14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
October 23, 2007
revised October 22, 2009

Lecture 17

Unit Roots

Review from last time

Let y_t be a random walk

$$y_t = \rho y_t + \epsilon_t \, , \rho = 1$$

where ϵ_t is a martingale difference sequence ($E[\epsilon_t|I_t] = 0$), with $\frac{1}{T}\sum E[\epsilon_t^2|I_{t-1}] \to \sigma^2$ a.s. and $E\epsilon_t^4 < K < \infty$. Then $\{\epsilon_t\}$ satisfy a functional central limit theorem,

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

We showed last time that:

$$\begin{pmatrix} \frac{1}{T}\sum y_{t-1}\epsilon_t \\ \frac{1}{T^{3/2}}\sum y_{t-1} \\ \frac{1}{T^2}\sum y_{t-1}^2 \end{pmatrix} \Rightarrow \begin{pmatrix} \sigma^2\int_0^1 W(t)dW(t) \\ \sigma\int_0^1 W(t)dt \\ \sigma^2\int_0^1 W(t)^2 dt \end{pmatrix}$$

and our OLS estimator and t-statistic have non-standard distributions:

$$\begin{aligned} T(\hat{\rho}-\rho) \Rightarrow & \frac{\int W dW}{\int W^2 dt} \\ t \Rightarrow & \frac{\int W dW}{\sqrt{\int W^2 dt}} \end{aligned}$$

This is quite different from the case with $|\rho| < 1$. If

$$x_t = \rho x_{t-1} + \epsilon_t \, , |\rho| < 1$$

then,

$$\begin{pmatrix} \frac{1}{\sqrt{T}}\sum x_{t-1}\epsilon_t \\ \frac{1}{T}\sum x_{t-1} \\ \frac{1}{T}\sum x_{t-1}^2 \end{pmatrix} \Rightarrow \begin{pmatrix} N(0, \frac{\sigma^4}{1-\rho^2}) \\ Ex_t = 0 \\ Ex_t^2 = \frac{\sigma^2}{1-\rho^2} \end{pmatrix}$$

and the OLS estimate and t-stat converge to normal distributions:

$$\begin{aligned} \sqrt{T}(\hat{\rho}-\rho) \Rightarrow & N(0, (1-\rho^2)) \\ t \Rightarrow & N(0,1) \end{aligned}$$

Adding a constant

Suppose we add a constant to our OLS regression of y_t. This is equivalent to running OLS on demeaned y_t. Let $y_t^m = y_t - \bar{y}$. Then,

$$\hat{\rho} - 1 = \frac{\sum y_{t-1}^m \epsilon_t}{\sum (y_{t-1}^m)^2}$$

Consider the numerator:

$$\begin{aligned} \frac{1}{T} \sum y_{t-1}^m \epsilon_t =& \frac{1}{T} \sum (y_{t-1} - \bar{y}) \epsilon_t \\ =& \frac{1}{T} \sum y_{t-1} \epsilon_t - \bar{y}\bar{\epsilon} \end{aligned}$$

$\bar{y}\bar{\epsilon} = \frac{1}{T^{3/2}} \sum y_{t-1} \frac{1}{\sqrt{T}} \epsilon_t$. We know that $\frac{1}{T^{3/2}} \sum y_{t-1} \Rightarrow \sigma \int W dt$, and $\frac{1}{\sqrt{T}} \sum \epsilon_t \Rightarrow \sigma W(1)$. Also, from before we know that $\frac{1}{T} \sum y_{t-1} \epsilon_t \Rightarrow \sigma^2 \int W dW$. Combining this, we have:

$$\frac{1}{T} \sum y_{t-1}^m \epsilon_t \Rightarrow \sigma^2 \left(\int W(s) dW(s) - W(1) \int W(t) dt \right)$$

We can think of this as the integral of a demeaned Brownian motion,

$$\begin{aligned} \int W(s) dW(s) - W(1) \int W(t) dt =& \int \left(W(s) - \int W(t) dt \right) dW(s) \\ =& \int W^m(s) dW(s) \end{aligned}$$

The limiting distributions of all statistics ($\hat{\rho}$, t, etc) would change. In most cases, the change is only in replacing W by W^m. One also can include a linear trend in the regression, then the limiting distribution would depend on a detrended Brownian motion.

