Lecture 19 — Diffusion Likelihood, Realized Volatility, and High-Frequency Covariance Estimation

Chapter 9: exact and approximate likelihood, microstructure noise, and covariance estimation

Jiajing Sun

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

Econometrics and Time Series Methods
Spring 2026

Why Lecture 19 matters

Lecture 19 completes the Chapter 9 arc by combining three themes:

- likelihood-based inference for discretely observed diffusion models;
- realized-volatility measurement from high-frequency data;
- variance–covariance estimation under microstructure noise and asynchronous trading.

Big message

Continuous-time models are only useful if we can estimate them from discrete data and if we can measure volatility and covariance robustly in noisy, high-frequency environments.

Where Lecture 19 fits in the course

- Lecture 18 introduced Brownian motion, stochastic differentials, and diffusion processes.
- **Lecture 19** studies how to estimate diffusion models and how to use high-frequency data to recover integrated variance and covariance.
- Lecture 20 will be review and problem class.

Teaching logic

We move from continuous-time probability to continuous-time inference. Theoretical objects such as transition densities, quadratic variation, and noise-robust covariance estimators now become empirical tools.

Learning goals

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

1. write the likelihood for discretely observed diffusion models;
2. derive exact MLE formulas for geometric Brownian motion;
3. explain why approximate likelihood is needed when transition densities are unavailable;
4. describe the Aït-Sahalia transformation-and-expansion strategy;
5. define quadratic variation, realized volatility, and integrated quarticity;
6. explain why naive realized volatility fails under microstructure noise;
7. write down the TSRV, RK, TSX, RKX, and PAVX constructions at a high level;
8. explain refresh-time synchronization and the Epps effect.

Three-hour plan

Hour 1

Likelihood and approximate likelihood for diffusion models.

Hour 2

R block: Brownian-path simulation and realized-volatility measurement from high-frequency data.

Hour 3

Volatility and covariance estimation under market microstructure noise, followed by course wrap-up.

From an SDE to a discrete-data likelihood

Suppose the latent continuous-time process satisfies

$$dX_t = \mu(X_t;\theta)\,dt + \sigma(X_t;\theta)\,dB_t.$$

We observe it only at discrete times

$$X_0, X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}.$$

Core estimation problem

How do we estimate θ when the model is continuous-time but the data are discrete-time?

Answer

Use the transition density

$$p_X(\Delta; x_i \mid x_{i-1}; \theta),$$

provided the process is Markov.

Likelihood for a discretely observed Markov diffusion

The Markov property implies

$$\mathbb{P}(X_{n\Delta}, \ldots, X_0; \theta) = \mathbb{P}(X_0; \theta) \prod_{i=1}^{n} p_X(\Delta; X_{i\Delta} \mid X_{(i-1)\Delta}; \theta).$$

Ignoring the initial-density term when appropriate, the log-likelihood is

$$\mathcal{L}_n(\theta) = \sum_{i=1}^{n} \log p_X(\Delta; X_{i\Delta} \mid X_{(i-1)\Delta}; \theta).$$

Interpretation

Inference for continuous-time models is therefore driven by the conditional transition law over the sampling interval Δ.

General MLE statement

The maximum likelihood estimator is

$$\widehat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, \mathcal{L}_n(\theta).$$

Under standard regularity conditions,

$$\sqrt{n}\big(\widehat{\theta} - \theta_0\big) \Rightarrow N\big(0, \mathcal{I}^{-1}(\theta_0)\big),$$

where

$$\mathcal{I}(\theta_0) = \lim_{n\to\infty} -\mathbb{E}\left[\frac{1}{n}\frac{\partial^2 \mathcal{L}_n(\theta_0)}{\partial\theta\,\partial\theta'}\right].$$

$$\operatorname{Var}(\widehat{\theta}) \approx \frac{1}{n}\mathcal{I}^{-1}(\theta_0).$$

Exact versus approximate transition densities

- If $p_X(\Delta; x \mid x_0; \theta)$ is known in closed form, exact MLE is feasible.
- If it is unavailable or intractable, we need an approximation.

Three benchmark cases

- GBM / Black–Scholes: exact lognormal transition.
- OU / Vasicek: exact Gaussian transition.
- CIR / square-root: exact noncentral chi-squared transition.

But most interesting diffusions are harder

For many nonlinear diffusions, the transition density is unknown, and direct MLE is not available in closed form.

Black–Scholes diffusion and log transformation

Consider

$$dX_t = \beta X_t\, dt + \sigma X_t\, dB_t.$$

Applying Itô's lemma to $\log X_t$ gives

$$d\log X_t = \left(\beta - \frac{\sigma^2}{2}\right) dt + \sigma\, dB_t = \alpha\, dt + \sigma\, dB_t,$$

where

$$\alpha = \beta - \frac{\sigma^2}{2}.$$

Payoff

Continuously compounded returns become Gaussian and i.i.d. over equal intervals.

Exact GBM transition density

For $\Delta > 0$,

$$\log X_{t+\Delta} \mid X_t = x \sim N\left(\log x + \left(\beta - \frac{\sigma^2}{2}\right)\Delta,\ \sigma^2\Delta\right).$$

Equivalently,

$$p_X(\Delta; x_1 \mid x_0; \beta, \sigma) = \frac{1}{x_1\sigma\sqrt{2\pi\Delta}} \exp\left(-\frac{\left[\log(x_1/x_0) - (\beta - \sigma^2/2)\Delta\right]^2}{2\sigma^2\Delta}\right).$$

Important point

This is an exact finite-Δ density, not merely a small-Δ approximation.

Exact GBM log-likelihood

Let

$$r_t(\Delta) = \log\left(\frac{X_t}{X_{t-\Delta}}\right).$$

Then

$$r_t(\Delta) \overset{i.i.d.}{\sim} N(\alpha\Delta, \sigma^2\Delta).$$

The log-likelihood is

$$\mathcal{L}_n(\alpha, \sigma) = -\frac{n}{2}\log(2\pi\sigma^2\Delta) - \frac{1}{2\sigma^2\Delta}\sum_{t=1}^{n}\left(r_t(\Delta) - \alpha\Delta\right)^2.$$

Interpretation

Once we transform to log returns, exact MLE becomes the familiar Gaussian likelihood problem.

