Lecture 14 — Fixed-*b* Asymptotics, Self-Normalization, and Robust Inference Beyond HAC

Chapter 6: advanced inference for dependent data

Jiajing Sun

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

Econometrics and Time Series Methods
Spring 2026

Why Lecture 14 matters

Lecture 13 established the classical HAC logic: estimate the long-run variance consistently and then rely on asymptotic normal or chi-squared critical values.

$$T_{\mathrm{HAC}} = \frac{\sqrt{T}(\widehat{\theta} - \theta_0)}{\widehat{\Omega}_{\mathrm{HAC}}^{1/2}}$$

is the familiar benchmark statistic behind that workflow.

Lecture 14 asks the next question

What should we do when the HAC estimator is asymptotically valid, but finite-sample inference is still unreliable because dependence is strong and bandwidth choice matters?

- Fixed-b keeps the bandwidth ratio visible in first-order asymptotics.
- Self-normalization avoids explicit long-run variance estimation.
- The goal is better size control and more honest uncertainty quantification.

Theme of the lecture

The target is not a prettier long-run variance estimator. The target is a more credible rejection rule and a more honest confidence interval.

Where Lecture 14 sits in the course

1. Lecture 13: spectrum, periodogram, HAC kernels, and bandwidth choice.
2. **Lecture 14:** advanced inference when classical HAC is not enough.
3. Lecture 15: bootstrap, GMM, and the bridge from robust inference to filtering.

Teaching logic

We are not abandoning HAC. We are learning what to do when "consistent" is not the same thing as "useful for finite-sample inference".

Review versus new material

Lecture 14 briefly reviews the HAC denominator because fixed-b and self-normalization are easiest to understand as two different responses to the same finite-sample problem.

Learning goals

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

1. explain why consistent HAC inference can still have poor size;
2. distinguish small-b and fixed-b asymptotics;
3. interpret the Brownian-motion and Brownian-bridge objects behind fixed-b theory;
4. explain the logic of self-normalization from Student's t onward;
5. define Shao's self-normalizer for the mean and interpret its role;
6. compare HAC, fixed-b, and self-normalization in practice;
7. implement a basic robustness workflow in R.

Three-hour plan

Hour 1

Fixed-b asymptotics: why the bandwidth ratio should remain visible in the limiting experiment.

Hour 2

Self-normalization: pivotal statistics without direct long-run variance estimation.

Hour 3

R block: HAC, fixed-b, and self-normalization in practice.

Quick recap: the classical HAC workflow

For a mean, regression coefficient, or smooth moment estimator, classical HAR inference starts from

$$\sqrt{T}(\widehat{\theta} - \theta_0) \Longrightarrow N(0, \Omega).$$

The corresponding long-run variance estimator is usually written as

$$\widehat{\Omega}_{\mathrm{HAC}} = \sum_{j=-(T-1)}^{T-1} k(j/m)\widehat{\gamma}(j), \qquad \widehat{\gamma}(j) = T^{-1} \sum_{t=|j|+1}^{T} \widehat{v}_t \widehat{v}'_{t-|j|}.$$

Then we:

1. estimate Ω by a HAC long-run variance estimator;
2. plug $\widehat{\Omega}_{\mathrm{HAC}}$ into a t, Wald, or J-type statistic;
3. use standard critical values from a Gaussian or chi-squared limit.

Classical asymptotic regime

$$m \to \infty, \qquad \frac{m}{T} \to 0.$$

Why consistent HAC can still disappoint

- The researcher must choose a kernel and a bandwidth.
- In realistic samples, different reasonable bandwidths can imply different standard errors.
- Strong persistence amplifies the effect of these choices.
- Size distortion can remain substantial even if $\widehat{\Omega}_{\mathrm{HAC}}$ is consistent.

What goes wrong mechanically

If $\widehat{\Omega}_{\mathrm{HAC}}$ is biased downward in the relevant sample size, then the denominator is too small, the t-statistic is too large in absolute value, and the null is rejected too often.

Main lesson

The object we care about is not just the MSE of the covariance estimator. We care about the rejection probabilities and coverage rates of the final inferential procedure.

Inferential accuracy is not covariance-estimation accuracy

Suppose two covariance estimators have similar asymptotic properties.

- The first may slightly underestimate low-frequency dependence and over-reject.
- The second may be noisier but yield more accurate finite-sample size.

$$P_{\theta_0}(|T_{\text{test}}| > c_\alpha) \neq \alpha \quad \text{can happen even when} \quad \widehat{\Omega} \xrightarrow{p} \Omega.$$

Why this matters

An estimator that looks "good" from a covariance-estimation perspective need not be the best denominator for a hypothesis test.

- Fixed-b responds by changing the asymptotic approximation.
- Self-normalization responds by changing the denominator entirely.

