



# Time Series Analysis and Forecasting

## Chapter 12: Spectral Analysis and Frequency-Domain Methods



Daniel Traian PELE

Bucharest University of Economic Studies

IDA Institute Digital Assets

Blockchain Research Center

AI4EFin Artificial Intelligence for Energy Finance

Romanian Academy, Institute for Economic Forecasting

MSCA Digital Finance

## Learning Objectives

By the end of this chapter, you will be able to:

1. **Explain** why frequency-domain analysis complements time-domain methods
2. **Derive** the Fourier representation of a stationary process
3. **Compute** and interpret the periodogram and spectral density function
4. **Apply** nonparametric and parametric spectral estimation methods
5. **Interpret** cross-spectral measures: coherence, phase, and gain
6. **Distinguish** Fourier analysis from wavelet analysis and their use cases
7. **Use** band-pass filters to extract business cycle components
8. **Implement** spectral analysis in Python for real economic and financial data

## Outline

- Motivation
- Fourier Analysis Foundations
- The Spectral Density Function
- The Periodogram
- Spectral Estimation
- Long Memory and the Spectrum
- Cross-Spectral Analysis
- Filtering in the Frequency Domain
- Wavelet Analysis
- Case Study: Extracting Business Cycles
- AI Use Case
- Summary
- Quiz

## Why Frequency-Domain Analysis?

### Time Domain vs. Frequency Domain

- ▣ **Time domain:** describes how a series evolves over time — ACF, ARIMA, GARCH
- ▣ **Frequency domain:** describes which *periodic components* drive variation
- ▣ These are **dual representations** — mathematically equivalent, but each reveals different structure

### When Frequency-Domain Analysis is Essential

- ▣ Detecting **hidden periodicities** (business cycles, seasonal patterns, astronomical cycles)
- ▣ Understanding **which frequencies** carry the most energy (variance)
- ▣ Designing **optimal filters** to extract trend, cycle, or seasonal components
- ▣ Analyzing **co-movement** between series at specific frequencies (coherence)

### Historical Milestones

- ▣ Fourier (1807): any periodic function can be decomposed into sine and cosine waves
- ▣ Schuster (1898): invented the periodogram for detecting hidden periodicities
- ▣ Wiener-Khintchine (1930s): linked autocovariance to spectral density

## Motivating Example: Sunspot Activity

### Time-Domain View

- Monthly sunspot numbers (1749–2024)
- Obvious cyclical pattern visible
- ACF decays slowly — suggests long memory or cyclical behavior
- But: what is the *exact* period? Are there multiple cycles?

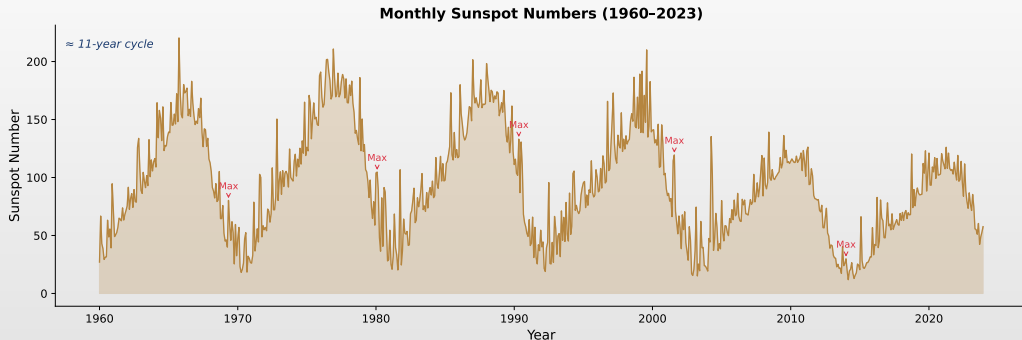
### Frequency-Domain View

- The periodogram reveals a **sharp peak** at frequency  $f \approx 0.0076$  cycles/month
- Corresponds to period  $T = 1/f \approx 131$  months  $\approx$  **11 years**
- Secondary peak at  $\approx 5.5$  years (harmonic)
- Directly identifies: “which cycles matter?”

**Key insight:** The frequency domain *decomposes variance by frequency*, revealing the dominant oscillatory components that drive the series.

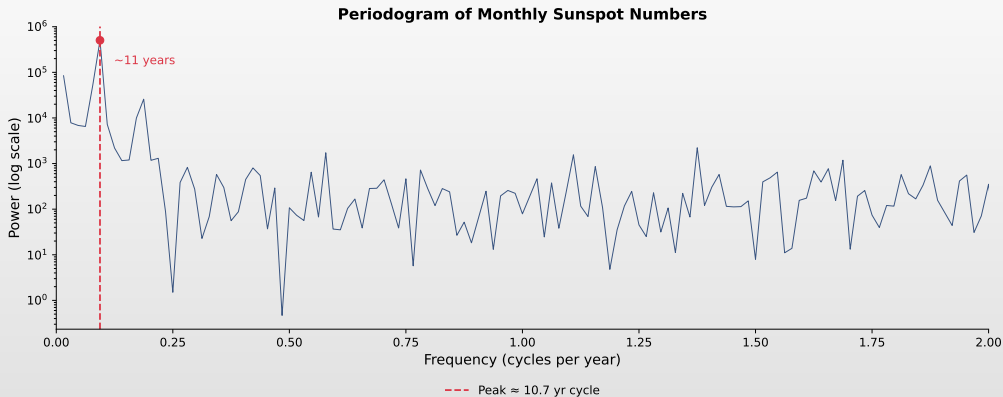
 TSA\_ch12\_sunspot

## Sunspot Time Series



- Monthly sunspot counts (1749–2024) with the  $\approx 11$ -year solar cycle
- Amplitude and shape of cycles vary — both time- and frequency-domain tools needed

## Sunspot Periodogram



- Dominant peak at  $f \approx 0.0076$  cycles/month  $\Rightarrow$  **11-year** sunspot cycle
- Secondary harmonic at  $\approx 5.5$  years confirms non-sinusoidal cycle shape

## The Spectral Representation: Big Picture

### Two Equivalent Descriptions of a Stationary Process

|                   | Time Domain                               | Frequency Domain                      |
|-------------------|-------------------------------------------|---------------------------------------|
| Summary statistic | Autocovariance $\gamma(h)$                | Spectral density $f(\omega)$          |
| Total variance    | $\gamma(0) = \text{Var}(Y_t)$             | $\int_{-\pi}^{\pi} f(\omega) d\omega$ |
| Relationship      | Fourier transform pair (Wiener–Khinchine) |                                       |
| Model class       | ARMA( $p, q$ )                            | Rational spectral density             |
| Estimation        | Sample ACF                                | Periodogram / spectral estimators     |
| Filter analysis   | Convolution (hard)                        | Multiplication (easy)                 |

### Why Learn Both?

- Filtering (trend extraction, seasonal adjustment) is **multiplication** in frequency domain but **convolution** in time domain
- Spectral density gives a complete, **visual summary** of the variance structure
- Some phenomena (business cycles, signal interference) are *naturally described* by frequency

## Fourier Series: Decomposing Periodic Functions

### Fourier's Fundamental Insight

Any periodic function  $g(t)$  with period  $T$  can be written as:

$$g(t) = a_0 + \sum_{k=1}^{\infty} \left[ a_k \cos\left(\frac{2\pi kt}{T}\right) + b_k \sin\left(\frac{2\pi kt}{T}\right) \right]$$

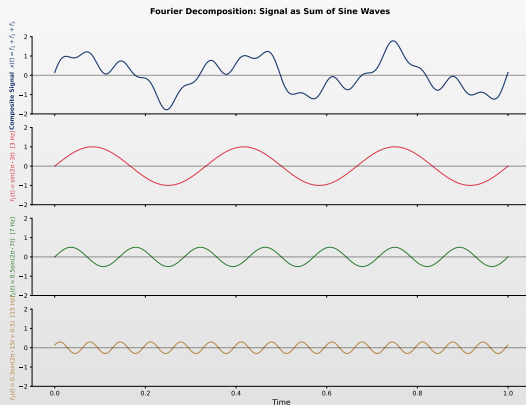
where the Fourier coefficients are:

$$a_0 = \frac{1}{T} \int_0^T g(t) dt, \quad a_k = \frac{2}{T} \int_0^T g(t) \cos\left(\frac{2\pi kt}{T}\right) dt, \quad b_k = \frac{2}{T} \int_0^T g(t) \sin\left(\frac{2\pi kt}{T}\right) dt$$

### Key Terminology

- ▣ **Frequency:**  $f_k = k/T$  cycles per unit time; **Angular frequency:**  $\omega_k = 2\pi f_k$  radians per unit time
- ▣ **Fundamental frequency:**  $f_1 = 1/T$ ; **Harmonics:**  $f_k = k \cdot f_1$  for  $k = 2, 3, \dots$
- ▣ **Amplitude:**  $R_k = \sqrt{a_k^2 + b_k^2}$ ; **Phase:**  $\phi_k = \arctan(b_k/a_k)$

## Fourier Decomposition Illustration



- Complex periodic signal (top) decomposed into fundamental + harmonic components
- Each row = one frequency component; the sum reconstructs the original exactly

## Complex Exponential Form

### Euler's Formula and Complex Fourier Series

Using Euler's formula  $e^{i\omega t} = \cos(\omega t) + i \sin(\omega t)$ , we can write:

$$g(t) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kt/T}, \quad c_k = \frac{1}{T} \int_0^T g(t) e^{-i2\pi kt/T} dt$$

where  $c_k = (a_k - ib_k)/2$  for  $k > 0$  and  $c_{-k} = \overline{c_k}$  (complex conjugate).

### Why Complex Notation?

- More compact: one formula instead of separate sine/cosine terms
- Algebraically simpler: convolution becomes multiplication
- Standard in signal processing, physics, and modern time series analysis

## Parseval's Theorem (Energy Conservation)

Total Energy is Preserved Across Representations

$$\frac{1}{T} \int_0^T |g(t)|^2 dt = \sum_{k=-\infty}^{\infty} |c_k|^2$$

Each  $|c_k|^2$  measures the **contribution of frequency  $k/T$  to total variance**.

### Implication

- Energy in time domain = energy in frequency domain
- This is the foundation for spectral decomposition of variance

## Discrete Fourier Transform (DFT) I: Definition

### From Continuous to Discrete

Given  $n$  observations  $y_1, y_2, \dots, y_n$ , the DFT is:

$$J(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^n y_t e^{-i\omega_j t}, \quad \omega_j = \frac{2\pi j}{n}, \quad j = 0, 1, \dots, n-1$$

The frequencies  $\omega_j$  are called the **Fourier frequencies** or **fundamental frequencies**.

### Orthogonality of the DFT

- The Fourier basis vectors are orthogonal:  $\sum_{t=1}^n e^{i(\omega_j - \omega_k)t} = n \cdot 1_{j=k}$
- Different frequencies are uncorrelated — each captures a distinct component of variation
- The DFT is a unitary transformation (up to scaling) of the data vector

## Discrete Fourier Transform (DFT) II: Properties

### Invertibility: Perfect Reconstruction

The DFT is perfectly invertible — the original data can be exactly reconstructed:

$$y_t = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} J(\omega_j) e^{i\omega_j t}, \quad t = 1, \dots, n$$

No information is lost in the frequency-domain representation.

