D.C. Economists Minicourse –
What's New in Econometrics: Time Series

Lecture 2:

September 26, 2008

The Functional Central Limit Theorem and "Breaks"

Outline

1. FCLT
2. Breaks – Testing
3. Breaks – Inference about Break Dates

1. FCLT

The Functional Central Limit Theorem

Problem: Suppose $\varepsilon_t \sim \text{iid}(0,\sigma_\varepsilon^2)$ (or weakly correlated with long-run variance σ_ε^2), $x_t = \sum_{i=1}^{t}\varepsilon_i$, and we need to approximate the distribution of a function of $(x_1, x_2, x_3, \ldots x_T)$, say $\sum_{t=1}^{T} x_t^2$.

Solution: Notice $x_t = \sum_{i=1}^{t}\varepsilon_i = x_{t-1} + \varepsilon_t$. This suggests an approximation based on a normally distributed (CLT for first equality) random walk (second equality). The tool used for the approximation is the Functional Central Limit Theorem.

Some familiar notions

1. Convergence in distribution or "weak convergence": ξ_T, $T = 1, 2, \ldots$ is a sequence of random variables. $\xi_T \xrightarrow{d} \xi$ means that the probability distribution function (PDF) of ξ_T converges to the PDF of ξ. As a practical matter this means that we can approximate the PDF of ξ_T using the PDF of ξ when T is large.

2. Central Limit Theorem: Let ε_t be a mds$(0,\sigma_\varepsilon^2)$ with $2+\delta$ moments and $\xi_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t$. Then $\xi_T \xrightarrow{d} \xi \sim \mathrm{N}(0,\ \sigma_\varepsilon^2)$.

3. Continuous mapping theorem. Let g be a continuous function and $\xi_T \xrightarrow{d} \xi$, then $g(\xi_T) \xrightarrow{d} g(\xi)$. (Example ξ_T is the usual t-statistic, and $\xi_T \xrightarrow{d} \xi \sim \mathrm{N}(0,\ 1)$, then $\xi_T^2 \xrightarrow{d} \xi^2 \sim \chi_1^2$.

These ideas can be extended to random functions:

A Random Function: The Wiener Process, a continuous-time stochastic process sometimes called Standard Brownian Motion that will play the role of a “standard normal” in the relevant function space.

Denote the process by $W(s)$ defined on $s \in [0,1]$ with the following properties

1. $W(0) = 0$

2. For any dates $0 \leq t_1 < t_2 < \ldots < t_k \leq 1$, $W(t_2) - W(t_1)$, $W(t_3) - W(t_4)$, … , $W(t_k) - W(t_{k-1})$ are independent normally distributed random variables with $W(t_i) - W(t_{i-1}) \sim \mathrm{N}(0, t_i - t_{i-1})$.

3. Realizations of W(s) are continuous w.p. 1.

From (1) and (2), note that $W(1) \sim \mathrm{N}(0,1)$.

Another Random Function: Suppose $\varepsilon_t \sim$ iidN(0,1), $t = 1, \dots, T$, and let $\xi_T(s)$ denote the function that linearly interpolates between the points $\xi_T(t/T) = \frac{1}{\sqrt{T}}\sum_{i=1}^{t}\varepsilon_i$.

Can we use W to approximate the probability law of $\xi_T(s)$ if T is large?

More generally, we want to know whether the probability distibution of a random function can be well approximated by the PDF of another (perhaps simpler, maybe Gaussian) function when T is large. Formally, we want to study weak convergence on function spaces.

Useful References: Hall and Heyde (1980), Davidson (1994), Andrews (1994)

Suppose we limit our attention to continuous functions on $s \in [0,1]$ (the space of such functions is denoted C[0,1]), and we define the distance between two functions, say x and y as $d(x,y) = sup_{0 \leq s \leq 1} |x(s) - y(s)|$.