Exact GBM MLE formulas

Maximizing the GBM likelihood gives

$$\widehat{\alpha} = \frac{1}{n\Delta}\sum_{t=1}^{n} r_t(\Delta), \qquad \widehat{\sigma}^2 = \frac{1}{n\Delta}\sum_{t=1}^{n}\big(r_t(\Delta) - \widehat{\alpha}\Delta\big)^2.$$

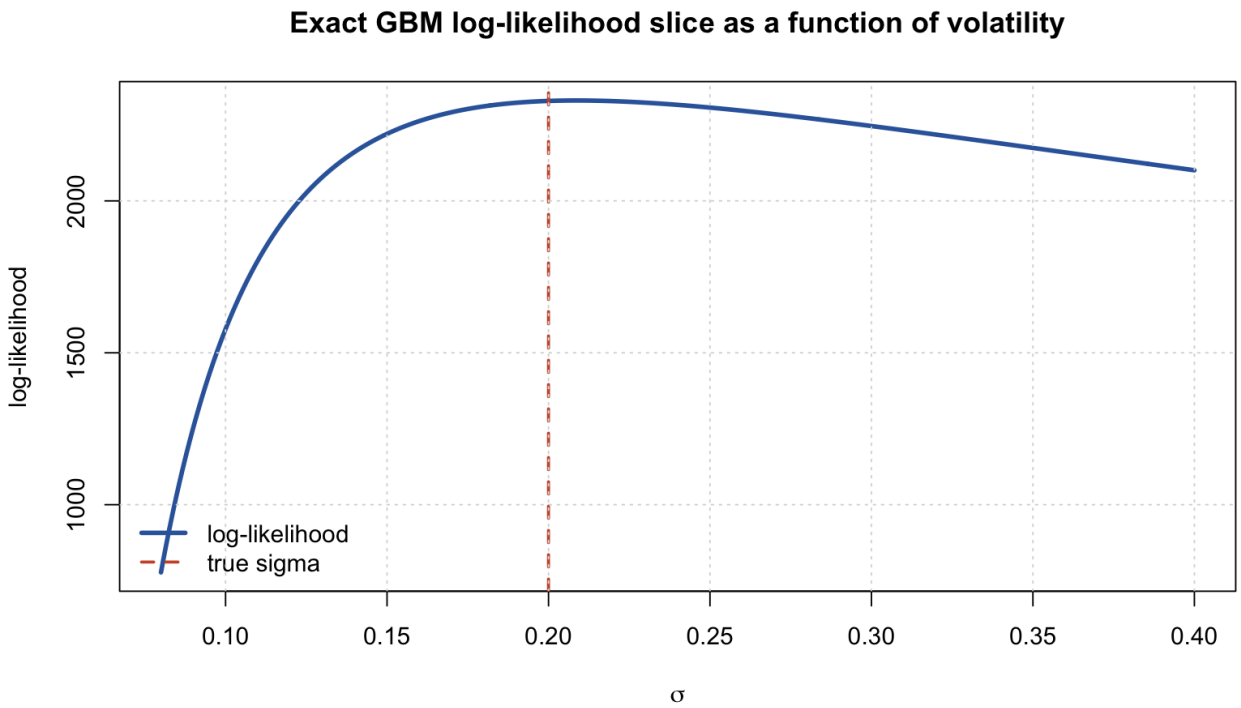
Recover the original drift by

$$\widehat{\beta} = \widehat{\alpha} + \frac{1}{2}\widehat{\sigma}^2.$$

Why this is useful

GBM is one of the rare cases where the exact finite-sample likelihood is simple enough to write and optimize directly.

Likelihood shape for GBM volatility



OU diffusion as another exact case

For the Ornstein–Uhlenbeck / Vasicek process

$$dX_t = \kappa(\alpha - X_t)\,dt + \sigma\,dB_t,$$

the exact conditional law is Gaussian:

$$X_{t+\Delta} \mid X_t = x \sim N\left(\alpha + e^{-\kappa\Delta}(x-\alpha),\ \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa\Delta}\right)\right).$$

Lesson

The OU model is analytically tractable because the drift is linear and the diffusion coefficient is constant.

CIR diffusion and noncentral chi-square law

For the square-root diffusion

$$dX_t = \kappa(\alpha - X_t)\,dt + \sigma\sqrt{X_t}\,dB_t,$$

the exact transition law is non-Gaussian but still known:

$$2cX_{t+\Delta} \mid X_t = x \sim \chi'^2_\nu(\lambda),$$

with

$$c = \frac{2\kappa}{\sigma^2(1 - e^{-\kappa\Delta})}, \quad \nu = \frac{4\kappa\alpha}{\sigma^2}, \quad \lambda = 2ce^{-\kappa\Delta}x.$$

Takeaway

Exact likelihood is possible for a few canonical models, but this tractability is exceptional.

Why approximate likelihood is needed

For many diffusion models:

- the transition density is unknown;
- the density exists but is not available in closed form;
- numerical solution of the Fokker–Planck equation is too expensive for repeated likelihood evaluation.

Problem

We still want likelihood-based inference, but we do not want to approximate the density badly when Δ is fixed and the process is non-Gaussian.

Textbook solution

Use the Aït-Sahalia transformation-and-expansion approach.

Aït-Sahalia strategy: three steps

The approximation strategy has three layers:

1. transform X into Y to remove state-dependent volatility;
2. standardize Y into Z, which is closer to Gaussian;
3. approximate the density of Z using Hermite expansions, then map back to X.

$$X \to Y \to Z, \qquad Z \to Y \to X.$$

Idea

Approximate the density in a transformed space where Gaussian expansions are more accurate.

First transformation: $X \to Y$

Start from

$$dX_t = \mu(X_t;\theta)\,dt + \sigma(X_t;\theta)\,dB_t.$$

Define

$$Y_t = \gamma(X_t;\theta) = \int^{X_t} \frac{du}{\sigma(u;\theta)}.$$

Then Itô's lemma gives

$$dY_t = \mu_Y(Y_t;\theta)\,dt + dB_t.$$

Payoff

The diffusion coefficient becomes one, so the transformed process has unit volatility.

Transformed drift after the $X \to Y$ map

The transformed drift is

$$\mu_Y(y;\theta) = \frac{\mu(\gamma^{-1}(y;\theta);\theta)}{\sigma(\gamma^{-1}(y;\theta);\theta)} - \frac{1}{2}\frac{\partial \sigma}{\partial x}(\gamma^{-1}(y;\theta);\theta).$$

Trade-off

- Variance structure becomes simpler.
- Drift becomes more nonlinear and model-specific.

But this is still progress

Gaussian approximation is much easier after the diffusion coefficient has been flattened.

Second transformation: $Y \to Z$

Standardize the transformed process by

$$Z_t = \Delta^{-1/2}(Y_t - y_0).$$

Why this helps

The new variable Z_t is scaled so that its short-horizon law is much closer to a standard normal density than the original diffusion.

