Lecture 12 — Nonparametric Applications and Spectral Analysis

Chapter 5 → Chapter 6: R workflow, nonlinear predictability, and the spectrum of ARMA processes

Jiajing Sun

School of Economics and Management, University of Chinese Academy of Sciences

Econometrics and Time Series Methods
Spring 2026

Why Lecture 12 is a natural continuation

Lecture 11 introduced the *theory* of kernel density estimation and nonparametric regression. Lecture 12 pushes those tools into three directions:

Direction 1: implementation

We move from definitions to an empirical workflow in R: density estimation, local regression, and simple nonlinear-predictability checks.

Direction 2: financial applications

We use nonparametric ideas to study market efficiency, runs, nonlinear dependence, nonparametric autoregression, and semiparametric volatility modelling.

Direction 3: transition to Chapter 6

We switch from *local smoothing in the time domain* to *variance decomposition in the frequency domain*. The bridge is the language of kernels, Fourier transforms, and second-order structure.

Learning goals

By the end of the lecture, students should be able to:

1. implement kernel density estimation and local regression in R and explain what the bandwidth controls;
2. distinguish linear unpredictability from broader nonlinear predictability;
3. explain how runs tests relate to weak-form efficiency and randomness of return signs;
4. define a nonparametric autoregression and state how it can be used to test martingale-difference restrictions;
5. write down the semiparametric GARCH likelihood with unknown innovation density;
6. define the population spectrum as the Fourier transform of the autocovariance function;
7. interpret white-noise, MA, AR, and ARMA spectra in economic and time-series terms.

Three-hour plan

Hour 1

R-oriented review of kernel density estimation, local regression, and nonlinear predictability.

Hour 2

Financial applications of nonparametrics: runs tests, nonparametric autoregression, and semiparametric volatility.

Hour 3

Spectral representation, the population spectrum, and the spectra of ARMA processes.

Main theme

The lecture asks how we detect structure when linear tools are too restrictive, and how the same second-order information can be read either through *lags* or through *frequencies*.

From nonparametric formulas to empirical workflow

The nonparametric estimators of Chapter 5 are conceptually simple but empirically delicate.

- We estimate an unknown object locally: a density $g(x)$ or a regression function $m(x)$.
- We must choose a kernel $K(\cdot)$ and, more importantly, a bandwidth h.
- We then inspect *shape*: skewness, heavy tails, multimodality, nonlinear mean reversion, asymmetric volatility, or local departures from a parametric benchmark.

The empirical logic

unknown object $\rightarrow$ local averaging $\rightarrow$ bandwidth choice $\rightarrow$ economic interpretation.

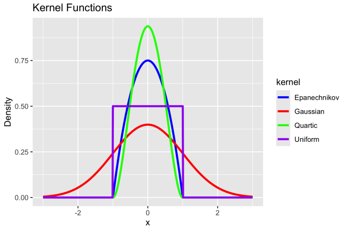
For financial time series, this often means asking whether returns are Gaussian, whether conditional means are nonlinear, and whether volatility reacts to lagged returns in ways not captured by a simple GARCH recursion.

Kernel density estimation: recall and implementation logic

For data $X_1, \ldots, X_T$, the kernel density estimator is

$$\widehat{g}(x) = \frac{1}{Th} \sum_{t=1}^{T} K\left(\frac{x - X_t}{h}\right).$$

- $K(\cdot)$ determines the shape of local weights.
- h determines the width of the neighborhood.
- Small h: low bias, high variance, wiggly estimate.
- Large h: high bias, low variance, oversmoothed estimate.

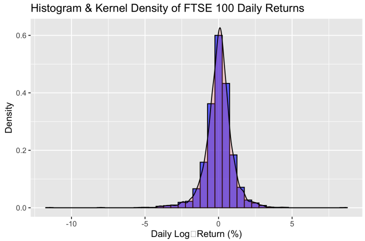


Key implementation fact

In practice, the choice of *bandwidth* matters much more than the choice of kernel family.

Financial returns are rarely Gaussian

- Return distributions are often skewed, heavy-tailed, and peaked around zero.
- A histogram plus KDE is a quick diagnostic for non-normality.
- In finance, the shape itself matters because tail risk, crash probability, and downside asymmetry are first-order issues.