Why the bandwidth ratio matters

Write the truncation lag as m, sample size as T, and bandwidth ratio as

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

- In actual empirical work, b is never literally zero.
- Analysts often choose m as a visible fraction of the sample length.
- The resulting statistic therefore depends materially on b, not just on T.

Empirical interpretation

When a paper says "we used 12 lags in a sample of length 80," it is already making a fixed-b-style statement, because $b = 12/80$ is economically non-negligible.

Fixed-b insight

If b matters in practice, it should matter in the first-order asymptotic approximation.

Small-b asymptotics in one line

The standard asymptotic story is:

$$\widehat{\Omega}_{\mathrm{HAC}} \xrightarrow{p} \Omega, \qquad \frac{\sqrt{T}(\widehat{\theta} - \theta_0)}{\widehat{\Omega}_{\mathrm{HAC}}^{1/2}} \Longrightarrow N(0, 1).$$

- This is elegant and convenient.
- But it treats the bandwidth effect as asymptotically negligible.
- In finite samples, that negligence can be costly.

Hidden simplification

Small-b effectively says that all tuning uncertainty washes out faster than the central-limit effect. Lecture 14 asks what happens if that simplification is too optimistic.

Why small-b can understate uncertainty

What small-b assumes

The bandwidth grows slowly enough that its randomness and first-order effect vanish.

What data often show

With moderate T, the bandwidth choice strongly affects the denominator and therefore the rejection decision.

Interpretation

Fixed-b is not a second-order refinement. It is a different first-order approximation designed to be closer to empirical practice.

Finite-sample mechanism

When the bandwidth materially affects the denominator, pretending that m/T is asymptotically invisible tends to make the usual normal reference law too sharp relative to the actual sampling distribution.

$$\text{classical rule: } c_\alpha = z_{\alpha/2} \qquad \Longrightarrow \qquad \text{fixed-}b \text{ rule: } c_\alpha = c_\alpha(b, k).$$

The fixed-b idea

Under fixed-b asymptotics, we let

$$\frac{m}{T} \to b \in (0, 1].$$

- The HAC estimator no longer converges to a deterministic constant in the usual way.
- Instead, it converges to a random matrix that is stochastically proportional to the true long-run variance.
- The resulting t- and Wald-type statistics have nonstandard reference laws indexed by b.

$$T_b \Longrightarrow \mathcal{L}_b \qquad \text{with} \qquad \mathcal{L}_b \neq N(0, 1).$$

Big change

Standard normal or chi-squared critical values are generally no longer correct.

Regression notation for fixed-b inference

Consider a linear regression with score process

$$v_t = X_t u_t, \qquad S_{\lfloor rT \rfloor} = \sum_{t=1}^{\lfloor rT \rfloor} v_t.$$

Then

$$\sqrt{T}(\widehat{\beta} - \beta_0) = \left(T^{-1} \sum_{t=1}^{T} X_t X_t' \right)^{-1} T^{-1/2} S_T.$$

Write

$$Q = \text{plim } T^{-1} \sum_{t=1}^{T} X_t X_t'.$$

- The numerator is still driven by the partial-sum process.
- The entire inferential issue lies in how we scale that numerator.

$$T_{\text{fixed-}b} = \frac{\sqrt{T}(\widehat{\beta} - \beta_0)}{\left[\widehat{\Omega}_v^{(b)}\right]^{1/2}}$$

is therefore the direct analog of the classical HAC t-statistic, except that its limit law is no longer Gaussian.

Partial-sum limit and Brownian motion

For the scalar benchmark, the functional CLT takes the form

$$T^{-1/2} S_{\lfloor rT \rfloor} \Longrightarrow \sigma B(r), \qquad \sigma^2 = \sum_{j=-\infty}^{\infty} \gamma(j).$$

- $B(r)$ is standard Brownian motion.
- σ^2 is the nuisance long-run variance.
- Fixed-b theory asks whether the denominator can be built so that σ cancels out.

How to read r

The index $r \in [0, 1]$ is rescaled sample time. It lets us study the entire cumulative path, not just its endpoint at $r = 1$.

Why functional limits matter

The denominator depends on the whole low-frequency path of the partial sums, not just on the terminal value S_T. That is why a process limit is more informative than an ordinary CLT here.

