Lecture 6: Non-stationary models: modeling trends and cycles - Part II

Dr. Joao B. Duarte[1]

[1]Nova School of Business and Economics
University of Cambridge

Masters, Economics: Macroeconometrics
Lisbon
Spring 2017

Lecture Objectives:

- Spurious regression
- Introduce the concept of cointegration
- Show that when a set of variables are cointegrated we can analyse short-run and long-term dynamics: we can have a long-run representation and error correction form.
- Introduce VECM
- Describe two tests for cointegration: the Engle-Granger test and the Johansen test.

Secondary Readings:

- Chapter 6, Applied Econometric Time Series, Enders, Walter, Fourth Edition
- Chapters 19 and 20, Time Series Analysis, Hamilton, James, first edition

Spurious Regression

- As we know by now, if some or all variables are non-stationary in a regression, the usual classical statistical results are usually no longer valid (unless they are cointegreated).
- One particular instance of such a case is called Spurious Regression.
- This case is worrisome as using classical statistics in this case suggest relation in variables that in reality do not exist!

Spurious Regression

- Granger and Newbold (1974) called the attention of the profession with a canonical example of such case. Suppose two series are by construction are I(1) and independent of each other:

$$y_t = y_{t-1} + \varepsilon_{yt}$$
$$z_t = z_{t-1} + \varepsilon_{zt}$$

If we regress y_t on z_t it turns out we find:

$$\begin{aligned} y_t &= \underset{(0.39)}{6.74} + \underset{(0.05)}{0.40} z_t, \quad R^2 = 0.21 \end{aligned}$$

- Hence, using the usual t-test that these variables are linearly related.

Spurious Regression

- We expect the coefficient to be zero. And indeed if we do a regression for the stationary first difference of the series we find:

$$\begin{aligned} \Delta y_t &= \underset{(0.07)}{-0.06} + \underset{(0.06)}{0.03}\Delta z_t, \quad R^2 = 0.00 \end{aligned}$$

- Hence, it gives the expected results.
- In the case of a spurious regression, the OLS estimate is not consistent and the variance goes to infinity as the sample size increases.
- Regression with non-stationary data only is valid when the series are cointegrated.

Intro to Cointegration

- Consider a case where a set of economic variables are in a long-run equilibrium:

$$\beta_1 x_{1t} + \beta_2 x_{2t} + ... + \beta_n x_{nt} = 0 \tag{1}$$

- Let x_t represent the vector of variables and β the vector representing the coefficients in (1). We can define deviations from this equilibrium by e_t"

$$e_t = \beta x_t \tag{2}$$

- For (1) to make sense as an equilibrium, e_t must necessarily be stationary.

Intro to Cointegration

- A set of non-stationary variables with the same order of integration say $I(d)$ are said to be cointegreated when a linear combination of them exists , $(\beta \neq 0)$, and is stationary.
- Most of the cointegration literature focus on cases where variables are I(1). The reason being that most economic variables are I(1).
- There can be multicointegration. Example: it is possible that a subset of variables is I(2) but that a linear combination of them is I(1). Hence, we can use this linear combination together with the other variables that are I(1) and have cointegration.
- The cointegration vector β is unique up to scalar. We need to use a normalization.

Normalization

- If β represents a cointegration vector, so does:

$$c\beta x_t = \beta^* x_t \sim I(0)$$

- Where c is a constant. We need some normalization. A typical normalization used is

$$\beta = (1, -\beta_2, ..., -\beta_n)$$

- So that

$$\beta x_t = x_{1t} - \beta_2 x_{2t} - ... - \beta_n x_{nt} \sim I(0) \quad (3)$$

or

$$x_{1t} = \beta_2 x_{2t} + ... + \beta_n x_{nt} + \varepsilon_t \quad (4)$$

Multiple Cointegration Vectors

- Moreover, we know from linear algebra that there can be at most $n-1$ linear independent vectors that span the entire linear cointegration space.
- Example: If there there are 3 variables, $n = 3$ there can be at most $r = 2$ linear independent cointegrating vectors.
- These linear independent vectors are the basis for the space of cointegrating vectors
- If we have 3 linearly independent cointegrating vectors, then we have recovered the entire space and any linear combination of the variables is stationary which can only occur if and only if all variables are stationary (hence they cannot be I(1) in the first place).

