Lecture 8: Nonstationarity, Unit Roots and Cointegration

R.G. Pierse

1 Introduction

Definition 1.1. *Weak stationarity*

A variable Y_t is weakly stationary if its mean and its variance are constant over time, and its autocovariances $\text{cov}(Y_t Y_{t-s})$ *are a function solely of s and not of t.*

The assumption of stationarity is necessary for econometric estimators and tests to have the standard distributions. Most economic variables do not satisfy the conditions of weak stationarity. In this case they need to be transformed in order to make them stationary.

1.1 The Autoregressive model

Consider the autoregressive model from lecture 6 with the addition of an intercept term:

$$Y_t = c + \rho Y_{t-1} + u_t \quad , \quad t = -\infty, \cdots, 0, 1, \cdots, T \tag{1.1}$$

where

$$-1 < \rho < 1\,.$$

Note that instead of defining an initial condition Y_0 as previously, we now assume that the process has been running indefinitely from $t = -\infty$. The error process u_t retains all the usual assumptions of the classical model. The parameter ρ is known as the *root* of the autoregressive process. Substituting for lagged Y_t, we get

$$Y_t = c(1 + \rho + \rho^2 + \rho^3 + \cdots) + u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \rho^3 u_{t-3} + \cdots$$

so that

$$\begin{aligned} \text{E}(Y_t) &= c(1 + \rho + \rho^2 + \rho^3 + \cdots) \\ &= \frac{c}{1-\rho} \end{aligned}$$

and

$$\begin{aligned}\text{cov}(Y_t Y_{t-s}) &= \sigma^2\rho^s(1+\rho^2+\rho^4+\rho^6+\cdots) \\ &= \frac{\sigma^2\rho^s}{1-\rho^2} \quad , \quad s \geq 0 \; .\end{aligned}$$

This model satisfies the conditions of weak stationarity since both mean and autocovariances are constant over time.

1.2 The Random Walk Model

Now consider what happens to the properties of the model if $\rho = 1$. We now revert to assuming that the process started at time $t = 0$ with fixed initial condition $Y_0 = 0$. Then

$$Y_t = c + Y_{t-1} + u_t \quad , \quad t = 1, \cdots, T \tag{1.2}$$

so that, substituting for lagged Y_t

$$Y_t = tc + \sum_{i=1}^{t} u_i + Y_0 \; .$$

In this model $\text{E}(Y_t) = tc$ so that it is not constant over time and

$$\text{E}(Y_t^2) = \sum_{i=1}^{t} \text{E}(u_i^2) = t\sigma^2 \; .$$

so that the variance increases as t increases. This process (1.2) is thus *no longer stationary* and is known as a *random walk model with drift.* The drift term is the intercept c and if $c = 0$ then the process is known as a *pure random walk model without drift.* In this case, the mean is constant although the variance still increases over time. Note that if we had assumed that the random walk process had been running indefinitely since $t = -\infty$, then the mean and the variance would be *infinite.* The random walk model is an example of a *unit root* process because the dynamic root ρ in the autoregressive model takes the value of unity.

Note that the random walk process can be made stationary by transforming the dependent variable by *first differencing* since

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

is a stationary process.

A series that can be made stationary by differencing is said to be *integrated*, or to possess a *unit root.*

Definition 1.2. *A time series Y_t is integrated of order d, denoted $I(d)$, if $\Delta^d Y_t$ is stationary. The series Y_t is said to have d unit roots.*

1.3 Two Alternative Models

1.3.1 Deterministic Trend

$$Y_t = \gamma + \beta t + u_t \quad , \quad u_t \sim iid(0, \sigma^2) \tag{1.3}$$

This model has a *non-constant* mean, and a *constant* variance. Stationarity caan be achieved by detrending.

1.3.2 Random walk with drift

$$Y_t = c + Y_{t-1} + u_t \quad , \quad u_t \sim iid(0, \sigma^2) \tag{1.4}$$

This model has a *non-constant* mean, and a *non-constant* variance. Stationarity can be achieved by first differencing.

These two models have very different implications for the effect of a one-off shock to u_t occurring in period t^*. In the deterministic trend model, the shock has no long run effect and the process continues on its former path in period $t^* + 1$. In the random walk with drift model however, the process switches to a new path and the effect of the shock is permanent.

2 Testing for Unit Roots

We want to be able to discriminate between the two alternative models (1.3) and (1.4). Because the model (1.4) is non-stationary, however, we cannot use standard t-tests to test the hypothesis that $\rho = 1$. Instead, new tests with non-standard distributions have had to be developed.

The framework for testing unit roots is the model

$$Y_t = \gamma + \beta t + u_t \tag{2.1}$$

and

$$u_t = (1 - \alpha)u_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma^2)\,. \tag{2.2}$$

A unit root test is then a test of the null of

$$H_0 : \alpha = 0 \quad \text{against} \quad H_1 : \alpha > 0$$

in this model.

