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

30 Jan-3 February and 26-30 March 2012

Lecture 2 Unit-root testing and the consequences of non-stationarity on regression analysis.

2.a. Why is non-stationarity a problem?

Most statistical tools are designed for stationary processes, that is, processes with finite moments. Once this condition does not hold, these methods are misleading.

Just think about the following: we have two processes that are independent of each other, but both have a positive trend. Now if you were to analyze their relationship with a traditional tool, say, linear correlation, you would find that they are positively related. This is a false (spurious) relationship.

The same is true for regression analysis: if you have two or more non-stationary variables and you run a regression on them you have high chance (in case on random walk processes about 80%) that you will find a relationship among them even though they are independent. (Granger and Newbold 1974)

What is the problem?

Generally, the linear combination of non-stationary series is also non-stationary. As such, we can observe that the residual from a time series regression is non-stationary. Now, all standard statistics from t-stats to R^2 are based on the idea that the residual is stationary. When should we get suspicious? Let's have an example:

I simulated two random walk processes: x and y. They are created independently, that is, x should not explain y or vice versa. Now I regress them on each other: $y_t = \beta_0 + \beta_1 x_t + u_t$. Knowing that the series are independent we expectat that $\beta_1 = 0$. Still, if we look at the output:

Dependent Variable: Y
Method: Least Squares
Date: 01/22/12 Time: 17:10
Sample: 0001 0100
Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.257769	0.300429	-7.515156	0.0000
X	0.405513	0.043703	9.278844	0.0000

R-squared	0.467672	Mean dependent var	-4.333821
Adjusted R-squared	0.462240	S.D. dependent var	2.734049
S.E. of regression	2.004936	Akaike info criterion	4.248899
Sum squared resid	393.9375	Schwarz criterion	4.301002
Log likelihood	-210.4450	Hannan-Quinn criter.	4.269986
F-statistic	86.09694	Durbin-Watson stat	0.255836
Prob(F-statistic)	0.000000		

We find a statistically significant coefficient! This should not be so. This result is false. This is a classical case of spurious regression. What are the signs that some problems are there?

1. First, if the residual is non-stationary then it should be strongly autocorrelated: $u_t = \rho u_{t-1} + \varepsilon_t$. Let us look at the Durbin Watson test statistics: $DW = \frac{\sum_{t=2}^{T}(u_t - u_{t-1})^2}{\sum_{t=1}^{T} u_t^2} \approx 2(\rho - 1)$. So if ρ=1, then DW=0. Hence, the first sign is that the DW is close to zero.

2. Secondly the R-squared should be near zero, since x does not explain anything of y. Still, in case of spurious regression the R^2 statistics is going to converge to a positive value, which can be rather high. Usually the rule of thumb is that if the DW is lower than the R-squared you are likely to have a spurious regression at hand.

3. None of the above methods can substitute a formal testing of the stationarity/non-stationarity of the residual, though.

2. b. Removing unit-root, order of integration

First of all, we need to know how to transform non-stationary series to become stationary. the obvious way is first-differencing, since if we have the following random-walk:

$y_t = y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$

then $\Delta y_t = y_t - y_{t-1} = \varepsilon_t$, which is stationary.

The same method works fine for two other type of non-stationary series as well:

a random-walk with drift: $y_t = \alpha + y_{t-1} + \varepsilon_t,\ \Delta y_t = \alpha + \varepsilon_t$

and a so-called deterministic trend process:

$y_t = \alpha + \delta t + \varepsilon_t, \Delta y_t = \delta + \Delta\varepsilon_t$. In the latter case, however, simply including a linear trend in the regression analysis would have been sufficient and preferable as well to get rid of the spurious regression problem.

For this reason the random walk processes are called difference stationary process, while the deterministic trend process is called trend stationary process.

If a series has to be differenced once in order to become stationary we call the series integrated of order one or I(1).

If a series has to be differenced d times to be stationary, we call it integrated of order d or I(d).

The difference operator is:

$\Delta^d = (1-L)^d$ where we normally assume that d is zero or a positive integer number.

Well, actually we assume that d=0,1,2,..., but it needs not to be so.