Brownian Background

Core definitions

A standard Brownian motion $B(r)$, or Wiener process, satisfies:

- $B(0) = 0$;
- $B(r) - B(s) \sim N(0, r - s)$ for $0 \leq s < r$;
- disjoint increments are independent;

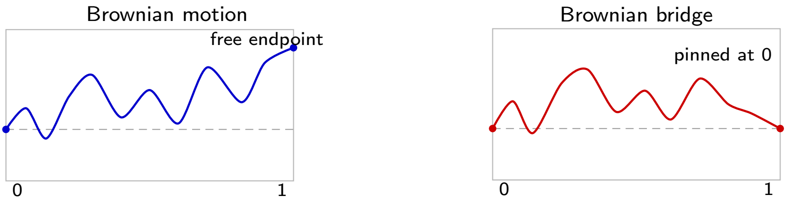
$$\mathbb{B}(r) = B(r) - rB(1)$$

is the associated Brownian bridge on $[0, 1]$.

Why it shows up here

- Brownian motion is the FCLT limit of standardized partial sums.
- Robert Brown supplied the name; Wiener supplied the modern construction.
- Brownian bridges appear after endpoint demeaning or residualization.
- Paths are continuous but not differentiable because short-run increments are order $\sqrt{\Delta}$, not order Δ.
- So dB_t denotes an increment, and $\int H_t \, dB_t$ is built from increment sums rather than from a derivative.

Motion Versus Bridge



- Brownian motion has a free endpoint; the bridge is tied down at zero at both ends.
- Both paths are continuous and jagged, so a pathwise derivative does not exist.
- Fixed-b denominators often involve the bridge because demeaning removes the endpoint component $rB(1)$.

The HAC estimator under fixed-b

With a kernel $k(\cdot)$ and bandwidth m, we write

$$\widehat{\Omega}_v^{(b)} = \sum_{j=-(T-1)}^{T-1} k(j/m)\widehat{\gamma}(j).$$

- Under small-b, this is consistent for Ω_v.
- Under fixed-b, it converges to a random Brownian functional.
- For the Bartlett kernel, that random limit can be expressed using Brownian bridges.

Key distinction

So the same formula is being interpreted under two different asymptotic regimes. The estimator has not changed, but the limit experiment has.

Brownian bridge representation

For Bartlett-type fixed-b asymptotics, the limiting matrix is of the form

$$Q_q(b) = \frac{2}{b}\int_0^1 \mathbb{B}_q(s)\mathbb{B}_q(s)' ds - \frac{1}{b}\int_0^{1-b} \{\mathbb{B}_q(s+b)\mathbb{B}_q(s)' + \cdots\} ds,$$

where $\mathbb{B}_q$ is a Brownian bridge.

Why Brownian bridges appear

Residualization and demeaning remove the endpoint component, leaving a bridge rather than an unrestricted Brownian motion.

$$\mathbb{B}(r) = B(r) - rB(1).$$

Intuition

The bridge records the interior wandering of the cumulative process after subtracting the straight line connecting the endpoints. That is exactly the fluctuation the HAC denominator is trying to summarize.

Why the fixed-b limit depends on b

Under the functional limit theory,

$$\widehat{\Omega}_v^{(b)} \Longrightarrow \Lambda Q_q(b)\Lambda'.$$

The same Λ appears in the numerator limit:

$$\sqrt{T}(\widehat{\beta} - \beta_0) \Longrightarrow Q^{-1}\Lambda B_q(1).$$

- The long-run covariance matrix is absorbed into the Brownian-function representation.
- But the reference law still depends on the chosen bandwidth ratio b and on the kernel choice.
- So fixed-b critical values are indexed by b; they are not universal Gaussian critical values.

Practical payoff

We do *not* gain a universal nuisance-free law. The payoff is that keeping b visible yields a more accurate first-order approximation and supports informative local power analysis for HAC-robust tests.

Fixed-b t statistics

Instead of converging to $N(0,1)$, a fixed-b t-type statistic converges to a nonstandard b-indexed distribution, often represented through Brownian-function ratios such as

$$\frac{B(1)}{\left(\int_0^1 \mathbb{B}(s)^2 ds\right)^{1/2}}$$

or a closely related b-dependent Brownian functional.

- Symmetry remains.
- Tails are typically heavier than Gaussian tails.
- Critical values depend on b and must be tabulated or simulated for the chosen kernel and bandwidth ratio.

Classroom shortcut

You can think of fixed-b as saying: the denominator is still random at first order, so the Gaussian benchmark is too optimistic.

Fixed-b confidence intervals

The practical implication is straightforward:

1. compute the fixed-b robust statistic;
2. use the appropriate nonstandard critical value for the chosen b;
3. invert the test to obtain a confidence interval.

$$\widehat{\theta} \pm \frac{c_{1-\alpha/2}(b)}{\sqrt{T}} \cdot \widehat{s}_{\text{fixed}-b}.$$

Interpretation

The interval is often wider than the classical HAC interval, but the gain is more reliable finite-sample coverage.

Inversion view

This interval is obtained by inverting all null values θ_0 that would not be rejected by the fixed-b test. The critical value changes because the null experiment changes.

