UNIT ROOT TESTS, COINTEGRATION, ECM, VECM, AND CAUSALITY MODELS

Compiled by Phung Thanh Binh[1]

(SG - 30/11/2013)

"EFA is destroying the brains of current generation's researchers in this country. Please stop it as much as you can. Thank you."

The aim of this lecture is to provide you with the key concepts of time series econometrics. To its end, you are able to understand time-series based researches, officially published in international journals[2] such as applied economics, applied econometrics, and the likes. Moreover, I also expect that some of you will be interested in time series data analysis, and choose the related topics for your future thesis. As the time this lecture is compiled, I believe that the Vietnam time series data[3] is long enough for you to conduct such studies. This is just a brief summary of the body of knowledge in the field according to my own understanding.

[1] School of Economics, University of Economics, HCMC. *Email:* ptbinh@ueh.edu.vn.

[2] Selected papers were compiled by Phung Thanh Binh & Vo Duc Hoang Vu (2009). You can find them at the H library.

[3] The most important data sources for these studies can be World Bank's *World Development Indicators*, IMF-IFS, GSO, and Reuters Thomson.

Therefore, it has no scientific value for your citations. In addition, researches using bivariate models have not been highly appreciated by international journal's editors and my university's supervisors. As a researcher, you must be fully responsible for your own choice in this field of research. My advice is that you should firstly start with the research problem of your interest, not with data you have and statistical techniques you know. At the current time, EFA becomes the most stupid phenomenon of young researchers that I've ever seen in my university of economics, HCMC. They blindly imitate others. I don't want the series of models presented in this lecture will become the second wave of research that annoys the future generation of my university. Therefore, just use it if you really need and understand it.

Some topics such as serial correlation, ARIMA models, ARCH family models, impulse response, variance decomposition, structural breaks[4], and panel unit root and cointegration tests are beyond the scope of this lecture. You can find them elsewhere such as econometrics textbooks, articles, and my lecture notes in Vietnamese.

The aim of this lecture is to provide you:

- An overview of time series econometrics
- The concept of nonstationary
- The concept of spurious regression

[4] My article about threshold cointegration and causality analysis in growth-energy consumption nexus (www.fde.ueh.edu.vn) did mention about this issue.

- The unit root tests
- The short-run and long-run relationships
- Autoregressive distributed lag (ARDL) model and error correction model (ECM)
- Single-equation estimation of the ECM using the Engle-Granger 2-step method
- Vector autoregressive (VAR) models
- Estimating a system of ECMs using vector error correction model (VECM)
- Granger causality tests (both cointegrated and non-cointegrated series)
- Optimal lag length selection criteria
- ARDL and bounds test for cointegration
- Basic practicalities in using Eviews and Stata
- Suggested research topics

1. AN OVERVIEW OF TIME SERIES ECONOMETRICS

In this lecture, we will mainly discuss single equation estimation techniques in a very different way from what you have previously learned in the basic econometrics course. According to Asteriou (2007), there are various aspects to time series analysis but the most common theme to them is to fully exploit the dynamic structure in the data. Saying differently, we will extract as much information as possible from the past history of the series. The analysis of time series is usually explored within two fundamental types, namely, **time series forecasting** and **dynamic modelling.** Pure time series forecasting, such as ARIMA and ARCH/GARCH family models, is often mentioned as *univariate* analysis. Unlike most other econometrics, in univariate analysis we do not concern much with building structural models, understanding the economy or testing hypothesis, but what we really concern is developing efficient models, which are able to forecast well. The efficient forecasting models can be empirically evaluated using various ways such as significance of the estimated coefficients (especially the longest lags in ARIMA), the positive sign of the coefficients in ARCH, diagnostic checking using the correlogram, Akaike and Schwarz criteria, and graphics. In these cases, we try to exploit the dynamic inter-relationship, which exists over time for any single variable (say, asset prices, exchange rates, interest rates, ect). On the other hand, dynamic modelling, including *bivariate* and *multivariate* time series analysis, is mostly concerned with understanding the structure of the economy and testing hypothesis. However, this kind of modelling is based on the view that most economic series are slow to adjust to any shock and so to understand the process must fully capture the adjustment process which may be long and complex (Asteriou, 2007). The dynamic modelling has become increasingly popular thanks to the works of two Nobel laureates in economics 2003, namely, Granger (for methods of analyzing economic time series with common trends, or cointegration) and

Engle (for methods of analyzing economic time series with time-varying volatility or ARCH)[5]. Up to now, dynamic modelling has remarkably contributed to economic policy formulation in various fields. Generally, the key purpose of time series analysis is to capture and examine the dynamics of the data.

In time series econometrics, it is equally important that the analysts should clearly understand the term **stochastic process**. According to Gujarati (2003), "a random or stochastic process is a collection of random variables ordered in time". If we let Y denote a random variable, and if it is continuous, we denote it a Y(t), but if it is discrete, we denote it as Y_t. Since most economic data are collected at discrete points in time, we usually use the notation Y_t rather than Y(t). If we let Y represent GDP, we have Y_1, Y_2, Y_3, …, Y_{88}, where the subscript 1 denotes the first observation (i.e., GDP for the first quarter of 1970) and the subscript 88 denotes the last observation (i.e. GDP for the fourth quarter of 1991). Keep in mind that each of these Y's is a random variable.

In what sense we can regard GDP as a stochastic process? Consider for instance the GDP of $2873 billion for 1970Q1. In theory, the GDP figure for the first quarter of 1970 could have been any number, depending on the economic and political climate then prevailing. The

[5] http://nobelprize.org/nobel_prizes/economics/laureates/2003/

figure of $2873 billion is just a particular **realization** of all such possibilities. In this case, we can think of the value of $2873 billion as the mean value of all possible values of GDP for the first quarter of 1970. Therefore, we can say that GDP is a stochastic process and the actual values we observed for the period 1970Q1 to 1991Q4 are a particular realization of that process. Gujarati (2003) states that "*the distinction between the stochastic process and its realization in time series data is just like the distinction between population and sample in cross-sectional data*". Just as we use sample data to draw inferences about a population; in time series, we use the realization to draw inferences about the underlying stochastic process.

The reason why I mention this term before examining specific models is that all basic assumptions in time series models relate to the stochastic process (population). Stock & Watson (2007) say that the assumption that the future will be like the past is an important one in time series regression. If the future is like the past, then the historical relationships can be used to forecast the future. But if the future differs fundamentally from the past, then the historical relationships might not be reliable guides to the future. Therefore, in the context of time series regression, the idea that historical relationships can be generalized to the future is formalized by the concept of **stationarity.**

2. STATIONARY STOCHASTIC PROCESSES

2.1 Definition

According to Gujarati (2003), a key concept underlying stochastic process that has received a great deal of attention and scrutiny by time series analysts is the so-called **stationary stochastic process.** Broadly speaking, "*a time series is said to be stationary if its* **mean** *and* **variance** *are constant over time and the value of the* **covariance**[6] *between the two periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed*" (Gujarati, 2011). In the time series literature, such a stochastic process is known as a **weakly stationary** or **covariance stationary.** By contrast, a time series is **strictly stationary** if all the moments of its probability distribution and not just the first two (i.e., mean and variance) are invariant over time. If, however, the stationary process is normal, the weakly stationary stochastic process is also strictly stationary, for the normal stochastic process is fully specified by its two moments, the mean and the variance. For most practical situations, the weak type of stationarity often suffices. According to Asteriou (2007), a time series is weakly stationary when it has the following characteristics:

[6] or the autocorrelation coefficient.

(a) exhibits mean reversion in that it fluctuates around a constant long-run mean;

(b) has a finite variance that is time-invariant; and

(c) has a theoretical correlogram that diminishes as the lag length increases.

In its simplest terms a time series Y_t is said to be weakly stationary (hereafter refer to stationary) if:

(a) Mean: $E(Y_t) = \mu$ (constant for all t);

(b) Variance: $Var(Y_t) = E(Y_t-\mu)^2 = \sigma^2$ (constant for all t); and

(c) Covariance: $Cov(Y_t,Y_{t+k}) = \gamma_k = E[(Y_t-\mu)(Y_{t+k}-\mu)]$

where γ_k, covariance (or autocovariance) at lag *k*, is the covariance between the values of Y_t and Y_{t+k}, that is, between two Y values *k* periods apart. If *k* = 0, we obtain γ_0, which is simply the variance of Y ($=\sigma^2$); if k = 1, γ_1 is the covariance between two adjacent values of *Y*.

Suppose we shift the origin of Y from Y_t to Y_{t+m} (say, from the first quarter of 1970 to the first quarter of 1975 for our GDP data). Now, if Y_t is to be stationary, the mean, variance, and autocovariance of Y_{t+m} must be the same as those of Y_t. In short, if a time series is stationary, its mean, variance, and autocovariance (at various lags) remain the same no matter at what point we measure them; that is, they are time invariant. According to Gujarati (2003), such time series will tend to return to its mean (called mean reversion) and fluctuations around this mean (measured by its variance) will have a broadly constant amplitude.

If a time series is not stationary in the sense just defined, it is called a **nonstationary** time series. In other words, a nonstationary time series will have a time-varying mean or a time-varying variance or both.

Why are stationary time series so important? According to Gujarati (2003, 2011), there are at least two reasons. *First*, if a time series is nonstationary, we can study its behavior only for the time period under consideration. Each set of time series data will therefore be for a particular episode. As a result, it is not possible to generalize it to other time periods. Therefore, for the purpose of forecasting or policy analysis, such (nonstationary) time series may be of little practical value. *Second*, if we have two or more nonstationary time series, regression analysis involving such time series may lead to the phenomenon of **spurious** or **nonsense regression** (Gujarati, 2011; Asteriou, 2007).

In addition, a special type of stochastic process (or time series), namely, a **purely random**, or **white noise**, **process**, is also popular in time series econometrics. According to Gujarati (2003), we call a stochastic process purely random if it has zero mean, constant variance σ^2, and is serially uncorrelated. This is similar to what we call the error term, u_t, in the classical normal linear regression model, once discussed in the phenomenon of serial correlation topic. This error term is often denoted as $u_t \sim iid(0,\sigma^2)$.

2.2 Random Walk Process

According to Stock and Watson (2007), time series variables can fail to be stationary in various ways, but two are especially relevant for regression analysis of economic time series data: (1) the series can have persistent, long-run movements, that is, the series can have trends; and, (2) the population regression can be unstable over time, that is, the population regression can have breaks. For the purpose of this lecture, I only focus on the first type of nonstationarity.

A **trend** is a persistent long-term movement of a variable over time. A time series variable fluctuates around its trend. There are two types of trends often seen in time series data: deterministic and stochastic. A **deterministic trend** is a nonrandom function of time (i.e. $Y_t = A + B*Time + u_t$, $Y_t = A + B*Time + C*Time^2 + u_t$, and so on)[7]. For example, the LEX [the logarithm of the dollar/euro daily exchange rate, **TABLE13-1.wf1**, Gujarati (2011)] is a nonstationary seris (Figure 2.1), and its detrended series (i.e. residuals from the regression of

[7] $Y_t = a + bT + e_t \Rightarrow e_t = Y_t - a - bT$ is called the detrended series. If Y_t is nonstationary, while e_t is stationary, Y_t is known as the **trend (stochastic) stationary (TSP)**. Here, the process with a deterministic trend is nonstationary but not a unit root process.

log(EX) on time: e_t = log(EX) – a – b*Time) is still nonstationary (Figure 2.2). This indicates that log(EX) is not a trend stationary series.

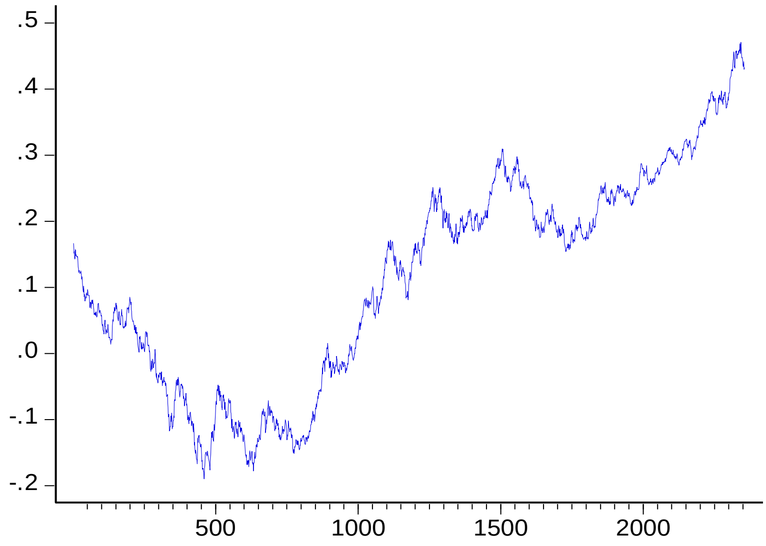


Figure 2.1: Log of the dollar/euro daily exchange rate.

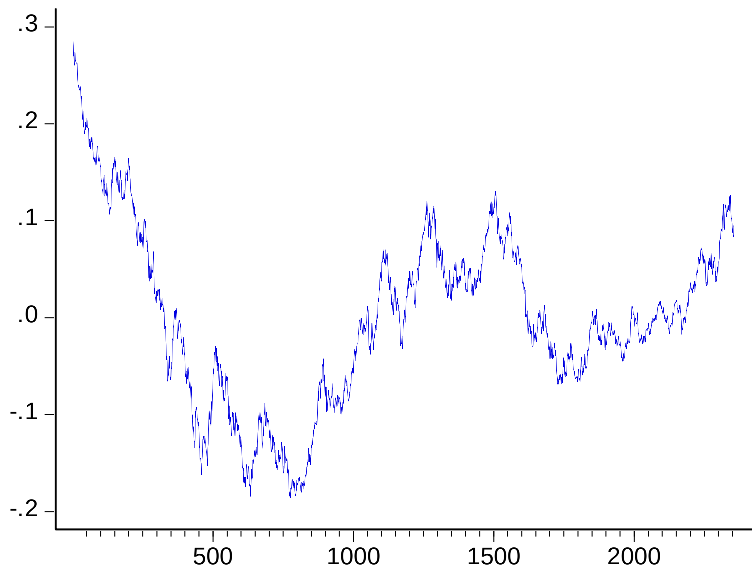


Figure 2.2: Residuals from the regression of LEX on time.

In contrast, a **stochastic trend** is random and varies over time. According to Stock and Watson (2007), it is more appropriate to model economic time series as having stochastic rather than deterministic trends. Therefore, our treatment of trends in economic time series focuses mainly on stochastic rather than deterministic trends, and when we refer to "trends" in time series data, we mean stochastic trends unless we explicitly say otherwise.

The simplest model of a variable with a stochastic trend is the **random walk.** There are two types of random walks: (1) random walk without drift (i.e. no constant or intercept term) and (2) random walk with drift (i.e. a constant term is present).

The **random walk without drift** is defined as follow. Suppose u_t is a white noise error term with mean 0 and variance σ^2. The Y_t is said to be a random walk if:

$$Y_t = Y_{t-1} + u_t \tag{1}$$

The basic idea of a random walk is that the value of the series tomorrow (Y_{t+1}) is its value today (Y_t), plus an unpredictable change (u_{t+1}).

From (1), we can write

$$Y_1 = Y_0 + u_1$$

$$Y_2 = Y_1 + u_2 = Y_0 + u_1 + u_2$$

$$Y_3 = Y_2 + u_3 = Y_0 + u_1 + u_2 + u_3$$

$$Y_4 = Y_3 + u_4 = Y_0 + u_1 + \ldots + u_4$$

…

$$Y_t = Y_{t-1} + u_t = Y_0 + u_1 + \ldots + u_t$$

In general, if the process started at some time 0 with a value Y_0, we have

$$Y_t = Y_0 + \sum u_t \quad (2)$$

therefore,

$$E(Y_t) = E(Y_0 + \sum u_t) = Y_0$$

In like fashion, it can be shown that

$$Var(Y_t) = E(Y_0 + \sum u_t - Y_0)^2 = E(\sum u_t)^2 = t\sigma^2$$

Therefore, the mean of Y_t is equal to its initial or starting value, which is constant, but as *t* increases, its variance increases indefinitely, thus violating a condition of stationarity. In other words, the variance of Y_t depends on *t*, its distribution depends on *t*, that is, it is nonstationary.

Interestingly, if we re-write (1) as

$$(Y_t - Y_{t-1}) = \Delta Y_t = u_t \quad (3)$$

where ΔY_t is the first difference of Y_t. It is easy to show that, while Y_t is nonstationary, its first difference is stationary. And this is very significant when working with time series data. This is widely known as the **difference stationary (stochastic) process (DSP)**.

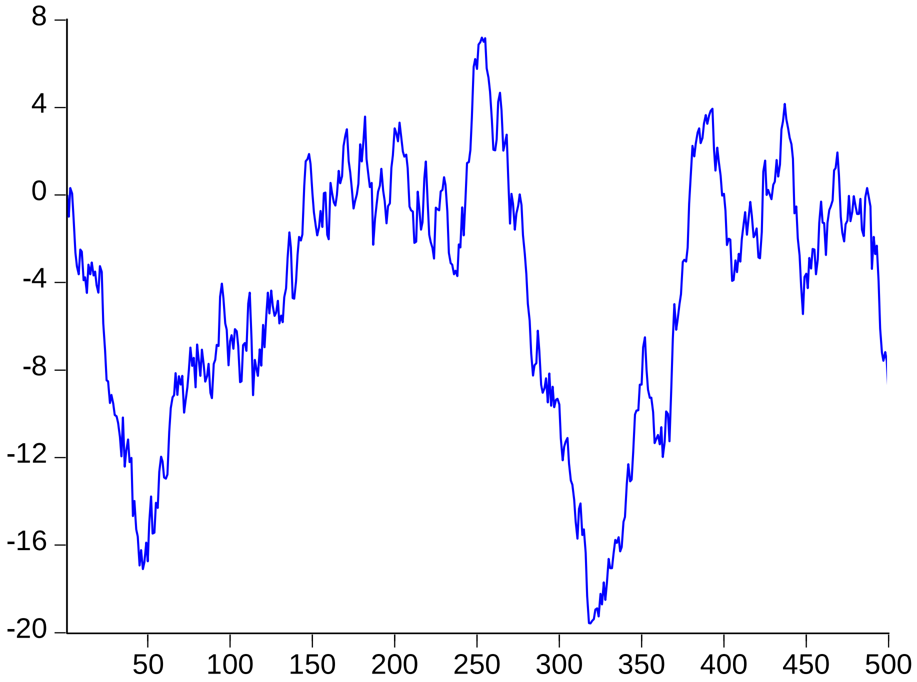


Figure 2.3: A random walk without drift.

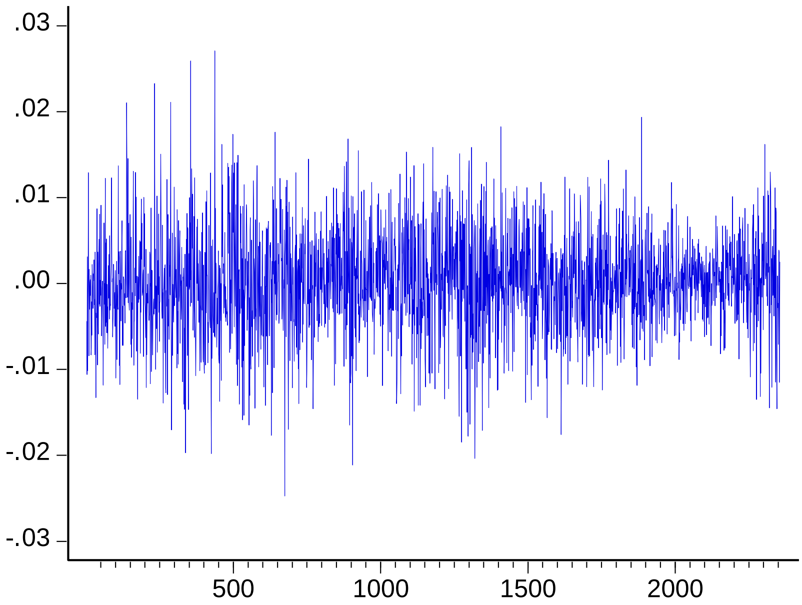


Figure 2.4: First difference of LEX.

The **random walk with drift** can be defined as follow:

$$Y_t = \delta + Y_{t-1} + u_t \qquad (4)$$

where δ is known as the **drift parameter.** The name drift comes from the fact that if we write the preceding equation as:

$$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t \qquad (5)$$

it shows that Y_t drifts upward or downward, depending on δ being positive or negative. We can easily show that, the random walk with drift violates both conditions of stationarity:

$$E(Y_t) = Y_0 + t.\delta$$

$$Var(Y_t) = t\sigma^2$$

In other words, both mean and variance of Y_t depends on *t*, its distribution depends on *t*, that is, it is nonstationary.

Stock and Watson (2007) say that because the variance of a random walk increases without bound, its population autocorrelations are not defined (the first autocovariance and variance are infinite and the ratio of the two is not well defined)[8].

[8] $Corr(Y_t, Y_{t-1}) = \dfrac{Cov(Y_t, Y_{t-1})}{\sqrt{Var(Y_t)Var(Y_{t-1})}} \sim \dfrac{\infty}{\infty}$

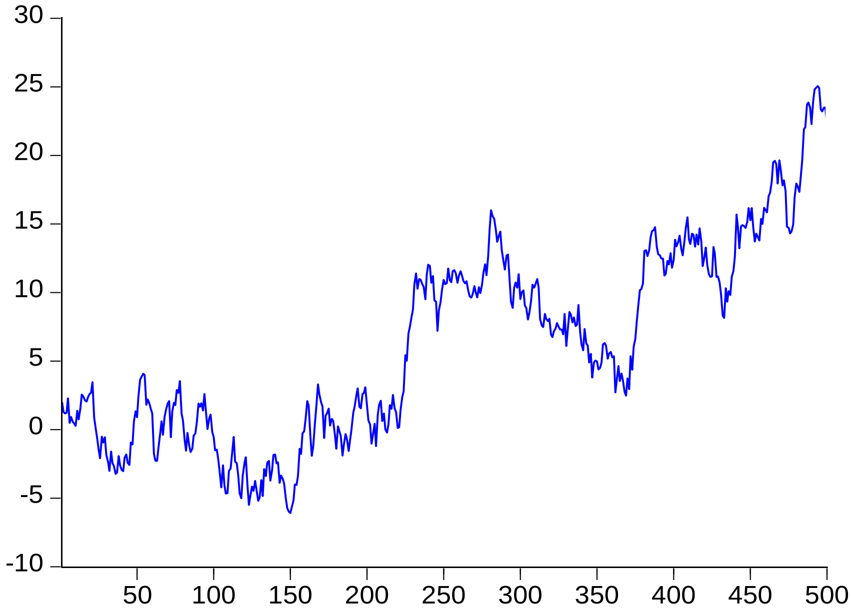


Figure 2.5: A random walk with drift ($Y_t = 2 + Y_{t-1} + u_t$).

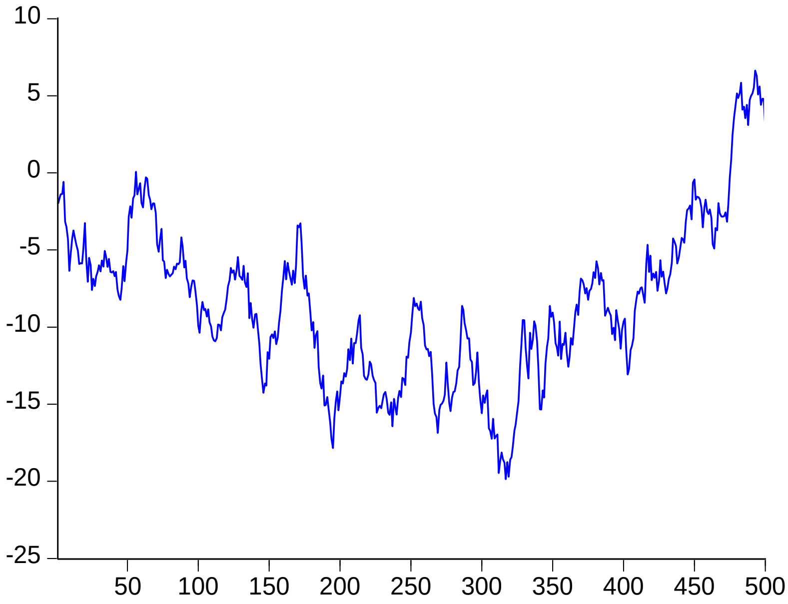


Figure 2.6: Random walk with drift ($Y_t = -2 + Y_{t-1} + u_t$).

2.3 Unit Root Stochastic Process

According to Gujarati (2003), the random walk model is an example of what is known in the literature as a **unit root process**.

Let us write the random walk model (1) as:

$$Y_t = \rho Y_{t-1} + u_t \quad (-1 \le \rho \le 1) \qquad (6)$$

This model resembles the Markov first-order autoregressive model [AR(1)], mentioned in the basic econometrics course, serial correlation topic. If $\rho = 1$, (6) becomes a random walk without drift. If ρ is in fact 1, we face what is known as the **unit root problem**, that is, a situation of nonstationarity. The name unit root is due to the fact that $\rho = 1$. Technically, if $\rho = 1$, we can write (6) as $Y_t - Y_{t-1} = u_t$. Now using the lag operator L so that $LY_t = Y_{t-1}$, $L^2Y_t = Y_{t-2}$, and so on, we can write (6) as $(1-L)Y_t = u_t$. If we set $(1-L) = 0$, we obtain, $L = 1$, hence the name unit root. Thus, the terms nonstationarity, random walk, and unit root can be treated as synonymous.

If, however, $|\rho| \le 1$, that is if the absolute value of ρ is less than one, then it can be shown that the time series Y_t is stationary.

2.4 Illustrative Examples

Consider the AR(1) model as presented in equation (6). Generally, we can have three possible cases:

Case 1: $|\rho| < 1$ and therefore the series Y_t is stationary. A graph of a stationary series for $\rho = 0.67$ is presented in Figure 2.7.

Case 2: $|\rho| > 1$ where in this case the series explodes. A graph of an explosive series for $\rho = 1.26$ is presented in Figure 2.8.

Case 3: $\rho = 1$ where in this case the series contains a unit root and is non-stationary. Graph of stationary series for $\rho = 1$ are presented in Figure 2.9.

In order to reproduce the graphs and the series which are stationary, exploding and nonstationary, we type the following commands in Eviews:

Step 1: Open a new workfile (say, undated type), containing 200 observations.