### The Nyquist Frequency

- The highest detectable frequency is  $\omega_{n/2} = \pi$  radians per period
- Requires at least 2 observations per cycle — faster oscillations cannot be resolved
- Frequencies above Nyquist are “aliased” (folded back) — see next slides

### Fast Fourier Transform (FFT)

- Naïve DFT:  $O(n^2)$  operations. FFT (Cooley–Tukey, 1965):  $O(n \log n)$
- One of the most important algorithms of the 20th century

## Aliasing and the Nyquist Frequency

### The Sampling Theorem (Shannon–Nyquist)

A continuous signal with maximum frequency  $f_{\max}$  can be perfectly reconstructed from discrete samples if the sampling rate satisfies:

$$f_s > 2f_{\max} \quad (\text{Nyquist criterion})$$

The **Nyquist frequency**  $f_N = f_s/2$  is the highest frequency we can detect.

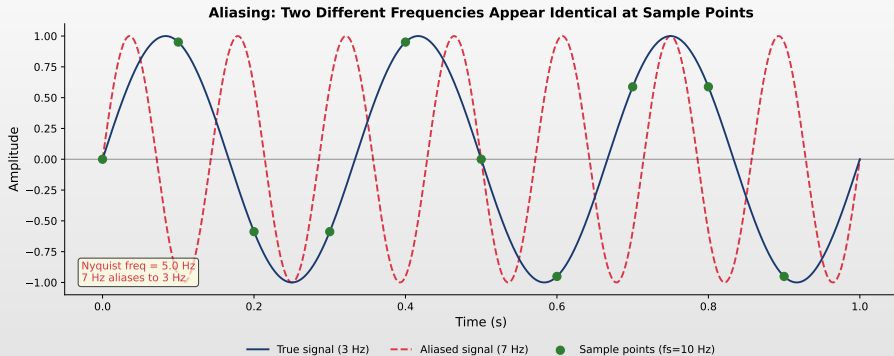
### Aliasing: What Goes Wrong

- If a signal contains frequencies *above*  $f_N$ , they are “folded back” and appear as lower frequencies
- **Example:** daily data ( $f_s = 1/\text{day}$ ,  $f_N = 0.5/\text{day}$ ). An intraday cycle at 0.7 cycles/day appears as 0.3 cycles/day
- Aliasing is **irreversible** — once data is sampled, aliased frequencies cannot be separated

### Practical Implication for Economic Data

- Monthly GDP data:  $f_N = 0.5$  cycles/month  $\Rightarrow$  cannot detect cycles shorter than 2 months
- Quarterly data: can only detect cycles  $\geq 6$  months

## Aliasing Illustrated



- High-frequency sinusoid (dashed) sampled too sparsely  $\Rightarrow$  indistinguishable from lower-frequency alias (solid)
- Anti-aliasing filters must be applied *before* sampling to prevent irreversible distortion

## Section Summary: Fourier Analysis

### Key Concepts

- ▣ **DFT** decomposes a series into orthogonal sinusoids at frequencies  $\omega_j = 2\pi j/n$
- ▣ **Parseval's theorem**: total energy preserved across domains
- ▣ **Aliasing**: frequencies above  $\omega_N = \pi$  are aliased — sample at  $\geq 2f_{\max}$
- ▣ **Next**: How to estimate the spectral density from data

## Definition and the Wiener–Khintchine Theorem

### Theorem 1 (Wiener–Khintchine Theorem)

Let  $\{Y_t\}$  be a weakly stationary process with autocovariance function  $\gamma(h)$  satisfying  $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ . Then the **spectral density function** (SDF) exists and is given by:

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h}, \quad -\pi \leq \omega \leq \pi$$

The **inverse** relationship is:

$$\gamma(h) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega h} d\omega$$

### Properties of the Spectral Density

- ▣  $f(\omega) \geq 0$  for all  $\omega$  (non-negative — it represents variance contribution)
- ▣  $f(\omega) = f(-\omega)$  (symmetric for real-valued processes)
- ▣  $\gamma(0) = \text{Var}(Y_t) = \int_{-\pi}^{\pi} f(\omega) d\omega$  (total variance = area under  $f$ )
- ▣  $f$  is **periodic** with period  $2\pi$ , so we only need  $\omega \in [-\pi, \pi]$  or  $[0, \pi]$

## Spectral Density: White Noise and AR(1)

White Noise:  $Y_t = \varepsilon_t, \varepsilon_t \sim \text{WN}(0, \sigma^2)$

$$\gamma(h) = \sigma^2 1_{h=0} \quad \implies \quad f(\omega) = \frac{\sigma^2}{2\pi} \quad (\text{flat — equal power at all frequencies})$$

The spectral density is constant: white noise contributes equally at every frequency.

AR(1):  $Y_t = \phi Y_{t-1} + \varepsilon_t$

$$f(\omega) = \frac{\sigma^2}{2\pi} \cdot \frac{1}{|1 - \phi e^{-i\omega}|^2} = \frac{\sigma^2}{2\pi(1 - 2\phi \cos \omega + \phi^2)}$$

- ▣  $\phi > 0$ : **red spectrum** — peaks at  $\omega = 0$ , power concentrated at low frequencies (trend-like)
- ▣  $\phi < 0$ : **blue spectrum** — peaks at  $\omega = \pi$ , power at high frequencies (alternating)
- ▣  $|\phi| \rightarrow 1$ : spectrum becomes more concentrated, approaching a spike

## Spectral Density: MA(1) and General ARMA

MA(1):  $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$

$$f(\omega) = \frac{\sigma^2}{2\pi} \cdot |1 + \theta e^{-i\omega}|^2 = \frac{\sigma^2}{2\pi} (1 + 2\theta \cos \omega + \theta^2)$$

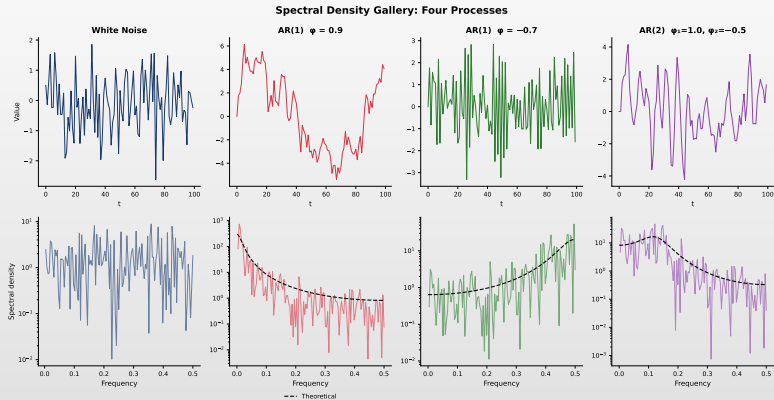
- ▣  $\theta > 0$ : more power at low frequencies;  $\theta < 0$ : more power at high frequencies
- ▣ Spectral shape is a simple cosine modulation of the white noise spectrum

General ARMA( $p, q$ ):  $\phi(B)Y_t = \theta(B)\varepsilon_t$

$$f(\omega) = \frac{\sigma^2}{2\pi} \cdot \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} \quad (\text{rational spectral density})$$

- ▣ AR poles  $\Rightarrow$  peaks in the spectrum; MA zeros  $\Rightarrow$  troughs (spectral zeros)
- ▣ The rational form makes parametric spectral estimation natural for ARMA processes

## Spectral Density Gallery: Visual Comparison



- White noise, AR(1) ( $\phi = 0.9$ ), AR(1) ( $\phi = -0.7$ ), AR(2), and seasonal processes
- Each model's autocorrelation structure maps to a distinctive spectral shape

## The Spectral Representation Theorem

### Theorem 2 (Spectral Representation / Cramér Representation)

Any weakly stationary process  $\{Y_t\}$  with  $\mathbb{E}[Y_t] = 0$  can be represented as:

$$Y_t = \int_{-\pi}^{\pi} e^{i\omega t} dZ(\omega)$$

where  $Z(\omega)$  is an orthogonal-increment process satisfying:

- ▣  $\mathbb{E}[dZ(\omega)] = 0$
- ▣  $\mathbb{E}[|dZ(\omega)|^2] = dF(\omega)$  where  $F(\omega)$  is the spectral distribution function
- ▣  $\mathbb{E}[dZ(\omega_1) dZ(\omega_2)] = 0$  for  $\omega_1 \neq \omega_2$

### Interpretation

- ▣ Every stationary process is a **superposition of uncorrelated oscillations** at different frequencies
- ▣  $dZ(\omega)$  gives the (random) amplitude and phase at frequency  $\omega$
- ▣ The spectral density  $f(\omega) = dF(\omega)/d\omega$  measures **variance per unit frequency**
- ▣ This is the **frequency-domain analog of the Wold decomposition**

## Definition and Interpretation

### Definition 1 (Periodogram)

The **periodogram** of  $\{y_t\}_{t=1}^n$  at frequency  $\omega_j = 2\pi j/n$  is:

$$I(\omega_j) = \frac{1}{2\pi} |J(\omega_j)|^2 = \frac{1}{2\pi n} \left| \sum_{t=1}^n y_t e^{-i\omega_j t} \right|^2$$

Equivalently, using trigonometric form:

$$I(\omega_j) = \frac{n}{4\pi} (\hat{a}_j^2 + \hat{b}_j^2)$$

where  $\hat{a}_j = \frac{2}{n} \sum_t y_t \cos(\omega_j t)$  and  $\hat{b}_j = \frac{2}{n} \sum_t y_t \sin(\omega_j t)$ .

### Interpretation

- ▣  $I(\omega_j)$  measures the **contribution of frequency  $\omega_j$  to total sample variance**
- ▣ Large  $I(\omega_j) \Rightarrow$  strong oscillation at period  $T_j = 2\pi/\omega_j = n/j$
- ▣ The periodogram decomposes variance:  $\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2 = \sum_{j=1}^{\lfloor n/2 \rfloor} I(\omega_j)$  (up to constants)

## Periodogram and Sample Autocovariance

### Theorem 3 (Periodogram–Autocovariance Relationship)

The periodogram is the Fourier transform of the sample autocovariance function:

$$I(\omega) = \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{-i\omega h}$$

where  $\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (y_t - \bar{y})(y_{t+|h|} - \bar{y})$  is the sample autocovariance at lag  $h$ .

### This is the Sample Analog of Wiener–Khinchine

- ▣ Population:  $f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h}$  (spectral density)
- ▣ Sample:  $I(\omega) = \frac{1}{2\pi} \sum_h \hat{\gamma}(h) e^{-i\omega h}$  (periodogram)
- ▣ The periodogram is a **natural estimator** of the spectral density — but is it a *good* one?

### Key Limitation: The Periodogram is Inconsistent

- ▣  $\mathbb{E}[I(\omega_j)] \rightarrow f(\omega)$  as  $n \rightarrow \infty$  (asymptotically unbiased)
- ▣ But  $\text{Var}[I(\omega_j)] \not\rightarrow 0$  — the variance does *not* shrink with more data!

## Reading a Periodogram: Common Patterns

### Periodogram Signatures of Common Processes

| Process                  | Spectral Shape                                                   | Periodogram Pattern              |
|--------------------------|------------------------------------------------------------------|----------------------------------|
| White noise              | Flat: $f(\omega) = \sigma^2/(2\pi)$                              | Roughly uniform, no peaks        |
| AR(1), $\phi > 0$        | Peaks at $\omega = 0$ , decays                                   | Large values at low frequencies  |
| AR(1), $\phi < 0$        | Peaks at $\omega = \pi$ , decays                                 | Large values at high frequencies |
| AR(2), complex roots     | Peak at $\omega_0 = \cos^{-1}(\phi_1/(2\sqrt{-\phi_2}))^\dagger$ | Hump at cyclical frequency       |
| Random walk <sup>‡</sup> | $f(\omega) \propto \omega^{-2}$                                  | Strong low-frequency dominance   |
| Seasonal ( $s = 12$ )    | Peaks at $\omega = 2\pi k/12$                                    | Spikes at seasonal harmonics     |

<sup>†</sup> Approximate — derived from the characteristic roots of the AR(2) polynomial. <sup>‡</sup> Non-stationary;  $f(\omega) \propto \omega^{-2}$  is the pseudo-spectral density.