Three important theorems (Hall and Heyde (1980) and Davidson (1994, part VI):

Important Theorm 1: (Hall and Heyde Theorem A.2) Weak Convergence of random functions on C[0,1]

Weak convergence (denoted "$\xi_T \Rightarrow \xi$") follows from (i) and (ii), where

(i) Let $0 \le s_1 < s_2 \ldots < s_k \le 1$, a set of k points. Suppose that $(\xi_T(s_1), \xi_T(s_2), \ldots, \xi_T(s_k)) \overset{d}{\rightarrow} (\xi(s_1), \xi(s_2), \ldots, \xi(s_k))$ for any set of k points, $\{s_i\}$.

(ii) The function $\xi_T(s)$ is "tight" (or more generally satisfies "stochastic equicontinuity" as discussed in Andrews (1994)), meaning

(a) For each $\varepsilon > 0$, $\text{Prob}[\sup_{|s-t|<\delta}|\xi_T(s) - \xi_T(t)| > \varepsilon] \rightarrow 0$ as $\delta \rightarrow 0$ uniformly in T. (This says that the function ξ_T does not get too "wild" as T grows.)

(b) $\text{Prob}[|\xi_T(0)| > \delta] \rightarrow 0$ as $\delta \rightarrow \infty$ uniformly in T. (This says the function can't get too crazy at the origin ast T grows.)

Important Theorem 2: (Hall on Heyde Theorem A.3) Continuous Mapping Theorem

Let g: C[0,1] $\rightarrow \mathbb{R}$ be a continuous function and suppose $\xi_T(.) \Rightarrow \xi(.)$.

Then $g(\xi_T) \Rightarrow g(\xi)$.

Important Theorem 3: (Hall and Heyde) Functional Central Limit Theorem:

Suppose $\varepsilon_t \sim$ mds with variance σ_ε^2 and bounded 2+δ moments for some $\delta > 0$.

(a) Let $\xi_T(s)$ denote the function that linearly interpolates between the points $\xi(t/T) = \frac{1}{\sqrt{T}}\sum_{i=1}^{t}\varepsilon_i$. Then $\xi_T \Rightarrow \sigma_\varepsilon W$, where W is a Wiener process (standard Brownian motion).

(b) The results can be extended to $\xi_T(s) = \frac{1}{\sqrt{T}}\sum_{i=1}^{[sT]}\varepsilon_i$, the step-function interpolation, where [.] is the "less than or equal to integer function" (so that [3.1] = 3, [3.0] = 3, [3.9999] = 3, and so forth).

See Davidson Ch. 29 for extensions.

An Example:

(1): Let $x_t = \sum_{i=1}^{t} \varepsilon_i$, where ε_i is $mds(0, \sigma_\varepsilon^2)$, and let $\xi_T(s) = \frac{1}{\sqrt{T}} \sum_{i=1}^{[sT]} \varepsilon_i = \frac{1}{\sqrt{T}} x_{[sT]}$ be a step function approximation of $W(s)$.

Then

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} x_t = \frac{1}{T} \sum_{t=1}^{T} \left[\frac{1}{T^{1/2}} \sum_{i=1}^{t} \varepsilon_i \right] = \sigma_\varepsilon \int_0^1 \xi_T(s) ds \Rightarrow \sigma_\varepsilon \int_0^1 W(s) ds$$

Additional persistence …

Suppose $a_t = \varepsilon_t - \theta\varepsilon_{t-1} = \theta(\mathrm{L})\varepsilon_t$, and $x_t = x_{t-1} + a_t$.

Then

$$T^{-1/2}x_t = T^{-1/2}\sum_{i=1}^{t} a_i = T^{-1/2}\sum_{i=1}^{t}(\varepsilon_i - \theta\varepsilon_{i-1}) = (1-\theta)T^{-1/2}\sum_{i=1}^{t}\varepsilon_i + \theta T^{-1/2}(\varepsilon_t - \varepsilon_0)$$

But $\theta T^{-1/2}(\varepsilon_t - \varepsilon_0)$ is negligible, so that $T^{-1/2}x_{[sT]} \Rightarrow (1-\theta)\sigma_\varepsilon W(s)$.

