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

Lecture 9

November 14, 2008

Stochastic time variation and low-frequency models

Outline

1. Models of Time Variation
    a. Breaks
    b. Markov Switching
    c. Martingale Variation

2. Testing and Estimation

3. Bridges between I(0) and I(1) Models

4. Can you discriminate between different models?

1. Break Models (Lecture 2)

Example (a special case of the linear regression):

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

- 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

Stochastic Breaks: Markov Switching

(a) A 2-state version of Hamilton's Markov-Switching Model:

$$y_t = \mu(s_t) + \sigma(s_t)\varepsilon_t,\ \ s_t = 0 \text{ or } 1 \text{ with } P(s_t = i \mid s_{t-1} = j) = p_{ij}$$

Rewrite as $y_t = \mu_0(1 - s_t) + \mu_1 s_t + \{\sigma_0(1 - s_t) + \sigma_1 s_t\}\,\varepsilon_t$

Issues:

(i) Filtering and Smoothing given parameters

(ii) Testing for Markov Switching:

$$H_o\!: \mu_1 = \mu_0 \text{ and } \sigma_1 = \sigma_0$$

p_{ij}'s are unidentified (Andrews-Ploberger (1994), Hansen (1992))

(iii) Estimation: MLE (easy via data augmentation/EM)

(v) Lots of extensions (use your imagination)

(iv) Changes are "recurrent"

(b) A non-recurrent model (inspired by Pesaran, Pettenuzzo, and Timmerman (2006))

Motivation: (i) Non-recurrent model; (ii) Deterministic break model is not useful for forecasting (it says nothing about post-sample breaks).

$$y_t = \mu(s_t) + \sigma(s_t)\varepsilon_t\ ,\ \ K \text{ states: } \ s_t = 1,2,\ldots, K$$

State Dynamics:

Initial condition: $s_1 = 1$.

At other dates there are two possibilities: $s_t = \begin{cases} s_{t-1} \text{ with prob } p \\ s_{t-1} + 1 \text{ with prob } 1 - p \end{cases}$

$\mu(s_t) = \mu + \eta(s_t)$ where $\eta(s_t) \sim \mathrm{N}(0, \sigma_\eta^2)$ (similar for $\sigma(s_t)$)

Or, perhaps $\mu(s_t) = \mu(s_t - 1) + \eta(s_t)$

(PTT use hierarchical Bayes method for estimation)

Other ways of incorporating instability into forecasting models:

Add factors and intercept shifts:

(i) Subjective/Judgment

(ii) Differencing (e.g., Clements and Hendry (1999))

(iii) Martingale variation in intercept: Cooley and Prescott (1973a, 1973b, 1976)

Martingale Variation: Linear Models, say $y_t = \beta_t' x_t + \varepsilon_t$ with $\beta_t = \beta_{t-1} + \eta_t$.
(Textbook References: Hamilton (1994), Harvey (1989).)

Running example: $y_t = \beta_t + \varepsilon_t$

$$\beta_t = \beta_{t-1} + \gamma\eta_t$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim iidN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right)$$

Parameters (β_0, σ, and γ)

Testing: H_0: $\gamma = 0$ (constant coefficients) versus H_a: $\gamma > 0$

Tests for martingale time variation

$$y_t = \beta_t + \varepsilon_t, \quad \beta_t = \beta_{t-1} + \gamma\eta_t$$

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 $Y = (y_1, \ldots y_T)'$, so that $Y \sim N(0, \sigma_\varepsilon^2 \Omega(\gamma))$, where $\Omega(\gamma) = I + \gamma^2 A$, 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 $\xi(s) = \sigma W(s)$. 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 $T^{-1}\psi \overset{H_o}{\Rightarrow} \int_0^1 (W(s)-sW(1))^2\,ds$

Estimation (Part 1):

$$y_t = \beta_t + \varepsilon_t, \quad \beta_t = \beta_{t-1} + \gamma\eta_t, \quad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim iidN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right)$$

Parameters (β_0, σ, and γ)

β_0: Initialize Kalman Filter with "Vague" prior $\beta_{0/0} \sim N(0, \kappa)$, where $\kappa \approx \infty$. Then $\beta_{0/T}$ is the GLS (Gaussian MLE) estimator of β_0 (give σ_ε and σ_η).

σ, and γ): Nonlinear maximization of log-likelihood (which can be computed using Kalman Filter as described in Lecture 6).

