

Advanced time-series analysis (University of Lund, Economic History Department)

30 Jan-3 February and 26-30 March 2012

Lecture 6 Econometric techniques for non-stationary series 1: Cointegration and Error-Correction models.

6.a. Cointegration

Generally, the linear combination (for example the sum) of non-stationary time series is also non-stationary. The order of integration of the resulting series is usually the highest of the component series. So: generally if you add an $I(2)$ series with an $I(1)$ series you receive an $I(2)$ series.

This is the reason why regressing a non-stationary series on one or more series will often be a problem: the resulting residuals will be a linear combination of the variables of the regression and should be non-stationary, resulting in a spurious regression.

But there is an important exception! What there is linear combination of variables integrated of the same order d that is integrated of a lower order ($d-p$), then these variables are **cointegrated** of order $CI(d,p)$. The scalars (coefficients) of the linear combination are called the elements of the cointegrating vector. In case of cointegration the OLS estimates are super-consistent, meaning that you can not only trust them, but they tend to converge to the population parameters even faster than in case of an OLS on stationary variables.

Most often we handle the $CI(1,1)$ case when we have a number of $I(1)$ variables for which there exist a linear combination that is stationary $I(0)$.

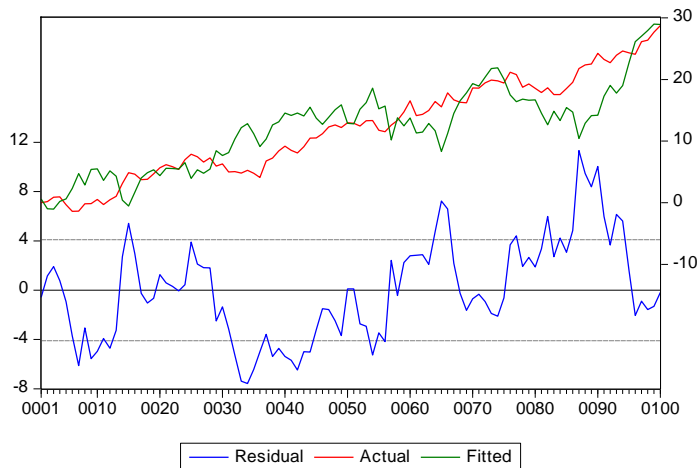
In order to see clearly let us take a simple case with the lack of cointegration:

We have two series, both are $I(1)$ but they are not cointegrated. We run a regression between them and obtain seemingly significant coefficient:

Dependent Variable: Y
Method: Least Squares
Date: 01/28/12 Time: 13:59
Sample: 0001 0100
Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.584111	0.802722	0.727663	0.4686
X	1.893089	0.113788	16.63705	0.0000
R-squared	0.738521	Mean dependent var		12.07146
Adjusted R-squared	0.735853	S.D. dependent var		7.965914
S.E. of regression	4.094100	Akaike info criterion		5.676768
Sum squared resid	1642.642	Schwarz criterion		5.728872
Log likelihood	-281.8384	Hannan-Quinn criter.		5.697855
F-statistic	276.7913	Durbin-Watson stat		0.288199
Prob(F-statistic)	0.000000			

Let us look at the graph:



You can see the original y and the fitted values from the regression. The residual is a linear combination of vector y , a unit vector (the intercept...) and vector x . You can see that there is no obvious tendency that the series are sharing some common movement: for that you should expect that the residual should have a tendency to return to its mean (zero). But this particular residual does not have such a tendency: since it is non-stationary, once it experiences a shock, its effect will not fade away and will have a permanent effect. Non-stationary series are non-mean reverting processes. So we can say that there is no mechanism that would keep y and x bound together: they are not cointegrated. Indeed, once you difference the series and re-run your regression you will find no relationship:

Dependent Variable: $D(Y)$
 Method: Least Squares
 Date: 01/28/12 Time: 14:12
 Sample (adjusted): 0002 0100
 Included observations: 99 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
$D(X)$	-0.103777	0.102763	-1.009862	0.3151
C	0.305953	0.101573	3.012137	0.0033
R-squared	0.010404	Mean dependent var		0.290263
Adjusted R-squared	0.000202	S.D. dependent var		0.998851
S.E. of regression	0.998750	Akaike info criterion		2.855371
Sum squared resid	96.75764	Schwarz criterion		2.907797
Log likelihood	-139.3408	Hannan-Quinn criter.		2.876582
F-statistic	1.019822	Durbin-Watson stat		2.080417
Prob(F-statistic)	0.315074			

Let us see now a case of cointegration:

We have quarterly data between 1950 and 2000 for the real consumption expenditures and the real GDP for the USA. (Greene, Table 5.1). We find that both variables are $I(1)$.

