14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
Novemeber 1, 2007

Lecture 20

Cointegration

We think, or at least we cannot reject the null hypothesis, that many macro series have unit roots. For example, log consumption and log output are both non-stationary. At the same time often we can argue that some linear combination of them is stationary,say, the difference between log consumption and log output is stationary. This situation is called cointegration. We will see on today lecture that asymptotic behavior of estimates change completely depending on whether the series are cointegrated. The practical problem will be though that we often do not have enough data to definitively tell whether or not we have cointegrated series.

Multi-dimensional Random Walk

I start with summarizing multivariate convergence results. I am not providing proofs, but they can be easily found in the literature. Each of the statements are just multi-dimensional variants of single dimension results that we've already seen.

Let ϵ_t be i.i.d. $k \times 1$-vectors with zero mean, $E[\epsilon_t \epsilon_t'] = I_k$, and finite fourth moments. Then,

$$\xi_T(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[\tau T]} \epsilon_t \Rightarrow W$$

where W is a k-dimensional Brownian motion. We also want to allow for serial correlation, so we need to look at the behavior of

$$v_t = F(L)\epsilon_t \ , \sum i|F_i| < \infty$$

The longterm variance of v_t will be $F(1)F(1)'$. Finally, let

$$\eta_t = \eta_{t-1} + v_t$$

We will use the following results:

(a) $\frac{\eta_{[T\tau]}}{\sqrt{T}} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} v_t \Rightarrow F(1)W(\tau)$

(b) $\frac{1}{T} \sum \eta_{t-1}\epsilon_t' \Rightarrow F(1) \int_0^1 W(t)dW(t)'$

(c) $\frac{1}{T^{3/2}} \sum \eta_t \Rightarrow F(1) \int W dt$

(d) $\frac{1}{T^2} \sum \eta_t \eta_t' \Rightarrow F(1) \int W W' dt F(1)'$

(e) $\frac{1}{T^{5/2}} \sum t\eta_t \Rightarrow F(1) \int tW(t)dt$

Intro to Cointegration

Last time we started the discussion of OLS estimates on persistent regressors. We've seen on a case of spurious regression that if there is no cointegration, then the estimate is not consistent, and $|t| \to^p \infty$. Now we consider an example of regression with persistent regressors when there is a cointegration.

Let us have two random walks, x_t and y_t, such that a linear combination of them is stationary. This situation is called co-integration.

$$\begin{cases} y_t = \beta x_t + e_t \\ x_t = x_{t-1} + u_t \end{cases} ; \quad (e_t, u_t) \sim i.i.d.$$

Assume that $Ee_t^2 = Eu_t^2 = 1; \text{cov}(e_t, u_t) = \phi$, and forth moments are finite. For simplicity assume that $e_t = \phi u_t + \sqrt{1-\phi^2} z_t$, and apply convergence above to $\varepsilon_t = (u_t, z_t)'$.

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \begin{pmatrix} u_t \\ z_t \end{pmatrix} \Rightarrow \begin{pmatrix} W_1(\tau) \\ W_2(\tau) \end{pmatrix}$$

here W_1 and W_2 are two independent Brownian motions.

We are interested in OLS estimator of β in a regression of y_t on x_t. Notice that x_t is not exogenous (correlated with the error term in the first regression!). Despite this correlation (and contrary to your cross-section-OLS intuition), β will be consistently estimated. Moreover, $\hat{\beta}$ is super-consistent.

$$T(\hat{\beta} - \beta) = \frac{\frac{1}{T}\sum x_t e_t}{\frac{1}{T^2}\sum x_t^2} = \frac{\frac{1}{T}\sum x_{t-1} e_t + \frac{1}{T}\sum u_t e_t}{\frac{1}{T^2}\sum x_t^2}$$

From what we know the denominator converges to $\int_0^1 W_1^2 dt$, the second term in the numerator satisfies the Law of Large Numbers. The statement (b) implies:

$$\frac{1}{T}\sum x_{t-1} e_t = \frac{\phi}{T}\sum x_{t-1} u_t + \frac{\sqrt{1-\phi^2}}{T}\sum x_{t-1} z_t \Rightarrow \phi \int_0^1 W_1 dW_1 + \sqrt{1-\phi^2} \int_0^1 W_1 dW_2$$

We conclude that

$$T(\hat{\beta} - \beta) \Rightarrow \frac{\phi \int W_1 dW_1 + \sqrt{1-\phi^2} \int W_1 dW_2 + \phi}{\int W_1^2 dt}.$$

Notice that if $\phi = 0$ (x_t is exogenous), then the limiting distribution $T(\hat{\beta} - \beta) \Rightarrow \frac{\int W_1 dW_2}{\int W_1^2 dt}$ is centered at zero and free of nuisance parameters. However, since $\phi \neq 0$, the limiting distribution is shifted. As a result, $\hat{\beta}$ has a finite sample bias of order $\frac{1}{T}$, which could be large. In addition, one can show that the limit of the t-statistic would also depend on ϕ, and thus, not be free of nuisance parameters. This makes inference difficult.