Limiting Distribution

Let's return to the case without a constant. The limiting distribution of the OLS esitmate is $T(\hat{\rho} - 1) \Rightarrow \frac{\int W dW}{\int W^2 dt}$. This distribution is skewed and shifted to the left. If we include a constant, the distribution is even more shifted. If we also include a trend, the distribution shifts yet more. For example, the 2.5%-tile without a constant is -10.5, with a constant is -16.9, and with a trend is -25. The 97.5%-tiles are 1.6, 0.41, and -1.8, respectively. Thus, estimates of ρ have bias of order $\frac{1}{T}$. This bias is quite large in small samples.

The distribution of the t-statistic is also shifted to the left and skewed, but less so than the distribution of the OLS estimate. The 2.5%-tiles are -2.23, -3.12, -3.66 and of the 97.5%-tiles are 1.6, 0.23, and -0.66 for no constant, with a constant, and with a trend, respectively.

Allowing Auto-correlation

So far, we have assumed that ϵ_t are not auto-correlated. This assumption is too strong for empirical work. Let,

$$y_t = y_{t-1} + v_t$$

where v_t is a zero-mean stationary process with an MA representation,

$$v_t = \sum_{j=0}^{\infty} c_j \epsilon_j$$

where ϵ_j is white noise. Now we need to look at the limiting distributions of each of the terms we looked at before ($\sum v_t$, $\sum y_{t-1}v_t$, $\sum y_t^2$, etc). $\sum v_t$ is a term we already encountered in HAC estimation:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_t \Rightarrow N(0,\omega^2)$$

where ω^2 is the long-term variance,

$$\omega^2 = \sum_{j=-\infty}^{\infty} \gamma_j = \sigma^2 c(1)^2 \equiv \omega^2$$

For linearly dependent process, v_t, we have the following central limit theorem:

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

This can be proven by using the Beveridge-Nelson decomposition. All the other terms converge to things similar to the uncorrelated case, except with σ replaced by ω. For example,

$$\frac{y}{\sqrt{T}} = \xi_T(t/T) \Rightarrow \omega W(t/T)$$
$$\frac{1}{T^{3/2}}\sum y_{t-1} \Rightarrow \omega \int W(s)ds$$

An exception is:

$$\begin{aligned}\frac{1}{T}\sum y_{t-1}v_t =& \frac{1}{2T}y_T^2 - \frac{1}{2T}\sum v_t^2 \\ \Rightarrow& \frac{1}{2}(\omega^2 W(1)^2 - \sigma_v^2) \\ \Rightarrow& \omega^2 \int W dW + \frac{\omega^2 - \sigma_v^2}{2}\end{aligned}$$

This leads to an extra constant in the distribution of $\hat{\rho}$:

$$T(\hat{\rho}-1) \Rightarrow \frac{\int W dW - \frac{1}{2}\frac{\omega^2-\sigma_v^2}{\omega^2}}{\int W^2 dt}$$

This additional constant is "nuisance" parameter, and thus, the statistic is *not pivotal*. So it is impossible to just look up critical values in a table. Phillips & Perron (1988) suggested corrected statistics:

$$T(\hat{\rho}-1) + \frac{\frac{1}{2}\hat{\omega}^2 - \hat{\sigma}_v^2}{\frac{1}{T^2}\sum y_t^2} \Rightarrow \frac{\int W dW}{\int W^2 dt}$$

where $\hat{\omega}^2$ is an estimate of the long-run variance, say Newey-West, and $\hat{\sigma}_v^2$ is the variance of the residuals.

Augmented Dickey Fuller

Another approach is the augmented Dickey-Fuller test. Suppose,

$$y_t = \sum_{j=1}^{p} a_j y_{t-j} + \epsilon_t$$

where $z = 1$ is a root,

$$1 - \sum_{j=1}^{p} a_j z^j = 0$$
$$1 = \sum a_j$$

Suppose we factor $a(L)$ as

$$1 - \sum^{p} a_j L^j = (1 - L)b(L)$$

with $b(L) = 1 - \sum_{j=1}^{p-1} \beta_j L^j$. We can write the model as

$$\Delta y_t = \sum_{j=1}^{p-1} \beta_j \Delta y_{t-j} + \epsilon_t$$

This suggests estimating:

$$y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \beta_j \Delta y_{t-j} + \epsilon_t$$

and testing whether $\rho = 1$ to see if we have a unit root. This is another way of allowing auto-correlation because we can write the model as:

$$a(L)y_t = \epsilon_t$$
$$b(L)\Delta y_t = \epsilon_t$$
$$\Delta y_t = b(L)^{-1}\epsilon_t$$
$$y_t = y_{t-1} + v_t$$

where $v_t = b(L)^{-1}\epsilon_t$ is auto-correlated.