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

becomes the natural reference density for the approximation.

Hermite expansion for the density of Z

Define Hermite polynomials by

$$H_j(z) = e^{z^2/2} \frac{d^j}{dz^j} \left(e^{-z^2/2} \right).$$

Then the order-J approximation is

$$p_Z^{(J)}(\Delta, z \mid y_0; \theta) = \phi(z) \sum_{j=0}^{J} \eta_j(\Delta, y_0; \theta) H_j(z).$$

The coefficients satisfy

$$\eta_j(\Delta, y_0; \theta) = \frac{1}{j!} \int_{-\infty}^{\infty} H_j(z) p_Z(\Delta, z \mid y_0; \theta) \, dz.$$

Mapping the approximation back to X

Once $p_Z^{(J)}$ is available, define

$$p_Y^{(J)}(\Delta, y \mid y_0; \theta) = \Delta^{-1/2} p_Z^{(J)}\left(\Delta, \Delta^{-1/2}(y - y_0) \mid y_0; \theta\right).$$

Then map back to the original diffusion:

$$p_X^{(J)}(\Delta, x \mid x_0; \theta) = \sigma(x; \theta)^{-1} p_Y^{(J)}\left(\Delta, \gamma(x; \theta) \mid \gamma(x_0; \theta); \theta\right).$$

Main message

The approximation is carried out in a space where it behaves well, then translated back to the original variables used in the likelihood.

Closed-form density expansion

The textbook's closed-form expansion is

$$\widetilde{p}_Z^{(K)}(\Delta, z \mid y_0; \theta) = \Delta^{-1/2}\phi\left(\frac{y - y_0}{\Delta^{1/2}}\right)\exp\left(\int_{y_0}^{y}\mu_Y(\varpi;\theta)\,d\varpi\right)\sum_{k=0}^{K}c_k(y \mid y_0;\theta)\frac{\Delta^k}{k!}.$$

Why this matters

It yields a tractable approximate transition density that can be plugged directly into a numerical log-likelihood.

Key point

The approximation is explicit enough for computation, but rich enough to inherit the asymptotic properties of the true MLE as the expansion order grows.

Recursive coefficients in the expansion

The recursion starts from

$$c_0(y \mid y_0; \theta) = 1.$$

For $j \geq 1$,

$$c_j(y \mid y_0; \theta) = j(y - y_0)^{-j} \int_{y_0}^{y} (\varpi - y_0)^{j-1} \left\{ \lambda_Y(\varpi; \theta) c_{j-1}(\varpi \mid y_0; \theta) + \frac{1}{2} \frac{\partial^2 c_{j-1}(\varpi \mid y_0; \theta)}{\partial \varpi^2} \right\} d\varpi,$$

where

$$\lambda_Y(y; \theta) = -\frac{1}{2} \left[\mu_Y^2(y; \theta) + \frac{\partial \mu_Y(y; \theta)}{\partial y} \right].$$

Reading this slide

The recursive structure is what makes the approximation computational rather than merely symbolic.

Practical approximate-MLE workflow in R

```
# 1. Observe a discretely sampled diffusion path
x <- observed_series
Delta <- 1 / 252

# 2. Choose a model class for mu(x;theta), sigma(x;theta)

# 3. Build exact likelihood if p_X is known,
# otherwise build an approximate transition density p_X^(J)

# 4. Numerically maximize
theta_hat <- argmax_theta sum(log p_X_or_approx(Delta, x[i], x[i-1], theta))

# 5. Compute score, Hessian, information matrix,
# and standard errors
```

Software message

The statistical difficulty is analytical, not conceptual: once a usable transition density is available, likelihood evaluation is straightforward.

Diagnostics for approximate likelihood

In practice we check:

- sensitivity to starting values;
- score vector near the optimum;
- Hessian and information matrix stability;
- approximate versus exact derivatives when a benchmark case is available;
- conventional and sandwich standard errors.

Why diagnostics matter

Approximate likelihood is powerful, but numerical reliability still has to be checked. Good inference depends on both the quality of the approximation and the stability of the optimization.

Hour 1 summary

- Discrete-data likelihood for a diffusion is built from the transition density.
- Exact MLE is straightforward when the transition law is known, as in GBM and OU.
- Approximate likelihood is needed once exact densities become unavailable.
- Aït-Sahalia's method works by transforming the model to a nearly Gaussian space and then expanding the density there.

Bridge to Hour 2

Once the model is continuous-time, volatility is naturally summarized by quadratic variation. High-frequency data give us a direct way to estimate it.

Quadratic variation and integrated variance

For a continuous semimartingale X_t, the quadratic variation over $[0, t]$ is

$$\langle X, X \rangle_{0:t} = \text{plim}_{\max(t_{k+1} - t_k) \to 0} \sum_{t_k \le t} \left(X_{t_{k+1}} - X_{t_k} \right)^2.$$

For

$$dX_t = \mu_t \, dt + \sigma_t \, dB_t,$$

we have

$$\langle X, X \rangle_{0:t} = \int_0^t \sigma_u^2 \, du.$$

Interpretation

Quadratic variation is the cumulative continuous-time variance, or integrated variance.

Brownian semimartingale and realized volatility

A Brownian semimartingale can be written as

$$X_t = \int_0^t \mu_u \, du + \int_0^t \sigma_u \, dB_u.$$

If we observe n equally spaced log prices on $[0, 1]$, the realized volatility is

$$RV_X^n = \sum_{l=1}^{n-1} \left(X_{\frac{l+1}{n}} - X_{\frac{l}{n}} \right)^2.$$

Key fact

$$RV_X^n \xrightarrow{p} QV = \int_0^1 \sigma_u^2 \, du.$$

So realized volatility is the natural high-frequency estimator of integrated variance.