Estimation (Part 2): When γ is small, there is not much time variation in β and MLE does not work well. An alternative is "back out" an estimate of γ from the value of the TVP test statistic. (Stock and Watson (1998).)

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

Consider distribution under "local alternative" $\gamma = g/T$. Then

$$T^{-1/2}\sum_{t=1}^{[sT]} y_t = T^{-1/2}\sum_{t=1}^{[sT]} \beta_t + T^{-1/2}\sum_{t=1}^{[sT]} \varepsilon_t = gT^{-1}\sum_{t=1}^{[sT]}\left[T^{-1/2}\sum_{j=1}^{t}\eta_j\right] + T^{-1/2}\sum_{t=1}^{[sT]}\varepsilon_t$$

$$\Rightarrow A_g(s) = g\int_0^s \xi_1(\tau)d\tau + \xi_2(s)$$

so that $T^{=1}\psi = \dfrac{\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{T}}\sum_{i=t}^{T} y_i)^2}{\frac{1}{T}\sum_{t=1}^{T} y_t^2} \overset{H_a}{\Rightarrow} \int_0^1 (A_g(1)-A_g(s))^2 ds$, making it possible

to "invert" the statistic to find a confidence interval for g.

TVPs as nuissance parameters: $\theta = (\theta_1, \theta_2)$, where θ_1 is the parameter of interest and θ_2 is possibly TVP.

Question: When can you ignore possible TVP in θ_2 when conducting inference about θ_1 ?

Answer: When TVP in θ_2 is sufficiently small. But how small is small? Müller and Li (2008) (general nonlinear GMM) , Li (2008) (linear model and NKPC).

Basic idea: $y_t = \alpha + x_t\beta + \varepsilon_t$

β is the parameter of interest

α is a nuissance parameter with $\alpha_t = \alpha_{t-1} + \gamma\eta_t$

Rewrite as $y_t = (\alpha_t + \bar{x}\beta) + (x_t - \bar{x})\beta + \varepsilon_t$

or $y_t = \tilde{\alpha}_t + \tilde{x}_t\beta + \varepsilon_t$

where $\tilde{\alpha}_t = \tilde{\alpha}_{t-1} + \eta_t$

And the OLS estimator of β is $\hat{\beta} = \dfrac{\sum \tilde{x}_t y_t}{\sum \tilde{x}_t^2} = \beta + \dfrac{\sum \tilde{x}_t \tilde{\alpha}_t}{\sum \tilde{x}_t^2} + \dfrac{\sum \tilde{x}_t \varepsilon_t}{\sum \tilde{x}_t^2}$

and $$\sqrt{T}(\hat{\beta} - \beta) = \frac{\frac{1}{\sqrt{T}}\sum \tilde{x}_t \tilde{\alpha}_t}{\frac{1}{T}\sum \tilde{x}_t^2} + \frac{\frac{1}{\sqrt{T}}\sum \tilde{x}_t \varepsilon_t}{\frac{1}{T}\sum \tilde{x}_t^2}$$

The second term on the rhs is the usual source of sampling variability in the OLS estimator. Thus, the key new term is the first term on the rhs.

Term of interest: $\frac{1}{\sqrt{T}}\sum \tilde{x}_t\tilde{\alpha}_t$

Process for $\tilde{\alpha}_t$: $\tilde{\alpha}_t = \tilde{\alpha}_{t-1} + \gamma\eta_t$

Recall MUB discussion: Suppose $\gamma = g/T$, then $\frac{1}{\sqrt{T}}\sum \tilde{\alpha}_t \xrightarrow{d} Normal$

but (because $\tilde{x}$ has mean zero), if x process is "nice" $\frac{1}{\sqrt{T}}\sum \tilde{x}_t\tilde{\alpha}_t \xrightarrow{p} 0$,

So that $\sqrt{T}(\hat{\beta}-\beta) = \dfrac{\frac{1}{\sqrt{T}}\sum \tilde{x}_t\varepsilon_t}{\frac{1}{T}\sum \tilde{x}_t^2} + o_p(1)$