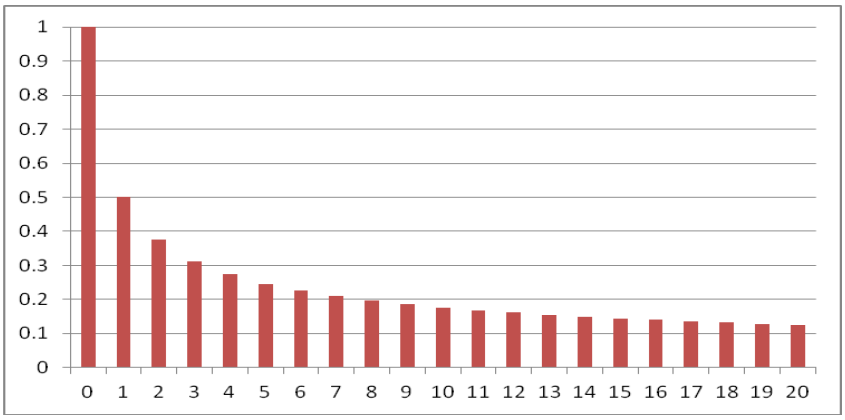
What if I said that d=1/2 ? Differencing half? Does it sound crazy? Actually, this type of time-series can exist and they are called fractionally integrated series. These are also long-memory processes.

$$(1-L)^{0.5} y_t = \varepsilon_t \rightarrow y_t = (1-L)^{-0.5} \varepsilon_t$$

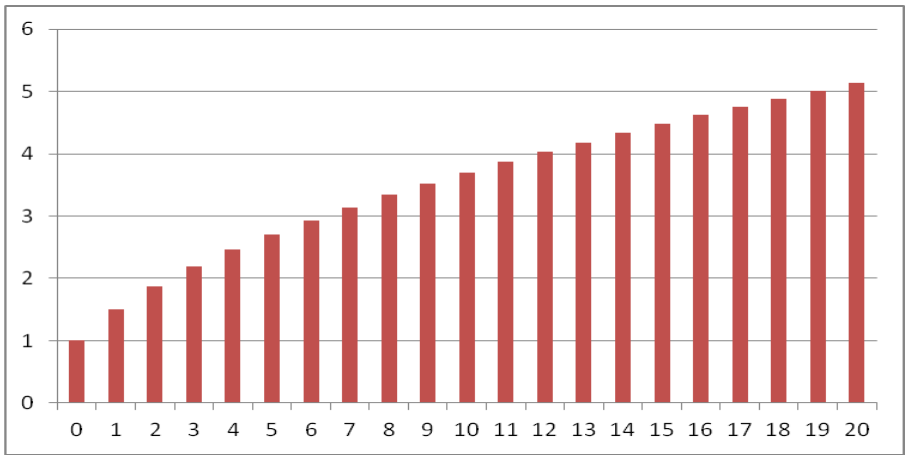
See Hamilton (1997) for the solution: $(1-L)^{-d} = 1 + dL + \frac{(d+1)dL^2}{2!} + \frac{(d+2)(d+1)dL^3}{3!} + \ldots$

$y_t = \varepsilon_t + 0.5\varepsilon_{t-1} + \frac{0.75}{2}\varepsilon_{t-2} + \frac{1.875}{6}\varepsilon_{t-3} + \ldots$, that is, it leads to an infinite MA model. The process remembers to every single shock that occurred to it, with decreasing weight:

IRF of a unit innovation in ε $(1-L)^{0.5} y_t = \varepsilon_t$



cumulative IRF of a unit innovation in ε $(1-L)^{0.5} y_t = \varepsilon_t$



If d<0.5 the provess is covariance stationary, since the IRF at high lags (j) is going to be $(j+1)^{d-1}$. This converges to zero once we have d lower then 0.5. If 0.5<d<1, it is advised to difference the series first, since:

$$(1-L)^{d} y_t = (1-L)^{d-1}(1-L) y_t = \varepsilon_t$$

Nevertheless, this problem falls out of the scope of this course, but you can understand the importance of fractionally intergate series: they are long-memory processes with decaying effect of past shocks.

In economics and social sciences stationary I(0) and non-stationary I(1) series are the most common. Sometimes we find I(2)series. I(3) are extremely rare (price level during extreme hyperinflations).

2.c Testing for the order of integration

By simply looking at a series you will not be able to decide if it is stationary or not or if it needs to be differenced. Let us take the following two series:

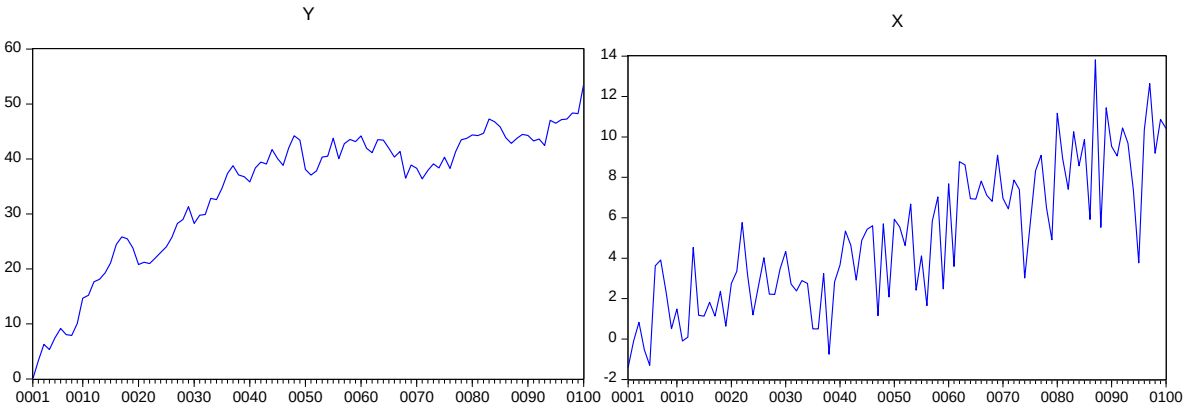
$y_t = \alpha + y_{t-1} + \varepsilon_t$ a random walk with drift (also called stochastic trend model)

