NBER Summer Institute Minicourse –
What's New in Econometrics: Time Series

Lecture 2:

July 14, 2008

The Functional Central Limit Theorem and Testing for Time Varying Parameters

Outline

1. FCLT
2. Overview of TVP topics (Models, Testing, Estimation)
3. Testing problems
4. Tests

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

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

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

Another Random Function: Suppose $\varepsilon_t \sim$ iidN(0,1), $t = 1, \ldots, T$, 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 $d(x,y) = sup_{0 \le s \le 1} |x(s) - y(s)|$.

Three key 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 $0 \leq s_1 < s_2 \ldots < s_k \leq 1$, a set of k points. Suppose that $(\xi_T(s_1), \xi_T(s_2), \ldots, \xi_T(s_k)) \xrightarrow{d} (\xi(s_1), \xi(s_2), \ldots, \xi(s_k))$ 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 $\varepsilon > 0$, $\text{Prob}[\sup_{|s-t|<\delta}|\xi_T(s) - \xi_T(t)| > \varepsilon] \to 0$ as $\delta \to 0$ uniformly in T. (This says that the function ξ_T does not get too "wild" as T grows.)

(b) $\text{Prob}[|\xi_T(0)| > \delta] \to 0$ as $\delta \to \infty$ 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] $\to \mathbb{R}$ 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 δ > 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.

2. Overview of TVP topics

Models

Linear regression: $y_t = x_t'\beta_t + \varepsilon_t$

IV Regression (Linear GMM): $E\{z_t\,(y_t - x_t'\beta_t)\} = 0$

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

Discrete Breaks:

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

Two Breaks: $\beta_t = \begin{cases} \beta \text{ for } t \le \tau_1 \\ \beta + \delta_1 \text{ for } \tau_1 < t \le \tau_2 \\ \beta + \delta_1 + \delta_2 \text{ for } t > \tau_2 \end{cases}$

Multiple Breaks: ….

Stochastic Breaks/Markov Switching (Hamilton (1989)):

2-Regime Model:

$$\beta_t = \begin{Bmatrix} \beta \text{ when } s_t = 0 \\ \beta + \delta \text{ when } s_t = 1 \end{Bmatrix},$$

s_t follows a Markov process $P(s_t = i | s_{t-1} = j) = p_{ij}$

Multiple Regimes …

Other Stochastic Breaks: …

"Continuous" Evolution:

Random Walk/Martingale: $\beta_t = \beta_{t-1} + \eta_t$

ARIMA Model: $\beta_t \sim \text{ARIMA}(p,d,q)$

In simple model: $y_t = \beta_t + \varepsilon_t$,

these are "unobserved component" models (Harvey (1989), Nerlove, Grether and Carvalho (1995))

with the familiar simple forecast functions

$$y_{t+1/t} = (1-\theta)^{-1}\sum_{i=0}^{\infty}\theta^{i+1}y_{t-i}\,.$$

Relationship between models:

(1) “Discrete” vs. “Continuous” – Not very important

(a) $\beta_t = \beta_{t-1} + \eta_t$. If distribution of η has point mass at zero, this is a model with occasional discrete shifts in β.

(b) Elliott and Müller (2006). Optimal tests are very similar if number of “breaks” is large (say, 3).

(c) Discrete breaks and long-memory (fractional process for β): Diebold and Inoue (2001) and Davidson and Sibbertsen (2005)

(2) Strongly mean-reverting or not – can be more important

β_t ~ ARMA with small AR roots, β_t ~ recurrent Markov Switching

vs.

β_t ~ RW (or AR with large root), β_t has "breaks" with little structure.

(a) Difference important for forecasting

(b) Not important for "tracking" (smoothing)

(3) Deterministic vs. Stochastic …

Important for forecasting (we will return to this)

What does TVP mean?

- Suppose Y_t is a covariance stationary vector. Then subsets of Y are covariance stationary
  - $Y_t \sim$ ARMA, then subsets of Y are $\sim$ ARMA (where order depends on the subset chosen). Thus, finding TVP in univariate or bivariate models indicates TVP in larger models that include Y.