Mixed-normal CLT for realized volatility

In the no-leverage case,

$$n^{1/2}\left(RV_X^n - QV\right) \Rightarrow \sqrt{2}\int_0^1 \sigma_u^2\, dB_u \sim MN\left(0,\ 2\int_0^1 \sigma_u^4\, du\right).$$

Define integrated quarticity

$$IQ = \int_0^1 \sigma_u^4\, du,$$

with estimator

$$\widehat{IQ} = \frac{n}{3}\sum_{i=1}^{n-1}\left(X_{\frac{i+1}{n}} - X_{\frac{i}{n}}\right)^4.$$

Then

$$\frac{n^{1/2}(RV_X^n - QV)}{\sqrt{2}\,\widehat{IQ}^{1/2}} \Rightarrow N(0,1).$$

R block: simulate Brownian or GBM paths

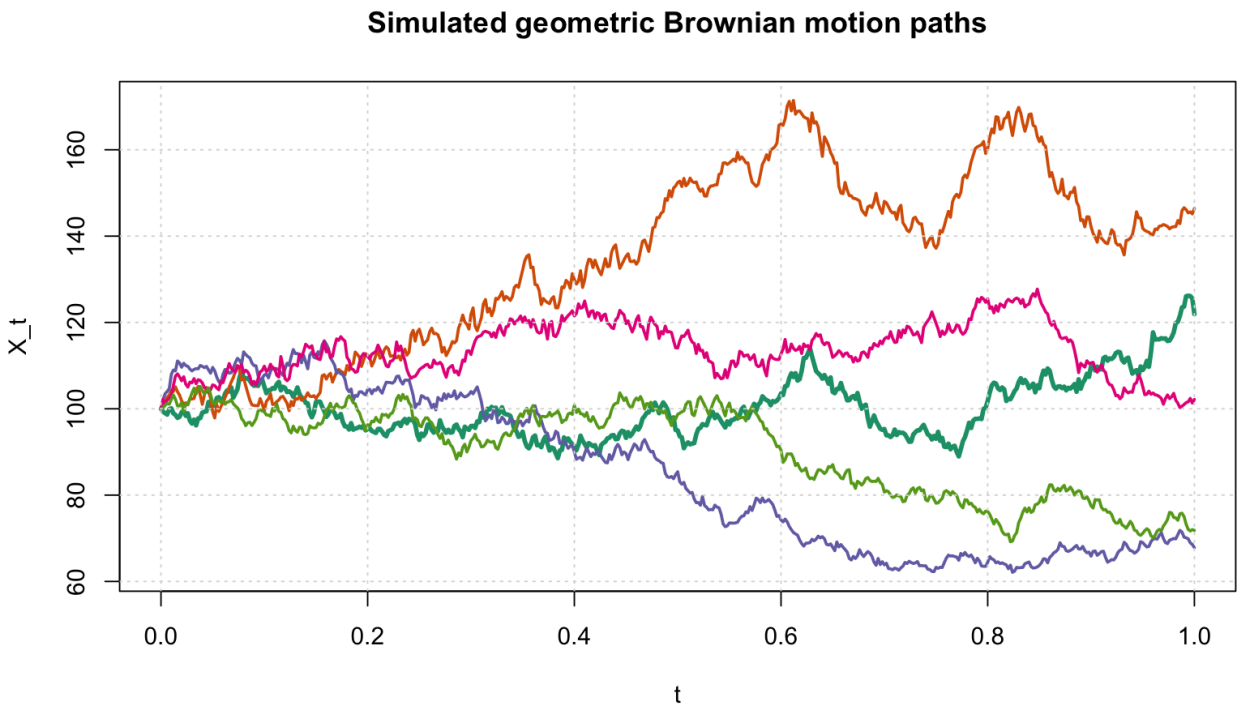
```
set.seed(1)
n <- 5000
Delta <- 1 / n
B <- c(0, cumsum(rnorm(n, sd = sqrt(Delta))))

mu <- 0.08
sigma <- 0.25
X <- numeric(n + 1)
X[1] <- 100
for (i in 2:(n + 1)) {
  X[i] <- X[i - 1] * exp((mu - 0.5 * sigma^2) * Delta +
                         sigma * (B[i] - B[i - 1]))
}
```

Purpose

Simulation lets us compare latent quadratic variation, realized variance at different grids, and the effect of adding observation noise.

Simulated GBM paths



Realized variance on a clean path

Given a sampling interval k,

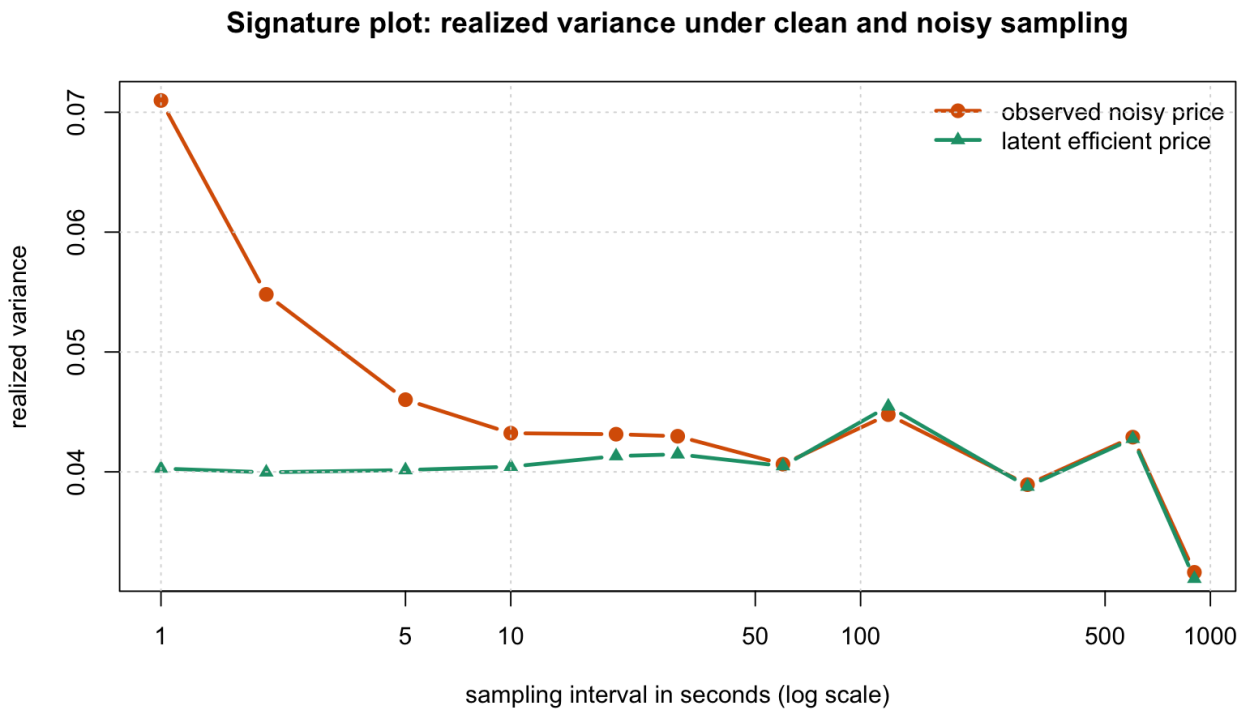
$$RV^{(k)} = \sum_j \left(X_{t_j} - X_{t_{j-k}}\right)^2.$$

