WARWICK
Business School

Empirical Finance
Lecture 8: Analysis of non-stationary processes II: long-run relationships in empirical finance

Module Leader: Dr Stuart Fraser

stuart.fraser@wbs.ac.uk

Room D1.18 (Social Studies)

THE UNIVERSITY OF WARWICK

Today

Multivariate analysis with nonstationary variables

1. Spurious regressions
2. Cointegration
3. Analyzing long-run relationships in empirical finance

   Seminar 7: Cointegration analysis of long-run PPP.

(Brooks Chps 7.4-7.8; Verbeek Chp 9)

Spurious regressions

The problem: Independent non-stationary processes can appear to be related:

For example suppose...

$$Y_t = Y_{t-1} + \varepsilon_{1t}, \qquad \varepsilon_{1t} \sim IID\left(0, \sigma_1^2\right)$$
$$X_t = X_{t-1} + \varepsilon_{2t}, \qquad \varepsilon_{2t} \sim IID\left(0, \sigma_2^2\right), \quad \operatorname{cov}\left(\varepsilon_{1t}, \varepsilon_{2t}\right) = 0.$$

Y and X are independent random walks (I(1) processes).

...then a regression of Y on X...

$$Y_t = \mu + \beta X_t + \varepsilon_t$$

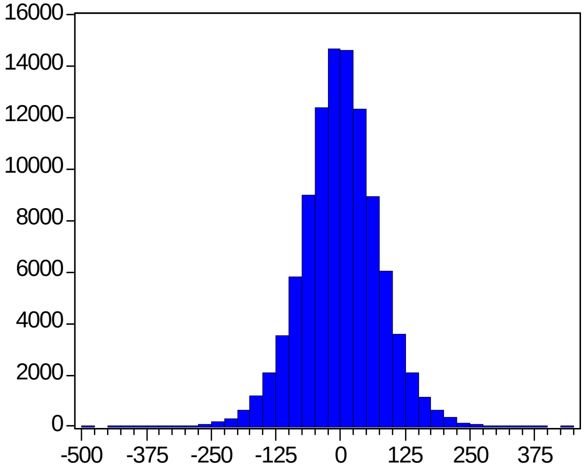
...will typically display:

- A high R-sq (even though it should be zero).
- A significant t-stat for β (even though it should be insignificant).

Clearly this can be very misleading in empirical work. However the regression also contains a clue that there is a problem with the equation:

- The error term is highly autocorrelated (in fact it's I(1) – see below).

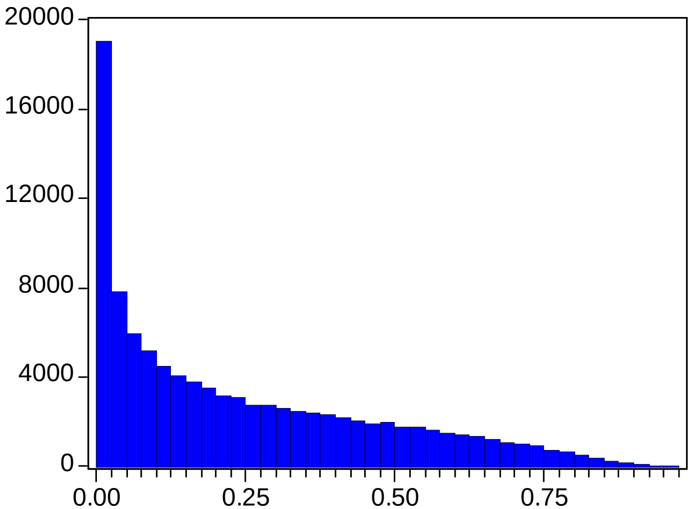
Spurious regression: sampling distributions of the t-stat and R-sq from a simulation with 100,000 samples of independent random walks with T=10,000



Series: T_SPUR	
Sample 1 100000	
Observations 100000	
Mean	0.023541
Median	0.088570
Maximum	448.0761
Minimum	-480.9837
Std. Dev.	74.15763
Skewness	-0.033843
Kurtosis	3.963160
Jarque-Bera	3884.412
Probability	0.000000

If we run regressions with independent series then we'd expect the absolute value of the t-stat for β to be >1.96 only 5% of the time in repeated samples (based on a normal dist.).