Null Hypothesis: REALCONS is stationary
 Exogenous: Constant, Linear Trend
 Bandwidth: 11 (Newey-West automatic) using Bartlett kernel

	LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic	0.400620
Asymptotic critical values*:	
1% level	0.216000
5% level	0.146000
10% level	0.119000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	69390.85
HAC corrected variance (Bartlett kernel)	691723.4

Null Hypothesis: D(REALCONS) is stationary
 Exogenous: Constant, Linear Trend
 Bandwidth: 8 (Newey-West automatic) using Bartlett kernel

	LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic	0.100475
Asymptotic critical values*:	
1% level	0.216000
5% level	0.146000
10% level	0.119000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	468.5064
HAC corrected variance (Bartlett kernel)	1144.830

Null Hypothesis: D(REALGDP) is stationary
 Exogenous: Constant, Linear Trend
 Bandwidth: 6 (Newey-West automatic) using Bartlett kernel

	LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic	0.082381
Asymptotic critical values*:	
1% level	0.216000
5% level	0.146000
10% level	0.119000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	1558.027
HAC corrected variance (Bartlett kernel)	2933.181

In both cases we found that the first-difference of the variables were trend stationary, since the trend coefficient in the test regression were significant (not reported in the tables here). Since both variables are integrated of the same order, there is a chance they are cointegrated.

The regression of real consumption of real GDP yields:

Dependent Variable: REALCONS
 Method: Least Squares
 Date: 01/28/12 Time: 14:50
 Sample: 1950Q1 2000Q4
 Included observations: 204

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-149.9919	6.490555	-23.10925	0.0000
REALGDP	0.690263	0.001291	534.5518	0.0000
R-squared	0.999294	Mean dependent var		2999.436
Adjusted R-squared	0.999290	S.D. dependent var		1459.707
S.E. of regression	38.89290	Akaike info criterion		10.16926
Sum squared resid	305556.9	Schwarz criterion		10.20179
Log likelihood	-1035.264	Hannan-Quinn criter.		10.18242
F-statistic	285745.7	Durbin-Watson stat		0.323727
Prob(F-statistic)	0.000000			

The Durbin-Watson test is very small, so we cannot be certain if the series are indeed cointegrated. A unit-root test quickly decides the case:

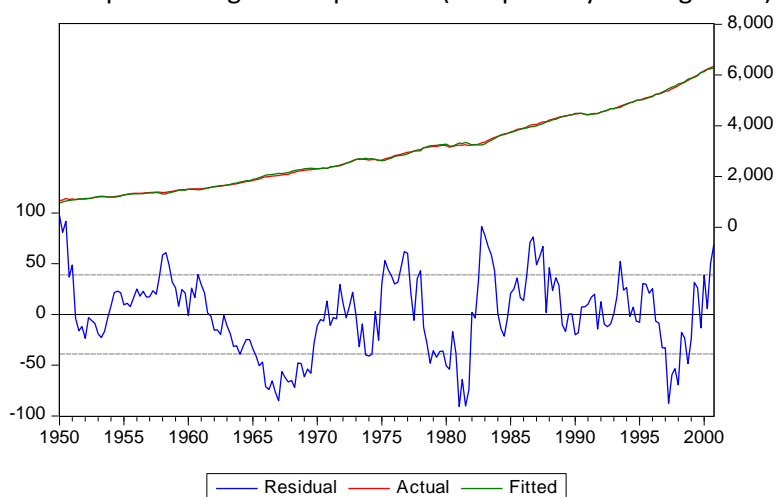
Null Hypothesis: RESID01 is stationary
 Exogenous: Constant
 Bandwidth: 10 (Newey-West automatic) using Bartlett kernel

	LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic	0.118084
Asymptotic critical values*:	
1% level	0.739000
5% level	0.463000
10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	1497.828
HAC corrected variance (Bartlett kernel)	8795.024

The residual is stationary, so the two variables are cointegrated. This should be a lesson to you that the rule of thumb based on the Durbin-Watson statistics is very unreliable. Actually it indicates the lack of spurious regression problem (and possibly cointegration) with high probability if it exceeds 1.