The nice thing about the augemented Dickey-Fuller test is that the coefficients on Δy_{t-j} each converge to a normal distribution at rate $\sqrt{T}$, and the coefficient on y_{t-1} converges to a non-standard distribution without nuisance parameters. To see this, let $x_t = [y_{t-1}, \Delta y_{t-1}, ..., \Delta y_{t-p+1}]$ and let $\theta = [\rho, \beta_1, ..., \beta_{p-1}]'$. Consider: $\hat{\theta} - \theta = (X'X)^{-1}(X'\epsilon)$. We need to normalize this so that it converges. Our normalizing matrix will be

$$Q = \begin{bmatrix} T & ... & 0 \\ 0 & \sqrt{T} & 0 \\ \vdots & \ddots & \\ & 0 & \sqrt{T} \end{bmatrix}$$

Multiplying by Q gives:

$$Q(\hat{\theta} - \theta) = (Q^{-1}(X'X)^{-1}Q^{-1})^{-1}(Q^{-1}X'\epsilon)$$

The denominator is:

$$Q^{-1}(X'X)^{-1}Q^{-1} = \begin{pmatrix} \frac{1}{T^2}\sum y_{t-1}^2 & \frac{1}{T^{3/2}}\sum y_{t-1}\Delta y_{t-1} & ... \\ \frac{1}{T^{3/2}}\sum y_{t-1}\Delta y_{t-1} & \frac{1}{T}\tilde{X}'\tilde{X} & \\ \vdots & & \end{pmatrix}$$

Note that $\frac{1}{T^{3/2}}\sum y_{t-1}\Delta y_{t-j} \xrightarrow{p} 0$ (since we proved that $\frac{1}{T}\sum y_{t-1}\Delta y_{t-j} \Rightarrow \frac{1}{2}(\omega^2 W(1)^2 - \sigma_v^2))$. Also, we know that $\frac{1}{T^2}\sum y_{t-1}^2 \Rightarrow \omega^2 \int W^2 dt$, so

$$Q^{-1}(X'X)^{-1}Q^{-1} \Rightarrow \begin{pmatrix} \omega^2 \int W^2 dt & 0 & \dots \\ 0 & E\tilde{X}'\tilde{X} & \\ \vdots & & \end{pmatrix}$$

Similarly,

$$Q^{-1}X'\epsilon = \begin{pmatrix} \frac{1}{T}\sum y_{t-1}\epsilon_t \\ \frac{1}{\sqrt{T}}\sum \Delta y_{t-j}\epsilon_t \end{pmatrix} \Rightarrow \begin{pmatrix} \sigma\omega \int W dW \\ N(0, E[\tilde{X}'\tilde{X}]\sigma^2) \end{pmatrix}$$

Thus,

$$T(\hat{\rho}-1) \Rightarrow \frac{\sigma}{\omega}\frac{\int W dW}{\int W^2 dt}$$

or

$$T(\hat{\rho}-1)\frac{\hat{\omega}}{\hat{\sigma}} = \frac{T(\hat{\rho}-1)}{1-\sum \hat{\beta}_j} \Rightarrow \frac{\int W dW}{\int W^2 dt}$$

Other Tests

Sargan-Bhargrava (1983) For a non-stationary process, the variance of y_t grows with time, so we could look at:

$$\frac{\frac{1}{T^2}\sum y_{t-1}^2}{\frac{1}{T}\sum \Delta y_{t-1}^2}$$

when ϵ_t is an m.d.s., this converges to a distribution, $\frac{\frac{1}{T^2}\sum y_{t-1}^2}{\frac{1}{T}\sum \Delta y_{t-1}^2} \Rightarrow \int W^2(dt)$. For a stationary process, this converges to 0 in probability.

Range (Mandelbrot and coauthors) You could also look at the range of y_t, this should blow up for non-stationary processes for a random walk, we have:

$$\frac{\frac{1}{\sqrt{T}}(\max y_t - \min y_t)}{\sqrt{\frac{1}{T}\sum \Delta y_t^2}} \Rightarrow \sup_{\lambda} W(\lambda) - \inf_{\lambda} W(\lambda)$$

Test Comparison

The Sargan-Bhargrava (1983) test has optimal asymptotic power against local alternatives, but has bad size distortions in small samples, especially if errors are negatively autocorrelated. The augmented Dickey-Fuller test with the BIC used to choose lag-length has overall good size in finite samples.