Lecture 13 — Sample Periodogram, HAC Estimation, and Kernel/Bandwidth Choice

Chapter 6: frequency-domain intuition, long-run variance estimation, and robust inference under dependence

Jiajing Sun

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

Econometrics and Time Series Methods
Spring 2026

Why Lecture 13 matters

Lecture 12 ended by introducing the *spectrum* as a frequency-domain description of second-order dependence. Lecture 13 turns that idea into *robust inference*.

Core econometric logic

$$\text{serial dependence} \Longrightarrow \text{long-run variance} \Longrightarrow \text{HAC estimation.}$$

- The *periodogram* tells us how sample variance is distributed across frequencies.
- The *long-run variance* is the spectral density at frequency zero, up to the factor 2π.
- HAC estimators are therefore best understood as *smoothed low-frequency estimators*.

Big picture

The lecture moves from *frequency-domain intuition* to *robust standard errors* and finally to the practical questions of *positive semi-definiteness, kernels, and bandwidth choice*.

Learning goals

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

1. define the sample periodogram and explain its relationship to the sample autocovariance function;
2. interpret the long-run variance as 2π times the spectrum at zero frequency;
3. derive the asymptotic variance of OLS estimators under heteroskedasticity and autocorrelation;
4. write down the Newey–West/HAC long-run variance estimator in both the time and frequency domains;
5. explain why raw covariance sums can fail to be positive semi-definite;
6. explain how nonnegative spectral windows guarantee positive semi-definite HAC estimators;
7. compare truncated, Bartlett, Parzen, Daniell, and quadratic-spectral kernels;
8. discuss the bias–variance trade-off induced by the bandwidth choice.

Three-hour plan

Hour 1

Sample periodogram, Fourier frequencies, the interpretation of spectra, and why low frequencies matter for long-run variance estimation.

Hour 2

HAC inference in linear regression: asymptotic variance, Newey–West estimation, weighted sums of covariances, and weighted periodograms.

Hour 3

Positive semi-definiteness, kernel choice, bandwidth choice, practical implementation, and the transition to fixed-b and self-normalization.

Theme of the lecture

The lecture is not just about "robust standard errors." It is about learning to see dependence through two equivalent lenses:

$$\text{lags/autocovariances} \quad \Longleftrightarrow \quad \text{frequencies/spectral power.}$$

Three equivalent languages for the same second-order object

For a covariance-stationary process $\{Y_t\}$ with autocovariance function $\gamma(h)$, the second-order structure can be expressed in three equivalent ways:

Time domain

$$\{\gamma(h) : h \in \mathbb{Z}\}.$$

Frequency domain

$$s_Y(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\lambda h}, \qquad \lambda \in [-\pi, \pi].$$

Inference object

$$\Omega_Y = \sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi s_Y(0).$$

Interpretation

The long-run variance is simply the *low-frequency content* of the process, evaluated at frequency zero.

From autocovariances to the spectrum

If $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, the spectral density exists and is

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

The inverse Fourier relationship is

$$\gamma(h) = \int_{-\pi}^{\pi} e^{i\lambda h} s_Y(\lambda)\, d\lambda.$$

- Large power near $\lambda = 0$ means persistent low-frequency movement.
- Power concentrated away from zero means oscillatory or cyclical movement.
- A flat spectrum corresponds to white noise.

Econometric message

The spectrum contains the same second-order information as the autocovariance function, but reorganized by *frequency* rather than *lag*.

Why zero frequency matters for robust inference

Set

$$S_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_t.$$

Then, under weak dependence,

$$\mathrm{Var}(S_T) \to \Omega_Y \qquad \text{with} \qquad \Omega_Y = \sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi s_Y(0).$$

Interpretation of $s_Y(0)$

Frequency zero corresponds to movements with *infinite period*, that is, very slow accumulation over time. Statistics built from *partial sums* are therefore governed by the spectrum *near zero*.

- If the data are positively autocorrelated, Ω_Y exceeds $\gamma(0)$.
- If the data are negatively autocorrelated, Ω_Y can be smaller than $\gamma(0)$.
- Classical standard errors implicitly replace Ω_Y by $\gamma(0)$, which is generally wrong under dependence.