You can observe that the residual has a tendency to return to zero, so there must be a mechanism that keep the two series bounded together. It is not difficult to imagine what the underlying reason in this particular case: there is an equilibrium relationship between aggregate consumption and aggregate income. If the variables are cointegrated you can trust your OLS estimates so you already

have an estimate for the cointegrating vector: $\beta'=(1, 149.99, -0.69)$. The reason for this type of notation is that

$y\beta = (y_t, 1, x_t)(1, -\beta_0, -\beta_1)' = u_t \rightarrow y\beta = y_t - \beta_0 - \beta_1 x_t = u_t$. In this case we say that y was normalized to one, so the fitted value of the regression express the equilibrium value of y. There is no reason not to allow for x to be normalized to one: in this case you can choose x as dependent variable. Still, with this methodology you will not necessarily obtain what you would expect based on

renormalizing the cointegrating vector: $\left(-\frac{1}{\beta_1}, \frac{\beta_0}{\beta_1}, 1\right)' (y_t, 1, x_t) = v_t$. So this methodology is

sensitive to your choice of dependent variable. Nevertheless, in this particular case you get the renormalized cointegrating vector quite close to the expected values:

$\beta'_{renorm}=(-1.448, 220.4, 1)$.

Note: it is possible to have multiple cointegrating vectors, but the above methodology can only estimate a single one. Theoretically, if you have k variables you can have k-1 cointegrating vectors. It does not have to be so, however: it is perfectly all right if you find that say 4 variables are connected via a single cointegrating relationship. You should think about each cointegrating vectors as describing a rule (equilibrium relationship) that bind a group of variables together.

Now, if you look at a regression on first differences, you find a significant relationship:

Dependent Variable: D(REALCONS)
 Method: Least Squares
 Date: 01/28/12 Time: 15:06
 Sample (adjusted): 1950Q2 2000Q4
 Included observations: 203 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.332042	2.446247	0.544525	0.5867
D(REALGDP)	0.337358	0.030569	11.03592	0.0000
@TREND	0.116699	0.022091	5.282620	0.0000
R-squared	0.524857	Mean dependent var		26.02069
Adjusted R-squared	0.520105	S.D. dependent var		24.81678
S.E. of regression	17.19169	Akaike info criterion		8.541397
Sum squared resid	59110.82	Schwarz criterion		8.590361
Log likelihood	-863.9518	Hannan-Quinn criter.		8.561206
F-statistic	110.4628	Durbin-Watson stat		2.266090
Prob(F-statistic)	0.000000			

The linear trend was necessary to include since the first differences of the variables were trend stationary. But this is not the end of our analysis. The relationship between the first differences of the two variables reflects only the effect of a unit change in real GDP on real consumption: if GDP increased by one dollar, real consumption will increase by about 34 cents. But as you remember the long-run effect between x and y is estimated under the assumption that $\Delta x=0$ and $\Delta y=0$. So a specification including differences will not be able to capture the long-run relationship between x and y. The long-run relationship is reflected by the cointegrating vector(s).

For above reason it is a mistake to automatically run a regression on differenced variables once we find out that the original variables were non-stationary. If you have non-stationary variables, always look for a possible long-run relationship (cointegration). Only if you find none, you should go for regression on differences.

6.b. Differenced OLS (Stock and Watson, 1993)

Just above I suggested you to trust your OLS estimates of the cointegrating vector. Still since we do not know the exact structure of causality among (between) your variables, it can be more efficient in small samples to include the present lagged and future values of the first difference of your variables. Now the coefficients of the level of the variables can be interpreted as elements of the cointegration vector.

