\therefore$ γ is not a short or long run parameter. Interpretation is problematic. Range depends on size of β_0 and long run multiplier.
   - DGP is constant, so no dynamics in Y.

Problems:

- Serial correlation is quite likely, given DGP is dynamics and the estimated model is static.

1. Tests for parameter constancy as well as t-tests will be prone to falsely reject the null.
   $\therefore$ Proper "interpretation" of test information unlikely.

2: AR(1)

$$Y_t = \rho Y_{t-1} + u_{2t}$$

Restriction on (7): $\beta_0 = \beta_1 = 0$
Implications:

- Exclusion of X_t: Recall that $X_t = \lambda X_{t-1} + v_t, \ |\lambda| < 1$

$$X_t = \frac{v_t}{1 - \lambda B}$$

- Substitute and rearrange the DGP:

$$\begin{aligned}(1-\alpha B)\,Y_t &= \beta_0 \frac{v_t}{1-\lambda B} + \beta_1 \frac{v_{t-1}}{1-\lambda B} + u_t \\ &= \beta_0 \left[\frac{\left(1+\frac{\beta_1}{\beta_0}B\right)}{(1-\lambda B)}\right] v_t + u_t \quad \textit{if} \;\; \beta_0 \neq 0\end{aligned}$$

Now if $\frac{\beta_1}{\beta_0} = -\lambda$

$$Y_t = \rho Y_{t-1} + (u_t + \beta_0 v_t)$$

But ...

Continued

Problems:

1. σ_v^2 may not be constant (x_t changes).
2. If $\frac{\beta_1}{\beta_0} \neq -\lambda$, then the AR model will have autocorrelated error (v_{t-i}).
3. λ must be constant as well, else parameters in the model are not constant.

3: First differences model

$$\Delta Y_t = \phi \Delta X_t + u_{3t}$$

Restriction on (7): $\alpha = 1, \ \beta_0 = -\beta_1$.
Implications:

- Loss of long run relationship: K

$$\Rightarrow \quad \Delta Y_t = K \Delta X_t + u_{3t}$$
$$\textit{where } u_{3t} \textit{ is ``white noise''}$$

Put back into levels

$$Y_t = K X_t + u_t^*$$

where

$$u_t^* = u_{t-1}^* + \epsilon_t \Rightarrow \ \textit{random walk}$$

Problems

- Proportional relationship with a "permanent disequilibrium."
- Growth rate models with "white noise" residuals are internally inconsistent with long-run proportional relationships (errors in levels, equilbria are multiplicative, thus errors should be heteroskedastic).

4: Leading Indicator model

$$Y_t = CX_{t-1} + u_{4t}$$

Restrictions on (7): $\beta_0 = \alpha = 0$
Implications:

- No feedback, no constancy in prediction. If Y_t changes, $X_t' s$ effect is different (ΔC). Therefore, X_t is not a good predictor (specification).
- Goodness of fit will be poor.
- Lack of noise model will lead to likely serial correlation in the errors.
- These are everywhere, "Lagged one year to avoid endogeneity."

5: Dead start model

$$Y_t = b_1 X_{t-1} + b_2 Y_{t-1} + u_{5t}$$

Restrictions on (7): $\beta_0 = 0$

1. **Implications:**
   - $X_t = \lambda X_{t-1} + v_t \Rightarrow b_1 = \beta_0 + \beta_1 \lambda$. $\therefore$ to gauge usefulness of b_1 need to know λ.
   - "Lagged one year," again.

GLS

Serial correlation poses a threat to inference from using $\frac{\beta}{s.e.}$ "efficiency." GLS is applicable to situations where σ^2 is not constant or is in very special cases where errors are not independent across time (MA(1)).

Major Symptoms:

1. Heteroscedasticity
2. Serial Correlation

GLS II

- **Intuition:** We often must assume Y_t are independently drawn when using maximum likelihood and ordinary least squares.
- **Consequence:** Since independence is necessary for deriving Gauss-Markov for OLS, the estimates will have larger variance as a result, which we call inefficiency.
- **Correction:** We can correct the problem with GLS, or treat serial correlation as a more serious issue involving modeling the dynamics in the Y_t.

GLS III

Consider the model: $Y_t = \alpha + \beta X_t + \epsilon_t$.

- **Consistency and Unbiasedness:**

$$\hat{\beta} = \frac{\sum X_t (\beta X_t + \epsilon_t)}{\sum X_t^2} = \beta + \frac{\sum X_t \epsilon_t}{\sum X_t^2}$$

if $(X_t \epsilon_t) = 0$, then consistent and unbiased.