and

$y_t = \alpha t + \eta_t$ a deterministic trend model (also called trend stationary)

The expected value of both series is going to be:

$Ey_t = \alpha t$, provided the initial value is zero.



Would you be able to tell which of the above series are produced by a deterministic or a stochastic trend model if you had not been told the truth? The honest is answer is no. You can guess but that is not enough.

Actually the distinction between trend stationary and random walk with drift series came into the focus of debates in macroeconomics during the '70s and remained there as a crucial point of research until the '90s.

The traditional view on business cycles was motivated by Keynesian models, where the long-run output was determined by technology (supply side) leading to some smooth growth of output, and

business cycles were attributed to short-run deviations from equilibrium. This idea can be modeled by a trend stationary process:

$\ln y_t = \alpha + \delta t + u_t$, where y denotes per capita income and the parameter δ is the growth rate of per capita income per period. The residual u has all the shocks that would capture the deviations from the long-run growth path. In this model one would expect that the shocks have just an immediate effect so the economy should return to its long-run growth path.

If you do not like this much, you can assume that it takes some time to return to equilibrium, by making the model autoregressive:

$\ln y_t = \alpha_0 + \alpha_1 \ln y_{t-1} + \delta t + u_t$ ($0<|\alpha_1|<1$), but the model is still trend stationary.

Kydland and Prescott, followed by many, argued that cycles may actually arise even in equilibrium. That is, the equilibrium output of the economy, that had been believed to follow a smooth path, may also exhibit cyclical behavior. Because these are shocks in real (not nominal) variables, this school is called the Real Business Cycle School or RBC.

This idea is best captured by a stochastic trend model:

$\ln y_t = \ln y_{t-1} + \delta + u_t$

The economy will have a tendency to grow at a rate δ, but any shocks (in this context it is called technological or productivity shock) could have a lasting effect. Remember: having a unit-root means exactly the same. So cycles are not always signs of disequilibrium.

This is the reason why the article of Granger and Newbold in 1974 became so influential. People started testing if macroeconomic series playing a role in RBC theory (GDP, employment, TFP) were trend stationary or not.

Dickey Fuller test:

The first and most fundamental unit-root test:

There are three possible null-hypotheses:

1. The DGP is a random-walk: $y_t = y_{t-1} + u_t$ with $u_t \sim IID(0, \sigma_u^2)$.

   In this case the test equation is $\Delta y_t = \rho y_{t-1} + u_t$, where if $\rho = 0$ then the null hypothesis is accepted, if $\rho < 0$ then it is rejected. The test equation is estimated by OLS, but the distribution of the coefficient is not going to converge to normality so you should not use Student's t-distribution to find critical values. Instead critical values are supplied in a different table or built in into some packages.

2. The DGP is either a random-walk with a constant or a random-walk with drift: $y_t = \alpha + y_{t-1} + u_t$ with $u_t \sim IID(0, \sigma_u^2)$. In this case the test equation is $\Delta y_t = \beta + \rho y_{t-1} + u_t$, where if $\rho = 0$ then the null hypothesis is accepted, if $\rho < 0$ then it is rejected.

3. The DGP is a random-walk with a constant and a deterministic trend: $y_t = \alpha + y_{t-1} + \delta t + u_t$ with $u_t \sim IID(0, \sigma_u^2)$. In this case the test equation is $\Delta y_t = \beta_0 + \rho y_{t-1} + \beta_1 t + u_t$, where if $\rho = 0$ then the null hypothesis is accepted, if $\rho < 0$ then it is rejected. If we reject the null-hypothesis but find that $\beta_1 \neq 0$, then the series are trend-stationary.

So that the DF test works properly the residual should exhibit no serial correlation. But this assumption is too bold. Let us see why:

You can observe that the DF test uses an AR(1) model to test is a series is non-stationary. But what if the real DGP is a higher-order AR model?

$y_t = 0.7y_{t-1} + 0.3y_{t-2} + \varepsilon_t$, this model is also non-stationary since the sum of the two coefficients equals one (alternatively you can estimate the roots that are -1.43 and 1).