Common Trend Interpretation

- Cointegrated series share a common stochastic trend. To see why, lets see an example with two non-stationary variables that are decomposed into their stochastic trend (random walk) plus a stationary irregular component:

$$\begin{aligned} y_t &= \mu_{yt} + e_{yt} && (5) \\ z_t &= \mu_{zt} + e_{zt} && (6) \end{aligned}$$

And the linear combination of them is:

$$\begin{aligned} \beta_1 y_t + \beta_2 z_t &= \beta_1(\mu_{yt} + e_{yt}) + \beta_2(\mu_{zt} + e_{zt}) \\ &= (\beta_1 \mu_{yt} + \beta_2 \mu_{zt}) + (\beta_1 e_{yt} + \beta_2 e_{zt}) \end{aligned}$$

Common Trend Interpretation

- Given that e_t are both stationary, for y_t and z_t to be cointegrated it must be the case that

$$(\beta_1 \mu_{yt} + \beta_2 \mu_{zt}) = 0 \tag{7}$$

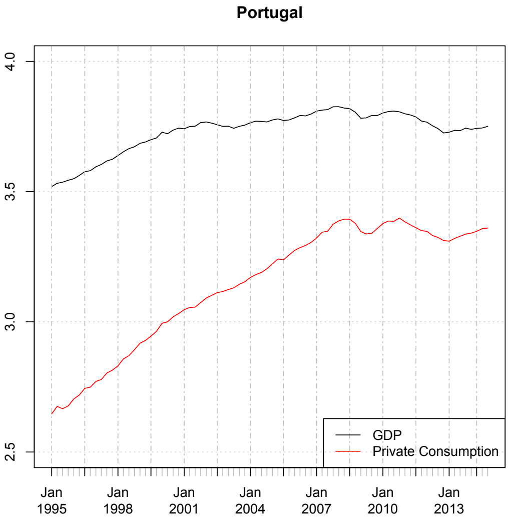
Hence,

$$\mu_{yt} = -\frac{\beta_2}{\beta_1} \mu_{zt} \tag{8}$$

- That is, for cointegration to exist between y_t and z_t, their stochastic trend must be identical up to a scalar.
- The essential insight of Stock and Watson (1988) is that the parameters of the cointegrating vector must be such that they purge the trend from the linear combination.

Example of Cointegration

- Income and consumption share a common trend:



Error Correction Model

- Its clear that for both to be in a long-run equilibrium, one of them or both need to adjust to movements of the other.
- That is, if consumption rises, either GDP has to rise next period, or consumption needs to fall next period or both or even GDP would have to rise by more than consumption next period.
- Hence, for a cointegration to exist, a dynamic adjustment must also exist. Such dynamic adjustment is called the error correction representation.
- Granger representation theorem: For any set of I(1) variables, error correction and cointegration are equivalent representations.

Error Correction Model

- Lets start with a 2 variable example of consumption and income and then we extend our analysis to a n variable case.
- Note that with two variables there can be at most $r = 1$ cointegration vector. Consumption and income have the following long-run equilibrium and cointegration vector $(1, -\beta)$:

$$y_t = \beta c_t + \varepsilon_t \tag{9}$$

We know they must adjust and one error-correction model is:

$$\begin{aligned} \Delta y_t &= -\alpha_y(y_{t-1} - \beta c_{t-1}) + \varepsilon_{yt} && (10) \\ \Delta c_t &= \alpha_c(y_{t-1} - \beta c_{t-1}) + \varepsilon_{ct} && (11) \end{aligned}$$