Clean-data intuition

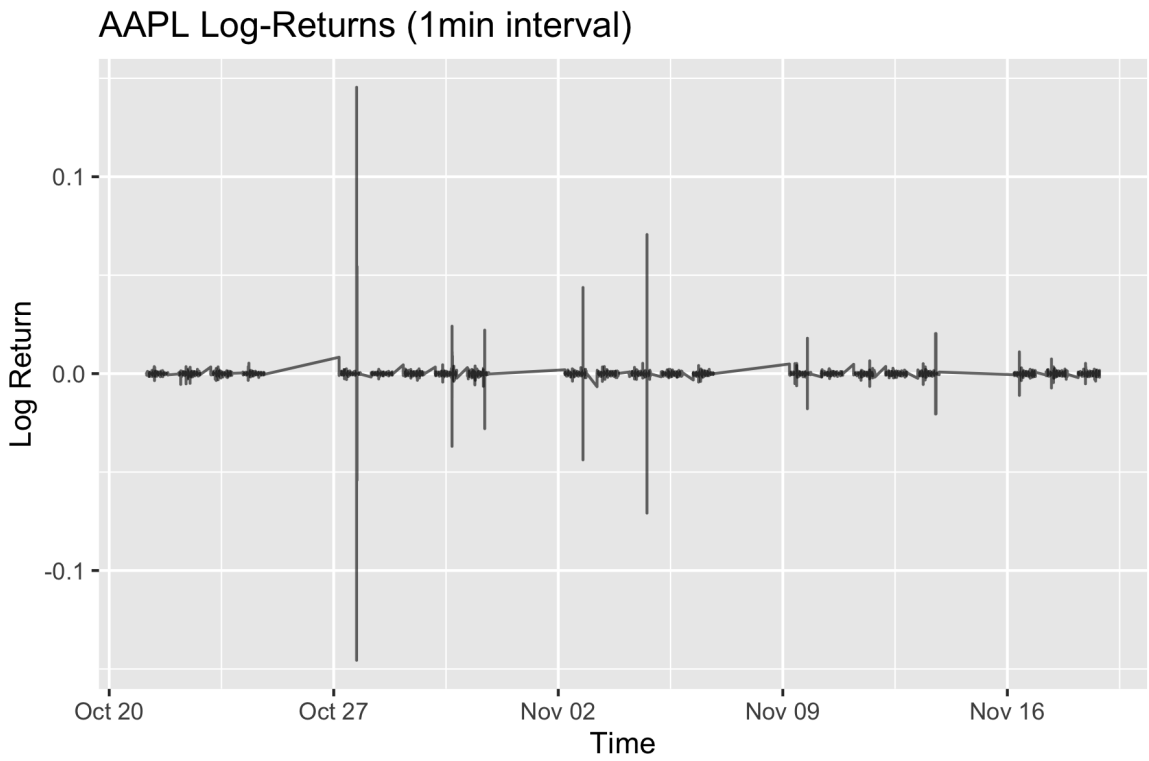
Without microstructure noise, realized variance stabilizes as we sample more finely because it converges to quadratic variation.

- Coarse sampling misses short-horizon variation.
- Fine sampling captures more of the latent integrated variance.

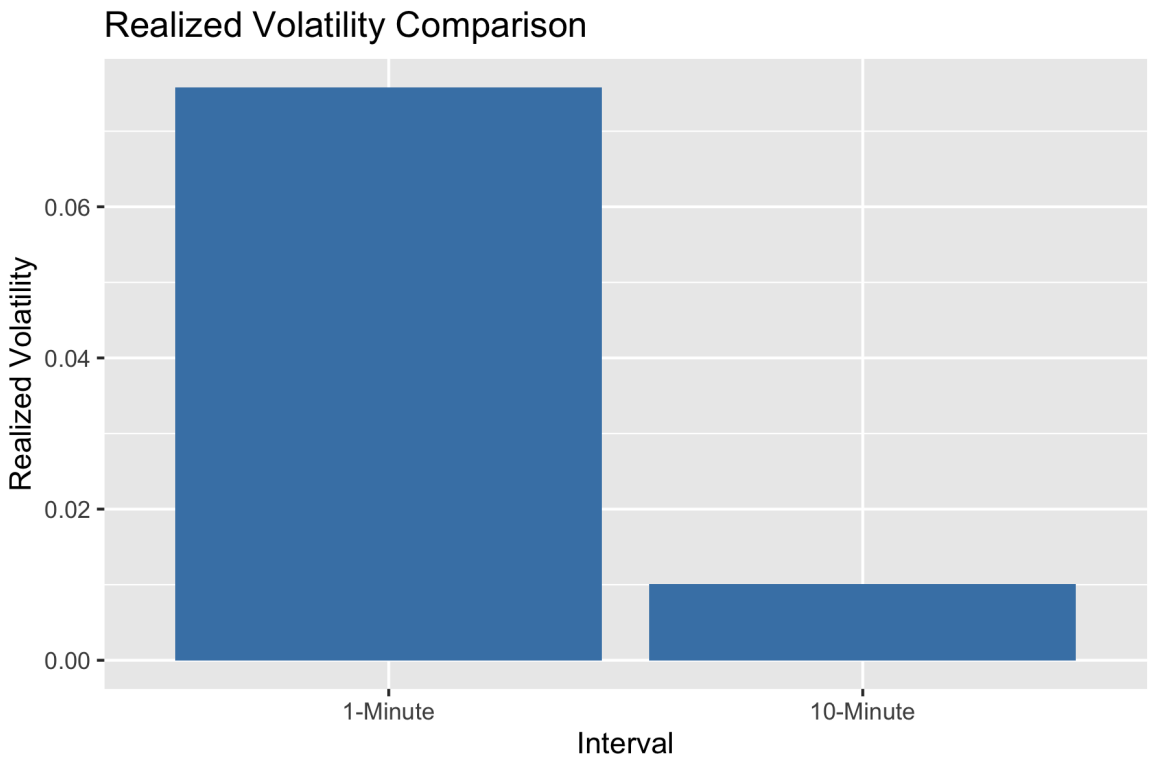
Signature plot under clean and noisy sampling