This generalizes: suppose $a_t = \theta(\mathrm{L})\varepsilon_t$ and $\sum_{i=0}^{\infty} i\,|\theta_i| < \infty$ (so that the MA coefficients are "one-summable"), then $T^{-1/2}x_{[sT]} \Rightarrow \theta(1)\sigma_\varepsilon W(s)$.

Note: $\theta(1)\sigma_\varepsilon$ is the "long-run" standard deviation of a.

What does this all mean?

Suppose I want to approximate the 95th quantile of the distribution of, say,

$v_T = \frac{1}{T^{3/2}} \sum_{t=1}^{T} x_t$. Because $v_T \Rightarrow v = \sigma_\varepsilon \int_0^1 W(s)ds$, I can use the 95th quantile of v are the approximator.

How do I find (or approximate) the 95th quantile of v?

Use Monte Carlo draws of $\sigma_\varepsilon N^{-3/2} \sum_{t=1}^{N} \sum_{i=1}^{t} z_i$ where $z_i \sim$ iidN(0,1) and N is very large.

This approximation works well when T is reasonably large, and does not require knowledge of the distribution of x.

TVP: Models

Linear regression: $y_t = x_t'\beta_{\mathbf{t}} + \varepsilon_t$

IV Regression (Linear GMM): $E\{z_t\,(y_t - x_t'\beta_{\mathbf{t}})\} = 0$

Nonlinear Model (GMM, NLLS, Stochastic Volatility, …)

Simple model as leading case (variables are scalars):

$$y_t = \beta_t + \varepsilon_t$$

β_t is the local level ("mean") of y_t.

Time variation in β

Nonstochastic Breaks: $\beta_t = \begin{cases} \beta \text{ for } t \le \tau \\ \beta + \delta \text{ for } t > \tau \end{cases}$ (or 2, 3 breaks, …)

Stochastic Breaks:

- Markov Switching (Hamilton (1989))
- $\beta_t \sim$ ARIMA. These are unobserved components (UC) models (Harvey (1989), Nerlove, Grether and Carvalho (1995))

Today we will focus on a model with single nonstochastic break.

Tests for a break

Model: $y_t = \beta_t + \varepsilon_t$, where $\varepsilon_t \sim \text{iid}\ (0,\ \sigma_\varepsilon^2)$

$$\beta_t = \begin{cases} \beta \text{ for } t \le \tau \\ \beta + \delta \text{ for } t > \tau \end{cases}$$

Null and alternative: H_o: $\delta = 0$ vs. H_o: $\delta \neq 0$

Tests for H_o vs. H_a depends on whether τ is known or unknown.

Chow Tests (known break date)

Least squares estimator of δ: $\hat{\delta} = \bar{Y}_2 - \bar{Y}_1$

where $\bar{Y}_1 = \frac{1}{\tau}\sum_{t=1}^{\tau} y_t$ and $\bar{Y}_2 = \frac{1}{T-\tau}\sum_{t=\tau+1}^{T} y_t$

Wald statistic: $\xi_W = \frac{1}{\hat{\sigma}_\varepsilon^2}\frac{\hat{\delta}^2}{(\frac{1}{\tau}+\frac{1}{T-\tau})}$

Follows from $\bar{Y}_1 \overset{a}{\sim} N(\beta, \frac{\sigma_e^2}{\tau})$ and $\bar{Y}_2 \overset{a}{\sim} N(\beta+\delta, \frac{\sigma_\varepsilon^2}{T-\tau})$ and they are independent.

Under H_o ξ_W is distributed as a χ_1^2 random variable in large (τ and $T-\tau$) samples. Thus, critical values for the test can be determined from the χ^2 distribution.