However in this case the distribution of t is highly non-normal: in fact 97.6% of the t-values are greater than 1.96 in absolute value! The t-ratios in these regression tend to indicate that there is a significant relationship even though there is none.



Series: RSQ_SPUR	
Sample 1 100000	
Observations 100000	
Mean	0.241821
Median	0.172911
Maximum	0.958967
Minimum	1.78e-12
Std. Dev.	0.226885
Skewness	0.843509
Kurtosis	2.674488
Jarque-Bera	12299.95
Probability	0.000000

We'd expect the R-sq to be close to 0 since the series are independent.

However here we find that:

- R-sq>0.5 in 16.4% of the samples.
- R-sq>0.9 in 0.1% of the samples!

Spurious regression: simulation results continued

Sample ACF of residuals

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
|********	|*****	1	0.998	0.998	9867.4	0
|********	|	2	0.996	0.01	19700	0
|********	|	3	0.995	-0.003	29498	0
|********	|	4	0.993	0.007	39261	0
|********	|	5	0.991	0.006	48991	0
|********	|	6	0.989	-0.011	58686	0
|********	|	7	0.987	0.001	68347	0
|********	|	8	0.986	0.008	77975	0
|********	|	9	0.984	-0.02	87568	0
|********	|	10	0.982	0.015	97127	0
|********	|	11	0.98	0.013	106654	0
|********	|	12	0.979	-0.013	116147	0
|********	|	13	0.977	0	125607	0
|*******|	|	14	0.975	-0.008	135033	0
|*******|	|	15	0.973	0.006	144426	0
|*******|	|	16	0.971	-0.008	153786	0
|*******|	|	17	0.97	0.005	163112	0
|*******|	|	18	0.968	0.01	172405	0
|*******|	|	19	0.966	0.003	181667	0
|*******|	|	20	0.964	-0.002	190896	0
|*******|	|	21	0.963	0.004	200094	0
|*******|	|	22	0.961	0.011	209261	0
|*******|	|	23	0.959	-0.019	218396	0
|*******|	|	24	0.958	0.012	227500	0

The residuals are highly autocorrelated. In fact they look <u>non-stationary</u>!

This is a major clue that the relationship between Y and X is spurious.

It indicates that Y and X do <u>not</u> move together in the long-run.

Spurious regression: the cause

The OLS estimator is:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{T^{-1}\sum x_i(\beta x_i + \varepsilon_i)}{T^{-1}\sum x_i^2}$$

$$= \beta + \frac{T^{-1}\sum x_i \varepsilon_i}{T^{-1}\sum x_i^2}$$

Sample covariance between X and ε

Sample variance of X

If Y and X are stationary and cov(X,ε)=0 (X is exogenous) then the OLS estimator is consistent:

$$T^{-1}\sum x_i \varepsilon_i \xrightarrow{p} \mathrm{cov}(X, \varepsilon)$$

$$\Rightarrow \hat{\beta} \xrightarrow{p} \beta$$

$'\xrightarrow{p}' \equiv$ convergence in probability

However with Y and X~I(1) and β=0 then ε~I(1). In that case the OLS estimator is inconsistent – it does not converge on β=0 ⇒Y and X appear to be related (even as T→∞).

The reason is that the stochastic trends in X and ε (both are I(1)) causes the sample covariance between X and ε to diverge (in probability) ⇒ it does not tend to cov(X,ε)=0.

Cointegration

In general linear combinations of I(1) variables form spurious relationships:

$$z_t = \beta_0 + \sum_{j=1}^{n} \beta_j X_{jt} \sim I(1)$$

The relationship is spurious because there is no tendency for the series to move together in the long-run (z is I(1)).

We get the case just analyzed when:

$$z_t = Y_t - \mu - \beta X_t \sim I(1), \quad \beta = 0.$$

Therefore, in OLS estimation with I(1) variables, typically:

1. Point estimators are inconsistent because the error term/z is I(1)
2. The t-stats follow non –normal distributions (see slide 4).

