








Modelling Financial Markets

Seminar 9: Realised Volatility — Hands-on



Daniel Traian PELE
Drd. Antoaneta Amza, antoaneta.amza@csie.ase.ro

Bucharest University of Economic Studies
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Romanian Academy, Institute for Economic Forecasting
MSCA Digital Finance

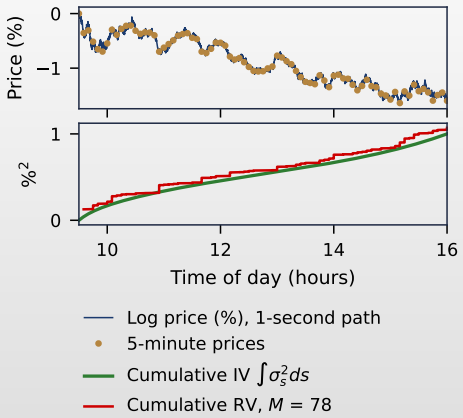
Does Intraday Measurement Improve a Volatility Forecast?

- Question of the seminar
  - **how precisely does one day of intraday prices measure volatility, and does the measurement improve a forecast?**
- Two parts, in this order
  - Part I, measurement and diagnosis: A1, A2, B1, B2, A3
  - Part II, models and forecasts: A4, B3, A5, A6, C1
- Deliverables
  - the answers to the A exercises
  - one figure and one table per B exercise
  - the C1 correction table
  - one interpretation sentence per exercise
- Notebook for today: open the seminar notebook in Colab; each task names its notebook section
- Background: Chapter 9 and Franke, Härdle & Hafner (2019)
- Seminar: Prof. Daniel Traian Pele, danpele@ase.ro; Antoaneta Amza, PhD candidate, antoaneta.amza@csie.ase.ro

Which Market Clock and Measurement Units?

- SPY (SPDR S&P 500 ETF Trust) is an exchange-traded fund (ETF) tracking the Standard & Poor's 500 index (S&P 500)
- Bitcoin prices are in United States dollars (USD)
- Timestamps are in Coordinated Universal Time (UTC); SPY bars are converted to New York time
- Files in data/market/intraday: `SPY.US_5m.csv` and `BTC-USD.CC_5m.csv`, with `datetime_utc`
- A bar timestamp is the start of its interval
- Intraday chart axes show interval ends
- Annualisation uses 252 days for SPY and 365 for Bitcoin
- Returns are logarithmic, in percent; their squares are in percent squared
- C1 checks an artificial intelligence (AI) answer: for every number, state the asset, interval, unit and forecast origin

Realised Variance



- Details: Chapter 9
- **Integrated variance** (IV): the variance the day actually delivered

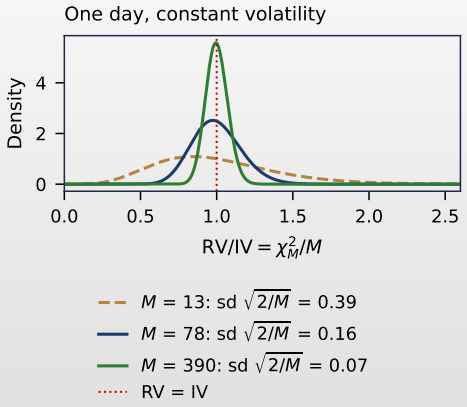
$$\mathrm{IV}_t = \int_{t-1}^{t} \sigma_s^2 \, ds$$

  - σ_s: spot (instantaneous) volatility at time s inside day t
- **Realised variance** (RV) Andersen et al. (2001): the sum of the M squared intraday returns

$$\mathrm{RV}_t = \sum_{i=1}^{M} r_{t,i}^2 \xrightarrow{p} \mathrm{IV}_t \quad (M \to \infty)$$

  - $r_{t,i} = 100\,(\ln P_{t,i} - \ln P_{t,i-1})$: i-th return of day t, in %; SPY: $M = 78$ five-minute returns
  - convergence requires a continuous price without noise; annualised volatility $\sqrt{252\,\mathrm{RV}_t}$ (%)
  - chart: a simulated day with U-shaped intraday volatility; RV tracks IV step by step

How Precise Is One Day's RV?

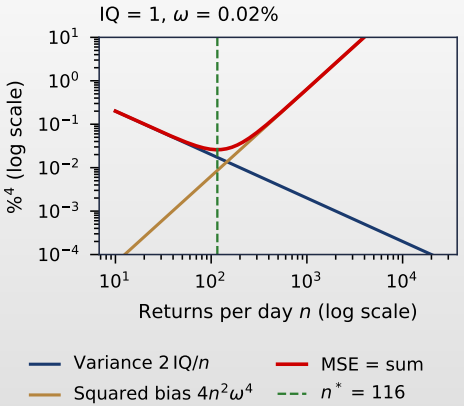


- Measurement error of RV for large M

$$\mathrm{RV}_t - \mathrm{IV}_t \;\approx\; N\Big(0,\; \frac{2\,\mathrm{IQ}_t}{M}\Big), \qquad \mathrm{IQ}_t = \int_{t-1}^{t} \sigma_s^4\, ds$$

  - IQ_t: integrated quarticity; estimated by the realised quarticity $\mathrm{RQ}_t = \frac{M}{3}\sum_i r_{t,i}^4$
  - constant volatility: $\mathrm{RV}/\mathrm{IV} = \chi^2_M/M$, a chi-square with M degrees of freedom divided by M, relative sd $\sqrt{2/M}$
- **Delta method**: for a smooth g, the standard error $\mathrm{se}(g(X)) \approx |g'(X)|\,\mathrm{se}(X)$
  - $g = \ln$: $\mathrm{se}(\ln \mathrm{RV}) \approx \mathrm{se}(\mathrm{RV})/\mathrm{RV}$; an interval built on the log scale and exponentiated stays positive
  - chart: the spread of RV/IV for 13, 78 and 390 returns a day

Microstructure Noise and the Sampling Frequency



- Observed log price = efficient price + noise

$$p_i^{\text{obs}} = p_i^* + \varepsilon_i, \qquad \varepsilon_i \text{ i.i.d. } (0, \omega^2)$$

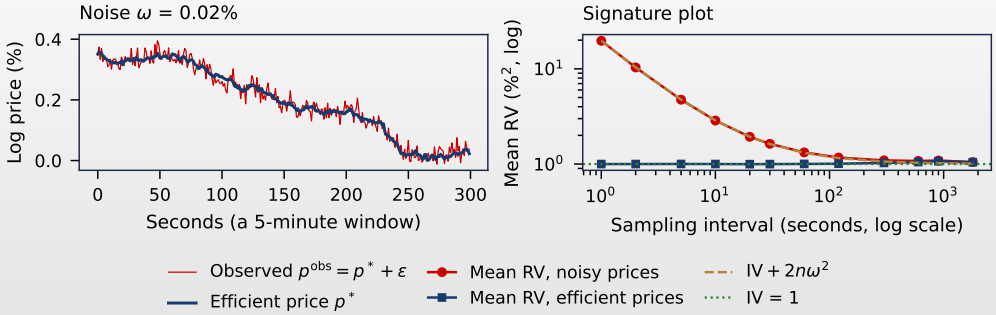
  - ω^2: noise variance (bid–ask bounce, price ticks), ε independent of p^*
  - each observed return $r_i = r_i^* + \varepsilon_i - \varepsilon_{i-1}$ carries two noise terms

- With n returns: bias grows with n, sampling variance falls like $1/n$ Bandi & Russell (2008)

$$\mathbb{E}[\mathrm{RV}^{(n)}] = \mathrm{IV} + 2n\omega^2, \qquad \mathrm{MSE}(n) \approx \frac{2\,\mathrm{IQ}}{n} + 4n^2\omega^4$$

  - MSE (Mean Squared Error) = variance + squared bias; the minimising n^* balances the two terms (chart: $\omega = 0.02\%$)
  - in practice 5 minutes is hard to beat Liu, Patton & Sheppard (2015)

Noise in Prices and the Signature Plot



- Simulation: 150 days, IV $= 1\,\%^2$, noise $\omega = 0.02\%$
  - left: tick by tick the observed price oscillates around the efficient one
  - right, **signature plot**: mean RV against the sampling interval; noisy prices follow $\mathrm{IV} + 2n\omega^2$, efficient prices stay at IV
  - an approximately flat signature plot is compatible with a small noise bias at the intervals analysed; it does not rule out noise

Sparse Grids, Averaging and the Block Bootstrap

- **Sparse grid** of $k \times 5$ minutes starting at offset $o \in \{0, \dots, k-1\}$

$$\mathrm{RV}_t^{(k,o)} = \sum_j \left(100\,\Delta \ln P\right)^{\mathbf{2}} \text{ on that grid}, \qquad \overline{\mathrm{RV}}_t^{(k)} = \frac{1}{k} \sum_{o=\mathbf{0}}^{k-\mathbf{1}} \mathrm{RV}_t^{(k,o)}$$

  - averaging over the k shifted grids uses all prices and lowers the sampling variance, without more noise bias
  - ratio of means $\sum_t \mathrm{RV}_t^{(k)} / \sum_t \mathrm{RV}_t^{(\mathbf{5\ min})}$: near 1 when the noise bias at 5 minutes is negligible

- **Moving-block bootstrap** Künsch (1989): a confidence interval for a statistic of dependent data
  - draw random starting days, join blocks of b consecutive days (here $b = 20$) until the original length T
  - recompute the statistic B times; the 2.5% and 97.5% percentiles give the 95% interval
  - blocks keep the autocorrelation inside each block; i.i.d. (independent and identically distributed) resampling would destroy it and give too narrow intervals

Jumps and Bipower Variation

- With jumps J inside the day, RV collects their squares

$$\mathrm{RV}_t \xrightarrow{p} \mathrm{IV}_t + \sum_{s \in \mathbf{day}\ t} J_s^{\mathbf{2}}$$