Size-power trade-off

Potential gain

Fixed-b often improves empirical size because it takes bandwidth uncertainty seriously.

Potential cost

The same conservatism can reduce power under local alternatives.

Empirical message

Advanced inference is not about maximizing rejection rates. It is about making rejection decisions more trustworthy.

$$\text{better size control} \quad \Longleftrightarrow \quad \text{often weaker local power.}$$

Local-alternative intuition

Because the first-order limit keeps b fixed, fixed-b provides a direct framework for local power analysis under the chosen bandwidth ratio. Heavier critical values typically imply weaker local power.

Choosing and reporting b in practice

- A fixed-b analysis still needs a bandwidth ratio or truncation rule.
- The researcher should explain whether b was chosen by rule-of-thumb, plug-in logic, or sensitivity analysis.
- It is good practice to show at least two plausible values of b.

Good reporting template

"We report Bartlett fixed-b confidence intervals for $b \in \{0.10, 0.20\}$ and compare them with the classical HAC interval."

Reporting rule

Never report a fixed-b result without stating the kernel, the implied $b = m/T$, and the source of critical values.

Useful sensitivity display

A simple table with $b \in \{0.10, 0.15, 0.20\}$ is often more informative than a single reported standard error because it shows whether the inferential conclusion is structurally stable.

When fixed-b is especially useful

- Moderate sample size with visible dependence.
- Inference that is obviously sensitive to bandwidth choices.
- Applied work where the researcher wants a first-order approximation closer to actual implementation.
- Situations where local power analysis for HAC-robust tests is substantively useful.
- HAC settings where size control matters more than squeezing out local power.

Less compelling case

If T is very large and the conclusion is already stable across bandwidth choices, the practical gain from fixed-b may be smaller.

Rule for applied judgment

If two plausible bandwidths give noticeably different t-statistics or p-values, that is already evidence that the small-b approximation may be too optimistic for the sample at hand.

$$\text{visible tuning sensitivity} \quad \Longrightarrow \quad \text{fixed-}b \text{ deserves a place in the robustness table.}$$

Student's t and the Old Idea

Self-normalization is not a new trick. Its oldest prototype is Student's t-statistic:

$$\frac{\sqrt{T}(\bar{X}-\mu)}{S} \sim t_{T-1}, \qquad S^2 = \frac{1}{T-1}\sum_{t=1}^{T}(X_t-\bar{X})^2.$$

- In small samples, S^2 is random and is not the population variance itself.
- But S is stochastically proportional to the unknown scale σ.
- That is enough to self-normalize the numerator and produce an exact finite-sample pivot in the IID Gaussian case.

$$t_{T-1} \Longrightarrow N(0,1) \qquad \text{as} \qquad T \to \infty.$$

Main teaching point

The denominator does not need to equal the true variance. It only needs to track the unknown scale in the right stochastic way.

From Student's t to Time-Series Self-Normalization

Fixed-b still begins with a HAC-style estimator. Self-normalization goes one step further:

Time-series version of the same idea

Instead of estimating the long-run variance directly, construct a random denominator that is automatically proportional to the unknown scale.

- In time series, the nuisance scale is the long-run variance rather than the IID variance.
- Historically, this line of work was relatively underdeveloped because so much effort went into HAC long-run variance estimation.
- Self-normalization revives the Student-t logic in a dependent-data setting.

$$\text{replace } \widehat{\Omega}^{1/2} \text{ by } \widehat{D}_T \text{ such that } \widehat{D}_T \asymp \Omega^{1/2} \text{ stochastically.}$$

Conceptual contrast

Fixed-b says "keep the classical denominator but change the asymptotic reference law." Self-normalization says "change the denominator itself so that nuisance scale cancels pathwise."

What self-normalization is trying to do

We want a denominator that:

1. is built from the same data as the numerator;
2. tracks the stochastic size of the numerator under dependence;
3. yields a limit law free of nuisance scale parameters.

Difference from HAC

HAC aims to estimate the long-run variance accurately. Self-normalization aims to construct a pivotal ratio even when the denominator is not a consistent estimator of that variance.

$$\frac{\sqrt{T}(\widehat{\theta} - \theta_0)}{\widehat{D}_T} \Longrightarrow \mathcal{L}_{\mathrm{SN}}, \qquad \mathcal{L}_{\mathrm{SN}} \text{ pivotal.}$$

Three-object summary

The numerator still measures estimation error, the denominator now measures path instability rather than long-run covariance directly, and the limit law absorbs the remaining randomness into a non-Gaussian but nuisance-free reference distribution.