Note that on H_0 :

$$\Delta Y_t = \beta + \varepsilon_t$$

which is the random walk with drift model, whereas, on H_1:

$$Y_t = (1 - \alpha)Y_{t-1} + \alpha\beta t + (\gamma\alpha + (1 - \alpha)\beta) + \varepsilon_t \tag{2.3}$$

or

$$\Delta Y_t = -\alpha Y_{t-1} + bt + c + \varepsilon_t \tag{2.4}$$

which is an autoregressive model (with autoregressive root $\rho = (1-\alpha)$) with intercept and deterministic trend.

2.1 The Dickey-Fuller (DF) Test

Fuller (1976) and Dickey and Fuller (1979, 1981) propose a test based on the t-ratio $t(\alpha)$ in the OLS regression (2.4). The distribution of this statistic is *non-standard* and depends on the presence of the *nuisance parameters*, β and γ. Two special cases need to be considered: (i) $\beta = 0 \Rightarrow b = 0$ and (ii) $\beta = \gamma = 0 \Rightarrow b = c = 0$. In the case (i), there is no drift term under the null H_0 and no trend term under the alternative H_1. In the case (ii) there is no drift term under the null H_0 and neither intercept nor trend term under the alternative H_1.

The general case and the two special cases have different distributions and so have different critical values. Critical values of the statistic for all three cases are given in Fuller (1976) Table 8.5.2 and in Banerjee *et al.* (1993).

2.2 The Durbin-Watson test

Sargan and Bhargava (1983) develop an alternative test for a unit root based on the Durbin-Watson statistic in the equation (2.3). They show that, on the null hypothesis of a unit root, then

$$DW \to 0$$

and construct a test using this statistic. Note that although this test uses the same statistic as the Durbin-Watson test *it is a completely different test* with a different null hypothesis. In particular it is designed for equations with a lagged dependent variable where the conventional Durbin-Watson test is not valid. Critical values for the test are given in Table 1 of Sargan and Bhargava (1983).

2.3 Dealing with autocorrelation

The tests in the previous section are based on the assumption that ε_t is 'white noise' *i.e. serially uncorrelated.* If ε_t is serially correlated then the serial correlation needs to be corrected before the unit root test is performed.

2.3.1 The Augmented Dickey-Fuller Test

Assume that the serial correlation in ε_t can be represented by an $AR(p)$ process. Then it can be corrected by adding the p lagged terms $\Delta Y_{t-1}, \cdots, \Delta Y_{t-p}$ to the regression (2.4) to give

$$\Delta Y_t = -\alpha Y_{t-1} + bt + c + \gamma_1 \Delta Y_{t-1} + \cdots + \gamma_p \Delta Y_{t-p} + \varepsilon_t \, . \tag{2.5}$$

The distribution of the test statistic is unaffected by the addition of these lagged differences.

3 Spurious Regression

Suppose that two variables Y_t and X_t both follow random walk processes with drift:

$$Y_t = c + Y_{t-1} + u_t \quad , \quad u_t \sim iid(0, \sigma^2) \tag{3.1}$$

and

$$X_t = d + X_{t-1} + u_t \quad , \quad v_t \sim iid(0, \omega^2) \tag{3.2}$$

where the error processes u_t and v_t are uncorrelated.

Suppose that an investigator runs the regression

$$Y_t = a + bX_t + \eta_t \, . \tag{3.3}$$

What would we expect to happen? We might expect to find that $\widehat{b}$ was insignificant since the two processes X and Y are completely unrelated. However, Granger and Newbold (1974) showed that in fact this regression will most likely produce a significant coefficient for $\widehat{b}$ and a very high explanatory power R^2. It will also have a very low DW statistic. Such regressions are known as *spurious regressions.* The reason for the misleading results is that in general the conventional statistical tests are just not valid in a regression with non-stationary variables since the error process η_t will also be non-stationary which violates the basic assumptions.

How can spurious regressions be spotted? Granger and Newbold suggest using a rule of thumb that $R^2 > DW$ is indication of spurious regression. However, a better strategy for avoiding spurious regression is to test the order of integration of all series in a regression and, if they are found to be integrated, then test for cointegration among the variables.

4 Cointegration

Suppose that Y_{1t}, Y_{2t}, $\cdots$, Y_{kt} are a set of $I(1)$ variables. In general, any linear combination of them such as

$$\sum_{i=1}^{k} w_i Y_{it}$$

will also be $I(1)$ for all set of weights $w_i \neq 0$. However, suppose there exists some linear combination such that

$$\sum_{i=1}^{k} w_i^* Y_{it} \text{ is } I(0) \quad , \quad w_i^* \neq 0\,.$$

Then we say that the variables Y_{it} are *cointegrated* and that the weights w_i^* form a *cointegrating vector.*

Definition 4.1. *If*

$$Y_{it} \sim I(d) \quad \textit{and} \quad \sum_{i=1}^{k} w_i^* Y_{it} \;\sim\; I(d-b) \quad , \quad w_i^* \neq 0$$

then

$$\{Y_{1t}, Y_{2t}, \cdots, Y_{kt}\} \sim CI(d,b) \quad , \quad d \geq b > 0\,.$$

There can be r different cointegrating vectors, where $0 \leq r < k$. Note that r must be less than the number of variables k.