Error Correction Model

- In this case, if the income is above the long-run equilibrium level, then income would fall next period because of $-\alpha_y$ and consumption would increase, α_c.
- These forces help reestablish equilibrium. How fast? α_y, α_c determine the **speed of adjustment**.
- Note that the name error-correction comes form the fact that $y_{t-1} - \beta c_{t-1} = \varepsilon_{t-1}$. Hence the variables in difference react to deviations of the long-run equilibrium.
- Also, note that both (10) and (11) are I(0). Since y_t and c_t are both I(1), their first difference is stationary. Moreover, because of cointegration , $y_{t-1} - \beta c_{t-1} = \varepsilon_{t-1}$ is also stationary.
- We conclude that (10) and (11) can only exist if and only if there exists cointegration (9).

Error Correction Model

- Without loss of generality we can add lagged terms to (10) and (11) and we can generalize for n variables:

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \varepsilon_t \tag{12}$$

- Where π is the matrix with the cointegrating vectors. In our previous example we had the same cointegrating vectors, and so the rank of π was 1.

- Take a moment to look carefully at (12). Its looks very similar to a VAR in differences but now with an additional term πx_{t-1}. (12) is in fact the Vector Error-Correction Model (VECM).

VECM

- Lets start with a VAR(p):

$$x_t = A_1 x_{t-1} + ... + A_p x_{t-p} + \varepsilon_t \tag{13}$$

- Adding and Subtracting $A_p x_{t-p+1}$ on both sides we have:

$$x_t = A_1 x_{t-1} + ... + A_{p-2} x_{t-p+2} + (A_{p-1} + A_p) x_{t-p+1} - A_p \Delta x_{t-p+1} + \varepsilon_t$$

- Now Add and Subtract $(A_{p-1} + A_p) x_{t-p+2}$:

$$x_t = A_1 x_{t-1} + ... + A_{p-2} x_{t-p+2} - (A_{p-1} + A_p) \Delta x_{t-p+2} - A_p \Delta x_{t-p+1} + \varepsilon_t$$

$$\vdots$$

VECM

- Continue until:

$$x_t = \pi^* x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \varepsilon_t$$

- Finally subtract x_{t-1} on both sides and we have a VECM:

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \varepsilon_t \tag{14}$$

- Where $\pi = -(I - \sum_{i=1}^{p} A_i)$. Again, a VECM representation only makes sense if the rank of π, is greater than zero and less than n, e.g. $(0 < r < n)$. If it is zero, it means all elements of *pi* are zero and there is no cointegrating vector so that we just have a VAR in differences. If the rank is n, then all variables are stationary and cannot be cointegrated.

Testing for Cointegration

- There are two main ways to test for cointegration:

    1. The Engle-Granger methodology. It seeks to determine if the residuals of the long-term equilibrium are stationary.
    2. The Johansen methodology. It uses the VECM form and tests the rank of π

Engle-Granger Test for Cointegration - 4 steps

1. Pretest the variables for their order of integration.

    - By definition, cointegration requires that the variables are integrated of the same order.
    - Hence, first we need to test for the order of integration of each variable.
    - We can use the ADF test to infer the number of roots in each series.
    - If all are stationary we can stop here and use the standard stationary analysis we saw so far. If they are of different order, they are probably not cointegrated. We just need to double check for multicointegration.

Engle-Granger Test for Cointegration - 4 steps

2. Estimate the long-term relationship.

- After we confirm in step 1 that both variables are I(1), then we estimate the long-term equilibrum equation:

$$y_t = \beta c_t + \varepsilon_t \tag{15}$$

- In case of cointegreation, OLS can be used as the estimates of β are super-consistent. However, beware they do not have standard distributions and the usual *t*-tests do not apply. Hence, just use estimates. Do not use it to test β.

- Then we use the ADF test (critical values are different because β is estimated) on the residuals $\hat{\varepsilon}_t$ and check if they are stationary. If they are stationary, then there is cointegration. If not, probably spurious regression.