Step 2: Generate X, Y, Z as the following commands:

```
smpl 1 1
genr X=0
genr Y=0
genr Z=0
smpl 2 200
genr X=0.67*X(-1)+nrnd
genr Y=1.26*Y(-1)+nrnd
genr Z=Z(-1)+nrnd
```

```
smpl 1 200
```

Step 3: Plot X, Y, Z using the line plot type (Figures 2.7, 2.8, and 2.9).

```
plot X
plot Y
plot Z
```

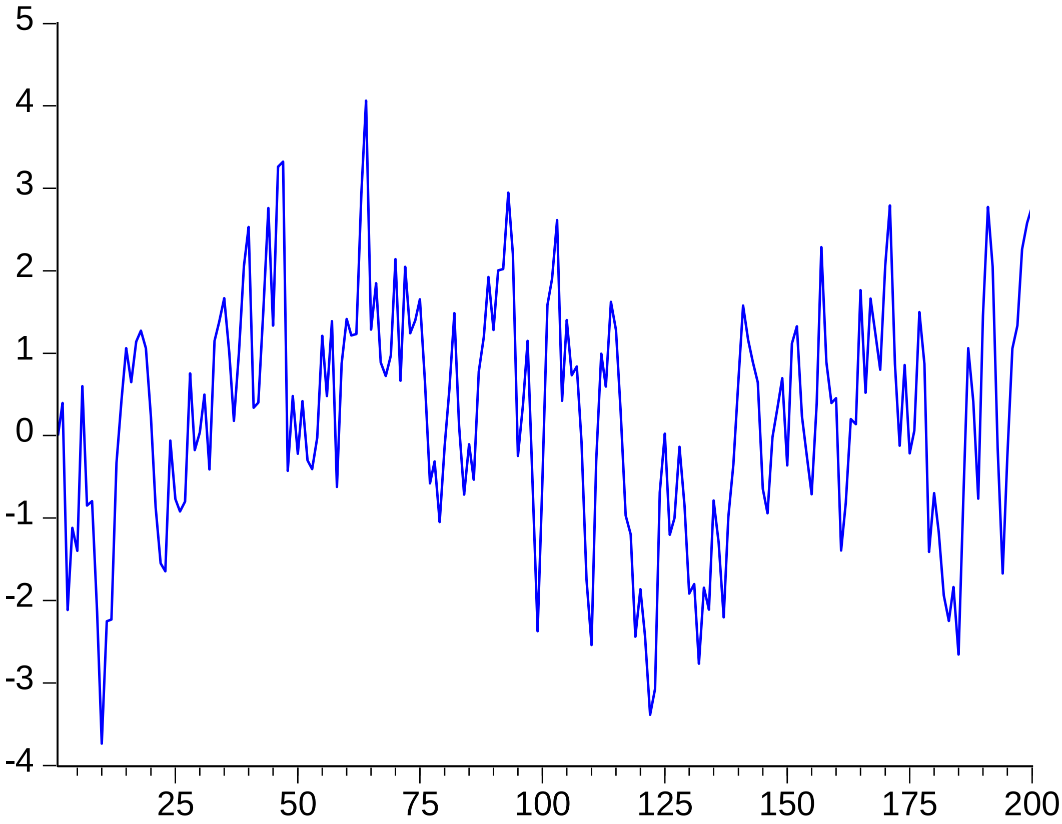


Figure 2.7: A stationary series

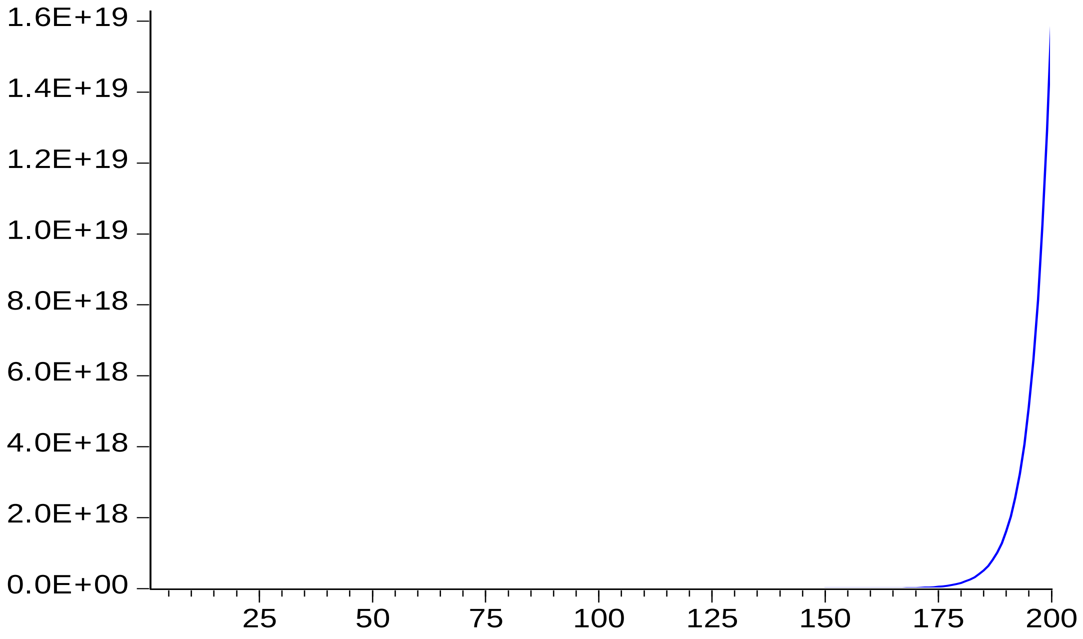


Figure 2.8: An explosive series

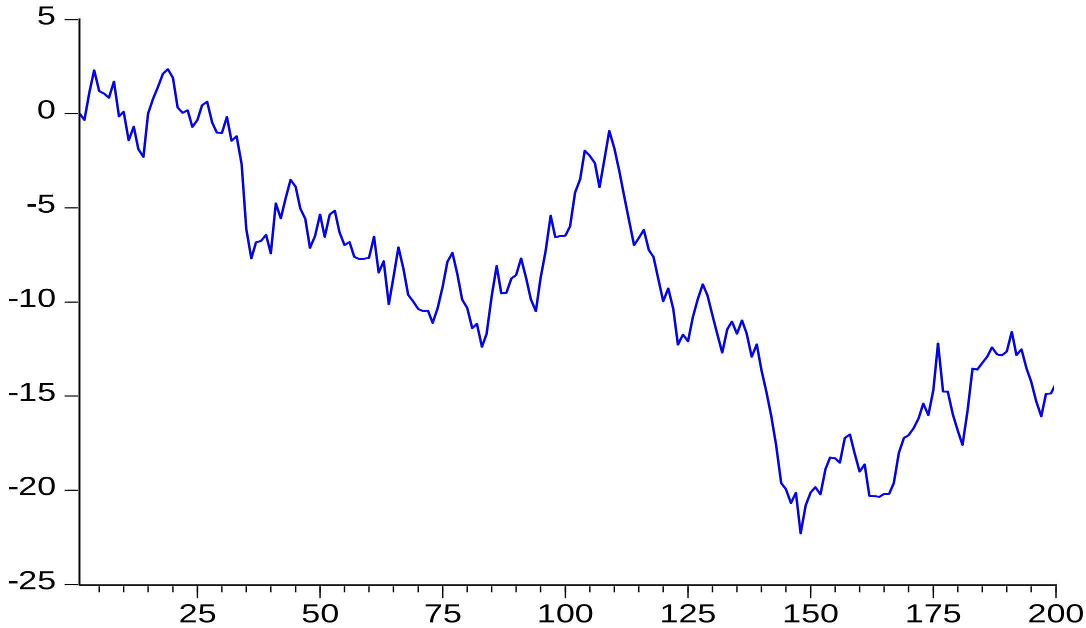


Figure 2.9: A nonstationary series

3. THE UNIT ROOTS AND SPURIOUS REGRESSIONS

3.1 Spurious Regressions

Most macroeconomic time series are trended and therefore in most cases are nonstationary. The problem with nonstationary or trended data is that the standard ordinary least squares (OLS) regression procedures can easily lead to incorrect conclusions. According to Asteriou (2007), it can be shown in these cases that the regression results have very high value of R^2 (sometimes even higher than 0.95) and very high values of *t*-ratios (sometimes even higher than 4), while the variables used in the analysis have no real interrelationships.

Asteriou (2007) states that many economic series typically have an underlying rate of growth, which may or may not be constant, for example GDP, prices or money supply all tend to grow at a regular annual rate. Such series are not stationary as the mean is continually rising however they are also not integrated as no amount of differencing can make them stationary. This gives rise to one of the main reasons for taking the logarithm of data before subjecting it to formal econometric analysis. If we take the logarithm of a series, which exhibits an average growth rate we will turn it into a series which follows a linear trend and which is integrated. This can be easily seen formally. Suppose we have a series X_t, which increases by 10% every period, thus:

$$X_t = 1.1X_{t-1}$$

If we then take the logarithm of this we get

$$\log(X_t) = \log(1.1) + \log(X_{t-1})$$

Now the lagged dependent variable has a unit coefficient and each period it increases by an absolute amount equal to log(1.1), which is of course constant. This series would now be *I*(1).

More formally, consider the model:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (7)$$

where u_t is the error term. The assumptions of classical linear regression model (CLRM) require both Y_t and X_t to have zero and constant variance (i.e., to be stationary). In the presence of nonstationarity, then the results obtained from a regression of this kind are totally **spurious**[9] and these regressions are called spurious regressions.

The intuition behind this is quite simple. Over time, we expect any nonstationary series to wander around (see Figure 3.1), so over any reasonably long sample the series either drift up or down. If we then consider two completely unrelated series which are both nonstationary, we would expect that either they will both go up or down together, or one will go up while the other goes down. If we then performed a regression of one series on the other, we would then find either a significant positive

[9] This was first introduced by Yule (1926), and re-examined by Granger and Newbold (1974) using the Monte Carlo simulations.

relationship if they are going in the same direction or a significant negative one if they are going in opposite directions even though really they are both unrelated. This is the essence of a spurious regression.

It is said that a spurious regression usually has a very high R^2, t statistics that appear to provide significant estimates, but the results may have no economic meaning. This is because the OLS estimates may not be consistent, and therefore all the tests of statistical inference are not valid.

Granger and Newbold (1974) constructed a Monte Carlo analysis generating a large number of Y_t and X_t series containing unit roots following the formulas:

$$Y_t = Y_{t-1} + e_{Yt} \quad (8)$$

$$X_t = X_{t-1} + e_{Xt} \quad (9)$$

where e_{Yt} and e_{Xt} are artificially generated normal random numbers (as the same way performed in section 2.4).

Since Y_t and X_t are independent of each other, any regression between them should give insignificant results. However, when they regressed the various Y_ts to the X_ts as show in equation (8), they surprisingly found that they were unable to reject the null hypothesis of $\beta_2 = 0$ for approximately 75% of their cases. They also found that their regressions had very high R^2s and very low values of DW statistics.

To see the spurious regression problem, we can type the following commands in Eviews (after opening the new workfile, say, undated with 500 observations) to see how many times we can reject the null hypothesis of $\beta_2 = 0$. The commands are:

```
smpl @first @first+1 (or smpl 1 1)
genr Y=0
genr X=0
smpl @first+1 @last (or smpl 2 500)
genr Y=Y(-1)+nrnd
genr X=X(-1)+nrnd
scat(r) Y X
smpl @first @last
ls Y c X
```

An example of a plot of Y against X obtained in this way is shown in Figure 3.1. The estimated equation between these two simulated series is:

Table 3.1: Spurious regression

Dependent Variable: Y
Method: Least Squares
Date: 02/11/10 Time: 18:20
Sample: 1 500
Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-22.02419	0.603511	-36.49346	0.0000
X	-0.505983	0.064798	-7.808589	0.0000

R-squared	0.109082	Mean dependent var	-25.31996
Adjusted R-squared	0.107293	S.D. dependent var	10.20899
S.E. of regression	9.645779	Akaike info criterion	7.374910
Sum squared resid	46334.44	Schwarz criterion	7.391768
Log likelihood	-1841.727	Hannan-Quinn criter.	7.381525
F-statistic	60.97406	Durbin-Watson stat	0.014435
Prob(F-statistic)	0.000000		

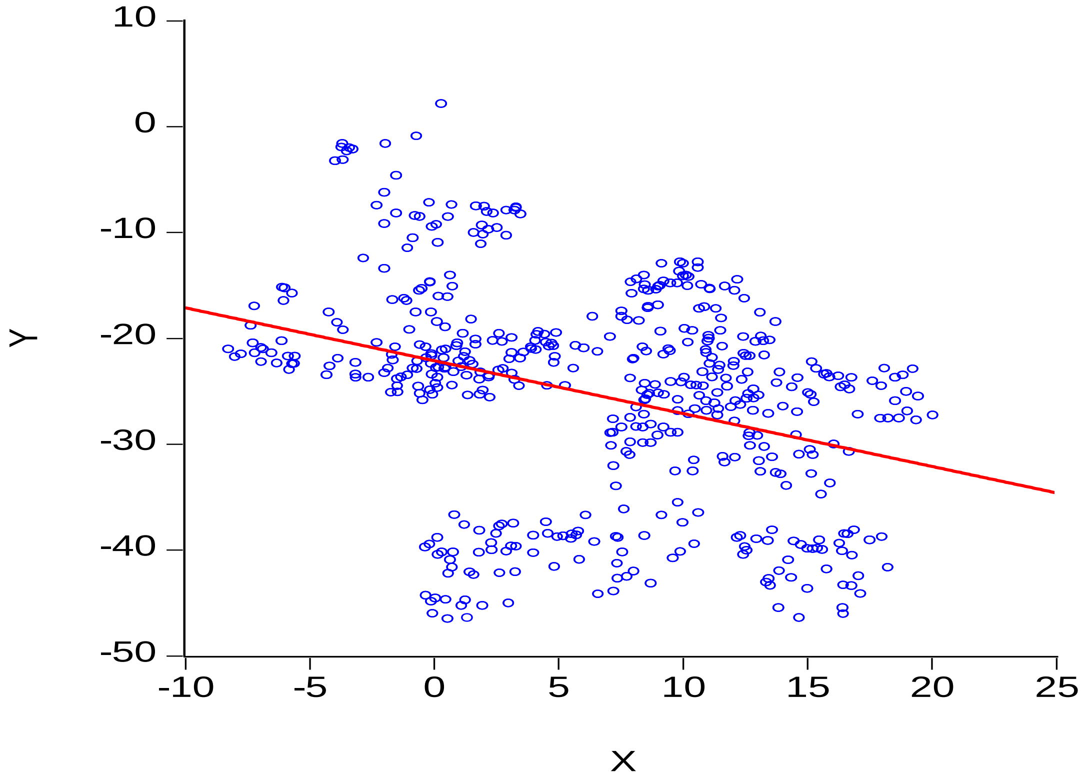


Figure 3.1: Scatter plot of a spurious regression

Granger and Newbold (1974) proposed the following "rule of thumb" for detecting spurious regressions: If $R^2 > DW$ statistic or if $R^2 \approx 1$ then the estimated regression 'must' be spurious.

To understand the problem of spurious regression better, it might be useful to use an example with real economic data. This example was conducted by Asteriou (2007). Consider a regression of the logarithm of real GDP (Y_t) to the logarithm of real money supply (M_t) and a constant. The results obtained from such a regression are the following:

$$Y_t = 0.042 + 0.453M_t;\ R^2 = 0.945;\ DW = 0.221$$
$$(4.743)\ (8.572)$$

Here we see very good *t*-ratios, with coefficients that have the right signs and more or less plausible magnitudes. The coefficient of determination is very high (R^2 = 0.945), but there is a high degree of autocorrelation (*DW* = 0.221). This shows evidence of the possible existence of spurious regression. *In fact, this regression is totally meaningless because the money supply data are for the UK economy and the GDP figures are for the US economy.* Therefore, although there should not be any significant relationship, the regression seems to fit the data very well, and this happens because the variables used in the example are, simply, trended (i.e. nonstationary). So, Asteriou (2007) recommends that econometricians should be very careful when working with trended variables. You can find more such examples in Gujarati (2011, pp.224-226)

3.2 Explaining the Spurious Regression Problem

According to Asteriou (2007), in a slightly more formal way the source of the spurious regression problem comes from the fact that if two variables, X and Y, are both stationary, then in general any linear combination of them will certainly be stationary. One important linear combination of them is of course the equation error, and so if both variables are stationary, the error in the equation will also be stationary and have a well-behaved distribution. However, when the variables become nonstationary, then of course we can not guarantee that the errors will be stationary and in fact as a general rule (although not always) the error itself be nonstationary and when this happens, we violate the basic CLRM assumptions of OLS regression. If the errors were nonstationary, we would expect them to wander around and eventually get large. But OLS regression because it selects the parameters so as to make the sum of the squared errors as small as possible will select any parameter which gives the smallest error and so almost any parameter value can result.

The simplest way to examine the behaviour of u_t is to rewrite (7) as:

$$u_t = Y_t - \beta_1 - \beta_2 X_t \tag{10}$$

or, excluding the constant β_1 (which only affects u_t sequence by rescaling it):

$$u_t = Y_t - \beta_2 X_t \tag{11}$$

If Y_t and X_t are generated by equations (8) and (9), then if we impose the initial conditions $Y_0 = X_0 = 0$ we get that:

$$u_t = Y_0 + e_{Y1} + e_{Y2} + \ldots + e_{Yi} + \beta_2(X_0 + e_{X1} + e_{X2} + \ldots + e_{Xi})$$

or

$$u_t = \sum_{i=1}^{t} e_{Yi} + \beta_2 \sum_{i=1}^{t} e_{Xi} \tag{12}$$

From equation (12), we realize that the variance of the error term will tend to become infinitely large as t increases. Hence, the assumptions of the CLRM are violated, and therefore, any t test, F test or R^2 are unreliable.

In terms of equation (7), there are four different cases to discuss:

Case 1: Both Y_t and X_t are stationary[10], and the CLRM is appropriate with OLS estimates being BLUE.

Case 2: Y_t and X_t are integrated of different orders. In this case, the regression equations are meaningless.

Case 3: Y_t and X_t are integrated of the same order [often $I(1)$]and the u_t sequence contains a stochastic trend. In this case, we have spurious regression and it is often recommended to re-estimate the regression equation in the semi-difference methods (such as the FGLS method: Orcutt-Cochrane procedure, AR(1), and Newey-West standard error). These methods did mention in my lectures about serial correlation and a brief review of basic econometrics.

Case 4: Y_t and X_t are integrated of the same order and the u_t is stationary. In this special case, Y_t and X_t

[10] Based on the statistical tests such as ADF, PP, and KPSS.

are said to be **cointegrated.** The concept of cointegration will be examined in detail later.

4. TESTING FOR UNIT ROOTS

4.1 Graphical Analysis

According to Gujarati (2003), before one pursues formal tests, it is always advisable to plot the time series under study. Such a plot (line graph of the level) and correlogram [of both the level (ACF) and the first difference (ACF and PACF)] gives an initial clue about the likely nature of the time series. Such a intuitive feel is the starting point of formal tests of stationarity (i.e. choose the appropriate test equation).

4.2 Autocorrelation Function and Correlogram

Autocorrelation is the correlation between a variable lagged one or more periods and itself. The correlogram or autocorrelation function is a graph of the autocorrelations for various lags of a time series data. According to Hanke (2005), the **autocorrelation coefficients**[11] for different time lags for a variable can be used to answer the following questions:

(1) Are the data random? (This is usually used for the diagnostic tests of forecasting models).

(2) Do the data have a trend (nonstationary)?

[11] This is not explained in this lecture. You can make references from either Gujarati (2003: 808-813), Hanke (2005: 60-74), or Nguyen Trong Hoai *et al* (2009: Chapter 3, 4, and 8).

(3) Are the data stationary?

(4) Are the data seasonal?

Besides, the correlogram is very useful when selecting the appropriate *p* and *q* in the ARIMA models and ARCH family models[12].

(1) If a series is random, the autocorrelations (i.e. ACF) between Y_t and Y_{t-k} for any lag *k* are close to zero (i.e. the autocorrelation coefficient is statistically insignificant). The successive values of a time series are not related to each other (Figure 4.1). In other words, Y_t and Y_{t-k} are independent.

(2) If a series has a (stochastic) trend, successive observations are highly correlated, and the autocorrelation coefficients are typically significantly different from zero for the first several time lags and then gradually drop toward zero as the number of lags increases. The autocorrelation coefficient for time lag 1 is often very large (close to 1). The autocorrelation coefficient for time lag 2 will also be large. However, it will not be as large as for time lag 1 (Figure 4.2).

(3) If a series is stationary, the autocorrelation coefficients for lag 1 or lag 2 are significantly

[12] See Nguyen Trong Hoai *et al*, 2009 and my lecture about ARIMA models.

different from zero and then suddenly die out as the number of lags increases (Figure 4.3). In other words, Y_t and Y_{t-1}, Y_t and Y_{t-2}, Y_t and Y_{t-3} are weakly correlated; but Y_t and Y_{t-k} [as *k* increases] are completely independent.

(4) If a series has a seasonal pattern, a significant autocorrelation coefficient will occur at the seasonal time lag or multiples of seasonal lag (Figure 4.4). This pattern is not important within this lecture context.

Figure 4.1: Correlogram of a random series

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.019	0.019	0.1737	0.677
		2	-0.032	-0.033	0.6974	0.706
		3	-0.053	-0.052	2.1143	0.549
		4	0.027	0.028	2.4824	0.648
		5	-0.020	-0.025	2.6868	0.748
		6	-0.058	-0.059	4.3992	0.623
		7	0.057	0.061	6.0509	0.534
		8	0.004	-0.005	6.0585	0.641
		9	-0.000	-0.002	6.0586	0.734
		10	-0.028	-0.019	6.4641	0.775
		11	0.008	0.003	6.4932	0.839
		12	-0.038	-0.042	7.2543	0.840

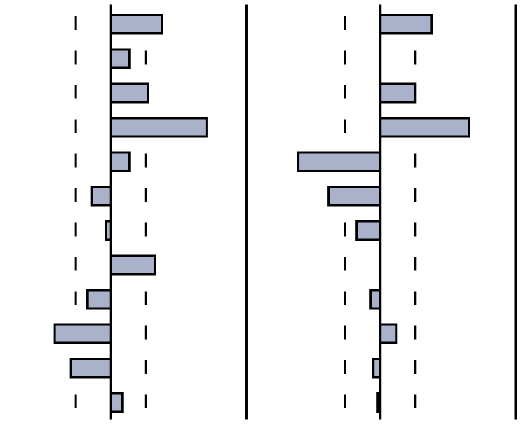
Figure 4.2: Correlogram of a nonstationary series

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.948	0.948	44.115	0.000
		2	0.919	0.201	86.528	0.000
		3	0.865	-0.226	124.96	0.000
		4	0.816	-0.068	160.00	0.000
		5	0.750	-0.164	190.28	0.000
		6	0.699	0.071	217.25	0.000
		7	0.621	-0.215	239.10	0.000
		8	0.543	-0.199	256.23	0.000
		9	0.466	0.033	269.18	0.000
		10	0.381	-0.125	278.06	0.000

Figure 4.3: Correlogram of a stationary series

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.408	0.408	8.5211	0.004
		2	0.155	-0.014	9.7759	0.008
		3	-0.027	-0.102	9.8141	0.020
		4	-0.021	0.034	9.8378	0.043
		5	-0.062	-0.058	10.050	0.074
		6	-0.113	-0.092	10.785	0.095
		7	-0.054	0.042	10.954	0.141
		8	-0.053	-0.043	11.123	0.195
		9	-0.076	-0.072	11.480	0.244
		10	-0.073	-0.013	11.818	0.297

Figure 4.4: Correlogram of a seasonal series



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.393	0.393	8.4970	0.004
		2	0.154	-0.000	9.8280	0.007
		3	0.294	0.276	14.773	0.002
		4	0.744	0.687	47.114	0.000
		5	0.151	-0.633	48.469	0.000
		6	-0.153	-0.400	49.898	0.000
		7	-0.047	-0.179	50.036	0.000
		8	0.347	0.002	57.720	0.000
		9	-0.183	-0.081	59.897	0.000
		10	-0.435	0.120	72.533	0.000
		11	-0.315	-0.058	79.332	0.000
		12	0.091	-0.023	79.916	0.000

The correlogram becomes very useful for time series forecasting and other practical (business) implications. If you conduct academic studies, however, it is necessary to provide more formal statistics such as *t* statistic[13], Box-Pierce *Q* statistic, Ljung-Box (*LB*) statistic, or especially unit root tests.

4.3 Simple Dickey-Fuller Test for Unit Roots

Dickey and Fuller (1979, 1981) devised a procedure to formally test for nonstationarity (hereafter refer to DF test). The key insight of their test is that testing for nonstationarity is equivalent to testing for the existence of a unit root. Thus the obvious test is the following which is based on the following simple AR(1) model:

[13] See Nguyen Trong Hoai *et al*, 2009 and my lecture about ARIMA models to understand the standard error in time series econometrics s.e. = $1/\sqrt{n}$.

$$Y_t = \rho Y_{t-1} + u_t \quad (13)$$

What we need to examine here is $\rho = 1$ (unity and hence 'unit root'). Obviously, the null hypothesis is H_0: $\rho = 1$, and the alternative hypothesis is H_1: $\rho < 1$ (why?).

We obtain a different (more convenient) version of the test by subtracting Y_{t-1} from both sides of (13):

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + u_t$$

$$\Delta Y_t = (\rho - 1)Y_{t-1} + u_t$$

$$\Delta Y_t = \delta Y_{t-1} + u_t \quad (14)$$

where $\delta = (\rho - 1)$. Then, now the null hypothesis is H_0: $\delta = 0$, and the alternative hypothesis is H_1: $\delta < 0$ (why?). In this case, if $\delta = 0$, then Y_t follows a pure random walk (and, of course, Y_t is nonstationary).

Dickey and Fuller (1979) also proposed two alternative regression equations that can be used for testing for the presence of a unit root. The first contains a constant in the random walk process as in the following equation:

$$\Delta Y_t = \alpha + \delta Y_{t-1} + u_t \quad (15)$$

According to Asteriou (2007), this is an extremely important case, because such processes exhibit a definite trend in the series when $\delta = 0$, which is often the case for macroeconomic variables.

The second case is also allow, a non-stochastic time trend in the model, so as to have:

$$\Delta Y_t = \alpha + \gamma T + \delta Y_{t-1} + u_t \qquad (16)$$

The Dickey-Fuller test for stationarity is the simply the normal '*t*' test on the coefficient of the lagged dependent variable Y_{t-1} from one of the three models (14, 15, and 16). This test does not, however, have a conventional '*t*' distribution and so we must use special critical values which were originally calculated by Dickey and Fuller. This is also known as the Dickey-Fuller *tau* statistic (Gujarati, 2003; 2011). However, most modern statistical packages such as Stata and Eviews routinely produce the critical values for Dickey-Fuller tests at 1%, 5%, and 10% significant levels.

MacKinnon (1991,1996) tabulated appropriate critical values for each of the three above models and these are presented in Table 4.1.

Table 4.1: Critical values for DF test

Model	**1%**	**5%**	**10%**
$\Delta Y_t = \delta Y_{t-1} + u_t$	-2.56	-1.94	-1.62
$\Delta Y_t = \alpha + \delta Y_{t-1} + u_t$	-3.43	-2.86	-2.57
$\Delta Y_t = \alpha + \gamma T + \delta Y_{t-1} + u_t$	-3.96	-3.41	-3.13
Standard critical values	*-2.33*	*-1.65*	*-1.28*

Source: Asteriou (2007)

In all cases, the test concerns whether $\delta = 0$. The DF test statistic is the *t* statistic for the lagged dependent variable. If the DF statistical value is smaller in absolute terms than the critical value then we reject the null hypothesis of a unit root and conclude that Y_t is a stationary process.

4.4 Augmented Dickey-Fuller Test for Unit Roots

As the error term is unlikely to be white noise, Dickey and Fuller extended their test procedure suggesting an augmented version of the test (hereafter refer to ADF test) which includes extra lagged terms of the dependent variable in order to eliminate autocorrelation in the test equation.

The lag length[14] on these extra terms is either determined by Akaike Information Criterion (AIC) or Schwarz Bayesian/Information Criterion (SBC, SIC), or more usefully by the lag length necessary to whiten the residuals (i.e. after each case, we check whether the residuals of the ADF regression are autocorrelated or not through *LM* tests and not the *DW* test (why?)).

The three possible forms of the ADF test are given by the following equations:

$$\Delta Y_t = \delta Y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta Y_{t-i} + u_t \qquad (17)$$

[14] This issue will be discussed later in this lecture.

$$\Delta Y_t = \alpha + \delta Y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta Y_{t-i} + u_t \qquad (18)$$

$$\Delta Y_t = \alpha + \gamma T + \delta Y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta Y_{t-i} + u_t \qquad (19)$$

The difference between the three regressions concerns the presence of the deterministic elements α and γT. The critical values for the ADF test are the same as those given in Table 4.1 for the DF test.

According to Asteriou (2007), unless the econometrician knows the actual data-generating process, there is a question concerning whether it is most appropriate to estimate (17), (18), or (19). Daldado, Jenkinson and Sosvilla-Rivero (1990) suggest a procedure which starts from estimation of the most general model given by (19) and then answering a set of questions regarding the appropriateness of each model and moving to the next model. This procedure is illustrated in Figure 4.1. It needs to be stressed here that, although useful, this procedure is not designed to be applied in a mechanical fashion. Plotting the data and observing the graph is sometimes very useful because it can clearly indicate the presence or not of deterministic regressors. However, this procedure is the most sensible way to test for unit roots when the form of the data-generating process is typically unknown.

