Cointegration and common trends

- Chapter 17 of Asteriou and Hall
- When we difference a time series variable, we lose (sometimes a lot of) information.
- We can no longer determine the level of the variable unless we have the first undifferenced observation at least:
- The variable could be $x_t = 10000$ and $x_{t+1} = 10010$, or it could be $x_t = 2$ and $x_{t+1} = 12$.
- When we difference we focus on the **short run** component of the (stochastic) process generating the data, and we lose information on the long run movement (i.e. the trend - whether deterministic or stochastic).

Dynamic models

- Another way to say this is that, when we difference we can no longer obtain a unique long-run solution.
- E.g., suppose $y_t = 0.5x_{t-1} + e_t \;\; e_t \sim N(0, \sigma^2)$, with σ^2 relatively small, or even zero so that y_t is a deterministic process.
- Given a value for x_t, we can predict y_t accurately if σ^2 is relatively small (and perfectly if $\sigma^2 = 0$).
- If $x_t = 2$, then $y_{t+1} \approx 1$. If, on the other hand, we difference, so we have

  $$\triangle y_t = 0.5\triangle x_{t-1} + u_t.$$

  then we can never obtain y_{t+1} without more information (namely, some way to obtain y_t, since $y_{t+1} = y_t + \triangle y_{t+1}$).

Dynamic regression model

- We would like to be able to build dynamic models that incorporate both the short-run **and** long-run properties of the process, and at the same time satisfy the requirement that all the variables in the model be stationary.
- If two variables, say y and x, are related in the long run, then we would expect them to 'move together' or 'in tandem', at least approximately and **in the long run.**
- That is, we think they have a **common trend** (= long run comovement). Note that this may be a **stochastic** trend (they may both be random walk type processes - 'integrated of order one').
- In this case, it is usually possible to find a linear combination of the variables (often the difference between them at a point in time) that is stationary. This captures the idea that they never 'get too far apart'.

The drunk and her dog

- This example is due to Murray (1994) A Drunk and Her Dog: An Illustration of Cointegration and Error Correction, *American Statistician, 48,1.*
- Imagine a drunk is staggering around town randomly - i.e. in an unpredictable manner. If you don't know where the drunk was within the last few time periods (say minutes), you may have little to no idea about her location.
- Now suppose you know that the drunk has a dog, and you also know that the dog tends to stay reasonably close to its owner (so as not to get locked out at night perhaps).
- If you are looking for the drunk and you see her dog, you know that (with pretty high probability) the drunk is nearby, and further, the dog will likely lead you to the drunk in the near future.
- You have learned a lot about the location of the drunk from observing the dog.

A time series = trend + variations around trend

- If a time series variable, X_t is nonstationary, then we can decompose it into a trend component, and short run variations around the trend.

  $$X_t = Trend_t + x_t$$

  where $Trend_t$ may be stochastic, and x_t is stationary.
- The main idea: if we can find another variable, say Z_t that is **cointegrated** with X_t, then we can model the long run trend component using this other variable (or variables).
- If two (or more) variables are cointegrated, then they "move together" in the long run, or "tend not to diverge" or they have "a common trend."
- Note that cointegration is a **long run** concept about **non**stationary variables.

Definition of cointegration

- Two (or more) variables are cointegrated if

1. They are nonstationary and integrated of the same order (usually all $I(1)$), and
2. There exists a linear combination of the variables that is stationary.

- Example, suppose X_t and Z_t are both found to be $I(1)$. If we can find a linear combination, call it Y_t,

$$Y_t = \alpha_1 X_t + \alpha_2 Z_t = X_t + \alpha Z_t$$

such that $Y_t \sim I(0)$. Then X_t and Z_t are cointegrated.

The Engle-Granger cointegration methodology

- First determine the order of integration of all the variables.
- To be cointegrated the variables must be integrated of the **same order** and **not stationary**.
- OLS minimizes the sum of squared residuals, so it finds the smallest linear combination (with the dependent variable) of the variables in the regression model.
- So if we estimate a linear regression using the potentially cointegrated variables, the residuals from this regression will be an estimate of the smallest (on average) set of values possible in a linear combination.
- If there is a stationary combination of variables, these will be the smallest values (since a nonstationary variable does not stay around zero), so OLS will find it!

Engle-Granger (cont.)

- Suppose X_t and Y_t are both found to be $I(1)$. Then an approximate, long run linear combination can be written as

$$Z_t = Y_t + \alpha_0 + \alpha X_t$$

or, rearranging,

$$Y_t = \alpha_0 + \alpha X_t + Z_t$$

- So if we estimate a linear regression in **levels** form (i.e. using the **non**stationary variables), the estimated residual will be an estimate of the cointegrating relationship (if it exists).
- Then we determine if the residuals from this regression are stationary. If so, that is evidence that the variables are cointegrated.
- The choice for dependent variable is **arbitrary** (the normalization in the linear combination).