### Practical Rule of Thumb

- ▣ **Red spectrum** (power at low  $\omega$ ): trend-dominated, positively autocorrelated
- ▣ **Blue spectrum** (power at high  $\omega$ ): noise-dominated, negatively autocorrelated
- ▣ **Sharp peaks**: deterministic or near-deterministic periodic components

## Statistical Properties of the Periodogram

### Asymptotic Distribution (Gaussian Process)

For a stationary Gaussian process with spectral density  $f(\omega)$ :

$$\frac{2 I(\omega_j)}{f(\omega_j)} \xrightarrow{d} \chi_2^2 \quad \text{for } 0 < \omega_j < \pi$$

- ▣  $\mathbb{E}[I(\omega_j)] \approx f(\omega_j)$  — approximately unbiased
- ▣  $\text{Var}[I(\omega_j)] \approx f(\omega_j)^2$  — variance equals the squared mean
- ▣ Periodogram values at **different Fourier frequencies are asymptotically independent**

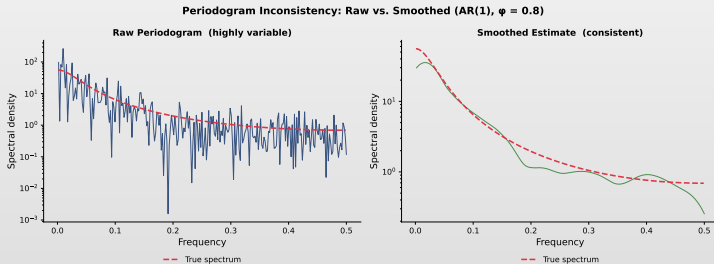
### The Inconsistency Problem

- ▣ The coefficient of variation  $\text{CV} = \text{SD}/\text{Mean} = 1$  regardless of sample size
- ▣ Adding more data adds more frequencies but does **not** reduce noise at any given frequency
- ▣ The raw periodogram is **not a consistent estimator** of the spectral density
- ▣ **Solution:** smooth the periodogram (see Section 5)

## Periodogram Inconsistency Illustrated

### Key Result: Distribution of the Periodogram

- For large  $n$ , at non-zero, non-Nyquist frequencies:  $I(\omega_j) \stackrel{d}{\approx} f(\omega_j) \cdot \chi^2_2/2$
- $\text{Var}[I(\omega_j)] \approx f(\omega_j)^2$  — **does not shrink as  $n \rightarrow \infty$** : periodogram is *inconsistent*



## Fisher's Test for Hidden Periodicities

### Is That Peak Real or Just Noise?

Fisher (1929) proposed testing  $H_0$ : the series is white noise against  $H_1$ : there is a hidden periodicity. The test statistic is:

$$g = \frac{\max_j I(\omega_j)}{\sum_{j=1}^{\lfloor n/2 \rfloor} I(\omega_j)}$$

Under  $H_0$ , the exact distribution of  $g$  is known. For large  $n$ :

$$P(g > x) \approx m(1-x)^{m-1} - \binom{m}{2}(1-2x)_+^{m-1} + \dots$$

where  $m = \lfloor n/2 \rfloor$  and  $(a)_+ = \max(a, 0)$ .

### Interpretation

- If the largest periodogram value is significantly larger than expected under white noise, reject  $H_0$
- Can be applied iteratively: remove detected frequency, test again
- Modern alternative: test individual frequencies using the  $\chi^2_2$  distribution

## The Estimation Problem

Goal: Estimate  $f(\omega)$  from Data

- We need a **consistent** estimator: both bias  $\rightarrow 0$  and variance  $\rightarrow 0$  as  $n \rightarrow \infty$ . Two main approaches:

### Nonparametric Methods

- Smooth the raw periodogram
- No model assumptions
- **Lag-window** (Blackman–Tukey)
- **Segment-averaging** (Bartlett, Welch)
- **Multitaper** (Thomson)
- Bias-variance tradeoff via bandwidth

### Parametric Methods

- Fit a model, derive its spectral density
- Assumes a specific model form
- **AR spectral estimation**
- Higher resolution for short series
- Risk of misspecification
- AIC/BIC for model selection

**Key insight:** Consistent spectral estimation requires sacrificing frequency resolution. This is a fundamental bias-variance tradeoff — you cannot have both perfect resolution and low variance.

## Smoothed Periodogram (Daniell Kernel)

### Idea: Average Neighboring Periodogram Values

The **smoothed periodogram** at frequency  $\omega$  is:

$$\hat{f}(\omega) = \sum_{k=-m}^m w_k I(\omega + \omega_k)$$

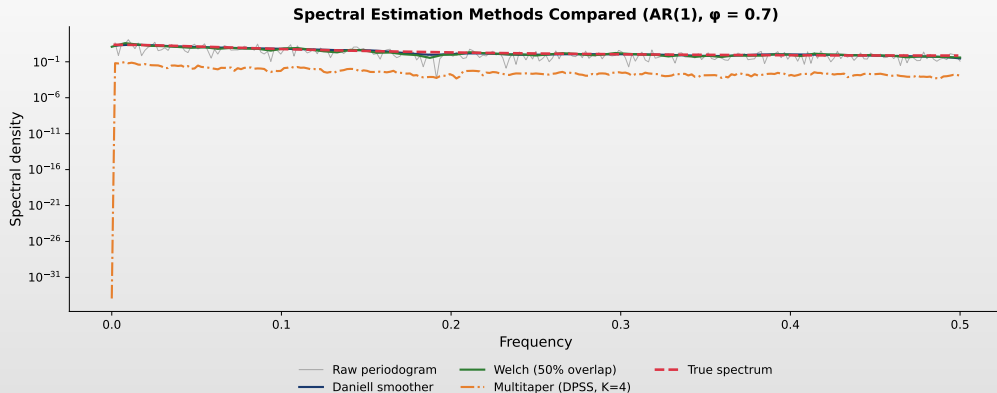
where  $\{w_k\}$  are non-negative weights summing to 1 and  $m$  is the **bandwidth**. The simplest choice is the Daniell (uniform) kernel:  $w_k = 1/(2m+1)$  for  $|k| \leq m$ . Equivalent degrees of freedom:  $\nu \approx 2(2m+1)$ .

### Bandwidth Choice: The Fundamental Tradeoff

- ▣ Small  $m$ : low bias, high variance (close to raw periodogram)
- ▣ Large  $m$ : low variance, high bias (oversmoothed, peaks blurred)
- ▣ Approximately  $\hat{f}(\omega) \sim f(\omega) \cdot \chi_\nu^2 / \nu$

Note: "Bandwidth" in smoothing ( $m$ , the number of averaged frequencies) differs from "bandwidth"  $W$  in multitaper analysis (the spectral concentration half-width). Context always clarifies which is meant.

## Smoothing Comparison



- Raw periodogram vs. Daniell-smoothed estimates with bandwidth  $m = 3, 8, 20$
- Wider bandwidth reduces variance but blurs peaks — bias-variance tradeoff

## Lag-Window (Blackman–Tukey) Estimation

### Alternative: Taper the Autocovariances

Instead of smoothing in frequency domain, apply a window to the sample autocovariances:

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{h=-(M-1)}^{M-1} w(h/M) \hat{\gamma}(h) e^{-i\omega h}$$

where  $w(\cdot)$  is a **lag window** (kernel) and  $M$  is the **truncation point**.

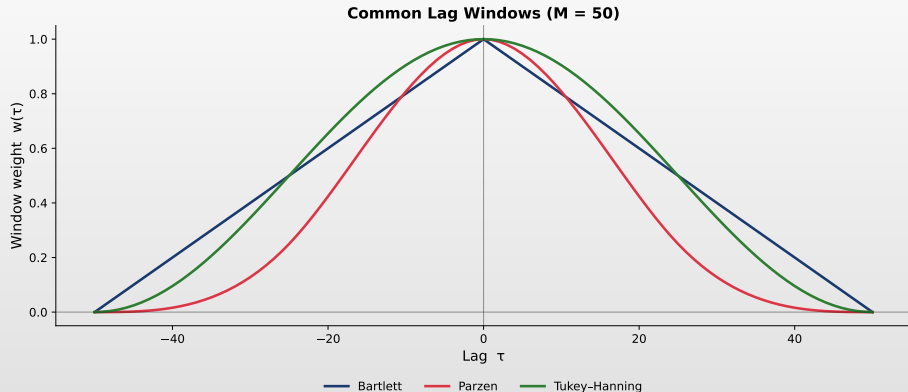
### Common Lag Windows

| Window              | Formula $w(x),  x  \leq 1$                                                   | Properties                                                                                             |
|---------------------|------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| Bartlett (triangle) | $1 -  x $                                                                    | Simple, non-negative spectrum<br>Smooth, non-negative<br>Good resolution<br>Optimal for smooth spectra |
| Parzen              | $1 - 6x^2 + 6 x ^3$ for $ x  \leq 1/2$ , else $2(1 -  x )^3$                 |                                                                                                        |
| Tukey–Hanning       | $\frac{1}{2}(1 + \cos(\pi x))$                                               |                                                                                                        |
| Bartlett–Priestley  | $\frac{3}{(\pi x)^2} \left( \frac{\sin(\pi x)}{\pi x} - \cos(\pi x) \right)$ |                                                                                                        |

### Choosing $M$

- Rule of thumb:  $M \approx 2\sqrt{n}$ . Cross-validation or plug-in selectors for automatic choice.

## Lag Windows Illustrated



- Bartlett, Parzen, Tukey-Hanning, and Bartlett-Priestley lag windows (left) and spectral windows (right)
- Smoother lag windows  $\Rightarrow$  better-behaved spectral estimates with non-negative values

## Welch's Method (Segment Averaging)

### Idea: Divide and Average

Welch (1967) proposed:

1. Divide the series of length  $n$  into  $K$  overlapping segments of length  $L$
2. Apply a data taper (e.g., Hanning window) to each segment
3. Compute the periodogram of each tapered segment
4. Average the  $K$  periodograms:  $\hat{f}_{\text{Welch}}(\omega) = \frac{1}{K} \sum_{k=1}^K I^{(k)}(\omega)$

### Advantages

- ▣ Variance reduced by factor  $\approx 1/K$
- ▣ Tapering reduces spectral leakage
- ▣ 50% overlap is standard

### Tradeoff

- ▣ Frequency resolution  $\Delta f = 1/L$
- ▣ Shorter segments  $\Rightarrow$  worse resolution
- ▣ Longer segments  $\Rightarrow$  fewer averages  $\Rightarrow$  higher variance
- ▣ Typical:  $L = n/4$  to  $n/8$  with 50% overlap

## Multitaper Spectral Estimation

### Thomson's Multitaper Method (1982)

Instead of one taper (window), use  $K$  **orthogonal tapers**  $\{h_k(t)\}_{k=0}^{K-1}$ :

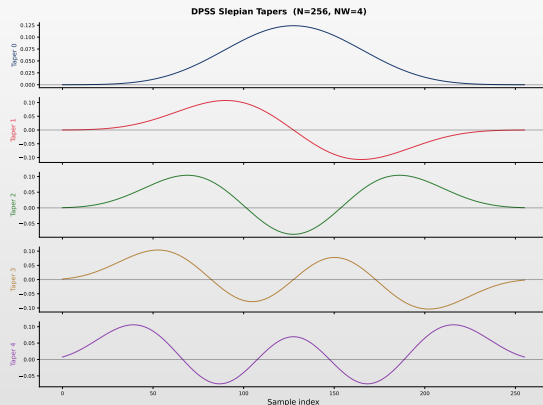
$$\hat{f}_{\text{MT}}(\omega) = \frac{1}{2\pi K} \sum_{k=0}^{K-1} \left| \sum_{t=1}^n h_k(t) y_t e^{-i\omega t} \right|^2$$

The tapers are **discrete prolate spheroidal sequences** (DPSS / Slepian sequences), which optimally concentrate energy in the frequency band  $[-W, W]$ .