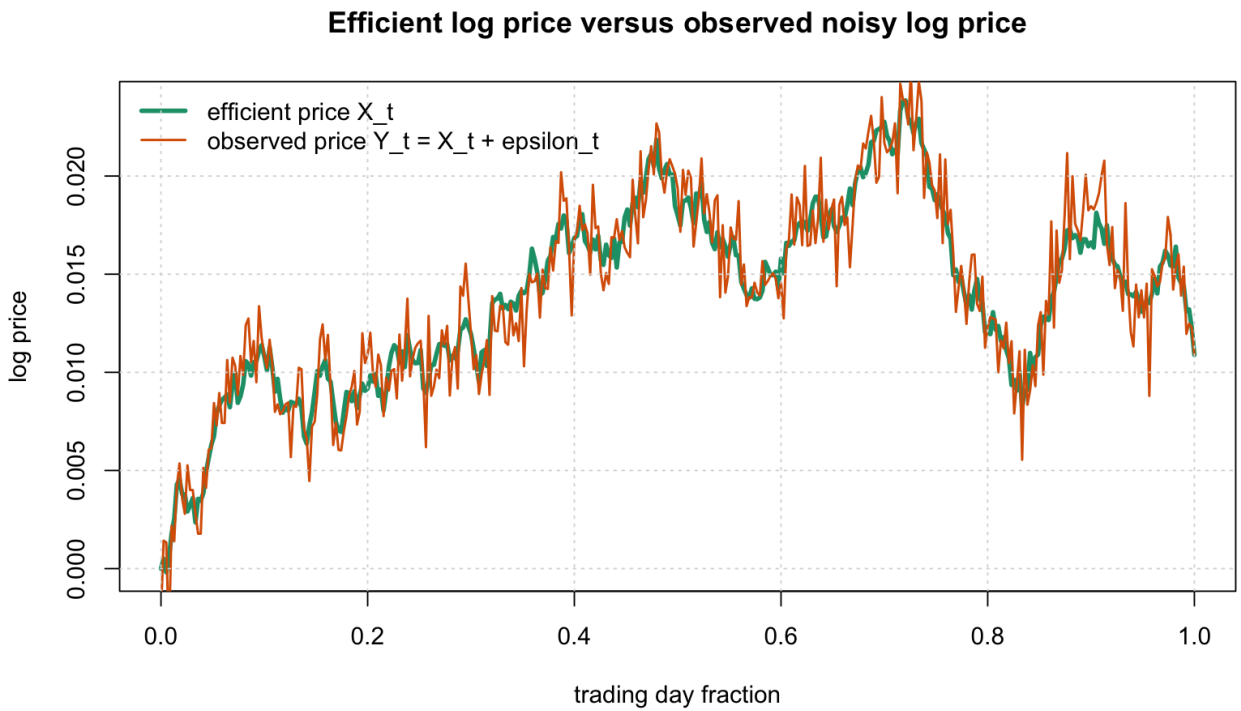
Textbook illustration: one-minute log returns



Textbook illustration: one-minute versus ten-minute RV



Efficient price versus noisy observed price



Measurement error model for microstructure noise

Let the observed log price be

$$Y_{t_j} = X_{t_j} + \varepsilon_{t_j},$$

where ε_{t_j} is i.i.d. noise with mean zero and variance σ_ε^2, independent of the latent efficient price X_{t_j}.

Interpretation

At lower frequencies, X_t dominates. At ultra-high frequencies, the noise can dominate short-run increments.

Empirical meaning

Bid–ask bounce, discreteness, execution frictions, and other microstructure effects contaminate naive realized volatility.

Why naive realized volatility fails under noise

The realized volatility of observed prices is

$$RV_Y^n = \sum_{i=1}^{n-1} \left(Y_{\frac{i+1}{n}} - Y_{\frac{i}{n}} \right)^2$$
$$= \sum_{i=1}^{n-1} \left(X_{\frac{i+1}{n}} - X_{\frac{i}{n}} \right)^2 + \sum_{i=1}^{n-1} \left(\varepsilon_{\frac{i+1}{n}} - \varepsilon_{\frac{i}{n}} \right)^2 + \text{cross term}.$$

Moreover,

$$\frac{1}{n} \sum_{i=1}^{n-1} \left(\varepsilon_{\frac{i+1}{n}} - \varepsilon_{\frac{i}{n}} \right)^2 \xrightarrow{p} 2\sigma_\varepsilon^2.$$

$$RV_Y^n \xrightarrow{p} \infty.$$

TSRV idea: subsampling and bias cancellation

Let $K(m+1) = n$. For each subsample $j = 1, \ldots, K$,

$$RV_{\text{sub}_j} = \sum_{i=1}^{m} \left(Y_{\frac{j+iK}{n}} - Y_{\frac{j+(i-\mathbf{1})K}{n}} \right)^2 .$$

Key idea

Each coarse-grid subsample reduces the impact of noise, while the full-sample RV tells us how large the bias is at the highest frequency.

The leading noise bias is proportional to m in the coarse subsample and proportional to n in the full sample, so a linear combination can remove it.

TSRV formula and related estimators

The two-scale realized volatility estimator is

$$\widehat{\theta}_{\text{TSRV}} = \frac{1}{K}\sum_{j=1}^{K} RV_{\text{sub}_j} - \frac{m}{n} RV_n.$$

Interpretation

Average coarse-grid RV gives the signal plus bias. The scaled full-sample RV estimates the bias. Subtracting the second from the first delivers a consistent estimator.

- MSRV generalizes TSRV to many scales.
- Realized-kernel and pre-averaging methods are more flexible under richer noise structures.

Hour 2 summary

- Quadratic variation is the continuous-time volatility object.
- Realized volatility consistently estimates it under clean high-frequency sampling.
- Mixed-normal limit theory uses integrated quarticity for studentization.
- Under microstructure noise, naive RV explodes and becomes inconsistent.
- TSRV corrects the leading noise bias using two sampling scales.

Bridge to Hour 3

Variance estimation is only the univariate part of the story. Covariance estimation is harder because asynchronicity enters in addition to noise.

Why covariance is harder than variance

Variance estimation uses one asset. Covariance estimation uses at least two assets, and therefore inherits extra complications:

- non-synchronous trading times;
- microstructure noise in each series;
- the need to preserve positive semi-definiteness of the covariance matrix.

Practical point

At high frequency, covariance is not just "variance twice". Synchronization and PSD constraints matter a lot.