- **Bipower variation** (BV) Barndorff-Nielsen & Shephard (2004): products of adjacent absolute returns

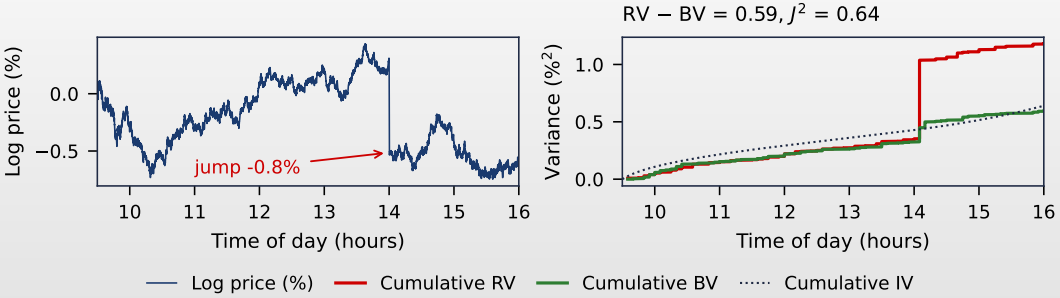
$$\mathrm{BV}_t = \mu_{\mathbf{1}}^{-\mathbf{2}} \frac{M}{M-1} \sum_{i=\mathbf{2}}^{M} |r_{t,i}|\,|r_{t,i-\mathbf{1}}| \xrightarrow{p} \mathrm{IV}_t, \qquad \mu_{\mathbf{1}} = \mathbb{E}|Z| = \sqrt{2/\pi}$$

  - $Z \sim N(0,1)$, so $\mu_{\mathbf{1}}^{-\mathbf{2}} = \pi/2$; with finite-activity jumps and no noise (the base case), a jump enters BV only multiplied by a small neighbouring return; other settings need extra conditions
  - jump part $J_t = \mathrm{RV}_t - \mathrm{BV}_t$; jump share in total variation $\sum_t J_t / \sum_t \mathrm{RV}_t$
- **Tripower quarticity** (TQ): a jump-robust estimate of IQ_t, used by the jump test

$$\mathrm{TQ}_t = M \mu_{\mathbf{4/3}}^{-\mathbf{3}} \frac{M}{M-2} \sum_{i=\mathbf{3}}^{M} |r_{t,i}|^{\mathbf{4/3}} |r_{t,i-\mathbf{1}}|^{\mathbf{4/3}} |r_{t,i-\mathbf{2}}|^{\mathbf{4/3}}, \qquad \mu_{\mathbf{4/3}} = \mathbb{E}|Z|^{\mathbf{4/3}}$$

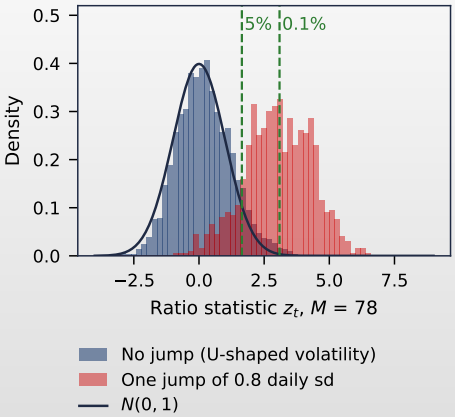
- Tiny example, $M = 4$, returns $0.1, 0.1, 0.1, 1.0$ (%)
  - $\mathrm{RV} = 1.03$; $\mathrm{BV} = \frac{\pi}{\mathbf{2}} \cdot \frac{\mathbf{4}}{\mathbf{3}} (0.01 + 0.01 + 0.10) = 0.25$; jump part $\mathrm{RV} - \mathrm{BV} = 0.78$

A Jump Seen by RV and by BV



- Simulated day, IV $= 0.64\,\%^2$, one jump of -0.8% at 14:00, 78 five-minute returns
  - cumulative RV jumps by about $J^2 = 0.64$ at 14:00; cumulative BV keeps following IV
  - at the close $\mathrm{RV} - \mathrm{BV} = 0.59$: close to the squared jump

The Ratio Jump Test



- H_0: no jump on day t; Huang–Tauchen ratio statistic Huang & Tauchen (2005)

$$z_t = \frac{(\mathrm{RV}_t - \mathrm{BV}_t)/\mathrm{RV}_t}{\sqrt{\theta\, M^{-1} \max\left(1,\ \mathrm{TQ}_t/\mathrm{BV}_t^2\right)}}$$

  - $(\mathrm{RV}_t - \mathrm{BV}_t)/\mathrm{RV}_t$: relative jump measure, near 0 under H_0
  - $\theta = \pi^2/4 + \pi - 5 = 0.609$; $\mathrm{TQ}_t/\mathrm{BV}_t^2$ measures how uneven volatility was in the day, floored at 1

- Under H_0, $z_t \approx N(0,1)$; one-sided test
  - reject when $z_t > 1.645$, 2.326, 3.09 at 5%, 1%, 0.1%; p-value $p_t = 1 - \Phi(z_t)$, Φ: standard Normal cdf
  - N days tested: about αN rejections by chance alone (multiple testing, Chapter 9)
  - chart: with U-shaped volatility and no jump, 9% of days exceed 1.645 instead of 5%

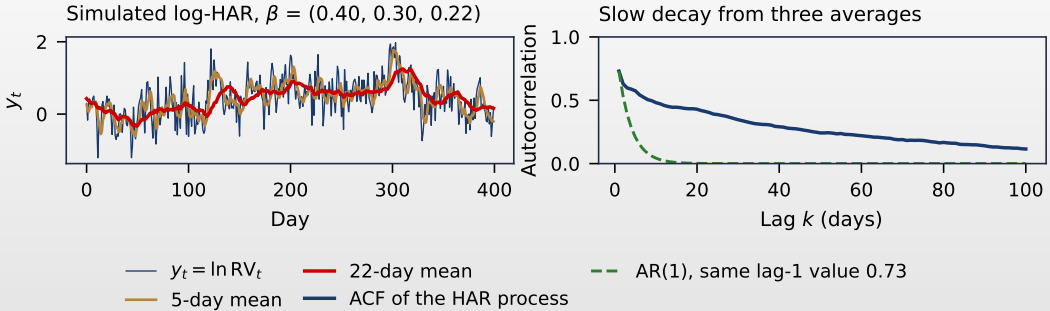
The HAR Model

- Volatility is highly persistent, consistent with long memory (Chapter 1): the autocorrelation function (ACF) of $|r_t|$ decays slowly, while a GARCH(1,1) forecast (generalised autoregressive conditional heteroskedasticity, Chapter 5) reverts exponentially
- HAR (Heterogeneous Autoregressive) Corsi (2009): tomorrow's RV on three averages of past RV

$$\mathrm{RV}_{t+\mathbf{1}} = \beta_{\mathbf{0}} + \beta_d \mathrm{RV}_t + \beta_w \mathrm{RV}_t^{(\mathbf{5})} + \beta_m \mathrm{RV}_t^{(\mathbf{22})} + \varepsilon_{t+\mathbf{1}}, \qquad \mathrm{RV}_t^{(h)} = \frac{1}{h} \sum_{j=\mathbf{0}}^{h-\mathbf{1}} \mathrm{RV}_{t-j}$$

  - β_d, β_w, β_m: weights on yesterday, the last week (5 days) and the last month (22 days); $\beta_{\mathbf{0}}$: constant
  - HAR is an autoregressive AR(22) model whose 22 coefficients are tied to three slopes: a stationary HAR has short memory
  - with non-negative slopes, persistence $\beta_d + \beta_w + \beta_m < 1$ is sufficient for stationarity: shocks die out
  - in general (e.g. a negative slope), check that the roots of the AR(22) polynomial lie outside the unit circle
- **log-HAR**: the same regression for $y_t = \ln V_t$
  - forecast of the level $F_{t+\mathbf{1}} = \exp(\hat{y}_{t+\mathbf{1}} + s^{\mathbf{2}}/2)$, $s^{\mathbf{2}}$: residual variance (lognormal correction)
  - positive forecasts, residuals closer to the Normal distribution
- Tiny example: $\hat{\beta} = (0.1, 0.4, 0.3, 0.2)$, $\mathrm{RV}_t = 2$, $\mathrm{RV}_t^{(\mathbf{5})} = 1.5$, $\mathrm{RV}_t^{(\mathbf{22})} = 1$
  - forecast $0.1 + 0.8 + 0.45 + 0.2 = 1.55\,\%^{\mathbf{2}}$; persistence 0.9

Three Averages Approximate Persistence over Finite Horizons



- Simulated log-HAR, $\beta_d = 0.40$, $\beta_w = 0.30$, $\beta_m = 0.22$ (persistence 0.92)
  - left: the 22-day mean moves slowly and anchors the forecast; the daily term reacts fast
  - right: the ACF is still 0.24 at lag 50; an AR(1) with the same lag-1 value has forgotten everything by lag 15

Inference in Time-Series Regressions

- Ordinary least squares (OLS): $\hat{\beta} = (X^\top X)^{-1} X^\top y$; with autocorrelated, heteroskedastic errors: the Newey–West heteroskedasticity and autocorrelation consistent (HAC) covariance

$$\hat{V}_{\mathbf{NW}} = (X^\top X)^{-1}\Big[\hat{\Gamma}_0 + \sum_{k=1}^{q}\Big(1 - \frac{k}{q+1}\Big)(\hat{\Gamma}_k + \hat{\Gamma}_k^\top)\Big](X^\top X)^{-1}$$

  - $\hat{\Gamma}_k = \sum_t \hat{u}_t \hat{u}_{t-k}\, x_t x_{t-k}^\top$, $\hat{u}_t$: residuals, x_t: regressor row; lags $q = \lfloor 4(T/100)^{2/9} \rfloor$