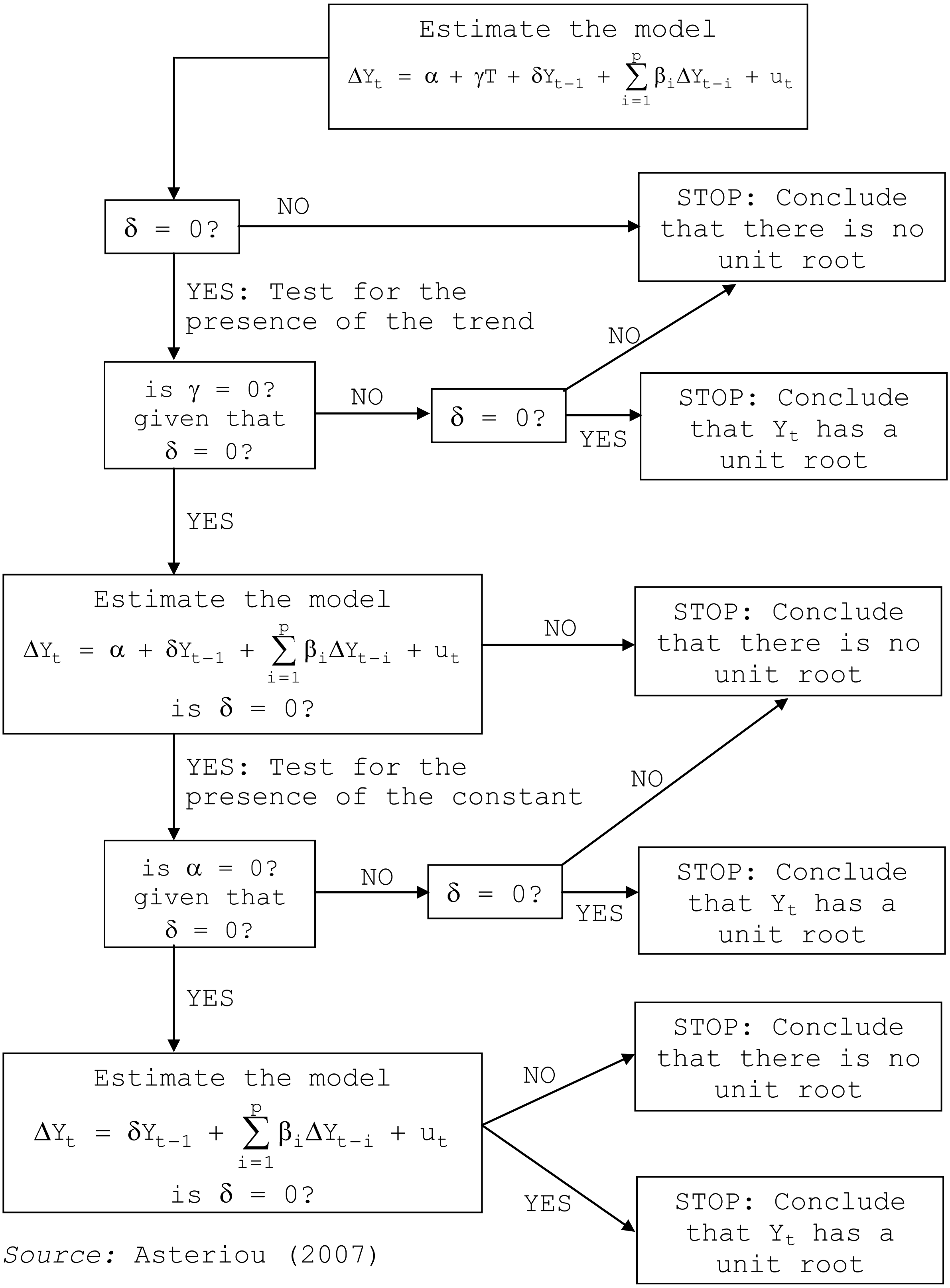
In practical studies, researchers mostly use both the ADF and the Phillips-Perron (PP) tests. Because the distribution theory that supporting the Dickey-Fuller tests is based on the assumption of random error terms [$iid(0,\sigma^2)$], when using the ADF methodology we have to make sure that the error terms are uncorrelated and they really have a constant variance. Phillips and Perron (1988) developed a generalization of the ADF test procedure that allows for fairly mild assumptions concerning the distribution of errors. The regression for the PP test is similar to equation (15).

$$\Delta Y_t = \alpha + \delta Y_{t-1} + e_t \qquad (20)$$

While the ADF test corrects for higher order serial correlation by adding lagged differenced terms on the right-hand side of the test equation, the PP test makes a correction to the *t* statistic of the coefficient δ from the AR(1) regression (semi-difference method) to account for the serial correlation in e_t.

So, the PP statistics are just modifications of the ADF *t* statistics that take into account the less restrictive nature of the error process. The expressions are extremely complex to derive and are beyond the scope of this lecture. Luckily, since most statistical packages have routines available to calculate these statistics, it is good for researcher to test the order of integration of a series performing the PP test as well. The asymptotic distribution of the PP *t* statistic is the same as the ADF *t* statistic and therefore the MacKinnon (1991,1996) critical values are still applicable.

Figure 4.1: Procedure for testing for unit roots



Source: Asteriou (2007)

As with the ADF test, the PP test can be performed with the inclusion of a constant and linear trend, or neither in the test regression.

Dickey-Fuller tests may have low power (H_0 of unit root not rejected, whereas in reality there may be no unit root) when ρ is close to one. This could be the case of trend stationarity (H_0). An alternative test is KPSS (Kwiatkowski-Phillips-Schmidt-Shin). Its test procedure is briefly summarized as:

(1) Regress Y_t on intercept and time trend and obtain OLS residuals e_t.

(2) Calculate partial sums $S_t = \sum_{s=1}^{t} e_s$ for all t.

(3) Calculate the test statistic $KPSS = T^{-2}\sum_{s=1}^{T}\frac{S_t^2}{\hat{\sigma}^2}$, and compare with critical value.

The critical values are routinely produced by statistical packages such as Stata and Eviews. The null hypothesis is rejected if the KPSS test statistic is larger than the selected critical value.

4.5 Performing Unit Root Tests in Eviews

4.5.1 The DF/ADF test

Table 4.2: DF/ADF test procedure

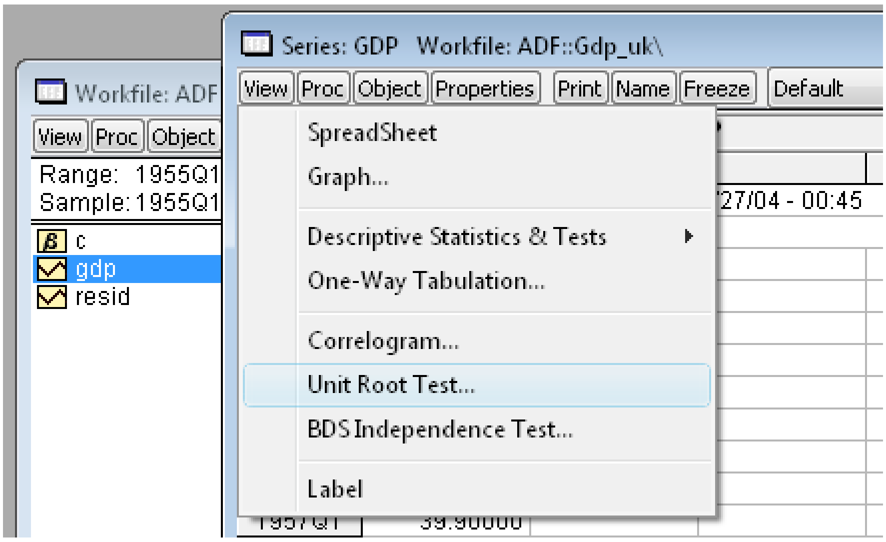
Step 1	Open the file **ADF.wf1** by clicking **File/Open/Workfile** and then choosing the file name from the appropriate path.

Step 2	Let's assume that we want to examine whether the series named GDP contains a unit root. Double click on the series named 'GDP' to open the series window and choose **View/Unit Root Test** … In the unit-root test dialog box that appears, choose the type test (i.e. the **Augmented Dickey-Fuller**) by clicking on it.
Step 3	We then specify whether we want to test for a unit root in the level, first difference, or second difference of the series. We can use this option to determine the number of unit roots in the series. However, we usually start with the level and if we fail to reject the test in levels we continue with testing the first difference and so on.
Step 4	We also have to specify which model of the three ADF models we wish to use (i.e. whether to include a constant, a constant and linear trend, or neither in the test regression). For the model given by equation (17) click on **'none'** in the dialog box; for the model given by equation (18) click on **'intercept'** in the dialog box; and for the model given by equation (19) click on **'intercept and trend'** in the dialog box.
Step 5	Finally, we have to specify the number of lagged dependent variables to be included in the model in order to correct the presence of serial correlation. In practice, we just click the 'automatic selection' on the **'lag length'** dialog box.
Step 6	Having specified these options, click <**OK**> to carry out the test. Eviews reports the test

	statistic together with the estimated test regression.
Step 7	We reject the null hypothesis of a unit root against the one-sided alternative if the ADF statistic is less than (lies to the left of) the critical value, and we conclude that the series is stationary.
Step 8	After running a unit root test, we should examine the estimated test regression reported by Eviews, especially if unsure about the lag structure or deterministic trend in the series. We may want to rerun the test equation with a different selection of right-hand variables (add or delete the constant, trend, or lagged differences) or lag order.

Source: Asteriou (2007)

Figure 4.2: Illustrative steps in Eviews (ADF)



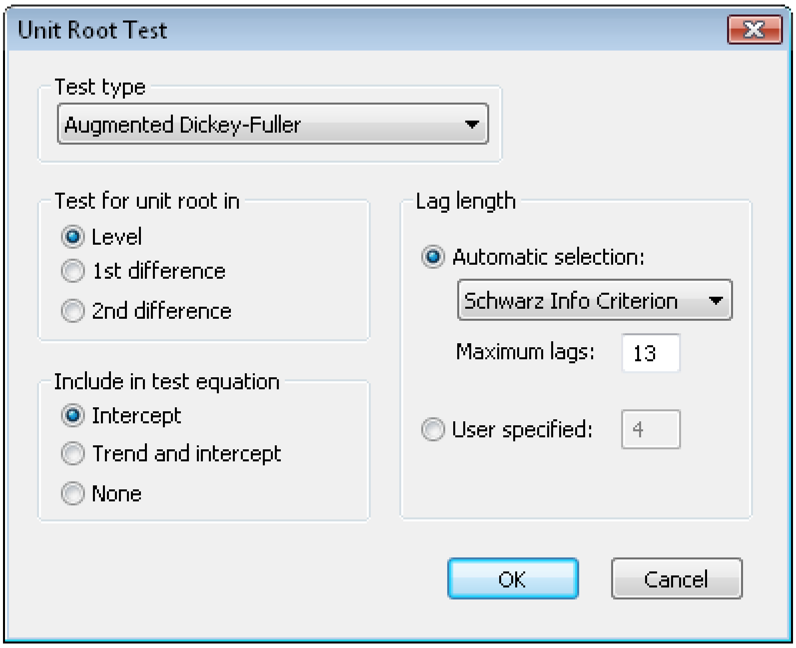


Table 4.3: ADF test of GDP series

Null Hypothesis: GDP has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=13)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		1.838806	0.9998
Test critical values:	1% level	-3.468072	
	5% level	-2.878015	
	10% level	-2.575632	

*MacKinnon (1996) one-sided p-values.

This figure is positive, so the selected model is incorrect (see Gujarati (2003)).

4.5.2 The PP/KPSS test

Table 4.4: PP/KPSS test procedure

Step 1	Open the file **PP.wf1** by clicking **File/Open/Workfile** and then choosing the file name from the appropriate path.
Step 2	Let's assume that we want to examine whether the series named GDP contains a unit root. Double click on the series named 'GDP' to open the series window and choose **View/Unit Root Test** … In the unit-root test dialog box that appears, choose the type test (i.e. the **Phillipd-Perron**/**Kwiatkowski-Phillips-Schmidt-Shin**) by clicking on it.
Step 3	We then specify whether we want to test for a unit root in the level, first difference, or second difference of the series. We can use this option to determine the number of unit roots in the series. However, we usually start with the level and if we fail to reject the test in levels we continue with testing the first difference and so on.
Step 4	We also have to specify which model of the three we need to use (i.e. whether to include a constant, a constant and linear trend, or neither in the test regression). For the random walk model click on **'none'** in the dialog box; for the random with drift model click on **'intercept'** in the dialog box; and for the random walk with drift and with deterministic trend model click on **'intercept and trend'** in the dialog box.

Step 5	Finally, for the **PP/KPSS** test we specify the lag truncation to compute the **Newey-West standard error** consistent estimate of the spectrum at zero frequency.
Step 6	Having specified these options, click <**OK**> to carry out the test. Eviews reports the test statistic together with the estimated test regression.
Step 7	We reject the null hypothesis of a unit root against the one-sided alternative if the ADF statistic is less than (lies to the left of) the critical value, and we conclude that the series is stationary.

Source: Asteriou (2007)

Figure 4.3: Illustrative steps in Eviews (PP)

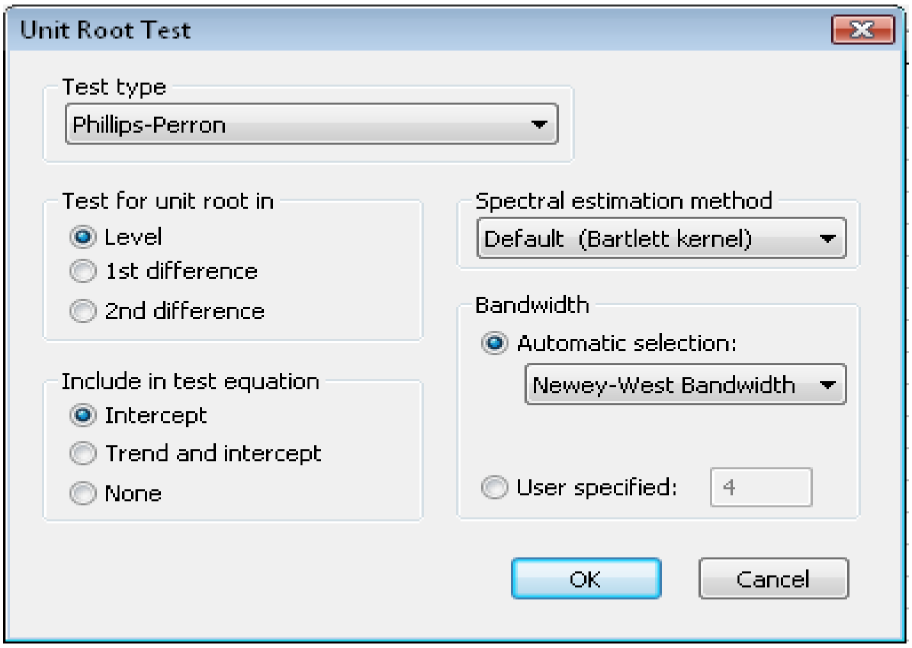


Table 4.5: PP test of GDP series

This figure is positive, so the selected model is incorrect (see Gujarati (2003)).

Null Hypothesis: GDP has a unit root
Exogenous: Constant
Bandwidth: 6 (Newey-West using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		1.389577	0.9990
Test critical values:	1% level	-3.468072	
	5% level	-2.878015	
	10% level	-2.575632	

*MacKinnon (1996) one-sided p-values.

Table 4.6: ADF test of log(EX)

Null Hypothesis: LOG(EX) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=26)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.026489	0.1251
Test critical values:	1% level	-3.961944	
	5% level	-3.411717	
	10% level	-3.127739	

(**TABLE13-1.wf1**, Gujarati, 2011)

Table 4.7: ADF test of Δlog(EX)

Null Hypothesis: D(LOG(EX)) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=26)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-48.47526	0.0000
Test critical values:	1% level	-3.961945	
	5% level	-3.411718	
	10% level	-3.127739	

(**TABLE13-1.wf1**, Gujarati, 2011)

Table 4.8: PP test of Δlog(EX)

Null Hypothesis: D(LOG(EX)) has a unit root			
Exogenous: Constant, Linear Trend			
Bandwidth: 3 (Newey-West using Bartlett kernel)			
		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-48.47612	0.0000
Test critical values:	1% level	-3.961945	
	5% level	-3.411718	
	10% level	-3.127739	
*MacKinnon (1996) one-sided p-values.			
Residual variance (no correction)			3.50E-05
HAC corrected variance (Bartlett kernel)			3.46E-05

(**TABLE13-1.wf1**, Gujarati, 2011)

5. SHORT-RUN AND LONG-RUN RELATIONSHIPS

5.1 Understanding Concepts

In case of bivariate model, you have once known the static or short-run causal relationship between two time series Y_t and X_t, where Y_t is dependent variable and X_t is independent variable. The OLS regression often experiences the serial correlation, and we perform various remedies such as semi-difference methods (Cochrane-Orcutt, Prais-Winsten, AR(1)), first difference method, and Newey-West standard error. By any way, the purpose of your study is just to know the short-run slope or elasticity of Y_t with respect to X_t [$\partial Y_t/\partial Y_t$]. However, the nature of the structural modeling is to discover the dynamic causal relationship between Y_t and X_t. In such model, you must at least distingush between short-run and long-run relationship [slope or elasticity]. To simplify our analysis, we consider the simple autoregressive distributed lag model [ARDL(1,1)] in the following form:

$$Y_t = A_0 + A_1Y_{t-1} + B_0X_t + B_1X_{t-1} + u_t \qquad (21)$$

We can analyse both short-run and long-run effects (slopes or elasticities) as follows:

(1) Short-run or static effect:

$$\frac{\partial Y_t}{\partial X_t} = B_0 \qquad (22)$$

(2) Long-run or dynamic or equilibrium effect:

$$\frac{\partial Y_T}{\partial Xt} = \frac{B_0 + B_1}{1 - A_1} \qquad (23)$$

Proof:

$$\frac{\partial Y_t}{\partial X_t} = B_0$$

$$\frac{\partial Y_{t+1}}{\partial X_t} = A_1\frac{\partial Y_t}{\partial X_t} + B_1 = A_1.B_0 + B_1 \qquad \text{(why?)}$$

$$\frac{\partial Y_{t+2}}{\partial X_t} = A_1\frac{\partial Y_{t+1}}{\partial X_t} = A_1(A_1.B_0 + B_1) \qquad \text{(why?)}$$

$$\frac{\partial Y_{t+3}}{\partial X_t} = A_1\frac{\partial Y_{t+2}}{\partial X_t} = A_1^2(A_1.B_0 + B_1)] \qquad \text{(why?)}$$

…

$$\frac{\partial Y_{t+\infty+1}}{\partial X_t} = A_1\frac{\partial Y_{t+\infty}}{\partial X_t} = A_1^\infty(A_1.B_0 + B_1)] \text{ (why?)}$$

If $|A_1| < 1$, the cummulative effect or long-run slope (S_{1r}) will be the sum of all derivatives:

$$S_{1r} = B_0 + [A_1B_0 + B_1] + A_1[A_1B_0 + B_1] + A_1^2(A_1.B_0 + B_1)] + ... + A_1^{\infty}(A_1.B_0 + B_1)] \quad (24)$$

Multiply both sides of (24) by A_1, we have:

$$A_1S_{1r} = A_1B_0 + A_1[A_1B_0 + B_1] + A_1^2(A_1.B_0 + B_1)] + ... + A_1^{\infty}(A_1.B_0 + B_1)] \quad (25)$$

By substract (25) from (24), we obtain:

$$S_{1r} - A_1S_{1r} = B_0 + B_1$$

$$Slr = \frac{B_0 + B_1}{1 - A_1} = \text{equation (23)}$$

We can also take expectations to derive the long-run relation between Y_t and X_t:

$$E(Y_t) = A_0 + A_1E(Y_{t-1}) + B_0E(X_t) + B_1E(X_{t-1})$$

$$E(Y_t) = A_0 + A_1E(Y_t) + B_0E(X_t) + B_1E(X_t)$$

$$E(Y_t) - A_1E(Y_t) = A_0 + (B_0 + B_1)E(X_t)$$

$$(1-A_1)E(Y_t) = A_0 + (B_0 + B_1)E(X_t)$$

$$=> \quad E(Y_t) = \frac{A_0}{1 - A_1} + \frac{(B_0 + B_1)}{(1 - A_1} E(X_t)$$

$$= \alpha + \beta E(X_t)$$

or simply to write:

$$Y^* = \alpha + \beta X^* \quad (26)$$

Here, $\beta = (B_0+B_1)/(1-A_1)$ is the long-run effect of a **lasting shock** in X_t. And the short-run effect of a **change** in X_t is B_0.

In the same token, we can expand the model ARDL(p,q):

(1) Short-run or static effect:

$$\frac{\partial Y_t}{\partial X_t} = B_0 \tag{27}$$

(2) Long-run or dynamic or equilibrium effect:

$$\frac{\partial Y_T}{\partial Xt} = \frac{B_0 + B_1 + B_2 ... + B_q}{1 - A_1 - A_2 ... - A_p} \tag{28}$$

5.2 ARDL and Error Correction Model (ECM)

By substracting Y_{t-1} both sides of equation (23), and re-arraning, we have:

$$Y_t - Y_{t-1} = A_0 + A_1Y_{t-1} - Y_{t-1} + B_0X_t - B_0X_{t-1} + B_0X_{t-1} + B_1X_{t-1} + u_t$$

$$\Delta Y_t = A_0 - (1-A_1)Y_{t-1} + B_0\Delta X_t + (B_0+B_1)X_{t-1} + u_t$$

$$= B_0\Delta X_t - (1-A_1)\left(Y_{t-1} - \frac{A_0}{(1 - A_1)} - \frac{(B_0 + B_1)}{(1 - A_1)}X_{t-1}\right) + u_t$$

$$= B_0\Delta X_t - (1-A_1)(Y_{t-1} - \alpha - \beta X_{t-1}) + u_t$$

$$= B_0\Delta X_t - \pi(Y_{t-1} - \alpha - \beta X_{t-1}) + u_t \tag{29a}$$

$$= B_0\Delta X_t - \pi ECT_{t-1} + u_t \tag{29b}$$

This is applicable with all ARDL models. Part in brackets of equation (29a) is error-correction term (equilibrium error). Equations (29a or 29b) are widely known as the error correction model (ECM). Therefore, ECM and ARDL are basically the same if the series Y_t and X_t are integrated of the same order [often *I*(1)] and cointegrated.

In this model, Y_t and X_t are assumed to be in long-run equilibrium, i.e. changes in Y_t relate to changes in X_t

according B_1. If Y_{t-1} deviates from the optimal value (i.e. its equilibrium), there is a correction. Speed of adjustment is given by $\pi = (1-A_1)$, which is between > 0 and <1. This will be further discussed later.

6. ENGLE-GRANGER 2-STEP METHOD OF COINTEGRATION

6.1 Cointegration

According to Asteriou (2007), the concept of cointegration was first introduced by Granger (1981) and elaborated further Engle and Granger (1987), Engle and Yoo (1987), Phillips and Ouliaris (1990), Stock and Watson (1988), Phillips (1986 and 1987), and Johansen (1988, 1991, and 1995).

It is known that trended time series can potentially create major problems in empirical econometrics due to spurious regressions. One way of resolving this is to difference the series successively until stationary is achieved and then use the stationary series for regression analysis. According to Asteriou (2007), this solution, however, is not ideal because it not only differences the error process in the regression, but also no longer gives a unique long-run solution.

If two variables are nonstationary, then we can represent the error as a combination of two cumulated error processes. These cumulated error processes are often called stochastic trends and normally we could expect that they would combine to produce another non-stationary process. However, in the special case that two variables, say X_t and Y_t, are really related, then we would expect them to move together and so the two stochastic trends would be very similar to each other and when we combine them together it should be possible to find a combination of them which eliminates the nonstationarity. In this special case, we say that the variables are cointegrated (Asteriou, 2007). Cointegration becomes an overriding requirement for any economic model using nonstationary time series data. If the variables do not co-integrate, we usually face the problems of spurious regression and econometric work becomes almost meaningless. On the other hand, if the stochastic trends do cancel to each other, then we have cointegration.

Suppose that, if there really is a genuine long-run relationship between Y_t and X_t, the although the variables will rise overtime (because they are trended), there will be a common trend that links them together. For an equilibrium, or long-run relationship to exist, what we require, then, is a linear combination of Y_t and X_t that is a stationary variable [an $I(0)$ variable]. A linear combination of Y_t and X_t can be directly taken from estimating the following regression:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (30)$$

And taking the residuals:

$$\hat{u}_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t \qquad (31)$$

If $\hat{u}_t \sim I(0)$, then the variables Y_t and X_t are said to be co-integrated.

6.2 An example of cointegration

TABLE14-1.wf1 gives quarterly data on personal consumption expenditure (PCE) and personal disposible (i.e. after-tax) income (PDI) for the USA for the period 1970-2008 (Gujarati, 2011: pp.226). Both graph (Figure 6.1) and ADF tests (Table 6.1) indicate that these two series are not stationary. They are $I(1)$, that is, they have stochastic trends. In addition, the regression of log(PCE) on log(PDI) seems to be spurious (Table 6.3).

Since both series are trending, let us see what happens if we add a trend variable to the model. The elasticity coefficient is now changed, but the regression is still spurious (Table 6.4). However, after estimating the regression of log(PCE) on log(PDI) and trend, we realize that the obtained residuals is a stationary series [$I(0)$]. This implies that a linear combination (e_t = log(PCE) – b_1 – $b_2$log(PDI) – b_3T) cancels out the stochastic trends in the two variables. Therefore, this regression is, in fact, not spurious. In other words, the variables log(PCE) and log(PDI) are cointegrated.

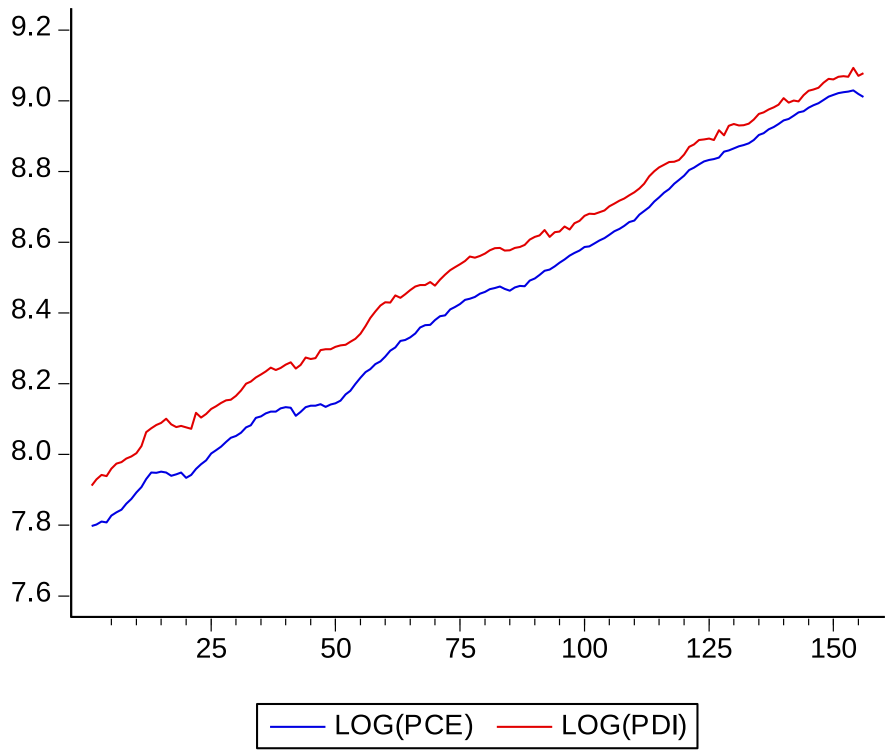


Figure 6.1: Logs of PDI and PCE, USA 1970-2008.