Regression with persistent regressors.

Three facts from the example above may look strange at first: 1) the rate of convergence on persistent regressor is stronger; 2) if regressor is persistent small endogenuity does not preclude us from having a consistent estimator; 3) behavior for cointegrated case is strickingly different from no-cointegration.

Let us consider a more general regression that may have different stochastic and non-stochastic trends. Let us write the model as

$$y_t = \gamma z_t + e_t$$

where $z_t = [z_{1t}\, z_{2t}\, z_{3t}\, z_{4t}]$ where z_{1t} is zero-mean and stationary, z_{2t} is a constant, z_{3t} is $I(1)$, and $z_{4t} = t$. We can write z_t as

$$z_t = \begin{pmatrix} F_1(L) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ F_2(L) & G & H & 0 \\ F_3(L) & \mathcal{T} & k & 1 \end{pmatrix} \begin{pmatrix} \epsilon_t \\ 1 \\ \eta_t \\ t \end{pmatrix}$$

where ϵ_t (i.i.d) and η_t (non-stationary process) are defined in the first section of this lecture. Assume that e_t is a stationary process (we may think of it as a component of v_t). The regression of such a for is called canonical.

We show that the coefficients on the different components of z_t will converge at different rates. $\widehat{\gamma}$ is the OLS estimate. We are interested in the asymptotic behavior of

$$\hat{\gamma} - \gamma = (z'z)^{-1} z'e$$

We need some normalization for this to converge to a distribution. Our normalizing matrix will be:

$$Q = \begin{pmatrix} \sqrt{T} I_{k_1} & 0 & 0 & 0 \\ 0 & \sqrt{T} & 0 & 0 \\ 0 & 0 & T I_{k_3} & 0 \\ 0 & 0 & 0 & T^{3/2} \end{pmatrix},$$

here $k_1 = dim(z_{1t}); k_3 = dim(z_{3t})$ We look at:

$$Q(\hat{\gamma} - \gamma) = (Q^{-1} z'zQ^{-1})^{-1} Q^{-1} z'e$$

Consider the first part:

$$Q^{-1}z'zQ^{-1} \Rightarrow \begin{pmatrix} \text{const } \; k_1 \times k_1\text{-matrix} & 0 \quad 0 \quad 0 \\ \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} & \begin{pmatrix} \text{The elements of this block converge} \\ \text{to functions of Brownian motions. The exact} \\ \text{form of it is not important for the current point} \end{pmatrix} \end{pmatrix}$$

Explanation:

- the convergence in the upper-left corner is due to Law of Large Numbers (applied to stationary process);
- $\frac{1}{T} \sum z_{1t} \to^p E z_{1t} = 0$ due to LLN and z_1 is mean zero;
- $\frac{1}{T} \sum z_{1t} z_{3t}$ is a sum of the form described in (b) so it converges to some linear function of $\int W(t) dW(t)'$; then, $\frac{1}{T^{3/2}} \sum z_{1t} z_{3t} \to^p 0$.
- one can heck using Chebyshev's inequality that $\frac{1}{T^2} \sum z_{1t} t \to^p 0$
- you should check also that the normalization is right for all other elements of low-left matrix.

The important thing is that the matrix has a block diagonal structure, so we can look at the blocks separately when inverting. Now, let's look at the second part. We assume that z_{1t} is exogenous, that is, $E(e_t | z_{1t}, z_{1,t-1} ...) = 0$. We do not assume that z_{3t} is exogenous though(it may be that innovations to z_{3t} and error term e_t are correlated).

$$Q^{-1} z'e = \begin{pmatrix} \frac{1}{\sqrt{T}} \sum z_{1t} e_t \\ \frac{1}{\sqrt{T}} \sum e_t \\ \frac{1}{T} \sum z_{3t} e_t \\ \frac{1}{T^{3/2}} \sum t e_t \end{pmatrix} \Rightarrow \begin{pmatrix} N(0, 1?) \\ N(0, 2?) \\ 3?(\int W dW) 4? \\ N(0, 5?) \end{pmatrix}$$

1? is used for the long-run variance of $z_{1t}e_t$; 2? - for the long-run variance of e_t; 5? - for the limit variance (a stronger Limit Theorem is needed here!); ?3 and ?4 denote some linear combination of stochastic integrals. It is also known (Chen & Wei) that the first component is independent of the others.