- **Wald test** of q_R linear restrictions $H_0 : R\beta = r$

$$W = (R\hat{\beta} - r)^\top \big(R\hat{V}R^\top\big)^{-1}(R\hat{\beta} - r) \xrightarrow{d} \chi^2_{q_R}$$

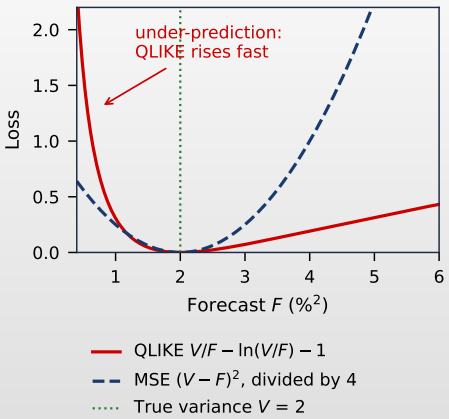
  - R: one row per restriction; e.g. $\beta_w = \beta_m$ is the row $(0, 0, 1, -1)$, $r = 0$

- **Ljung–Box**: are the first h autocorrelations jointly zero?

$$Q(h) = T(T+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{T-k}$$

  - $\hat{\rho}_k$: sample autocorrelation at lag k; on residuals, a large Q = memory left unexplained
  - for an i.i.d. series, $Q(h) \to \chi^2_h$; on the residuals of an estimated dynamic model (HAR) this reference is not valid in general: calibrate by a bootstrap that re-estimates the model
  - fit: $R^2 = 1 - \sum_t \hat{u}_t^2 / \sum_t (y_t - \bar{y})^2$, the share of the variance of y explained

Loss Functions for Variance Forecasts

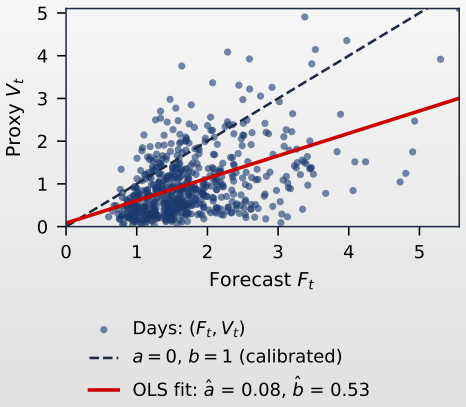


- The true variance is never observed: score F_t against a proxy V_t, conditionally unbiased, $\mathbb{E}[V_t \mid \mathcal{F}_{t-1}] = \mathbb{E}[\mathrm{IV}_t \mid \mathcal{F}_{t-1}]$
- Two losses that rank forecasts as the true variance would Patton (2011)

$$\mathrm{QLIKE}(V,F) = \frac{V}{F} - \ln\frac{V}{F} - 1, \qquad \mathrm{MSE}(V,F) = (V-F)^2$$

  - both are 0 at $F = V$; QLIKE (Quasi-Likelihood) penalises under-prediction more
- **Diebold–Mariano** Diebold & Mariano (1995): $d_t = L(V_t, F_t^A) - L(V_t, F_t^B)$
  - $t = \bar{d}/\mathrm{se}_{\mathrm{NW}}(\bar{d}) \approx N(0,1)$ under equal expected loss; $t < -1.96$: A significantly better

Calibration: the Mincer–Zarnowitz Regression



- Regress the proxy on the forecast

$$V_t = a + b\,F_t + u_t$$

  - a: intercept, b: slope, u_t: error; a calibrated forecast satisfies $a = 0$ and $b = 1$ in this linear regression; $\mathbb{E}[V_t - F_t] = 0$ alone does not imply it
  - $b < 1$: high forecasts come with realised means below the 45° line (chart: $\hat{b} = 0.53$)

- Joint Wald test of $(a, b) = (0, 1)$, with $\delta = (\hat{a},\ \hat{b} - 1)^\top$

$$W = \delta^\top \hat{V}^{-1} \delta \xrightarrow{d} \chi^2_2$$

  - $\hat{V}$: 2×2 covariance of $(\hat{a}, \hat{b})$, built from the standard errors and their correlation
  - 95% confidence ellipse: the (a, b) with $W \le \chi^2_{2;0.95} = 5.99$
  - not rejecting $(a, b) = (0, 1)$ does not prove calibration

Benchmark Forecasts, Out of Sample

- GARCH(1,1)-t on daily returns (Chapter 5): a model-based forecast

$$F_t = \sigma_t^2 = \omega + a\,\varepsilon_{t-1}^2 + b\,\sigma_{t-1}^2, \qquad \varepsilon_t = r_t - \mu = \sigma_t z_t,\ z_t \sim t_\nu$$

  - $\omega > 0$: the GARCH constant (not ω, the standard deviation of microstructure noise); $a, b \geq 0$, $a + b < 1$
  - z_t: unit-variance Student-t shocks with ν degrees of freedom

- EWMA (Exponentially Weighted Moving Average, RiskMetrics): exponentially weighted squared returns, $\lambda = 0.94$

$$\sigma_t^2 = \lambda\,\sigma_{t-1}^2 + (1-\lambda)\,r_{t-1}^2$$

  - a GARCH(1,1) with $\omega = 0$ and $a + b = 1$: no mean reversion
  - RW (random walk): yesterday's measure as today's forecast, $F_t = V_{t-1}$ (column `RW`)

- Out of sample, expanding window
  - at each origin $t-1$ every model is re-estimated on data up to $t-1$, then F_t is compared with V_t
  - losses are averaged over the common dates of all models

Data Used

Series	Content	Sample
SPY, 5-minute bars	prices of the regular session, UTC time stamps, each bar labelled by its start	2020-10-12 – 2026-09-18, 1 489 days
Bitcoin (USD), 5-minute bars	24-hour UTC days	659 of 747 days kept
SPY and Bitcoin, daily prices	adjusted closes (SPY) and closes (Bitcoin), for GARCH	since 2000 (SPY), 2014 (Bitcoin)
SPY daily variances (derived)	intraday RV, overnight return and total variance V_t ($\%^2$, %)	one row per SPY day
One-day forecasts (derived)	log-HAR, HAR, GARCH(1,1)-t, EWMA, yesterday's RV (column `RW`) and the target V_t	from 3 Jan 2022 (SPY), 1 Mar 2025 (Bitcoin)

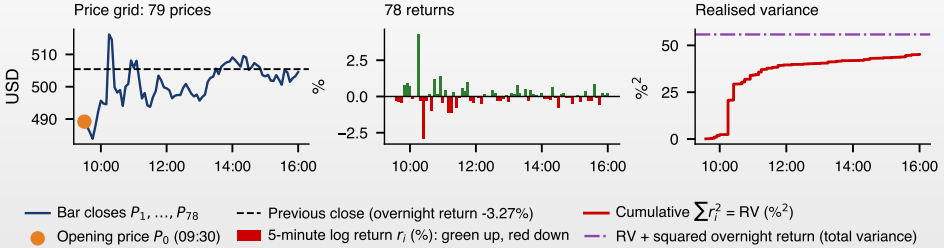
- SPY: UTC converted to New York time; bars 09:30–15:55; missing bars inside a day carry the last price; 78 returns a day, 42 (09:30–13:00) on the 12 early-close days, where the bar labelled 13:00 lies after the close and is not an interval
- Bitcoin: days with fewer than 95% of the 288 bars are dropped; gaps carry the last price, so a missing interval merges with the next one; the price at 00:00 is the open of the first bar or the previous day's 23:55 close; the day ends at the last observed price
- Units: returns $r = 100\,\Delta \ln P$ in %, variances in $\%^2$; annualise with 252 (SPY) or 365 (Bitcoin) days; all data end on 18 September 2026

Setup: Run and Verify

```
# notebook section "Data" defines
# spy_intraday, btc_intraday,
# intraday_returns, spy_overnight
P = spy_intraday()         # day x 79 prices
R = intraday_returns(P)    # day x 78 returns, %
print(P.shape)
print(R.notna().sum(axis=1).value_counts())
Pb = btc_intraday()        # day x 289 prices, UTC
print(Pb.shape, Pb.index[0], Pb.index[-1])
```

- Expected output
  - (1489, 79): SPY days, from 2020-10-12 to 2026-09-18
  - 1 477 days with 78 returns, 12 with 42
  - (659, 289): Bitcoin, 2024-09-01 to 2026-09-17
- Column 0 of `P` is the opening price of the 09:30 bar; columns 1–78 are the closes of the bars
- If your counts differ, stop and check the time zone, the session filter and the cut-off date

From Bars to Returns: One SPY Day



- 7 April 2025: opening price $P_0 = 489.19$, then bar closes $P_1 = 487.60$, $P_2 = 485.88$, …, $P_{78} = 504.54$: 79 prices give 78 returns; differencing the 78 closes alone would give 77
- $r_1 = 100 \ln(P_1/P_0) = -0.33\%$, $r_2 = -0.35\%$; $\mathrm{RV} = \sum_i r_i^2 = 45.1\,\%^2$; the overnight return -3.27% is outside the bars, total variance $V = \mathrm{RV} + r_{\mathrm{on}}^2 = 55.8\,\%^2$
- Overnight return: the adjusted close-to-close return minus the open-to-close return, so dividends fall in the overnight part

MFM_ch9_seminar

Notation Used in Every Exercise

- Day t has M intraday returns $r_{t,i}$, $i = 1, \ldots, M$, in %
- Total daily variance $V_t = \mathrm{RV}_t + r_{\mathrm{on},t}^2$ for SPY ($r_{\mathrm{on},t}$: the overnight return); $V_t = \mathrm{RV}_t$ for Bitcoin (24 hours)
- Forecast F_t of V_t, made at the close of day $t-1$; every loss compares F_t with V_t