Engle-Granger Test for Cointegration - 4 steps

TABLE C: Critical Values for the Engle–Granger Cointegration Test

T	1%	5%	10%	1%	5%	10%
	Two Variables			***Three Variables***		
50	−4.123	−3.461	−3.130	−4.592	−3.915	−3.578
100	−4.008	−3.398	−3.087	−4.441	−3.828	−3.514
200	−3.954	−3.368	−3.067	−4.368	−3.785	−3.483
500	−3.921	−3.350	−3.054	−4.326	−3.760	−3.464
	Four Variables			***Five Variables***		
50	−5.017	−4.324	−3.979	−5.416	−4.700	−4.348
100	−4.827	−4.210	−3.895	−5.184	−4.557	−4.240
200	−4.737	−4.154	−3.853	−5.070	−4.487	−4.186
500	−4.684	−4.122	−3.828	−5.003	−4.446	−4.154

The critical values are for cointegrating relations (with a constant in the cointegrating vector) estimated using the Engle–Granger methodology.

Source: Critical values are interpolated using the response surface in MacKinnon (1991)

Figure: Caption

Engle-Granger Test for Cointegration - 4 steps

3. Estimate the error-correction model.

$$\Delta y_t = -\alpha_y(y_{t-1} - \beta c_{t-1}) + \varepsilon_{yt} \quad (16)$$

$$\Delta c_t = \alpha_c(y_{t-1} - \beta c_{t-1}) + \varepsilon_{ct} \quad (17)$$

- They propose a way to estimate them by replacing the residual of (15) in (16) and (17) because $\hat{\varepsilon}_t = y_t - \hat{\beta} c_t$

$$\Delta y_t = -\alpha_y(\hat{\varepsilon}_{t-1}) + \varepsilon_{yt} \quad (18)$$

$$\Delta c_t = \alpha_c(\hat{\varepsilon}_{t-1}) + \varepsilon_{ct} \quad (19)$$

- Hence, we can estimate (18) and (19) with OLS and use the usual *t*-tests as all of the variable are stationary. From this we get estimates of the speed of convergence $\hat{\alpha}_y, \hat{\alpha}_c$

Engle-Granger Test for Cointegration - 4 steps

4. Model checking and reports.

- Check if the residuals in (18) and (19) are I(0). IF they are not, then there might be a mispecification in the number of lags. That is, we might need to estimate:

$$\begin{aligned}
\Delta y_t &= -\alpha_y(\hat{\varepsilon}_{t-1}) + a_{11}\Delta y_{t-1} + a_{12}\Delta c_{t-1} + \varepsilon_{yt} \qquad (20) \\
\Delta c_t &= \alpha_c(\hat{\varepsilon}_{t-1}) + a_{21}\Delta y_{t-1} + a_{22}\Delta c_{t-1} + \varepsilon_{ct} \qquad (21)
\end{aligned}$$

- Test the speed of adjustment. Suppose α_y is not significant. In this case, all the adjustment to equilibrium comes from consumption.
- Finally, we can use the usual error decomposition (Choleski) to get the impulse responses coming from the VECM.

Engle-Granger Test for Cointegration - Limitations

- Which variable goes in the left-hand side makes a difference in small samples.
- There is no way to test for multiple cointegrating vectors with three variables or more.
- It relies in a two-step procedure. Uncertainty in step 1 is carried over to step two. (e.g. uncertainty in the parameters of the long-run regression carry over can compound with the short-run parameters uncertainty)

Johansen Test for Cointegration

- This tests used the general VECM form and tests the rank of π.
- Its main idea is to explore how the eigenvalues of π are related to the rank of π.

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \varepsilon_t \tag{22}$$

- The number of independent cointegrating vectors of π can be checked by the number of significant eigenvalues.