Econometric message

A flexible density estimate is often the first sign that a Gaussian likelihood may be misspecified.

Local regression: Nadaraya–Watson and local polynomial fits

For a scalar regressor X_t and response Y_t, the Nadaraya–Watson estimator is

$$\widehat{m}(x) = \frac{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right) Y_t}{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right)}.$$

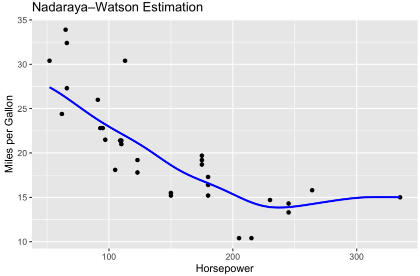
This is a *local weighted average*. The local polynomial estimator generalizes this by fitting a polynomial around x via weighted least squares:

$$(\widehat{\beta}_0, \ldots, \widehat{\beta}_p) = \arg\min_{\beta_0,\ldots,\beta_p} \sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right) \left[Y_t - \sum_{j=0}^{p} \beta_j (X_t - x)^j\right]^2.$$

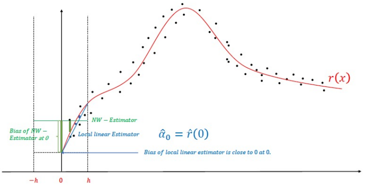
Interpretation

- $p = 0$ gives the local-constant/Nadaraya–Watson estimator.
- $p = 1$ gives the local-linear estimator.
- Higher-order local polynomials reduce approximation bias when the true regression function bends locally.

Why local linear smoothing improves the boundary



The local estimate is driven by nearby observations with kernel weights.



Local linear smoothing corrects part of the boundary bias of a local-constant fit.

- Near the boundary, a symmetric kernel effectively becomes one-sided.
- A local constant fit then inherits a systematic edge bias.
- A local linear fit compensates by allowing the local mean to tilt.

R workflow: density estimation and local regression

```
# Daily return density and local regression workflow
library(quantmod)
library(KernSmooth)
library(ggplot2)

getSymbols("^FTSE", src = "yahoo", from = "2015-01-01")
r <- na.omit(100 * dailyReturn(Cl(FTSE), type = "log"))

# 1. Kernel density estimate
kd <- density(as.numeric(r), kernel = "gaussian")

# 2. Local regression example on lagged returns
lagr <- stats::lag(r, -1)
df <- na.omit(data.frame(y = as.numeric(r), x = as.numeric(lagr)))

bw <- dpill(df$x, df$y)
fit_nw <- locpoly(df$x, df$y, bandwidth = bw, degree = 0)
fit_ll <- locpoly(df$x, df$y, bandwidth = bw, degree = 1)
```

What this code is doing

First estimate the unconditional shape of returns; then estimate a local conditional mean of r_t given r_{t-1} and compare local-constant and local-linear fits.

Linear unpredictability is not the same as nonlinear unpredictability

A time series can have zero linear autocorrelation and still be predictable in nonlinear ways.

Classical pitfall

If $\mathrm{Corr}(Y_t, Y_{t-1}) = 0$, this does *not* imply that Y_t is independent of Y_{t-1}, nor that $\mathbb{E}(Y_t \mid Y_{t-1})$ is constant.

Consider a transformation $g_{t-1} = g(Y_{t-1}, Y_{t-2}, \ldots)$ such as

$$g_{t-1} = Y_{t-1}^2, \qquad \mathrm{sign}(Y_{t-1}), \qquad \sum_{j=1}^{p} Y_{t-j}^2.$$

Then one can examine

$$\gamma(g) = \mathrm{Cov}(Y_t, g_{t-1}).$$

Key idea: if Y_t is uncorrelated with the transformed lag information, then this particular form of nonlinear predictability is absent. If not, there is predictive content beyond the usual linear AR view.