Shao's self-normalizer for the mean

Let

$$\bar{X}_j = \frac{1}{j}\sum_{i=1}^{j} X_i, \qquad \bar{X}_T = \frac{1}{T}\sum_{i=1}^{T} X_i.$$

Define

$$W_T^2 = T^{-2}\sum_{j=1}^{T} j^2(\bar{X}_j - \bar{X}_T)^2.$$

$$W_T^2 \Longrightarrow \sigma^2 \int_0^1 \mathbb{B}(s)^2\, ds, \qquad \mathbb{B}(s) = B(s) - sB(1).$$

$$\mathrm{S}_T = \frac{\sqrt{T}(\bar{X}_T - \mu)}{W_T} \Longrightarrow \frac{B(1)}{\left[\int_0^1 \mathbb{B}(s)^2 ds\right]^{1/2}}.$$

- W_T^2 is not a consistent long-run variance estimator.
- It is stochastically proportional to the long-run variance σ^2.
- No kernel or bandwidth enters the construction.

Why recursive means appear

If cumulative averages fluctuate a lot before settling near $\bar{X}_T$, the underlying long-run uncertainty is large. The normalizer turns that path variation into a denominator.

Shao's critical values for the mean

For the scalar statistic S_T, the textbook reports the following upper critical values for the *right tail*:

α	5.0%	2.5%	1.0%	0.5%	0.1%
c_α^S	5.3210	6.7805	8.5458	9.7029	13.0598

How to use the table

For a two-sided $100(1-\alpha)\%$ confidence interval, use the upper $\alpha/2$ critical value because the limit law of S_T is symmetric around zero.

Interpretation

These values are much larger than Gaussian critical values because the denominator remains random in the limit. That is the price of the self-normalization logic.

Influence Function and Approximately Linear Statistics

Following the textbook, let $\mathsf{Y}_t = (X_t, \ldots, X_{t+m-1})'$ and $\theta_0 = \mathsf{T}(\mathsf{F}^m)$.

Influence function

$$\mathsf{IF}(\mathsf{y}; \mathsf{F}^m) = \lim_{\epsilon \downarrow 0} \frac{\mathsf{T}\{(1-\epsilon)\mathsf{F}^m + \epsilon\delta_{\mathsf{y}}\} - \mathsf{T}(\mathsf{F}^m)}{\epsilon}$$

Approximately linear statistic

$$\mathsf{T}(\rho_{1,N}) - \mathsf{T}(\mathsf{F}^m) = N^{-1}\sum_{t=1}^{N} \mathsf{IF}(\mathsf{Y}_t; \mathsf{F}^m) + \mathsf{R}_{1,N}, \qquad \mathsf{R}_{1,N} = o_p(N^{-1/2})$$

- The influence function gives the first-order effect of a small contamination at y.
- Once this expansion holds, self-normalization is applied to the influence-function or score path.
- Examples: means, autocovariances, quantiles, and regression estimators.

Kiefer's original construction

For regression, define partial sums of residual scores and then set

$$\widehat{C} = T^{-2} \sum_{t=1}^{T} \widehat{S}_t \widehat{S}_t'.$$

$$\widehat{S}_t = \sum_{j=1}^{t} \widehat{v}_j, \qquad \widehat{v}_j = X_j \widehat{u}_j.$$

Then construct a denominator from $\widehat{C}^{1/2}$.

- The resulting transformed statistic converges to a Brownian functional.
- The nuisance covariance matrix again drops out.

$$T_{\text{Kiefer}} = T\,(\widehat{\beta} - \beta_0)'\,\widehat{C}^{-1}(\widehat{\beta} - \beta_0)$$

is the corresponding Wald-style object.

Brownian motion and Brownian bridge

Why does the theory keep returning to Brownian bridges?

- The numerator reflects the full partial-sum process.
- The denominator uses de-meaned or residualized cumulative paths.
- Demeaning removes the endpoint component and produces a bridge.

Useful fact

For Gaussian processes, the terminal Brownian motion value and the corresponding Brownian bridge are independent.

$$B(r) = \mathbb{B}(r) + rB(1).$$

Why that independence is useful

It separates the endpoint uncertainty that drives the numerator from the interior path fluctuation that drives the self-normalizing denominator.

Shao's general self-normalized Wald statistic

For an approximately linear q-vector $\widehat{\theta}_N$, define

$$W_N^2 = N^{-2} \sum_{t=1}^{N} t^2 (\widehat{\theta}_t - \widehat{\theta}_N)(\widehat{\theta}_t - \widehat{\theta}_N)'.$$

Then

$$N(\widehat{\theta}_N - \theta)'(W_N^2)^{-1}(\widehat{\theta}_N - \theta) \Longrightarrow U_q,$$

where

$$U_q = \mathrm{B}_q(1)' V_q^{-1} \mathrm{B}_q(1), \qquad V_q = \int_0^1 \mathbb{B}_q(s)\mathbb{B}_q(s)'\, ds.$$

$\alpha \backslash q$	1	2	3	5
10%	28.26	70.65	127.71	280.45
5%	46.15	105.10	177.95	358.88
1%	99.62	197.44	319.76	555.00

Why these are so large

This is a quadratic-form statistic, not the scalar S_T statistic. So its critical values live on a completely different scale.