An important exception to 1. is where there are values of the β's such that z~I(0):

$$z_t = \beta_0 + \sum_{j=1}^{n} \beta_j X_{jt} \sim I(0)$$

z is a linear combination of I(1) variables which is I(0).
This combination of the variables is called a:
COINTEGRATING RELATIONSHIP
In this case z is CI(1,1) ('cointegrated of order one-one')
In general if z is a linear combination of I(d) variables which is I(d-b) (b>0) then z is CI(d,b).

Cointegration

The intuition behind cointegration is that the I(1) variables share the same fundamentals/long-run components:

- ⇒ The variables share a common stochastic trend

Individually the variables vary widely in the long-run:

- ⇒ Their variance is infinite.
- ⇒ Their spectra are infinite at frequency zero (⇒∞ long-run variation).

But in combination the variables move together in the long-run:

- ⇒ The variance of the combination is finite.
- ⇒ The spectrum of the combination at frequency zero is finite.

In effect the dominant long-run components of the individual variables 'cancel out' in the cointegrating relationship.

Cointegration

Cointegration is a <u>very</u> important concept in empirical finance because it means that variables which:

- Have <u>no equilibrium</u> tendency <u>individually</u> (because they are I(1))
- Are nonetheless bound together <u>in equilibrium as a group</u> (because a linear combination of the variables is I(0))

Cointegration analysis is therefore very important in analyzing the long-run/equilibrium properties of a system of non-stationary variables.

Clive Granger shared the Nobel Prize in Economics in 2003 (with Robert Engle who got it for ARCH):

"for methods of analyzing economic time series with common trends (cointegration)".

Cointegration: properties of OLS estimator

The cointegrating relationship can be written as a linear regression equation:

$$Y_t = b_1 + b_2 X_{2,t} + \ldots + b_n X_{n,t} + \varepsilon_t$$

Here we've just normalized the cointegrating relationship on one of the variables (X_1) which we've then made the dependent variable (Y).

The error term is the same as z in the previous formulation. Cointegration implies that the error term in a regression with I(1) variables is I(0).

The fact that $\varepsilon \sim I(0)$ makes a big difference to the OLS estimator compared to the spurious regressions analysis:

- Not only are the OLS estimators consistent they are SUPER-CONSISTENT.
- This means the estimator converges on the population coefficients much faster than in the stationary case.

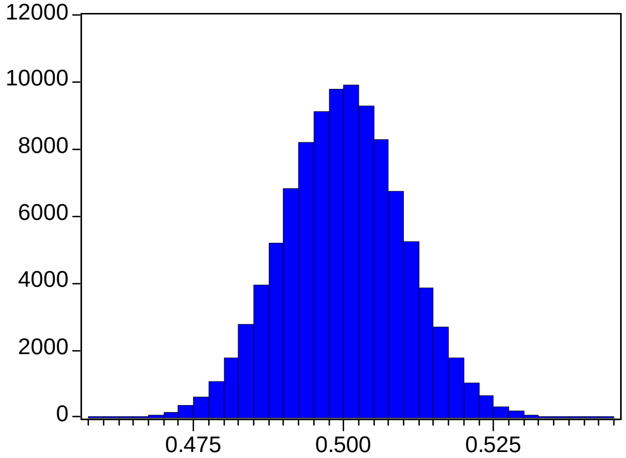
Why? OLS chooses parameter values which minimize the residual variance:

- Only the cointegrating relationship will have a finite variance: $\varepsilon \sim I(0)$.
- All other linear combinations are associated with an infinite residual variance: $\varepsilon \sim I(1)$.

Therefore OLS is very efficient at finding the cointegrating relationships (if they exist) from amongst all the other (non-stationary) linear combinations.