### Advantages of the Multitaper Method

- ▣ Uses **all the data** (no segmenting) — maximum frequency resolution
- ▣  $K$  orthogonal tapers provide  $K$  approximately independent spectral estimates
- ▣ Near-optimal bias-variance tradeoff controlled by bandwidth  $W$ ;  $K = 2nW - 1$
- ▣  $NW = 3-4$  is standard ( $K = 5-8$ ). Python: `scipy.signal.windows.dpss()`

## Multitaper DPSS Tapers



- First DPSS (Slepian) tapers for  $NW = 4$ ; each successive taper has one extra zero crossing
- Together they provide nearly independent spectral estimates concentrated in  $[-W, W]$

## Parametric Spectral Estimation

### AR-Based Spectral Estimation

If we fit an  $AR(p)$  model:  $Y_t = \sum_{k=1}^p \hat{\phi}_k Y_{t-k} + \varepsilon_t$ , the spectral density is:

$$\hat{f}_{AR}(\omega) = \frac{\hat{\sigma}^2}{2\pi |1 - \hat{\phi}_1 e^{-i\omega} - \dots - \hat{\phi}_p e^{-ip\omega}|^2}$$

#### Advantages

- ▣ High resolution even for short series
- ▣ Smooth estimate (no erratic fluctuations)
- ▣ Can resolve closely-spaced peaks
- ▣ Yule–Walker or Burg's algorithm

#### Disadvantages

- ▣ Assumes AR model is correct
- ▣ May miss features not captured by AR
- ▣ Order  $p$  selection critical (AIC, BIC)
- ▣ Cannot represent sharp spectral zeros

**Recommendation:** Use AR spectral estimation for short series ( $n < 100$ ) or when the process is believed to be autoregressive. Use nonparametric methods for longer series or unknown spectral shape.

## Comparison of Spectral Estimation Methods

### Method Comparison Summary

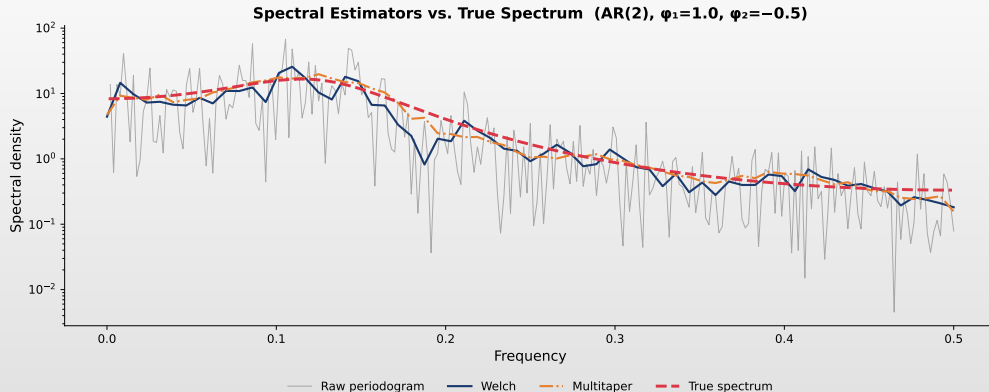
| Method            | Resolution  | Variance               | Assumptions        | Best For                |
|-------------------|-------------|------------------------|--------------------|-------------------------|
| Raw periodogram   | Highest     | Highest (inconsistent) | None               | Initial exploration     |
| Daniell smoothing | Medium      | Medium                 | Locally smooth $f$ | Quick smoothed estimate |
| Blackman–Tukey    | Medium      | Low–medium             | Smooth $f$         | Classical analysis      |
| Welch's method    | Low–medium  | Low                    | Locally stationary | Long series, real-time  |
| Multitaper        | Medium–high | Low                    | None               | Gold standard           |
| AR parametric     | Very high   | Very low               | AR model correct   | Short series            |

### Practical Workflow

1. Plot the raw periodogram (on log scale) for initial exploration
2. Apply multitaper estimation for the primary analysis
3. Cross-check with AR spectral estimate if peaks need higher resolution
4. Use confidence intervals:  $\hat{f}(\omega) \cdot \chi^2_{\nu}/\nu$  with  $\nu$  effective degrees of freedom



## Spectral Estimation Comparison



- Six estimation methods applied to the same AR(2) series for direct comparison
- Multitaper achieves the best resolution-variance balance among nonparametric methods

## Spectral Leakage and Data Tapering

### The Leakage Problem

- Computing the DFT implicitly assumes the data is periodic (wraps around)
- If the series does not start and end at the same value, this creates **discontinuities**
- Energy from strong spectral peaks “leaks” into neighboring frequencies
- Result: periodogram shows broad humps instead of sharp peaks

### Solution: Data Tapering (Windowing)

Apply a taper  $\{d_t\}$  that smoothly brings the data to zero at both ends:

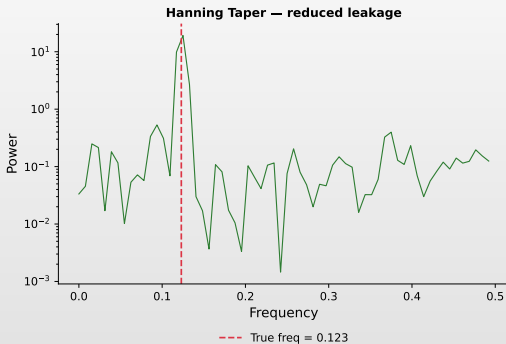
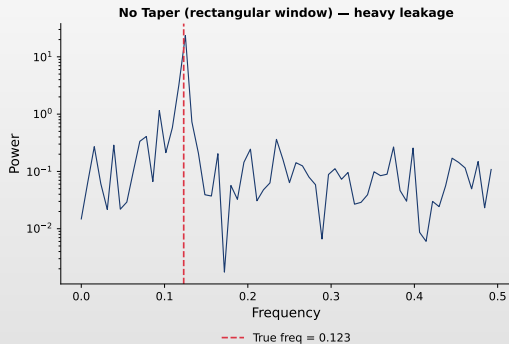
$$y_t^{\text{tapered}} = d_t \cdot y_t, \quad \text{where } d_1 \approx d_n \approx 0, \quad d_{n/2} \approx 1$$

Common tapers:

- **Hanning**:  $d_t = 0.5(1 - \cos(2\pi t/n))$  — good general-purpose taper
- **Hamming**:  $d_t = 0.54 - 0.46 \cos(2\pi t/n)$  — slightly less leakage reduction
- **Blackman**:  $d_t = 0.42 - 0.5 \cos(2\pi t/n) + 0.08 \cos(4\pi t/n)$  — excellent leakage suppression
- **DPSS (Slepian)**: optimal energy concentration (used in multitaper)

## Spectral Leakage Illustrated

**Spectral Leakage: Effect of Tapering**



- Without tapering (left): severe leakage of spectral energy into nearby frequencies
- With Hanning taper (right): peak is well-concentrated, leakage floor greatly reduced

## Section Summary: Spectral Estimation

### Estimator Comparison

- ▣ **Periodogram**: unbiased, inconsistent — high variance
- ▣ **Smoothed (Daniell/BT)**: bias–variance tradeoff via bandwidth  $M$
- ▣ **Welch**: reduces variance by segment averaging; lower resolution
- ▣ **Multitaper**: optimal bias–variance via DPSS tapers; preferred in practice
- ▣ **Parametric**: efficient if model correct; sensitive to misspecification

## Long Memory Processes

### ARFIMA and Spectral Behavior Near the Origin

An ARFIMA( $p, d, q$ ) process with fractional difference parameter  $d \in (-0.5, 0.5)$ :

$$\phi(B)(1 - B)^d Y_t = \theta(B)\varepsilon_t$$

has spectral density satisfying:

$$f(\omega) \sim C_f |\omega|^{-2d} \quad \text{as } \omega \rightarrow 0$$

For  $0 < d < 0.5$ : long memory — the spectrum **diverges** at the origin (unbounded low-frequency power).

### Connection to the Hurst Exponent

- ▣ Hurst exponent:  $H = d + 1/2$ , so  $H > 1/2$  indicates long memory
- ▣ Autocovariances decay hyperbolically:  $\gamma(h) \sim C |h|^{2d-1}$  (not summable for  $d > 0$ )
- ▣ The Wiener–Khinchine theorem still applies, but the spectrum is unbounded at  $\omega = 0$

### Examples in Economics and Finance

- ▣ Realized volatility ( $d \approx 0.4$ ), inflation persistence, interest rate spreads, trading volume.

## Spectral Regression Estimators of $d$

### GPH Estimator (Geweke & Porter-Hudak, 1983)

Regress  $\log I(\omega_j)$  on  $\log |2 \sin(\omega_j/2)|^2$  at the lowest  $m$  Fourier frequencies:

$$\log I(\omega_j) = \beta_0 - d \cdot \log(4 \sin^2(\omega_j/2)) + \varepsilon_j, \quad j = 1, \dots, m$$

where  $m = n^\alpha$  with  $\alpha \in (0, 1)$ , typically  $\alpha = 0.5$ – $0.8$ .

- Simple log-periodogram regression; consistent but inefficient
- $\hat{d}_{\text{GPH}} \xrightarrow{d} N(d, \pi^2/(24m))$  under regularity conditions

### Local Whittle Estimator (Robinson, 1995)

Maximizes the local Whittle likelihood over the  $m$  lowest frequencies:

$$\hat{d}_{\text{LW}} = \arg \min_d \left[ \log \left( \frac{1}{m} \sum_{j=1}^m \frac{I(\omega_j)}{\omega_j^{-2d}} \right) - \frac{2d}{m} \sum_{j=1}^m \log \omega_j \right]$$

- More efficient than GPH;  $\sqrt{m}(\hat{d}_{\text{LW}} - d) \xrightarrow{d} N(0, 1/4)$
- Extended versions: exact local Whittle (Shimotsu & Phillips, 2005) handles non-stationarity

## Whittle Likelihood Estimation

### The Whittle Likelihood

For a parametric spectral model  $f_\theta(\omega)$ , the Whittle log-likelihood approximation is:

$$\ell_W(\theta) = - \sum_{j=1}^{\lfloor n/2 \rfloor} \left[ \log f_\theta(\omega_j) + \frac{I(\omega_j)}{f_\theta(\omega_j)} \right], \quad \hat{\theta}_W = \arg \max_{\theta} \ell_W(\theta)$$

### Asymptotic Properties

- ▣  $\sqrt{n}(\hat{\theta}_W - \theta_0) \xrightarrow{d} N(0, \Sigma^{-1})$  where  $\Sigma = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left( \frac{\partial \log f_\theta}{\partial \theta} \right) \left( \frac{\partial \log f_\theta}{\partial \theta} \right)^\top d\omega$
- ▣ Asymptotically equivalent to time-domain Gaussian MLE

### Advantages Over Time-Domain MLE

- ▣ Computation:  $O(n \log n)$  vs.  $O(n^2)$  for exact time-domain likelihood
- ▣ Robustness: only requires correct spectral specification, not full distributional assumptions
- ▣ Modularity: straightforward for ARFIMA or models with complex spectral structure

## Long Memory: Practical Application

### Example: Financial Volatility Persistence

- ▣ Realized volatility of stock returns exhibits long memory with  $d \approx 0.35\text{--}0.45$
- ▣ The spectrum diverges near  $\omega = 0$ : extremely slow mean-reversion
- ▣ Standard ARMA models require very high order  $p$  to approximate this behavior
- ▣ ARFIMA captures the same dynamics parsimoniously with a single parameter  $d$

### Practical Workflow

1. Plot the log-periodogram: if approximately linear near origin, suspect long memory
2. Estimate  $d$  using GPH and local Whittle; compare for robustness
3. Fit ARFIMA( $p, d, q$ ); compare AIC/BIC with standard ARMA
4. Whittle likelihood for efficient parametric estimation

**Caution:** Structural breaks and deterministic trends can mimic long memory in the periodogram — always test for both (Qu, 2011; Shimotsu, 2006).