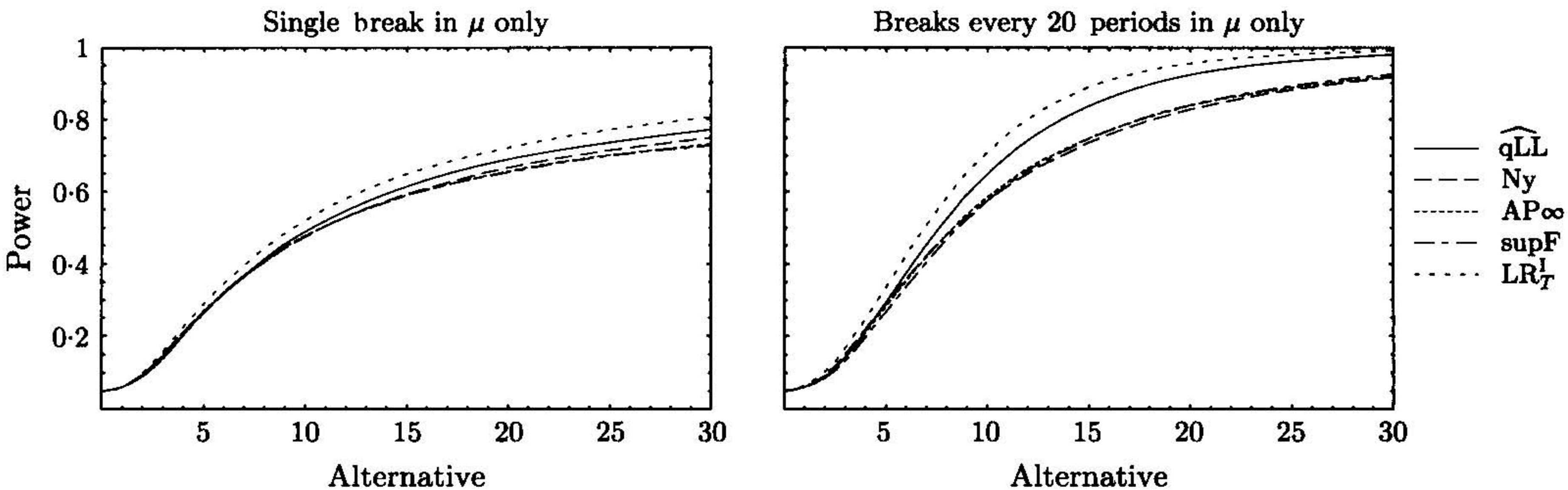
Thus, if TVP in α is not so large that you would detect it with very high probability, it doesn't matter for inference about β.

Suppose you test for TVP and you reject the null, what do you conclude?

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: $y_t = \beta_t + \varepsilon_t$