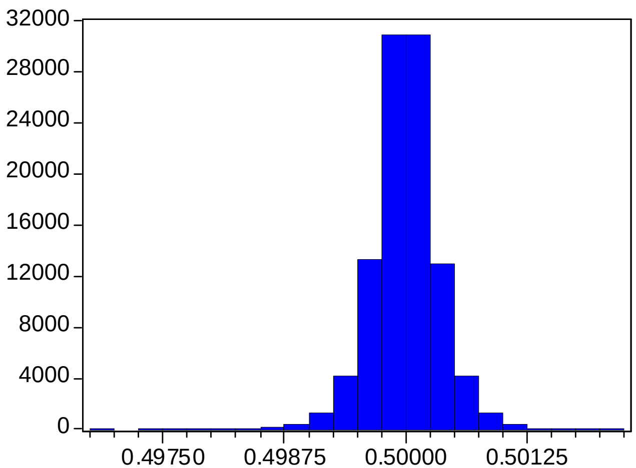
Simulation evidence on the properties of the OLS estimator: sampling distributions of the OLS estimator from a regression of Y on X in two instances

Stationary model



Series: SLOPES_S	
Sample 1 100000	
Observations 100000	
Mean	0.499972
Median	0.500006
Maximum	0.542554
Minimum	0.457939
Std. Dev.	0.010025
Skewness	-0.002035
Kurtosis	2.982246
Jarque-Bera	1.382321
Probability	0.500994

Cointegrating model



Series: SLOPES_NS	
Sample 1 100000	
Observations 100000	
Mean	0.499999
Median	0.499999
Maximum	0.502237
Minimum	0.496996
Std. Dev.	0.000332
Skewness	-0.017694
Kurtosis	4.905678
Jarque-Bera	15136.92
Probability	0.000000

Y is a simulated series with the following DGP:

$$Y_t = 0.5X_t + \varepsilon_t, \quad t = 1,\ldots,10{,}000$$
$$\varepsilon_t \sim NID(0,1)$$

100,000 samples of Y and X were obtained (X is either a random walk or an I(0) process – see below). The model was estimated by OLS 100,000 times to estimate the sampling distribution of $\hat{\beta}$.

The simulation was conducted in two instances:

1. Y,X and ε are I(0) (a classical stationary model).
2. Y and X are I(1); ε is I(0) (a cointegrating model).

The simulation evidence highlights…

Both the sampling distributions are centred very close to 0.5 (the true parameter value). Recall, the sample size is large: T=10,000.

However the mean is closer to 0.5 in the cointegrating model.

- The relative error is 0.0056% in the stationary model versus 0.0002% in the cointegrating model.

Relative error

$$\frac{|\hat{\beta}-\beta|}{\beta}\times 100$$

Also the estimates are less dispersed about the mean in the cointegrating model:

- The std dev is 0.01 in the stationary model versus 0.0003 in the cointegrating model.
- The range (max value-min value) is 0.085 in the stationary model versus 0.005 in the cointegrating model.

This highlights that the OLS estimator is collapsing on the true value faster for the cointegrating model than the stationary model ⇒ OLS is SUPER-CONSISTENT.

- In fact the sampling distribution is collapsing on β at the rate T for the cointegrating model versus $\sqrt{T}$ for the stationary model.

However the sampling distribution in the cointegrating model is highly non-normal (unlike the distribution in the stationary model which is normal).

This highlights again that classical inferences do not apply with non-stationary models (see also the DF distribution in lecture 7).

Examples of cointegration in empirical finance

<u>Relationship between spot and future prices</u>

Spot (s) and forward (f) prices are I(1).

However for a given asset we would expect s and f to be driven by the same fundamentals (share a common stochastic trend).

In that case there should exist a cointegrating relationship between f and s

$$f_t - s_t \sim CI(1,1)$$

The <u>cointegrating vector</u> is (1 -1)

For example in FX markets CIP $\Rightarrow$ $f_t^h - s_t = r_t - r_t^*$ and UIP$\Rightarrow \Delta s_{t+h} = r_t - r_t^*$(see lecture 4). Therefore if CIP and UIP hold then

$$f_t^h - s_t = \Delta s_{t+h} \sim I(0)$$

Examples of cointegration in empirical finance

The Expectations Hypothesis (EH) of the term structure (Brooks 7.12; Cuthbertson & Nitzsche 20.2 and 22.1)

The EH says that the expected one period holding yield on bonds of different maturities m should be equalized:

$$E\left(R_t^{(m)}\right) = R_t + \mathrm{T}, \quad \text{for all maturities } m.$$
$$\Rightarrow R_t^{(m)} = R_t + \mathrm{T} + \varepsilon_t^{(m)}$$

T≡Term premium (constant over time and independent of m).
R≡Known return on 1 period bond.