Benchmark spectra: white noise, MA(1), AR(1), and ARMA

For a white-noise process, the spectrum is flat:

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

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

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

For an AR(1), $Y_t = \phi Y_{t-1} + \varepsilon_t$ with $|\phi| < 1$,

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

More generally,

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

Intuition

Positive AR persistence piles up power near zero; negative serial correlation shifts power toward higher frequencies.

Sample autocovariances and the sample periodogram

Given observations $Y_1, \dots, Y_T$, define

$$\widehat{\gamma}(j) = \begin{cases} T^{-1} \sum_{t=j+1}^{T} (Y_t - \bar{Y})(Y_{t-j} - \bar{Y}), & j = 0, 1, \dots, T-1, \\ \widehat{\gamma}(-j), & j = -1, -2, \dots, -T+1. \end{cases}$$

The sample periodogram is the plug-in estimator

$$\widehat{I}(\lambda) = \frac{1}{2\pi} \sum_{j=-T+1}^{T-1} \widehat{\gamma}(j) e^{-i\lambda j} = \frac{1}{2\pi} \left[\widehat{\gamma}(0) + 2 \sum_{j=1}^{T-1} \widehat{\gamma}(j) \cos(\lambda j) \right].$$

Key interpretation

The periodogram is the sample analogue of the spectrum. It attributes sample variance to different frequencies.

In spectral analysis, the raw periodogram is the basic descriptive object, but not yet the final estimator for inference.

The discrete Fourier transform (DFT) view

The same object can be written in terms of the DFT:

$$d_T(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} (Y_t - \bar{Y}) e^{-i\lambda t}, \qquad \widehat{I}(\lambda) = d_T(\lambda)\overline{d_T(\lambda)}.$$

In the scalar case,

$$\widehat{I}(\lambda) = |d_T(\lambda)|^2.$$

- The real part of $d_T(\lambda)$ captures correlation with cosine waves.
- The imaginary part captures correlation with sine waves.
- Squaring and adding the two parts gives the power at frequency λ.

Econometric insight

The periodogram is the squared size of the sample Fourier coefficient. It tells us how strongly the data line up with a sinusoid of frequency λ.

Fourier frequencies, orthogonality, and the Nyquist limit

For a sample of size T, the canonical Fourier frequencies are

$$\lambda_j = \frac{2\pi j}{T}, \qquad j = 0, 1, \ldots, T-1.$$

At these frequencies, the sample sine and cosine functions are orthogonal. This is why spectral analysis is computationally convenient.

Two practical limits

- **Record length** determines the lowest detectable frequency.
- **Sampling interval** determines the highest detectable frequency, the *Nyquist frequency*.

If the data are sampled weekly, you cannot identify cycles shorter than two weeks. If the sample is short, you cannot distinguish very close low-frequency components.

Takeaway

Spectral resolution is a sample-design issue, not just a statistical one.

The area under the periodogram equals the sample variance

The raw periodogram satisfies

$$\int_{-\pi}^{\pi} \widehat{I}(\lambda)\, d\lambda = \widehat{\gamma}(0), \qquad 2\int_{0}^{\pi} \widehat{I}(\lambda)\, d\lambda = \widehat{\gamma}(0).$$

Meaning of this identity

- $\widehat{\gamma}(0)$ is the sample variance.
- The periodogram distributes that variance across frequencies.
- Integrating the spectral power over all frequencies recovers total variance.

So the periodogram should be read as a *variance decomposition by frequency*. Large spikes mark frequencies that account for a disproportionate share of the sample variance.

Periodogram as a finite Fourier regression

For even T, one may view the time series as the dependent variable in the regression

$$Y_t = a_0 + \sum_{j=1}^{T/2-1} \left[a_j \cos\left(\frac{2\pi jt}{T}\right) + b_j \sin\left(\frac{2\pi jt}{T}\right) \right] + a_{T/2} \cos(\pi t).$$