Table 6.1: Unit root tests for log(PCE)

Null Hypothesis: LOG(PCE) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=1)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.038450	0.5754
Test critical values:	1% level	-4.018748	
	5% level	-3.439267	
	10% level	-3.143999	

Table 6.2: Unit root tests for log(PDI)

Null Hypothesis: LOG(PDI) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on AIC, MAXLAG=1)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.774824	0.2089
Test critical values:	1% level	-4.018748	
	5% level	-3.439267	
	10% level	-3.143999	

Table 6.3: OLS regression of log(PCE) on log(PDI)

Dependent Variable: LOG(PCE)
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.842511	0.033717	-24.98752	0.0000
LOG(PDI)	1.086822	0.003949	275.2415	0.0000

R-squared	0.997971	Mean dependent var	8.430699
Adjusted R-squared	0.997958	S.D. dependent var	0.366642
S.E. of regression	0.016567	Akaike info criterion	-5.350028
Sum squared resid	0.042269	Schwarz criterion	-5.310927
Log likelihood	419.3022	Hannan-Quinn criter.	-5.334147
F-statistic	75757.87	Durbin-Watson stat	0.367185

Table 6.4: Regression of log(PCE) on log(PDI) and trend

Dependent Variable: LOG(PCE)
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.672982	0.487339	3.432888	0.0008
LOG(PDI)	0.770240	0.061316	12.56175	0.0000
TIME	0.002366	0.000457	5.172301	0.0000

R-squared	0.998273	Mean dependent var	8.430699
Adjusted R-squared	0.998251	S.D. dependent var	0.366642
S.E. of regression	0.015335	Akaike info criterion	-5.498352
Sum squared resid	0.035978	Schwarz criterion	-5.439701
Log likelihood	431.8714	Hannan-Quinn criter.	-5.474530
F-statistic	44226.63	Durbin-Watson stat	0.261689

Economically speaking, two variables will be cointegrated if they have a long-run, or equilibrium, relationship between them. In the present context, economic theory tells us that there is a strong relationship between consumption expenditure and personal disposible income.

In the language of cointegration theory, the equation $\log(PCE) = B_1 + B_2\log(PDI) + B_3T$ is known as a **cointegrating regression** and the slope parameters B_2 and B_3 are known as **cointegrating parameters**.

6.3 Engle-Granger Tests for Cointegration

For single equation, the simple tests of cointegration are DF and ADF unit root tests on the residuals estimated from the cointegrating regression. These modified tests by the Engle-Granger (EG) and Augmented Engle-Granger (AEG) tests. Notice the difference between the unit root and cointegration tests. Tests for unit roots are performed on single time series, whereas simple cointegration deals with the relationship among a group of variables, each having a unit root (Gujarati, 2011).

6.4 Interpretation of the ECM

According to Asteriou (2007), the concepts of cointegration and the error correction mechanism are very closely related. To understand the ECM, it is better to think first of the ECM as a convenient reparametrization of the general linear autoregressive distributed lag (ARDL) model (see Section 5).

Consider the very simple dynamic ARDL model describing the behaviour of Y_t in terms of X_t as equation (21):

$$Y_t = A_0 + A_1Y_{t-1} + B_0X_t + B_1X_{t-1} + u_t \qquad (21)$$

where $u_t \sim iid(0,\sigma^2)$.

In this model[15], the parameter B_0 denotes the short-run reaction of Y_t after a change in X_t. The long-run effect is given when the model is in equilibrium where:

$$Y^* = \alpha + \beta X^* \qquad (26)$$

Recall that he long-run effect (slope oe elasticity) between Y_t and X_t is captured by $\beta=(B_0+B_1)/(1-A_1)$. It is noted that, we need to make the assumption that $|A_1| < 1$ (why?) in order that the short-run model (21) converges to a long-run solution [equation (26), see Section 6].

The ECM is shown in equation (29a or 29b):

$$\Delta Y_t = B_1\Delta X_t - \pi(Y_{t-1} - \alpha - \beta X_{t-1}) + u_t \qquad (29a)$$

or

$$\Delta Y_t = B_1\Delta X_t - \pi ECT_{t-1} + u_t \qquad (29b)$$

According to Asteriou (2007), what is of importance here is that when the two variables Y_t and X_t are cointegrated,

[15] We can easily expand this model to a more general case for large numbers of lagged terms [ARDL(p,q)].

the ECM incorporates not only short-run but also long-run effects. This is because the long-run equilibrium $[Y_{t-1} - \alpha - \beta X_{t-1}]$ is included in the model together with the short-run dynamics captured by the differenced term. Another important advantage is that all the terms in the ECM model are stationary and the standard OLS estimation is therefore valid. This is because if Y_t and X_t are $I(1)$, then ΔY_t and ΔX_t are $I(0)$, and by definition if Y_t and X_t are cointegrated then their linear combination $[Y_{t-1} - \alpha - \beta X_{t-1}] \sim I(0)$.

A final important point is that the coefficient $\pi = (1-A_1)$ provides us with information about the speed of adjustment in cases of disequilibrium. To understand this better, consider the long-run condition. When equilibrium holds, then $[Y_{t-1} - \alpha - \beta X_{t-1}] = 0$. However, during periods of disequilibrium this term is no longer be zero and measures the distance the system is away from equilibrium. For example, suppose that due to a series of negative shocks in the economy in period t-1. This causes $[Y_{t-1} - \alpha - \beta X_{t-1}]$ to be negative because Y_{t-1} has moved below its long-run equilibrium path. However, since $\pi = (1-A_1)$ is positive (why?), the overall effect is to boost ΔY_t back towards its long-run path as determined by X_t in equation (26). Notice that the speed of this adjustment to equilibrium is dependent upon the magnitude of $\pi = (1-A_1)$.

The coefficient π in equation (29a,b) is the error-correction coefficient and is also called the adjustment coefficient. In fact, π tells us how much of the adjustment to equilibrium takes place each period, or how much of the equilibrium error is corrected each period. According to Asteriou (2007), it can be explained in the following ways:

(1) If $\pi \sim 1$, then nearly 100% of the adjustment takes place within the period[16], or the adjustment is very fast.

(2) If $\pi \sim 0.5$, then about 50% of the adjustment takes place each period.

(3) If $\pi \sim 0$, then there seems to be no adjustment.

According to Asteriou (2007), the ECM is important and popular for many reasons:

(1) Firstly, it is a convenient model measuring the correction from disequilibrium of the previous period which has a very good economic implication.

(2) Secondly, if we have cointegration, ECM models are formulated in terms of first difference, which typically eliminate trends from the variables involved; they resolve the problem of spurious regressions.

[16] Depending on the kind of data used, say, annually, quarterly, or monthly.

(3) A third very important advantage of ECM models is the ease with they can fit into the general-to-specific (or Hendry) approach to econometric modeling, which is a search for the best ECM model that fits the given data sets.

(4) Finally the fourth and most important feature of ECM comes from the fact that the disequilibrium error term is a stationary variable. Because of this, the ECM has important implications: the fact that the two variables are cointegrated implies that there is some automatically adjustment process which prevents the errors in the long-run relationship becoming larger and larger.

6.5 Engle-Granger 2-Step Method of Cointegration

Granger (1981) introduced a remarkable link between nonstationary processes and the concept of long-run equilibrium; this link is the concept of cointegration. Engle and Granger (1987) further formulized this concept by introducing a very simple test for the existence of co-integrating (i.e. long-run equilibrium) relationships.

This approach involves the following steps:

Table 6.5: Engle-Granger 2-Step Method: Step-by-Step

Step 1	Test the variables for their order of integration. The first step is to test each variable to determine its order of integration. The Dickey-

<table>
<tr><td></td><td>Fuller and the augmented Dickey-Fuller tests can be applied in order to infer the number of unit roots in each of the variables. We might face three cases:
a) If both variables are stationary (I(0)), it is not necessary to proceed since standard time series methods apply to stationary variables.
b) If the variables are integrated of different orders, it is possible to conclude that they are not cointegrated.
c) If both variables are integrated of the same order, we proceed with step two.</td></tr>
<tr><td>Step 2</td><td>Estimate the long-run (possible co-integrating) relationship.
If the results of step 1 indicate that both X_t and Y_t are integrated of the same order (usually I(1)) in economics, the next step is to estimate the long-run equilibrium relationship of the form:
$$Y_t = \hat{\alpha} + \hat{\beta}X_t + \hat{u}_t$$
and obtain the residuals of this equation.
If there is no cointegration, the results obtained will be spurious. However, if the variables are cointegrated, OLS regression yields consistent estimators for the co-</td></tr>
</table>

	integrating parameter $\hat{\beta}$.
Step 3	Check for (cointegration) the order of integration of the residuals. In order to determine if the variables are actually cointegrated, denote the estimated residual sequence from the equation by $\hat{u}_t$. Thus, $\hat{u}_t$ is the series of the estimated residuals of the long-run relationship. If these deviations from long-run equilibrium are found to be stationary, the X_t and Y_t are cointegrated.
Step 4	Estimate the error correction model. If the variables are cointegrated, the residuals from the equilibrium regression can be used to estimate the error correction model and to analyse the long-run and short-run effects of the variables as well as to see the adjustment coefficient, which is the coefficient of the lagged residual terms of the long-run relationship identified in step 2. At the end, we always have to check for the accuracy of the model by performing diagnostic tests.

Source: Asteriou (2007)

According to Asteriou (2007), one of the best features of the Engle-Granger 2-step method is that it is both very easy to understand and to implement. However, it also remains some caveats:

(1) One very important issue has to do with the order of the variables. When estimating the long-run relationship, one has to place one variable in the left-hand side and use the others as regressors. The test does not say anything about which of the variables can be used as regressors and why. Consider, for example, the case of just two variables, X_t and Y_t. One can either regress Y_t on X_t (i.e. $Y_t = C + DX_t + u_{1t}$) or choose to reverse the order and regress X_t on Y_t (i.e. $X_t = D + EY_t + u_{2t}$). It can be shown, which asymptotic theory, that as the sample goes to infinity the test for cointegration on the residuals of those two regressions is equivalent (i.e. there is no difference in testing for unit roots in u_{1t} and u_{2t}). However, in practice, in economics we rarely have very big samples and it is therefore possible to find that one regression exhibits cointegration while the other doesn't. This is obviously a very undesirable feature of the EG approach. The problem obviously becomes far more complicated when we have more than two variables to test.

(2) A second problem is that when there are more than two variables there may be more than one integrating relationship, and Engle-Granger 2-step method using residuals from a single relationship

can not treat this possibility. So, the most important problem is that it does not give us the number of co-integrating vectors.

(3) A third and final problem is that it replies on a two-step estimator. The first step is to generate the residual series and the second step is to estimate a regression for this series in order to see if the series is stationary or not. Hence, any error introduced in the first step is carried into the second step.

The Engle-Granger 2-step method in Eviews

The Engle-Granger 2-step method is very easy to perform and does not require any more knowledge regarding the use of Eviews. For the first step, ADF and PP tests on all variables are needed to determine the order of integration of the variables. If the variables (let's say X and Y) are found to be integrated of the same order, then the second step involves estimating the long-run relationship with simple OLS procedure. So the command here is simply:

```
ls X c Y
```

or

```
ls Y c X
```

depending on the relationship of the variables. We then need to obtain the residuals of this relationship which are given by:

```
genr res1=resid
```

The third step (the actual test for cointegration) is a unit root test on the residuals, the command for which is:

```
adf res1
```

for no lags, or

```
adf(4) res1
```

for 4 lags in the augmentation term, and so on.

6.6 Engle-Granger 2-step Method: An Example

Use the example in Section 6.2, we have the following error correction model:

Table 6.6: Error correction model of LPCE and LPDI

Dependent Variable: D(LOG(PCE))
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005530	0.000626	8.831372	0.0000
D(LOG(PDI))	0.306402	0.051588	5.939372	0.0000
RES(-1)	-0.065246	0.033504	-1.947393	0.0533

R-squared	0.189783	Mean dependent var	0.007825
Adjusted R-squared	0.179122	S.D. dependent var	0.006764
S.E. of regression	0.006128	Akaike info criterion	-7.332580
Sum squared resid	0.005709	Schwarz criterion	-7.273675
Log likelihood	571.2749	Hannan-Quinn criter.	-7.308654
F-statistic	17.80204	Durbin-Watson stat	1.707358

All coeeficients in the table are individually statistically significant at 6% or lower level. The coefficient of about 0.31 shows that a 1% increase in $\log(PDI_t/PDI_{t-1})$ will lead on average to a 0.31% increase in $\ln(PCE_t/PCE_{t-1})$. This is the short-run consumption-income elasticity. The long-run value is given by the cointegrating regression (Table 6.4), which is 0.77.

The coefficient of the error-correction term of about -0.06 suggests that about 6% of the discrepancy between long-term and short-term PCE is corrected within a quarter (quarterly data), suggesting a slow rate of adjustment to equilibrium.

7. VECTOR AUTOREGRESSIVE MODELS

According to Asteriou (2007), it is quite common in economics to have models where some variables are not only explanatory variables for a given dependent variable, but they are also explained by the variables that they are used to determine. In those cases, we have models of simultaneous equations, in which it is necessary to clearly identify which are the endogenous and which are the exogenous or predetermined variables. The decision regarding such a differentiation among variables was heavily criticized by Sims[17] (1980).

According to Sims (1980), if there is simultaneity among a number of variables, then all these variables should be treated in the same way. In other words, these should be no distinction between endogenous and exogenous

[17] Nobel prize in economics 2012.

variables. Therefore, once this distinction is abandoned, all variables are treated as endogenous. This means that in its general reduced form, each equation has the same set of regressors which leads to the development of the VAR models.

VAR is defined as a system of ARDL equations describing dynamic evolution of a set of variables from their common history (here **vector** implies multiple variables involved. The VAR model is defined as follow. Suppose we have two series, in which Y_t is affected by not only its past (or lagged) values but current[18] and lagged values of X_t, and simultaneously, X_t is affected by not only its lagged values but current and lagged values of Y_t. This simple bivariate VAR model is given by:

$$Y_t = A_1 + B_1X_t + \sum_{j=1}^{p} C_jY_{t-j} + \sum_{j=1}^{p} D_jX_{t-j} + u_{1t} \qquad (32)$$

$$X_t = A_2 + B_2Y_t + \sum_{j=1}^{p} E_jY_{t-j} + \sum_{j=1}^{p} F_jX_{t-j} + u_{2t} \qquad (33)$$

where we assume the u_{1t} and u_{2t} are uncorrelated white-noise error terms, called **impulses** or **innovations** or **shocks** in the language of VAR (Gujarati, 2011, pp.266). Note that these equations are not reduced-form equations

[18] Gujarati (2011, pp.266) said that [from the point of view of forecasting] each equation in VAR contains only its own lagged values and the lagged values of the other variables in the system. Similarly, Wooldridge (2003, pp.620-621) said that whether the contemporaneous (current) value is included or not depends partly on the purpose of the equation. In forecasting, it is rarely included.

since Y_t has a contemporaneous impact on X_t, and X_t has a contemporaneous impact on Y_t.

The bivariate VAR often has the following features (according to Gujarati, 2011, pp.266):

(1) Although the number of lagged values of each variable can be different, in most cases we use the same number of lagged terms in each equation.

(2) The bivariate VAR system given above is known as a VAR(p) model, because we have p lagged values of each variable on the right-hand side. If we have only one lagged value of each variable on the right-hand side, it would be a VAR(1) model; if two lagged terms, it would be a VAR(2) model; and so on.

(3) Although we are dealing with only two variables, the VAR system can be extended to several variables.

(4) In the two-variable system, there can be at most one cointegrating, or equilibrium, relationship between them. If we have a three-variable VAR system, there can be at most two cointegrating relationships between the three variables.

Note that all variables have to be of the same order of integration. The following cases are distinct:

(1) All the variables are $I(0)$ (stationary): one is in the standard case, i.e. a VAR in level.

(2) All the variables are I(d) (non-stationary) with d > 0 (usually d = 1):

- If there are nonstationary [I(1)] variables, we estimate a VAR using first differences of variables [that are I(0)] to remove common trends.
- The variables are cointegrated: the error correction term has to be included in the VAR model. The model becomes a vector error correction model (**VECM**) which can be seen as a **restricted VAR.** Why (if possible) a VECM instead of a VAR on differenced variables? VECM gives long-run structural relations plus information on adjustment, which provides better insight in economic processes.
- The variables are not cointegrated: the variables have first to be differenced d times and one has a VAR in difference.

According to Asteriou (2007), the VAR model has some good characteristics. First, it is very simple because we do not have to worry about which variables are endogenous or exogenous. Second, estimation is very simple as well, in the sense that each equation can be estimated with the usual OLS method separately. Third, forecasts obtained from VAR models are in most cases better than those obtained from the far more complex simultaneous equation models (see Mahmoud, 1984; McNees, 1986). Besides forecasting purposes, VAR models also provide framework for causality tests, which will be presented shortly.

However, on the other hand the VAR models have faced severe criticism on various different points. According to Asteriou (2007), the VAR models have been criticised by the following aspects. First, they are a-theoretic since they are not based on any economic theory. Since initially there are no restrictions on any of the parameters under estimation, in effect 'everything causes everything'. However, statistical inference is often used in the estimated models so that some coefficients that appear to be insignificant can be dropped, in order to lead models that might have an underlying consistent theory. Such inference is normally carried out using what are called causality tests. Second, they are criticised due to the loss of degrees of freedom. Thus, if the sample size is not sufficiently large, estimating that large a number of parameters, say, a three-variable VAR model with 12 lags for each, will consume many degrees of freedom, creating problems in estimation. Third, the obtained coefficients of the VAR models are difficult to interpret since they totally lack any theoretical background.

8. VECM AND COINTEGRATION

8.1 Rank of Cointegrating Matrix

In this section, we extend the single-equation error correction model to a multivariate one. Let's assume that we have three variables, Y_t, X_t and W_t, which can all be endogenous, i.e. we have that (using matrix notation for $Z_t = [Y_t, X_t, W_t]$)

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + \dots + A_p Z_{t-p} + u_t \qquad (34)$$

A VAR(p) can be reformulated in a vector error correction model as follows:

$$\Delta Z_t = \Gamma_1 \Delta Z_{t-1} + \Gamma_2 \Delta Z_{t-2} + \dots + \Gamma_{p-1} \Delta Z_{t-p+1} + \Pi Z_{t-1} + u_t \qquad (35)$$

where the matrix Π contains information regarding the long-run relationships. We can decompose $\Pi = \pi\beta'$ where π will include the speed of adjustment to equilibrium coefficients, while β' will be the long-run matrix of coefficients.

Therefore, the $\beta' Z_{t-1}$ term is equivalent to the error correction term $[Y_{t-1} - \alpha - \beta X_{t-1}]$ in the single-equation case, except that now $\beta' Z_{t-1}$ contains up to $(p - 1)$ vectors in a multivariate framework.

For simplicity, we assume that $p = 2$, so that we have only two lagged terms, and the model is then the following:

$$\begin{pmatrix} \Delta Y_t \\ \Delta X_t \\ \Delta W_t \end{pmatrix} = \Gamma_1 \begin{pmatrix} \Delta Y_{t-1} \\ \Delta X_{t-1} \\ \Delta W_{t-1} \end{pmatrix} + \Pi \begin{pmatrix} Y_{t-1} \\ X_{t-1} \\ W_{t-1} \end{pmatrix} + u_t \qquad (33)$$

or

$$\begin{pmatrix}\Delta Y_t\\ \Delta X_t\\ \Delta W_t\end{pmatrix} = \Gamma_1\begin{pmatrix}\Delta Y_{t-1}\\ \Delta X_{t-1}\\ \Delta W_{t-1}\end{pmatrix} + \begin{pmatrix}\pi_{11} & \pi_{12}\\ \pi_{21} & \pi_{22}\\ \pi_{31} & \pi_{32}\end{pmatrix}\begin{pmatrix}\beta_{11} & \beta_{21} & \beta_{31}\\ \beta_{12} & \beta_{22} & \beta_{32}\end{pmatrix}\begin{pmatrix}Y_{t-1}\\ X_{t-1}\\ W_{t-1}\end{pmatrix} + u_t \quad (34)$$

Let us now analyse only the error correction part of the first equation (i.e. for ΔY_t on the left-hand side) which gives:

$$\Pi_1 Z_{t-1} = ([\pi_{11}\beta_{11} + \pi_{12}\beta_{12}][\ \pi_{11}\beta_{21} + \pi_{12}\beta_{22}]$$

$$[\pi_{11}\beta_{31} + \pi_{12}\beta_{32}])\begin{pmatrix}Y_{t-1}\\ X_{t-1}\\ W_{t-1}\end{pmatrix} \quad (35)$$

Equation (35) can be rewritten as:

$$\Pi_1 Z_{t-1} = \pi_{11}(\beta_{11}Y_{t-1} + \beta_{21}X_{t-1} + \beta_{31}W_{t-1}) +$$

$$\pi_{12}(\beta_{12}Y_{t-1} + \beta_{22}X_{t-1} + \beta_{32}W_{t-1}) \quad (36)$$

which shows clearly the two co-integrating vectors with their respective speed of adjustment terms π_{11} and π_{12}.

What are advantages of the multiple equation approach?

(1) From the multiple equation approach we can obtain estimates for both co-integrating vectors (36), while the simple equation we have only a linear combination of the two long-run relationships.

(2) Even if there is only one co-integrating relationship (for example the first only) rather than two, with the multiple equation approach we can calculate all three differing speeds of adjustment coefficients $(\pi_{11}\ \ \pi_{21}\ \ \pi_{31})'$.

(3) Only when $\pi_{21} = \pi_{31} = 0$, and only one co-integrating relationship exists, can we then say that the multiple equation method is the same (reduces to) as the single equation approach, and therefore, there is no loss from not modelling the determinants of ΔX_t and ΔW_t. Here, it is good to mention as well that when $\pi_{21} = \pi_{31} = 0$, is equivalent to X_t and W_t being weakly exogenous.

Suppose that we have k variables in a VECM, the $k \times k$ matrix Π contains the error correction terms (linear combinations of k variables in Z_{t-1} that are $I(0)$.

Table 8.1: Rank of matrix Π and its implications

Rank of Π	Implications
r = 0	There is no cointegration. No stable long-run relations between variables. VECM is not possible (only VAR in first differences).
0 < r < k	There are r cointegrating vectors. These vectors describe the long-run relationships between variables. VECM is o.k.
r = k	All variables are already stationary. No need to estimate the model as VECM. VAR on untransformed data is o.k.

8.2 Johansen's Test for Cointegration

8.2.1 Test Procedure

According to Asteriou (2007), if we have more than two variables in the model, then there is a possibility of having more than one co-integrating vector. By this we mean that the variables in the model might form several equilibrium relationships. In general, for *k* number of variables, we can have only up to *k-1* co-integrating vectors. To find out how many cointegrating relationships exist among *k* variables requires the use of Johansen's methodology[19].

This method involves the following steps:

Table 8.2: Johansen's approach

Step 1	Testing the order of integration of all variables.
Step 2	Setting the appropriate lag length of the model. Setting the value of the lag length is affected by the omission of variables that might affect only the short-run behaviour of the model. This is due to the fact that omitted variables instantly become part of the error term. Therefore, very careful inspection of the data and the functional relationship is necessary before proceeding with estimation in order to decide whether to include additional variables. It is white common to use dummy variables to take into account short-run 'shocks' to the system, such as political events that had important effects on macroeconomic conditions. The most common procedure in choosing the

[19] Similar to EG approach, the Johansen's approach also requires all variables in the system are integrated of the same order 1 [*I*(1)].

	optimal lag length is to estimate a VAR model including all variables in levels (non-differenced). This VAR model should be estimated for a large number of lags, then reducing down by re-estimating the model for one lag less until we reach zero lag. In each of these models, we inspect the values of the AIC and the SBC criteria, as well as the diagnostics concerning autocorrelation, heteroskedasticity, possible ARCH effects and normality of the residuals. In general, the model that minimizes AIC and SBC is selected as the one with the optimal lag length. This model should also pass all the diagnostic checks.
Step 3	Choosing the appropriate model regarding the deterministic components in the multivariate system. Another important aspect in the formulation of the dynamic model is whether an intercept and/or trend should enter either the short-run or the long-run model, or both models. The general case of the VECM including all the various options that can possibly happen is given by the following equation: $\Delta Z_t = \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_{k-1} \Delta Z_{t-p+1} + \pi \begin{pmatrix} \beta \\ \mu_1 \\ \delta_1 \end{pmatrix} (Z_{t-1}\ 1\ \ t)\, Z_{t-1}$ $\quad + \mu_2 + \delta_2 t + u_t \qquad (37)$ In general five distinct models can be considered. Although the first and the fifth

	model are not that realistic, we present all of them for reasons of complementarity. **Model 1:** No intercept or trend in CE (co-integrating equation) or VAR ($\delta_1 = \delta_2 = \mu_1 = \mu_2 = 0$). **Model 2:** Intercept (no trend) in CE, no intercept or trend in VAR ($\delta_1 = \delta_2 = \mu_2 = 0$). **Model 3:** Intercept in CE and VAR, no trend in CE and VAR ($\delta_1 = \delta_2 = 0$). **Model 4:** Intercept in CE and VAR, linear trend in CE, no trend in VAR ($\delta_2 = 0$). **Model 5:** Intercept and quadratic trend in CE, intercept and linear trend in VAR.
Step 4	Determining the rank of Π or the number of co-integrating vectors. There are two methods for determining the number of co-integrating relations, and both involve estimation of matrix Π. (1) One method tests the null hypothesis, that $Rank(\Pi) = r$ against the hypothesis that the rank is $r+1$. So, the null in this case is that there are co-integrating vectors and that we have up to r co-integrating relationships, with the alternative suggesting there is ($r+1$) vectors. The test statistics are based on the characteristic roots (also called

eigenvalues) obtained from the estimation procedure. The test consists of ordering the largest eigenvalues in descending order and considering whether they are significantly different from zero. To understand the test procedure, suppose we obtained *n* characteristic roots denoted by $\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_n$. If the variables under examination are not cointegrated, the rank of Π is zero and all the characteristic roots will equal zero. Therefore, $(1 - \hat{\lambda}_i)$ will be equal to 1 and since ln(1) = 0. To test how many of the numbers of the characteristic roots are significantly different from zero, this test uses the following statistic:

$$\lambda_{max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1}) \qquad (38)$$

As we said before, the test statistic is based on the *maximum eigenvalue* and because of that is called the *maximal eigenvalue statistic* (denoted by λ_{max}).

(2) The second method is based on a likelihood ratio test about the trace of the matrix (and because of that it is called the *trace statistic*). The trace statistic considers whether the trace is increased by adding more eigenvalues beyond the r^{th} eigenvalue. The null hypothesis in this case is that the number of co-integrating vectors is less

	than or equal to *r*. From the previous analysis, it is clear that when all $\hat{\lambda}_i$ = 0, the trace statistic is equal to zero as well. This statistic is calculated by: $$\lambda_{trace}(r) = -T\sum_{i=r+1}^{n}\ln(1-\hat{\lambda}_{r+1}) \quad (39)$$ Critical values for both statistics are provided by Johansen and Juselius (1990). These critical values are directly provided from Eviews after conducting a cointegration test.

Source: Asteriou (2007)

8.2.2 A Numerical Example

Remind that we reject the null hypothesis that r, the number of cointegrating vectors, is less than *k* if the test statistic is greater than the critical values specified.

Table 8.3: Trace test

H_0	H_1	Statistic	95% Critical	Decision
r = 0	r = 1	62.18	47.21	Reject H_0
r ≤ 1	r = 2	19.55	29.68	Accept H_0
r ≤ 2	r = 3	8.62	15.41	Accept H_0
r ≤ 3	r = 4	2.41	3.76	Accept H_0

Phùng Thanh Bình
ptbinh@ueh.edu.vn

We conclude that this data exhibits one cointegrating vector.

8.2.3 The Johansen approach in Eviews

Eviews has a specific command for testing for cointegration using Johansen approach under group statistics. Consider the file **JOHANSEN.wf1,** which has quarterly data for three macroeconomic variables: X, Y, and Z.

Step 1: Determine the order of integration for the variables.

To do this, we apply the unit-root tests on all three variables. We apply the Doldado, Jenkinson and Sosvilla-Rivero (1990) procedure for choosing the appropriate model and we determine the number of lags according to the SBC criterion.

Table 8.4: Integration of the variables at level

Null Hypothesis: X has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic based on SIC, MAXLAG=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.438259	0.1344
Test critical values:	1% level	-3.505595	
	5% level	-2.894332	
	10% level	-2.584325	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: Y has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.132096	0.5209
Test critical values:	1% level	-4.063233	
	5% level	-3.460516	
	10% level	-3.156439	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: Z has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.166369	0.5021
Test critical values:	1% level	-4.062040	
	5% level	-3.459950	
	10% level	-3.156109	

*MacKinnon (1996) one-sided p-values.

