14.384 Time Series Analysis
Fall 2008

14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
October 25, 2007

Lecture 14

More Non-Stationarity

We have seen that there's a discrete difference between stationarity and non-stationarity. When we have a non-stationary process, limiting distributions are quite different from in the stationary case. For example, let ϵ_t be a martingale difference sequence, with $E(\epsilon_t^2|I_{t-1}) = 1$, $E\epsilon_t^4 < k < \infty$. Then $\xi_T(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[\tau T]} \epsilon_t \Rightarrow W(\cdot)$. Then there is a sort of discontinuity in the limiting distribution of an $AR(1)$ at $\rho = 1$:

	Unit Root	Stationary
Process	$y_t = y_{t-1} + \epsilon_t$	$x_t = \rho x_{t-1} + \epsilon_t$
Limiting distribution of ρ	$T(\hat{\rho} - 1) \Rightarrow \frac{\int W dW}{\int W^2 dt}$	$T(\hat{\rho} - \rho) \Rightarrow N(0, 1 - \rho^2)$
Limiting distribution of t	$t \Rightarrow \frac{\int W dW}{\sqrt{\int W^2 dt}}$	$t \Rightarrow N(0, 1)$

In finite samples, the distribution of the t-stat is continuous in $\rho \in [0, 1]$. However, the limit distribution is discontinuous at $\rho = 1$. This must mean that the convergence is not uniform. In particular, the convergence of the t-stat to a normal distribution is slower, the closer ρ is to 1. Thus, in small samples, when ρ is close to 1, the normal distribution badly approximates the unknown finite sample distribution of the t-statistic. A more precise statement is that we have pointwise convergence, *i.e.*

$$\sup_x |P(t(\rho, T) \leq x) - \Phi(x)| \to 0 \ \forall \rho < 1$$

but not uniform convergence, *i.e.*

$$\sup_{\rho \in (0,1)} \sup_x |P(t(\rho, T) \leq x) - \Phi(x)| \not\to 0$$

where $\Phi(\cdot)$ is the normal cdf.

Just how bad is the normal approximation? If you construct a 95% confidence interval based on a normal approximation, then without a constant the coverage is 90%, with a constant 70%, and with a linear treand 35%.

Local to Unity Asymptotics

Local to unity asymptotics is one way to try to construct a better approximation. Let:

$$\begin{aligned} x_t &= \rho x_{t-1} + \epsilon_t \ , \ t = 1, ..., T \\ \rho &= \exp(c/T) \approx 1 + c/T \ , \ c < 0 \end{aligned}$$

This model is not meant to be a literal way of describing the world. It is just a device for building a better approximating limiting distribution. It can be shown that:

$$\frac{x_{[\tau T]}}{\sqrt{T}} \Rightarrow \Im_c(\tau) \tag{1}$$

where $\Im_c(\tau)$ is an Ornstein-Ulenbeck process.

Definition 1. *Ornstein-Ulenbeck* process: $\Im_c(\tau) = \int_0^\tau e^{c(\tau - s)} dW(s)$, so $\Im_c(\tau) \sim N(0, \frac{e^{2\tau c} - 1}{2c})$

We will not prove (1), but we will sketch the idea. First, observe that

$$\begin{aligned}\frac{x_t}{\sqrt{T}} &= \sum_{j=1}^{t} \rho^j \frac{\epsilon_j}{\sqrt{T}} \\ &= \sum_{j=1}^{t} e^{c(t/T-j/T)} \frac{\epsilon_j}{\sqrt{T}}\end{aligned}$$

Defining $\xi_T(\tau)$ as usual we have:

$$\frac{x_t}{\sqrt{T}} = \sum_{j=1}^{t} e^{c(t/T-j/T)} \Delta\xi_T(j/T)$$

then taking $\tau = t/T$ we have:

$$\frac{x_{[\tau T]}}{\sqrt{T}} = \int_0^\tau e^{c(\tau-s)} d\xi_T(s) ds$$

Finally, assuming convergence of the stochastic integral (which we could prove if we took care of some technical details), gives:

$$\frac{x_t}{\sqrt{T}} \Rightarrow \int_0^\tau e^{c(\tau-s)} dW(s) ds \equiv \Im_c(\tau)$$

Using this result, the limiting distribution of OLS will be (omitting several technical steps):

$$\begin{aligned}T(\hat{\rho}-\rho) &\Rightarrow \frac{\int \Im_c(s) dW(s)}{\int \Im_c^2(s) ds} \\ t_{\rho=e^{c/T}} &\Rightarrow t^c = \frac{\int \Im_c(s) dW(s)}{\sqrt{\int \Im_c^2(s) ds}}\end{aligned}$$

If $c = 0$, t^c is a Dickey-Fuller distribution. If $c \to -\infty$, the $t^c \Rightarrow N(0,1)$. This was shown by Phillips (1987).