Quandt Tests (Sup Wald or QLR) (unknown break date)

Quandt (1960) suggested computing the Chow statistic for a large number of possible values of τ and using the largest of these as the test statistics.

$$\text{QLR statistic: } \xi_Q = \max_{\tau_1 \le \tau \le \tau_2} \xi_W(\tau)$$

where the Chow statistic ξ_W is now indexed by the break date.

The problem is then to find the distribution of ξ_Q under the null (it will not be χ^2), so that the critical value for the test can be determined.

Let $s = \tau/T$. Under the null $\delta = 0$, and (now using s as the index), we can then write ξ_W as

$$\xi_{W,T}(s) \overset{H_o}{=} \frac{1}{\hat{\sigma}_e^2} \frac{\left[\frac{1}{[sT]}\sum_{t=1}^{[sT]} \varepsilon_t - \frac{1}{[(1-s)T]}\sum_{t=[sT]+1}^{T} \varepsilon_t\right]^2}{\frac{1}{[sT]} + \frac{1}{[(1-s)T]}}$$

$$= \frac{1}{\hat{\sigma}_e^2} \frac{\left[\frac{1}{s}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \varepsilon_t - \frac{1}{(1-s)}\frac{1}{\sqrt{T}}\sum_{t=[sT]+1}^{T} \varepsilon_t\right]^2}{\frac{1}{s} + \frac{1}{(1-s)}}$$

$$= \frac{\left[\frac{1}{s}W_T^a(s) - \frac{1}{(1-s)}(W_T^a(1) - W_T^a(s))\right]^2}{\frac{1}{s} + \frac{1}{(1-s)}} = \frac{[W_T^a(s) - sW_T^a(1)]^2}{s(1-s)}$$

where $W_T^a(s) = \frac{1}{\hat{\sigma}_\varepsilon}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \varepsilon_t$, and the last equality follows from multiplying the numerator and denominator by $s^2(1-s)^2$ and simplifying.

Thus, using FCLT, $\xi_{W,T}(\ ) \Rightarrow \xi(\ )$, where $\xi(s) = \frac{[W(s) - sW(1)]^2}{s(1-s)}$.

Suppose that τ_1 is chosen as $[\lambda T]$ and τ_2 is chosen as $[(1-\lambda)T]$, where $0 < \lambda < 0.5$. Then

$$\xi_Q = \sup_{\lambda \leq s \leq (1-\lambda)} \xi_{W,T}(s), \text{ and } \xi_Q \Rightarrow \sup_{\lambda \leq s \leq (1-\lambda)} \xi(s)$$

It has become standard practice to use a value of $\lambda = 0.15$. Using this value of λ, the 1%, 5% and 10% critical values for the test are: 12.16, 8.68 and 7.12. (These can be compared to the corresponding critical values of the χ^2 distribution of 6.63, 3.84 and 2.71).

The results have been derived here for the case of a single constant regressor. Exensions to the case of multiple (non-constant) regressors can be found in Andrews (1993) (Critical values for the test statistic are also given in Andrews (1993) with corrections in Andrews (2003), reprinted in Stock and Watson (2006).)

Optimal Tests when the break point (τ) is unknown

The QLR test seems very sensible, but is there a more powerful procedure? Andrews and Ploberger (1993) develop optimal tests (most powerful) for this (and related) problems.

Recall the Neyman-Pearson (NP) lemma: consider two simple hypotheses

$$H_o: Y \sim f_o(y) \quad \text{vs.} \quad H_a: Y \sim f_a(y),$$

then the most powerful test rejects H_o for large values of the Likelihood Ratio, LR = $f_a(Y)/f_o(Y)$, the ratio of the densities evaluated at the realized value of the random variable.

Here: likelihoods depend on parameters δ, β, σ_ε, and τ, where $\delta = 0$ under H_o). (δ, β, σ_ε) are easily handled in the NP framework. τ is more of a problem. It is unknown, and if H_o is true, it is irrelevant (τ is "unidentified" under the null). Andrews and Ploberger (1994) (AP) attack this problem.