Second line follows assuming EH and rational expectations: $\varepsilon_t^{(m)}$ is a martingale difference error term.

If the yields are I(1) (and empirically they are) then the EH implies that the spreads are I(0).

$$R_t^{(n)} - R_t^{(m)} = \varepsilon_t^{(n)} - \varepsilon_t^{(m)} \sim CI(1,1).$$

The yields at different maturities are cointegrated. The cointegrating vector is (1 -1).

The implication is that the yields of different maturities are being driven by the same fundamentals (R_t+T): the yields share a common stochastic trend.

Examples of cointegration in empirical finance

Purchasing Power Parity (PPP) (see Seminars 6-8)

$$P_t = S_t P_t^*$$
$$\Rightarrow \ln S_t - \ln P_t + \ln P_t^* = 0$$

The log real exchange rate=0 in equilibrium (or a non-zero constant if the price indices are based In different years)

Adding a shock to the equilibrium at time *t* gives a stochastic equation

$$\ln S_t - \ln P_t + \ln P_t^* = \varepsilon_t$$

Equilibrium error:

$$E(\varepsilon_t) = 0$$
$$\varepsilon_t \sim I(0)$$

The cointegrating vector is: (1 -1 1)

lnS, lnP and lnP* are I(1) variables: see Seminar 6

For cointegration the equilibrium error must be stationary.

Conversely if ε~I(1) then shocks to the equilibrium will have a permanent effect

- No tendency for the system to revert back to equilibrium.
- No long-run PPP (spurious relationship).

Dynamic equations: short run dynamics versus long run equilibrium

The cointegrating relationship is a static equation which relates only to the long run equilibrium.

We need to look at a dynamic model to gain information about the short-run and other dynamics in the system:

Autoregressive distributed lag (ADL) model

$$Y_t = \mu + \delta_0 X_t + \delta_1 X_{t-1} + \phi Y_{t-1} + v_t$$

The immediate impact of X on Y is δ_0

In equilibrium: $Y_t = Y_{t-1} = Y_{t-2} = \ldots \equiv Y^*, \quad X_t = X_{t-1} = X_{t-2} = \ldots \equiv X^*$

So the dynamic model has the following long-run form

$$Y^* = \mu + \delta_0 X^* + \delta_1 X^* + \phi Y^*$$

$$\Rightarrow (1-\phi)Y^* = \mu + (\delta_0 + \delta_1)X^*$$

$$\Rightarrow Y^* = \frac{\mu}{1-\phi} + \frac{\delta_0 + \delta_1}{1-\phi} X^*$$

The long run impact of X on Y is:

$$(\delta_0 + \delta_1)/(1-\phi)$$

There exists a stable long run relationship if $\phi<1$. If $\phi=1$ then y|x has a unit root. In that case there is no cointegrating relationship between Y and X.

Dynamic equations: impact, interim and long-run multipliers

$$Y_t = \mu + \delta_0 X_t + \delta_1 X_{t-1} + \phi Y_{t-1} + v_t$$

$$\frac{\partial Y_t}{\partial X_t} = \delta_0$$

The immediate impact of a unit change in X on Y:
Impact multiplier

$$\frac{\partial Y_{t+1}}{\partial X_t} = \delta_1 + \phi \frac{\partial Y_t}{\partial X_t} = \delta_1 + \phi\delta_0$$

The impact after one period:
Interim multiplier (after 1-period)

$$\frac{\partial Y_{t+2}}{\partial X_t} = \phi \frac{\partial Y_{t+1}}{\partial X_t} = \phi(\delta_1 + \phi\delta_0)$$

The impact after two periods:
Interim multiplier (after 2-periods)