- Time variation in conditional mean or variance?
  - $\Phi(\mathrm{L})Y_t = \varepsilon_t$, $\Phi = \Phi_t$ and/or $\Sigma_\varepsilon = \Sigma_{\varepsilon,t}$
  - Suppose $\Gamma(\mathrm{L})X_t = e_t$, and Y is a subset of X. Then

    $\mathrm{C}(\mathrm{L})Y_t = \sum_{i=1}^{n_x} A_i(L)e_{i,t}$. Changes in the relative variances of $e_{i,t}$ will induce changes in both Φ and Σ_ε in the marginal representation for Y. Thus, finding Φ-TVP in Y model does not imply Γ-TVP in X model (but it does imply Γ and/or Σ_e TVP).

Evidence on TVP in Macro

- VARs
  - SW (1996): 5700 Bivariate VARs involving Post-war U.S. macro series. Reject null of constant VAR coefficient (Φ) in over 50% of cases using tests with size = 10%.
  - Many others …

- Volatility (Great Moderation)

2. Testing Problems

Tests for a break

Model: $y_t = \beta_t + \varepsilon_t$, where $\varepsilon_t \sim$ 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: $\hat{\delta} = \overline{Y}_2 - \overline{Y}_1$

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

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

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

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

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).)

<u>Optimal</u> 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 (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:

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

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

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

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.

Tests for martingale time variation

Write the model as

$$y_t = \beta_t + \varepsilon_t \text{ with } \beta_t = \beta_{t-1} + \gamma e_t$$

where e_t is iidN$(0,\sigma_\varepsilon^2)$ and is independent of ε_i for all t and i. (As a normalization, the variances of ε and e are assumed to be equal.) For simplicity suppose that $\beta_0 = 0$. (Non-zero values are handled by restricting tests to be invariant to adding a constant to the data.)

Let $Y = (y_1, \dots y_T)'$, so that $Y \sim \mathrm{N}(0,\sigma_\varepsilon^2\Omega(\gamma))$, where $\Omega(\gamma) = \mathrm{I} + \gamma^2 A$, where $A = [a_{ij}]$ with $a_{ij} = \min(i,j)$.

From King (1980), the optimal test of H_o: $\gamma = 0$ vs. H_a: $\gamma = \gamma_a$, can be constructed using the likelihood ratio statistic . The LR statistic is given by

$$LR = |\Omega(\gamma_a)| \exp\{-\frac{1}{2\sigma_\varepsilon^2}[Y'\Omega(\gamma_a)^{-1}Y - Y'Y]\}$$

so that the test rejects the null for large values of $\dfrac{Y'\Omega(\gamma_a)^{-1}Y}{Y'Y}$.

Optimal tests require a choice of γ_a, which measures the amount of time variation under the alternative. A common way to choose γ_a is to use a value so that, when $\gamma = \gamma_a$, the test has a pre-specified power, often 50%.

Generally, this test (called a "point optimal" or "point optimal invariant" test) has near optimal power for a wide range of values of γ. A good rule of thumb (from Stock-Watson (1998) is to set $\gamma_a = 7/T$.)

A well known version of this test uses the local (γ_a^2 small) approximation

$$\Omega(\gamma_a)^{-1} = [I + \gamma_a^2 A]^{-1} \approx 1 - \gamma_a^2 A.$$

In this case, the test rejects for large values of

$$\psi = \frac{Y' A Y}{Y' Y}$$

which is a version of the locally best test of Nyblom (1989).

Because $A=PP'$ where P is a lower triangular matrix of 1's, the test statistic can be written as $\psi = \frac{Q'Q}{Y'Y}$, where $Q=P'Y$ (so that $q_t = \sum_{i=t}^{T} y_i$). The statistic can then be written as $\psi = \frac{\sum_{t=1}^{T} p_t^2}{\sum_{t=1}^{T} y_t^2} = \frac{\sum_{t=1}^{T} (\sum_{i=t}^{T} y_i)^2}{\sum_{t=1}^{T} y_t^2}$.