Table 8.5: Integration of the variables at difference

Null Hypothesis: D(X) has a unit root
Exogenous: Constant
Lag Length: 2 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.444517	0.0005
Test critical values:	1% level	-3.506484	
	5% level	-2.894716	
	10% level	-2.584529	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(Y) has a unit root
Exogenous: Constant
Lag Length: 4 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.317332	0.0008
Test critical values:	1% level	-3.508326	
	5% level	-2.895512	
	10% level	-2.584952	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(Z) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.911884	0.0007
Test critical values:	1% level	-4.065702	
	5% level	-3.461686	
	10% level	-3.157121	

*MacKinnon (1996) one-sided p-values.

Step 2: Determine the optimal lag length

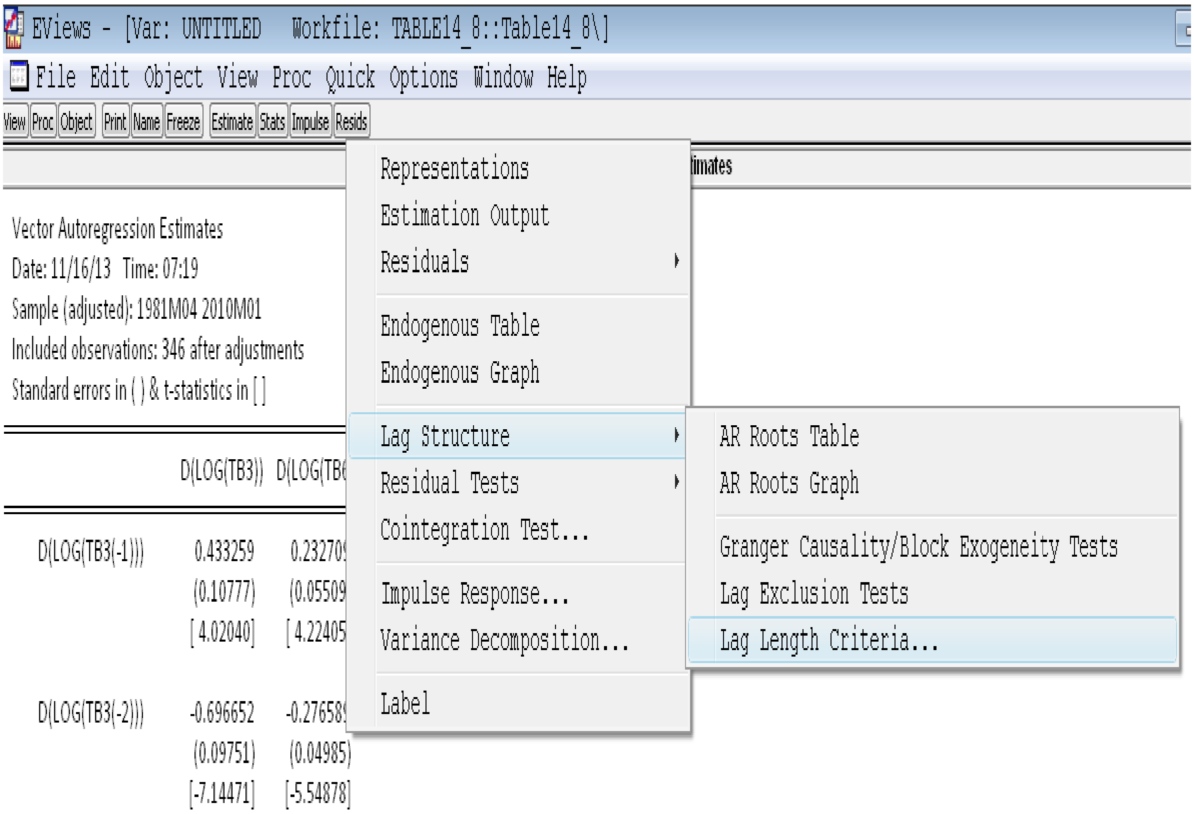
Unfortunately, Eviews does not allow us to automatically detect the lag length (in cointegrating test equation), so we need to estimate the model for a large number of lags and then reduce down to check for the optimal value of AIC and SBC. A rule of thumb is to choose the lag length according to the data interval (year, quarter, month). By doing this, we found that the optimal lag length is 4 lags (quarterly data).

In Eviews, we can automatically determine the lag length as the following steps.

(1) **Estimate unrestricted VAR** (using $I(0)$ variables in the VAR model, i.e. first differences) with the defaut lag length (routinely 2).

(2) At the 'VAR estimates', select **View/Lag Structure/Lag Length Criteria**

Figure 8.1: Lag length criteria



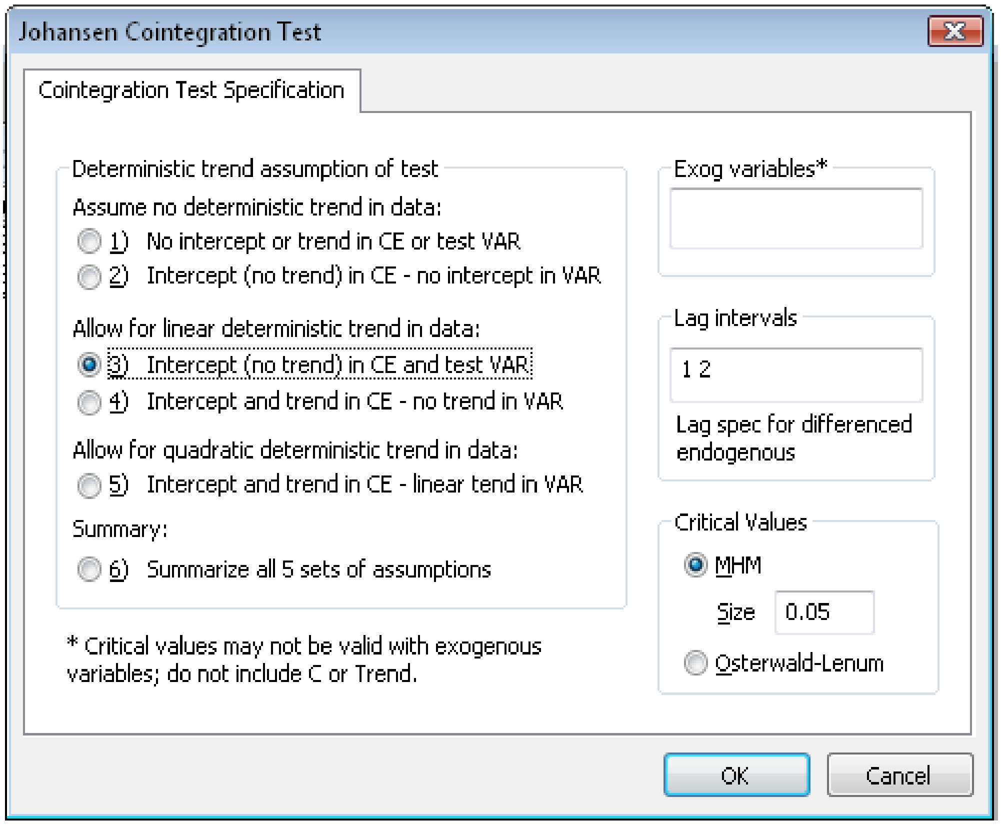
Step 3: Perform the test equation for cointegration with optimal lag length determined above.

We test each one of the models for cointegration in Eviews by opening **Quick/Group Statistics/Cointegration Test**. Then in the **series list** window, we enter the names of the series to check for cointegration, for example:

X Y Z

then press <**OK**>. The five alternative models explained in **Step 3** above are given under labels 1, 2, 3, 4, and 5. There is another option (option 6 in Eviews) that compares all these models together.

Figure 8.2: Johansen test specification



In our case, we wish to estimate models 2, 3, and 4 (because as noted earlier models 1 and 5 occur only very rarely). To estimate model 2, we select that model, and specify the number of lags in the bottom-right corner box that has the (default by Eviews) numbers '1 2' for inclusion of two lags. We change the '1 2' ro '1 4' for four lags, and click <**OK**> to get the results.

Note that there is another box that allows us to include (by typing their names) variables that will be treated as exogenous. Here we usually put variables that are either found to be $I(0)$ or dummy variables that possibly affect the behaviour of the model.

Doing the same for models 3 and 4 (in the **untitled group window** select **View/Cointegration Test**) and simply change the model by clicking next to 3 or 4. We get the results as the following tables.

Table 8.6: Cointegration test results (model 2)

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.286013	51.38016	35.19275	0.0004
At most 1 *	0.139113	22.07070	20.26184	0.0279
At most 2	0.098679	9.038752	9.164546	0.0528

Trace test indicates 2 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Table 8.7: Cointegration test results (model 3)

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None	0.166219	25.79093	29.79707	0.1351
At most 1	0.108092	9.975705	15.49471	0.2826
At most 2	0.000271	0.023559	3.841466	0.8779

Trace test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Table 8.8: Cointegration test results (model 4)

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.319369	52.02666	42.91525	0.0048
At most 1	0.137657	18.55470	25.87211	0.3079
At most 2	0.063092	5.669843	12.51798	0.5033

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Step 4: Decide which of the estimated models to choose in testing for cointegration.

We can use 'Option 6' in Figure 8.2 to select the best model to use in the VECM.

Table 8.9: Johansen Cointegration Test Summary

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Trace	1	2	0	1	1
Max-Eig	1	1	0	1	1

*Critical values based on MacKinnon-Haug-Michelis (1999)

	Akaike Information Criteria by Rank (rows) and Model (columns)				
0	-8.899905	-8.899905	-9.125069	-9.125069	-9.119250
1	-9.048621	-9.075876	-9.168922	-9.348885*	-9.321664
2	-9.037019	-9.064749	-9.145384	-9.336067	-9.330688
3	-8.914076	-9.007723	-9.007723	-9.240318	-9.240318

	Schwarz Criteria by Rank (rows) and Model (columns)				
0	-7.879530	-7.879530	-8.019662	-8.019662	-7.928812
1	-7.858182	-7.857094	-7.893453	-8.045071*	-7.961163
2	-7.676518	-7.647561	-7.699851	-7.833847	-7.800125
3	-7.383513	-7.392128	-7.392128	-7.539692	-7.539692

From this summary, we see that Model 4 (Linear, Intercept, and Trend) seems to be the best model (one cointegrating vector).

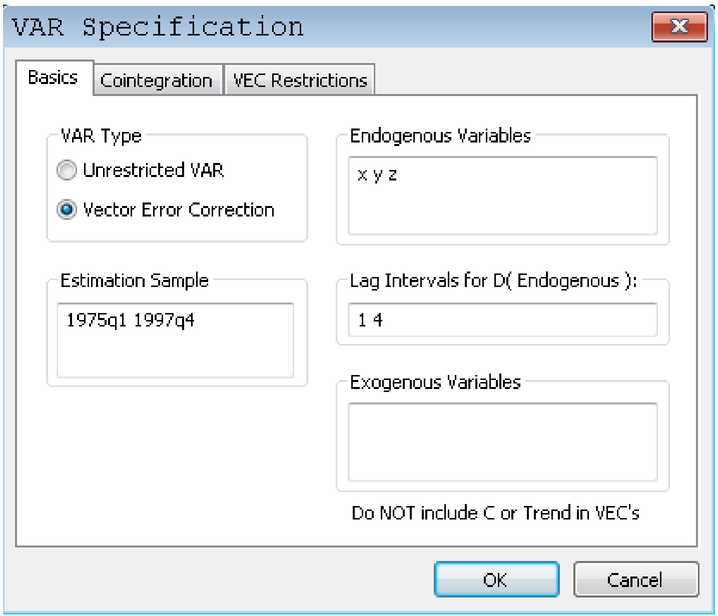
8.3 Estimation of VECM in Eviews

After determining optimal lag length, and number of cointegrating vectors, we start estimating the VECM:

(1) Quick/Estimate VAR

(2) In 'VAR Specification', choose Vector Error Correction in 'VAR type', add all variables in 'Endogeneous Variables', then choose 'Lag Length'.

Figure 8.3: VAR specification



(3) We then enter number of cointegrating in 'Cointegration':

Figure 8.4: Cointegration

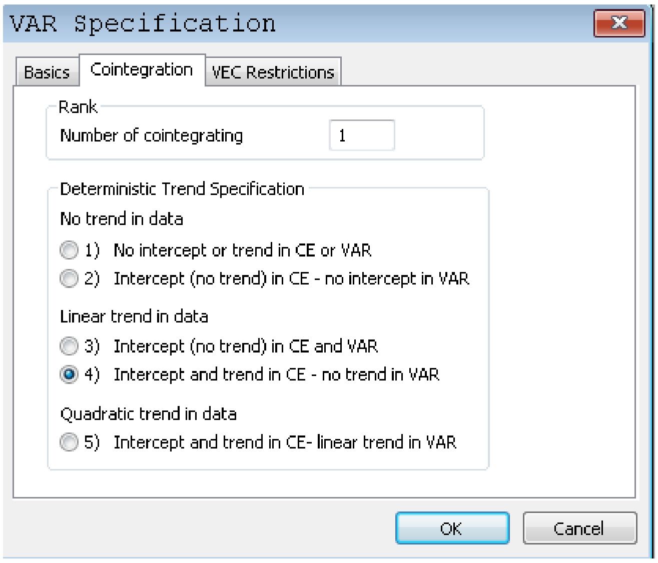


Table 8.10: VECM estimates

Vector Error Correction Estimates
Date: 11/17/13 Time: 02:31
Sample (adjusted): 1976Q2 1997Q4
Included observations: 87 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1		
X(-1)	1.000000		
Y(-1)	-1.635290 (0.12464) [-13.1206]		
Z(-1)	0.013053 (0.00192) [6.79987]		
@TREND(75Q1)	0.012697 (0.00102) [12.5082]		
C	1.316243		
Error Correction:	**D(X)**	**D(Y)**	**D(Z)**
CointEq1	-0.337203 (0.06479) [-5.20443]	-0.011388 (0.04697) [-0.24246]	-4.735796 (5.15504) [-0.91867]

9. CAUSALITY TESTS

According to Asteriou (2007), one of the good features of VAR models is that they allow us to test the direction of causality. Causality in econometrics is somewhat different to the concept in everyday use (take examples?); it refers more to the ability of one variable to predict (and therefore cause) the other.

Suppose two stationary variables, say Y_t and X_t, affect each other with distributed lags. The relationship between Y_t and X_t can be captured by a VAR model. In this case, it is possible to have that (a) Y_t causes X_t (Unidirectional Granger causality from Y to X), (b) X_t causes Y_t (Unidirectional Granger causality from X to Y), (c) there is a bi-directional feedback (causality among the variables), and finally (d) the two variables are independent. The problem is to find an appropriate procedure that allows us to test and statistically detect the cause and effect relationship among variables.

Granger (1969) developed a relatively simple test that defined causality as follows: a variable Y_t is said to Granger-cause X_t, if X_t can be predicted with greater accuracy by using past values of the Y_t variable rather than not using such past values, all other terms remaining unchanged. This test has been widely applied in economic policy analysis.

9.1 The Granger Causality Test

The Granger causality test for the of two *stationary* variables, say, ΔY_t and ΔX_t, involves as a first step the estimation of the following VAR model:

$$\Delta Y_t = A_1 + \sum_{j=1}^{p} C_j \Delta Y_{t-j} + \sum_{j=1}^{p} D_j \Delta X_{t-j} + u_{1t} \qquad (40)$$

$$\Delta X_t = A_2 + \sum_{j=1}^{p} E_j \Delta Y_{t-j} + \sum_{j=1}^{p} F_j \Delta X_{t-j} + u_{2t} \qquad (41)$$

where it is assumed that both u_{1t} and u_{1t} are uncorrelated white-noise error terms, and Y_t and X_t are integrated of order 1. In this model, we can have the following different cases:

Case 1 The lagged ΔX terms in equation (40) are statistically different from zero as a group, and the lagged ΔY terms in equation (41) are not statistically different from zero. In this case, we have that X_t causes Y_t.

Case 2 The lagged ΔY terms in equation (41) are statistically different from zero as a group, and the lagged ΔX terms in equation (40) are not statistically different from zero. In this case, we have that Y_t causes X_t.

Case 3 Both sets of lagged ΔX and lagged ΔY terms are statistically different from zero as a group in equation (40) and (41), so that we have bi-directional causality between Y_t and X_t.

Case 4 Both sets of lagged ΔX and lagged ΔY terms are not statistically different from zero in equation (40) and (41), so that X_t is independent of Y_t.

The Granger causality test, then, involves the following procedures. First, estimate the VAR model given by equations (40) and (41). Then check the significance of the coefficients and apply variable deletion tests first in the lagged X terms for equation (40), and then in the lagged Y terms in equation (41). According to the result of the variable deletion tests, we may conclude about the direction of causality based upon the four cases mentioned above.

More analytically, and for the case of one equation (i.e. we will examine equation (40)), it is intuitive to reverse the procedure in order to test for equation (41), and we perform the following steps:

Step 1 Regress ΔY_t on lagged ΔY terms as in the following model:

$$\Delta Y_t = A + \sum_{j=1}^{p} C_j \Delta Y_{t-j} + u_{1t} \qquad (42)$$

and obtain the *RSS* of this regression (which is the restricted one) and label it as RSS_R.

Step 2 Regress ΔY_t on lagged ΔY terms plus lagged ΔX terms as in the following model:

$$\Delta Y_t = A_1 + \sum_{j=1}^{p} C_j \Delta Y_{t-j} + \sum_{j=1}^{p} D_j \Delta X_{t-j} + u_{1t} \qquad (40)$$

and obtain the *RSS* of this regression (which is the unrestricted one) and label it as RSS_U.

Step 3 Set the null and alternative hypotheses as below:

$$H_0 : \sum_{j=1}^{p} D_j = 0 \text{ or } X_t \text{ does not cause } Y_t$$

$$H_1 : \sum_{j=1}^{p} D_j \neq 0 \text{ or } X_t \text{ does cause } Y_t$$

Step 4 Calculate the *F* statistic for the normal Wald test on coefficient restrictions given by:

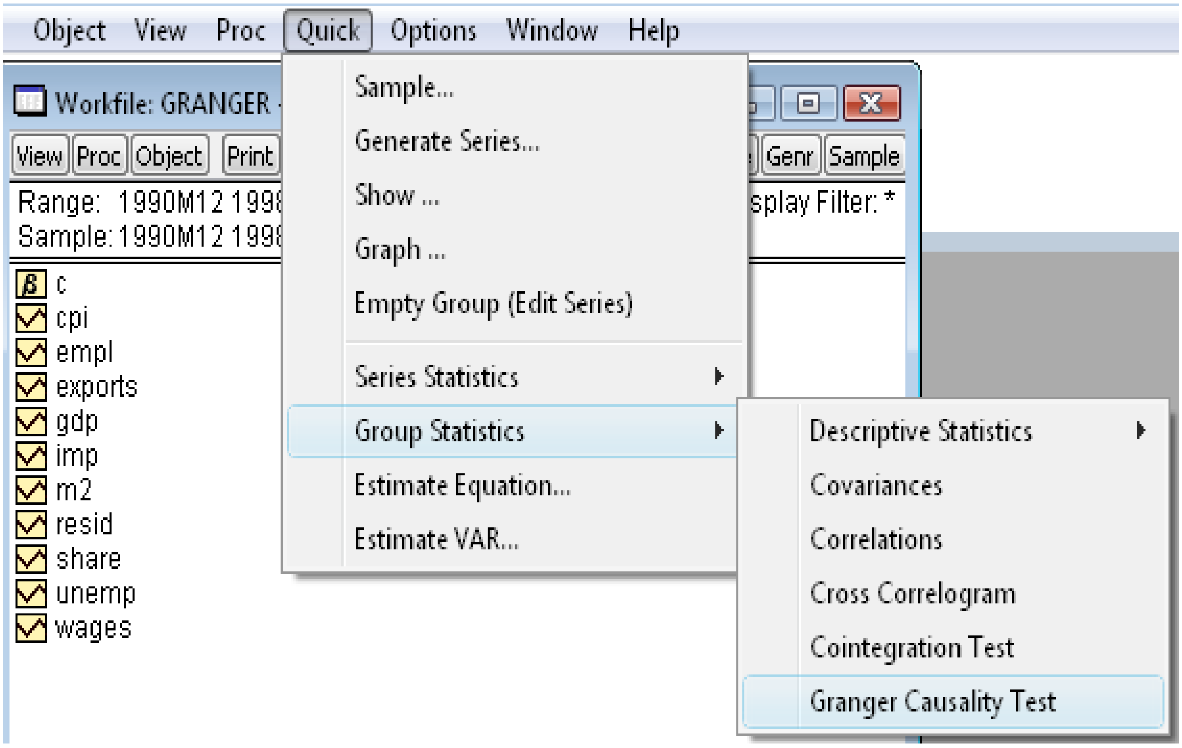
$$F = \frac{(RSS_R - RSS_U)/p}{RSS_u/(N-k)}$$

where N is the included observations and $k = 2p + 1$ is the number of estimated coefficients in the unrestricted model.

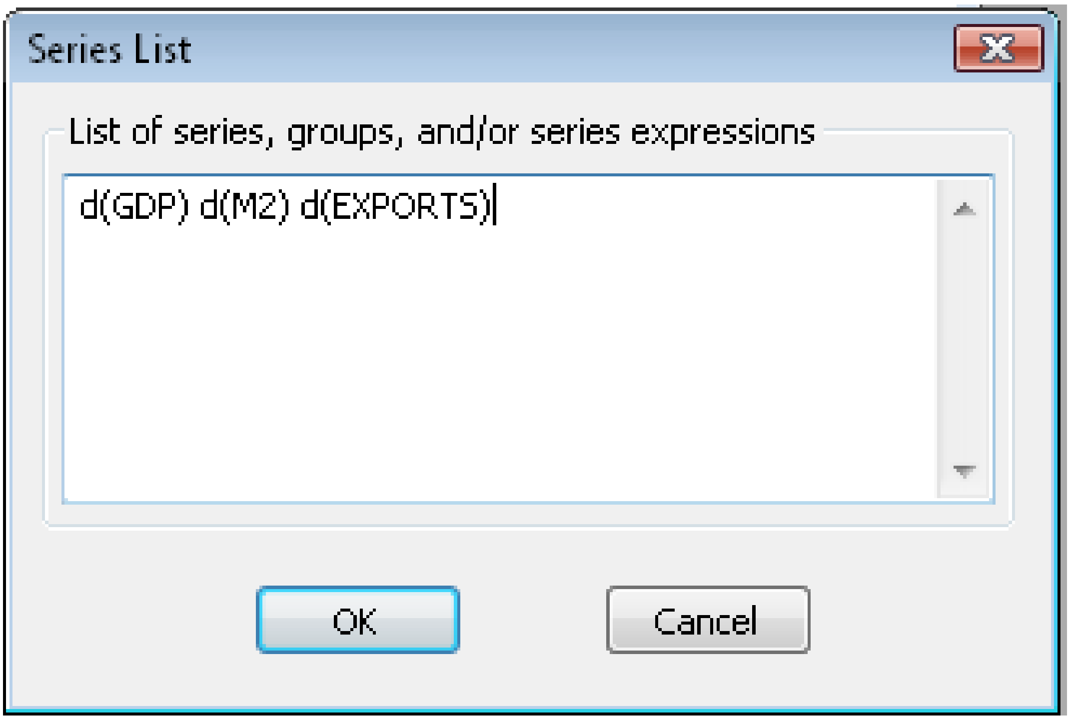
Step 5 If the computed F value exceeds the critical F value, reject the null hypothesis and conclude that X_t causes Y_t.

Open the file **GRANGER.wf1** and then perform as follows:

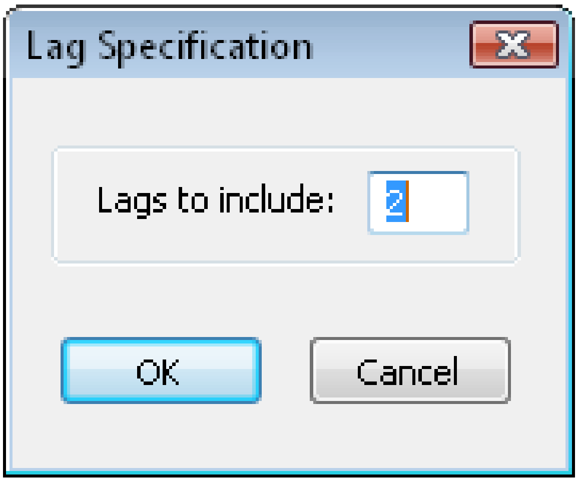
Figure 9.1: An illustration of GRANGER in Eviews[20]



[20] However, this is not a good way of conducting Granger causality test (why?)



Why I use the first differenced series?



Note that this 'lag specification' is not highly appreciated in empirical studies (why?).

Table 9.2: A result of the Granger causality tests

Pairwise Granger Causality Tests
Date: 02/20/10 Time: 16:16
Sample: 1990M12 1998M12
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
D(M2) does not Granger Cause D(GDP)	94	6.98481	0.0015
D(GDP) does not Granger Cause D(M2)		1.52055	0.2242
D(EXPORTS) does not Granger Cause D(GDP)	94	9.76799	0.0001
D(GDP) does not Granger Cause D(EXPORTS)		2.27818	0.1084
D(EXPORTS) does not Granger Cause D(M2)	94	0.40254	0.6698
D(M2) does not Granger Cause D(EXPORTS)		8.99303	0.0003

9.2 The Sims Causality Test

Sims (1980) proposed an alternative test for causality making use of the fact that in any general notion of causality, it is not possible for the future to cause the present. Therefore, when we want to check whether a variable Y_t causes X_t, Sims suggests estimating the following VAR model:

$$\Delta Y_t = A_1 + \sum_{j=1}^{p} C_j \Delta Y_{t-j} + \sum_{j=1}^{p} D_j \Delta X_{t-j} + \sum_{i=1}^{m} \alpha_i \Delta X_{t+i} + u_{1t} \quad (43)$$

$$\Delta X_t = A_2 + \sum_{j=1}^{p} E_j \Delta X_{t-j} + \sum_{j=1}^{p} F_j \Delta Y_{t-j} + \sum_{i=1}^{m} \beta_i \Delta Y_{t+i} + u_{2t} \quad (44)$$

Assume that Y_t and X_t are nonstationay [$I(1)$]. The new approach here is that apart from lagged values of X and Y, there are also leading values of X included in the first equation (and similarly leading values of Y in the second equation).

Examining only the first equation, if Y_t causes X_t, then we will expect that there is some relationship between Y and the leading values of X. Therefore, instead of testing for the lagged values of X_t, we test for $\sum_{i=1}^{m} \alpha_i = 0$. Note that if we reject the restriction, then the causality runs from Y_t to X_t, and not vice versa, since the future cannot cause the present.

To carry out the test, we simply estimate a model with no leading terms (which is the restricted model) and then the model as appears in (43), which is the unrestricted model, and the obtain the *F* statistic as in the Granger test above.

It is unclear which version of the two tests is preferable, and most researchers use both. The Sims test, however, using more regressors (due to the inclusion of the leading terms), leads to a bigger loss of degrees of freedom.

10. LAG LENGTH SELECTION CRITERIA

This section discusses statistical methods for choosing the number of lags, first in an autoregression, then in a time series regression model with multiple predictors.

10.1 Determining the Order of an Autoregression

According to Stock and Watson (2007), choosing the order *p* of an autoregression requires balancing the marginal benefit of including more lags against the marginal cost of additional estimation uncertainty. On the one hand, if the order of an estimated autoregression is too low, you will omit potentially valuable information contained in the more distant lagged values. On the other hand, if it is too high, you will be estimating more coefficients than necessary, which in turn introduces additional estimation error into your forecasts. Various statistical methods can be used, but two most important ones are SIC and AIC[21].

The SIC. A way around this problem is to estimate *p* by minimizing an 'information criterion'. One such information is the **Schwarz Information Criterion** (SIC), which is:

$$SIC(p) = \ln\left(\frac{RSS(p)}{T}\right) + (p + 1)\frac{\ln T}{T} \qquad (45)$$

where RSS(p) is the sum of squared residuals of the estimated AR(p). The SIC estimator of p, $\hat{p}$, is the value that minimizes SIC(p) among the possible choices p = 0, 1, …, p_{max}, where p_{max} is the largest value of p considered.