The periodogram ordinates are

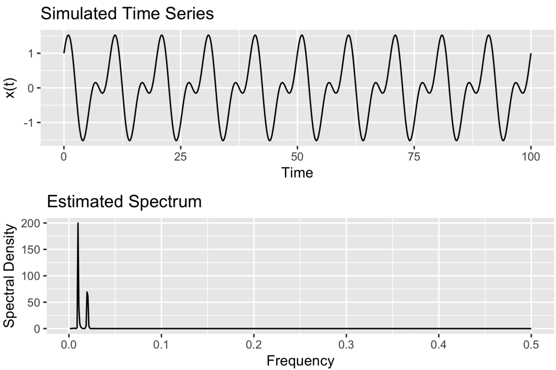
$$I_j = a_j^2 + b_j^2, \qquad j = 1, \ldots, T/2 - 1.$$

- a_j and b_j tell us how much the data correlate with a sinusoid of frequency $2\pi j/T$.
- I_j measures the variance explained by that oscillation.
- A line spectrum is just the plot of these ordinates against frequency.

Intuition

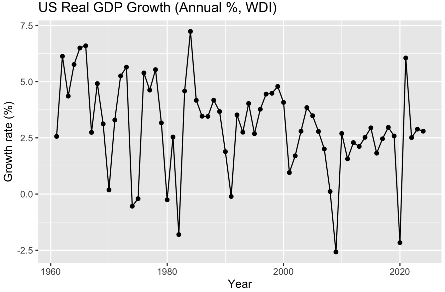
A time series can be read as a superposition of waves. The periodogram tells us which waves matter most.

Example: a simple sinusoidal series and its estimated spectrum

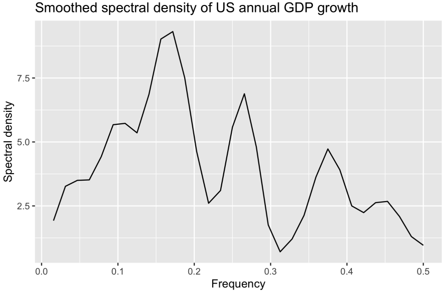


- The simulated signal combines a cosine with period 10 and a sine with period 5.
- The estimated spectrum exhibits peaks at the corresponding frequencies.
- Spectral methods therefore recover *dominant oscillatory components* very naturally.

Example: annual GDP growth and low-frequency power



Annual GDP growth fluctuates over medium- and long-run horizons.



The smoothed spectrum concentrates substantial mass at low frequencies.

Why this matters

When spectral mass is concentrated near zero, partial sums accumulate strongly. This is exactly why long-run variance estimation matters for inference.

Three drawbacks of the raw periodogram

The raw periodogram is useful for intuition, but it has serious limitations:

1. **Resolution depends on sample length.** The available frequencies change with T.
2. **It becomes more spiky as T grows.** More data do not automatically produce a smoother graph.
3. **It is not a consistent estimator of the spectral density.**

Why the third point matters

Inference requires estimators whose variability vanishes asymptotically. The raw periodogram fails that requirement.

- Its expectation is close to the spectrum.
- But its variance does not disappear.
- Therefore we must *smooth* it.

Why the raw periodogram is not consistent

In the IID case with mean zero,

$$s_Y(\lambda) = \frac{\gamma(0)}{2\pi}, \qquad \mathbb{E}\{\widehat{I}(\lambda)\} = s_Y(\lambda),$$

so the raw periodogram is asymptotically unbiased.
However,

$$\mathrm{Var}\{\widehat{I}(\lambda)\} = O(1),$$

not $o(1)$. Hence

$$\widehat{I}(\lambda) \not\xrightarrow{p} s_Y(\lambda).$$

Core reason

The raw periodogram uses a very large collection of noisy sample autocovariances or, equivalently, noisy sample Fourier coefficients. More data add more ordinates, not automatic smoothing.

Lesson

To estimate a spectrum consistently, one must *average nearby frequencies* or equivalently *taper the sample autocovariances*.

Smoothing the periodogram

A generic smoothed spectral estimator takes the form

$$\widehat{s}_Y^{sm}(\lambda) = \sum_{\ell=-(T-1)}^{T-1} K_b(\lambda - \lambda_\ell)\,\widehat{I}(\lambda_\ell),$$

where $K_b(\cdot)$ is a spectral window and b is a bandwidth or smoothing parameter. Equivalent time-domain form:

$$\widehat{s}_Y^{sm}(\lambda) = \frac{1}{2\pi} \sum_{|h|<T} k_m(h)\,\widehat{\gamma}(h) e^{-i\lambda h}.$$

- *Frequency-domain smoothing*: average nearby ordinates.
- *Time-domain smoothing*: downweight distant autocovariances.