The convergence to this distribution is uniform (Mikusheva (2007)),

$$\sup_{\rho\in[0,1]} \sup_x |P(t(\rho,T) \leq x) - P(t^c \leq x | \rho = e^{c/T})| \to 0 \text{ as } T \to \infty$$

Figure 1 illustrates this convergence

Confidence Sets

We usually construct confidence sets by inverting a test. Consider testing $H_0 : \rho = \rho_0$ vs $\rho \neq \rho_0$. We construct a confidence set as $C(x) = \{\rho_0 : \text{hypothesis accepted}\}$. So, for example in OLS, we take $t = \frac{\hat{\rho}-\rho}{s.e.(\hat{\rho})}$ and

$$\begin{aligned}C(x) =& \{\rho_0 : -1.96 \leq \frac{\hat{\rho}-\rho}{se(\hat{\rho})} \leq 1.96\} \\ =& [\hat{\rho} - 1.96 se(\hat{\rho}), \hat{\rho} + 1.96 se(\hat{\rho})]\end{aligned}$$

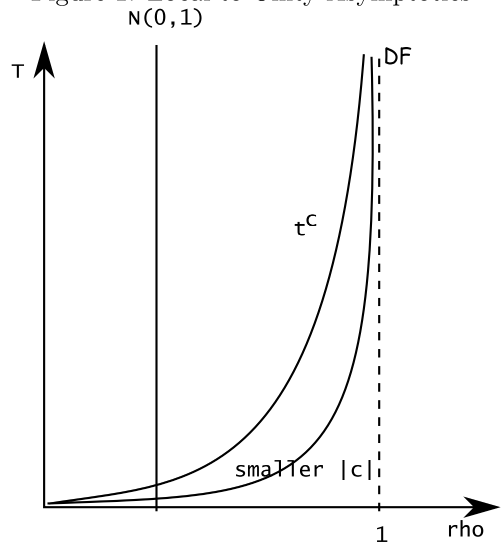
To construct confidence sets using local to unity asymptotics, we do the exact same thing, except the quantiles of our limiting distribution depend on ρ_0, *i.e.*

$$C(x) = \{\rho_0 : q_1(\rho_0, T) \leq \frac{\hat{\rho}-\rho}{se(\hat{\rho})} \leq q_1(\rho_0, T)\}$$

where $q_1(\rho_0, T)$ and $q_2(\rho_0, T)$ are quantiles of t^c for $c = T \log \rho_0$.

This approach was developed by Stock (1991). It only works when we have an $AR(1)$ with no autocorrelation. Some correction could be done in $AR(p)$ to construct a confidence set for the largest autoregressive root.

Figure 1: Local to Unity Asymptotics



Grid Bootstrap

This was an approach developed by Hansen (1999). It has a local to unity interpretation. Suppose

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

where ρ will be the sum of AR coefficients; it is a measure of persistence. For the grid bootstrap we:

- Choose grid on $[0,1]$
- Test $H_0 : \rho = \rho_0$ vs $\rho \neq \rho_0$ for each point on grid
  1. Regress x_t on x_{t-1} and $\Delta x_{t-1}, ...\Delta x_{t-p+1}$ to get $\hat{\rho}$, t_{ρ_0}-stat
  2. Regress $x_t - \rho_0 x_{t-1}$ on $\Delta x_{t-1}, ...\Delta x_{t-p+1}$ to get $\hat{\beta}_j$
  3. Bootstrap:
     - ϵ_t^* from residuals of step 1
     - Form $x_t^* = \rho_0 x_{t-1}^* + \sum \hat{\beta}_j \Delta x_{t-j} + \epsilon_t^*$ do OLS as in step 1
     - Repeat, use quantiles of bootstrapped t-stats as critical values to form test
- All ρ_0 for which the hypothesis is accepted form a confidence set

Bayesian Perspective

From a Bayesian point of view, there is nothing special about unit roots if one assumes a flat prior. Sims and Uhlig (1991) argue that all the attention paid to unit roots is non-productive. Phillips (1991) has a reply that looks more carefully at the idea of uninformative priors. Sims and Uhlig (1991) had put a uniform prior on $[0,1]$. Phillips points out that this puts all weight on the stationary case. He argues that a uniform prior is not necessarily uninformative, and point out that a Jeffreys prior would put much more weight (asymptotically almost unity weight) on the non-stationary case. In this case Bayesian conclusions look more like frequentists'. There is a Journal of Applied Econometrics issue about this debate.