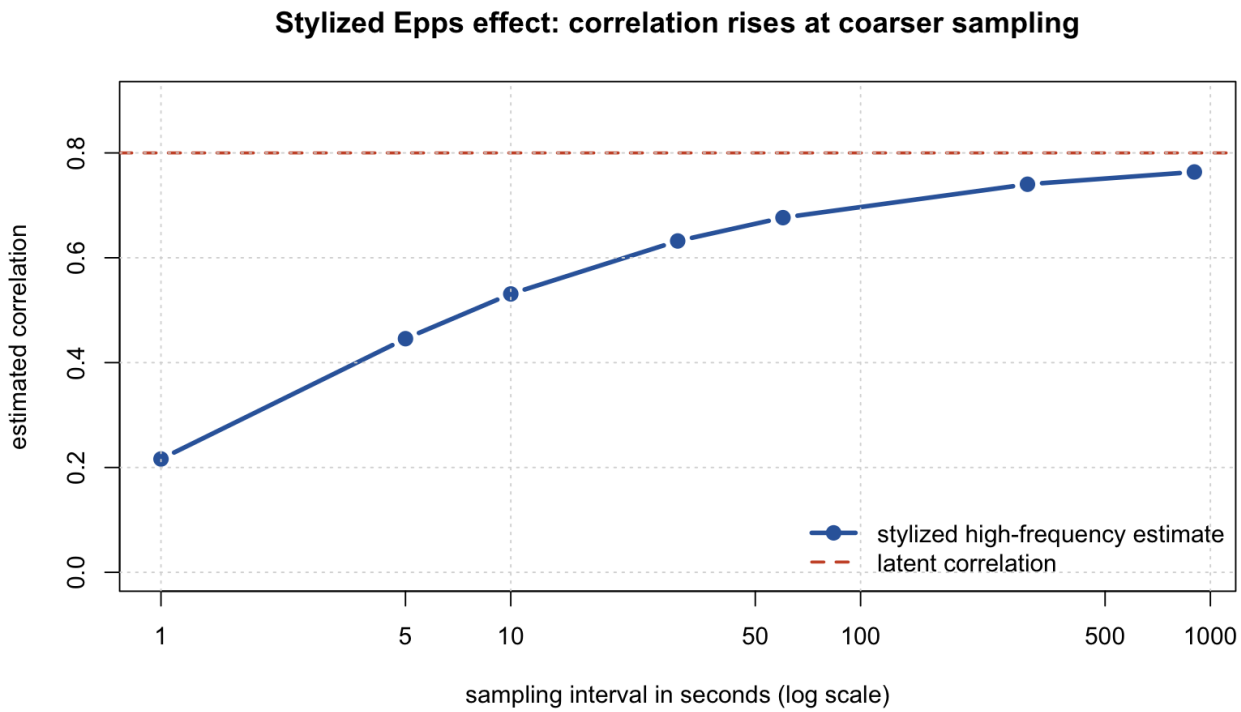
Three sources of bias

1. **Bid–ask bounce and noise:** observed prices differ from efficient prices.
2. **Discreteness:** prices move on a grid, not continuously.
3. **Asynchronicity:** different assets trade at different times.

Consequence

If we sample too finely and ignore these frictions, off-diagonal covariance estimates are biased downward and can even become numerically unstable.

A stylized Epps effect



Refresh-time synchronization

Refresh times solve the synchronization problem by keeping only times at which every asset has traded at least once since the previous refresh time.
For two assets,

$$\phi_1 = \max(\tau_1, \theta_1), \qquad \phi_{j+1} = \max(\tau_{N^X_{\phi_j}+1}, \theta_{N^Y_{\phi_j}+1}).$$

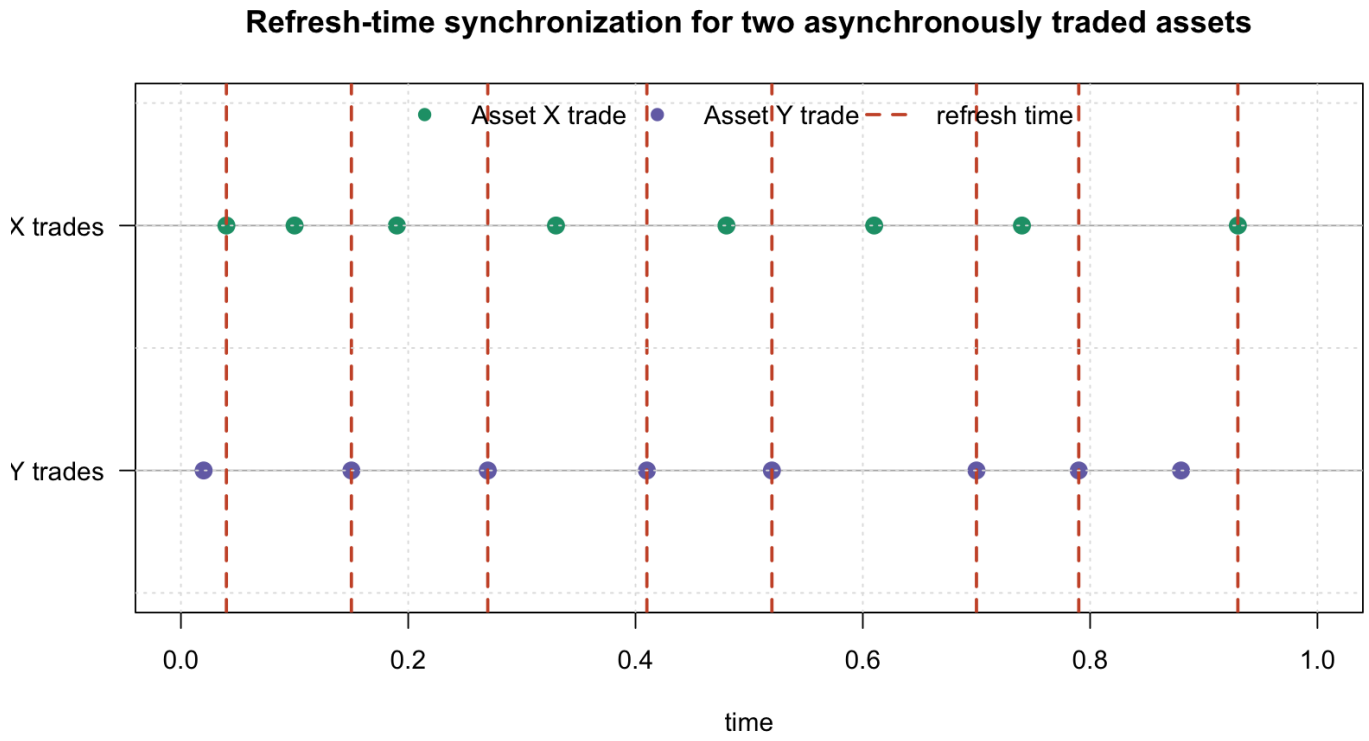
Trade-off

You get a synchronized panel, but at the cost of discarding observations.

$$p = \frac{dN}{\sum_{i=1}^{d} n^{(i)}}$$

measures the retention ratio.