Engle-Granger (cont.)

- **Procedure**
- (1) Determine which (if any) variables satisfy the first condition (integrated of the same order),
- (2) Estimate a linear regression in **levels** form (i.e. using the **non**stationary variables, with **no lags**) for those that satisfy the condition.
- (3) Test whether the residuals from this regression are stationary or not.
- However, we only have an **estimate** of the cointegrating relationship, we don't observe the actual cointegrating relationship.
- So we can use the ADF test, but **the critical value is different** (when testing for cointegration), critical-t $\simeq -2.9$.

Error correction models

- If the variables are found to be cointegrated, then we modify the dynamic regression model by adding an error correction term.
- The error correction term is **one** lag of the residual from the cointegrating regression (i.e. the one we tested to see if it was stationary).
- It is important to note that there are **no** lags included in the cointegrating equation (regression), and we include only **one** lag of the error correction term in the dynamic regression model.
- We still include as many lags as appears appropriate for the other (stationary transformed) variables in the model (including lags of the dependent variable).

Example ECM

- For example, in the dynamic regression model,

$$\triangle y_t = \alpha + \sum_{i=1}^{p} \phi_i \triangle y_{t-i} + \sum_{i=0}^{q} \beta_i \triangle x_{t-i} + e_t\,,$$

- Suppose we found that y_t and x_t are both I(1) and cointegrated. The model then becomes an ECM,

$$\triangle y_t = \alpha + \sum_{i=1}^{p} \phi_i \triangle y_{t-i} + \sum_{i=0}^{q} \beta_i \triangle x_{t-i} + \gamma ecm_{t-1} + e_t\,,$$

  where

$$ecm_{t-1} = y_{t-1} - \hat{a} - \hat{b} x_{t-1}\,,$$

  i.e., one lag of the residual from a regression of y on x (or x on y).

Another example

- Example 2, in the dynamic regression model,

$$\triangle y_t = \alpha + \sum_{i=1}^{p} \phi_i \triangle y_{t-i} + \sum_{i=0}^{q} \beta_{1i} \triangle x_{1t-i} + \sum_{i=0}^{s} \beta_{2i} x_{2t-i} + e_t,$$

- Suppose we found that y_t and x_{1t} are both I(1) and cointegrated, and x_{2t} is I(0). The model then becomes an ECM,

$$\triangle y_t = \alpha + \sum_{i=1}^{p} \phi_i \triangle y_{t-i} + \sum_{i=0}^{q} \beta_{1i} \triangle x_{1t-i} + \ldots + \sum_{i=0}^{s} \beta_{ki} x_{kt-i} + \gamma ecm_{t-1} + e_t,$$

where $ecm_{t-1} = y_{t-1} - \hat{a} - \hat{b} x_{t-1}$, one lag of the residual from a regression of y on x (or x on y).

- What if the dependent variable, y, is I(0)?

ECMs

- The error correction component captures the long run comovement in the variables.
- The larger the ‘distance’ between y_{t-1} and x_{t-1}, the bigger the adjustment back towards each other or “equilibrium”.
- An ECM can be viewed as a model of short run disequilibrium, the market adjusting towards equilibrium each period (if other things, in the model, remain unchanged).
- Just because two variables are cointegrated, that doesn’t mean that the error correction component will be significant in the dynamic equation.
- The dependent variable may be a ‘follower’ rather than a leader. E.g., one company may be a price setter, and other companies may adjust their price to match whenever the leader(s) changes the price . Also, since we are estimating the coefficient on an estimate (the residual), there is reduced precision and so we are more likely to observe statistically insignificant results.

ADF test for cointegration

- Suppose we have two variables, y_t and x_t, and both are I(1).
- First we estimate,

$$y_t = \alpha + \beta x_t + e_t$$

- Notice that there are **no lags** in this equation, and **both the variables are undifferenced**.
- Estimate the ADF equation using the estimated residuals, $\hat{e}_t$,

$$\triangle \hat{e}_t = \alpha + \phi \hat{e}_{t-1} + \sum_{i=1}^{p} \beta_i \triangle \hat{e}_{t-i} + u_t.$$

- Perform a t-test for $H_0 : \phi = 0$ vs. $H_0 : \phi < 0$ using the critical value -2.9.