Johansen Test for Cointegration

- Because if there n linear independent cointegrating vectors, the determinant $|\pi| \neq 0$
- Moreover, since the determinant of a matrix is equal to the product of the eigenvalues $\prod_{i=1}^{n} \lambda_i$, we know that for the determinant to be non-zero, the eigeinvalues must all be non-zero as well.
- Hence, the rank of π is equal to the number of its characteristic roots that differ from zero.
- The Johansen methodology allows you to determine the number of roots that are statistically different from zero.

Johansen Test for Cointegration

- In practice, we need to estimate π and λ_i. We will not go deep into this estimation process. It suffices to say that OLS is not appropriate for this purpose.
- Two statistics allow you to test for cointegration:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} \ln\left(1 - \hat{\lambda}_i\right) \tag{23}$$

$$\lambda_{max}(r+1) = -T \ln\left(1 - \hat{\lambda}_{r+1}\right) \tag{24}$$

- The first one is based on the idea that if there are r linear independent vectors, then the rest $n - r$ eigenvalues should be all zero.
- The second statistics only uses the next one and checks if it is close to zero.

Johansen Test for Cointegration

- Formally, the first statistics λ_{trace} tests:

$$\begin{array}{ll} H_0: & \text{the number of distinct cointegrating vectors is} \leq r \\ H_1: & > r \end{array}$$

- The further the estimated characteristic roots are from zero, the higher is the statistic (23) and we more likely to reject the null hypothesis.

- The second statistic λ_{max} tests:

$$\begin{array}{ll} H_0: & \text{the number of cointegrating vectors is } r \\ H_1: & r + 1 \text{ cointegrating vectors} \end{array}$$

- Again the closer to zero the eigenvalue is the larger is the statistic and the more likely we are of rejecting the null.

Johansen Test for Cointegration

- Formally, the first statistics λ_{trace} tests:

$$\begin{array}{ll} H_0: & \text{the number of distinct cointegrating vectors is} \leq r \\ H_1: & > r \end{array}$$

- The further the estimated characteristic roots are from zero, the higher is the statistic (23) and we more likely to reject the null hypothesis.

- The second statistic λ_{max} tests:

$$\begin{array}{ll} H_0: & \text{the number of cointegrating vectors is } r \\ H_1: & r + 1 \text{ cointegrating vectors} \end{array}$$

- Again the closer to zero the eigenvalue is the larger is the statistic and the more likely we are of rejecting the null.

Johansen Test for Cointegration - Procedure

1. Select the VAR lag using the undifferenced data.
2. Estimate the model and determine the rank of π.
3. If there is more then one cointegrating vector, we may need to check for possible ways to impose restrictions. If we find only one vector, we can move to step 3.
4. Analyze the speed of adjustment and the normalized cointegrating vector.
5. IRF and FEVD.

Should we Difference the data?

- This question is open for debate. Some authors argue against it like Sims and other for it.
- The reason is that there are advantages and disadvantages in both cases.
- The advantages of Not differencing are the following:
  1. There can be cointegrating relationships. In this case, the estimates are in fact superconsistent, and estimating a model in differences is clearly mispecified.
  2. Even if there is no cointegration, the parameters are in generally still consistent. However, the tests have non-standard distributions.
  3. From a Baysian perspective, there is no issue with unit-roots for inference.
- Start in levels and if the results are the same with the VAR in differences, then the results are stronger since they do not depend on unit root assumptions.

Summary

- Many macro variables are non-stationary processes I(1).
- However, there are important equilibrium conditions that connect the stochastic trends of these processes together.
- We can test for cointegration using the Engle-Granger method or the Johansen method.
- The Johansen method is particularly useful for more than 2 variables systems.
- We should start with a VAR in levels and if we find non-sationary series, we should try to find cointegration. If not, we can see how the results look like for a VAR in differences.
- If they are consistent, the results do not depend on unit root assumptions.

Questions to think about

- Is cointegration necessarily an economic equilibrium?
- How is the significance of the speed of adjustment coefficients related to Granger Causality?
- Furthermore, are they related to IRF and FEVD?
- When should we difference the data?
- Can we use standard tests for the cointegrating coefficients β in the long-run equation?