Dependent Variable: REALCONS
 Method: Least Squares
 Date: 01/28/12 Time: 15:48
 Sample (adjusted): 1950Q4 2000Q2
 Included observations: 199 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-152.5927	5.495649	-27.76609	0.0000
REALGDP	0.688580	0.001527	451.0226	0.0000
D(REALCONS)	0.477525	0.148951	3.205924	0.0016
D(REALGDP)	-0.328967	0.079783	-4.123266	0.0001
D(REALCONS(-1))	0.555500	0.154991	3.584083	0.0004
D(REALGDP(-1))	-0.300890	0.079286	-3.795010	0.0002
D(REALCONS(1))	-0.435703	0.150211	-2.900608	0.0042
D(REALGDP(1))	0.342620	0.079310	4.320019	0.0000
D(REALCONS(2))	-0.309653	0.143458	-2.158496	0.0322
D(REALGDP(2))	0.305808	0.078895	3.876153	0.0001
D(REALCONS(-2))	0.379313	0.152877	2.481159	0.0140
D(REALGDP(-2))	-0.242008	0.075853	-3.190476	0.0017
R-squared	0.999533	Mean dependent var	2994.904	
Adjusted R-squared	0.999506	S.D. dependent var	1420.577	
S.E. of regression	31.57926	Akaike info criterion	9.801285	
Sum squared resid	186485.6	Schwarz criterion	9.999876	
Log likelihood	-963.2278	Hannan-Quinn criter.	9.881660	
F-statistic	36407.90	Durbin-Watson stat	0.127427	
Prob(F-statistic)	0.000000			

The difference is marginal here. Still DOLS is quite popular, especially in panel analysis of cointegrated variables. How many lags and leads should you use? There is no established rule, but in doubtful cases you can rely on what you learned about model selection: e.g. use some information criteria.

6.c. Error-correction models

We actually know everything already to arrive at the error-correction representation (also referred to as EC or ECM).

We know that the residual from a cointegrating regression can be interpreted as the deviation of the dependent variable (the one whose coefficient is chosen to be normalized to one) from its

equilibrium (or long-run) value. If it is true than there should be a mechanism that works to eliminate this deviation. This mechanism is called the error-correction.

The simplest error-correction representation is as follows:

$\Delta y_t = -\gamma \hat{u}_{t-1} + e_t$ where \hat{u}_t is the residual from the cointegrating regression. If $\hat{u}_t > 0$ we can conclude that in period t the dependent variable is above its equilibrium value. If $\hat{u}_t < 0$ we can conclude that in period t the dependent variable is below its equilibrium value. The coefficient γ reflect the speed of adjustment, that is, the speed of convergence to the equilibrium value. It can be interpreted as the share of deviation from the equilibrium value eliminated during a single period.

The speed of adjustment is most often measured by the half-life, that is, the time needed in order to eliminate 50% of the deviation. This is calculated as follows: $t_{half\ time} = \frac{\ln 2}{\gamma} \approx \frac{0.69}{\gamma}$.

You can also include the first difference of the right-hand side variables as well, in order to estimate the immediate effects. This yields the classic Error Correction Model:

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t - \gamma \hat{u}_{t-1} + e_t$$

If γ is significantly different from zero, we can argue that there is a valid error-correction representation. If two or more non-stationary variables have a valid error correction representation, they are cointegrated. This is the **Granger representation theorem**.

Obviously, a valid adjustment coefficient means that the process y has the tendency to eliminate deviations from the cointegrating relationship, that is, it returns to the equilibrium value. This is cointegration. Actually, the ECM representation makes sense only, if the variables are cointegrated. if y and x are both I(1), then you have a stationary variable on the left side, and another one of the right side of the equation (Δx_t). Now the only way you can get e_t stationary if \hat{u}_{t-1} is also stationary. So actually the ECM indeed works as a cointegration test.

6.d. Estimation of ECM

You already know the first way to estimate an ECM as 6.c. basically introduced this method. This is called the **Engle-Granger two-step method**.

Provided you have two or more variables integrated of order one.

Step 1. You regress one of your variables on the rest. If your residual is stationary, your coefficients are your estimate for the cointegrating vector. Save your residuals.

Step 2. Estimate a regression of the first difference of your dependent variable on the first-difference of the rest of your variables, plus the lagged value of your residual from step 1.

You could make the whole process even simpler, however. Since $\hat{u}_{t-1} = y_{t-1} - \beta_0 - \beta_1 x_{t-1}$, you can just plug this into the ECM to obtain:

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t - \gamma (y_{t-1} - \beta_0 - \beta_1 x_{t-1}) + e_t = \alpha_0 + \gamma \beta_0 + \alpha_1 \Delta x_t - \gamma y_{t-1} + \gamma \beta_1 x_{t-1} + e_t$$

This is the **one-step Engle-Granger method**. Theoretically, you should be able to arrive at exactly the same cointegrating vector from the one-step method as from the two-step method, but you should be aware that even though the two methods are algebraically the same, statistically they are not. So, do not be surprised if the two methods lead to somewhat different results.