After differencing the real DGP becomes:

$$\Delta y_t = -0.3y_{t-1} + 0.3y_{t-2} + \varepsilon_t$$

If you approximate this equation by an AR(1) model you will get a coefficient close to zero but the omitted lags will cause an autocorrelation in u.

There are two solutions:

1. Correcting the test statistics of the DF test for serial correlation and possible heteroscedasticity: this is the Phillips-Perron (PP) test.
2. Adding lags of the dependent variable in the test equation of the DF test to capture autocorrelation: Augmented Dickey-Fuller test (ADF test).

The augmented Dickey Fuller test has the following test equations:

$$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^{q} \Delta y_{t-i} + u_t, \Delta y_t = \beta + \rho y_{t-1} + \sum_{i=1}^{q} \Delta y_{t-i} + u_t, \Delta y_t = \beta_0 + \rho y_{t-1} + \beta_1 t + \sum_{i=1}^{q} \Delta y_{t-i} + u_t$$

Choosing the q can be done based on model selection statistics, like the AIC or BIC. Smarter softwares select q for you, based on some criteria.

Other modifications for the DF test are also available: the DF-GLS test for example, detrend the series before running a standard ADF procedure.

The main problem of DF type test is the low power. This means that the probability of Type II error is high. For the DF type tests this means that you have a high chance that when you have an autoregressive DGP with a high degree of positive autocorrelation, close to but less than one, the test will not reject the null hypothesis.

Solution:

Kwiatkowski, Phillips, Schmidt and Shin suggested a test with stationarity as null hypothesis (KPSS-test).

The null hypotheses can be that the series is either stationary or trend stationary. Depending on this the test equations are:

$y_t = \alpha + \hat{u}_t$ or $y_t = \alpha + \delta t + \hat{u}_t$

The statistics is based on the residual:

$S_t = \sum_{s=1}^{t} \hat{u}_s$ which is the recursive sum of the residual.

The test statistics is: $KPSS = \frac{T^{-2}\sum_{t=1}^{T} S_t^2}{\hat{\sigma}^2}$ where $\hat{\sigma}^2$ is the estimated long-run variance of $\hat{u}$.

Using the unit roots, determining the order of integration is not difficult:

1. We carry out a unit-root test on the level of y: if it is stationary then the process is I(0). If not, we take the difference of the series.
2. We carry out a unit-root test on Δy. If it is stationary then y is I(1), if not, we difference Δy further.
3. The process is followed as long as it needed to achieve stationarity.

2.d Overdifferencing

A word of caution: it is possible to overdifference the series, that is, to take the difference of a non-stationary series, which will lead to a special autoregressive pattern.

Let us assume that y is stationary:

$y_t = \varepsilon_t$ with $\varepsilon_t \sim IID(0, \sigma_u^2)$

taking first difference yields:

$x_t = \Delta y_t = \varepsilon_t - \varepsilon_{t-1}$

The first consequence is obvious:

$\sigma_x^2 = \sigma_{\varepsilon_t}^2 + \sigma_{\varepsilon_{t-1}}^2 = 2\sigma_\varepsilon^2$ So the differenced series is going to have higher variance!

The second consequence is that x is going to be serially correlated:

$$Cov(x_t, x_{t-1}) = E\left(\left(\varepsilon_t - \varepsilon_{t-1}\right)\left(\varepsilon_{t-1} - \varepsilon_{t-2}\right)\right) = E\left(\varepsilon_t\varepsilon_{t-1} - \varepsilon_t\varepsilon_{t-2} - \varepsilon_{t-1}^2 + \varepsilon_{t-1}\varepsilon_{t-2}\right) = -\sigma_\varepsilon^2$$

$$\gamma_1 = \frac{Cov(x_t, x_{t-1})}{\sigma_x^2} = -0.5\,,\ \gamma_j = \frac{Cov(x_t, x_{t-j})}{\sigma_x^2} = 0 \text{ for all j>1}$$

That is, if we obtain a variable after differencing that has a first order autocorrelation around -0.5

then we have reason to believe that we overdifferenced it.

ACF and PACF of the first difference of a standard normal variable.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.493	-0.493	24.835	0.000
		2	-0.071	-0.416	25.355	0.000
		3	0.118	-0.218	26.810	0.000
		4	-0.124	-0.299	28.424	0.000
		5	0.054	-0.266	28.738	0.000
		6	0.113	-0.067	30.121	0.000
		7	-0.099	-0.031	31.179	0.000
		8	0.036	0.057	31.325	0.000
		9	-0.139	-0.175	33.462	0.000
		10	0.153	-0.041	36.076	0.000
		11	-0.054	-0.102	36.407	0.000
		12	-0.046	-0.188	36.650	0.000
		13	0.129	-0.066	38.580	0.000
		14	-0.066	0.007	39.096	0.000
		15	-0.064	-0.017	39.589	0.001

Note: observe that the ACF(1) is close to -0.5.