The impact after n periods:
Interim multiplier (after n-periods)

$$\frac{\partial Y_{t+n}}{\partial X_t} = \phi \frac{\partial Y_{t+n-1}}{\partial X_t} = \phi^{n-1}(\delta_1 + \phi\delta_0)$$

The **long-run (or equilibrium) multiplier** is the cumulative impact of a unit change in X on Y

$$\begin{aligned} &\delta_0 + (\delta_1 + \phi\delta_0) + \phi(\delta_1 + \phi\delta_0) + \dots \\ &= (\delta_0 + \delta_1) + \phi(\delta_0 + \delta_1) + \phi^2(\delta_0 + \delta_1) + \dots \\ &= \frac{\delta_0 + \delta_1}{1 - \phi} \end{aligned}$$

Same result as before (see the previous slide).

Error Correction Model (ECM)

The ECM is a popular representation of the ADL model which incorporates both long-run equilibrium and short run dynamics in the system.

From ADL to ECM

$$Y_t = \mu + \delta_0 X_t + \delta_1 X_{t-1} + \phi Y_{t-1} \qquad \text{(ADL)}$$

$$\Rightarrow \Delta Y_t = \mu + \delta_0 X_t + \delta_1 X_{t-1} + (\phi - 1) Y_{t-1} \qquad \text{(subtract } Y_{t-1} \text{ from both sides)}$$

$$\Rightarrow \Delta Y_t = \mu + \delta_0 \Delta X_t + (\delta_0 + \delta_1) X_{t-1} + (\phi - 1) Y_{t-1} \qquad \text{(subtract and add } \delta_0 X_{t-1} \text{ on the RHS)}$$

$$\Rightarrow \Delta Y_t = \delta_0 \Delta X_t + (\phi - 1)\left[Y_{t-1} - \frac{\mu}{1-\phi} - \frac{\delta_0 + \delta_1}{1-\phi} X_{t-1} \right] \qquad \text{(ECM)}$$

Short run dynamics

The speed of adjustment to dis-equilibrium is measured by $\phi - 1$

The Error Correction Term (ECT)
This is the equilibrium error in the previous period: ε_{t-1}

Comments on ECM

1. The ECM incorporates both long-run and short-run effects assuming $\phi<1$ (i.e., assuming cointegration).
2. If $\phi=1$ (no cointegration) then only the differenced variables appear in the model (corresponding to short run effects).
3. If there is cointegration $\phi-1<0$ parameterizes the speed of adjustment of *Y* to dis-equilibrium in the previous period.
   - If Y is above the long-run equilibrium in the previous period $\left(\Rightarrow \varepsilon_{t-1} > 0\right)$ then Y will fall in the following period (and vice-versa)

$$\varepsilon_{t-1} > 0 \Rightarrow (\phi - 1)\varepsilon_{t-1} < 0 \Rightarrow \Delta Y_t < 0$$
$$\varepsilon_{t-1} < 0 \Rightarrow (\phi - 1)\varepsilon_{t-1} > 0 \Rightarrow \Delta Y_t > 0$$

For example, if $\phi-1=-0.5$ then 50% of the dis-equilibrium in period t-1 is corrected in period t. A further $(\phi-1)\,\phi$ (=25%) of this dis-equilibrium is corrected in period t+1…
…a further $(\phi-1)\,\phi^n$ is corrected in period t+n.

The cumulative sum of the adjustments

$$(\phi-1)+(\phi-1)\phi+(\phi-1)\phi^2 \ldots$$

is -1 $\Rightarrow$100% adjustment back to equilibrium (in the long-run).

Comments on ECM

4. All the variables in the ECM are stationary
   - If Y and X are I(1) then ΔY are ΔX are I(0)
   - If Y and X are cointegrated then ε~I(0)
   - If Y and X are not cointegrated then only the differenced (stationary) variables appear in the model (it's no longer an ECM but a differenced model of Y).

Since all the variables are stationary, the ECM can be estimated by OLS with classical t and F tests being valid for inferences.

Granger Representation Theorem

This is a fundamental theorem in cointegration analysis. It states that:

- If there exists a linear combination z of I(1) variables such that z~CI(1,1) then there must exist an ECM for the data.
- If there exists an ECM for a group of I(1) variables then they must be cointegrated CI(1,1).