Two views, one estimator

The smoothed periodogram and the HAC estimator are the same idea, just expressed in two different languages.

Transition to HAC: estimate the spectrum near zero

For inference on means, regression coefficients, and moment conditions, the target is not the full spectrum but the value at zero:

$$\Omega_Y = 2\pi s_Y(0).$$

Hence a HAC estimator is simply a *smoothed estimator of the spectrum at zero frequency*:

$$\widehat{\Omega}_Y = 2\pi \,\widehat{s}_Y^{sm}(0).$$

Practical reading

- If there is strong low-frequency power, the long-run variance is large.
- If the process is close to white noise, the long-run variance is close to the one-period variance.
- Robust inference is therefore fundamentally about *estimating low-frequency power well*.

Why OLS t-tests fail under dependence

Consider the regression

$$Y_t = X_t'\beta + u_t, \qquad t = 1, \ldots, T,$$

with $E(u_t \mid X_t) = 0$, but where $\{u_t\}$ may be heteroskedastic and autocorrelated.

- OLS can still be consistent under suitable exogeneity.
- The problem is the *variance formula*, not necessarily the point estimate.
- Classical OLS standard errors assume

$$E(u_t u_s \mid X) = 0 \quad \text{for } t \neq s, \qquad E(u_t^2 \mid X) = \sigma^2.$$

When these assumptions fail, the usual t-statistics need not have the stated asymptotic distribution, and nominal significance levels can be badly misleading.

Regression setup and score process

The OLS estimator is

$$\widehat{\beta} = \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{T} X_t Y_t.$$

Substituting $Y_t = X_t'\beta + u_t$ gives

$$\sqrt{T}(\widehat{\beta} - \beta) = \left(\frac{1}{T} \sum_{t=1}^{T} X_t X_t' \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t u_t \right).$$

Define the score-like process

$$v_t = X_t u_t.$$

Immediate implication

Robust inference about β reduces to inference about the partial sum

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_t.$$

So the entire problem is a long-run variance problem for $\{v_t\}$.

Asymptotic variance of OLS under heteroskedasticity and autocorrelation

Assume a WLLN and a CLT:

$$\frac{1}{T}\sum_{t=1}^{T} X_t X_t' \xrightarrow{p} Q, \qquad \frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_t \Longrightarrow N(0,\Omega_v).$$

Then

$$\sqrt{T}(\widehat{\beta}-\beta) \Longrightarrow N(0, Q^{-1}\Omega_v Q^{-1}),$$

where

$$\Omega_v = \sum_{j=-\infty}^{\infty} \Gamma_v(j), \qquad \Gamma_v(j) = E(v_t v_{t-j}').$$

Sandwich form

$$\underbrace{Q^{-1}}_{\text{bread}} \quad \text{long-run} \underbrace{\Omega_v}_{\text{variance of } X_t u_t} \quad \underbrace{Q^{-1}}_{\text{bread}}.$$

Under IID homoskedastic errors, Ω_v collapses to the familiar one-period variance matrix. Under serial correlation, it does not.

The long-run variance matrix in time and frequency domains

For $v_t = X_t u_t$, define

$$\Omega_v = \sum_{j=-\infty}^{\infty} \Gamma_v(j).$$

This is the long-run variance–covariance matrix of the partial sums of v_t.
Equivalent frequency-domain representation:

$$\Omega_v = 2\pi S_v(0),$$

where

$$S_v(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma_v(j) e^{-i\lambda j}.$$

Interpretation

- The covariance matrix of a *sum* depends on all autocovariances, not just the variance at lag 0.
- Positive serial correlation increases Ω_v.
- Negative serial correlation can decrease Ω_v.