Decision rule

- If $t \leq -2.9$ then there is evidence that $\hat{e}_t$ is stationary, so the evidence indicates that y_t and x_{1t} **are** cointegrated.
- If $t > -2.9$ then there is evidence that $\hat{e}_t$ is **not** stationary, so the evidence indicates that y_t and x_{1t} are **not** cointegrated.
- As with the ADF test for stationarity, the test statistic, t, must be negative **and** statistically significant.
- Remember: If it is **not** significant, then the variables are **not** cointegrated.

Examples using the Engle-Granger approach

- Generate some data where there are two or three variables, at least two are nonstationary and (a) at least two are cointegrated and (b) none are cointegrated.
- Test for cointegration and build an ECM, if appropriate for the variables in R.
- see **coint.R**
- Consider the **aggmacro.dat** data as an example using real data.

Drawbacks of the Engle-Granger approach

- In theory (asymptotically!), which variable is selected as the dependent variable in the regression equation to estimate the cointegration relationship (the one in levels with no lags) should not matter. In practice it often does matter, since we rarely have large enough samples to be able to safely presume asymptotic results will hold.
- This is especially a problem with time series data because the data are **dependent** across time, so there is **less information in a new observation** than in an independent cross section.
- This problem becomes more complicated as the number of potentially cointegrated variables increases.

Drawbacks of the Engle-Granger approach

- When there are more than two variables, we have to consider the possibility that only a subset of the variables belong in a cointegrating relationship (e.g. all the pairs of variables), and there also may be more than one cointegrating relationship between them. The Engle-Granger cannot identify all of these possibilities.
- The Engle-Granger approach employs "two-step estimator" - the residuals are estimated first, then they are used in the second stage to determine if the variables are cointegrated and again in the error correction component. Any error present in the first step is carried over into the second stage.
- An alternative to the Engle-Granger approach was developed by Johansen (*JEDC*, 1988, and *Econometrics*, 1991).
- The approach is essentially to extend the ADF testing model to a vector of variables instead of just one variable.

Johansen approach to cointegration

- Suppose we have 2 variables, which we stack in a matrix $Z_t = [X_t, Y_t]$.
- The model 1st order (vector) AR model (VAR), is

$$Z_t = \Phi Z_{t-1} + e_t$$

- This can be written as

$$\triangle Z_t = (\Phi - I)Z_{t-1} + e_t$$

  which is a vector version of the Dickey-Fuller unit root testing equation.
- If we test the characteristic roots of this equation, which is determined by the eigenvalues of the matrix $\Pi = (\Phi - I)$, we can determine whether there exists a linear combination of the variables that is stationary .

Vector error correction model

- Suppose we have 3 variables, which we stack in a matrix $Z_t = [X_{1t}, X_{2t}, X_{3t}]$.

$$Z_t = \mu + A_1 Z_{t-1} + \ldots + A_k Z_{t-k} + \varepsilon_t, \quad (t = 1, \ldots, T)$$

  is again comparable to the single equation dynamic regression model, except now we have 3 equations, one for each variable.
- This can be reformulated as a **vector** error correction model, **VECM**:

$$\Delta Z_t = \mu + \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_{k-1} \Delta Z_{t-k-1} + \Pi Z_{t-1} + \varepsilon_t$$

where

Parameters of interest for cointegration

$$\Gamma_i = -(I - A_1 - \ldots - A_i), \quad (i = 1, \ldots, k-1),$$

and

$$\Pi = -(I - \Pi A_1 - \ldots - A_k)$$

- The Γ_i matrices contain the short-run or transitory effects, and the Π matrix contains the long-run impacts - the error correction component.
- We can decompose $\Pi = \alpha\beta'$, where αinclude the speed of adjustment to equilibrium coefficients (the error correction) and the β' will be the long-run (cointegrating relationship) coefficients.
- That is, $\beta' Z_{t-1}$ is the multivariate equivalent to the error correction term, $\hat{e}_{t-1}$ in the single equation case. However, it contains up to $n-1$ linear combinations of the variables in Z.
- See p. 369 of Asteriou and Hall.

Johansen approach - when cointegration occurs

- Given that Z_t is a vector of nonstationary I(1) variables, then $\triangle Z_{t-i}$ are I(0), and we must have

$$\Pi Z_{t-1} \sim I(0)$$

  for the error term, ε_t to be stationary.
- Three cases when $\Pi Z_{t-1} \sim I(0)$:
- (1) All the variables in Z_t are already stationary.
- (2) There is no cointegration, so there is no linear combination of the variables that is I(0), and $\Pi = 0$.
- (3) When there is up to $n - 1$ cointegrating relationships of the form $\beta' Z_t \sim I(0)$. In this case there will be $r \leq n - 1$ cointegrating vectors in β.