One way to motivate the AP approach: suppose τ is a random variable with a known distribution, say F_τ. Then the density of Y is a mixture:

$Y \sim f_a(y)$ where $f_a(y) = E_\tau[f_a(y|\tau)]$.

The LR test (ignoring (δ, β, σ_ε) for convenience) is then

LR = $E_\tau[f_a(y|\tau)]/f_o(Y)$.

The interpretation of this test is (equivalently) that it is (i) the most powerful for $\tau \sim F_\tau$, or (ii) it maximizes F_τ- weighted power for fixed values of τ.

This approach to dealing with nuisance parameters (here τ) that are unidentified under H_o is now standard.

The specific form of the test depends on the weight function F_τ, (AP suggest a uniform distribution on τ) and how large δ is assumed to be under the alternative.

When δ is "small," the test statistic turns out to be simple average of $\xi_W(\tau)$ over all possible break dates. Sometimes this test statistic is called the "Mean Wald" statistic.

When δ is "large," the test statistic turns out to be a simple average of $\exp(0.5\times\xi_W(\tau))$ over all possible break dates. Sometimes this test statistic is called the "Exponential Wald" statistic.

Importantly, as it turns out, the AP exponential Wald test statistic is typically dominated by the largest values of $\xi_W(\tau)$. This means the QLR statistic behaves very much like the exponential Wald test statistic and is, in this sense, essentially an optimal test.

Inference about Break Dates

(Bai (1997), Hansen (2001))

Same Example:

$$y_t = \beta_t + \varepsilon_t \qquad \beta_t = \begin{cases} \beta \text{ for } t \le \tau \\ \beta + \delta \text{ for } t > \tau \end{cases}$$

$\pi = \tau/T$ = Break "Fraction" (sorry for the change in notation)

τ_o = true break date

π_o = true break fraction

$\hat{\tau}$ = Least squares estimator of τ, $\hat{\pi} = \hat{\tau}/T$

Some results that are useful for inference:

Bai(1997) shows $\hat{\pi} - \pi_o \sim O_p(T^{-1}\delta^{-2})$, so that

$$T\delta^2(\hat{\pi} - \pi_o) \sim O_p(1)$$

$$\delta^2(\hat{\tau} - \tau_o) \sim O_p(1)$$

Thus, $\hat{\pi}$ is consistent for π_o, but $\hat{\tau}$ is not consistent for τ_o.

The speed at which $\hat{\pi}$ converges to π_o depends on δ.

In general, the distribution of $\hat{\pi}$ and $\hat{\tau}$ depends on the distribution of the errrors ε_t. This is true even when T is large. Thus, robust inference is problematic.

There are approximations that can be used when δ is appropriately small.

An asymptotic approximation: Recall $T\delta^2(\hat{\pi}-\pi_o) \sim O_p(1)$, so assume $T\delta^2 \to\infty$. Also, δ is small, so assume $\delta \to 0$. (More formally, $\delta = \delta_T$ which approaches zero as T grows large.). (Example, both of these are satisfied if $\delta_T = a/T^{1/3}$)

The challenge is to compute a convenient expression $\hat{\pi}$. The trick is to use empirical process methods like those used for the FCLT. The main ideas can be understood in a situation in which β and δ are known, so that estimating τ (equivalently π) is the only problem.

The least squares objective function is

$$\text{SSR}(\tau) = \sum_{t=1}^{\tau}(y_t-\beta)^2 + \sum_{t=\tau+1}^{T}(y_t-\beta-\delta)^2$$

and the trick is to study the behavior of this function as T gets large.