Newey–West / Bartlett HAC estimator

The best-known HAC estimator is the Newey–West estimator:

$$\widehat{\Omega}_v^{NW} = \sum_{j=-m}^{m} \left(1 - \frac{|j|}{m+1}\right) \widehat{\Gamma}_v(j), \qquad \widehat{\Gamma}_v(j) = \frac{1}{T} \sum_{t=|j|+1}^{T} \widehat{v}_t \widehat{v}_{t-j}'.$$

Here $\widehat{v}_t = X_t \widehat{u}_t$ and m is the bandwidth (truncation lag).

- The Bartlett weights taper linearly to zero.
- Nearby lags receive larger weight.
- Very distant lags are discarded.

Consistency conditions

For standard HAC asymptotics, we require

$$m \to \infty, \qquad m/T \to 0.$$

Historical note: White, Newey–West, and Andrews

- **White (1980):** heteroskedasticity-robust covariance estimation.
- **Newey–West (1987):** HAC covariance estimation with a simple PSD Bartlett lag window.
- **Andrews (1991, 1992):** general kernels, bandwidth choice, and prewhitening refinements.

Why these papers matter

They made robust inference operational in empirical econometrics by turning abstract long-run variance formulas into implementable estimators.

HAC as a weighted sum of sample autocovariances

The generic time-domain HAC estimator is

$$\widehat{\Omega}_v^{SC} = \sum_{j=-(T-1)}^{T-1} k_m(j)\widehat{\Gamma}_v(j),$$

where $k_m(j)$ is a lag window or kernel.

Time-domain reading

- Start from the infinite sum $\Omega_v = \sum_j \Gamma_v(j)$.
- Replace population autocovariances by sample autocovariances.
- Downweight high lags to stabilize variance.

Common choices:

$$\text{Truncated: } k_m(j) = 1(|j| \le m), \qquad \text{Bartlett: } k_m(j) = \left(1 - \frac{|j|}{m+1}\right) 1(|j| \le m).$$

Limitation

Not every lag-window estimator is automatically positive semi-definite.

HAC as a weighted periodogram

The same estimator may be written in the frequency domain as

$$\widehat{\Omega}_v^{WP} = 2\pi \sum_{\ell=-(T-1)}^{T-1} K_b(\lambda_\ell)\widehat{I}_v(\lambda_\ell), \qquad \lambda_\ell = \frac{2\pi\ell}{T}.$$

Here $\widehat{I}_v(\lambda_\ell)$ is the periodogram of v_t and $K_b(\cdot)$ is a spectral window centered at zero frequency.

Frequency-domain reading

- The long-run variance depends on $S_v(0)$.
- So we average periodogram ordinates *near zero*.
- The bandwidth controls how much low-frequency averaging we perform.

This is why HAC estimation is often described as *smoothed low-frequency periodogram estimation*.

Why the time-domain and frequency-domain estimators are the same

The link is the Fourier transform:

$$k_m(j) = \sum_{\ell=-(T-1)}^{T-1} K_b(\lambda_\ell) e^{-ij\lambda_\ell}.$$

Hence

$$\widehat{\Omega}_v^{WP} = \widehat{\Omega}_v^{SC}.$$

Meaning

- Tapering the autocovariances in the time domain is the same as smoothing the periodogram in the frequency domain.
- The choice of kernel can be described either as a *lag window* or as a *spectral window*.

Important econometric lesson

Many apparently different robust estimators are just different representations of one underlying smoothing problem.

Scalar mean example: robust inference for μ

If Y_t is stationary with mean μ, then

$$\sqrt{T}(\bar{Y} - \mu) \Longrightarrow N(0, \Omega_Y), \qquad \Omega_Y = \sum_{h=-\infty}^{\infty} \gamma(h).$$

A HAC-t statistic replaces Ω_Y by a consistent estimator:

$$t_{HAC} = \frac{\sqrt{T}(\bar{Y} - \mu_0)}{\sqrt{\widehat{\Omega}_Y}}.$$

Why this is the prototype

- Every regression coefficient can be treated similarly after defining the appropriate score process.
- If you understand HAC inference for the mean, you understand the core logic for OLS and GMM as well.

The only real difficulty is estimating the long-run variance accurately in finite samples.