Rank of Π

- Since $\Pi = \alpha\beta'$, if there are $r \leq n - 1$ cointegrating relationships, then α will be $n \times r$ and β' will be $r \times n$.
- The Π matrix is made up of a set of only r cointegrating vectors.
- Therefore, the rank of Π will be r (since it will have only r linearly independent rows).

Using the rank of Π for cointegration testing

- So if we can determine $\text{rank}(\Pi)$, we can determine how many cointegrating relationships exist between the variables in Z.
- (1) When all the variables are stationary, $\text{rank}(\Pi) = n$.
- (2) When there are no cointegrating relationships, $\Pi = 0$. so $\text{rank}(\Pi) = 0$.
- (3) When there are $r \leq n - 1$ cointegrating relationships, $\text{rank}(\Pi) = r$.

Johansen methodology

- Johansen (1988) developed two tests for the rank of Π and estimators for α and β.
- This is a rather complicated reduced rank maximum likelihood procedure, so we will skip the details.
- The idea is the same as the ADF test however: to estimate the eigenvalues (characteristic roots) of the dynamic model (essentially the VECM) and examine if these are larger than one or not (remember, roots that lie outside the unit circle => stationarity).
- The Johansen tests calculate maximum likelihood estimates of these eigenvalues, λ and test if $(1 - \lambda)$ is less than zero or not.

Johansen procedure in practice

- Determine the order of integration of the variables - they must all be nonstationary and integrated of the same order.
- Determine the appropriate lag length - just as with the ADF test, we must find the best choice for number of lags of all the variables to include to increase the power of the tests. This is usually accomplished by estimating a VAR in levels.
- You may need to include seasonal dummy variables or other exogenous variables, but you typically should **not** include a deterministic time trend unless you are convinced it makes sense. (The "intercept no trend" specification is usually the one to use).
- Calculate the max eigenvalue and trace statistics and determine if statistically significant using the critical values.

The test statistics

- The "lambda max" statistic

$$\hat{\lambda}_{max}(r, r+1) = -n \ln(1 - \hat{\lambda}_{r+1})$$

  tests the null hypothesis that $\text{rank}(\Pi) = r$, vs. the alternative that $\text{rank}(\Pi) = r + 1$.

- the "lambda trace" statistic

$$\hat{\lambda}_{trace}(r) = -n \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_{r+1})$$

  test the null hypothesis that $\text{rank}(\Pi) \leq r$, vs. the alternative that $\text{rank}(\Pi) > r$.

Decision rule

- The asymptotic distributions of these tests is not the usual chi-square distribution and depends on the number of lags and cointegrating relationships present. Critical values are generally determined by simulation methods.
- When there is no cointegration present, both statistics will be approximately zero (since all eigenvalues will be zero). The larger the statistics, the more evidence of cointegration.
- Most of the time we really only care whether or not there is cointegration, our model will be essentially the same regardless of exactly how many cointegrating relationships are present, as long as there is at least one.

Johansen v. Engle-Granger

- The traditional Engle-Granger two-step procedure and the asymptotic distribution of the "ADF test statistic" follows the distribution described by Phillips and Ouliaris (see Hamilton's Time Series Analysis book for a description of this distribution).
- Under the null of no-cointegration (unit root) this distribution incorporates the fact that the estimated cointegrating relationship is spurious and is similar to the Johansen distribution.
- In this case, the Engle-Granger and Johansen procedures may give different results if the OLS cointegrating vector is quite a bit different than the cointegrating vector estimated by the Johansen MLE.

Warning: low power area

- The Johansen MLE can sometimes give very strange results.
- One reason for this is that the finite sample distn of the Johansen MLE does not have any moments (as proved by Phillips in a Journal of Econometrics article) and so the tails of the finite sample distribution are very fat which can produce extreme values of the cointegrating vector.
- This is similar to the situation with the LIML estimator in the traditional simultaneous equation model (the 2SLS estimator has moments in overidentified models; the LIML estimator does not).

The package urca, and the ca.jo function in R

- There is an R package called urca, which stands for something like "unit root and cointegration analysis."
- Suppose we have 3 variables, which we stack in a matrix $Z_t = [X_{1t}, X_{2t}, X_{3t}]$. Load the package urca,

```
library(urca)
# To run the max eigenvalue test
ca.jo(Z, type="eigen", ecdet="const")
# To run the trace test
ca.jo(Z, type="trace", ecdet="const")
```

- There is an adf test function in this package also:
- ur.df(y, type = c("none", "drift", "trend"), lags = 1, selectlags = c("Fixed", "AIC", "BIC"))