Exercise Map

Exercise	Question	Type	Worked example
A1	how precise is one day's RV?	Solved	–
A2	how often should a noisy price be sampled?	Solved	–
B1	does SPY show a noise bias at 5 minutes?	Solved	A2
B2	which SPY days contain a jump?	Solved	–
A3	can six returns reveal a jump?	Proposed	B2, A1
A4	what does HAR impose on an AR(22)?	Solved	–
B3	how persistent is SPY's log variance?	Solved	A4
A5	which loss ranks variance forecasts correctly?	Solved	–
A6	is GARCH less accurate, biased, or both?	Proposed	A5; Wald example
C1	what is wrong in an AI answer?	Proposed	A2, B3
Extension packs	A2–A5, B1–B3 extended; A7, A8, B4–B9, C2, C3 (optional, after the seminar)	Solved and Proposed	named on each task

- **[Solved]**: full solution in the slides and in the notebook, a model to copy; **[Proposed]**: you solve it, the solution is discussed in class
- A exercises: derivations on paper; B exercises: computation and inference on SPY and Bitcoin, each ending with an interpretation question; C exercises: open questions and AI critique

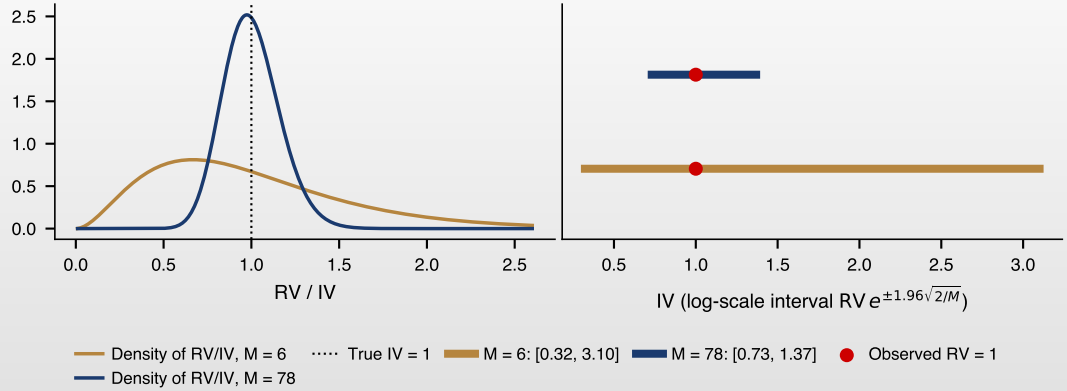
A1 [Solved]: Precision of Realised Variance — Task

- **Question**: how precisely does one day of 5-minute returns measure integrated variance?
- **Inputs**: constant volatility, M intervals of length $\Delta = 1/M$, $r_i = \sigma\sqrt{\Delta}\, Z_i$ with Z_i i.i.d. $N(0,1)$, $\mathrm{IV} = \sigma^2$; use $\mathbb{E}Z^2 = 1$, $\mathrm{Var}(Z^2) = 2$, $\mathbb{E}Z^4 = 3$; χ^2_M: chi-square with M degrees of freedom (a sum of M squared i.i.d. $N(0,1)$); no data file
- **Tasks**
  1. Show that $\mathrm{RV} = \sigma^2 \chi^2_M / M$ and derive $\mathbb{E}[\mathrm{RV}]$ and $\mathrm{Var}(\mathrm{RV})$.
  2. Show that $\mathrm{RQ} = \frac{M}{3}\sum_i r_i^4$ is unbiased for $\mathrm{IQ} = \sigma^4$ and write the estimated standard error of RV.
  3. Use the delta method for $\ln \mathrm{RV}$ and write the 95% interval for IV built on the log scale.
  4. Compute the interval factors $e^{\pm 1.96\sqrt{2/M}}$ for $M = 78$ and $M = 6$, and the relative width for $M = 78$.
- **Report**: the formulas, the two pairs of factors and one sentence on how precise one day's RV is
- Notebook: section A1

A1 [Solved]: Solution (1/3) — Derivation

- $\mathrm{RV} = \sum_i \sigma^2 \Delta Z_i^2 = \sigma^2 \frac{1}{M} \sum_i Z_i^2 = \sigma^2 \chi^2_M / M$, so $\mathbb{E}[\mathrm{RV}] = \sigma^2$ (unbiased) and $\mathrm{Var}(\mathrm{RV}) = \sigma^4 \cdot 2M/M^2 = 2\sigma^4/M$
- $\mathbb{E}[r_i^4] = \sigma^4 \Delta^2 \mathbb{E} Z^4 = 3\sigma^4/M^2$, so $\mathbb{E}[\mathrm{RQ}] = \frac{M}{3} \cdot M \cdot 3\sigma^4/M^2 = \sigma^4$; hence $\widehat{\mathrm{Var}}(\mathrm{RV}) = 2\,\mathrm{RQ}/M$ and $\mathrm{se}(\mathrm{RV}) = \sqrt{\frac{2}{3} \sum_i r_i^4}$
- Delta method with $g(x) = \ln x$, $g'(x) = 1/x$: $\mathrm{se}(\ln \mathrm{RV}) \approx \mathrm{se}(\mathrm{RV})/\mathrm{RV} = \sqrt{2\,\mathrm{RQ}/M}/\mathrm{RV}$
- 95% interval: $\mathrm{RV} \exp(\pm 1.96\, \mathrm{se}(\ln \mathrm{RV}))$, always positive and asymmetric in levels; with constant volatility $\mathrm{RQ}/\mathrm{RV}^2 \approx 1$, so the factors are $e^{\pm 1.96\sqrt{2/M}}$
- For volatility $\sqrt{\mathrm{RV}}$, $g'(x) = 1/(2\sqrt{x})$ halves the relative standard error: the interval for volatility is narrower than the one for variance

A1 [Solved]: Solution (2/3) — Density of RV/IV and Interval Factors



MFM_ch9_seminar

A1 [Solved]: Solution (3/3) — Simulated Coverage and Interpretation

- $M = 78$: $\sqrt{2/78} = 0.160$, factors $[0.73, 1.37]$: IV lies between $0.73\,\mathrm{RV}$ and $1.37\,\mathrm{RV}$, a total relative width of 0.64
- $M = 6$ (six equal intervals over the observation period): factors $[0.32, 3.10]$: almost uninformative
- Simulated coverage of the nominal 95% interval with RQ (200 000 days): 94.2% for $M = 78$, 86.7% for $M = 6$: the 95% is an asymptotic statement
- With stochastic volatility the same studentised limit holds asymptotically, for a continuous price and negligible noise, because the convergence is stable (Chapter 9)
- **Interpretation**: one day of 5-minute data pins IV down to roughly $-27\%/+37\%$

MFM_ch9_seminar

A2 [Solved]: Noise Bias and the Optimal Frequency — Task

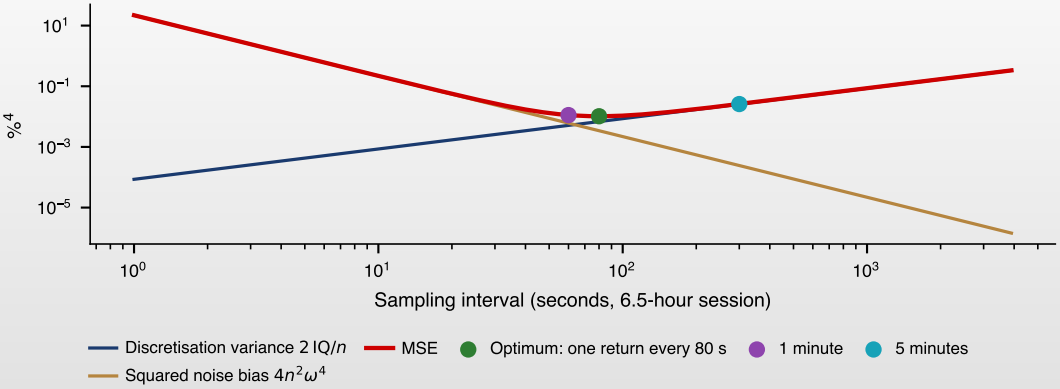
- **Question**: how often should we sample a price that contains noise?
- **Inputs**: the noise model of Noise and frequency; zero drift, constant volatility; $\mathrm{IV} = 1\,\%^2$, IQ $= 1$, $\omega = 0.01\%$, so $\omega^2 = 0.0001\,\%^2$; session of 23 400 seconds
- **Tasks**
  1. Write the observed return, expand $\mathbb{E}[r_i^2]$ and derive $\mathbb{E}[\mathrm{RV}^{(n)}] = \mathrm{IV} + 2n\omega^2$ for n returns.
  2. Derive $\mathrm{cov}(r_i, r_{i-1}) = -\omega^2$.
  3. From MSE(n) (slide Noise and frequency; Bandi & Russell (2008)), derive the MSE-optimal n^* and its sampling interval.
  4. Fill a table for 5 minutes, 1 minute, 1 second and n^*: n, bias $2n\omega^2$, $\mathbb{E}[\mathrm{RV}]$, MSE.
- **Report**: the derivations, the table and which interval you would use in practice, with the reason
- Notebook: section A2