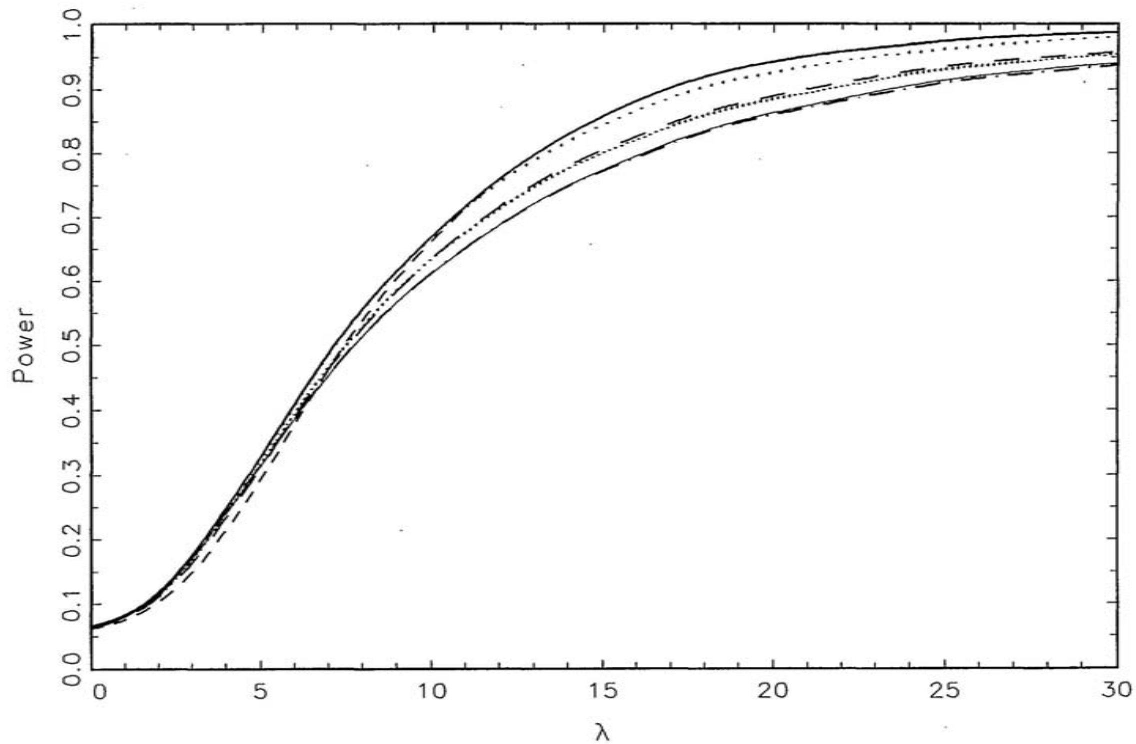


Figure 1. Asymptotic Power Functions of 5% Tests of $\tau = 0$ Against Alternatives $\tau = \lambda/T$. ——, envelope; ——, L; · — ·, MW; · · · · ·, EW; — — —, QLR; · · ·, POI(7); - - - , POI(17).

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

Models of low-frequency variability:

I(0): Same as $y_t = \varepsilon_t$

I(1): Same as $y_t = y_{t-1} + \varepsilon_t$

Three Bridges:

Local-level Model: $y_t = \beta_t + \varepsilon_t$ with $\beta_t = \beta_{t-1} + \gamma\eta_t$ $(\gamma = g/T)$

Local-to-unity AR model: $y_t = \rho y_{t-1} + \varepsilon_t$ with $\rho = (1 - c/T)$

Fractional/Long-Memory Model: $(1-\mathrm{L})^d y_t = \varepsilon_t$

Local-to-unity AR model:

$y_t = \rho y_{t-1} + \varepsilon_t$

I(0) model $|\rho| < 1$, so that $\rho^T \to 0$

I(1) model $\rho = 1$, so that $\rho^T = 1$

Local-to-unity model: $\rho = (1 - c/T)$, so that $\rho^T \to e^{-c}$

Approximations: $W(\frac{t}{T}) = \frac{1}{\sqrt{T}}\sum_{j=1}^{t}\varepsilon_j$, and denoting $J_{c,T}(\frac{t}{T}) = \frac{1}{\sqrt{T}}y_t$ we can

rewrite AR(1) model as $y_t = (1 - c/T)y_{t-1} + \varepsilon_t$

so that $J_{c,T}(\frac{t}{T}) = (1-\frac{c}{T})J_{c,T}(\frac{t-1}{T}) + W(\frac{t}{T}) - W(\frac{t-1}{T})$

or $J_{c,T}(\frac{t}{T}) - J_{c,T}(\frac{t-1}{T}) = (\frac{c}{T})J_{c,T}(\frac{t-1}{T}) + W(\frac{t}{T}) - W(\frac{t-1}{T})$

so as T grows large, the scaled y process behaves like the diffusion:

$$dJ_c(s) = cJ_c(s)ds + dW(s).$$

This can be used the compute, for example, the asymptotic distribution of the "Dickey-Fuller" test statistic under the alternative that $c = c_a$. This in turn can be used to construct a confidence interval for c. (Stock (1987)).

Largest autogressive roots for real exchange rates

$e_t = \rho e_{t-1} + \varepsilon_t$

Let h solve $\rho^h = ½$ (h is the "half-life" of a shock)

$H_0 : \rho = 1$ cannot be rejected for many real exchange rates, so $h = \infty$ cannot be rejected. But, what is a lower bound on the 95% confidence interval for h ?

Rossi (2005)

Country	Lower bound of 95% CI (Months)
Australia	61.3
Canada	60.6
Germany	4.5
Japan	21.9
Switzerland	4.0
UK	8.1

Fractional/Long-Memory Model: $(1-\mathrm{L})^d y_t = \varepsilon_t$

Where $(1-\mathrm{L})^d = 1 - d\mathrm{L} + d(d-1)\mathrm{L}^2/2 - d(d-1)(d-2)\mathrm{L}^3/3! + \ldots$

Which generates a long-AR process that is an I(1) process if $d = 1$ and an I(0) process if $d = 0$.

What can we know about low frequency variability?