A simple transformed-predictor test

Estimate

$$\widehat{\gamma}(g) = \frac{1}{T}\sum_{t=2}^{T}(Y_t - \bar{Y})(g_{t-1} - \bar{g}), \qquad \bar{g} = \frac{1}{T}\sum_{t=1}^{T} g_t.$$

Under suitable regularity conditions and under the null $\gamma(g) = 0$,

$$\sqrt{T}(\widehat{\gamma}(g) - \gamma(g)) \Longrightarrow N(0, V(g)),$$

with

$$V(g) = E\big[(Y_t - EY)^2(g_{t-1} - Eg_{t-1})^2\big].$$

The studentized statistic is

$$\widehat{S}(g) = \frac{\sqrt{T}(\widehat{\gamma}(g) - \gamma(g))}{\sqrt{T^{-1}\sum_{t=1}^{T}(Y_t - \bar{Y})^2(g_{t-1} - \bar{g})^2}} \Longrightarrow N(0, 1).$$

Limitation

This is only as rich as the chosen transformation g. A single function cannot capture the whole space of nonlinear alternatives.

R workflow: a simple nonlinear-predictability check

```
# Example: test whether squared lagged returns predict returns
rnum <- as.numeric(r)
y <- rnum[-1]
g <- rnum[-length(rnum)]^2

Sg <- sqrt(length(y)) * cov(y, g) /
     sqrt(mean((y - mean(y))^2 * (g - mean(g))^2))

p_value <- 2 * (1 - pnorm(abs(Sg)))

# A richer approach: compare parametric AR with local regression
ar_fit <- lm(y ~ g)
# or use np::npreg for a nonparametric conditional mean
```

Interpretation

This is not a universal test of efficiency. It is a diagnostic that asks whether one specific nonlinear signal helps forecast the next return.

Applications map for the second hour

The next block uses nonparametric ideas in three financial directions.

1. **Runs tests**: randomness of signs and weak-form efficiency.
2. **Nonparametric autoregression**: flexible conditional means and tests of martingale-difference restrictions.
3. **Semiparametric volatility**: parametric dynamics for σ_t^2 but nonparametric shape for the innovation density, or fully flexible variance functions.

Unifying idea

Each method relaxes a parametric restriction: Gaussianity, linearity, or a rigid variance recursion.

Weak-form efficiency and distribution-free evidence

The Efficient Market Hypothesis asks whether past information can be used to forecast future price changes. In weak form, the question is whether lagged price or return information generates systematic predictive power.

Why nonparametric methods are attractive here

Parametric tests usually focus on *linear* predictability. But markets may violate efficiency through nonlinear dependence, sign asymmetries, or state-dependent mean reversion.

- Runs tests use only the ordering of signs.
- Nonparametric autoregression estimates the entire conditional mean function.
- Semiparametric volatility relaxes restrictive assumptions on the shock density or variance function.

Runs test: the basic intuition

A *run* is a consecutive sequence of identical outcomes, for example a string of positive returns or a string of negative returns.

Null idea

If returns are generated by a random sign process, then the number and the lengths of positive and negative runs should look like what an IID sign sequence would generate.

- Too few runs suggests persistence or trend-following in signs.
- Too many runs suggests rapid alternation or negative dependence.
- Extremely long runs may be evidence against randomness of sign changes.

This is conceptually close to the coin-toss problem, but the application is economic: do gains and losses arrive in a way consistent with unpredictability?

Run-length construction and asymptotic benchmark

Define the signed run length recursively by setting $Z_0 = 0$ and, for each t,

$$Z_t = \begin{cases} Z_{t-1} + \operatorname{sign}(X_t), & \text{if } \operatorname{sign}(X_t) = \operatorname{sign}(X_{t-1}), \\ 0, & \text{if } \operatorname{sign}(X_t) = -\operatorname{sign}(X_{t-1}). \end{cases}$$

Then Z_t records the signed length of the current run.
If X_t is IID with median zero, then $\operatorname{sign}(X_t)$ is IID Bernoulli on $\{\pm 1\}$. For long samples, the maximum run length obeys the approximation

$$Z_n = \left\lfloor \frac{W}{\log 2} + \frac{\log n}{\log 2} - 1 \right\rfloor, \qquad \mathbb{P}(W \le t) = \exp[-\exp(-t)].$$

Interpretation

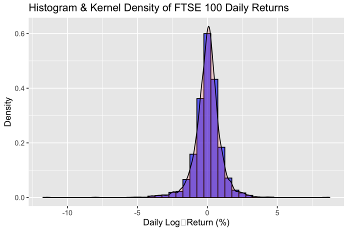
The longest run grows only logarithmically with sample size, so extremely long sign streaks are unusual under randomness.

