14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
October 18, 2007

Lecture 16

Empirical Processes

Introduction

References: Hamilton ch 17, Chapters by Stock and Andrews in Handbook of Econometrics vol 4

Empirical process theory is used to study limit distributions under non-standard conditions. Applications include:

1. Unit root, cointegration and persistent regressors. For example if $y_t = \rho y_{t-1} + e_t$, with $\rho = 1$, then $T(\hat{\rho} - 1)$ converges to some non-standard distribution
2. Structural breaks at unknown date (testing with nuisance parameters). For example, $y_t = \begin{cases} \mu + e_t & t \leq \tau \\ \mu + k + e_t & t > \tau \end{cases}$.
   Want to test H_0: no break $k = 0$ with τ being unknown. A test statistic for this hypothesis is $S = \max_\tau |t_\tau|$ where t_τ is the t-statistic for testing $k = 0$ with the break at time τ. S will have a non-standard distribution.
3. Weak instruments & weak GMM.
4. Simulated GMM with non-differentiable objective function – e.g. Berry & Pakes
5. Semi-parametrics

We will discuss 1 & 2. We will cover simulated GMM later.

Empirical Process Theory

Let x_t be a real-valued random $k \times 1$ vector. Consider some $\mathbb{R}^n$ valued function $g_t(x_t, \tau)$ for $\tau \in \Theta$, where Θ is a subset of some metric space. Let

$$\xi_T(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (g_t(x_t, \tau) - Eg_t(x_t, \tau))$$

$\xi_T(\tau)$ is a random function; it maps each $\tau \in \Theta$ to an $\mathbb{R}^n$ valued random variable. $\xi_T(\tau)$ is called an empirical process. Under very general conditions (some limited dependence and enough finite moments), standard arguments (like Central Limit Theorem) show that $\xi_T(\tau)$ converges point-wise, i.e. $\forall \tau_0 \in \Theta$, $\xi_T(\tau_0) \Rightarrow N(0, \sigma^2(\tau_0))$. Also, standard arguments imply that on a finite collection of points, $(\tau_1, ..., \tau_p)$,

$$\begin{bmatrix} \xi_T(\tau_1) \\ \vdots \\ \xi_T(\tau_p) \end{bmatrix} \Rightarrow N(0, \Sigma(\tau_1, ..., \tau_p)) \tag{1}$$

We would like to generalize this sort of result so that we talk about the convergence of $\xi_T()$ as a random function in a functional space. For that we have to introduce a metric in a space of right-continuous functions.

We define a *metric* for functions on Θ as $d(b_1, b_2) = \sup_{\tau \in \Theta} |b_1(\tau) - b_2(\tau)|$. Let $\mathcal{B} =$ be a space of bounded functions on Θ, and $\mathcal{U}(\mathcal{B}) =$ be a class of uniformly continuous (wrt $d()$) bounded functionals from $\mathcal{B}$ to $\mathbb{R}$.

Definition 1. *Weak convergence in* $\mathcal{B}$: $\xi_T \Rightarrow \xi$ iff $\forall f \in \mathcal{U}(\mathcal{B})$ we have $Ef(\xi_T) \to Ef(\xi)$ as $T \to \infty$.

Definition 2. ξ is *stochastically equicontinuous* if $\forall \epsilon > 0$, $\forall \eta > 0$, there exists $\delta > 0$ s.t.

$$\lim_{T\to\infty} P(\sup_{|\tau_1-\tau_2|<\delta} |\xi(\tau_1) - \xi(\tau_2)| > \eta) < \epsilon$$

Theorem 3. Empirical Processes Theorem*: If*

1. *Θ is bounded*
2. *there exists a finite-dimensional distribution convergence of ξ_T to ξ (as in (1))*
3. *$\{\xi_T\}$ are stochastically equicontinuous*

then $\xi_T \to \xi$

Condition 2 is usually easy to check checked. Main difficulty is usually in checking condition 3.