The formula for SIC might look a bit mysterious at first, but it has an intuitive appeal. Consider the first term in equation (45). Because the regression coefficients are estimated by OLS, the sum of squared residuals necessarily decreases (or at least does not increase)

[21] Others including FPE, HQ, and LR are also used in empirical studies.

when you add a lag. In contrast, the second term is the number of estimated regression coefficients (the number of lags, *p*, plus one for the intercept) times the factor (ln*T*)/*T*. This second term increases when you add a lag. The SIC trades off these two forces so that the number of lags that minimizes the SIC is a consistent estimator of the true lag length.

The AIC. The SIC is not the only information criterion; another is the **Akaike Information Criterion** (AIC), which is:

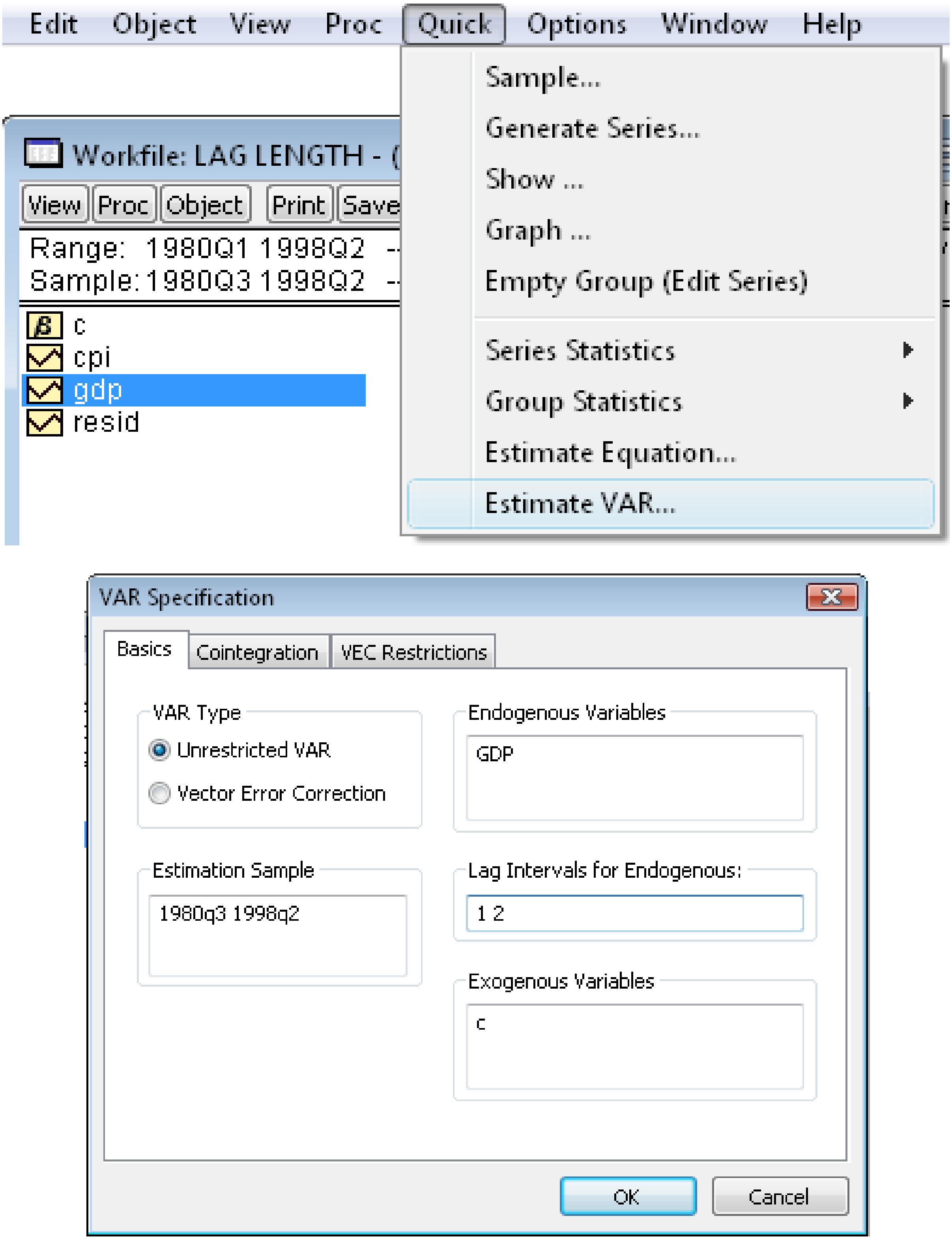
$$AIC(p) = \ln\left(\frac{RSS(p)}{T}\right) + (p+1)\frac{2}{T} \qquad (46)$$

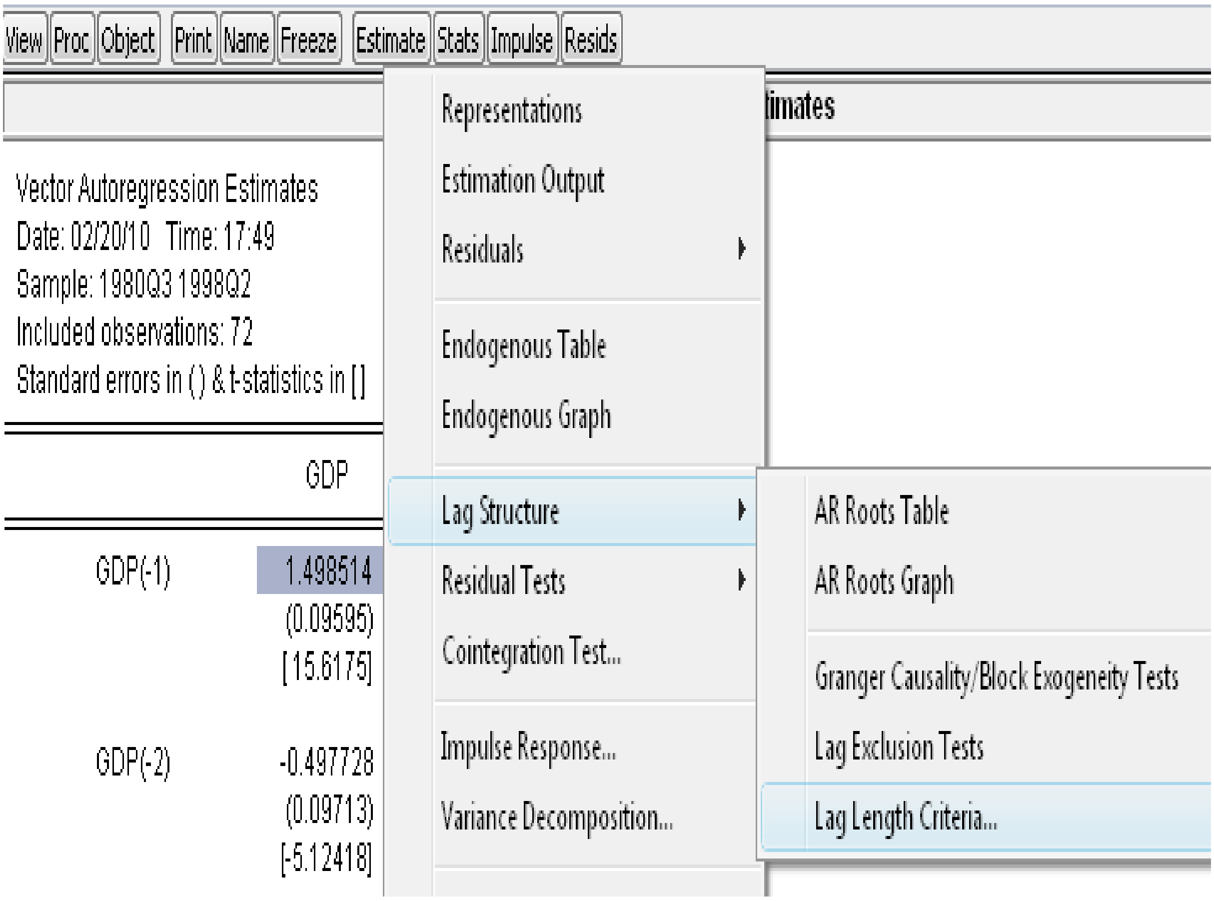
The difference between the AIC and the SIC is that the term "ln*T*" in the SIC is replaced by "2" in the AIC, so the second term in the AIC is smaller (why?). Stock and Watson (2007) state that the second term in the AIC is not large enough to ensure that the correct lag length is chosen, even in large samples, so the AIC estimator of *p* is not consistent.

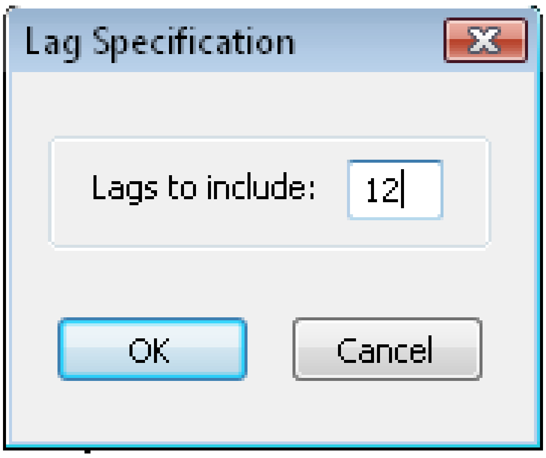
Despite this theoretical shortcoming, the AIC is widely used in practice. If you are concerned that the SIC might yield a model with two few lags, the AIC provides a reasonable alternative.

How to choose the optimal lag length in Eviews?

Figure 10.1: Lag length of a VAR: Step-by-step




View
Proc
Object
Print
Name
Freeze
Estimate
Stats
Impulse
Resids
Representations
Estimation Output
Residuals
Endogenous Table
Endogenous Graph
Lag Structure
Residual Tests
Cointegration Test...
Impulse Response...
Variance Decomposition...
timates
Vector Autoregression Estimates
Date: 02/20/10 Time: 17:49
Sample: 1980Q3 1998Q2
Included observations: 72
Standard errors in ( ) & t-statistics in [ ]
GDP
GDP(-1) 1.498514
(0.09595)
[15.6175]
GDP(-2) -0.497728
(0.09713)
[-5.12418]
AR Roots Table
AR Roots Graph
Granger Causality/Block Exogeneity Tests
Lag Exclusion Tests
Lag Length Criteria...


Lag Specification
Lags to include:
12
OK
Cancel

VAR Lag Order Selection Criteria
Endogenous variables: GDP
Exogenous variables: C
Date: 02/20/10 Time: 17:50
Sample: 1980Q3 1998Q2
Included observations: 62

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-228.6805	NA	96.65737	7.409047	7.443355	7.422517
1	-49.70910	346.3962	0.310423	1.668036	1.736653	1.694976
2	-40.66688	17.20938	0.239502	1.408609	1.511535*	1.449020*
3	-40.64148	0.047524	0.247177	1.440048	1.577282	1.493930
4	-37.97270	4.907120*	0.234263*	1.386216*	1.557759	1.453568
5	-37.95867	0.025345	0.241896	1.418022	1.623873	1.498844
6	-37.78489	0.308309	0.248519	1.444674	1.684834	1.538967
7	-37.61696	0.292531	0.255402	1.471515	1.745984	1.579278
8	-35.40017	3.790002	0.245724	1.432263	1.741041	1.553497
9	-34.30781	1.832336	0.245183	1.429284	1.772370	1.563989
10	-34.19246	0.189771	0.252521	1.457821	1.835216	1.605996
11	-34.19048	0.003201	0.261083	1.490015	1.901719	1.651661
12	-33.68053	0.806050	0.265606	1.505823	1.951835	1.680939

* indicates lag order selected by the criterion

10.2 Lag Length in Multiple-Predictor Regression

As in an autoregression, the SIC and AIC can be used to estimate the number of lags and the variables in the time series regression model with multiple predictors. After we determine (and fix) the optimal lags for the autoregression, we then continue add successive lags of explanatory variables. If the regression model has *K* coefficients (including the intercept), the SIC and AIC are defined as:

$$\mathrm{SIC}(K) = \ln\left(\frac{\mathrm{RSS}(K)}{T}\right) + K\,\frac{\ln T}{T} \qquad (47)$$

$$\mathrm{AIC}(K) = \ln\left(\frac{\mathrm{RSS}(K)}{T}\right) + K\,\frac{2}{T} \qquad (48)$$

There are two important practical considerations when using an information criterion to estimate the lag length. First, as is the case for the autoregression, all the candidate models must be estimated over the same sample. Second, when there are multiple predictors, this approach is computationally demanding because it requires many different models. According to Stock and Watson (2007), in practice, a convenient shortcut is to require all the regressors to have the same number of lags, that is, to require that $p = q_1 = \ldots = q_k$. This does, therefore, not provide the best lag length structure.

11. BOUNDS TEST FOR COINTEGRATION

Another way to test for cointegration and causality is the Bounds Test for Cointegration within ARDL modelling approach. This model was developed by Pesaran *et al.* (2001) and can be applied irrespective of the order of integration of the variables (irrespective of whether regressors are purely $I(0)$, purely $I(1)$ or mutually cointegrated). This is specially linked with the ECM models and called VECM as specified in Section 5.

11.1 The Model

The ARDL modelling approach involves estimating the following error correction models:

$$\Delta Y_t = \alpha_{0y} + \alpha_{1y}Y_{t-1} + \alpha_{2y}X_{t-1} + \sum_{i=1}^{m}\beta_i\Delta Y_{t-i} + \sum_{j=1}^{m}\gamma_j\Delta X_{t-j} + u_{1t} \quad (49)$$

$$\Delta X_t = \alpha_{0x} + \alpha_{1x}Y_{t-1} + \alpha_{2x}X_{t-1} + \sum_{i=1}^{m}\theta_i\Delta X_{t-i} + \sum_{j=1}^{m}\delta_j\Delta Y_{t-j} + u_{2t} \quad (50)$$

11.2 Test Procedure

Suppose we have Y_t and X_t are nonstationary.

Step 1: Testing for the unit root of Y_t and X_t

(using either DF, ADF, or PP tests)

Suppose the test results indicate that Y_t and X_t have different orders of integration [$I(0)$ and/or $I(1)$].

Step 2: Testing for cointegration between Y_t and X_t

(using Bounds test approach)

For equations 1 and 2, the F-test (normal Wald test) is used for investigating one or more long-run relationships. In the case of one or more long-run relationships, the F-test indicates which variable should be normalized.

In equation (49), when Y is the dependent variable, the null hypothesis of no cointegration is H_0: $\alpha_{1y} = \alpha_{2y} = 0$ and the alternative hypothesis of cointegration is H_1: $\alpha_{1y} \neq \alpha_{2y} \neq 0$.

On the other hand, in equation (50), when X is the dependent variable, the null hypothesis of no cointegration is H_0: $\alpha_{1x} = \alpha_{2x} = 0$ and the alternative hypothesis of cointegration is H_1: $\alpha_{1x} \neq \alpha_{2x} \neq 0$.

In this model, we still apply the Granger causality test following the same procedure presented above.

12. SUGGESTED RESEARCH TOPICS

From previous studies, I would like to suggest the following topics that you can consider for your coming research proposal.

Saving, Investment and Economic Development

- An analysis of the interaction among savings, investments and growth in Vietnam
- Are saving and investment cointegrated? The case of Vietnam
- Causal relationship between domestic savings and economic growth: Evidence from Vietnam
- Does saving really matter for growth? Evidence from Vietnam
- The relationship between savings and growth: Cointegration and causality evidence from Vietnam
- The saving and investment nexus for Vietnam: Evidence from cointegration tests
- Do foreign direct investment and gross domestic investment promote economic growth?
- Foreign direct investment and economic growth in Vietnam: An empirical study of causality and error correction mechanisms
- The interactions among foreign direct investment, economic growth, degree of openness and unemployment in Vietnam

Trade and Economic Development

- How trade and foreign investment affect the growth: A case of Vietnam?
- Trade, foreign direct investment and economic growth in Vietnam
- A cointegration analysis of the long-run relationship between black and official foreign exchange rates: The case of the Vietnam dong
- An empirical investigation of the causal relationship between openness and economic growth in Vietnam
- Export and economic growth in Vietnam: A Granger causality analysis
- Export expansion and economic growth: Testing for cointegration and causality for Vietnam
- Is the export-led growth hypothesis valid for Vietnam?
- Is there a long-run relationship between exports and imports in Vietnam?
- On economic growth, FDI and exports in Vietnam
- Trade liberalization and industrial growth in Vietnam: A cointegration analysis

Stock Market and Economic Development

- Causality between financial development and economic growth: An application of vector error correction to Vietnam

- Financial development and the FDI growth nexus: The Vietnam case
- Macroeconomic environment and stock market: The Vietnam case
- The relationship between economic factors and equity market in Vietnam
- Modelling the linkages between the US and Vietnam stock markets
- The long-run relationship between stock returns and inflation in Vietnam
- The relationship between financial deepening and economic growth in Vietnam
- Testing market efficient hypothesis: The Vietnam stock market
- Threshold adjustment in the long-run relationship between stock prices and economic activity

Energy and the Economy

- The dynamic relationship between the GDP, imports and domestic production of crude oil: Evidence from Vietnam
- Causal relationship between gas consumption and economic growth: A case of Vietnam
- Causal relationship between energy consumption and economic growth: The case of Vietnam

- Causality relationship between electricity consumption and GDP in Vietnam
- The causal relationship between electricity consumption and economic growth in Vietnam
- A cointegration analysis of gasoline demand in Vietnam
- Cointegration and causality testing of the energy-GDP relationship: A case of Vietnam
- Does more energy consumption bolster economic growth?
- Energy consumption and economic growth in Vietnam: Evidence from a cointegration and error correction model
- The causality between energy consumption and economic growth in Vietnam
- The relationship between the price of oil and macroeconomic performance: Empirical evidence for Vietnam

Fiscal Policy and Economic Development

- A causal relationship between government spending and economic development: An empirical examination of the Vietnam economy
- Economic growth and government expenditure: Evidence from Vietnam
- Government revenue, government expenditure, and temporal causality: Evidence from Vietnam

- The relationship between budget deficits and money demand: Evidence from Vietnam

Monetary Policy and Economic Development

- Granger causality between money and income for the Vietnam economy
- Money, inflation and casuality: Evidence from Vietnam
- Money-output Granger causality: An empirical analysis for Vietnam
- Time-varying parameter error correction models: The demand for money in Vietnam
- Monetary transmission mechanism in Vietnam: A VAR analysis

Tourism and Economic Development

- Cointegration analysis of quarterly tourism demand by international tourists: Evidence from Vietnam
- Does tourism influence economic growth? A dynamic panel data approach
- International tourism and economic development in Vietnam: A Granger causality test
- Tourism demand modelling: Some issues regarding unit roots, co-integration and diagnostic tests
- Tourism, trade and growth: the case of Vietnam

Agriculture and Economic Development

- Dynamics of rice prices and agricultural wages in Vietnam
- Macroeconomic factors and agricultural production linkages: A case of Vietnam
- Is agriculture the engine of growth?
- The causal relationship between fertilizer consumption and agricultural productivity in Vietnam
- Macroeconomics and agriculture in Vietnam

Others

- Hypotheses testing concerning relationships between spot prices of various types of coffee
- The relationship between wages and prices in Vietban
- An error correction model of luxury goods expenditures: Evidence from Vietnam
- The relationship between macroeconomic variables and housing price index: A case of Vietnam
- Explaining house prices in Vietnam
- Long-term trend and short-run dynamics of the Vietnam gold price: an error correction modelling approach
- Macroeconomic adjustment and private manufacturing investment in Vietnam: A time-series analysis

- Testing for the long run relationship between nominal interest rates and inflation using cointegration techniques
- The long-run relationship between house prices and income: Evidence from Vietnam housing markets

It is noted that many empirical studies use the nonstationary panels, typically characterised by panel unit root tests and panel cointegration tests. However, they are beyond the scope of this lecture.

REFERENCES

Asteriou, D. and Hall, S.G. (2007) *Applied Econometrics: A Modern Approach Using Eviews and Microfit*, Revised Edition. Palgrave Macmillan.

Cheung, Y.W. and Lai, K.S. (1995) 'Lag Order and Critical Values of the Augmented Dickey-Fuller Test', *Journal of Business & Economic Statistics*, Vol.13, No.3, pp.277-280.

Dickey, D.A. and Fuller, W.A. (1979) 'Distribution of the Estimators for Autoregressive Time Series with a Unit Root', *Journal of the American Statistical Association*, Vol.74, No.366, pp.427- 431.

Dickey, D.A. and Fuller, W.A. (1981) 'Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root', *Econometrica*, Vol.49, p.1063.

Diebold, F.X. (2004) *Elements of Forecasting*, 3rd Edition, Thomson.

Dolado, J., T.Jenkinson and S.Sosvilla-Rivero. (1990) 'Cointegration and Unit Roots', *Journal of Economic Surveys*, Vol.4, No.3.

Durbin, J. (1970) 'Testing for Serial Correlation in Least Squares Regression When Some of the Variables Are Lagged Dependent Variables', *Econometrica*, Vol.38, pp.410-421.

Engle, R.F. and Granger, C.W.J. (1987) 'Co-integration and Error Correction Estimates: Representation, Estimation, and Testing', *Econometrica*, Vol.55, p.251-276.

Granger, C.W.J. (1981) 'Some Properties of Time Series Data and Their Use in Econometric Model Specification', *Journal of Econometrics*, Vol.16, pp.121-130.

Granger, C.W.J. (1988) 'Some Recent Developments in the Concept of Causality', *Journal of Econometrics*, Vol.39, pp.199-211.

Granger, C.W.J. (2004) 'Time Series Analysis, Cointegration, and Applications', *The American Economic Review*, Vol.94, No.3, pp.421-425.

Granger, C.W.J. and Newbold, P. (1977) 'Spurious Regression in Econometrics', *Journal of Econometrics*, Vol.2, pp.111-120.

Griffiths, W.E., R.C.Hill and G.C.Lim. (2008) *Using Eviews for Principles of Econometrics*, 3rd Edition, John Wiley & Sons.

Gujarati, D.N. (2003) *Basic Econometrics*, 4th Edition, McGraw-Hill.

Gujarati, D.N. (2011) *Econometrics by Example*, 1st Edition, Palgrave Macmillan.

Hanke, J.E. and Wichern, D.W. (2005) *Business Forecasting*, 8th Edition, Pearson Education.

Holton, W.J. and Keating, B. (2007) *Business Forecasting With Accompanying Excel-Based ForecastXTM Software*, 5th Edition, McGraw-Hill.

Johansen, S. (1991) 'Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models', *Econometrica*, Vol.59, pp.1551-1580.

Johansen, S. and Juselius, K. (1990) 'Maximum Likelihood Estimation and Inference on Cointegration, with

Applications for the Demand for Money', *Oxford Bulletin of Economics and Statistics*, Vol.52, pp.169-210.

Kang, H. (1985) 'The Effects of Detrending in Granger Causality Tests', *Journal of Business & Economic Statistics*, Vol.3, No.4, pp.344-349.

Katircioglu, S. (2009) 'Tourism, Trade and Growth: The Case of Cyprus', *Applied Economics*, Vol.41, pp.2741-2750.

Li, X. (2001) 'Government Revenue, Government Expenditure, and Temporal Causality: Evidence from China', *Applied Economics*, Vol.33, pp.485-497.

Ljung, G.M. and Box, G.E.P. (1978) 'On a measure of Lack of Fit in Times Series Models', *Biometrica*, Vol.65, pp.297-303.

Mackinnon, J.G. (1994) 'Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration', *Journal of Business & Economic Statistics*, Vol.12, No.2, pp.167-176.

Mackinnon, J.G. (1996) 'Numerical Distribution Functions for Unit Root and Cointegration Tests', *Journal of Applied Econometrics*, Vol.11, No.6, pp.601-618.

Mackinnon, J.G., Alfred A. Haug and Leo Michelis. (1999) 'Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration', *Journal of Applied Econometrics*, Vol.14, No.5, pp.563-577.

Mahmoud, E. (1984) 'Accuracy in Forecasting: A Survey', *Journal of Forecasting*, Vol.3, pp.139-159.

McNees, S. (1986) 'Forecasting Accuracy of Alternative Techniques: A Comparison of US Macroeconomic

Forecasts', *Journal of Business and Economic Statistics*, Vol.4, pp.5-15.

Mehrara, M. (2007) 'Energy-GDP Relationship for Oil-Exporting Countries: Iran, Kuwait and Saudi Arabia', *OPEC Review*, Vol.3.

Nguyen Trong Hoai, Phung Thanh Binh, and Nguyen Khanh Duy. (2009) *Forecasting and Data Analysis in Economics and Finance*, Statistical Publishing House.

Phillips, P.C.B. (1987) 'Time Series Regression with a Unit Root', *Econometrica*, Vol.55, No.2, pp.277-301.

Phillips, P.C.B. (1998) 'New Tools for Understanding Spurious Regressions', *Econometrica*, Vol.66, No.6, pp.1299-1325.

Phillips, P.C.B. and Perron, P. (1988) 'Testing for a Unit Root in Time Series Regression', *Biometrica*, Vol.75, No.2, pp.335-346.

Pindyck, R.S. and Rubinfeld, D.L. (1998) *Econometric Models and Economic Forecasts*, 4th Edition, McGraw-Hill.

Ramanathan, R. (2002) *Introductory Econometrics with Applications*, 5th edition, Harcourt College Publisher.

Sims, C.A. (1980) 'Macroeconomics and Reality', *Econometrica*, Vol.48, No.1, pp.1-48.

Stock, J.H. and Watson, M.W. (2007) *Introduction to Econometrics*, 2nd Edition, Pearson Education.

Studenmund, A.H. (2001) *Using Econometrics: A Practical Guide*, 4th Edition, Addison Wesley Longman.

Toda, H.Y. and Yamamoto, T. (1995) 'Statistical Inference in Vector Autoregressive with Possibly Integrated

Processes’, *Journal of Econometrics*, Vol.66, No.1, pp.225-250.

Vogelvang, B. (2005) *Econometrics: Theory and Applications with Eviews*, Pearson Education.

Wooldridge, J.M. (2003) *Introductory Econometrics: A Modern Approach*, 2nd Edition, Thomson.

APPENDIX

STATA COMMANDS

Source: Practical exercises, Advanced Econometrics course 2012, Wageningen University, The Netherlands.

EXAMPLE 1

Use the data set **WHEATOIL.dta**, which contains (nominal) prices of wheat (*pwht*), nominal oil prices (*poil*) and a time indicator (*t*). The data are monthly and available for the period Jan 1990 till December 2008 (19 years*12 months = 228 obs.). In this example, we will investigate whether there is a long-run relationship between wheat prices and oil prices. There may be all kinds of reasons for such a relationship: oil is an important input in fertilizer production, is used for applying machinery, drives transportation costs, etc.

Please declare the data to be time-series data using the following command:

tsset t

1. **Create one graph with line plots of both** *pwht* **and** *poil* **against** *t*. **Considering the line plot for** *pwht*, **do you think this variable is stationary? Motivate your answer.**

```
twoway (tsline pwht) (tsline poil, yaxis(2))
```

The graphs for both *pwht* and *poil* indicate that there are stochastic trends (means are not constant) and their variances are also not constant. For the *pwht*, it first increases and highly fluctuates (from observation 1 to about 70), followed by a declining period (from observation about 70 to about 120) with less fluctuation, then it tends to increase and especially decline very quickly in the last months. Therefore, we might say that *pwht* is not stationary.

2. Use appropriate tests to find out the orders of integration of both *pwht* and *poil*?

In order to check the order of integration for *pwht* we perform the Augmented Dickey Fuller (ADF) test and the KPSS test on *pwht* until finding a stationary time series.

2.1 The *pwht* series.

ADF test

H_0: The *pwht* series is non-stationary (the *pwht* series has a unit root)

As this is a monthly series, we starts with 12 lags. In addition, we include the trend in the test equation.