Back Exercise list

A2 [Solved]: Solution (1/3) — Derivation

- $r_i = p_i^{\text{obs}} - p_{i-1}^{\text{obs}} = r_i^* + \varepsilon_i - \varepsilon_{i-1}$
- $\mathbb{E}[r_i^2] = \mathbb{E}[r_i^{*2}] + \mathbb{E}[\varepsilon_i^2] + \mathbb{E}[\varepsilon_{i-1}^2] = \mathbb{E}[r_i^{*2}] + 2\omega^2$ (cross terms have mean 0 by independence); summing over n returns: $\mathbb{E}[\mathrm{RV}^{(n)}] = \mathrm{IV} + 2n\omega^2$
- Adjacent returns share only ε_{i-1}, with opposite signs: $\mathrm{cov}(r_i, r_{i-1}) = \mathrm{cov}(-\varepsilon_{i-1}, \varepsilon_{i-1}) = -\omega^2$
- $\frac{d}{dn}\mathrm{MSE} = -2\,\mathrm{IQ}/n^2 + 8n\omega^4 = 0$ gives $n^3 = \mathrm{IQ}/(4\omega^4)$, $n^* = (\mathrm{IQ}/(4\omega^4))^{1/3}$; the second derivative is positive: a minimum
- Numbers: $n^* = (1/(4 \times 10^{-8}))^{1/3} = 292$, one return every $23\,400/292 = 80$ seconds

A2 [Solved]: Solution (2/3) — MSE against the Sampling Interval



MFM_ch9_seminar

A2 [Solved]: Solution (3/3) — Bias and MSE at Four Frequencies

Interval	n	$2n\omega^2$	$\mathbb{E}[\mathrm{RV}]$	MSE
5 minutes	78	0.0156	1.0156	0.0259
1 minute	390	0.0780	1.0780	0.0112
1 second	23 400	4.6800	5.6800	21.9025
n^*, every 80 s	292	0.0585	1.0585	0.0103

- $\mathrm{MSE} = 2/n + 4n^2 \times 10^{-8}$ in $\%^4$; at one second RV is almost six times IV
- In this model 1 minute is close to the optimum; 5 minutes has 2.5 times the minimal MSE
- **Interpretation**: the usual 5-minute choice rests on arguments outside this model
  - ω is uncertain; real noise is not i.i.d.
  - 5-minute RV is hard to beat empirically Liu, Patton & Sheppard (2015)

MFM_ch9_seminar

B1 [Solved]: Which Sampling Frequency? — Task

- **Prerequisites and examples**: A2 [Solved]
- **Questions**
  - does SPY show a measurable noise bias at 5 minutes?
  - which grid gives the most useful daily measure?
- **Data**: the SPY price matrix `P` of the setup (1 489 days); grid of $k \times 5$ minutes, $k = 1, 3, 6, 13$ (5, 15, 30, 65 minutes): prices at columns $o, o+k, o+2k, \ldots$ plus the open and the close, for offsets $o = 0, \ldots, k-1$ (helper `sparse_points`)
- **Tasks**
  1. Compute daily RV on one grid ($o = 0$) and the average over the k shifted grids, and report $\sqrt{252\,\overline{\mathrm{RV}}}$ for each interval.
  2. For each interval compute the ratio of mean one-grid RV to mean 5-minute RV and a 95% moving-block bootstrap interval (Grids and bootstrap): 20-day blocks of the paired daily series, 2 000 resamples.
  3. Compute the correlation of today's log RV (one grid and averaged grids) with tomorrow's 5-minute log RV.
- **Report**: one chart and one table with a row per interval: volatility, ratio with its confidence interval (CI), correlation
- **Interpretation**
  - is there a measurable noise bias at 5 minutes?
  - why a block bootstrap and not an i.i.d. one?
  - which grid is most useful for forecasting?
- Notebook: section B1

Back Exercise list

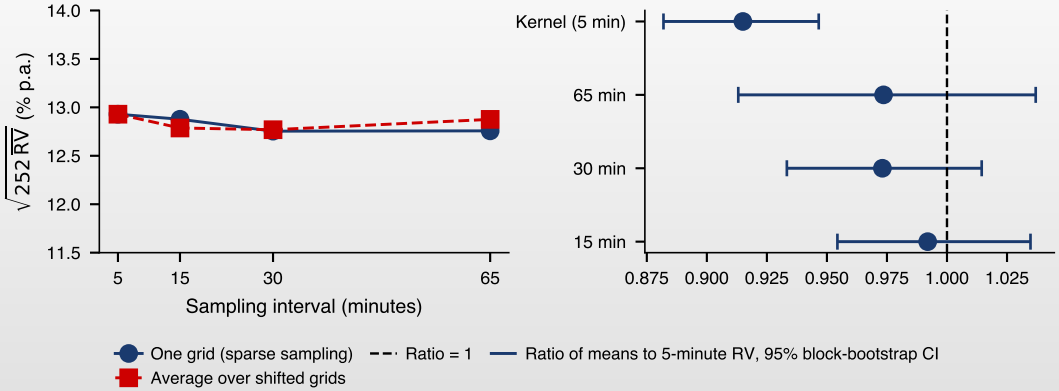
B1 [Solved]: Solution (1/3) — Procedure and Table

- Procedure: for each day take the grid prices, compute $\mathrm{RV}^{(o)} = \sum (100\,\Delta \ln P)^2$ on each offset o, average over o; the bootstrap resamples the pairs $(\mathrm{RV}_t^{(k)}, \mathrm{RV}_t^{(\mathbf{5\ min})})$ in 20-day blocks and recomputes $\sum \mathrm{RV}^{(k)} / \sum \mathrm{RV}^{(\mathbf{5\ min})}$

Interval	Vol. (one grid)	Vol. (avg. grids)	Ratio [95% CI]	Corr. next day	Corr. (avg.)
5 min	12.93	12.93	1.000 [1.000, 1.000]	0.73	0.73
15 min	12.88	12.79	0.992 [0.954, 1.035]	0.71	0.72
30 min	12.75	12.77	0.973 [0.933, 1.014]	0.70	0.72
65 min	12.76	12.88	0.974 [0.913, 1.037]	0.64	0.70

- Block bootstrap: daily RVs are strongly autocorrelated; i.i.d. resampling would destroy this dependence and make the interval too narrow

B1 [Solved]: Solution (2/3) — Signature Plot and Kernel Ratios



MFM_ch9_seminar

B1 [Solved]: Solution (3/3) — Noise Bias and Predictive Correlation

- All intervals contain 1: the coarse grids detect no bias relative to 5-minute RV; the signature is flat
- The predictive correlation falls from 0.73 (5 minutes) to 0.64 (65 minutes, one grid): sparse RV is noisier; averaging the shifted grids recovers part of it (0.70)
- **Interpretation**: 5-minute RV is the most useful measure here
- The realised kernel, a noise-robust estimator (extension pack), averages 0.915 of the 5-minute RV (95% CI $[0.882, 0.947]$): a difference between two estimators, not necessarily a noise bias

MFM_ch9_seminar

B2 [Solved]: Jump Days in SPY — Task

- **Questions**
  - on which SPY days does the ratio test reject?
  - how many rejections would chance alone produce?
- **Data**: the SPY return matrix `R` (1 489 days; $M_t = 78$, or 42 on early-close days); ratio test Huang & Tauchen (2005) (Bipower variation and Ratio jump test); helper `jump_test`
- **Tasks**
  1. For every day compute RV, BV, TQ, z_t and the one-sided $p_t = 1 - \Phi(z_t)$; work 27 August 2021 step by step.
  2. Count the days that reject at 5%, 1% and 0.1% and compare each count with αN.
  3. On the 0.1% days find the interval with the largest $|r_{t,i}|$ and plot the histogram of its end time.
  4. Compute the jump share $\sum_t J_t / \sum_t \mathrm{RV}_t$, with $J_t = \mathrm{RV}_t - \mathrm{BV}_t$ on 0.1% days and 0 otherwise.
- **Report**: the worked day, a table of counts against αN, the histogram and the jump share
- **Interpretation**
  - are there more rejections than the level alone would produce?
  - what could explain them?
- Notebook: section B2

B2 [Solved]: Solution (1/3) — One Day Step by Step

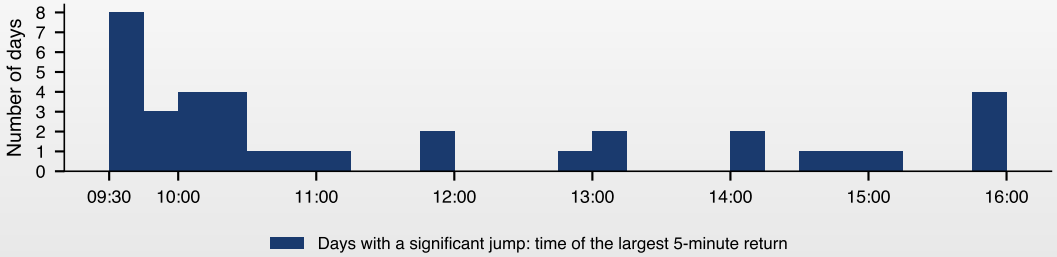
- 27 August 2021, $M = 78$: $\mathrm{RV} = 0.1354$, $\mathrm{BV} = 0.0687$, $\mathrm{TQ} = 0.00539$ (%^{2}, %^{2}, %^{4})
- $(\mathrm{RV} - \mathrm{BV})/\mathrm{RV} = 0.493$; $\mathrm{TQ}/\mathrm{BV}^2 = 1.143 > 1$, so the floor does not bind
- $z = 0.493/\sqrt{0.609 \times 1.143/78} = 5.22$, $p = 9.0 \times 10^{-8}$: reject at 0.1%
- The largest return, $+0.24$% in the interval ending 10:05, is 44% of RV: 10:00–10:05 is when the Federal Reserve chair's Jackson Hole speech was released
- The same four steps, vectorised over days, are `jump_test(R)`

B2 [Solved]: Solution (2/3) — Counts and Jump Share

Level	Rejections	αN
5%	236	74.5
1%	110	14.9
0.1%	36	1.5

- z_t has mean 0.50 and standard deviation 1.15 instead of 0 and 1: far more rejections than the level implies
- Jump share of total RV with 0.1% days: 2.8%
- Explanations: real jumps, and the intraday U-shape, which biases BV down; the extension pack applies false discovery rate (FDR) control and the periodicity correction

B2 [Solved]: Solution (3/3) — When Do the Largest Returns Occur?