Refresh-time illustration



TSX: two-scale realized covariance

For a univariate log-price series X, the TSX variance component is

$$\left(1-\frac{\bar{n}_K}{\bar{n}_J}\right)^{-1}\left([X,X]_T^{(K)}-\frac{\bar{n}_K}{\bar{n}_J}[X,X]_T^{(J)}\right),$$

where

$$\bar{n}_K=\frac{n-K+1}{K},\qquad [X,X]_T^{(K)}=\frac{1}{K}\sum_{i=1}^{n-K+1}(X_{t_{i+K}}-X_{t_i})^2.$$

The covariance component after refresh-time synchronization is

$$c_N\left([X,Y]_T^{(K)}-\frac{\bar{n}_K}{\bar{n}_J}[X,Y]_T^{(J)}\right).$$

TSX: strengths and weaknesses

- It addresses microstructure bias by differencing across time scales.
- It handles off-diagonal covariances through synchronized returns.
- It is easy to motivate from the univariate TSRV logic.

Limitation

TSX is not automatically positive semi-definite, so practical implementations often replace negative eigenvalues by zero.

Implementation note

In `highfrequency`, the function `rTSCov(..., makePsd = TRUE)` is commonly used.

RKX construction and jittering

After refresh-time synchronization, define N refresh points and n, m such that

$$n - 1 + 2m = N.$$

Then boundary prices are jittered:

$$X_0 = \frac{1}{m}\sum_{j=1}^{m} X(\tau_j), \qquad X_n = \frac{1}{m}\sum_{j=1}^{m} X(\tau_{N-m+j}),$$

and interior returns are

$$r_j = X_j - X_{j-1}, \qquad j = 1, \ldots, n.$$

Purpose of jittering

Averaging boundary observations reduces noise and stabilizes the realized-kernel construction.

RKX formula

The multivariate realized kernel estimator is

$$K(X) = \sum_{h=-n}^{n} k\left(\frac{h}{H+1}\right)\Gamma_h,$$

where

$$\Gamma_h = \begin{cases} \sum_{j=|h|+1}^{n} \mathrm{r}_j \mathrm{r}'_{j-h} & h \geq 0, \\ \sum_{j=|h|+1}^{n} \mathrm{r}_{j-h} \mathrm{r}'_j & h < 0. \end{cases}$$

Interpretation

RKX is a weighted sum of lagged realized autocovariance matrices, with weights supplied by a kernel function.

Kernel conditions and bandwidth for RKX

The kernel k should satisfy:

- $k(0) = 1$ and $k'(0) = 0$;
- sufficient smoothness, typically twice differentiable;
- nonnegative Fourier transform to guarantee PSD.

$$\int_{-\infty}^{\infty} k(x) e^{i\lambda x}\, dx \geq 0.$$

Why RKX is attractive

It is PSD by construction and can handle endogenous or serially correlated microstructure noise better than simpler two-scale estimators.

PAVX: pre-averaging covariance estimator

Pre-averaged returns are

$$\bar{r}^{(k)}_{\tau_j} = \sum_{h=1}^{k_N-1} g\left(\frac{h}{k_N}\right) r^{(k)}_{\tau_{j+h}}, \qquad g(x) = \min(x, 1-x), \qquad k_N = \lfloor \theta N^{1/2} \rfloor.$$

The multivariate PAVX estimator is

$$\mathrm{PAVX} = \frac{N}{N-k_N+2} \cdot \frac{1}{\psi_2 k_N} \sum_{i=0}^{N-k_N+1} \bar{\mathrm{r}}_{\tau_i} \bar{\mathrm{r}}'_{\tau_i} - \frac{\psi_1^{k_N}}{\theta^2 \psi_2^{k_N}} \widehat{\Psi}_N,$$

where

$$\widehat{\Psi}_N = \frac{1}{2N} \sum_{i=1}^{N} \mathrm{r}_{\tau_i} \mathrm{r}'_{\tau_i}.$$

Idea

Smooth first, then bias-correct.

TSX, RKX, and PAVX compared

Estimator	Main idea	Strength	Caveat
TSX	Two scales + refresh time	simple noise correction	needs PSD fix
RKX	Kernel lags	PSD by design	bandwidth sensitive
PAVX	Pre-average	fast asymptotics	more tuning

Practical reading

There is no universally best estimator. The right choice depends on noise structure, data dimension, computational constraints, and whether PSD is essential for the application.

R workflow with the highfrequency package

```
library(highfrequency)

# 1. Synchronize prices
ref_data <- refreshTime(list(asset1 = x1, asset2 = x2))

# 2. Estimate covariance matrices
Sigma_tsx <- rTSCov(ref_data, makePsd = TRUE)
Sigma_rkx <- rKernelCov(ref_data)
Sigma_pavx <- rMRCov(ref_data, makePsd = TRUE)

# 3. Compare covariance estimates and downstream portfolios
```

Workflow

Synchronize first, estimate second, then use the covariance matrix for inference, forecasting, or portfolio construction.

From covariance estimation to portfolio choice

Given an estimated covariance matrix Σ, the global minimum variance portfolio solves

$$\min_{w} \; w'\Sigma w \qquad \text{subject to} \qquad 1'w = 1.$$

The solution under full investment and no other constraints is

$$w^{\star} = \frac{\Sigma^{-1}1}{1'\Sigma^{-1}1}.$$

Why high-frequency covariance matters

If the covariance matrix is biased because of noise or asynchronicity, the resulting portfolio weights can be badly distorted.

Course wrap-up

This course moved through six major layers:

1. univariate and multivariate time-series foundations;
2. inference with dependence and long-run variance issues;
3. bootstrap, GMM, and filtering;
4. state-space models and Kalman methods;
5. Brownian motion, diffusion processes, and continuous-time likelihood;
6. high-frequency volatility and covariance estimation.

Common thread

The whole course has been about extracting reliable dynamic information from dependent, noisy, and often only indirectly observed data.

What to review before Lecture 20

- Exact versus approximate likelihood for diffusions.
- The meaning of quadratic variation and realized volatility.
- Why microstructure noise invalidates naive high-frequency estimators.
- The logic of TSRV, RKX, and PAVX.
- The role of synchronization and PSD covariance matrices in empirical work.

Final takeaway

Continuous-time econometrics is useful only when theory, computation, and noisy real data are all brought into the same framework. That is what Lecture 19 has been about.