4.1 The Meaning of Cointegration

If $I(1)$ variables are cointegrated, this means that although they are individually non-stationary, they are moving together so that there is some long run relationship between them. Consider again the static equation between two $I(1)$ variables which now may possibly be cointegrated:

$$Y_t = a + bX_t + \eta_t\,. \tag{4.1}$$

. If Y_t and X_t are *not* cointegrated then there is no possible value of the parameters a and b such that η_t can be stationary. If they are cointegrated however, then there is a *single* value for the two parameters such that the linear combination $Y_t - a + bX_t$ is stationary. This is when the parameters are the weights of a cointegrating vector. For this unique value of the parameters, (4.1) is a valid econometric equation with stationary error term η_t. It represents the long run equilibrium relationship between the two variables and this can only exist when there is cointegration.

Cointegration can thus be seen as the existence of a long run relationship between variables and economic theory leads us to expect that cointegration should exist. Cointegration is a long run property of variables. In the short-run, the variables can be moving in different ways, driven by different dynamic processes. However, cointegration ties the variables together in the long run.

4.2 Testing Cointegration

If a set of variables are cointegrated, then the residuals from a static regression of any one of the variables on all the others will be stationary. If not, then the residuals will be integrated. Thus Dickey-Fuller tests on the *OLS* residuals e_t from a static regression provide a way of testing cointegration. This was proposed by Engle and Granger (1987).

The critical values will be different from those from the standard Dickey-Fuller tests because e_t is based on *estimated* parameters. The null hypothesis in the test is that $e_t \sim I(1)$, i.e. *zero cointegrating vectors*, and the alternative is that $e_t \sim I(0)$, i.e. *one cointegrating vector*. Critical values for the *ADF* tests are given in MacKinnon (1991).

The unit root Durbin-Watson test can also be used to test cointegration in the residuals from a static regression and is described in Sargan and Bhargava (1983).

5 Error Correction Mechanisms

The cointegrating vector represents the long run relationship between two cointegrated variables. What about the short-run relationship? Granger and Engle (1987) show that this can be represented by an error correction model or *ECM*.

Suppose that two $I(1)$ variables Y_t and X_t are cointegrated with cointegrating relationship

$$\eta_t = Y_t - a - bX_t \,. \tag{5.1}$$

Then the short run relationship can be represented by

$$\Delta Y_t = a_0 + a_1 \Delta Y_{t-1} + \cdots + a_p \Delta Y_{t-p} + b_0 \Delta X_t + \cdots + b_l \Delta X_{t-l} + \gamma \eta_{t-1} + u_t \tag{5.2}$$

which is an *ECM* representation. Note that all the terms in the representation (5.2) are $I(0)$ so that the coefficients in the equation will all have standard distributions.

The *ECM* can be given an economic interpretation as an adjustment mechanism whereby deviations from the equilibrium relationship in the previous period, as measured by η_{t-1}, lead to adjustments in Y_t. This is the reason why it is known as an error correction mechanism.

It can be shown that the *ECM* representation (5.2) is simply a reparameterisation of the general dynamic model

$$Y_t = a_0 + a_1 Y_{t-1} + \cdots + a_{p+1} Y_{t-p-1} + \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_{l+1} X_{t-l-1} + u_t \tag{5.3}$$

but one that makes explicit the long-run cointegrating relationship (5.1) and which is expressed entirely in terms of stationary variables.

References

[1] Banerjee, A., J. Dolado, J.W. Galbraith and D.F. Hendry (1993), *Cointegration, Error-Correction, and the Analysis of Non-stationary Data*, Oxford University Press, Oxford.

[2] Dickey, D.A. and W.A. Fuller (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association*, 74, 427–431.

[3] Dickey, D.A. and W.A. Fuller (1981), 'Likelihood ratio statistics for autoregressive time series with a unit root', *Econometrica*, 49, 1057–1072.

[4] Engle, R.F. and C.W. J. Granger (1987), 'Cointegration and error correction: representation, estimation and testing', *Econometrica*, 55, 251–276.

[5] Engle, R.F. and C.W.J. Granger (1991), *Long Run Economic Relationships: Readings in Cointegration*, Oxford University Press, Oxford.

[6] Fuller, W.A. (1976), *Introduction to Statistical Time Series*, Wiley, New York.

[7] Granger, C.W.J. and P. Newbold (1974), 'Spurious regressions in econometrics', *Journal of Econometrics*, 2, 111–120.

[8] MacKinnon, J.G. (1991), 'Critical values of cointegration tests', in R.F. Engle and C.W.J. Granger (eds.) *Long Run Economic Relationships: Readings in Cointegration*, Oxford University Press, Oxford.

[9] Sargan, J.D. and A.S. Bhargava (1983), 'Testing residuals from least squares regression for being generated by the Gaussian random walk', *Econometrica*, 51, 153–174.