- Time of the largest $|r|$ on flagged days: it peaks after the open; attributing a flag to news needs separate evidence

MFM_ch9_seminar

A Large Return and Evidence of a Jump

- **What do you think**: does one large return on a calm day prove a jump?
- **Answer**: no; the jump test asks whether a continuous price is implausible, given the sampling frequency
- The test statistic cannot exceed $\sqrt{M/\theta}$: $\mathrm{BV} \geq 0$ and the maximum in the denominator is at least 1
- Compare this cap with the critical value before interpreting a non-rejection
- RV, BV and TQ use the same return unit; the test statistic does not depend on it

A3 [Proposed]: Six Returns and One Jump — Task

- **Prerequisites and examples**: B2 [Solved] (one day step by step) and A1
- **Question**: can six 5-minute returns reveal a jump?
- **Inputs**: the returns in the table (%); the last one is a news jump; $\mathrm{TQ} = 0.0097$ (supplied); $\theta = 0.609$; critical value at 0.1%: 3.09; formulas: Bipower variation and Ratio jump test
- **Tasks**
  1. Complete the table and compute RV and BV with $M = 6$.
  2. Compute the jump part $\mathrm{RV} - \mathrm{BV}$ and its share of RV.
  3. Compute $\mathrm{TQ}/\mathrm{BV}^2$, apply the floor $\max(1, \mathrm{TQ}/\mathrm{BV}^2)$ and compute the ratio statistic z.
  4. Decide at 0.1% and compare z with its largest possible value $\sqrt{M/\theta}$.
- **Report**: the completed table, RV, BV, z, the decision and one sentence of interpretation

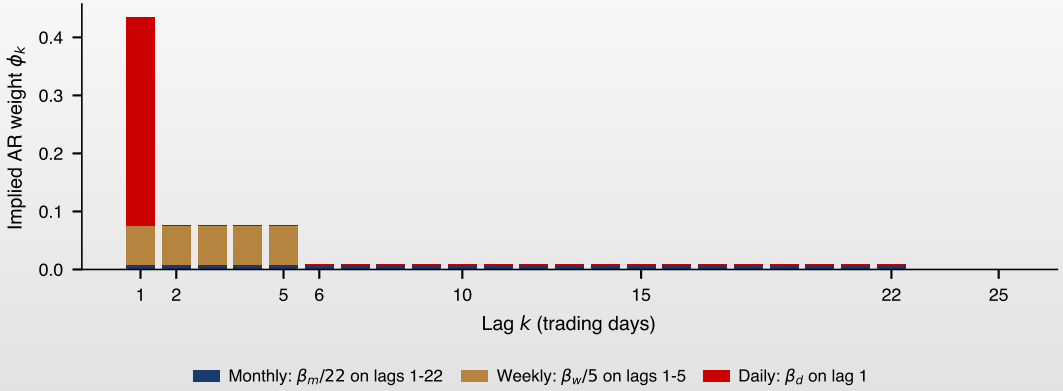
i	r_i	r_i^2	$\lvert r_i r_{i-1} \rvert$
1	0.08	0.0064	–
2	−0.15		
3	0.03		
4	0.10		
5	−0.04		
6	−0.90		

A4 [Solved]: HAR as a Restricted AR(22) — Task

- **Question**: which restrictions does HAR impose on an AR(22) for log variance?
- **Inputs**: the log-HAR of The HAR model Corsi (2009), for $y_t = \ln V_t$: daily, weekly (5-day) and monthly (22-day) means of y; rounded SPY estimates $\hat{\beta} = (-0.08, 0.36, 0.34, 0.17)$
- **Tasks**
  1. Expand the two means and write the implied AR coefficients ϕ_1, $\phi_2 = \ldots = \phi_5$ and $\phi_6 = \ldots = \phi_{22}$.
  2. With the vector $\phi = (\phi_0, \phi_1, \ldots, \phi_{22})$, write two rows of the matrix R in $R\phi = 0$ and count the restrictions.
  3. Compute ϕ_1, $\phi_{2..5}$, $\phi_{6..22}$ and the persistence $\sum_k \phi_k$ for $\hat{\beta}$, and draw the 22 weights.
- **Report**: the three weight formulas, two rows of R, the numbers and what the persistence implies for shocks
- Notebook: section A4

Back Exercise list

A4 [Solved]: Solution (1/2) — Implied AR(22) Weights



MFM_ch9_seminar

A4 [Solved]: Solution (2/2) — Restrictions and Persistence

- Lag 1 is in all three components: $\phi_{\mathbf{1}} = \beta_d + \beta_w/5 + \beta_m/22$; lags 2–5 in the weekly and monthly ones: $\phi_k = \beta_w/5 + \beta_m/22$; lags 6–22 only in the monthly one: $\phi_k = \beta_m/22$
- Rows of R (column 0: intercept): $(0, 0, 1, -1, 0, \ldots, 0)$ for $\phi_{\mathbf{2}} = \phi_{\mathbf{3}}$; $(0, \ldots, 0, 1, -1, 0, \ldots)$ at columns 6, 7 for $\phi_{\mathbf{6}} = \phi_{\mathbf{7}}$
- Count: 3 equalities among $\phi_{\mathbf{2}}, \ldots, \phi_{\mathbf{5}}$ plus 16 among $\phi_{\mathbf{6}}, \ldots, \phi_{\mathbf{22}}$: 19 rows, R is 19×23 (22 lags minus 3 free slopes)
- $\phi_{\mathbf{1}} = 0.36 + 0.34/5 + 0.17/22 = 0.436$; $\phi_{\mathbf{2..5}} = 0.076$; $\phi_{\mathbf{6..22}} = 0.008$; $\sum_k \phi_k = \beta_d + \beta_w + \beta_m = 0.87$
- **Interpretation**: persistence $0.87 < 1$: shocks to log variance decay slowly but die out; the data test of the 19 restrictions is in the extension pack

MFM_ch9_seminar

B3 [Solved]: log-HAR for SPY with Robust Inference — Task

- **Prerequisites and examples**: A4 [Solved]
- **Questions**
  - how persistent is SPY's log total variance?
  - do robust standard errors change any conclusion?
- **Data**: SPY total daily variance from the setup, $V_t = \mathrm{RV}_t + r^2_{\mathrm{on},t}$, 1 489 days; regressors dated $t-1$: $\ln V_{t-1}$ and the means of $\ln V$ over days $t-5..t-1$ and $t-22..t-1$ (mean of logs); the first 22 days are lost
- **Tasks**
  1. Build the design matrix and check your row for 8 April 2025 against the solution.
  2. Estimate log-HAR by OLS; report the usual and the Newey–West (HAC) standard errors with 7 lags, the rule of Regression inference.
  3. Test $\beta_w = \beta_m$ with a Wald test using the Newey–West covariance.
  4. Run Ljung–Box tests with 22 lags on $\ln V_t$ and on the residuals, plot both ACFs, and compare R^2 with an AR(1); for the residuals, obtain the p-value from a bootstrap that re-estimates HAR.
- **Report**: the coefficient table with both standard errors, the tests with p-values and the chart
- **Interpretation**
  - which component matters most?
  - do robust errors change a conclusion?
  - has HAR captured the persistence?
- Notebook: section B3

Back Exercise list

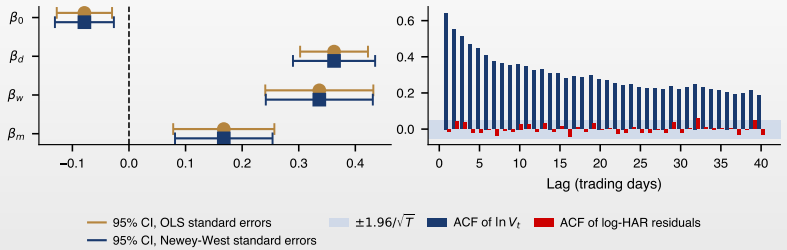
B3 [Solved]: Solution (1/2) — Design Row and Estimates

- Row for 8 April 2025: $y_t = \ln V_t = 2.976$; regressors $\ln V_{t-\mathbf{1}} = 4.022$ (7 April), weekly mean 2.068, monthly mean 0.612
- Fitted value $\hat{y}_t = 2.177$: the model expects a partial return towards the monthly level
- Usable observations: $1\,489 - 22 = 1\,467$; $R^{\mathbf{2}} = 0.470$ against 0.413 for an AR(1) ($\hat{\phi} = 0.643$); persistence 0.866

	$\hat{\beta}$	SE OLS	SE NW	$t_{\mathbf{NW}}$
$\beta_{\mathbf{0}}$	−0.079	0.025	0.027	−3.0
β_d	0.362	0.031	0.037	9.8
β_w	0.336	0.049	0.048	7.0
β_m	0.168	0.046	0.044	3.8

- Newey–West standard errors (SE) are up to 20% larger than the OLS ones; all slopes stay significant