## Bivariate Spectral Analysis

### From Univariate to Bivariate

For two jointly stationary processes  $\{X_t\}$  and  $\{Y_t\}$ , the **cross-spectral density** is:

$$f_{XY}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{XY}(h) e^{-i\omega h}$$

where  $\gamma_{XY}(h) = \text{Cov}(X_{t+h}, Y_t)$  is the cross-covariance function.

### Key Difference from Univariate

- $f_{XY}(\omega)$  is **complex-valued**:  $f_{XY}(\omega) = c_{XY}(\omega) - i q_{XY}(\omega)$
- $c_{XY}(\omega)$ : **co-spectrum** (in-phase co-movement at frequency  $\omega$ )
- $q_{XY}(\omega)$ : **quadrature spectrum** (out-of-phase co-movement)
- These are harder to interpret directly  $\Rightarrow$  use derived measures below

## Squared Coherence

### Squared Coherence: Frequency-Domain $R^2$

$$\kappa_{XY}^2(\omega) = \frac{|f_{XY}(\omega)|^2}{f_X(\omega) f_Y(\omega)}, \quad 0 \leq \kappa_{XY}^2(\omega) \leq 1$$

### Interpretation

- ▣ Measures the **strength of linear association** between  $X$  and  $Y$  at frequency  $\omega$
- ▣  $\kappa^2 = 1$ : perfect linear relationship at this frequency;  $\kappa^2 = 0$ : no relationship
- ▣ Analog of  $R^2$  in regression, but **frequency-specific**
- ▣ Unlike time-domain correlation, coherence can reveal that two series are strongly related at business cycle frequencies but unrelated at higher frequencies

### Estimation

- ▣ Raw cross-periodogram gives  $\hat{\kappa}^2 = 1$  everywhere — must smooth first
- ▣ Use multitaper or kernel-smoothed cross-spectral estimates

## Phase Spectrum and Gain

### Phase Spectrum

$$\phi_{XY}(\omega) = \arctan\left(\frac{-q_{XY}(\omega)}{c_{XY}(\omega)}\right)$$

- **Lead/lag** between  $X$  and  $Y$  at frequency  $\omega$
- $\phi > 0$ :  $X$  leads  $Y$
- Time delay:  $\tau(\omega) = \phi(\omega)/\omega$
- Phase is meaningful only where coherence is significant

### Gain (Amplitude Ratio)

$$G_{XY}(\omega) = \frac{|f_{XY}(\omega)|}{f_X(\omega)}$$

- Measures how much frequency  $\omega$  in  $X$  is **amplified or attenuated** in  $Y$
- If  $Y = \beta X + \varepsilon$ , the gain is  $|\beta|$  at all frequencies
- Useful for assessing frequency-dependent transmission

**Together:** Coherence tells you *how strongly* two series are related at each frequency, phase tells you *who leads*, and gain tells you *by how much* oscillations are amplified.

## Phase: Numeric Example

### Setup

- ▣ Let  $Y_t = X_{t-2} + \varepsilon_t$  (pure lag-2 relationship)
- ▣ Cross-spectrum:  $S_{XY}(\omega) = e^{-2i\omega} S_{XX}(\omega)$
- ▣ Phase:  $\phi(\omega) = -2\omega$

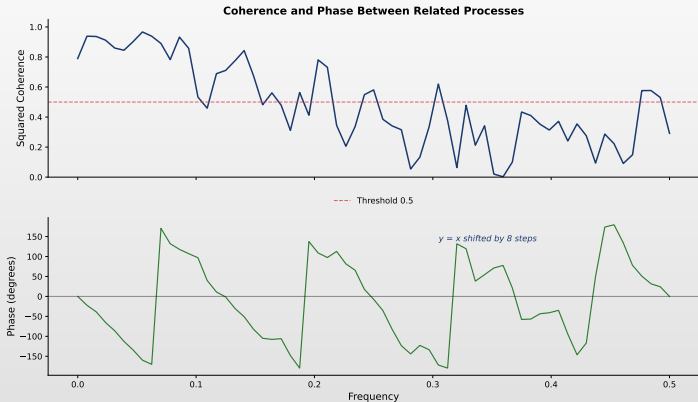
### Interpretation

- ▣ At  $\omega = \pi/4$  (8-period cycle):  $\phi = -\pi/2$  (quarter-cycle lag)
- ▣ At  $\omega = \pi/8$  (16-period cycle):  $\phi = -\pi/4$  (eighth-cycle lag)
- ▣ **Phase slope**  $= d\phi/d\omega = -2$  gives the group delay (lag in periods)

### Key insight

- ▣ Negative phase  $\Rightarrow Y$  lags  $X$
- ▣ Positive phase  $\Rightarrow Y$  leads  $X$
- ▣ Phase  $= -\omega k$  means exactly  $k$ -period lag at all frequencies

## Coherence Example



- Squared coherence, phase spectrum, and gain between two simulated economic series
- High coherence at BC frequencies (6–32 quarters) with near-zero phase  $\Rightarrow$  synchronised co-movement

## Cross-Spectral Analysis: Economic Applications

### Application 1: Business Cycle Synchronization

- Coherence between two countries' GDP at business cycle frequencies ( $\omega$  corresponding to 6–32 quarters)
- High coherence  $\Rightarrow$  synchronized business cycles (relevant for monetary union)
- Phase  $\Rightarrow$  which country leads/lags

### Application 2: Leading Indicators

- Coherence and phase between a leading indicator (e.g., yield curve slope) and GDP
- If coherence is high at business cycle frequencies and phase is positive, the indicator truly “leads”
- More informative than simple cross-correlation, which mixes all frequencies

### Application 3: Monetary Policy Transmission

- Coherence between policy rate and inflation at different frequencies
- May be high at low frequencies (long-run) but low at high frequencies (short-run noise)
- Phase reveals the **transmission lag** at each frequency

## Linear Filters and Transfer Functions I

### A Linear Filter

A linear time-invariant filter transforms  $\{X_t\}$  into  $\{Y_t\}$ :

$$Y_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j} = A(B) X_t$$

The **frequency response function** (transfer function) is:

$$A(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j}$$

### The Fundamental Result: Filtering = Multiplication

If  $X_t$  has spectral density  $f_X(\omega)$ , then  $Y_t = A(B)X_t$  has spectral density:

$$f_Y(\omega) = |A(\omega)|^2 \cdot f_X(\omega)$$

**A filter multiplies the spectrum by its squared gain** — this is why the frequency domain makes filter analysis trivial!

## Linear Filters and Transfer Functions II

### Classic Filter Examples

- ▣ **First difference**  $(1 - B)$ :  $|A(\omega)|^2 = 4 \sin^2(\omega/2)$   
 $\Rightarrow$  amplifies high frequencies, kills  $\omega = 0$  (removes trend)
- ▣ **Moving average of order  $m$** :  $|A(\omega)|^2 = \left( \frac{\sin(m\omega/2)}{m \sin(\omega/2)} \right)^2$   
 $\Rightarrow$  low-pass filter (passes slow components, kills fast oscillations)
- ▣ **Seasonal difference**  $(1 - B^{12})$ : peaks of  $|A(\omega)|^2 = 0$  at seasonal harmonics  
 $\Rightarrow$  removes monthly seasonal pattern

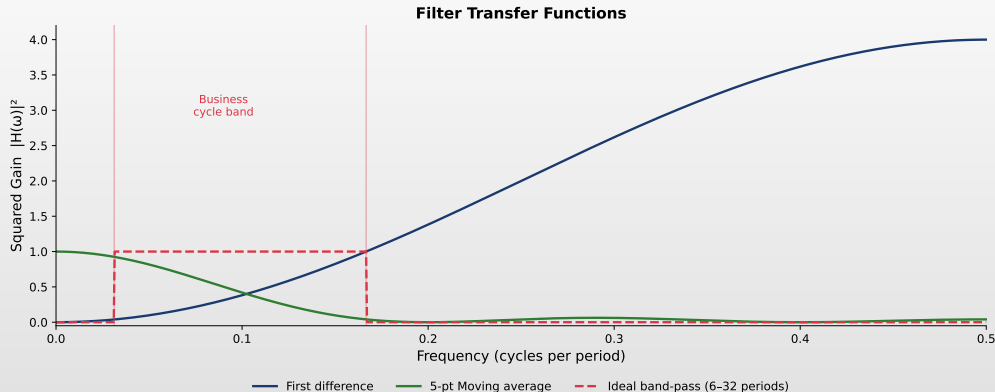
### Ideal Band-Pass Filter

Passes frequencies in  $[\omega_L, \omega_H]$  and blocks all others:

$$A(\omega) = \begin{cases} 1 & \text{if } \omega_L \leq |\omega| \leq \omega_H \\ 0 & \text{otherwise} \end{cases} \quad a_j = \frac{\sin(\omega_H j) - \sin(\omega_L j)}{\pi j}$$

**Problem:** requires infinitely many leads/lags  $\Rightarrow$  must truncate in practice.

## Filter Transfer Functions



- Squared gain  $|A(\omega)|^2$  for four filters: first difference, 12-month MA, ideal band-pass, HP cycle
- Each filter's frequency-selectivity is immediately visible in the frequency domain

## Baxter–King and Christiano–Fitzgerald Filters

### Baxter–King (BK) Filter (1999)

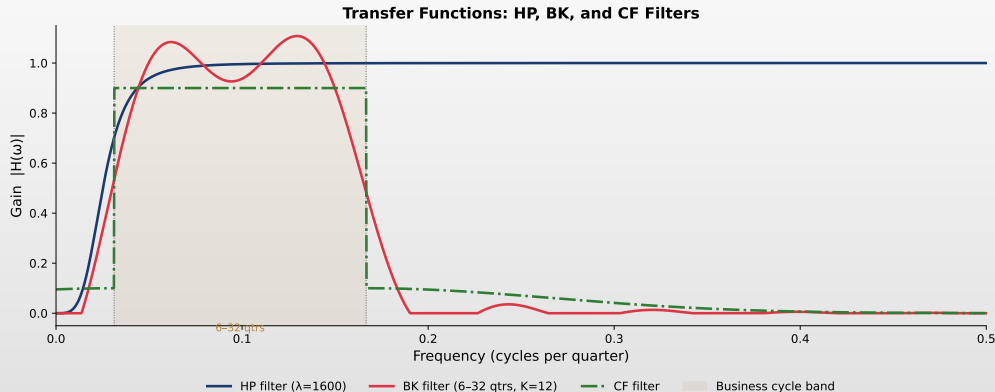
- Symmetric, fixed-length approximation to the ideal band-pass filter
- Uses  $K$  leads and lags (typically  $K = 12$  for quarterly data)
- Coefficients: truncated ideal filter + constrained to sum to zero (removes trend)
- **Drawback:** loses  $K$  observations at each end of the series

### Christiano–Fitzgerald (CF) Filter (2003)

- Asymmetric, time-varying filter — uses all observations (no data loss)
- Optimal approximation under the assumption that  $X_t$  is a random walk
- Filter weights adapt near the endpoints of the series
- Generally preferred over BK for most economic applications

**HP filter:** ( $\lambda = 1600$ ): high-pass filter with  $|A_{\text{HP}}(\omega)|^2 = \frac{4\lambda(1-\cos \omega)^2}{1+4\lambda(1-\cos \omega)^2}$ , passes cycles  $< 32$  quarters.