A self-normalized t-type statistic

For the mean, the self-normalized statistic is

$$S_T = \frac{\sqrt{T}(\bar{X}_T - \mu)}{W_T}.$$

- No kernel is chosen.
- No bandwidth is chosen.
- The denominator is entirely path-based.

$$S_T \Longrightarrow \frac{B(1)}{\left(\int_0^1 \mathbb{B}(r)^2 dr\right)^{1/2}} \quad \text{under standard weak-dependence conditions.}$$

Trade-off

Self-normalized procedures are often robust and size-stable, but they can be conservative.

Adjusted-range self-normalization

For the mean problem, Hong et al. replace W_T by the adjusted range

$$R_T = \max_{1 \le k \le T} T^{-1/2} \sum_{t=1}^{k} (X_t - \bar{X}_T) - \min_{1 \le k \le T} T^{-1/2} \sum_{t=1}^{k} (X_t - \bar{X}_T).$$

$$\mathrm{M}_T = \frac{\sqrt{T}(\bar{X}_T - \mu)}{R_T} \Longrightarrow \frac{B(1)}{\sup_{s \in [0,1]} \mathbb{B}(s) - \inf_{s \in [0,1]} \mathbb{B}(s)}.$$

- The denominator is the peak-to-trough span of the centered partial-sum path.
- The method keeps the nuisance-free logic of self-normalization.
- The goal is a better size-power trade-off than the earlier quadratic-path normalizer.
- Range functionals can be more robust to persistence, outliers, and heavy-tailed behavior.

Why range helps

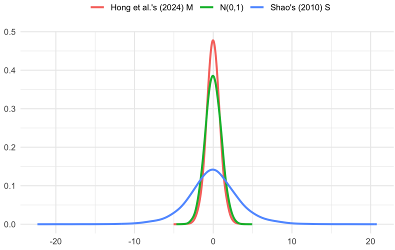
Range-based normalizers react to path excursions directly, so they can be less sensitive to the quadratic weighting built into earlier self-normalizers.

Mandelbrot connection

Mandelbrot used range-based statistics in highly non-Gaussian and long-range dependent settings, which is one reason the range has a natural robustness interpretation here.

Selected scalar critical values: adjusted range versus Shao

α	c_α^M	c_α^S
5.0%	1.4100	5.3210
2.5%	1.7171	6.7805
1.0%	2.0682	8.5458
0.5%	2.3535	9.7029
0.1%	3.0706	13.0598



Reading the table

The adjusted-range statistic M_T is much closer to normal critical values, whereas Shao's S_T is far more conservative in the tails.

Reading guide

The figure and the table tell the same story: M_T stays closer to the Gaussian benchmark, while S_T has much heavier tails.

Adjusted-range for approximately linear statistics

For a scalar approximately linear statistic, the adjusted-range normalizer becomes

$$R_T = \max_{1\le k\le T} T^{-1/2}\sum_{j=1}^{k} j(\widehat{\theta}_{1,j}-\widehat{\theta}_{1,T}) - \min_{1\le k\le T} T^{-1/2}\sum_{j=1}^{k} j(\widehat{\theta}_{1,j}-\widehat{\theta}_{1,T}).$$

For $q \geq 2$, after partial prewhitening one obtains

$$N(\widehat{\theta}_{1,N}-\theta_0)'(V_N^R)^{-1}(\widehat{\theta}_{1,N}-\theta_0) \Longrightarrow \mathrm{M}_q^2,$$

$$\mathrm{M}_q^2 = \mathrm{B}_q(1)'\mathrm{R}_q^{-2}\mathrm{B}_q(1), \qquad \mathrm{R}_q = \mathrm{diag}[\sup \mathbb{B}_q - \inf \mathbb{B}_q].$$

$\alpha\backslash q$	1	2	3	5
10%	2.08	3.48	4.72	7.00
5%	3.11	4.67	6.22	8.55
1%	5.62	7.72	9.63	12.43

What changes in higher dimensions

Cross-dependence matters, so the textbook first partially prewhitens the recursive path and then applies the range normalizer component by component.

When Shao-style self-normalization can lose power

- The quadratic recursive-path denominator can fluctuate substantially in moderate samples.
- That extra randomness can widen intervals and reduce rejection probability under local alternatives.
- This is why Shao's self-normalizer is often more conservative than a well-tuned HAC, fixed-b, or adjusted-range procedure.