What runs tests can and cannot detect

- They are robust because they use only signs, not distributional details.
- They are useful when outliers and heavy tails make moment-based tests unstable.
- They speak directly to the question of directional persistence.

But they are limited

A runs test ignores magnitudes. A market may have random signs but predictable volatility or predictable tail risk. So passing a runs test is not the same as full market efficiency.



Sign-based procedures are robust to heavy-tailed return magnitudes.

Complete Granger causality and predictability in distribution

Linear Granger causality asks whether lagged X_t helps predict the conditional *mean* of Y_t within a VAR. A more general notion is predictability in distribution.

Definition

Let $\mathcal{F}_{t-1}^X = \{X_{t-1}, X_{t-2}, \ldots\}$ and $\mathcal{F}_{t-1}^Y = \{Y_{t-1}, Y_{t-2}, \ldots\}$. We say that X_t Granger-causes Y_t in distribution if

$$P(Y_t \leq y \mid \mathcal{F}_{t-1}) \neq P(Y_t \leq y \mid \mathcal{F}_{t-1}^Y)$$

for some y.

- This is stronger than mean predictability.
- It allows variance, skewness, and tail behavior to depend on lagged information.
- Nonparametric methods are natural because the conditional distribution is rarely known parametrically.

Nonparametric autoregression: the object of interest

For horizon k and lag order p, define the conditional mean surface

$$m_k(y_1, \ldots, y_p) = E(Y_{t+k} \mid Y_t = y_1, \ldots, Y_{t+1-p} = y_p) .$$

This is the nonparametric analogue of a linear AR model.

Compare with AR(p)

A linear AR(p) imposes

$$E(Y_{t+1} \mid Y_t, \ldots, Y_{t+1-p}) = c + \phi_1 Y_t + \cdots + \phi_p Y_{t+1-p}.$$

A nonparametric autoregression replaces the linear function by an unknown smooth surface.

- More flexible, but much more data-hungry.
- Suffers from the curse of dimensionality as p grows.

Univariate kernel estimator for the conditional mean

In the simplest case $p = 1$, a local weighted estimator of $m_k(y)$ is

$$\widehat{m}_k(y) = \frac{\sum_{t=1}^{T-k} K\left(\frac{Y_t - y}{h}\right) Y_{t+k}}{\sum_{t=1}^{T-k} K\left(\frac{Y_t - y}{h}\right)}.$$

This is just a Nadaraya–Watson estimator with response Y_{t+k} and regressor Y_t.

- If $\widehat{m}_k(y)$ is flat in y, then the conditional mean does not vary with the current state.
- If it bends, then the data suggest state-dependent predictability.

Econometric trade-off

A flexible $\widehat{m}_k(y)$ can reveal nonlinear structure, but one pays with larger variance and bandwidth sensitivity.

Testing the martingale-difference restriction

A relaxed efficient-markets restriction is the martingale-difference condition

$$E(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = 0 \quad \text{a.s.}$$

Under this null and suitable regularity conditions, the conditional mean surface is constant:

$$m_k(y) = \mu \quad \text{for all } k \text{ and } y.$$

For fixed k and y,

$$\sqrt{Th}\,(\widehat{m}_k(y) - \widehat{\mu}) \Longrightarrow N(0, \omega),$$

where

$$\omega = \int K^2(u)\,du\,\frac{\sigma_k^2(y)}{f_Y(y)}.$$

A feasible studentized statistic is

$$\widehat{S}_k(y) = \frac{\widehat{m}_k(y) - \widehat{\mu}}{\widehat{\text{se}}\{\widehat{m}_k(y)\}} \Longrightarrow N(0, 1).$$

A grid-based nonparametric test of no predictability

If the null implies $m_k(y) = \mu$ for all k and y, then we can test the null over a grid of lags and evaluation points:

$$\sum_{k=1}^{K} \sum_{\ell=1}^{L} \widehat{S}_k(y_\ell)^2 > \chi^2_{KL}(\alpha).$$

- Choose a set of horizons $k = 1, \ldots, K$.
- Choose evaluation points $y_1, \ldots, y_L$ over the support of the process.
- Large values indicate that the conditional mean surface deviates from a constant at one or more states.