GLS IV

- **Efficiency:**

$$
\begin{aligned}
Var\left(\hat{\beta}\right) = E\left[\beta - \hat{\beta}\right]^2 &= E\left[\frac{\left(\sum X_t \epsilon_t\right)^2}{\left(\sum X_t^2\right)^2}\right] \\
&= \frac{\sum X_t^2 \epsilon_t^2}{\left(\sum X_t^2\right)^2} + \frac{\sum_t \sum_s X_t X_s \epsilon_t \epsilon_s}{\left(\sum X_t^2\right)^2} \\
&= \frac{\sum X_t^2 \sigma^2}{\left(\sum X_t^2\right)^2} + \frac{\sum_t \sum_s X_t X_s \sigma_{ts}}{\left(\sum X_t^2\right)^2} \\
&= \frac{\sigma^2}{\left(\sum X_t^2\right)^2} + \frac{\sum_t \sum_s X_t X_s \sigma_{ts}}{\left(\sum X_t^2\right)^2}
\end{aligned}
$$

If $\sigma_{ts} = 0$, then OLS variance is best.

DW Test

$$
\begin{aligned}
d &= \frac{\sum_t^n (\epsilon_t - \epsilon_{t-1})^2}{\sum_t^n \epsilon_t^2} \\
&= \left[\epsilon_t^2 - 2\epsilon_t\epsilon_{t-1} + \epsilon_{t-1}^2\right] * \sigma_\epsilon^{-2} \\
&= \left[2\hat{\sigma}_\epsilon^2 - 2Cov\left(\epsilon_t\epsilon_{t-1}\right)\right] * \sigma_\epsilon^{-2} \\
&= \left[2\hat{\sigma}_\epsilon^2 - 2\rho_\epsilon\right] * \sigma_\epsilon^{-2} \\
&= \frac{2\hat{\sigma}_\epsilon^2 (1-\rho)}{\sigma_\epsilon^2} \\
&= 2(1-\rho)
\end{aligned}
$$

Elementary My Dear Durbin-Watson

therefore

$$\begin{aligned} \rho = 1 &\quad\Rightarrow\quad d = 0 \\ \rho = -1 &\quad\Rightarrow\quad d = 4 \\ \rho = 0 &\quad\Rightarrow\quad d = 2 \end{aligned}$$

Consult a table for critical values in Durbin-Watson tables located in friendly econometrics texts. Not valid for equations with lagged left-hand-side variables.

Psuedo-GLS

1. Estimate ρ, the correlation coefficient for errors.
2. Multiply both sides of equation by $(1 - B\rho)$. This implies taking generalized differences of data so that X and Y equal $Y_t - \rho Y_{t-1}$ and $X_t - \rho X_{t-1}$.
3. Estimate using OLS. Be sure to compute all statistics except the coefficients and standard errors from the original data. Stata does this with *prais* command.

Psueo-GLS II

Some observations:

1. The above rountine assumes that ρ is known. A better procedure is the Hildreth-Lu method that uses a grid search to find the value of $\hat{\rho}$ that maximizes the likelihood. Similar routine in Stata under *prais* command with *ssesearch* option.
2. The GLS model is the same as having MA(1) noise component. This sometimes causes problems because computer programs and individuals sometimes refer to the dynamics as an AR process. It is an AR(1) for disturbances but not Y.
3. Compared to ARIMA models, the GLS model is a special case.
4. GLS is seen as a fix to a problem. Modeling the dynamics more directly is preferable, if every case where the underlying model is not an MA(1) noise model.

Serial Correlation with LDV

$$
\begin{aligned}
Y_t &= \beta Y_{t-1} + u_t \qquad (8)\\
\text{assume} : u_t &= \rho u_{t-1} + \epsilon_t
\end{aligned}
$$

Implications: Serial correlation results in violating the condition $E\left[Y_{t-1}, u_t\right] = 0$, and the result is an *inconsistent* estimate of β.