Prewhitening: reduce serial correlation before smoothing

If v_t is strongly autocorrelated, one can first fit a short VAR(p):

$$v_t = \Phi_1 v_{t-1} + \cdots + \Phi_p v_{t-p} + \varepsilon_t,$$

compute a HAC estimator for the residual process ε_t, and then recolor:

$$\widehat{\Omega}_v = \left(I - \sum_{i=1}^{p} \widehat{\Phi}_i\right)^{-1} \widehat{\Omega}_\varepsilon \left(I - \sum_{i=1}^{p} \widehat{\Phi}_i\right)^{-1\prime}.$$

- Prewhitening can reduce small-sample bias.
- It is especially useful when persistence is substantial.
- But it introduces an additional modelling step, so implementation must be careful.

Idea

Rather than asking the kernel to do all the work, remove part of the persistence first, and then smooth what remains.

Why positive semi-definiteness matters

A covariance estimator should be positive semi-definite (PSD):

$$a'\widehat{\Omega}a \geq 0 \qquad \text{for all vectors } a.$$

- Estimated variances must not be negative.
- Wald statistics require inversion of covariance matrices.
- Portfolio and GMM applications often need a well-behaved covariance matrix.

Problem

A naive weighted sum of sample autocovariances can violate PSD, even when the true long-run variance matrix is PSD by construction.

So kernel choice is not just a smoothing decision; it is also an *admissibility* decision.

A simple failure of naive covariance summation

Consider the scalar estimator

$$\widetilde{\Omega} = \sum_{j=-(m-1)}^{m-1} \widehat{\gamma}(j).$$

If $m = 2$, then

$$\widetilde{\Omega} = \widehat{\gamma}(0) + 2\widehat{\gamma}(1) = \widehat{\gamma}(0)\left(1 + 2\frac{\widehat{\gamma}(1)}{\widehat{\gamma}(0)}\right).$$

Failure condition

If the first sample autocorrelation satisfies

$$\frac{\widehat{\gamma}(1)}{\widehat{\gamma}(0)} < -\frac{1}{2},$$

then $\widetilde{\Omega} < 0$.

Lesson

Even in the scalar case, an apparently natural covariance sum can produce an impossible "variance." In the multivariate case, the issue is even more serious.

How nonnegative spectral windows guarantee PSD

For any vector a,

$$a'\widehat{\Omega}^{WP}a = 2\pi \sum_{\ell=-(T-1)}^{T-1} K_b(\lambda_\ell)\, a'\widehat{I}_v(\lambda_\ell)a.$$

Because

$$a'\widehat{I}_v(\lambda_\ell)a = \left|a'd_v(\lambda_\ell)\right|^2 \geq 0,$$

we obtain

$$a'\widehat{\Omega}^{WP}a = 2\pi \sum_{\ell} K_b(\lambda_\ell)\left|a'd_v(\lambda_\ell)\right|^2 \geq 0$$

whenever

$$K_b(\lambda_\ell) \geq 0 \quad \text{for all } \ell.$$

Practical criterion

Choose a nonnegative spectral window. Then the weighted periodogram estimator is automatically PSD.

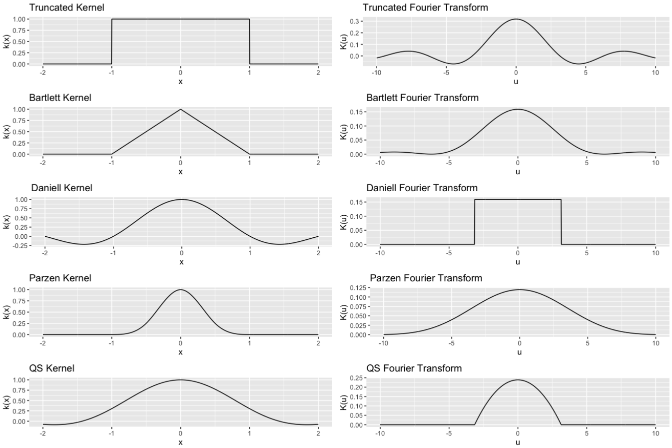
Common kernels in the time domain and their spectral counterparts