Economic interpretation

Rejection means that the conditional mean is state dependent. That does not automatically mean an exploitable trading strategy exists after costs, but it does mean the martingale-difference view is too restrictive for the data.

Practical issues: dimensionality and overfitting

Nonparametric autoregression is powerful, but it is not free.

- As the lag dimension p rises, the effective data per neighborhood collapses.
- Bandwidth choice becomes harder because under-smoothing produces noise, while over-smoothing hides structure.
- Results can be highly sensitive in the tails, where few observations are available.

Applied strategy

In finance, nonparametric AR methods are often used as diagnostic tools or as low-dimensional local alternatives to parametric AR, TAR, or STAR models rather than as unrestricted high-dimensional forecasting engines.

Semiparametric volatility: why go beyond Gaussian QMLE?

Suppose returns satisfy

$$y_t = \nu_t \sigma_t, \qquad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha y_{t-1}^2,$$

where ν_t are IID but their density f is unknown.

- Gaussian QMLE is consistent for many parameter targets, but it is not generally efficient when the innovation density is far from Gaussian.
- Heavy tails and skewness are common in financial returns.
- A semiparametric procedure keeps the GARCH dynamics but learns the innovation density from the standardized residuals.

Main idea

Parametric dynamics for dependence, nonparametric flexibility for the shock distribution.

Semiparametric GARCH likelihood and score

Given a candidate density f and parameter vector θ, the conditional log-likelihood can be written as

$$\mathcal{L}(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\log\sigma_t^2(\theta) - \sum_{t=1}^{T}\log f\left(\frac{y_t}{\sigma_t(\theta)}\right).$$

The corresponding score is

$$\frac{\partial\mathcal{L}}{\partial\theta}(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left(\nu_t(\theta)\frac{f'}{f}(\nu_t(\theta)) + 1\right)\frac{\partial\log\sigma_t^2(\theta)}{\partial\theta}.$$

What changes relative to Gaussian MLE?

Instead of fixing f as standard normal or Student-t, we estimate f from the data, usually from standardized residuals obtained in a first step.

Two-step semiparametric volatility estimation

A standard procedure is:

1. Estimate θ consistently, for example by Gaussian QMLE.
2. Form standardized residuals $\widehat{\nu}_t = y_t / \widehat{\sigma}_t$.
3. Estimate the density f nonparametrically by KDE applied to $\widehat{\nu}_t$.
4. Re-estimate θ using the semiparametric likelihood with $\widehat{f}$.

Why this helps

The second-step likelihood adapts to the empirical shape of the shocks, so it can deliver gains in robustness and efficiency when Gaussianity is badly violated.

Caution

The first-step residuals are generated regressors, so the asymptotic theory is more delicate than in a one-step parametric MLE.

Fully flexible volatility functions

The parametric GARCH recursion can itself be relaxed by writing

$$\sigma_t^2 = g(y_{t-1}, \dots, y_{t-p}),$$

where g is an unknown function. For $p = 1$, one kernel estimator is

$$\widehat{g}(y) = \frac{\sum_{t=2}^{T} K\left(\frac{y - y_{t-1}}{h}\right) y_t^2}{\sum_{t=2}^{T} K\left(\frac{y - y_{t-1}}{h}\right)}.$$

To adjust for the conditional mean, the textbook also introduces

$$\widehat{g}_m(y) = \widehat{E}(y_t^2 \mid y_{t-1} = y) - \widehat{E}(y_t \mid y_{t-1} = y)^2,$$

and the residual-based analogue $\widehat{g}_r(y)$.