B3 [Solved]: Solution (2/2) — Tests and Chart



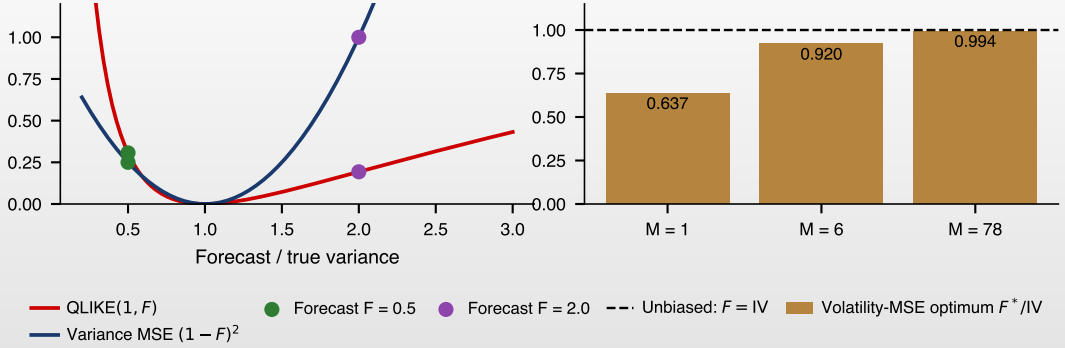
- Wald $\beta_w = \beta_m$: 4.13, $p = 0.042$: rejected at 5%, the weekly weight is larger
- Ljung–Box(22): $\ln V_t$ 4 637, residuals 17.6: three coefficients absorb the persistence; first-order residual autocorrelation −0.014
- Residual $Q(22)$: wild bootstrap (residuals with random signs) that re-estimates HAR (499 samples), $p = 0.59$; the χ^2_{22} reference ($p = 0.73$) ignores the estimated dynamics and the heteroscedasticity
- **Interpretation**: the weekly and daily components dominate; robust errors do not change a conclusion; HAR leaves no significant autocorrelation

MFM_ch9_seminar

A5 [Solved]: Which Loss Ranks Variance Forecasts? — Task

- **Question**: which loss ranks variance forecasts correctly when the target is a noisy proxy?
- **Inputs**: the conditionally unbiased proxy V_t and the losses QLIKE and variance MSE of Forecast losses Patton (2011)
- **Tasks**
  1. Minimise $\mathbb{E}[\text{QLIKE}(V_t, F) \mid \mathcal{F}_{t-1}]$ over F and show that the minimiser is $\mathbb{E}[\text{IV}_t \mid \mathcal{F}_{t-1}]$.
  2. For two forecasts F_A, F_B show that $\Delta L = V(1/F_A - 1/F_B) + \ln(F_A/F_B)$ is linear in V, so its conditional mean is the same with V_t or with IV_t: the ranking is preserved.
  3. With true variance 1, compute QLIKE and variance MSE for $F = 0.5$ and $F = 2$ and compare the two rankings.
- **Report**: the first-order condition, the ranking argument, four loss values and which loss you would use
- Notebook: section A5

A5 [Solved]: Solution (1/2) — QLIKE and MSE Loss Curves



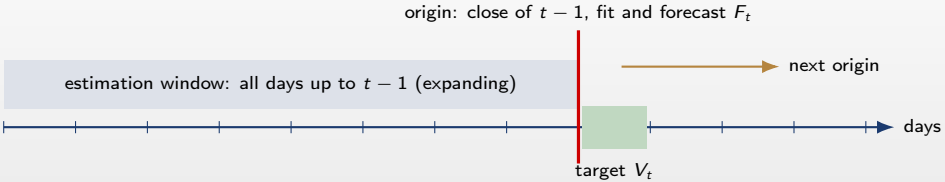
MFM_ch9_seminar

A5 [Solved]: Solution (2/2) — First-Order Condition and Ranking

- $\partial_F \mathbb{E}[V/F - \ln V + \ln F - 1] = -\mathbb{E}V/F^2 + 1/F = 0$ gives $F = \mathbb{E}[V_t \mid \mathcal{F}_{t-1}] = \mathbb{E}[\mathrm{IV}_t \mid \mathcal{F}_{t-1}]$; second derivative $1/F^2 > 0$
- $\mathrm{QLIKE}(V, F_A) - \mathrm{QLIKE}(V, F_B) = V(1/F_A - 1/F_B) + \ln(F_A/F_B)$: linear in V, and F_A, F_B are known at $t-1$, so replacing V_t by the conditionally unbiased IV_t leaves $\mathbb{E}[\Delta L \mid \mathcal{F}_{t-1}]$ unchanged
- QLIKE: $F = 0.5$ gives $2 - \ln 2 - 1 = 0.307$, $F = 2$ gives $0.5 + \ln 2 - 1 = 0.193$; variance MSE: 0.25 and 1, the opposite ranking
- Chart, right: the MSE-on-volatility optimum F^*/IV for $M = 1, 6, 78$ (A5 Extended)
- **Interpretation**: both losses are robust, but QLIKE penalises under-prediction more, which matters for risk; MSE on *volatility* is not robust (extension pack)

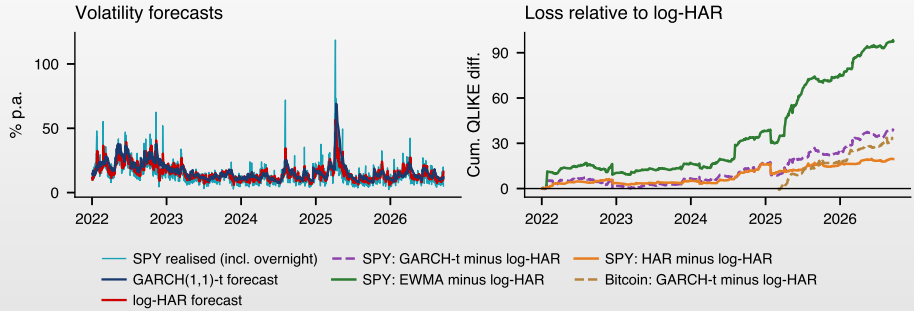
MFM_ch9_seminar

Forecast Evaluation: Origin, Target, Window



- The out-of-sample scheme of Out-of-sample forecasts: no value from day t enters the forecast F_t
- SPY: 1 180 origins from 3 January 2022 to 18 September 2026

One-Day Forecasts and Cumulative Losses



- Mean QLIKE: log-HAR 0.310, GARCH(1,1)-t 0.343; a rising cumulative line means log-HAR has the lower loss
- d_t = QLIKE(log-HAR) − QLIKE(GARCH) is the input of A6: is the gap more than chance?

MFM_ch9_forecasts

A Joint Calibration Test: A Numerical Example

- Toy estimates from the Mincer–Zarnowitz (MZ) regression: $\hat{a} = 0.10$, $\mathrm{SE}(\hat{a}) = 0.05$, $\hat{b} = 0.90$, $\mathrm{SE}(\hat{b}) = 0.08$, correlation $\rho = -0.5$. Test $a = 0$, $b = 1$.
- The covariance is $-0.5(0.05)(0.08) = -0.002$. Standardised deviations are $t_a = 2$, $t_b = -1.25$.
- $W = (t_a^2 + t_b^2 - 2\rho t_a t_b)/(1 - \rho^2) = (4 + 1.5625 - 2.5)/0.75 = 4.0833$. Against χ^2_2, $p = e^{-W/2} = 0.130$: do not reject jointly at 5%.
- If $|\rho|$ is near one, the covariance matrix is nearly singular: rounding can change the conclusion. In MZ, $b > 1$ means realised variance varies more than the forecasts in the calibration regression; these coefficients are not GARCH parameters.

A6 [Proposed]: Accuracy versus Calibration — Task

- **Prerequisites and examples**: the preceding joint calibration example and B3 step 3
- **Question**: is GARCH less accurate than log-HAR, biased, or both?
- **Inputs**: SPY one-day forecasts, January 2022 – September 2026 (keep four decimals)
  - accuracy: d_t = QLIKE(log-HAR) − QLIKE(GARCH) has mean $\bar{d} = -0.0327$, Newey–West SE 0.0142
  - calibration, MZ regression for GARCH, $V_t = a + bF_t + u_t$: $\hat{a} = 0.2418$ (SE 0.2310), $\hat{b} = 0.7046$ (SE 0.2382), $\mathrm{corr}(\hat{a}, \hat{b}) = -0.9750$
- **Tasks**
  1. Accuracy, H_0: equal expected loss. Compute the Diebold–Mariano t, its two-sided p-value, and decide at 5%.
  2. Calibration, H_0: $a = 0$, $b = 1$. Compute the separate t-statistics for $a = 0$ and $b = 1$.
  3. Build the 2×2 covariance matrix, compute the joint Wald statistic W of Calibration regression and its χ^2_2 p-value.
  4. Draw the 95% confidence ellipse, mark $(0, 1)$ and explain when separate and joint tests disagree.
- **Report**: three statistics, three decisions and one sentence separating accuracy from calibration

Back Exercise list

C1 [Proposed]: Critique an AI Answer — Task

- **Prerequisites and examples**: A2 and B3 [Solved]; a code-reading exercise, no tick data are needed
- **Setting**: a student gives an AI assistant a pandas Series `px` of clean, time-sorted SPY trade prices in New York time, 09:30–16:00, several days, in USD (so returns are decimal, not %)
- **Tasks**
  1. The answer on the next slide contains three conceptual errors; everything else is correct under this setting. Find them.
  2. For each error give the evidence: a derivation, a check on the output, or a number from the lecture or this seminar.
  3. Give the corrected formula and code line. For $\mathrm{RV}_t = 10^{-4}$, report the annualised volatility printed by the original and corrected code.
  4. Log each error you actually found in the format of `AI_ERRORS.md`: request, answer, error, how found, correction.
- **Report**: a table with one row per error (what, evidence, correction) and the two annualised volatilities of task 3
- Hints: A2 on $\mathbb{E}[\mathrm{RV}^{(n)}]$ with noise; what happens between 16:00 and 09:30; plug in $\mathrm{RV}_t = 1.0 \times 10^{-4}$ ($1\%^2$)