To derive the distribution of the statistic under the null, write (under the null) $y_t = \beta_0 + \varepsilon_t$, and $\frac{1}{\sqrt{T}}\sum_{t=1}^{[\bullet T]}\varepsilon_t \Rightarrow \xi(\bullet)$, where $\xi(s) = \sigma_\varepsilon W(s)$. Thus

$$T^{=1}\psi = \frac{\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{T}}\sum_{i=t}^{T} y_i)^2}{\frac{1}{T}\sum_{t=1}^{T} y_t^2} \overset{H_o}{\Rightarrow} \int_0^1 (W(1) - W(s))^2 ds.$$

In most empirical applications, β_0 is non-zero and unknown, and in this case the test statistic is

$$T^{=1}\psi = \frac{\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{T}}\sum_{i=t}^{T} \hat{\varepsilon}_i)^2}{\frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2},$$ where $\hat{\varepsilon}_t = y_t - \hat{\beta}$ is the OLS residual. In this case, one can show that $T^{=1}\psi \overset{H_o}{\Rightarrow} \int_0^1 (W(s) - sW(1))^2 ds$

(In Lecture 6, I'll denote $T^{-1}\psi$ by ξ^{Nyblom} because it is a version of the Nyblom test statistic.)

Regressors:

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

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

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

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

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

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.

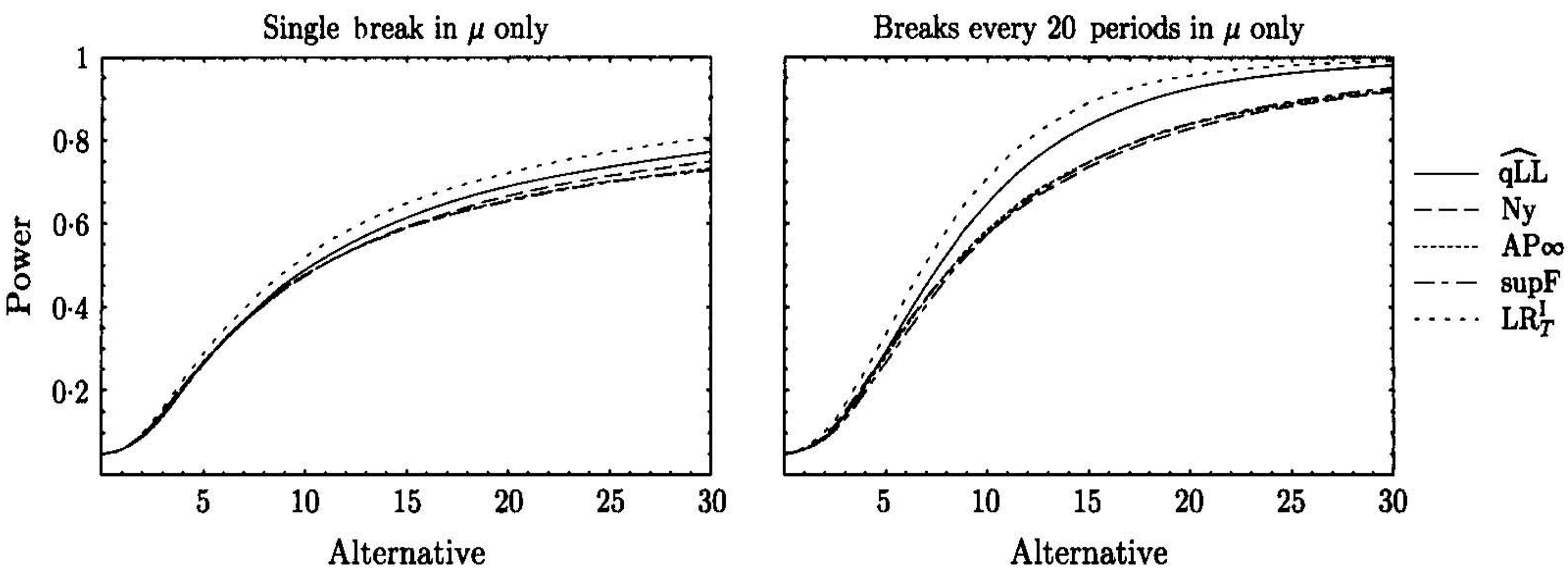
Hansen (2000) studies the effect of changes in the x process on standard TVP tests.