Interpretation

Now volatility is itself a local regression object. The gain is flexibility; the cost is that high-dimensional volatility dynamics become very hard to estimate nonparametrically.

What we would report in an applied R exercise

For a serious empirical exercise, the report should contain:

1. the bandwidth-selection rule and sensitivity checks;
2. the estimated density or regression curve together with the raw data backdrop;
3. the chosen evaluation grid for nonparametric predictability tests;
4. a comparison with a benchmark parametric model, such as AR, TAR, or GARCH;
5. economic interpretation, not only statistical significance.

Empirical discipline

Nonparametric methods are flexible, so it is easy to over-interpret noise. Always compare shapes across bandwidths and against simpler benchmark models.

Why move to the frequency domain?

So far, second-order dependence has been described through autocovariances $\gamma(h)$. The frequency-domain view asks a different question:

Question

How is the variance of the process distributed across low, medium, and high frequencies?

- Low frequencies correspond to slow, persistent movements.
- High frequencies correspond to rapid oscillation.
- Cycles show up as peaks in the spectrum.

Key equivalence

For a covariance-stationary process, the autocovariance function and the spectrum contain the same second-order information. They are just two languages for the same object.

From autocovariances to the spectrum

Let $\{Y_t\}$ be covariance stationary with autocovariance function $\gamma(h)$. The autocovariance-generating function is

$$g_Y(z) = \sum_{h=-\infty}^{\infty} \gamma(h) z^h.$$

Evaluating at $z = e^{-i\lambda}$ and dividing by 2π gives the population spectrum:

$$s_Y(\lambda) = \frac{1}{2\pi} g_Y(e^{-i\lambda}) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\lambda h}.$$

Interpretation

This is a Fourier transform of the autocovariance sequence. The lag index h is transformed into the frequency index λ.

Cosine representation and basic properties

Using Euler's formula $e^{-i\lambda h} = \cos(\lambda h) - i \sin(\lambda h)$ and the symmetry $\gamma(h) = \gamma(-h)$, the spectrum becomes

$$s_Y(\lambda) = \frac{1}{2\pi} \left\{ \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h) \cos(\lambda h) \right\}.$$

Hence, for a univariate covariance-stationary process:

- $s_Y(\lambda)$ is real valued and nonnegative;
- $s_Y(\lambda)$ is symmetric: $s_Y(-\lambda) = s_Y(\lambda)$;
- $s_Y(\lambda)$ is periodic with period 2π;
- only frequencies on $[0, \pi]$ are needed for a full description.

Variance decomposition

$$\gamma(0) = \mathrm{Var}(Y_t) = \int_{-\pi}^{\pi} s_Y(\lambda)\, d\lambda.$$

The spectrum distributes total variance across frequencies.

Cumulative spectrum and dominant frequencies

If we normalize $s_Y(\lambda)$ by its integral, it behaves like a density over frequencies. The cumulative spectrum is

$$F_Y(\lambda) = \int_{-\pi}^{\lambda} s_Y(u)\, du.$$

- Peaks of $s_Y(\lambda)$ indicate frequencies with large power.
- A dominant low-frequency peak means slow, persistent movement.
- A dominant high-frequency peak means rapid alternation.

Useful mental picture

The autocovariance function tells you how the series co-moves with its lagged copies. The spectrum tells you which sinusoidal components are most responsible for that co-movement.

Spectral Representation Theorem

A cornerstone result states that a zero-mean stationary process can be represented as

$$Y_t = \int_{-\pi}^{\pi} e^{i\lambda t}\, dF_Y(\lambda),$$

where $dF_Y(\lambda)$ is a complex-valued random measure.

Meaning

A stationary time series can be decomposed into an integral of sinusoidal components of all possible frequencies, each with a random amplitude and phase.

- In the time domain, we see lags and autocovariances.
- In the frequency domain, we see waves of different frequencies.
- The theorem is the bridge connecting the two views.