Kernel	Time-domain $k(x)$	Frequency-domain $K(u)$
Truncated	$\mathbf{1}(\lvert x\rvert \le 1)$	$\frac{1}{\pi}\frac{\sin u}{u}$
Bartlett	$(1-\lvert x\rvert)\mathbf{1}(\lvert x\rvert \le 1)$	$\frac{1}{2\pi}\left[\frac{\sin(u/2)}{u/2}\right]^2$
Daniell	$\frac{\sin(\pi x)}{\pi x}$	$\frac{1}{2\pi}\mathbf{1}(\lvert u\rvert \le \pi)$
Parzen	$\begin{cases} 1-6x^2+6\lvert x\rvert^3, & \lvert x\rvert \le 1/2, \\ 2(1-\lvert x\rvert)^3, & 1/2 < \lvert x\rvert \le 1, \\ 0, & \text{otherwise} \end{cases}$	$\frac{3}{8\pi}\left[\frac{\sin(u/4)}{u/4}\right]^4$
Quadratic spectral	$\frac{3}{(\pi x)^2}\left[\frac{\sin(\pi x)}{\pi x}-\cos(\pi x)\right]$	$\frac{3}{4\pi}\left[1-(u/\pi)^2\right]\mathbf{1}(\lvert u\rvert \le \pi)$

Important

Bandwidth usually matters more than the exact kernel family, but the kernel still matters for PSD and finite-sample behavior.

Kernel pairs in time and frequency domains



- Left panels: lag windows $k(\cdot)$ in the time domain.
- Right panels: spectral windows $K(\cdot)$ after Fourier transformation.

What the common kernels do intuitively

- **Truncated:** keeps all included lags equally and discards the rest abruptly; low bias, but oscillatory spectral behavior.
- **Bartlett:** tapers linearly to zero; simple, PSD, and widely used in practice.
- **Parzen:** smoother tapering, stronger attenuation of high-frequency noise, often lower variance but more bias.
- **Daniell:** corresponds to a flat frequency window; natural when one wants direct averaging of nearby periodogram ordinates.
- **Quadratic spectral:** often attractive in asymptotic MSE calculations because it balances smoothness and low-frequency fidelity.

General rule

Sharper windows preserve more detail but are noisier. Smoother windows suppress noise but risk over-smoothing the true low-frequency structure.

Bandwidth is the main tuning parameter

The bandwidth determines how aggressively we smooth:

$$m = \text{number of autocovariances retained,}$$
$$b = \text{number of periodogram ordinates averaged.}$$

For standard HAC consistency

$$m \to \infty, \qquad m/T \to 0 \qquad \Longleftrightarrow \qquad b \to 0, \qquad Tb \to \infty.$$

- Too small a bandwidth: noisy estimator, high variance.
- Too large a bandwidth: oversmoothing, high bias.
- The same old bias–variance trade-off from nonparametrics returns here.

Message

Kernel choice matters, but the bandwidth usually matters more.

Bias–variance trade-off in the frequency domain

A spectral-window estimator near zero takes the form

$$\widehat{\Omega}^{WP} = 2\pi \sum_{|\ell| \leq b} K_b(\lambda_\ell) \widehat{I}_v(\lambda_\ell).$$

- Larger b averages more nearby frequencies.
- This *reduces variance* by stronger smoothing.
- But it can *increase bias* if the true spectrum changes rapidly near zero.

Interpretation

A flat low-frequency spectrum can tolerate broader averaging. A sharply curved or peaked low-frequency spectrum cannot.

So bandwidth choice is really a judgement about how much local smoothness near $\lambda = 0$ one is willing to impose.

Bias–variance trade-off in the time domain

The time-domain estimator is

$$\widehat{\Omega}^{SC} = \sum_{|j|\leq m} k_m(j)\widehat{\Gamma}_v(j).$$

- Larger m includes more sample autocovariances.
- This *reduces truncation bias*, because more of the true long-run dependence is captured.
- But it *raises variance*, because high-lag sample autocovariances are noisy.

Equivalent reading

Frequency-domain smoothing and time-domain tapering are two manifestations of the same nonparametric trade-off.

Connection to Chapter 5

The bandwidth problem here is conceptually the same as in kernel density estimation and local regression.