Proof

Lag equation (8) one period and multiply by ρ (autocorrelation coefficient).

$$\rho Y_{t-1} = \rho\beta Y_{t-2} + \rho u_{t-1}$$

Solving for ρu_{t-1}:

$$\rho u_{t-1} = \rho Y_{t-1} - \rho\beta Y_{t-2} \tag{9}$$

Now rewrite equation (8) using equation (9):

$$\begin{aligned} Y_t &= \beta Y_{t-1} + \rho u_{t-1} + \epsilon_t \\ &= \beta Y_{t-1} + \rho Y_{t-1} - \rho\beta Y_{t-2} + \epsilon_t \\ & \quad (\beta + \rho)\, Y_{t-1} - \rho\beta Y_{t-2} + \epsilon_t \end{aligned} \tag{10}$$

Proof II

From (10) the “true” coefficient for Y_{t-1} is $(\beta + \rho)$, and the “true” coefficient for the omitted variable , Y_{t-2} is -$\rho\beta$.
Thus, the actual estimate, $\hat{\beta}$,does not converge asymptotically to the actual value, β.
Normally, a parameter estimate will equal the “true” parameter as the sample size goes to infinity. This is true, for example, for MLE when distributional assumptions are met.

GLS again

To find the asymptotic values of $\hat{\beta}$, multiply equation (10) by Y_{t-1}:

$$\sum Y_t Y_{t-1} = (\beta + \rho) \sum Y_{t-1}^2 - \beta\rho \sum Y_{t-1} Y_{t-2} + \sum u_t Y_{t-1}$$

Thus, dividing through by $\sum Y_{t-1}^2$,

$$\hat{\beta} = (\beta + \rho) - \beta\rho \frac{\sum Y_{t-1} Y_{t-2}}{\sum Y_{t-1}^2} + \frac{\sum u_t Y_{t-1}}{\sum Y_{t-1}^2}$$

and taking plims:

$$\begin{aligned} plim(\hat{\beta}) &= \beta + plim(\sum y_{t-1}^2)^{-1} plim(\sum u_t Y_{t-1}) \\ &= \beta + \frac{\rho(1-\beta^2)}{1-\rho\beta} \end{aligned}$$

So long as $\rho = 0$, $\hat{\beta} \to \beta$, so that OLS is consistent. Convergence if always confounded by ρ.

Not so Elementary,

In addition, detection for first-order serial correlation cannot be achieved by conventional D-W tests. This statistic is biased by

$$\frac{2\rho\left(1-\beta^{2}\right)}{\left(1-\rho\beta\right)}$$

If $\rho > 0$, the D-W statistic is biased upward–and may incorrectly indicate no serial correlation when it in fact exists.

The Solution: M-test

Use M-test when you have a model with a lagged endogenous variable. The M-test is

- Estimate OLS for the original model: $Y_t = \beta M_t + u_t$ where M includes all variables.
- Save the residuals, u_t.
- Regress contemporaneous residual u_t on original model and lags of itself:

$$u_t = \sum_{i=1}^{m} \gamma_i u_{t-i} + \beta M_t + v_t$$

Hypotheses for M-test

- Formulate hypotheses:

$$
\begin{aligned}
H_0 &: \textit{All } \gamma_i = 0 \ \text{``\textit{no serial correlation}''} \\
H_A &: \textit{At least one } \gamma_i \neq 0 \ \text{``\textit{serial correlation}''}
\end{aligned}
$$

- Test via t-distribution or F-distribution, depending on the value of m.

What to do now, Soapbox

1. Use theory to get your specification. If data is unavailable to test your theory (something is missing), understand how any parameterization of the partial process you do have data on will allow you to make inference on the DGP you are interested in.
2. Estimate (OLS, GLS, other)
3. Check to see that the statistical assumptions underlying your estimates are reasonable. This will include a DW if no LDV is present, and will include an M-test if a LDV is included.
4. If statistical assumption are not met, start again. Going back over how this new information changes your ability to make inference back to the parameters you are interested in.

Weak Stationarity Review

A series is weakly stationary if the expected value of the mean and the variance/covariance is constant for all time periods.

Random Walk: $y_t = 1.0y_{t-1} + \epsilon_t$

This means that $y_{t=i} = y_{t=0} + \Sigma_{i=0}^{i}\epsilon_i$

Shocks accumulate overtime.

Random Walk with drift: $y_t = \alpha + 1.0y_{t-1} + \epsilon_t$

This means that $y_{t=i} = y_{t=0} + \alpha t + \Sigma_{i=0}^{i}\epsilon_t$

Differencing

Differencing gets us back to stationarity.

Random Walk:

$\Delta y_t = (1.0 y_{t-1} + \epsilon_t) - y_{t-1}$

gives us $\Delta y_t = \epsilon_t$

Random Walk with drift:

$\Delta y_t = \alpha + 1.0 y_{t-1} + \epsilon_t - y_{t-1}$

gives us $\Delta y_t = \alpha + \epsilon_t$

note: drift is a constant in the change.