## HP, BK, and CF Transfer Functions



- HP cycle ( $\lambda=1600$ ), BK band-pass ( $K=12$ , 6-32q), and CF filter on quarterly data
- Ideal band-pass target (grey) shows how closely each filter approximates the ideal

## HP Filter: Critical Perspectives

- **Hamilton (2018) critique:** HP filter introduces spurious cycles not present in the data
  - ▶ The filtered series has an ARMA(4,4) structure regardless of the original DGP
  - ▶ Proposed alternative: regression-based detrending
- **Endpoint bias:** estimates near the start/end of sample are unreliable
  - ▶ HP filter is a two-sided filter; endpoints use asymmetric weights
  - ▶ Last 2–3 years of estimates should be interpreted with caution
- **$\lambda$  sensitivity:** results depend heavily on the smoothing parameter
  - ▶  $\lambda = 1600$  (quarterly) is a convention, not a universal optimum
  - ▶ BK and CF filters are less sensitive to bandwidth choice
- **Practical recommendation:** use HP as one tool among several; always compare with BK/CF

## From Fourier to Wavelets: Motivation

### Limitation of Fourier Analysis

- ▣ Fourier analysis decomposes a signal into *global* sine/cosine components
- ▣ Each frequency component extends over the **entire** time span
- ▣ Problem: **non-stationary signals** — the frequency content changes over time!
- ▣ Example: financial volatility is calm in some periods, turbulent in others

### The Uncertainty Principle (Heisenberg–Gabor)

- ▣ You **cannot** have perfect resolution in both time and frequency simultaneously
- ▣  $\Delta t \cdot \Delta f \geq \frac{1}{4\pi}$  (time-frequency uncertainty)
- ▣ Fourier:  $\Delta f$  is minimal but  $\Delta t = \infty$  (no time localization at all)
- ▣ Wavelets: adaptive — good time resolution at high  $f$ , good frequency resolution at low  $f$

### Wavelets: Localized in Both Time and Frequency

- ▣ Use **short, oscillatory** basis functions (“wavelets”) instead of infinite sinusoids
- ▣ Stretch wavelets for low frequencies, compress for high frequencies

## Continuous Wavelet Transform (CWT)

### Definition

The CWT of a signal  $x(t)$  with respect to a mother wavelet  $\psi(t)$  is:

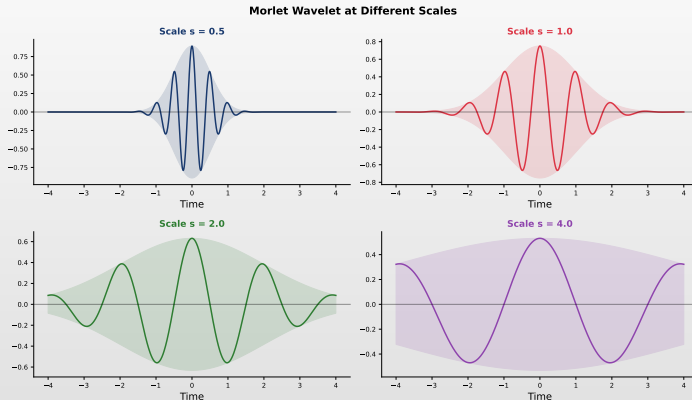
$$W(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt$$

- ▣  $a$ : **scale** parameter (inversely related to frequency: large  $a$  = low frequency)
- ▣  $b$ : **translation** parameter (time localization)
- ▣  $|W(a, b)|^2$ : **wavelet power** at scale  $a$  and time  $b$

### Common Mother Wavelets

- ▣ **Morlet**:  $\psi(t) = \pi^{-1/4} e^{i\omega_0 t} e^{-t^2/2}$  (Gaussian-modulated complex exponential,  $\omega_0 = 6$ )
- ▣ **Mexican hat**:  $\psi(t) = (1 - t^2)e^{-t^2/2}$  (second derivative of Gaussian)
- ▣ **Haar**: simplest wavelet — piecewise constant (good for jump detection)
- ▣ For time series: **Morlet is standard** — good balance of time/frequency resolution

## Morlet Wavelet Illustration



- Morlet wavelet at four scales: larger scales capture low-frequency oscillations
- Smaller scales capture high-frequency transients — adaptive time-frequency resolution

## Wavelet Power Spectrum and Scalogram

### The Wavelet Power Spectrum

$$\text{WPS}(a, b) = |W(a, b)|^2$$

Displayed as a **scalogram**: a 2D heatmap with time on the x-axis and scale (period) on the y-axis. Color intensity = power.

### Global Wavelet Spectrum

Time-averaged power at each scale:

$$\overline{W}^2(a) = \frac{1}{n} \sum_{b=1}^n |W(a, b)|^2$$

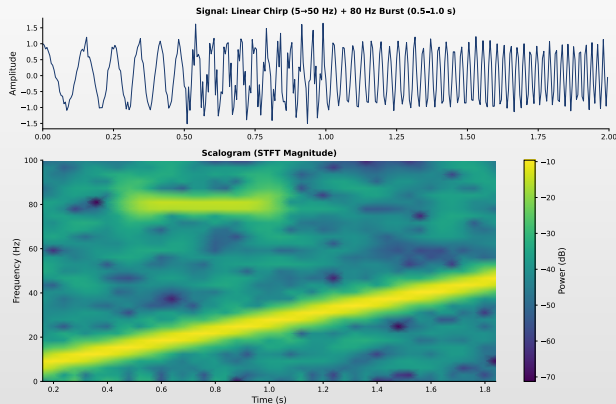
Comparable to the Fourier spectral density. Useful for identifying dominant periodicities.

### Cone of Influence (COI)

- ▣ Near edges of the time series, wavelet coefficients are unreliable (edge effects)
- ▣ The COI marks the region where edge effects are significant
- ▣ Only interpret power **inside** the COI

**Signif.:** Under  $H_0$  of red noise (AR(1)), wavelet power follows  $\chi^2_2$  scaled by the background spectrum. Contour lines at 95% significance are standard.

## Scalogram of a Synthetic Series



- Synthetic signal: 4-year cycle (first half) + 10-year cycle (second half)
- Scalogram reveals time-varying frequency content; COI (dashed) delimits reliable region

## Wavelet Coherence

### Cross-Wavelet Transform and Coherence

For two series  $X_t$  and  $Y_t$  with CWTs  $W_X(a, b)$  and  $W_Y(a, b)$ :

$$W_{XY}(a, b) = W_X(a, b) \overline{W_Y(a, b)} \quad (\text{cross-wavelet spectrum})$$

The **wavelet coherence** is:

$$R_{XY}^2(a, b) = \frac{|S(W_{XY}(a, b))|^2}{S(|W_X(a, b)|^2) \cdot S(|W_Y(a, b)|^2)}$$

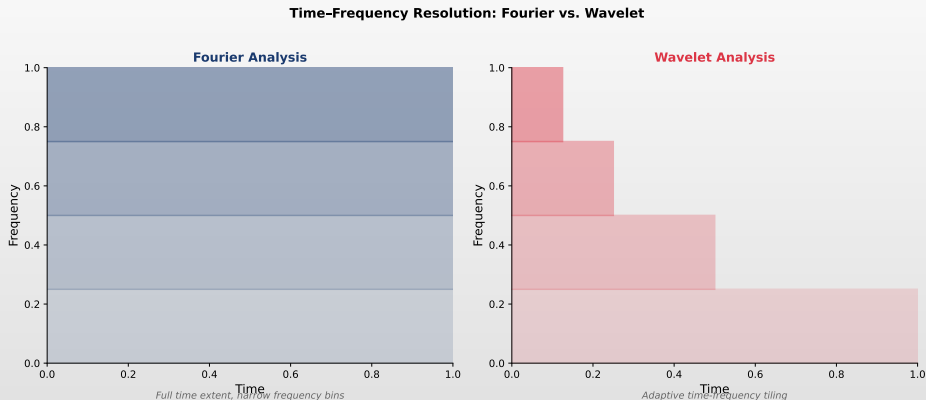
where  $S(\cdot)$  is a smoothing operator in both time and scale.

### Interpretation and Applications

- ▣  $R_{XY}^2(a, b) \in [0, 1]$ : time-varying, scale-varying measure of co-movement
- ▣ **Phase arrows** on the coherence plot indicate lead/lag relationships
- ▣ Application: oil prices and stock markets — coherence high at business cycle frequencies during crises
- ▣ Application: inflation and money supply — coherence varies across historical regimes



## Wavelet Coherence: Scalogram



- Wavelet coherence between two synthetic series with phase arrows
- Rightward = in-phase, leftward = anti-phase, upward = first series leads

## Locally Stationary Processes

### Dahlhaus (1997): A Framework for Non-Stationary Spectra

A **locally stationary process** has the representation:

$$X_{t,T} = \int_{-\pi}^{\pi} A\left(\frac{t}{T}, \omega\right) e^{i\omega t} dZ(\omega)$$

where  $A(u, \omega)$  is the **time-varying transfer function** and  $Z(\omega)$  is orthogonal-increment. The **time-varying spectral density** is:

$$f(u, \omega) = |A(u, \omega)|^2, \quad u \in [0, 1], \omega \in [-\pi, \pi]$$

### Key Idea

- At each rescaled time point  $u = t/T$ , the process “looks” stationary with spectrum  $f(u, \omega)$
- Generalizes the classical Cramér representation to allow time-varying structure
- Estimation via local periodogram or wavelet-based methods

## Spectral Change-Point Detection

### Testing for Spectral Stationarity

- $H_0: f(u, \omega) = f(\omega)$  for all  $u$  (spectrum is time-invariant)
- $H_1: f(u, \omega)$  changes at one or more time points (spectral change-points)

### Approaches

- **CUSUM on local periodograms:** Compare spectral estimates from sub-segments
- **Wavelet-based tests:** Use scale-by-scale variance to detect changes at specific frequency bands
- **Kolmogorov–Smirnov-type statistics:** Compare spectral distributions across windows

### Application: Regime Detection in Economics

- Detecting the onset and end of the Great Moderation via spectral changes
- Identifying structural breaks in inflation dynamics or financial volatility
- Changes in business cycle frequency or amplitude across policy regimes

## STFT vs. CWT: Bridging Fourier and Wavelets

### Short-Time Fourier Transform (STFT)

- Apply the DFT to windowed segments:  $\text{STFT}(\tau, \omega) = \sum_t x(t) w(t - \tau) e^{-i\omega t}$ , where  $w(\cdot)$  is a fixed-width window centered at time  $\tau$

|             | STFT                                    | CWT                                   |
|-------------|-----------------------------------------|---------------------------------------|
| Window      | Fixed width for all $\omega$            | Adaptive (narrow at high $\omega$ )   |
| Resolution  | Uniform $\Delta t \times \Delta \omega$ | Variable: $\Delta t \propto 1/\omega$ |
| Basis       | Windowed sinusoids                      | Scaled/shifted wavelets               |
| Best for    | Signals with stable frequency           | Signals with transient features       |
| Spectrogram | $ \text{STFT} ^2 = \text{spectrogram}$  | $ W(a, b) ^2 = \text{scalogram}$      |

### Practical Insight

- The STFT adds time localization to Fourier analysis; the CWT further adapts window width to frequency
- For economic data: CWT preferred because business cycle and high-frequency dynamics require different resolutions