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

Power Comparisons of Tests

1. Elliott-Müller (2006): Discrete Break DGP: $y_t = \mu_t + \alpha x_t + \varepsilon_t$, where μ_t follows a discrete break process.

928 REVIEW OF ECONOMIC STUDIES



Tests for martingale variation (qLL, Ny) have power that is similar to tests for discrete break (SupF=QLR, AP).

2. Stock-Watson (1998): Martingale TVP DGP: $y_t = \beta_t + \varepsilon_t$

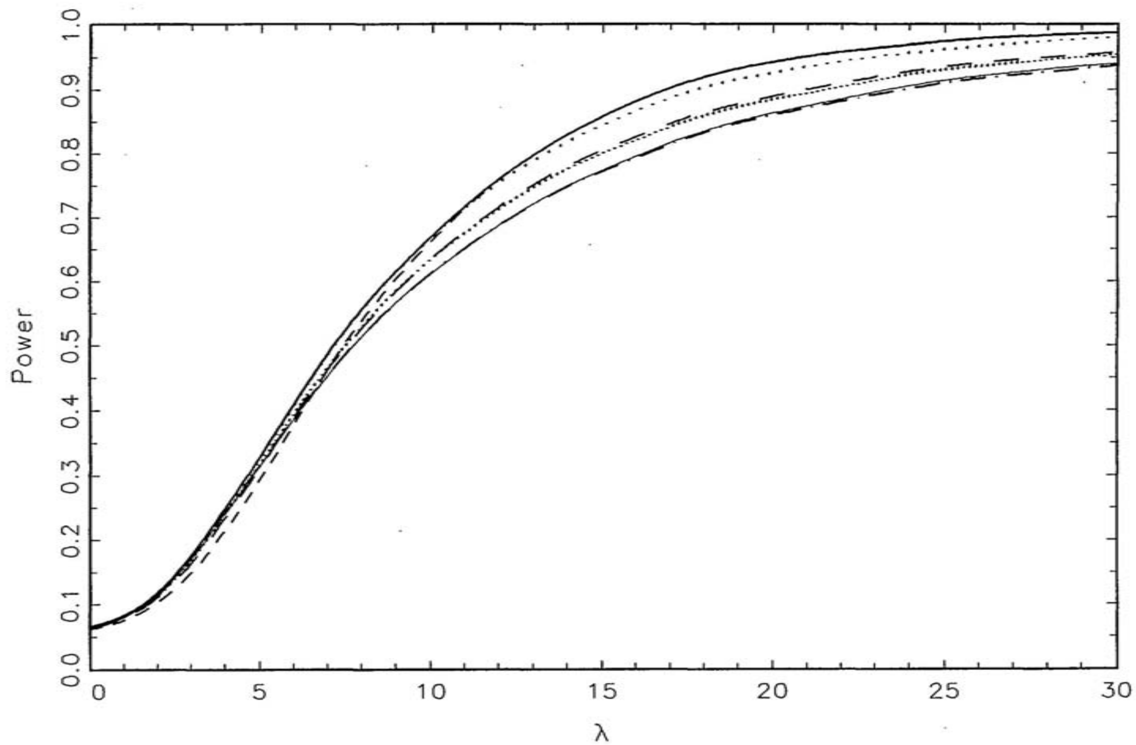


Figure 1. Asymptotic Power Functions of 5% Tests of $\tau = 0$ Against Alternatives $\tau = \lambda/T$. ——, envelope; ——, L; · – ·, MW; · · · · ·, EW; – – –, QLR; · · ·, POI(7); - - - , POI(17).

Tests for a discrete break (QLR, EW, MW) have power that is similar to tests for martingale variation (L, POI)

Summary

1. FCLT (tool)
2. Overview of TVP topics
   a. Persistent vs. mean reverting TVP
3. Testing problems
   a. Discrete break and AP approach
4. Tests
   a. Little difference between tests for discrete break and persistent "continuous" variation. What do you conclude when the null of stability is rejected?