Unit Roots

Most discussions of stationarity center on unit roots (Case 2 below). Consider the model,

$$Y_t = \phi Y_{t-1} + \epsilon_t \tag{11}$$

1. **Case 1:** $|\phi| < 1$.
   If $|\phi| < 1$, then $\rho_j = \phi^j$ (ACF declines exponentially).
2. **Case 2:** $|\phi| = 1$. If $|\phi| = 1$, then $\rho_j = \phi^j$ (ACF does not dampen as $Cov\left(Y_t, Y_{t-j}\right) \rightarrow \infty$).
   Asymptotic properties:

   $$Estimate\ of\ mean \Rightarrow\ Cauchy\ distribution\ (no\ mean)$$

   Proof: Spanos (1986: 70). And,

   $$variance \Rightarrow \infty$$

Implication: No t-distribution unless appropriate adjustments are made. In case 2, it is not useful for hypothesis testing when $Y_t \sim I\left(1\right)$.

Spurious Regression

Therefore, by ignoring unit roots, but assuming errors are independent, the following regression of $y_t \sim I(1)$ and $x_t \sim I(1)$
$y_t = \alpha + \beta x_t + \epsilon_t$
will produce a statistically significant estimate of β a large percentage of the time, even when the two variables are unrelated.
Problem increases as T increases and drift added.

T	Random Walk	Drift
25	0.53	0.645
100	0.76	0.945
500	0.89	1.00
1000	0.947	1.00

Problems for Multivariate Series

- Ignore unit roots: Spurious regression
- Difference($\Delta y_t = \alpha + \beta \Delta x_t + \epsilon_t$): Lose long-run information, such that we are assuming
  $y_t = y_{t-1} + \beta_1 x_t - \beta_2 x_{t-1} + \epsilon,$ *where* $\beta_1 = \beta_2 = \beta$
  What if $\beta_1 \neq \beta_2$
  Can we do better?

Cointegration

From Granger (1983) we can now induce "stationarity" via linear transformation. Such transformations do not exclude long-run relationships and they have stronger theoretical content.

Granger's Representation Theorem: Suppose Z_t and Y_t are $I(1)$. If Z_t and Y_t are "cointegrated" then their linear combination $u_t = Y_t - \alpha Z_t \Rightarrow I(0)$, where $u_t \sim I(0)$.

Cointegration Implications

1. Linear combination is stationary.
2. At least one variable Granger-causes the other (later).
3. There is a error-correction representation present and it is stationary.

Together Again: Cointegration's Theme Song

Recall that a series is said to be $I(d)$ is it has to be differences d times in order to be a stationary process.

A common correction procedure: differencing and the error correction model. Consider the following model:

$$\underset{I(1)}{Y_t} = \beta_1 \underset{I(1)}{Z_t} + \beta_2 \underset{I(1)}{Y_{t-1}} + \beta_3 \underset{I(1)}{Z_{t-1}} + \underset{I(1)}{\epsilon_t} \tag{12}$$

where

$$\frac{\beta_1 + \beta_3}{1 - \beta_2}$$

is the total multiplier.

Now if we difference all variables:

$$\underset{I(0)}{(1-B)\,Y_t} = \beta_1^* \underset{I(0)}{(1-B)\,Z_t} + \beta_2^* \underset{I(0)}{(1-B)\,Y_{t-1}} + \beta_3^* \underset{I(0)}{(1-B)\,Z_{t-1}} + \underset{I(0)}{u_t} \tag{13}$$

$$\underset{I(0)}{\Delta Y_t} = \beta_1^* \underset{I(0)}{\Delta Z_t} + \beta_2^* \underset{I(0)}{\Delta Y_{t-1}} + \beta_3^* \underset{I(0)}{\Delta Z_{t-1}} + \underset{I(0)}{u_t} \tag{14}$$

Together Still

So as an alternative to (12) and (13), (14) can be made stationary by constructing an ECM as follows:
Assume that Y_t and Z_t are $I(1)$.