## Fourier vs. Wavelets: When to Use Which

### Decision Guide

| Criterion            | Fourier / Spectral                 | Wavelet                          |
|----------------------|------------------------------------|----------------------------------|
| Stationarity         | Required (or assumed)              | Not required                     |
| Time localization    | No                                 | Yes                              |
| Frequency resolution | Uniform across all $\omega$        | Varies with scale                |
| Best for             | Stationary series, filter design   | Non-stationary, transient events |
| Hypothesis testing   | Well-developed (F-test, $\chi^2$ ) | Monte Carlo, AR(1) null          |
| Computational cost   | $O(n \log n)$                      | $O(n \log n)$ to $O(n^2)$        |
| Common in            | Economics, signal processing       | Geophysics, finance, climate     |

### Practical Recommendation

- ▣ Use **Fourier/spectral analysis** when you believe the frequency structure is *stable* over time
- ▣ Use **wavelets** when you suspect the frequency structure *changes* over time (regime shifts, evolving cycles)
- ▣ Often use **both**: Fourier for overall spectral shape, wavelets for time-varying behavior

## Section Summary: Wavelets vs. Fourier

### When to Use Which

- ▣ **Fourier**: stationary series, global frequency content, sharp frequency resolution
- ▣ **Wavelets**: non-stationary, time-varying frequencies, localization in time AND frequency
- ▣ **STFT**: compromise — fixed window; uniform time-frequency resolution
- ▣ **Scalogram**: preferred visualization for economic and financial cycles

## Practical Workflow: Spectral Analysis of a Real Series

1. **Preprocess:** remove trend (HP/first-difference), take logs if needed
2. **Stationarity check:** ADF/KPSS test; difference if needed
3. **Compute periodogram:** identify dominant frequencies visually
4. **Smooth:** apply Welch or multitaper for reliable spectral density estimate
5. **Test:** Fisher's test for significant hidden periodicities
6. **Long memory?:** estimate  $d$  via GPH or Local Whittle if  $\hat{f}(\omega) \rightarrow \infty$  near  $\omega = 0$
7. **Cross-spectral:** coherence + phase if analyzing two series
8. **Wavelet:** if non-stationarity suspected, use CWT scalogram

## Case Study Overview

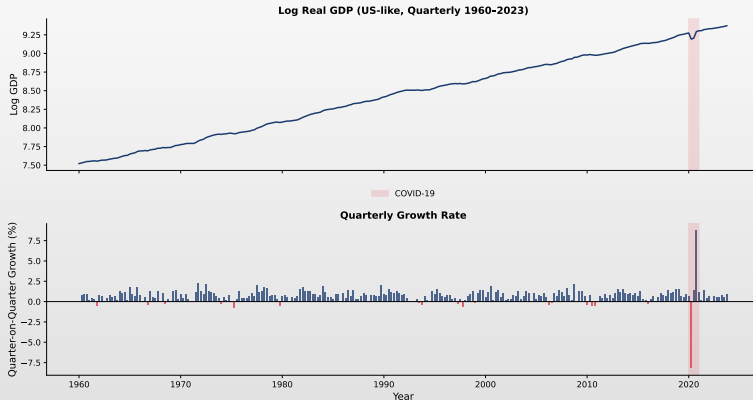
### Data Description

- ▣ **Data:** US Real GDP, quarterly, 1947Q1–2024Q4 (FRED series GDPC1)
- ▣ **Transformation:** log of real GDP
- ▣ **Task:** Extract the business cycle component using spectral methods
- ▣ **Methods compared:** Periodogram, Multitaper, HP filter, BK filter, CF filter, Wavelet

### Business Cycle Definition

- ▣ Burns & Mitchell (1946) / NBER: recurrent fluctuations with period 6–32 quarters
- ▣ Frequency band:  $\omega \in [2\pi/32, 2\pi/6] \approx [0.196, 1.047]$  radians/quarter
- ▣ Goal: isolate this band from trend (low  $\omega$ ) and irregular (high  $\omega$ ) components

## US GDP Time Series



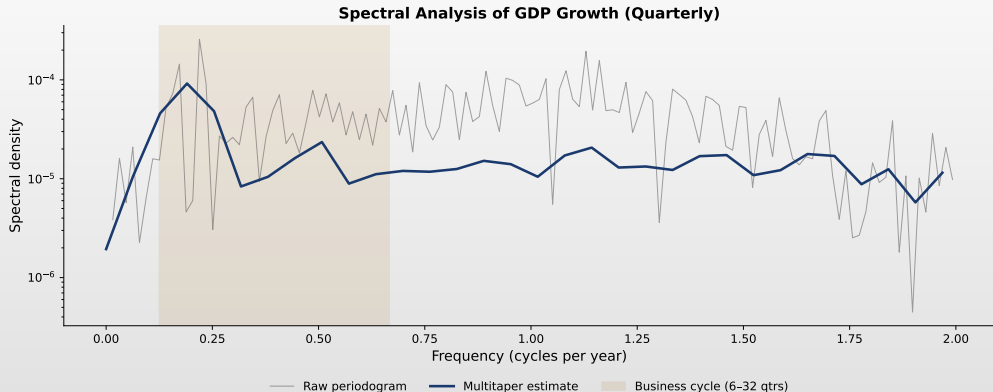
- Log US Real GDP (FRED GDPC1), 1947Q1–2024Q4, with quarterly growth rate
- Strong trend dominates time-domain view; spectral methods needed for cycle extraction

## Python: Periodogram and Multitaper Estimation

### Spectral Analysis with Python

```
import numpy as np; import pandas as pd
from scipy import signal; from spectrum import pmtm
# Load US Real GDP (quarterly, log-transformed)
gdp = pd.read_csv('GDPC1.csv', parse_dates=['DATE'], index_col='DATE')
y = np.log(gdp['GDPC1'].values)
y_dt = y - np.polyval(np.polyfit(np.arange(len(y)), y, 1), np.arange(len(y))
)
# Raw periodogram (fs=4 = quarterly)
freqs_p, Pxx_p = signal.periodogram(y_dt, fs=4)
# Multitaper spectral estimate (NW=4, K=7 tapers)
Pxx_mt, w, ev = pmtm(y_dt, NW=4, NFFT=1024, method='adapt')
freqs_mt = np.linspace(0, 2, len(Pxx_mt))
bc_low, bc_high = 1/8, 1/1.5 # business cycle band (cycles/year)
```

## GDP Periodogram and Multitaper Estimate



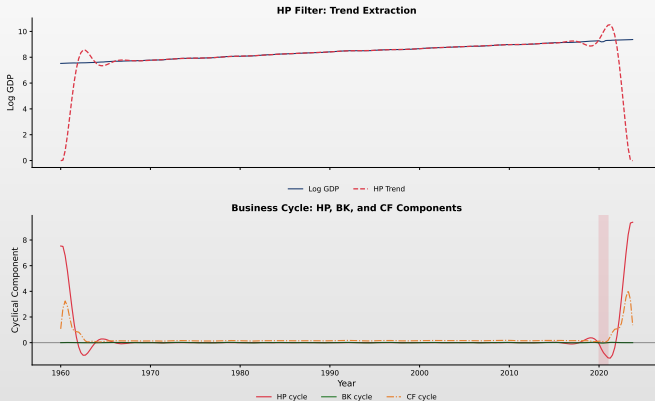
- ▣ Raw periodogram (grey) and multitaper estimate (blue) of detrended log GDP
- ▣ Shaded band = business cycle range (6-32q); dominant peak at  $\approx 16-20$  quarters

## Python: Band-Pass Filtering

### Extracting the Business Cycle Component

```
import statsmodels.api as sm
cycle_hp, trend_hp = sm.tsa.filters.hpfilter(y, lamb=1600) # HP filter
cycle_bk = sm.tsa.filters.bkfilter(y, low=6, high=32, K=12) # BK filter
cycle_cf, trend_cf = sm.tsa.filters.cffilter(y, low=6, high=32) # CF filter
# Compare extracted cycles
import matplotlib.pyplot as plt
fig, ax = plt.subplots(figsize=(12, 5))
ax.plot(gdp.index, cycle_hp, label='HP filter', alpha=0.8)
ax.plot(gdp.index[12:-12], cycle_bk, label='BK filter', alpha=0.8)
ax.plot(gdp.index, cycle_cf, label='CF filter', alpha=0.8)
ax.axhline(0, color='black', linewidth=0.5)
ax.legend(); ax.set_title('US Business Cycle Components')
```

## Band-Pass Filtered Business Cycles



- HP ( $\lambda=1600$ ), BK ( $K=12$ ), and CF business cycle components with NBER recession shading
- All three filters agree on timing of peaks and troughs; CF preserves endpoints

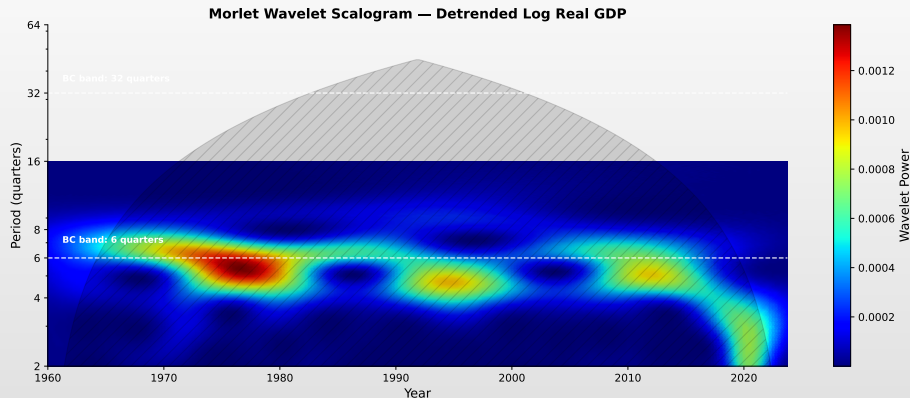
## Python: Wavelet Analysis

### Wavelet Power Spectrum and Scalogram

```
import pywt # PyWavelets
scales = np.arange(2, 128) # scales ~ periods in quarters
coefficients, frequencies = pywt.cwt(y_dt, scales, 'morl',
                                     sampling_period=0.25)

power = np.abs(coefficients) ** 2
periods = 1 / frequencies # convert to years
# Plot scalogram
fig, ax = plt.subplots(figsize=(12, 6))
im = ax.contourf(gdp.index, periods, power, levels=20, cmap='jet')
ax.set_ylim(1, 32); ax.set_yscale('log', base=2)
ax.set_ylabel('Period (years)')
ax.axhline(y=1.5, color='white', linestyle='--')
ax.axhline(y=8, color='white', linestyle='--')
plt.colorbar(im, label='Power')
```

## Wavelet Scalogram: US Real GDP



- Morlet CWT scalogram of HP-detrended log GDP; white dashed lines = 6–32q BC band
- “Great Moderation” (1985–2007) = reduced power; 2008 & 2020 = intense localised bursts

## Case Study Results: Spectral Density of US GDP

### Key Findings from Spectral Analysis

- ▣ **Dominant peak** in the periodogram at period  $\approx 16$ –20 quarters (4–5 years)
- ▣ Consistent with the average duration of NBER business cycles
- ▣ Secondary power concentration at period  $\approx 32$ –40 quarters (8–10 years) — Juglar cycle
- ▣ Very high power at lowest frequencies confirms strong trend component
- ▣ Multitaper estimate is smoother, confirms the same peak structure

### Filter Comparison

- ▣ HP, BK, and CF filters produce **similar** business cycle components (correlation  $> 0.85$ )
- ▣ BK loses 12 observations at each end; CF and HP preserve the full sample
- ▣ All three correctly identify NBER recessions as troughs in the cycle component
- ▣ CF filter slightly better at endpoints due to asymmetric adaptation
- ▣ **Confidence intervals:**  $\hat{f}(\omega) \cdot \chi^2_\nu / \nu$  with  $\nu$  effective d.o.f. provides pointwise CI for each spectral estimate