Interpretation

This is the same logic as Student's t: greater robustness to scale uncertainty usually means heavier tails in the reference distribution.

Important qualification

Adjusted-range self-normalization was proposed to keep the same philosophy while delivering a more balanced size-power trade-off.

Good empirical attitude

Treat Shao-style self-normalization as a robustness device, not as a universal replacement for every other inferential method.

HAC, fixed-b, and self-normalization variants compared

Method	Main input	Main strength	Main risk
HAC	kernel + bandwidth	standard and familiar	size sensitivity to tuning
Fixed-b	kernel + fixed b	better first-order size approximation; local power framework	b-indexed critical values
Shao SN	quadratic path denominator	no direct LRV estimation; strong robustness	often conservative
Adjusted-range SN	range-based path denominator	tuning-light, more balanced, and often more robust to persistence, outliers, and heavy tails	newer method; still needs non-Gaussian critical values

How to read the table

Move from left to right: what the method asks you to choose, what it buys you, and where it is most likely to disappoint.

Important distinction

Not all self-normalization variants behave the same way. Shao's quadratic normalizer is the more conservative benchmark, while adjusted-range self-normalization is designed to be closer to the Gaussian benchmark and more balanced.

A practical decision guide

1. Start with a classical HAC result for orientation.
2. If the result changes sharply with bandwidth, add a fixed-b analysis.
3. If dependence is persistent or the sample is short, add a self-normalized check.
4. Compare decisions across at least two robust procedures.

Suggested classroom workflow

Ask first whether the point estimate is stable, then whether the standard error is stable, and only then whether the rejection decision is stable.

Rule of thumb

Robust inference is most convincing when the empirical conclusion survives more than one reasonable denominator.

HAC fragile $\Rightarrow$ add fixed-b $\Rightarrow$ if still doubtful, add self-normalization.

What an empirical paper should report

- Which statistic is being tested?
- Which inferential device is used: HAC, fixed-b, or self-normalization?
- If HAC or fixed-b is used, what kernel and bandwidth rule were chosen?
- If self-normalization is used, which normalizer and which critical values were used?
- Are the main conclusions robust across at least two procedures?

Replication mindset

Another researcher should be able to reconstruct the reported p-value from the paper's description alone.

Minimum reproducibility set

Statistic, kernel, bandwidth or b, critical-value source, and robustness comparison should all be explicit. Otherwise the reader cannot tell whether "robust" means thoughtfully stress-tested or merely software-default.

Common mistakes in advanced inference

- Reporting only one robust standard error without explaining how it was built.
- Treating bandwidth choice as if it were innocuous.
- Mixing classical HAC standard errors with fixed-b critical values.
- Calling a self-normalized denominator a "variance estimator" in the usual sense.
- Over-interpreting one method's significance when other robust methods disagree.

Meta-mistake

The biggest mistake is to present robust inference as if it produced a unique, tuning-free truth. It almost never does.

R block roadmap

The implementation block has four goals:

1. compute a baseline HAC result;
2. vary the bandwidth and observe sensitivity;
3. approximate a fixed-b robustness check;
4. compute a self-normalized statistic for the same parameter.

Interpretive question

At each stage ask: did the statistic change, or did only the reference distribution change?

Pedagogical goal

Students should leave the room understanding not only the formulas, but also the workflow of robust empirical practice.

R block: simulate dependent data

We begin with a persistent AR(1) sample:

```
set.seed(123)
T <- 250
phi <- 0.75
u <- rnorm(T)
x <- numeric(T)
for (t in 2:T) x[t] <- phi * x[t - 1] + u[t]
```

- Persistence is high enough to make long-run variance estimation nontrivial.
- The sample is long enough to be realistic, but not long enough for asymptotics to solve everything.
- The true long-run variance for an AR(1) mean is $\sigma_u^2/(1-\phi)^2$, so $\phi = 0.75$ already creates a sizable low-frequency amplification.

R block: classical HAC

A baseline HAC analysis might look like:

```
theta_hat <- mean(x)
lrvar_hat <- sandwich::lrvar(x)
se_hac <- sqrt(lrvar_hat / T)
t_hac <- sqrt(T) * theta_hat / sqrt(lrvar_hat)
```

- This is the familiar starting point.
- It should never be the automatic ending point.
- The quantity $lrvar\_hat/T$ is the estimated variance of the sample mean.

Interpretive reminder

At this stage the point estimate has not changed at all. Only the denominator and therefore the uncertainty statement are being built.