. dfuller pwht, trend regress lags(12)

Augmented Dickey-Fuller test for unit root   Number of obs   =   215

	Test Statistic	Interpolated Dickey-Fuller 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-2.537	-4.002	-3.435	-3.135

MacKinnon approximate p-value for Z(t) = 0.3098

D.pwht	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pwht						
L1.	-.0623981	.0245977	-2.54	0.012	-.1109022	-.0138941
LD.	.3271641	.0702986	4.65	0.000	.1885425	.4657856
L2D.	-.0400395	.0746265	-0.54	0.592	-.1871952	.1071163
L3D.	.1565079	.0754853	2.07	0.039	.0076587	.305357
L4D.	-.0820566	.0767288	-1.07	0.286	-.2333578	.0692446
L5D.	.1622511	.0766431	2.12	0.035	.0111189	.3133832
L6D.	.0923137	.0766486	1.20	0.230	-.0588293	.2434568
L7D.	-.166064	.0769536	-2.16	0.032	-.3178086	-.0143194
L8D.	-.0144802	.0794344	-0.18	0.856	-.1711167	.1421562
L9D.	.0276799	.0831624	0.33	0.740	-.1363078	.1916676
L10D.	.003314	.0833536	0.04	0.968	-.1610506	.1676786
L11D.	.1434974	.0860173	1.67	0.097	-.0261197	.3131146
L12D.	-.0383418	.0863893	-0.44	0.658	-.2086925	.1320089
_trend	.0010636	.0113115	0.09	0.925	-.0212415	.0233686
_cons	8.805942	3.751807	2.35	0.020	1.407768	16.20412

Because the coefficients of trend and lag 12 are not statistically significant, so we can remove these in the test equation.

```
. dfuller pwht, regress lags(11)

Augmented Dickey-Fuller test for unit root         Number of obs   =       216

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.806            -3.471            -2.882            -2.572
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0574

------------------------------------------------------------------------------
      D.pwht |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        pwht |
         L1. |  -.0657226   .0234216    -2.81   0.006    -.1119034   -.0195417
         LD. |   .3268774   .0698223     4.68   0.000     .1892075    .4645473
        L2D. |  -.0362116   .0736867    -0.49   0.624     -.181501    .1090779
        L3D. |   .1595078   .0747156     2.13   0.034     .0121896     .306826
        L4D. |  -.0777765    .075759    -1.03   0.306    -.2271519    .0715989
        L5D. |   .1710043   .0740281     2.31   0.022     .0250416    .3169669
        L6D. |   .0975358   .0755654     1.29   0.198    -.0514579    .2465296
        L7D. |  -.1650262   .0762435    -2.16   0.032    -.3153569   -.0146954
        L8D. |  -.0048568   .0768696    -0.06   0.950    -.1564221    .1467085
        L9D. |   .0270703   .0820492     0.33   0.742    -.1347076    .1888483
       L10D. |   .0064812   .0821761     0.08   0.937    -.1555469    .1685092
       L11D. |   .1389518    .083426     1.67   0.097    -.0255408    .3034444
             |
       _cons |   9.359079   3.334838     2.81   0.005     2.783716    15.93444
------------------------------------------------------------------------------
```

If we choose the 5% significance level, the coefficients of lag 8 to lag 11 are not significant. Therefore, we try the test equation with 7 lags.

```
. dfuller pwht, regress lags(7)

Augmented Dickey-Fuller test for unit root         Number of obs   =       220

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.579            -3.470            -2.882            -2.572
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0974

------------------------------------------------------------------------------
      D.pwht |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        pwht |
         L1. |  -.0512845   .0198859    -2.58   0.011    -.0904849    -.012084
         LD. |   .3158331   .0661881     4.77   0.000     .1853584    .4463078
        L2D. |  -.0505144   .0695679    -0.73   0.469    -.1876515    .0866228
        L3D. |   .1458971   .0706096     2.07   0.040     .0067064    .2850877
        L4D. |    -.10859   .0710331    -1.53   0.128    -.2486154    .0314354
        L5D. |   .1595286   .0712431     2.24   0.026     .0190892     .299968
        L6D. |   .0949542   .0722266     1.31   0.190    -.0474239    .2373324
        L7D. |  -.1907707   .0705425    -2.70   0.007    -.3298291   -.0517123
             |
       _cons |   7.406046   2.888567     2.56   0.011     1.711898    13.10019
------------------------------------------------------------------------------
```

As the absolute value of the test statistics (2.579) is smaller than the absolute value of 5% critical value (2.882), we cannot reject the null hypothesis at 5% significance level. Therefore, ADF test suggests that *pwht* series is not stationary. To be sure, we apply the KPSS test (with H_0: the *pwht* series is stationary).

KPSS test

```
. kpss pwht

KPSS test for pwht

Maxlag = 14 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: pwht is trend stationary

10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order    Test statistic
    0            2.07
    1            1.06
    2             .72
    3            .552
    4            .452
    5            .385
    6            .339
    7            .304
    8            .277
    9            .256
   10            .239
   11            .225
   12            .213
   13            .203
   14            .194
```

All test statistics are greater than the 5% critical values, so we reject the null hypothesis. That means the *pwht* series is non-stationary.

We now examine stationarity of the first-differenced series of *pwht* without the constant term in the test equation because there is no trend in the original series of *pwht*. Here is the test result.

ADF test

H_0: The first-difference of *pwht* is not stationary.

```
. dfuller D.pwht, regress lags(12) nocons

Augmented Dickey-Fuller test for unit root         Number of obs   =       214

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -4.796            -2.585            -1.950            -1.618

------------------------------------------------------------------------------
     D2.pwht |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        pwht |
         LD. |  -1.107196   .2308406    -4.80   0.000    -1.562376   -.6520163
        LD2. |    .394732   .2208416     1.79   0.075    -.0407314    .8301954
       L2D2. |    .327055   .2122806     1.54   0.125    -.0915277    .7456377
       L3D2. |    .435902   .1998818     2.18   0.030     .0417678    .8300362
       L4D2. |   .3065757   .1877674     1.63   0.104    -.0636709    .6768222
       L5D2. |   .4198904   .1734387     2.42   0.016     .0778977    .7618832
       L6D2. |    .444228    .159587     2.78   0.006     .1295486    .7589074
       L7D2. |   .2308389   .1497876     1.54   0.125    -.0645178    .5261957
       L8D2. |   .1934566   .1393971     1.39   0.167    -.0814117    .4683249
       L9D2. |   .1609703   .1275091     1.26   0.208    -.0904568    .4123974
      L10D2. |   .1279276    .118637     1.08   0.282    -.1060053    .3618604
      L11D2. |   .2315775   .1015414     2.28   0.024     .0313544    .4318006
      L12D2. |    .190648   .0841474     2.27   0.025      .024723     .356573
------------------------------------------------------------------------------
```

As the absolute value of the test statistics (4.796) is larger than the 5% critical value (1.95), we reject the null hypothesis. That means the first-differenced series of *pwht* is stationary. We now examine the KPSS test for this first-differenced series.

KPSS test

H_0: The first-differenced series of *pwht* is stationary.

```
. kpss D.pwht

KPSS test for D.pwht

Maxlag = 14 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: D.pwht is trend stationary

10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order     Test statistic
    0            .0527
    1            .0418
    2            .0386
    3            .0361
    4            .0353
    5            .0342
    6            .0331
    7            .0333
    8            .0341
    9             .035
   10            .0361
   11            .0368
   12            .0376
   13            .0389
   14            .0406
```

The KPSS test results indicate that we fail to reject the null hypothesis.

In conclusion, the pwht series is integrated of order one [$I(1)$].

2.2 The poil series.

We proceed in the same way as in the case of *pwht*.

First I perform the ADF test and the KPSS test on the values of *poil*.

ADF test

H_0: The *poil* series is not stationary.

```
. dfuller poil, trend regress lags(12)

Augmented Dickey-Fuller test for unit root         Number of obs   =       215

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.706            -4.002            -3.435            -3.135
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.2334

------------------------------------------------------------------------------
D.poil       |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        poil |
         L1. |  -.0624086   .0230592    -2.71   0.007     -.107879   -.0169383
         LD. |   .2669757   .0699469     3.82   0.000     .1290476    .4049037
        L2D. |   .0974694   .0728268     1.34   0.182    -.0461375    .2410763
        L3D. |   .0962864   .0738565     1.30   0.194    -.0493509    .2419237
        L4D. |  -.0555913   .0755945    -0.74   0.463    -.2046558    .0934732
        L5D. |   .1087464   .0755929     1.44   0.152     -.040315    .2578077
        L6D. |  -.0251582   .0724683    -0.35   0.729    -.1680583    .1177418
        L7D. |  -.0056656   .0721355    -0.08   0.937    -.1479093    .1365781
        L8D. |   .0135717   .0720012     0.19   0.851    -.1284072    .1555505
        L9D. |  -.0741808    .070252    -1.06   0.292    -.2127104    .0643488
       L10D. |   .1606546   .0700959     2.29   0.023     .0224328    .2988765
       L11D. |  -.0106153   .0704357    -0.15   0.880    -.1495072    .1282765
       L12D. |  -.0806682   .0699725    -1.15   0.250    -.2186468    .0573103
      _trend |   .0094742   .0044587     2.12   0.035     .0006822    .0182663
       _cons |   1.356315   .6405548     2.12   0.035     .0932076    2.619423
------------------------------------------------------------------------------
```

Because the lag 11 and lag 12 are not significant, so we remove them in the test equation.

```
. dfuller poil, trend regress lags(10)

Augmented Dickey-Fuller test for unit root         Number of obs   =       217

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.370            -4.001            -3.434            -3.134
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0556

------------------------------------------------------------------------------
D.poil       |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        poil |
         L1. |  -.0734628    .021801    -3.37   0.001     -.116447   -.0304786
         LD. |   .2837988    .068133     4.17   0.000     .1494637     .418134
        L2D. |   .1035106   .0709027     1.46   0.146    -.0362855    .2433067
        L3D. |   .1061553   .0719452     1.48   0.142    -.0356963     .248007
        L4D. |  -.0767583    .070268    -1.09   0.276     -.215303    .0617864
        L5D. |   .0929341   .0707861     1.31   0.191    -.0466321    .2325004
        L6D. |   -.012689   .0716556    -0.18   0.860    -.1539696    .1285916
        L7D. |   .0042883   .0700411     0.06   0.951    -.1338089    .1423855
        L8D. |   .0333008   .0696247     0.48   0.633    -.1039755    .1705772
        L9D. |  -.0577865   .0689521    -0.84   0.403    -.1937367    .0781637
       L10D. |   .1716365   .0686519     2.50   0.013     .0362783    .3069948
      _trend |   .0112706   .0042534     2.65   0.009     .0028842    .0196569
       _cons |   1.538467    .609999     2.52   0.012     .3357554    2.741178
------------------------------------------------------------------------------
```

As the absolute value of the test statistics (3.37) is smaller than the 5% critical value (3.43), we cannot reject the null hypothesis. This implies that the *poil* series is not stationary. To be sure, we apply the KPSS test.

KPSS test

H_0: The *poil* series is stationary.

```
. kpss poil

KPSS test for poil

Maxlag = 14 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: poil is trend stationary

10%: 0.119  5% : 0.146  2.5%: 0.176  1% : 0.216

Lag order    Test statistic
    0           3.52
    1           1.81
    2           1.24
    3           .956
    4           .786
    5           .673
    6           .592
    7            .53
    8           .482
    9           .444
   10           .412
   11           .386
   12           .365
   13           .346
   14            .33
```

All test statistics are greater than the critical values (even at 1% significance level), so we reject the null hypothesis. That means the *poil* series is non-stationary.

We now examine stationarity of the first-differenced series of *poil* with the constant term in the test equation because there is trend in the original series of *poil*. Here is the test result.

ADF test

H_0: The first-difference of *poil* is not stationary.

```
. dfuller D.poil, regress lags(12)

Augmented Dickey-Fuller test for unit root         Number of obs   =       214

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -4.166            -3.472            -2.882            -2.572
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0008

------------------------------------------------------------------------------
     D2.poil |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        poil |
         LD. |  -1.077727   .2587189    -4.17   0.000    -1.587894   -.5675598
        LD2. |   .2995741   .2488476     1.20   0.230    -.1911275    .7902757
       L2D2. |   .3596402   .2398393     1.50   0.135    -.1132979    .8325784
       L3D2. |   .4342928   .2297527     1.89   0.060    -.0187556    .8873412
       L4D2. |   .3286435   .2136397     1.54   0.126    -.0926318    .7499188
       L5D2. |   .4128713   .1973281     2.09   0.038     .0237608    .8019818
       L6D2. |   .3685893    .181647     2.03   0.044     .0104002    .7267784
       L7D2. |   .3132068   .1645838     1.90   0.058    -.0113353     .637749
       L8D2. |   .2853127   .1470473     1.94   0.054    -.0046492    .5752746
       L9D2. |   .1569891     .12585     1.25   0.214    -.0911741    .4051524
      L10D2. |   .2772727   .1076319     2.58   0.011     .0650338    .4895115
      L11D2. |   .2238503   .0886082     2.53   0.012     .0491241    .3985764
      L12D2. |   .1339726   .0688075     1.95   0.053    -.0017086    .2696538
             |
       _cons |   .1043018   .1771545     0.59   0.557    -.2450284     .453632
------------------------------------------------------------------------------
```

As the absolute value of the test statistics (4.166) is larger than the 5% critical value (2.882), we reject the null hypothesis. The ADF test indicates that the first-differenced series of *poil* is stationary. We now confirm this by applying the KPSS test.

KPSS test

H_0: The first-difference of *poil* is stationary.

```
. kpss D.poil

KPSS test for D.poil

Maxlag = 14 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: D.poil is trend stationary

10%: 0.119  5% : 0.146  2.5%: 0.176  1% : 0.216

Lag order    Test statistic
    0            .0824
    1             .063
    2            .0562
    3            .0533
    4            .0536
    5            .0543
    6            .0558
    7            .0574
    8            .0589
    9            .0606
   10            .0614
   11            .0619
   12             .063
   13            .0648
   14            .0671
```

The KPSS test results indicate that the first-differenced series of *poil* is stationary.

Therefore, the poil series is integrated of order one [$I(1)$].

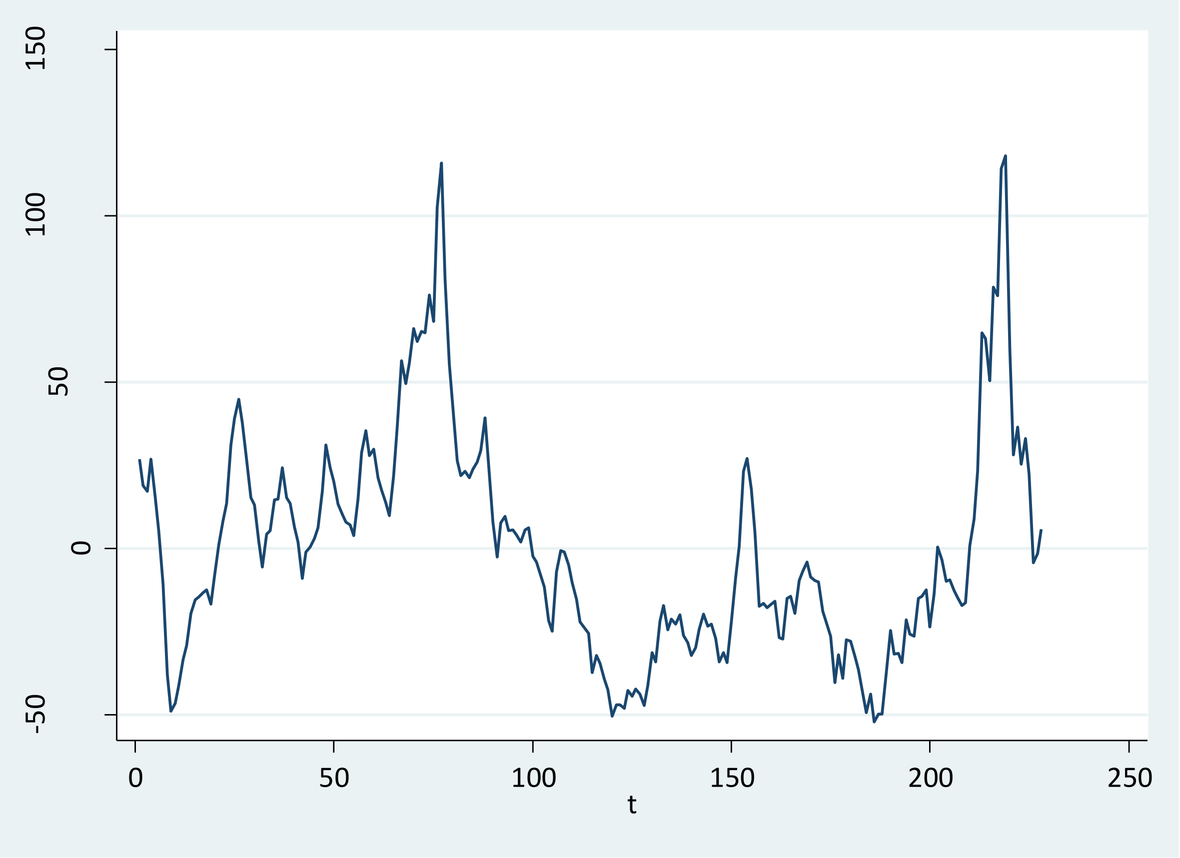
3. Given your findings at question 2, consider whether it is possible whether there exists a long-run (cointegrating) relationship between *pwht* and *poil*.

As both series are integrated of order one, there could exist a long-run relationship between *pwht* and *poil*.We must apply the cointegration tests to see whether there is really a long-run (or cointegrating) relationship between them.

4. Estimate the potential long-run relationship with *pwht* as dependent variable and *poil* as explanatory variable. Does this regression have the typical characteristics of a spurious regression or as a cointegrating relationship.

```
. regress pwht poil

      Source |       SS       df       MS              Number of obs =     228
-------------+------------------------------           F(  1,   226) =   61.51
       Model |  69028.652     1   69028.652           Prob > F      =  0.0000
    Residual |  253623.907   226  1122.22968           R-squared     =  0.2139
-------------+------------------------------           Adj R-squared =  0.2105
       Total |  322652.559   227  1421.37691           Root MSE      =    33.5

------------------------------------------------------------------------------
        pwht |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        poil |   1.273655    .162397     7.84   0.000     .9536489    1.593661
       _cons |   91.85221   6.870993    13.37   0.000      78.3128    105.3916
------------------------------------------------------------------------------

. predict ehat, residual

. dwstat

Durbin-Watson d-statistic(  2,   228) =  .1032901
```

The OLS estimation results seem to be spurious because of the following signals: The t-ratio is very high, while the Durbin-Watson test statistic is very small (0.103). The graph of residuals from this regression (see below) show that the residuals seem to be non-stationary. The R^2 is low (0.213); this is not really a phenomenon of spurious regression. This can be a signal of positive autocorrelation. To be sure, we must apply the statistical tests.



5. Investigate whether *pwht* and *poil* are cointegrated using two different tests. What do you conclude? How do you judge the estimation results of question 4: as a spurious regression or as a cointegration relationship.

We will apply two different tests: (i) residual-based test for no cointegration; and (ii) CRDW test for no cointegration. Both tests check for the cointegration between poil and pwht. *poil* and *pwht* are cointegrated if the residuals of the above estimated model $(pwht = \alpha + \beta poil)$ are stationary process.

5.1 Residual-based test for no cointegration (Engle-Granger approach)

First test: Dickey Fuller test on the residuals

Null Hypothesis: the residuals are not stationary (no cointegration)

```
. dfuller ehat, trend regress lags(12)

Augmented Dickey-Fuller test for unit root         Number of obs   =       215

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.827            -4.002            -3.435            -3.135
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.1873
```

D.ehat	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ehat						
L1.	-.0773964	.027379	-2.83	0.005	-.1313849	-.0234078
LD.	.2910427	.0701934	4.15	0.000	.1526285	.4294568
L2D.	-.0697509	.0732856	-0.95	0.342	-.2142626	.0747608
L3D.	.1273172	.074107	1.72	0.087	-.0188141	.2734485
L4D.	-.1184446	.0749147	-1.58	0.115	-.2661686	.0292795
L5D.	.1097472	.0749241	1.46	0.145	-.0379953	.2574897
L6D.	.0659669	.0744537	0.89	0.377	-.080848	.2127818
L7D.	-.1186408	.0745215	-1.59	0.113	-.2655894	.0283078
L8D.	-.0266299	.0766743	-0.35	0.729	-.1778237	.1245638
L9D.	.0579581	.0791436	0.73	0.465	-.0981048	.2140211
L10D.	.0713729	.0790632	0.90	0.368	-.0845315	.2272774
L11D.	.1582098	.0814371	1.94	0.053	-.0023757	.3187953
L12D.	-.0394107	.0815571	-0.48	0.629	-.2002328	.1214115
_trend	-.0147413	.0121056	-1.22	0.225	-.0386122	.0091297
_cons	1.823866	1.586048	1.15	0.252	-1.303656	4.951388

As coefficients of trend and lag 12 are not significant, so we remove them from the test equation.

```
. dfuller ehat, regress lags(11)

Augmented Dickey-Fuller test for unit root         Number of obs   =       216

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.791            -3.471            -2.882            -2.572
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0596

------------------------------------------------------------------------------
      D.ehat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ehat |
         L1. |  -.0691085   .0247637    -2.79   0.006    -.1179355   -.0202814
         LD. |    .285142   .0698773     4.08   0.000     .1473636    .4229205
        L2D. |  -.0757421   .0727558    -1.04   0.299    -.2191961    .0677118
        L3D. |   .1204753   .0735272     1.64   0.103    -.0244996    .2654502
        L4D. |  -.1255037   .0738708    -1.70   0.091    -.2711562    .0201488
        L5D. |   .1083555   .0728394     1.49   0.138    -.0352633    .2519743
        L6D. |   .0589951    .073434     0.80   0.423     -.085796    .2037863
        L7D. |  -.1273148    .073971    -1.72   0.087    -.2731648    .0185352
        L8D. |  -.0265417   .0738726    -0.36   0.720    -.1721976    .1191143
        L9D. |   .0438534   .0779145     0.56   0.574    -.1097722     .197479
       L10D. |    .058593   .0779147     0.75   0.453     -.095033    .2122189
       L11D. |   .1360306    .078904     1.72   0.086     -.019546    .2916071
             |
       _cons |   .0752368   .6860246     0.11   0.913    -1.277411    1.427885
------------------------------------------------------------------------------
```

The test results indicate that the residuals seem to be not stationary at 5% significance level. However, the constant in this test equation is not significant, so we try to remove it from the test equation.

```
. dfuller ehat, regress lags(11) nocons

Augmented Dickey-Fuller test for unit root          Number of obs   =       216

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.805            -2.585            -1.950            -1.618
```

D.ehat	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ehat						
L1.	-.0692292	.0246792	-2.81	0.006	-.1178883	-.0205701
LD.	.2853661	.0696781	4.10	0.000	.1479845	.4227477
L2D.	-.0754884	.0725427	-1.04	0.299	-.2185179	.0675412
L3D.	.1208098	.0732858	1.65	0.101	-.023685	.2653045
L4D.	-.125184	.0736343	-1.70	0.091	-.270366	.0199979
L5D.	.1085571	.0726397	1.49	0.137	-.0346637	.2517778
L6D.	.0592136	.073229	0.81	0.420	-.0851692	.2035963
L7D.	-.1272013	.0737844	-1.72	0.086	-.2726792	.0182766
L8D.	-.026396	.0736816	-0.36	0.721	-.1716711	.118879
L9D.	.0442853	.0776263	0.57	0.569	-.1087674	.1973381
L10D.	.0590092	.0776336	0.76	0.448	-.0940579	.2120763
L11D.	.1362575	.0786857	1.73	0.085	-.0188839	.2913989

The test equation without the constant term shows that the residuals become stationary even at 1% significance level. This confused results might be due to the less power of test of the ADF test. To avoid this, we now apply the KPSS test.

```
. kpss ehat

KPSS test for ehat

Maxlag = 14 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: ehat is trend stationary

10%: 0.119  5% : 0.146  2.5%: 0.176  1% : 0.216

Lag order    Test statistic
    0           1.48
    1           .759
    2            .52
    3           .401
    4            .33
    5           .283
    6            .25
    7           .225
    8           .206
    9            .19
   10           .178
```

The KPSS test results we reject the null hypothesis that the residuals are stationary. Therefore, there seems to be no-cointegration between *pwht* and *poil*.

5.2 CRDW test for no cointegration

```
. regress pwht poil

      Source |       SS       df       MS              Number of obs =     228
-------------+------------------------------           F(  1,   226) =   61.51
       Model |  69028.652     1   69028.652           Prob > F      =  0.0000
    Residual | 253623.907   226  1122.22968           R-squared     =  0.2139
-------------+------------------------------           Adj R-squared =  0.2105
       Total | 322652.559   227  1421.37691           Root MSE      =    33.5

------------------------------------------------------------------------------
        pwht |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        poil |   1.273655    .162397     7.84   0.000     .9536489    1.593661
       _cons |   91.85221   6.870993    13.37   0.000      78.3128    105.3916
------------------------------------------------------------------------------

. predict ehat, residual

. dwstat

Durbin-Watson d-statistic(  2,   228) =  .1032901
```

The Durbin-Watson test statistic is 0.103, which is smaller than the 5% critical value CRDW tests for no cointegration (~ 0.2, about 200 observations, 2 variables, Table 9.3, Verbeek, 2012). Therefore, we fail to reject the null hypothesis that the residuals is non-stationary. In other words, *pwht* and *poil* are not cointegrated.

In conclusion, there is no long-run relationship between *pwht* and *poil*. Therefore, the OLS estimation regression between *pwht* and *poil* is likely to be spurious regression.

5. Based on your findings at the previous questions, estimate a simple VAR or VECM (this choice depends upon finding at question 5). Specify only one lag. Based on this estimated model indicate how *pwht* and *poil* relate.

Because *pwht* and *poil* are not cointegrated, so we cannot apply the VECM model. It is just possible to use the VAR model for the first-differenced series of *pwht* and *poil*.

$$\Delta pwht_t = \delta_1 + \theta_{11}\, \Delta pwht_{t-1} + \theta_{12}\, \Delta poil_{t-1} + \varepsilon_{1t}$$

$$\Delta poil_t = \delta_2 + \theta_{21}\, \Delta pwht_{t-1} + \theta_{22}\, \Delta poil_{t-1} + \varepsilon_{2t}$$

```
. varbasic D.pwht D.poil, lags(1/1)

Vector autoregression

Sample:  3 - 228                              No. of obs      =        226
Log likelihood = -1391.377                    AIC             =   12.36617
FPE            =  804.657                     HQIC            =   12.40281
Det(Sigma_ml)  =  763.0439                    SBIC            =   12.45698

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_pwht               3     10.1186   0.0754   18.43269   0.0001
D_poil               3     2.77306   0.0971   24.29502   0.0000
----------------------------------------------------------------
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
D_pwht						
pwht LD.	.2543129	.0641665	3.96	0.000	.128549	.3800769
poil LD.	.3006002	.2354381	1.28	0.202	-.16085	.7620503
_cons	-.0346056	.6687405	-0.05	0.959	-1.345313	1.276102
D_poil						
pwht LD.	-.0006592	.0175851	-0.04	0.970	-.0351254	.0338071
poil LD.	.3168833	.064523	4.91	0.000	.1904206	.4433459
_cons	.0026927	.1832716	0.01	0.988	-.3565131	.3618985

The VAR model results indicate that the p-values of coefficients of the variable *poilLD.* in the first equation (0.202) and of the variable *pwhtLD* in the second equation (0.97) are very high. These suggest that neither *poil* affects *pwht*, nor *pwht* affects *poil*. However, the cofficients of *pwhtLD* in the first equation, and *poilLD*. in the second equation are highly significant. These indicate that the first-differenced series follow the AR process.

EXAMPLE 2

In this example, we analyse the dataset **TEXASHOUSING.dta** with monthly housing prices in four major cities in Texas (USA): Austin, Dallas, Houston and San Antonio. Natural logarithms of housing prices are available from January 1990 till December 2003 (168 observations). It is expected that there are regional linkages between these housing markets. If houses get very expensive in one city, people may decide to move to another city, creating upward pressure on housing prices in our cities. In other words it is assumed that there exist a long-run (spatial) equilibrium between these four housing prices series. That is what we will investigate here.

1. **Investigate for all four prices series the order of integration.**

We only report the Augment Dickey Fuller test and KPSS test of the last step.

We recall that:

Null hypothesis in Augmented Dickey Fuller test: time series is not stationary

Null Hypothesis in KPSS test: the time series is stationary

AUSTIN

```
. dfuller austin, regress lag(10)

Augmented Dickey-Fuller test for unit root          Number of obs   =       157

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.554            -3.491            -2.886            -2.576
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.5065
```

D.austin	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
austin						
L1.	-.0202429	.0130231	-1.55	0.122	-.0459825	.0054968
LD.	-.5293985	.0815978	-6.49	0.000	-.6906732	-.3681238
L2D.	-.4988033	.0916595	-5.44	0.000	-.6799646	-.3176419
L3D.	-.2712327	.099817	-2.72	0.007	-.4685169	-.0739484
L4D.	-.3146437	.1015649	-3.10	0.002	-.5153826	-.1139048
L5D.	-.1246215	.1040904	-1.20	0.233	-.330352	.081109
L6D.	-.1908387	.1036336	-1.84	0.068	-.3956663	.0139889
L7D.	-.1752066	.0977011	-1.79	0.075	-.3683087	.0178956
L8D.	-.1561306	.0907467	-1.72	0.087	-.3354878	.0232266
L9D.	-.1048884	.0820093	-1.28	0.203	-.2669764	.0571997
L10D.	-.1474517	.0699294	-2.11	0.037	-.2856643	-.0092392
_cons	.2587338	.154914	1.67	0.097	-.0474474	.564915

As the absolute value of the test statistics (1.554) is smaller than the 5% critical value (2.886), we cannot reject the null hypothesis. The time series is not stationary.