Suppose $\tau > \tau_o$. Then we can write

$$\begin{aligned} SSR(\tau) &= \sum_{t=1}^{\tau}(y_t-\beta)^2 + \sum_{t=\tau+1}^{T}(y_t-\beta-\delta)^2 \\ &= \sum_{t=1}^{\tau_o}(y_t-\beta)^2 + \sum_{t=\tau+1}^{T}(y_t-\beta-\delta)^2 + \sum_{t=\tau_o+1}^{\tau}((y_t-\beta-\delta)+\delta)^2 \\ &= \sum_{t=1}^{\tau_o}\varepsilon_t^2 + \sum_{t=\tau+1}^{T}\varepsilon_t^2 + \sum_{t=\tau_o+1}^{\tau}(\varepsilon_t+\delta)^2 \\ &= \sum_{t=1}^{T}\varepsilon_t^2 + (\tau-\tau_o)\delta^2 + 2\delta\sum_{t=\tau_o+1}^{\tau}\varepsilon_t \\ &= \sum_{t=1}^{T}\varepsilon_t^2 + (\pi-\pi_o)T\delta^2 + 2\delta\sum_{t=1}^{[T(\pi-\pi_o)]}\varepsilon_{[\pi_o T]+1} \end{aligned}$$

Where the last expression substitures $\pi = \tau/T$. The first term does depend on τ; ignore it when thinking about the function that is being maximizing.

Thus, we can think about choosing τ or π to minimize

$$(\tau - \tau_o)\delta^2 + 2\delta \sum_{t=\tau_o+1}^{\tau} \varepsilon_t = (\pi - \pi_o)T\delta^2 + 2\delta \sum_{t=1}^{[T(\pi-\pi_o)]} \varepsilon_{[\pi_o T]+1}$$

Let $\upsilon = (\pi - \pi_o)T\delta^2/\sigma_\varepsilon^2$.

Then minimizing *SSR* over τ is the same as minimizing g_T over υ, where

$$g_T(\upsilon) = \upsilon + 2(\delta/\sigma_\varepsilon) \sum_{t=1}^{[(\delta/\sigma_\varepsilon)^{-2}\upsilon]} (\varepsilon_{[\pi_o T]+1} / \sigma_\varepsilon)$$

and the division by σ_ε is for later convenience.

Recall $\delta \to 0$, so $\delta^{-2} \to \infty$, so that $(\delta/\sigma_\varepsilon)\sum_{t=1}^{[(\delta/\sigma_\varepsilon)^{-2}\upsilon]}(\varepsilon_{[\pi_o T]+1}/\sigma_\varepsilon) \xrightarrow{d} W(\upsilon)$.

(For analogy with the standard formula, think of $(\delta/\sigma_\varepsilon)^{-2}$ = *sample size*, so that $(\delta/\sigma_\varepsilon) = \dfrac{1}{\sqrt{sample\ size}}$).

Thus $g_T(\upsilon) = \upsilon + 2(\delta/\sigma_\varepsilon)\sum_{t=1}^{[(\delta/\sigma_\varepsilon)^{-2}\upsilon]}(\varepsilon_{[\pi_o T]+1}/\sigma_\varepsilon) \xrightarrow{d} g(\upsilon) = \upsilon + 2W(\upsilon)$.

and from arguments like those used for the FCLT, $g_T(\,.\,) \Rightarrow g(\,.\,)$.

The argument for $\tau < \tau_o$ is similar. Putting these together, the least squares problem for estimating τ (or $\pi = \tau/T$) is approximately the same as

$$\min_\upsilon G(\upsilon) \text{ where } G(\upsilon) = \begin{cases} |\upsilon| + 2W_1(\upsilon) \text{ for } \upsilon \geq 0 \\ |\upsilon| + 2W_2(-\upsilon) \text{ for } \upsilon < 0 \end{cases}$$

where W_1 and W_2 are independent Wiener processes.

The value of υ that minimizes this random function has a very non-gaussian shape.