Why complex exponentials appear: Euler and De Moivre

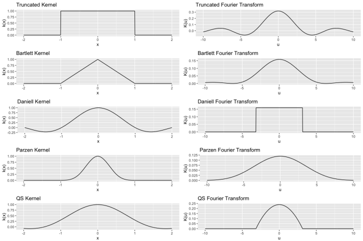
Euler's formula says

$$e^{i\theta} = \cos\theta + i\sin\theta,$$

so the complex exponential is just a convenient package for sine and cosine waves. De Moivre's theorem gives

$$(e^{i\theta})^n = e^{in\theta} = \cos(n\theta) + i\sin(n\theta).$$

- Frequency-domain analysis uses these exponentials because they are algebraically convenient eigenfunctions of linear filters.
- Lag polynomials act very simply on $e^{i\lambda t}$.



Two benchmark examples: white noise and pure periodicity

White noise

If $Y_t = \varepsilon_t$ with $\mathrm{Var}(\varepsilon_t) = \sigma^2$, then

$$s_Y(\lambda) = \frac{\sigma^2}{2\pi}.$$

Every frequency contributes equally: the spectrum is flat.

Pure periodic component

A perfectly periodic component concentrates power at one frequency. In spectral language, this appears as a spike or point mass at the corresponding λ_0.

Intuition

Flat spectrum means "no preferred frequency." A spike means "almost all variance comes from a cycle of one particular length."

Spectrum of an MA(∞) process

Suppose

$$Y_t = \mu + \psi(L)\varepsilon_t, \qquad \psi(L) = \sum_{j=0}^{\infty} \psi_j L^j, \qquad \sum_{j=0}^{\infty} |\psi_j| < \infty,$$

with white-noise innovations of variance σ^2. Then

$$s_Y(\lambda) = \frac{\sigma^2}{2\pi} \psi(e^{-i\lambda})\psi(e^{i\lambda}) = \frac{\sigma^2}{2\pi} \left|\psi(e^{-i\lambda})\right|^2.$$

Interpretation

The squared modulus of the transfer function tells us how the filter $\psi(L)$ amplifies or attenuates each frequency.

MA(1) spectrum and the sign of θ

For

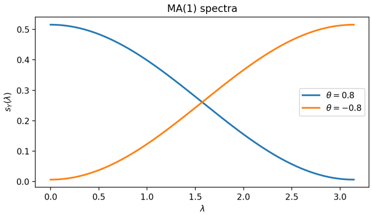
$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1},$$

the spectrum is

$$s_Y(\lambda) = \frac{\sigma^2}{2\pi}\left[1 + \theta^2 + 2\theta\cos(\lambda)\right].$$

- If $\theta > 0$, the spectrum is largest at low frequencies and declines with λ.
- If $\theta < 0$, the spectrum is largest at high frequencies and rises with λ.

Positive MA dependence smooths the series; negative MA dependence creates short-run alternation.



AR(1) spectrum and persistence

For

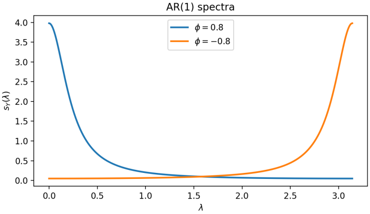
$$Y_t = c + \phi Y_{t-1} + \varepsilon_t, \qquad |\phi| < 1,$$

the spectrum is

$$s_Y(\lambda) = \frac{\sigma^2}{2\pi}\left[1 + \phi^2 - 2\phi\cos(\lambda)\right]^{-1}.$$

- If $\phi > 0$, the denominator is smallest near $\lambda = 0$, so low frequencies dominate.
- If $\phi < 0$, the denominator is smallest near $\lambda = \pi$, so high frequencies dominate.

As $\phi \to 1$, the spectrum piles up near zero frequency, which is the frequency-domain signature of strong persistence.



General ARMA(p, q) spectrum

For the stationary and invertible ARMA(p, q) model,

$$\phi(L)Y_t = \theta(L)\varepsilon_t,$$

with

$$\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j, \qquad \theta(z) = 1 + \sum_{j=1}^{q} \theta_j z^j,$$

the population spectrum is

$$s_Y(\lambda) = \frac{\sigma^2}{2\pi} \frac{\left|1 + \sum_{k=1}^{q} \theta_k e^{-ik\lambda}\right|^2}{\left|1 - \sum_{k=1}^{p} \phi_k e^{-ik\lambda}\right|^2}.$$

How to read it

The MA polynomial shapes the numerator, the AR polynomial shapes the denominator, and the spectrum is their frequency-by-frequency ratio.