```
. kpss austin

KPSS test for austin

Maxlag = 13 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: austin is trend stationary

10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order    Test statistic
     0           .54
     1          .347
     2          .267
     3          .221
     4           .19
     5          .167
     6          .149
     7          .135
     8          .125
     9          .116
    10          .109
    11          .102
    12         .0968
    13         .0923

. kpss austin, notrend

KPSS test for austin

Maxlag = 13 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: austin is level stationary

10%: 0.347   5% : 0.463   2.5%: 0.574   1% : 0.739

Lag order    Test statistic
     0          16.2
     1          8.22
     2          5.54
     3          4.19
     4          3.37
     5          2.83
     6          2.44
     7          2.15
     8          1.93
     9          1.74
    10           1.6
    11          1.47
    12          1.37
    13          1.28
```

The ADF test results indicate that housing price in Austin is non-stationary. However, the KPSS test indicates the non-stationarity (at 5% significance level) up to 5 lags only. When trying to use KPSS test without trend, the KPSS results turn out to provide a strong evidence of non-stationarity.

We now test the stationarity of the first-difference of the housing price in Austin.

```
. dfuller D.austin, regress lag(3)

Augmented Dickey-Fuller test for unit root         Number of obs   =       163

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)           -10.964            -3.489            -2.886            -2.576
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
   D2.austin |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      austin |
         LD. |  -2.421801   .2208801   -10.96   0.000     -2.85806   -1.985542
        LD2. |   .9347704   .1751922     5.34   0.000     .5887496    1.280791
       L2D2. |   .5136214   .1223197     4.20   0.000     .2720288     .755214
       L3D2. |   .2364427    .067627     3.50   0.001     .1028731    .3700122
             |
       _cons |   .0131558   .0034602     3.80   0.000     .0063216    .0199899
------------------------------------------------------------------------------
```

As the absolute value of the test statistics (10.964) is larger than the 5% critical value (2.886), we reject the null hypothesis. The time series of first differences is stationary.

```
. kpss D.austin

KPSS test for D.austin

Maxlag = 13 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: D.austin is trend stationary

10%: 0.119  5% : 0.146  2.5%: 0.176  1% : 0.216

Lag order    Test statistic
    0          .00751
    1           .0115
    2           .0165
    3           .0207
    4           .0279
    5           .0318
    6           .0335
    7            .038
    8           .0419
    9           .0443
   10           .0522
   11           .0533
   12            .051
   13           .0507
```

Both ADF and KPSS tests indicate that the first-differenced data of housing price in Austin is stationary. Therefore, the housing price in Austin is integrated of order on [$I(1)$].

DALLAS

```
. dfuller dallas, trend regress lag(10)

Augmented Dickey-Fuller test for unit root         Number of obs   =       157

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.792            -4.021            -3.442            -3.142
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.7085

------------------------------------------------------------------------------
D.dallas     |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dallas |
         L1. |   -.146736   .0818717    -1.79   0.075    -.3085615    .0150896
         LD. |    -.43757   .1050103    -4.17   0.000    -.6451309   -.2300092
        L2D. |   -.233615   .1039796    -2.25   0.026    -.4391385   -.0280915
        L3D. |  -.2844797   .0999016    -2.85   0.005    -.4819427   -.0870168
        L4D. |  -.1838233   .0958836    -1.92   0.057    -.3733444    .0056978
        L5D. |   -.268335   .0900026    -2.98   0.003    -.4462318   -.0904382
        L6D. |  -.3088513   .0867275    -3.56   0.001    -.4802746   -.1374279
        L7D. |  -.3587288   .0883186    -4.06   0.000    -.5332971   -.1841605
        L8D. |  -.2454506   .0883167    -2.78   0.006    -.4200152    -.070886
        L9D. |  -.2050867   .0855374    -2.40   0.018    -.3741578   -.0360157
       L10D. |  -.1745572   .0768247    -2.27   0.025    -.3264069   -.0227074
      _trend |   .0006188   .0003188     1.94   0.054    -.0000113     .001249
       _cons |   1.698752   .9423074     1.80   0.074    -.1637888    3.561294
------------------------------------------------------------------------------
```

As the absolute value of the test statistics (1.792) is smaller than the 5% critical value (3.442), we cannot reject the null hypothesis. The time series is not stationary.

```
. kpss dallas

KPSS test for dallas

Maxlag = 13 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: dallas is trend stationary

10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order    Test statistic
    0           1.58
    1           .939
    2            .69
    3           .563
    4           .484
    5           .434
    6           .398
    7           .371
    8           .348
    9           .327
   10           .308
   11           .289
   12           .272
   13           .256
```

Both ADF and KPSS tests indicate that the housing price in Dallas is non-stationary. We now test the stationarity of its first-differenced data.

```
. dfuller D.dallas, regress lag(12)

Augmented Dickey-Fuller test for unit root         Number of obs   =       154

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -5.683            -3.492            -2.886            -2.576
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
   D2.dallas |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dallas |
         LD. |  -4.654247   .8190314    -5.68   0.000    -6.273516   -3.034978
        LD2. |   3.050754   .7837061     3.89   0.000     1.501325    4.600183
       L2D2. |   2.642787   .7277283     3.63   0.000     1.204029    4.081545
       L3D2. |    2.20644   .6626479     3.33   0.001      .896349     3.51653
       L4D2. |   1.847904   .5933258     3.11   0.002     .6748668    3.020941
       L5D2. |   1.439479   .5255013     2.74   0.007     .4005351    2.478424
       L6D2. |   1.057621   .4565968     2.32   0.022      .154905    1.960338
       L7D2. |   .6548433   .3895031     1.68   0.095    -.1152252    1.424912
       L8D2. |   .3529426   .3249333     1.09   0.279     -.289468    .9953533
       L9D2. |   .1130304   .2644907     0.43   0.670    -.4098819    .6359427
      L10D2. |  -.1088403     .20048    -0.54   0.588    -.5052001    .2875195
      L11D2. |  -.2147775   .1389809    -1.55   0.125    -.4895503    .0599953
      L12D2. |  -.1436486   .0763759    -1.88   0.062    -.2946479    .0073507
             |
       _cons |   .0174704   .0039234     4.45   0.000     .0097137    .0252271
------------------------------------------------------------------------------
```

As the absolute value of the test statistics (5.683) is smaller than the 5% critical value (2.886), we reject the null hypothesis. The time series is stationary.

```
. kpss D.dallas

KPSS test for D.dallas

Maxlag = 13 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: D.dallas is trend stationary

10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order    Test statistic
    0          .00912
    1            .0135
    2            .0153
    3            .0184
    4            .0193
    5            .0227
    6             .027
    7            .0348
    8            .0418
    9            .0509
   10            .0627
   11            .0722
   12            .0695
   13            .0605
```

Both ADF and KPSS tests indicate that the first-differenced data of housing price in Dallas is stationary. Therefore, the housing price in Dallas is integrated of order on [$I(1)$].

HOUSTON

```
. dfuller houston, regress lag(11) nocons

Augmented Dickey-Fuller test for unit root         Number of obs   =       156

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)              5.368            -2.593            -1.950            -1.614

------------------------------------------------------------------------------
   D.houston |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     houston |
         L1. |   .0020627   .0003843     5.37   0.000     .0013032    .0028223
         LD. |  -.6934362   .0793837    -8.74   0.000    -.8503441   -.5365283
        L2D. |  -.3499416     .09481    -3.69   0.000    -.5373406   -.1625425
        L3D. |  -.5137478   .0965546    -5.32   0.000    -.7045953   -.3229003
        L4D. |  -.5202323   .0958276    -5.43   0.000    -.7096427   -.3308218
        L5D. |  -.5352645   .0993777    -5.39   0.000    -.7316921   -.3388369
        L6D. |  -.4398916   .1031044    -4.27   0.000    -.6436853    -.236098
        L7D. |  -.4007782   .1006201    -3.98   0.000    -.5996614    -.201895
        L8D. |  -.4875243   .0965438    -5.05   0.000    -.6783504   -.2966982
        L9D. |  -.2557234   .0956822    -2.67   0.008    -.4448464   -.0666005
       L10D. |  -.3193447   .0944027    -3.38   0.001    -.5059387   -.1327507
       L11D. |  -.3180061   .0790377    -4.02   0.000      -.47423   -.1617821
------------------------------------------------------------------------------

. kpss houston

KPSS test for houston

Maxlag = 13 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: houston is trend stationary

10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order    Test statistic
    0            1.2
    1           .801
    2           .598
    3           .501
    4           .439
    5           .398
    6           .365
    7           .339
    8           .317
    9           .296
   10           .278
   11           .261
   12           .244
   13            .23
```

Both ADF and KPSS tests indicate that the housing price in Houston is non-stationary. We now test the stationarity of its first-differenced data.

```
. dfuller D.houston, regress lag(10)

Augmented Dickey-Fuller test for unit root         Number of obs   =       156

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -8.323            -3.491            -2.886            -2.576
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
  D2.houston |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     houston |
         LD. |  -5.808183   .6978731    -8.32   0.000    -7.187581   -4.428784
        LD2. |    4.11735   .6614009     6.23   0.000     2.810042    5.424659
       L2D2. |   3.770199   .6064967     6.22   0.000     2.571412    4.968985
       L3D2. |   3.259248   .5457877     5.97   0.000     2.180458    4.338038
       L4D2. |   2.741879   .4861198     5.64   0.000     1.781026    3.702731
       L5D2. |   2.209587   .4208766     5.25   0.000     1.377693    3.041481
       L6D2. |   1.772522   .3491395     5.08   0.000     1.082421    2.462622
       L7D2. |   1.374148     .28023     4.90   0.000     .8202521    1.928043
       L8D2. |   .8887409   .2161295     4.11   0.000     .4615449    1.315937
       L9D2. |   .6348886   .1532458     4.14   0.000     .3319867    .9377905
      L10D2. |   .3170946    .079045     4.01   0.000     .1608562     .473333
             |
       _cons |   .0240483   .0044937     5.35   0.000     .0151661    .0329305
------------------------------------------------------------------------------
```

As the absolute value of the test statistics (8.323) is larger than the 5% critical value (2.886), we reject the null hypothesis. The time series of first differences is stationary.

```
. kpss D.houston

KPSS test for D.houston

Maxlag = 13 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: D.houston is trend stationary

10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order    Test statistic
    0          .00505
    1           .0103
    2           .0104
    3           .0135
    4           .0155
    5           .0198
    6           .0234
    7           .0259
    8           .0352
    9           .0362
   10           .0445
   11            .056
   12           .0495
   13           .0485
```

Both ADF and KPSS tests indicate that the first-differenced data of housing price in Houston is stationary. Therefore, the housing price in Houston is integrated of order on [$I(1)$].

SAN ANTONIO

```
. dfuller sa, trend regress lag(11)

Augmented Dickey-Fuller test for unit root         Number of obs   =       156

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.354            -4.021            -3.443            -3.143
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.4043

------------------------------------------------------------------------------
        D.sa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          sa |
         L1. |  -.3640841   .1546626    -2.35   0.020    -.6698228   -.0583455
         LD. |  -.2724428   .1565464    -1.74   0.084    -.5819055    .0370199
        L2D. |  -.3067109   .1523665    -2.01   0.046    -.6079107    -.005511
        L3D. |  -.2668833   .1445306    -1.85   0.067     -.552593    .0188265
        L4D. |  -.2654474   .1369049    -1.94   0.054    -.5360825    .0051878
        L5D. |  -.1484385   .1327877    -1.12   0.266    -.4109346    .1140577
        L6D. |  -.1266901   .1267752    -1.00   0.319    -.3773008    .1239206
        L7D. |  -.0668247   .1194138    -0.56   0.577    -.3028831    .1692338
        L8D. |  -.2136946   .1106637    -1.93   0.055    -.4324558    .0050665
        L9D. |  -.2532125   .1011489    -2.50   0.013    -.4531648   -.0532602
       L10D. |  -.1442041   .0894045    -1.61   0.109    -.3209398    .0325317
       L11D. |  -.2070834   .0766446    -2.70   0.008    -.3585952   -.0555715
      _trend |   .0012576   .0005553     2.26   0.025     .0001598    .0023554
       _cons |    4.10992   1.736144     2.37   0.019     .6778925    7.541948
------------------------------------------------------------------------------
```

```
. kpss sa

KPSS test for sa

Maxlag = 13 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: sa is trend stationary

10%: 0.119   5% : 0.146   2.5%: 0.176   1% : 0.216

Lag order    Test statistic
    0            .29
    1           .207
    2           .178
    3           .163
    4           .154
    5           .145
    6           .138
    7           .133
    8            .13
    9           .128
   10           .125
   11           .123
   12           .119
   13           .115

. kpss sa, notrend

KPSS test for sa

Maxlag = 13 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: sa is level stationary

10%: 0.347   5% : 0.463   2.5%: 0.574   1% : 0.739

Lag order    Test statistic
    0           14.9
    1           7.77
    2            5.3
    3           4.04
    4           3.27
    5           2.75
    6           2.38
    7            2.1
    8           1.88
    9           1.71
   10           1.57
   11           1.45
   12           1.35
   13           1.26
```

The ADF test results indicate that housing price in San Antonio is non-stationary. However, the KPSS test indicates the non-stationarity (at 5% significance level) up to 5 lags only. When trying to use KPSS test without trend, the KPSS results turn out to provide a strong evidence of non-stationarity. We now test the stationarity of the first-difference of the housing price in San Antonio.

```
. dfuller D.sa, regress lag(10)

Augmented Dickey-Fuller test for unit root         Number of obs   =       156

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -7.745            -3.491            -2.886            -2.576
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
        D2.sa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           sa |
          LD. |  -5.255941   .6785985    -7.75   0.000    -6.597242    -3.91464
         LD2. |   3.666462   .6430824     5.70   0.000     2.395361    4.937562
        L2D2. |   3.072758   .5921327     5.19   0.000     1.902363    4.243153
        L3D2. |   2.550517    .536451     4.75   0.000     1.490181    3.610852
        L4D2. |   2.058796    .478953     4.30   0.000      1.11211    3.005483
        L5D2. |   1.708308   .4137822     4.13   0.000     .8904367     2.52618
        L6D2. |   1.401999   .3464126     4.05   0.000     .7172881    2.086709
        L7D2. |   1.174429   .2752468     4.27   0.000     .6303827    1.718475
        L8D2. |   .8229273   .2062843     3.99   0.000     .4151908    1.230664
        L9D2. |   .4641046   .1384334     3.35   0.001     .1904806    .7377286
       L10D2. |   .2494536   .0754464     3.31   0.001     .1003282     .398579
              |
        _cons |   .0200444   .0046761     4.29   0.000     .0108017     .029287
------------------------------------------------------------------------------
```

```
. kpss D.sa

KPSS test for D.sa

Maxlag = 13 chosen by Schwert criterion
Autocovariances weighted by Bartlett kernel

Critical values for H0: D.sa is trend stationary

10%: 0.119  5% : 0.146  2.5%: 0.176  1% : 0.216

Lag order    Test statistic
    0          .00568
    1          .00786
    2           .0107
    3            .014
    4           .0181
    5           .0209
    6           .0237
    7            .024
    8            .028
    9           .0358
   10           .0397
   11           .0503
   12           .0511
   13           .0511
```

Both ADF and KPSS tests indicate that the first-differenced data of housing price in San Antonio is stationary. Therefore, the housing price in San Antonio is integrated of order on [$I(1)$].

In conclusion, all housing prices in these cities are integrated of the same order one. Therefore, there could be cointegrating relationships among these housing prices.

2. For the variables that have the same order of integration, use Johansen's rank and maximum eigenvalues tests to investigate the rank of the cointegrating matrix. Set the number of lags to 3. The maximum eigenvalues test can be obtain at the reporting tab. What do you conclude from these tests?

All the variables have the same order of integration (1), so they could be all cointegrated. In order to check that we perform two cointegration tests (Johansen tests for cointegration). Both test check the rank of the cointegrating matrix (matrix that contains the coefficients of cointegrating relationships). The number of cointegration relationships is equal to the rank of the matrix. The number of cointegration relationships cannot be larger than the number of variables minus one (in this case cannot be larger than 3).

Null hypothesis in both test: $r \leq r_0$ where r is the rank of the matrix of cointegrating relationships.

```
. vecrank austin dallas houston sa, trend(constant) lag(3) max

                     Johansen tests for cointegration
Trend: constant                                    Number of obs =     165
Sample:  1990m4 - 2003m12                                   Lags =       3
-------------------------------------------------------------------------------
                                                         5%
maximum                                     trace    critical
  rank    parms      LL       eigenvalue  statistic    value
    0      36    1107.7833         .       101.6070    47.21
    1      43    1137.7484     0.30456      41.6768    29.68
    2      48    1153.6435     0.17524       9.8865*   15.41
    3      51    1158.4191     0.05624       0.3354     3.76
    4      52    1158.5868     0.00203
-------------------------------------------------------------------------------
                                                         5%
maximum                                      max     critical
  rank    parms      LL       eigenvalue  statistic    value
    0      36    1107.7833         .        59.9302    27.07
    1      43    1137.7484     0.30456      31.7903    20.97
    2      48    1153.6435     0.17524       9.5511    14.07
    3      51    1158.4191     0.05624       0.3354     3.76
    4      52    1158.5868     0.00203
-------------------------------------------------------------------------------
```

In the trace test I reject the null hypothesis up to the rank of two (test statistics 9.88 smaller than 15.41). The eigenvalues test rejects the null hypothesis up to the rank of 2 (test statistics 9.55 smaller than 14.07). So I conclude that the rank of the cointegrating matrix is two; This implies that there exist two long-run relationships among housing prices in these cities.

3. Image that all variables are *I*(0). Explain how this would be reflected in the output of Johansen's rank and maximum eigenvalues tests.

If four variables were stationary at their original data [*I*(0)], the cointegrating matrix would have full rank (r_0 = 4). In the output of Johansen's rank and maximum eigenvalue tests, we should have found a rank of 4.

4. Estimate a VECM with the appropriate number of cointegrating relations (again set lag at 3). What are the long-run cointegrating relationships? Why are some of the adjustment parameters not significant for these cointegrating relations?

```
. vec austin dallas houston sa, trend(constant) rank(2) lag(3)

Vector error-correction model

Sample:  1990m4 - 2003m12                       No. of obs      =        165
                                                AIC             =  -13.40174
Log likelihood =  1153.644                      HQIC            =  -13.03496
Det(Sigma_ml)  =  9.93e-12                      SBIC            =  -12.49819

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_austin             11     .047995   0.2875   61.72788   0.0000
D_dallas             11     .038872   0.1954   37.15132   0.0001
D_houston            11     .042854   0.4676   134.3761   0.0000
D_sa                 11     .052892   0.3141   70.05815   0.0000
----------------------------------------------------------------
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
D_austin						
_ce1						
L1.	-.1480888	.0591085	-2.51	0.012	-.2639393	-.0322383
_ce2						
L1.	-.0402385	.1221438	-0.33	0.742	-.279636	.1991589
austin						
LD.	-.3678125	.086496	-4.25	0.000	-.5373415	-.1982836
L2D.	-.2289378	.0788415	-2.90	0.004	-.3834644	-.0744112
dallas						
LD.	.2717324	.1336168	2.03	0.042	.0098484	.5336165
L2D.	.0974838	.1153095	0.85	0.398	-.1285187	.3234862
houston						
LD.	.0536793	.1235346	0.43	0.664	-.188444	.2958027
L2D.	-.0437925	.0912195	-0.48	0.631	-.2225794	.1349945
sa						
LD.	-.2761228	.0817978	-3.38	0.001	-.4364436	-.1158019
L2D.	-.1549178	.0760511	-2.04	0.042	-.3039752	-.0058605
_cons	.0094995	.0038546	2.46	0.014	.0019448	.0170543

D_dallas						
_ce1						
L1.	.0728342	.0478729	1.52	0.128	-.0209949	.1666634
_ce2						
L1.	-.3089124	.0989262	-3.12	0.002	-.5028042	-.1150207
austin						
LD.	-.0007882	.0700544	-0.01	0.991	-.1380923	.136516
L2D.	.0149861	.063855	0.23	0.814	-.1101674	.1401397
dallas						
LD.	-.1582048	.1082183	-1.46	0.144	-.3703089	.0538992
L2D.	.012165	.093391	0.13	0.896	-.1708779	.195208
houston						
LD.	-.1811501	.1000526	-1.81	0.070	-.3772496	.0149494
L2D.	-.067546	.0738801	-0.91	0.361	-.2123484	.0772563
sa						
LD.	.1287132	.0662494	1.94	0.052	-.0011331	.2585596
L2D.	.0998227	.061595	1.62	0.105	-.0209012	.2205466
_cons	.0041647	.0031219	1.33	0.182	-.001954	.0102835

D_houston						
_ce1						
L1.	.1900478	.0527777	3.60	0.000	.0866053	.2934903
_ce2						
L1.	.6039838	.1090617	5.54	0.000	.3902267	.8177409
austin						
LD.	-.0719686	.0772319	-0.93	0.351	-.2233403	.0794032
L2D.	-.0486423	.0703973	-0.69	0.490	-.1866185	.0893339
dallas						
LD.	-.2710895	.1193059	-2.27	0.023	-.5049248	-.0372543
L2D.	-.2203616	.1029594	-2.14	0.032	-.4221583	-.0185648
houston						
LD.	-.1341298	.1103036	-1.22	0.224	-.3503208	.0820612
L2D.	.13256	.0814496	1.63	0.104	-.0270782	.2921982
sa						
LD.	.2176672	.073037	2.98	0.003	.0745173	.3608171
L2D.	.0781981	.0679057	1.15	0.249	-.0548946	.2112909
_cons	.0032671	.0034417	0.95	0.342	-.0034785	.0100128

D_sa						
_ce1						
L1.	.2826268	.0651401	4.34	0.000	.1549546	.410299
_ce2						
L1.	-.1783621	.1346077	-1.33	0.185	-.4421883	.085464
austin						
LD.	-.1212452	.0953222	-1.27	0.203	-.3080733	.0655829
L2D.	.1312372	.0868867	1.51	0.131	-.0390577	.3015321
dallas						
LD.	.2753613	.1472514	1.87	0.061	-.0132461	.5639687
L2D.	.0731444	.127076	0.58	0.565	-.1759199	.3222087
houston						
LD.	.0359778	.1361404	0.26	0.792	-.2308524	.302808
L2D.	.0904846	.1005278	0.90	0.368	-.1065462	.2875155
sa						
LD.	-.176293	.0901447	-1.96	0.051	-.3529734	.0003874
L2D.	-.138557	.0838115	-1.65	0.098	-.3028245	.0257105
_cons	.0017073	.0042479	0.40	0.688	-.0066184	.010033

```
Cointegrating equations

Equation           Parms    chi2      P>chi2
_ce1                 2    586.3044   0.0000
_ce2                 2    2169.826   0.0000

Identification:  beta is exactly identified

                 Johansen normalization restrictions imposed

        beta |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
_ce1         |
      austin |          1          .        .      .            .           .
      dallas |  (omitted)
     houston |  -.2623782   .1893625    -1.39  0.166    -.6335219    .1087655
          sa |  -1.241805    .229643    -5.41  0.000    -1.691897   -.7917128
       _cons |   5.577099          .        .      .            .           .
_ce2         |
      austin |  (omitted)
      dallas |          1          .        .      .            .           .
     houston |  -1.095652   .0669898   -16.36  0.000     -1.22695   -.9643545
          sa |   .2883986   .0812396     3.55  0.000     .1291718    .4476253
       _cons |  -2.351372          .        .      .            .           .
```

From the cointegrating equations results, (based on the significance of the estimated coefficients) we realize that there are two long-run cointegrating relationships between/among house prices of: (i) Austin and San Antonio; and (ii) Dallas, Houston, and San Antonio.

The adjustment parameters in the VECM model can be summarized as followed:

	Adj.Parameter	Coef.	P-value	Significance
D_austin	γ_{11}	-0.148	0.012	Yes
	γ_{12}	-0.040	0.742	No
D_dallas	γ_{21}	0.073	0.128	No

	γ_{22}	-0.309	0.002	Yes
D_houston	γ_{31}	0.190	0.000	Yes
	γ_{32}	0.604	0.000	Yes
D_sa	γ_{41}	0.283	0.000	Yes
	γ_{41}	-0.178	0.185	No

- For Austin: The adjustment parameter of the second cointegrating relation is not significant because Austin is omitted in this relation (see _ce2 in cointegrating equations).
- For Dallas: The adjustment parameter of the first cointegrating relation is not significant because Dallas is omitted in this relation (see _ce1 in the cointegrating equations).
- For Houston: Both adjustment parameters are highly significant because Houston exists in both relations (see _ce1 and _ce2 in the cointegrating equations).
- For San Antonia: The adjustment parameter of the second cointegrating relation is not significant (although it is included in both the cointegrating equations) because of lag selection (maybe). Say, when we change from lag(3) to lag(4), both adjustment parameters become significant at 5% significance level.

```
vec austin dallas houston sa, trend(constant) rank(2) lag(4)
```

D_sa						
_ce1						
L1.	.2480569	.0769452	3.22	0.001	.097247	.3988667
_ce2						
L1.	-.3265809	.156033	-2.09	0.036	-.6323999	-.0207619
austin						
LD.	-.1418052	.1027894	-1.38	0.168	-.3432688	.0596584
L2D.	.0572965	.1003203	0.57	0.568	-.1393277	.2539207
L3D.	-.1284946	.0892846	-1.44	0.150	-.3034891	.0464999
dallas						
LD.	.4500373	.1608877	2.80	0.005	.1347032	.7653714
L2D.	.2585537	.1521923	1.70	0.089	-.0397377	.5568451
L3D.	.2271957	.1272675	1.79	0.074	-.022244	.4766354
houston						
LD.	-.0150679	.1530394	-0.10	0.922	-.3150197	.2848839
L2D.	.0587159	.1359957	0.43	0.666	-.2078307	.3252626
L3D.	-.0507752	.1007152	-0.50	0.614	-.2481734	.146623
sa						
LD.	-.2797918	.0978819	-2.86	0.004	-.4716368	-.0879467
L2D.	-.2672216	.0944004	-2.83	0.005	-.4522429	-.0822003
L3D.	-.1552788	.0846439	-1.83	0.067	-.3211777	.0106202
_cons	.0028339	.0042404	0.67	0.504	-.0054771	.011145

Cointegrating equations

Equation	Parms	chi2	P>chi2
_ce1	2	693.5721	0.0000
_ce2	2	1907.879	0.0000

Identification: beta is exactly identified
Johansen normalization restrictions imposed

beta	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_ce1						
austin	1	.	.	.	.	.
dallas	(dropped)					
houston	-.5103764	.1733395	-2.94	0.003	-.8501155	-.1706372
sa	-.8806267	.2108319	-4.18	0.000	-1.29385	-.4674038
_cons	4.307426	.	.	.	.	.
_ce2						
austin	(dropped)					
dallas	1	.	.	.	.	.
houston	-1.098766	.0750085	-14.65	0.000	-1.24578	-.9517523
sa	.2814363	.0912325	3.08	0.002	.102624	.4602486
_cons	-2.235019	.	.	.	.	.

If the lags are 4, we can see that there are two long-run cointegrating relationships between/among house prices of: (i) Austin, Houston and San Antonio; and (ii) Dallas, Houston, and San Antonio.