R block: bandwidth sensitivity

To diagnose fragility, vary the bandwidth:

```
bw_grid <- c(4, 8, 12, 20, 30)
se_grid <- sapply(bw_grid, function(m) {
  sqrt(sandwich::NeweyWest(
    lm(x ~ 1), lag = m, prewhite = FALSE
  )[1, 1])
})
```

Interpretation

If significance flips across plausible bandwidths, the empirical conclusion is fragile and deserves a fixed-b or self-normalized check.

$$\text{same } \widehat{\theta}, \text{ different } \widehat{\Omega}_{\mathrm{HAC}} \quad \Rightarrow \quad \text{different inference.}$$

R block: a fixed-b robustness check

One practical way to think in fixed-b terms is:

```
b <- 0.20
m <- floor(b * T)
vcov_b <- sandwich::NeweyWest(
  lm(x ~ 1), lag = m, prewhite = FALSE
)
```

- The point is not that ‘NeweyWest‘ magically becomes fixed-b.
- The point is that the analyst holds m/T visible and then interprets inference through the fixed-b lens rather than the small-b lens.
- In practice, the missing ingredient is the correct fixed-b critical value or p-value approximation.

Good empirical habit

Store both m and $b = m/T$. The lag count alone is not enough once the sample length changes across specifications.

R block: self-normalized statistic

For the sample mean, a direct implementation is:

```
j <- seq_len(T)
cum_mean <- cumsum(x) / j
W2 <- sum((j * (cum_mean - mean(x)))^2) / T^2
sn_stat <- sqrt(T) * mean(x) / sqrt(W2)
```

- No kernel and no bandwidth enter.
- Critical values must come from the self-normalized reference law, not from $N(0, 1)$.
- The code makes the construction transparent: the denominator comes from recursive-mean instability.

What the code is measuring

If the cumulative means settle quickly, W_T is small. If they drift for a long time, W_T is larger and inference becomes more cautious.

R block: comparison table

A useful output table is:

```
data.frame(
  method = c("HAC", "fixed-b style", "self-normalized"),
  statistic = c(t_hac, NA_real_, sn_stat),
  note = c("normal critical values",
           "needs fixed-b critical values",
           "needs SN critical values")
)
```

Good practice

Store not only a statistic, but also the reference law required to interpret it.

Even better practice

Add the bandwidth, $b = m/T$, and the critical-value source as extra columns so that the table becomes self-documenting.

Mini empirical example

Suppose the target is the mean excess return, inflation persistence, or a regression slope.

1. estimate the model once;
2. compute HAC, fixed-*b*-style, and self-normalized statistics;
3. ask whether the sign, magnitude, and significance conclusions agree.

What to look for

The most informative empirical finding is often not a single p-value, but the extent to which that p-value is robust across inferential devices.

Compact rule

Same sign + similar magnitude + stable significance across procedures implies stronger empirical credibility.

If methods disagree

Disagreement is not failure. It is information.

- It tells us the result depends on how dependence is handled.
- It signals that the sample may not support a sharp inferential claim.
- It can motivate bootstrap methods, which is exactly where Lecture 15 begins.

How to write this in a paper

"The coefficient remains positive, but statistical significance is sensitive to the robust inferential device, so the null-rejection claim should be interpreted cautiously."

Referee-style interpretation

If three reasonable procedures disagree, the honest conclusion is usually that the sample does not pin down the inferential claim sharply enough.

Discussion prompts for class

1. Why can a consistent HAC estimator still yield poor test size?
2. In what sense is fixed-b closer to empirical practice than small-b?
3. Why is self-normalization conceptually linked to Student's t?
4. Why might a more robust inferential procedure also have lower power?

Bonus prompt

If you were refereeing a paper, which robustness table would most increase your confidence in its main claim?

Exercises for after class

1. Reproduce a HAC t-statistic for an AR(1) mean and study sensitivity to bandwidth.
2. Compute Shao's self-normalized statistic for the same data.
3. Explain, in words, why fixed-b critical values depend on b.
4. Compare the inferential conclusion across the three procedures.

Extension

Repeat the same exercise after increasing persistence from $\phi = 0.75$ to $\phi = 0.90$ and describe how the inferential disagreement changes.

Lecture 14 takeaways

- Classical HAC is foundational, but not always enough.
- Fixed-b changes the asymptotic approximation so the bandwidth ratio matters at first order.
- Self-normalization changes the denominator so explicit long-run variance estimation is avoided.
- Robust inference is strongest when conclusions survive multiple reasonable procedures.

One-line summary

Lecture 13 asked how to estimate the long-run variance well; Lecture 14 asked how to infer well even when that estimation step is fragile.

Preview of Lecture 15

Lecture 15 continues the same inferential story from three new angles:

1. bootstrap methods for dependent data;
2. GMM with time-series moment conditions and a C-CAPM illustration;
3. filtering in the time and frequency domains as the bridge to state-space methods.