Remark 4. Stochastic equicontinuity is equivalent to:

$$\forall \{\delta_T\} : \delta_T \to 0 \qquad \sup_{|\tau_1-\tau_2|<\delta_T} |\xi(\tau_1) - \xi(\tau_2)| \to^p 0$$

Theorem 5. ***Continuous Mapping Theorem****: if $\xi_T \Rightarrow \xi$, then $\forall$ continuous functionals, f, $f(\xi_T) \Rightarrow f(\xi)$*

Functional Central Limit Theorem

Let ϵ_t be a martingale difference sequence (*i.e.* $E(\epsilon_t|\epsilon_{t-1}, ...) = 0\ \forall t$) with $E(\epsilon_t^2|I_{t-1}) = \sigma^2$, $E\epsilon_t^4 < \infty$. Define

$$\xi_T(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \epsilon_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{1}_{\{t \le [T\tau]\}} \epsilon_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_t(\epsilon_t, \tau)$$

Consider some τ_0:

$$\xi_T(\tau_0) = \frac{\sqrt{T\tau_0}}{\sqrt{T}} \frac{1}{\sqrt{T\tau_0}} \sum_{t=1}^{[T\tau_0]} \epsilon_t$$

Standard Central Limit Theorem implies that $\frac{1}{\sqrt{T\tau_0}} \sum_{t=1}^{[T\tau_0]} \epsilon_t \Rightarrow N(0, \sigma^2)$, so

$$\xi_T(\tau_0) \Rightarrow N(0, \sigma^2\tau_0)$$

Similary, if we consider the joint distribution of $\xi_T(\tau_0)$ and $\xi_T(\tau_1) - \xi_T(\tau_0)$. These will be two non-overlappping sums of ϵ_t, so we have:

$$\begin{bmatrix} \xi_T(\tau_0) \\ \xi_T(\tau_1) - \xi_T(\tau_0) \end{bmatrix} \Rightarrow N\left(0, \begin{bmatrix} \tau_0\sigma^2 & 0 \\ 0 & (\tau_1-\tau_0)\sigma^2 \end{bmatrix}\right)$$

We can generalize this to any finite collection $\{\tau_i\}$.

Definition 6. *Brownian motion* or *Weiner process* is a stochastic process, $W()$, such that

1. $W(0) = 0$
2. For $0 \le t_1 < ... < t_k \le 1$, the increments, $W(t_2) - W(t_1), ..., W(t_k) - W(t_{k-1})$ are independent Gaussian with $W(t) - W(s) \sim N(0, t-s)$ for $t > s$
3. $W(t)$ is almost surely continuous

The Functional Central Limit Theorem implies $\xi_T \Rightarrow \sigma W$. (Surely, one has to prove stochastic equicontinuity, which happens to be true here).

Unit Root

Let $y_t = \rho y_{t-1} + \epsilon_t$ and $\rho = 1$(unit root process). The asymptotic behavior of y_t are very different from that of a stationary time series (or just autoregressive processes with $|\rho| < 1$). For example, $\xi_T(\tau) = \frac{1}{\sqrt{T}} y_{[T\tau]} \Rightarrow \sigma W(\tau)$ as a stochastic processes, while for $x_t = \rho x_{t-1} + e_t$, $|\rho| < 1$, we have $\frac{1}{\sqrt{T}} x_{[T\tau]} \rightarrow^p 0$. Another observation: $\bar{x} \xrightarrow{p} Ex_t = 0$ for a stationary process, while for a random walk, $\bar{y}/\sqrt{T}$ has a non-degenerate asymptotic distribution:

$$\begin{aligned}
\frac{1}{T^{3/2}} \sum_{t=1}^{T} y_t =& \frac{1}{T} \sum_{t=1}^{T} \frac{y_t}{\sqrt{T}} \\
=& \frac{1}{T} \sum_{t=1}^{T} \xi_T(t/T) \\
=& \int_0^1 \xi_T(s) ds \Rightarrow \sigma \int_0^1 W(t) dt
\end{aligned}$$

where we used the continuous mapping theorem in the last line. Integration is a continuous functional. Similar reasoning shows that:

$$T^{-1-k/2} \sum_{t=1}^{T} y_t^k \Rightarrow \sigma^k \int_0^1 W^k(s) ds \tag{2}$$

Now, let us consider a distribution of OLS estimates in an auto-regression (regression of y_t on y_{t-1}). In the stationary case, we have:

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

Now let's consider a non-stationary case. The asymptotic distribution of $\hat{\rho} - 1$ will be given by the distribution of $\frac{T^? \sum y_{t-1}\epsilon_t}{T^? \sum y_{t-1} y_{t-1}}$. We write $T^?$ because we're not yet sure of the rate of convergence. The rate will become clear after working with the expression a little bit. Let's examine the numerator and denominator of this expression separately. For the numerator,

$$\begin{aligned}
\sum_{t=1}^{T} y_{t-1}\epsilon_t =& \sum_{t=1}^{T} \epsilon_t(\epsilon_1 + ... + \epsilon_{t-1}) = \sum_{t=1, s<t}^{T} \epsilon_t \epsilon_s \\
=& \frac{1}{2} y_T^2 - \frac{1}{2} \sum \epsilon_t^2
\end{aligned}$$

since,

$$y_T^2 = (\sum_{t=1}^{T} \epsilon_t)^2 = \sum_{t=1}^{T} \epsilon_t^2 + 2 \sum_{t=1, s<t}^{T} \epsilon_t \epsilon_s$$

If we scale the numerator by $1/T$, then we have:

$$\begin{aligned}
\frac{1}{T} \sum y_{t-1}\epsilon_t =& \frac{1}{2} (\frac{1}{\sqrt{T}} y_T)^2 - \frac{1}{2} \frac{1}{T} \sum \epsilon_t^2 \\
=& \frac{1}{2} (\xi_T(1) - \hat{\sigma}^2) \\
\Rightarrow& \frac{1}{2} (W^2(1)\sigma^2 - \sigma^2) = \frac{1}{2} \sigma^2 (W^2(1) - 1)
\end{aligned}$$

If we scale the denominator by $1/T^2$, then we know from (2) above that $\frac{1}{T^2}\sum y_{t-1}^2 \Rightarrow \int_0^1 W(s)^2 ds$. Thus,

$$T(\hat{\rho}-1) \Rightarrow \frac{\frac{1}{2}(W^2(1)-1)}{\int_0^1 W(s)^2 ds}$$

Using Itō's lemma, we can modify this slightly.

Lemma 7. Itō's lemma *Suppose we have a diffusion process, S_s*

$$S_s = \int_0^s a_t dW(t) + \int_0^s b_t dt$$

or, informally,

$$dS = a_t dW + b_t dt$$

Let f be a three times differentiable function, then $df(S)$ is

$$df(S) = f'(a_t dW + b_t dt) + \frac{1}{2} f''(a_t^2 dt)$$

which means that

$$f(S_s) = \int_0^s f' a_t dW(t) + \int_0^s (f' b_t + \frac{1}{2} f'' a^2) dt$$

In our application, $S = W = \int dw$, $f(x) = x^2$, $f'(x) = 2x$, and $f''(x) = 2$. Applying Itō's lemma we have:

$$d(W^2(t)) = 2wdw + \frac{1}{2} 2dt$$

This means that

$$W^2(1) = \int_0^1 dW^2(t) = 2\int_0^1 W(s)dW(s) + \int_0^1 ds$$

so,

$$\int_0^1 W(s)dW(s) = \frac{1}{2}(W^2(1)-1)$$

Thus, we have:

$$T(\hat{\rho}-1) \Rightarrow \frac{\frac{1}{2}(W^2(1)-1)}{\int_0^1 W(s)^2 ds} = \frac{\int_0^1 W(s)dW(s)}{\int_0^1 W^2(s)ds}$$

Notice that:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^T y_{t-1}\epsilon_t &= \sum_{t=1}^T \frac{y_{t-1}}{\sqrt{T}} \frac{\Delta y_t}{\sqrt{T}} \\
&= \int_0^1 \xi_T(s) d\xi_T(s) \\
&\Rightarrow \int_0^1 W(s)dW(s)
\end{aligned}$$

However, the convergence in the last line does not follow from the continuous mapping theorem. Stochastic integration is not a continuous functional, *i.e.* if $f_n \to f$, in the uniform metric, $d()$, and g is bounded, then it does not necessarily imply that $\int g df_n \to \int g df$. Generally, showing convergence of stochastic integrals is a more delicate task. Nonetheless, it holds in our case, as we have just shown.