## Case Study Results: Wavelet Scalogram

### Time-Varying Spectral Structure

- The wavelet scalogram reveals that business cycle power is **not constant** over time
- **1970s–1980s**: Strong power at 4–8 year periods (oil shocks, Volcker disinflation)
- **1985–2007**: Reduced power at all business cycle frequencies — the “Great Moderation”
- **2008–2009**: Intense power concentration at 4–6 year periods (Great Recession)
- **2020**: Sharp, localized burst of power at 2–4 year periods (COVID shock)

### Insight: Why Wavelets Complement Fourier

- Fourier analysis tells us the *average* cycle is  $\approx 5$  years
- Wavelets tell us the cycle duration and intensity *change over time*
- The Great Moderation is visible as a “quiet zone” in the scalogram
- This cannot be detected with standard Fourier spectral analysis

## AI and Spectral Methods: Modern Intersections

### Neural Spectral Estimation

- ▣ Deep learning models can learn spectral representations directly from data
- ▣ Autoencoder architectures trained to reconstruct via frequency-domain loss functions
- ▣ Neural network-based parametric spectral estimators can adapt to non-standard spectral shapes

### Spectral Regularization in Forecasting

- ▣ Modern deep forecasting models (N-BEATS, TimesNet) use spectral decomposition layers
- ▣ **Fourier features** as input:  $\sin(2\pi kt/T)$ ,  $\cos(2\pi kt/T)$  capture known periodicities
- ▣ **Spectral loss**: penalize discrepancy between predicted and observed periodograms
- ▣ Helps prevent neural networks from ignoring important cyclical structure

### Classical vs. Learned: What to Verify

- ▣ Does the AI-generated spectrum respect known theoretical properties (non-negativity, symmetry)?
- ▣ Are the dominant frequencies consistent with classical periodogram/multitaper estimates?
- ▣ Does the model correctly handle the sampling frequency and Nyquist limit?

## Key Takeaways

### What We Learned

- ▣ **Spectral density**  $f(\omega)$  decomposes variance by frequency (Wiener–Khinchine:  $\gamma(h) \leftrightarrow f(\omega)$ )
- ▣ The **periodogram** is the natural estimator but *inconsistent* — must be smoothed (multitaper = gold standard)
- ▣ **Cross-spectral** measures (coherence, phase, gain) reveal frequency-specific relationships
- ▣ **Band-pass filters** (BK, CF) extract business cycles; HP is a high-pass filter
- ▣ **Wavelets** provide time-frequency localization for non-stationary signals

### Practical Recommendations

- ▣ Always detrend before spectral analysis; use log scale for plotting
- ▣ Choose multitaper for final estimates; raw periodogram only for exploration
- ▣ For non-stationary data, prefer wavelet analysis over Fourier methods

## Question 1

### Question

- What does the spectral density function  $f(\omega)$  measure, and what is its relationship to the autocovariance function  $\gamma(h)$ ?

### Answer Choices

- (A)**  $f(\omega)$  measures the mean of the process at frequency  $\omega$ ; it is the Laplace transform of  $\gamma(h)$
- (B)**  $f(\omega)$  measures the contribution of frequency  $\omega$  to total variance; it is the Fourier transform of  $\gamma(h)$  (Wiener–Khinchine theorem)
- (C)**  $f(\omega)$  measures the phase shift at frequency  $\omega$ ; it is the inverse of  $\gamma(h)$
- (D)**  $f(\omega)$  measures the trend component at frequency  $\omega$ ; it is the derivative of  $\gamma(h)$

## Question 1: Answer

Answer: (B)

- The spectral density function  $f(\omega) = \frac{1}{2\pi} \sum_h \gamma(h) e^{-i\omega h}$  decomposes the total variance of a stationary process across frequencies. The total variance equals  $\gamma(0) = \int_{-\pi}^{\pi} f(\omega) d\omega$ . This Fourier transform pair between  $\gamma(h)$  and  $f(\omega)$  is the Wiener–Khinchine theorem — the fundamental result of spectral analysis.

 TSA\_ch12\_quiz1

## Question 2

### Question

- Why is the raw periodogram an *inconsistent* estimator of the spectral density, and how can this be fixed?

### Answer Choices

- **(A)** The periodogram is biased; fix it by increasing the sample size
- **(B)** The periodogram's variance does not decrease with sample size ( $\text{Var}[I(\omega)] \approx f(\omega)^2$  for all  $n$ ); fix it by smoothing (e.g., multitaper, Welch, Daniell kernel)
- **(C)** The periodogram only works for white noise; fix it by differencing the data
- **(D)** The periodogram is inconsistent because it uses the FFT; fix it by computing the DFT directly

## Question 2: Answer

Answer: (B)

- The periodogram is approximately unbiased ( $\mathbb{E}[I(\omega)] \rightarrow f(\omega)$ ), but its variance  $\text{Var}[I(\omega)] \approx f(\omega)^2$  does not decrease as  $n \rightarrow \infty$ . This is because adding more data adds more Fourier frequencies rather than improving estimates at existing frequencies. Smoothing (averaging nearby periodogram values, multitaper, or lag-window methods) trades frequency resolution for reduced variance, yielding a consistent estimator.

 TSA\_ch12\_quiz2

### Question 3

#### Question

- ☐ You observe a spectral density with a sharp peak near  $\omega = 0$  and smaller peaks at  $\omega = \pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6, \pi$ . What type of process likely generated this data?

#### Answer Choices

- ☐ (A) White noise — flat spectrum with random fluctuations
- ☐ (B) AR(1) with  $\phi > 0$  — red spectrum with peak at zero
- ☐ (C) Monthly seasonal process with trend — peak at zero (trend) plus peaks at seasonal harmonics  $\omega_k = 2\pi k/12$
- ☐ (D) Random walk — all power at  $\omega = 0$  only

### Question 3: Answer

Answer: (C)

- The peaks at  $\omega = \pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6, \pi$  correspond exactly to  $2\pi k/12$  for  $k = 1, 2, \dots, 6$  — the seasonal harmonics of monthly data ( $s = 12$ ). The peak at  $\omega \approx 0$  indicates a strong trend or long-run component. This is the classic spectral signature of a monthly series with both trend and seasonality, such as airline passengers or retail sales.

 TSA\_ch12\_quiz3

## Question 4

### Question

- ☐ What is the main advantage of wavelet analysis over Fourier spectral analysis for economic time series?

### Answer Choices

- ☐ (A) Wavelets provide higher frequency resolution than Fourier for stationary series
- ☐ (B) Wavelets can localize spectral features in both time and frequency, making them suitable for non-stationary processes where the frequency content changes over time
- ☐ (C) Wavelets are computationally faster than the FFT
- ☐ (D) Wavelets do not require choosing any parameters, unlike Fourier methods

## Question 4: Answer

Answer: (B)

- Wavelet analysis provides simultaneous time-frequency localization. Unlike Fourier analysis, which produces a single global spectrum, wavelets create a scalogram showing how spectral power evolves over time. This is crucial for economic series subject to structural breaks, regime changes, and evolving business cycle dynamics. The Heisenberg uncertainty principle limits the joint resolution, but wavelets adaptively balance time and frequency resolution across scales.

 TSA\_ch12\_quiz4

## Question 5

### Question

- A researcher applies the Hodrick–Prescott (HP) filter with  $\lambda = 1600$  to quarterly GDP data. What is the spectral interpretation of this filter?

### Answer Choices

- (A) It is a band-pass filter that extracts cycles with period 6–32 quarters
- (B) It is a low-pass filter that extracts the trend by keeping only low frequencies
- (C) The HP cycle component is a high-pass filter:  $|A(\omega)|^2 = \frac{4\lambda(1-\cos \omega)^2}{1+4\lambda(1-\cos \omega)^2}$ , passing all cycles shorter than  $\approx 32$  quarters including irregular noise
- (D) It is a notch filter that removes only the seasonal frequency

## Question 5: Answer

Answer: (C)

- The HP filter's cycle component is a high-pass filter that passes all frequencies above a cutoff determined by  $\lambda$ . With  $\lambda = 1600$  and quarterly data, it passes cycles shorter than  $\approx 32$  quarters. Unlike a band-pass filter (BK, CF), the HP cycle includes both business cycle *and* high-frequency irregular components. This is a well-known limitation: the HP "cycle" is contaminated by noise, which is why dedicated band-pass filters are preferred for business cycle extraction.

 TSA\_ch12\_quiz5

## Bibliography I

### Foundational References

- ▣ Priestley, M.B. (1981). *Spectral Analysis and Time Series*. Academic Press.
- ▣ Brockwell, P.J. & Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer.
- ▣ Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press, Chapters 6, 10.
- ▣ Percival, D.B. & Walden, A.T. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press.

### Spectral Estimation

- ▣ Thomson, D.J. (1982). Spectrum Estimation and Harmonic Analysis, *Proceedings of the IEEE*, 70(9), 1055–1096.
- ▣ Welch, P.D. (1967). The Use of Fast Fourier Transform for the Estimation of Power Spectra, *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70–73.

## Bibliography II

### Band-Pass Filters and Business Cycles

- Baxter, M. & King, R.G. (1999). Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series, *Review of Economics and Statistics*, 81(4), 575–593.
- Christiano, L.J. & Fitzgerald, T.J. (2003). The Band Pass Filter, *International Economic Review*, 44(2), 435–465.
- Hodrick, R.J. & Prescott, E.C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit and Banking*, 29(1), 1–16.
- Burns, A.F. & Mitchell, W.C. (1946). *Measuring Business Cycles*. NBER.

### Wavelet Analysis

- Torrence & Compo (1998). A Practical Guide to Wavelet Analysis, *Bull. Amer. Meteorol. Soc.*, 79(1).
- Grinsted, Moore & Jevrejeva (2004). Cross Wavelet Transform and Wavelet Coherence, *Nonlin. Proc. Geophys.*, 11.
- Rua & Nunes (2009). Stock Market Returns: A Wavelet Analysis, *J. Empir. Finance*, 16(4).

## Bibliography III

### Long Memory and Spectral Regression

- Geweke & Porter-Hudak (1983). Estimation of Long Memory Time Series Models. *JTSAM*, 4(4), 221–238.
- Robinson (1995). Gaussian Semiparametric Estimation of Long Range Dependence. *Ann. Stat.*, 23(5).
- Shimotsu & Phillips (2005). Exact Local Whittle Estimation of Fractional Integration. *Ann. Stat.*, 33(4).
- Dahlhaus (1997). Fitting Time Series Models to Nonstationary Processes. *Ann. Stat.*, 25(1).

### Additional References

- Cooley & Tukey (1965). Machine Calculation of Complex Fourier Series. *Math. Comput.*, 19(90).
- Fisher (1929). Tests of Significance in Harmonic Analysis. *Proc. R. Soc. A*, 125(796).
- Granger & Hatanaka (1964). *Spectral Analysis of Economic Time Series*. Princeton UP.
- Stock & Watson (1999). Business Cycle Fluctuations in U.S. Macroeconomic Time Series. *Handbook Macro*. Vol. 1.

## Bibliography IV

### Online Resources and Code

- ▣ **Quantlet:** <https://quantlet.com> – Code platform for quantitative methods
- ▣ **Quantinar:** <https://quantinar.com> – Learning platform for quantitative methods
- ▣ **GitHub TSA:** [https://github.com/danpele/TSA/tree/main/TSA\\_ch12](https://github.com/danpele/TSA/tree/main/TSA_ch12) – Python code for this chapter

# Thank You!

## Questions?

- Course materials: <https://danpele.github.io/Time-Series-Analysis/>
-  Quantlet       Quantinar