Here are some values

	c	
Probability	$\Pr(\lvert \hat{\upsilon} \rvert < c)$	Standard Normal Prob($\lvert z \rvert < c$)
50%	2.8	
67%	4.4	1.0
80%	6.7	
90%	10.0	1.64
95%	13.8	1.96
99%	23.5	2.56

Constructing a confidence interval for τ: Ingredients, (T, $\hat{\sigma}_{\varepsilon}$, $\hat{\delta}$, $\hat{\pi}$ (or $\hat{\tau}$)):

We know (from the table on the last page) $\Pr(|\hat{\upsilon}| < 4.4) = 0.67$. Using $\hat{\upsilon} = \frac{1}{\hat{\sigma}_{\varepsilon}^{2}} T\hat{\delta}^{2}(\hat{\pi} - \pi_{o})$, a 67% confidence interval for π satisfies

$$\left|\frac{1}{\hat{\sigma}_{\varepsilon}^{2}} T\hat{\delta}^{2}(\hat{\pi} - \pi)\right| < 4.4 \text{ or } \hat{\pi} - 4.4\frac{\hat{\sigma}_{\varepsilon}^{2}}{T\hat{\delta}^{2}} < \pi < \hat{\pi} + 4.4\frac{\hat{\sigma}_{\varepsilon}^{2}}{T\hat{\delta}^{2}}$$

so that a 67% CI for τ is: $\hat{\tau} - 4.4\frac{\hat{\sigma}_{\varepsilon}^{2}}{\hat{\delta}^{2}} < \tau < \hat{\tau} + 4.4\frac{\hat{\sigma}_{\varepsilon}^{2}}{\hat{\delta}^{2}}$

(Programs: Bruce Hansen’s webpage)

Empirical Example … Great Moderation

$\phi_t(\mathrm{L})\Delta y_t = \mu_t + \varepsilon_t \quad (Y = \ln(\mathrm{GDP}))$

$$\varepsilon_t^2 = \sigma_t^2 + (\varepsilon_t^2 - \sigma_t^2)$$

(Numbers from SW(2002))
p-value for break in $\phi_t(\mathrm{L})$ and $\mu_t = 0.98$
p-value for break in $\sigma_t^2 = 0.00$ ($\hat{\tau} = 1983:2$)

67% CI for Break in σ_t^2: 1982:4 – 1985:3

Multiple Deterministic Breaks: Bai and Perron (1998)

Single Joint Deterministic Breaks in Multiple Processes: Bai, Lumsdaine, Stock (1998)

Multiple Joint Deterministic Breaks in Multiple Processes; Qu and Perron (2007)

Using “Breaks”: Historical Analysis vs. Forecasting

Regressors:

Example studied above: $y_t = \beta_t + \varepsilon_t$

Regressors: $y_t = x_t'\beta_t + \varepsilon_t$

Heuristic with $\beta_t = \beta_{t-1} + \gamma\eta_t,\ \beta_0 = 0$

$$x_ty_t = x_tx_t'\beta_t + x_t\varepsilon_t = \Sigma_{xx}\beta_t \quad + x_t\varepsilon_t \quad + \quad (x_tx_t' - \Sigma_{xx})\beta_t$$

$$= \Sigma_{xx}\beta_t \quad + \quad e_t \quad + \quad m_t\beta_t$$

Test statistic depends on $T^{-1/2}\sum_{i=1}^{t} y_t x_t = \Sigma_{xx} T^{-1/2}\sum_{i=1}^{t}\beta_t + T^{-1/2}\sum_{i=1}^{t} e_t + T^{-1/2}\sum_{i=1}^{t} m_t\beta_t$

With $m_t(x)$ “well-behaved,” the final term is negligible.

Hansen (2000) studies the effect of changes in the x process on standard TVP tests.

HAC Corrections: $y_t = x_t'\beta_t + \varepsilon_t$ where $Var(x_t\varepsilon_t) \neq \sigma_\varepsilon^2\Sigma_{XX}$.

Summary

1. FCLT (tool)
2. Breaks – Tests
3. Breaks – Break Dates