Result :

$$\sqrt{T}(\hat{\gamma}_1 - \gamma_1) \Rightarrow N(0, v)$$

So $\widehat{\gamma}_1$ asymptotically a normal distribution at the usual rate, and the OLS t-statistic for $\hat{\gamma}_1$ is asymptotically $N(0,1)$. All the other coefficients will converge to non-standard distributions at different rates. In particular, the speeds of convergence are:

$$\begin{aligned}
\sqrt{T}(\hat{\gamma}_2 - \gamma_2) &\Rightarrow \text{something 2} \\
T(\hat{\gamma}_3 - \gamma_3) &\Rightarrow \text{something 3} \\
T^{3/2}(\hat{\gamma}_4 - \gamma_4) &\Rightarrow \text{something 4}
\end{aligned}$$

Now let us consider a different regression

$$y_t = \beta x_t + e_t$$

where e_t is stationary, and regressors x_t may contain stationary and non-stationary components in different combinations. We can find a linear transformation $Dx_t = z_t$, that transform our regression into a canonical form ($z_t = [z_{1t}\, z_{2t}\, z_{3t}\, z_{4t}]$ is described above). Sims showed that such a decomposition always exists. Then

$$y_t = \gamma z_t + e_t$$

where $\gamma = \beta D^{-1}$.

So, what does this tell us about $\hat{\beta}$? We know that $\hat{\beta} = D\hat{\gamma}$. A component of $\hat{\beta}$ that is a linear combination of $\hat{\gamma}_1$, $\hat{\gamma}_3$, and $\hat{\gamma}_4$, and its distribution will be dominated by the behavior of $\hat{\gamma}_1$ (due to its slower convergence compared to the others). So, these components of $\hat{\beta}$ will be asymptotically normal and converge at $\sqrt{T}$. In particular, as long as we include a constant among the regressors, **the coefficients that can be represented as coefficients on the stationary regressors will be asymptotically normal and converge at rate $\sqrt{T}$.**

Example: Testing for Granger Causality

Suppose we have two series, y_1 and y_2. We want to test whether y_2 Granger causes y_1. We estimate

$$y_{1t} = \alpha + \sum_{i=1}^{p} \phi_i y_{1t-i} + \sum_{i=1}^{p} \alpha_i y_{2t-i} + \epsilon_t \tag{1}$$

and test $H_0 : \alpha_1 = ... = \alpha_p = 0$. Cases:

1. y_1 is $I(1)$, y_2 is stationary. Then the Wald stat $\Rightarrow \chi^2_p$, since $z_{1t} = y_{2t}$, $z_{1t} = 1$, and $z_{3t} = y_{1t}$

2. y_1 is $I(1)$, y_2 is $I(1)$, and cointegrated, i.e. a linear combination, $w_t = y_{2t} - \lambda y_{1t}$ is stationary. Then we can rewrite the regression as:

$$y_{1t} = \alpha + \sum_{i=1}^{p} \tilde{\phi}_i y_{1t-i} + \sum_{i=1}^{p} \alpha_i w_{t-i} + \epsilon_t \tag{2}$$

   where $\tilde{\phi}_i = \phi_i + \lambda\alpha_i$. Then applying the results from above, α_i converge to standard distributions, and so does the Wald statistic. Note that we do not need to actually know w_t and estimate (2). We estimate (1) and just need to know that (2) exists, i.e. y_1 and y_2 are cointegrated for standard asymptotics to apply.

3. y_1 and y_2 are $I(1)$, but not cointegrated. Then in general, we cannot transform α to coefficients on stationary regressors. As a result, the Wald statistic will not converge to a standard χ^2 distribution.

The lesson here is that it is important to know both whether or not series have unit roots and whether or not they are cointegrated.

Application: Permanent Income Hypothesis

Mankiw and Shapiro (1985) tested the permanent income hypothesis by regressing

$$\Delta c_t = \mu + \pi y^d_{t-1} + \delta t + \epsilon_t$$

where Δc_t is the change in consumption and y^d_t is disposable income. The null hypothesis is $H_0 : \pi = 0$. However, disposable income has a unit root, so the t-statistic associated with π has a non-standard distribution. We can get around this problem by estimating:

$$c_t = \mu + \alpha c_{t-1} + \pi y^d_{t-1} + \delta t + \epsilon_t$$

Under the null hypothesis c_t and y^d_t are cointegrated, so π will have a standard asymptotic distribution. Stock and West (1988) pointed this out. This version of the test is easier to carry out in practice, but it is not clear which version of the test is more powerful. Asymptotically, the first estimate of π is super-consistent, but it has a finite sample bias of order $\frac{1}{T}$. Because of this, the second version tends to perform better in small samples.