In other words, cointegration is both a necessary and sufficient condition for the existence of an ECM amongst I(1) variables:

$$\text{ECM for I(1) variables exists} \Leftrightarrow z\sim CI(1,1)$$

Given the ubiquity of ECMs in applied work the theorem has important empirical implications.

Notably, we need to test for cointegration before we can validly estimate an ECM for variables which are I(1) in levels.

Estimating Cointegrated Systems

Engle and Granger Two Step Estimator

EG 2 step estimator embodies the principle of the GRT:

- Firstly test for cointegration (estimate the long-run);
- Then, if there is cointegration, estimate the short-run dynamics in the ECM.

Step 1 (Estimate the long-run parameters)

a) Test the variables individually for unit roots (see seminar 7)

b) Estimate the cointegrating regression using OLS e.g.,

$$Y_t = \mu + \beta X_t + \varepsilon_t$$

c) Test the residuals for unit roots using an ADF test

$$\Delta\varepsilon_t = \rho\varepsilon_{t-1} + \sum_{j=1}^{p-1}\delta_j \Delta\varepsilon_{t-j} + v_t$$

No trend or intercept is required in this ADF test equation.

The alternative is that ε is stationary around a zero mean (the residuals have a zero mean by construction if the long-run equation includes a constant)

$$H_0 : \rho = 0 \Rightarrow \varepsilon \text{ is nonstationary} \Rightarrow \text{no cointegration}$$

$$H_1 : \rho < 0 \Rightarrow \varepsilon \text{ is stationary} \Rightarrow \text{cointegration}$$

Engle and Granger Two Step Estimator

Step 2 (Estimate the short run dynamics)

If $\varepsilon \sim I(0)$ then we can estimate an ECM using OLS e.g.,

Include an intercept in the ECM if there are trends in the data (analogous to including a drift term in a random walk).

The short-run dynamics are obtained from these parameters

$$\Delta Y_t = \alpha + \sum_{i=0}^{m} \delta_i \Delta X_{t-i} + \sum_{j=1}^{p} \gamma_j \Delta Y_{t-j} + \alpha \hat{\varepsilon}_{t-1} + v_t$$

Only include ΔX_t in the model if X is 'weakly exogenous for the short run parameters' $\Rightarrow cov(\Delta X_t, v_t) = 0$

Sometimes a model with just lagged values of ΔX (the 'reduced form') is estimated due to endogeneity issues.

Lagged long-run residuals from Step 1.

$\alpha < 0$ is the speed of adjustment to dis-equilibrium in the previous period.

Comments on EG 2 step

The OLS estimators of the long-run parameters are super-consistent (see slide 11).

However do not use t and F tests for inferences – they are invalid for inferences due to the nonstationarity of the variables.

Also standard DF critical values are invalid for the cointegration test

- The test involves estimated residuals rather than raw data.

OLS (which minimizes the residual variance) will tend to make these residuals appear stationary even if there is no cointegration.

- Standard DF critical values will tend to reject the null (no cointegration) too often.
- Use alternative critical values which take this issue into account (see Seminar 7).

Conclusions

- Distinguishing between spurious regressions and cointegrating relationships is crucial in multivariate analyses involving nonstationary variables.
  - Only in the presence of cointegration is there a meaningful relationship between the variables.
- Therefore cointegration analysis (testing for long-run relationships, estimating ECMs…) forms a fundamental (and commonplace) part of modern time-series econometrics.
- In this context the EG 2 Step estimator has intuitive appeal and is easily implemented (see Seminar 7).
- However there are major drawbacks with this approach which we will address in the final lecture.

References

Brooks (2002), Introductory econometrics for finance, CUP: Cambridge. Chps 7.4-7.8**.

Cuthbertson and Nitzsche (2004) Quantitative financial economics: stocks, bonds and foreign exchange, Wiley: Chichester. Chp 20.1 and 22.1

Verbeek (2004) A Guide to Modern Econometrics, 2nd Edition, Wiley: Chichester. Chps 9.1-9.3.