How long is the sample of data? (e.g, 60 year time span)

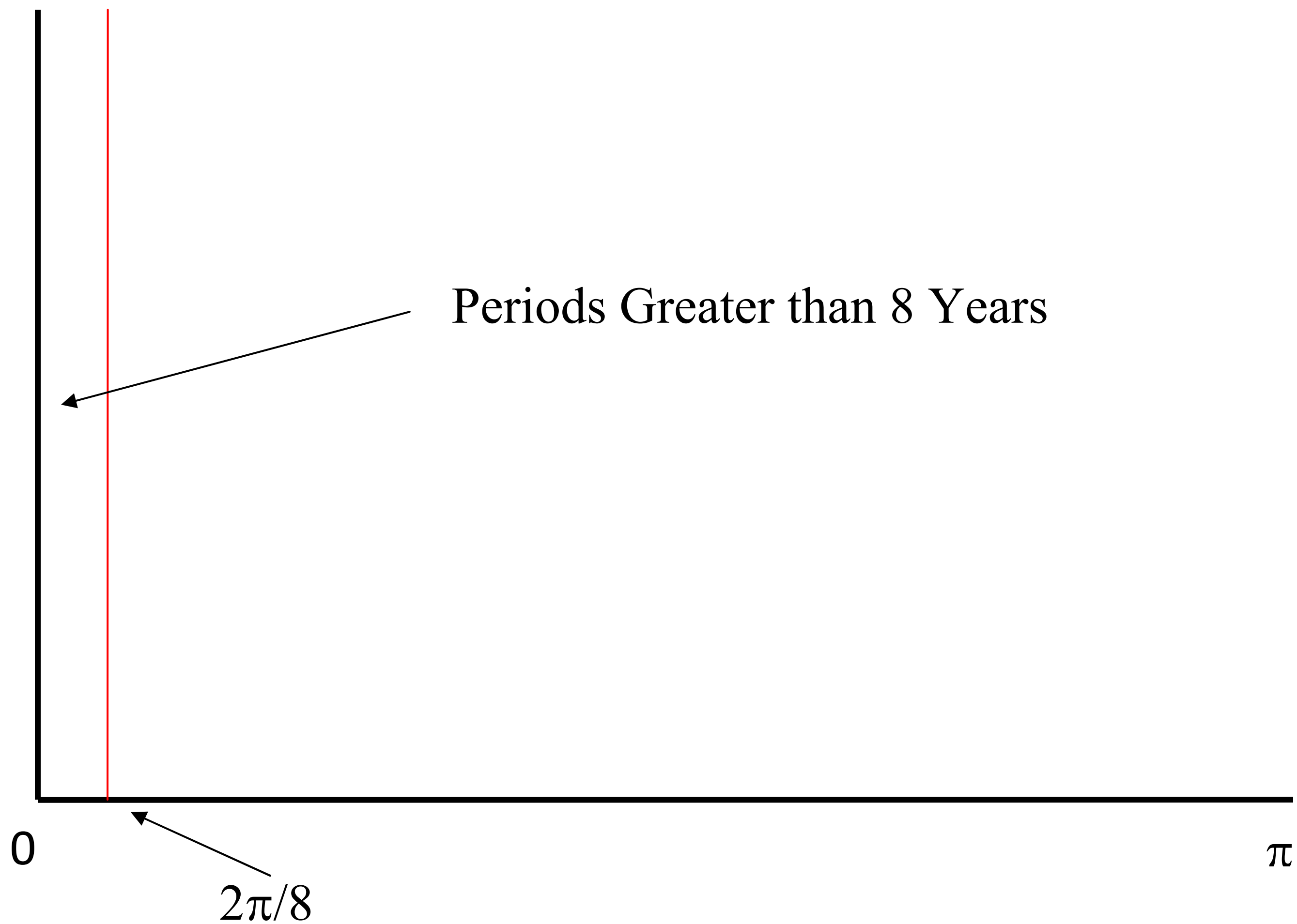
What is “low frequency” (e.g., periods longer than 8 years)

Roughly 8 non-overlaping periods of 8 years each.

Or recall periodogram analysis from our discussion of HAC. We have periodogram ordinates at frequencies $\frac{2\pi j}{T}$ for $j = 1, 2, \ldots T/2$.

Number of ordinates for periods $\geq p$ is $\frac{T}{p}$.

Low Frequency Band of the Spectrum



A calculation from Müller and Watson (2008)

Suppose you want to discriminate between a fractional model with d between 0 and 1 and a local-to-unity model using frequencies lower than $2\pi/8$.

To achieve 90% discrimination power you will need 480 years of data.