$$\begin{aligned}
\underset{I(0)}{(1-B)\,Y_t} &= \beta_1 \underset{I(1)}{Z_t} + (\beta_2 - 1)\underset{I(1)}{Y_{t-1}} + \beta_3 \underset{I(1)}{Z_{t-1}} + v_t \\
\underset{I(0)}{(1-B)\,Y_t} &= \beta_1 \underset{I(0)}{(1-B)\,Z_t} + (\beta_2 - 1)\underset{I(1)}{Y_{t-1}} + (\beta_1 + \beta_3)\underset{I(1)}{Z_{t-1}} + d_t \\
\underset{I(0)}{\Delta Y_t} &= \beta_1 \underset{I(0)}{\Delta Z_t} + (\beta_2 - 1)\underset{I(0)}{\left[Y_{t-1} + \left(\frac{\beta_1 + \beta_3}{\beta_2 - 1}\right) Z_{t-1}\right]} + g_t \\
\Delta Y_t &= \beta_1 \underset{I(0)}{\Delta Z_t} + (\beta_2 - 1)\, u_{t-1} + g_t
\end{aligned}$$

where:

$$u_{t-1} = \left[Y_{t-1} + \left(\frac{\beta_1 + \beta_3}{\beta_2 - 1}\right) Z_{t-1}\right]$$

$\therefore$ Retains long-run component.

Some comments:

1. If Y and Z are not cointegrated, then $\beta_2 - 1$ does not follow a t-distribution.
2. Cointegration provides balance to the equation.
3. There are two methods of estimation of the error correction model. One is a two step method that estimates the long-run constant first and then estimates the ECM. The one step method includes Y_{t-1} and Z_{t-1} in the equation, and once estimated, the long run component can be solved for analytically. There is some evidence that the latter method works better in small samples.

Unit Root Testing

Unit root testing centers on the following:

$$Y_t = \phi Y_{t-1} + \epsilon_t \tag{15}$$

Now difference (15):

$$\Delta Y_t = \Pi Y_{t-1} + \epsilon_t \ where \ \Pi = \phi - 1 \tag{16}$$

Hypotheses:

$$\begin{aligned} H_0 &: \Pi = 0 \Rightarrow unit\ root\ exists \\ H_A &: \Pi < 0 \Rightarrow trend\ stationary \end{aligned}$$

Dickey-Fuller Test (D-F)

Dickey-Fuller tests follow the model in (16). The DF distribution depends on the existence of "nuisance parameters" in the model:

1. No intercept, no trend.
2. Intercept, no trend.
3. Intercept, trend.

D-F: no intercept, no trend

$$\Delta Y_t = \Pi Y_{t-1} + \epsilon_t$$

D-F: intercept, no trend

$$\Delta Y_t = \alpha + \Pi Y_{t-1} + \epsilon_t$$

D-F: intercept, trend

$$\Delta Y_t = \alpha + \Pi Y_{t-1} + \delta t + \epsilon_t$$

The three models have alternative critical values.

Augmented D-F test

In addition, the model for the D-F tests can give inaccurate results when there is serial correlation.
So the D-F test has been modified to:

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{n} \rho_i \Delta Y_{t-i} + v_t$$

This is the *augmented* Dickey-Fuller (AD-F) representation. The lagged endogenous series is inserted to account for serial correlation.
The AD-F is still sensitive to nuisance parameters.

Testing for cointegration

The most prominent test for cointegration is the Engle-Granger (1987) two-step procedure. It makes use of the Dickey-Fuller tests.

Step 1: Ascertain using D-F procedures (AD-F is recommended here) how many series are $I(1)$. Then create a residual running a regression on the variable of interest:

$$\underset{I(1)}{Y_t} - \hat{\alpha}_1 - \beta_1 \underset{I(1)}{X_t} - \beta_2 \underset{I(1)}{Z_t} - \beta_3 \underset{I(1)}{P_t} = \underset{I(0)?}{u_t}$$

Step 2: Run an AD-F on u_t :

$$\Delta u_t = \Pi u_{t-1} + \sum_{i=1}^{n} \rho_i \Delta u_{t-i} + v_t$$

where the hypotheses are

$$\begin{aligned} H_0 &: \quad \Pi = 0 \;\Rightarrow\; u_t \sim I(1) \quad \text{“}\textit{not cointegrated}\text{”} \\ H_A &: \quad \Pi < 0 \;\Rightarrow\; u_t \sim I(0) \quad \text{“}\textit{cointegrated}\text{”} \quad \therefore \textit{can use ECM} \end{aligned}$$

Engle and Granger find the critical value of the AD-F for cointegration tests is -3.17 for $p = 0.05$.

Other Unit Root Tests

- KPSS Test (Kwiatkowski, Phillips Schmidt and Shin 1992)- Null is stationarity, as opposed to DF or ADF. Need evidence to reject stationarity
- Variance ratio test (Diebold 1989)- Null of a random walk with drift.

Many times these tests will give different answers. Remember they are testing different things, require different assumptions, and may have low power in finite samples.