Flat spectral windows, Daniell kernels, and the relation $mb = T$

Suppose the frequency-domain window is flat:

$$K(\ell) = \begin{cases} \frac{1}{2b+1}, & |\ell| \le b, \\ 0, & |\ell| > b. \end{cases}$$

Then

$$\widehat{\Omega}^{WP} = \frac{2\pi}{2b+1} \sum_{\ell=-b}^{b} \widehat{I}_v\left(\frac{2\pi\ell}{T}\right).$$

Its time-domain counterpart is approximately a Daniell/sinc-type kernel:

$$k(j) \approx \frac{\sin(2\pi j/m)}{2\pi j/m}, \qquad m = \frac{T}{b}.$$

Why this is useful

It makes the time-domain/frequency-domain duality very concrete:

$$\text{few frequencies averaged} \iff \text{many lags retained,}$$

and vice versa.

Automatic bandwidth choice and rules of thumb

A classic practical choice is the Andrews rule of thumb for Bartlett kernels:

$$m \propto T^{1/3}.$$

For example, a simple rule is

$$m \approx 0.75\, T^{1/3}.$$

- Plug-in methods estimate nuisance quantities and choose m to minimize an asymptotic MSE criterion.
- Software defaults often implement such rules.
- Prewhitening is sometimes combined with bandwidth selection.

But beware

Bandwidths that look good from an MSE perspective need not produce the best *test size* in finite samples. This is why Lecture 14 moves to fixed-b and self-normalization.

Kernel choice versus bandwidth choice

What usually matters more?

In many empirical applications:

bandwidth choice matters more than kernel choice.

Why?

- Different reasonable kernels produce qualitatively similar smoothers.
- Changing the bandwidth can radically change the amount of smoothing.
- But kernel choice still matters for PSD, edge behavior, and finite-sample performance.

Reasonable empirical strategy

Fix a defensible PSD kernel, choose the bandwidth transparently, and report sensitivity to nearby alternatives.

A practical workflow for empirical HAC inference

1. Identify the score or moment process v_t whose partial sums govern the statistic.
2. Check whether persistence is mild or substantial.
3. Decide whether prewhitening is needed.
4. Choose a PSD kernel.
5. Select a bandwidth by a transparent rule or plug-in method.
6. Compute the robust covariance matrix and the test statistic.
7. Report robustness to nearby bandwidth choices.

What should never be hidden

- the kernel,
- the bandwidth,
- whether prewhitening was used,
- and whether the covariance estimator is guaranteed to be PSD.

What can go wrong in finite samples?

Even consistent HAC estimators can perform poorly in moderate samples:

- size distortions under persistent dependence;
- oversensitivity to the bandwidth;
- noisy high-lag sample autocovariances;
- non-PSD estimates if the kernel/window is chosen badly;
- poor power when bandwidths are chosen conservatively.

Important warning

For inference, we care about *test size and coverage*, not just asymptotic MSE of the covariance estimator.

This explains why the later HAR-inference literature moved beyond classical HAC to fixed-b, self-normalization, bootstrap methods, and alternative critical values.

Transition to Lecture 14: fixed-b and self-normalization

Classical HAC asymptotics send the bandwidth to zero relative to sample size:

$$m \to \infty, \qquad m/T \to 0.$$

But in finite samples, the bandwidth is a fixed non-negligible number.

Next lecture

Lecture 14 asks: what happens if we treat the bandwidth ratio

$$b = \frac{m}{T}$$

as fixed rather than asymptotically negligible?

- fixed-b asymptotics use nonstandard but more accurate critical values;
- self-normalization avoids direct consistent LRV estimation;
- both approaches target better finite-sample inference under dependence.

Summary of Lecture 13

- The periodogram decomposes sample variance by frequency.
- The long-run variance is 2π times the spectrum at zero.
- HAC estimators are smoothed low-frequency spectral estimators.
- The same estimator can be written as a weighted sum of autocovariances or a weighted average of periodogram ordinates.
- Positive semi-definiteness is not automatic; nonnegative spectral windows help guarantee it.
- Kernels matter, but bandwidth matters even more.
- Good robust inference is about more than consistency: finite-sample size and admissibility matter too.

One-sentence takeaway

HAC inference is where spectral analysis becomes operational econometrics.