Let us try both methods on our example of real consumption and real GDP! The first step we already have, so we can go on to the second step:

Dependent Variable: D(REALCONS)
 Method: Least Squares
 Date: 01/29/12 Time: 06:09
 Sample (adjusted): 1950Q2 2000Q4
 Included observations: 203 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.13532	1.740758	5.822358	0.0000
D(REALGDP)	0.418397	0.031346	13.34755	0.0000
U(-1)	-0.084576	0.034486	-2.452448	0.0150
R-squared	0.474367	Mean dependent var		26.02069
Adjusted R-squared	0.469110	S.D. dependent var		24.81678
S.E. of regression	18.08205	Akaike info criterion		8.642384
Sum squared resid	65392.08	Schwarz criterion		8.691348
Log likelihood	-874.2020	Hannan-Quinn criter.		8.662193
F-statistic	90.24671	Durbin-Watson stat		2.026984
Prob(F-statistic)	0.000000			

Dependent Variable: D(REALCONS)
 Method: Least Squares
 Date: 01/29/12 Time: 06:12
 Sample (adjusted): 1950Q2 2000Q4
 Included observations: 203 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.182677	2.432800	0.486138	0.6274
D(REALGDP)	0.356581	0.032104	11.10708	0.0000
U(-1)	-0.061063	0.032928	-1.854431	0.0652
@TREND	0.110817	0.022185	4.995052	0.0000
R-squared	0.532928	Mean dependent var		26.02069
Adjusted R-squared	0.525887	S.D. dependent var		24.81678
S.E. of regression	17.08781	Akaike info criterion		8.534116
Sum squared resid	58106.68	Schwarz criterion		8.599401
Log likelihood	-862.2128	Hannan-Quinn criter.		8.560528
F-statistic	75.68618	Durbin-Watson stat		2.193543
Prob(F-statistic)	0.000000			

Since the first differences of both x and y were trend stationary, it is not a bad idea to include a linear trend as well in the ECM. We find a valid error-correction representation: the adjustment coefficient is of the right sign (that is we obtained a significant, negative coefficient for \hat{u}_{t-1} . It should be between 0 and -2).

The half-time based on the second ECM specification is 11.4 periods, that is about 11 quarters, almost three years. This is a slow adjustment: consumption may really wander off from its equilibrium path for extended periods.

Let us try the one-step approach now:

Dependent Variable: D(REALCONS)
 Method: Least Squares
 Date: 01/29/12 Time: 06:21
 Sample (adjusted): 1950Q2 2000Q4
 Included observations: 203 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-12.79770	6.282818	-2.036936	0.0430
D(REALGDP)	0.350302	0.032196	10.88022	0.0000
REALCONS(-1)	-0.059187	0.032809	-1.803999	0.0728
REALGDP(-1)	0.046227	0.022769	2.030263	0.0437
@TREND	-0.075893	0.115881	-0.654926	0.5133
R-squared	0.539198	Mean dependent var	26.02069	
Adjusted R-squared	0.529889	S.D. dependent var	24.81678	
S.E. of regression	17.01555	Akaike info criterion	8.530454	
Sum squared resid	57326.70	Schwarz criterion	8.612060	
Log likelihood	-860.8411	Hannan-Quinn criter.	8.563469	
F-statistic	57.92133	Durbin-Watson stat	2.213807	
Prob(F-statistic)	0.000000			

Now the trend is not significant, so you could remove it. Since it would not change much of the coefficients we stay with this for now.

The coefficient of the lagged real consumption is now our estimate of the adjustment parameter: -0.06, which yields a half-time estimate of 11.6 quarters. This is very similar to the two-step estimate. The long-run coefficient of the real GDP is $0.046/0.06=0.77$, higher than in case of the one-step estimate.

More complex methods are also available to estimate the cointegrating vectors and the adjustment coefficients. The most popular one was developed by Johansen and his colleagues at Copenhagen University. It is based on the estimation of a system of equations for each variables thought to be cointegrated (that is why it is called a vector regression). This is going to be a separate topic. The Engel-Granger method has drawbacks:

1. It is sensitive to the choice of the variable to be normalized.
2. It cannot estimate multiple cointegrating vectors.

The Johansen method does not have these weaknesses, but it requires a much more difficult procedure.