Roots, poles, and zeros

If the AR and MA polynomials are factored in terms of their roots, the spectrum can be written as

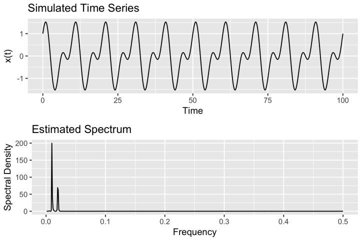
$$s_Y(\lambda) = \frac{\sigma^2 \prod_{j=1}^{q} \left[1 + \eta_j^2 - 2\eta_j \cos(\lambda)\right]}{2\pi \prod_{j=1}^{p} \left[1 + \omega_j^2 - 2\omega_j \cos(\lambda)\right]}.$$

- MA roots close to the unit circle create *zeros* or troughs at certain frequencies.
- AR roots close to the unit circle create *poles* or large peaks at certain frequencies.
- Near-unit roots in AR components create strong low-frequency concentration.

Economic interpretation

Persistence is a low-frequency phenomenon. Oscillatory alternation is a high-frequency phenomenon. The location of spectral mass tells us which one dominates.

From frequency-domain intuition to data analysis

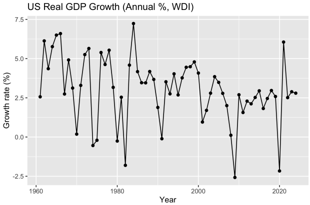


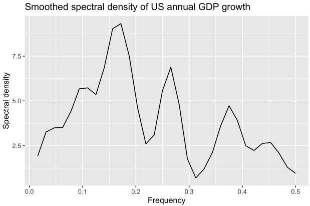
A periodic signal generates strong spectral peaks.

- A series that visually oscillates often produces a spectrum with clear peaks.
- A highly persistent series produces spectral mass near zero frequency.
- A noisy, weakly dependent series looks spectrally flatter.

The goal of spectral analysis is to make these statements quantitative.

Empirical preview: GDP growth and business-cycle frequencies





- Annual GDP growth has no seasonal frequency in the monthly or quarterly sense, but it may still concentrate variance at medium- and low-frequency business-cycle horizons.
- The spectrum summarizes which cycle lengths matter most.

Boundary with Lecture 13

Today we focus on the *population spectrum and interpretation*. Next lecture studies the *sample periodogram*, spectral estimation, HAC long-run variance estimators, and bandwidth choice.

A short R preview for the transition to Chapter 6

```
# Theoretical AR(1) spectrum on a frequency grid
phi <- 0.8
lam <- seq(0, pi, length.out = 400)
spec_ar1 <- (1 / (2*pi)) / (1 + phi^2 - 2*phi*cos(lam))

# Empirical spectrum of a stationary series (details next lecture)
y_ts <- ts(as.numeric(r))
sp <- spectrum(y_ts, plot = FALSE)
head(data.frame(freq = sp$freq, spec = sp$spec))
```

Purpose of this slide

Theoretical formulas let us understand the shape of spectra analytically. The `spectrum()` function is the empirical gateway, but we postpone estimation theory until Lecture 13.

Lecture summary

1. Nonparametric tools are useful because linear and Gaussian restrictions are often too narrow for financial data.
2. Runs tests, transformed-predictor tests, and nonparametric autoregressions all ask whether lagged information matters beyond a linear AR benchmark.
3. Semiparametric volatility modelling combines structured dynamics with flexible innovation densities or flexible variance functions.
4. The frequency-domain view repackages the same second-order information into a spectral density over frequencies.
5. AR and MA parameters have clear spectral signatures: low-frequency concentration, high-frequency alternation, peaks, troughs, poles, and zeros.

Next lecture

Lecture 13 turns this frequency-domain intuition into inference: the sample periodogram, HAC long-run variance estimation, positive semi-definiteness, kernels, and bandwidth choice.