Back Exercise list

C1 [Proposed]: Critique an AI Answer — the Answer

- **Request**: "Compute the daily realised variance, the annualised volatility and the HAR regressors from px."
- **Answer, text**
  - $\mathrm{RV}_t = \sum_i r_{t,i}^2$ with $r_{t,i}$ the intraday log returns of day t; use every tick: more returns always make RV_t more precise
  - RV_t is SPY's daily variance: compare it directly with squared close-to-close returns and with the GARCH variance
  - HAR: regress RV_{t+1} on RV_t and on its 5-day and 22-day averages Corsi (2009)

```
import numpy as np, pandas as pd
lp = np.log(px)
r = lp.groupby(lp.index.date).diff().dropna()
RV = (r**2).groupby(r.index.date).sum()
vol_ann = np.sqrt(RV) * 252
H = pd.DataFrame({'d': RV,
                  'w': RV.rolling(5).mean(),
                  'm': RV.rolling(22).mean(),
                  'y': RV.shift(-1)}).dropna()
```

Take-Home Choice

Take-home: choose one

- B4, jumps in Bitcoin (model: B2), or B7, forecasting the next week (model: B3, A6)
- Write it as a Quantlet folder (code, `Metainfo.txt`, the chart): the format the team project uses
- AI use is allowed; note the errors you actually catch, in the format of `AI_ERRORS.md`
- Extension packs are optional

Assessment criteria

- correct calculations and inference
- reproducible code
- interpretation of the results and their limits
- documented AI use

Extension Packs: What Each One Needs

Pack	Exercises	Needs
1. Noise	A2 extended [Solved], A7, B1 extended [Solved]	A2, B1
2. Jumps	A3 extended, B2 extended [Solved], B4	A3, B2
3. Distributions	B5	B1 (block bootstrap)
4. HAR	A4 extended [Solved], B3 extended [Solved], A8, B6	A4, B2, B3
5. Forecasting	A5 extended [Solved], B7, B7 extended, B8	A5, A6, the evaluation workflow
6. Roughness	B9	B1, Chapter 9 (rough volatility)
7. Projects	C2, C3	B3, A5, B8, the lecture case study

- The notebook supplies the worked examples and computational helpers; extension tasks also require your own code. Within a pack, work in order
- C2 and C3 can seed the team project

What Does Two-Scale Estimation Correct?

- Two-Scale Realised Variance (TSRV) subtracts an estimate of the noise bias from the variance averaged over sparse grids
- TSRV cancels the leading noise term and keeps the information of many observations
- Example: sparse grids 1.20, full grid 1.50, correction weight 0.20: $1.20 - 0.20 \times 1.50 = 0.90$
- The finite-sample factor corrects the remaining attenuation; state whether you applied it

A2 Extended [Solved]: Estimating the Noise Variance — Task

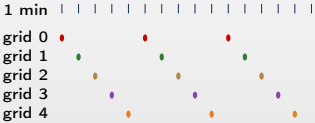
- **Questions**
  - how can ω^2 be estimated from the data?
  - in which limit is each estimator consistent?
- **Inputs**: the model of A2, n returns per day
- **Tasks**
  1. Show that $\mathbb{E}[\mathrm{RV}^{\mathrm{all}}/(2n)] = \omega^2 + \mathrm{IV}/(2n)$ and state its limit as $n \to \infty$.
  2. Propose a second estimator from the first-order autocovariance and state when it is consistent.
- **Report**: both derivations, the limit of each estimator and its consistency condition
- Notebook: section A2

A2 Extended [Solved]: Solution — Estimating the Noise Variance

- $\mathbb{E}[\mathrm{RV}^{\mathrm{all}} \mid \sigma] = \mathrm{IV} + 2n\omega^2$, so $\mathbb{E}[\mathrm{RV}^{\mathrm{all}}/(2n)] = \omega^2 + \mathrm{IV}/(2n)$; a law of large numbers for the noise terms gives $\hat{\omega}^2 = \mathrm{RV}^{\mathrm{all}}/(2n) \xrightarrow{p} \omega^2$ as $n \to \infty$ Bandi & Russell (2008)
- $-\widehat{\mathrm{cov}}(r_i, r_{i-1})$ estimates ω^2 (A2), consistent as the number of returns grows, provided the noise is i.i.d. and independent of the price
- In finite samples the first estimator is contaminated by $\mathrm{IV}/(2n)$; A7 compares it with the oracle value

A7 [Proposed]: Two-Scale RV — Task

- **Prerequisites and examples**: A2 [Solved]
- **Inputs**: a synthetic illiquid stock with a stipulated 390-minute session: $\mathrm{IV} = \mathrm{IQ} = 1$, $\omega = 0.03\%$; one day: 1-minute returns ($n = 390$) give $\mathrm{RV}^{\mathbf{all}} = 1.70$; $K = 5$ shifted 5-minute grids give $\mathrm{RV}^{\mathbf{avg}} = 1.10$



1. Compute the noise bias $2n\omega^2$ at 5 and 1 minutes, and n^* of A2 with its interval.
2. With $\bar{n} = (n - K + 1)/K$, compute TSRV $= \mathrm{RV}^{\mathbf{avg}} - \frac{\bar{n}}{n}\mathrm{RV}^{\mathbf{all}}$, with and without the factor $1/(1 - \bar{n}/n)$.
3. Estimate ω^2 twice: the oracle $(\mathrm{RV}^{\mathbf{all}} - \mathrm{IV})/(2n)$ and the feasible $\mathrm{RV}^{\mathbf{all}}/(2n)$ of A2 extended.
4. Show that TSRV removes the noise bias to first order. Optional: with $\mathrm{Var}(\text{error}) \approx aK/n + bn/K^2$, find the K that balances the terms and the rate of TSRV.

- **Report**: the biases and n^*, both TSRV values, both estimates of ω^2 and one sentence on when TSRV pays off

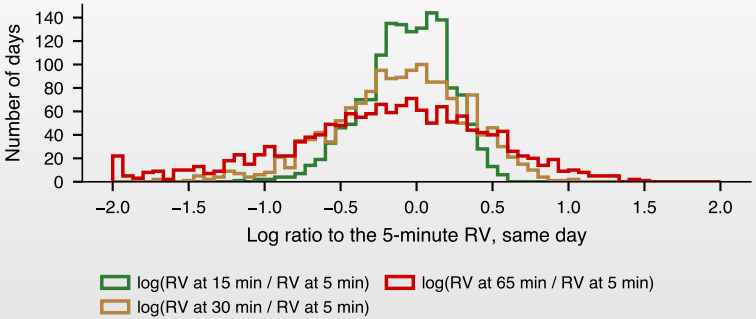
Back Exercise list

B1 Extended [Solved]: A Realised Kernel on 5-Minute Bars — Task

- **Question**: does the realised kernel, a noise-robust estimator, change SPY's 5-minute RV?
- **Data**: SPY returns `R` and prices `P`; helper `realized_kernel` Barndorff-Nielsen et al. (2008): Parzen weights, bandwidth $H = \lceil 3.5134\, \xi^{4/5} n^{3/5} \rceil$ per day ($\lceil \cdot \rceil$: round up), noise-to-signal ratio $\xi^2 = \hat{\omega}^2 / \sqrt{\widehat{\mathrm{IQ}}}$ (in practice $\sqrt{\widehat{\mathrm{IQ}}} \approx \widehat{\mathrm{IV}}$), at the ends, the session's open and close prices
- **Tasks**
  1. Compute the kernel for every day and the ratio of its mean to mean RV, with the block bootstrap of B1.
  2. Plot the distribution of $\ln(\mathrm{RV}^{(k)} / \mathrm{RV}^{(5\,\mathrm{min})})$ for 15, 30 and 65 minutes.
- **Report**: the kernel/RV ratio with its bootstrap CI, the median bandwidth H and the distribution chart
- Notebook: section B1

Back Exercise list

B1 Extended [Solved]: Solution — A Realised Kernel on 5-Minute Bars



- Kernel / RV $= 0.915$, CI $[0.882, 0.947]$ (median bandwidth $H = 7$): the kernel adds back small negative autocorrelations ($\hat{\rho}_1 = -0.008$) and averages the first and last two prices (jittering)
- Coarser grids scatter around the 5-minute value: sparse RV adds noise; the realised kernel is somewhat lower: a difference between estimators, not a measured noise bias

MFM_ch9_seminar

A3 Extended [Proposed]: Why BV Ignores a Jump — Task

- **Prerequisites and examples**: A1 [Solved]
- **Inputs**: one jump of size κ in an **interior** interval j ($1 < j < M$): $r_j = \kappa + r_j^c$; the other returns are diffusive, $O_p(\Delta^{1/2})$
- **Tasks**
  1. Show that to leading order the jump enters BV only through $|r_{j-1}|\,|\kappa| + |\kappa|\,|r_{j+1}|$, and that these products are $O_p(\Delta^{1/2})$.
  2. Conclude that $\mathrm{BV} \to \mathrm{IV}$ while $\mathrm{RV} \to \mathrm{IV} + \kappa^2$; how does a jump in the last interval (as in A3) differ?
- **Report**: the derivation, the limits of BV and RV, and the comparison with the last-interval jump in A3

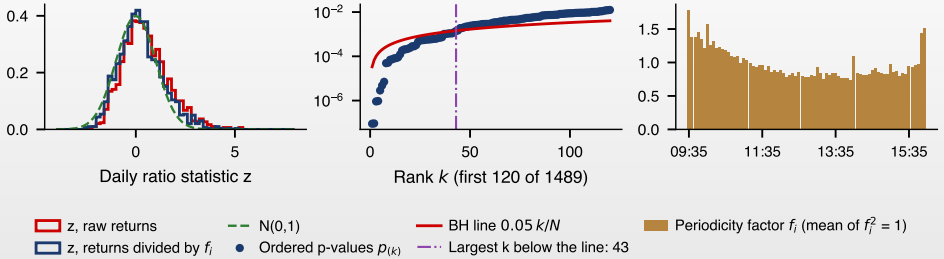
Back Exercise list

B2 Extended [Solved]: Multiplicity and Intraday Seasonality — Task

- **Question**: how many of B2's rejections survive a false-discovery-rate control and a correction for the intraday U-shape?
- **Data**: the $N = 1\,489$ daily p-values of B2; helpers `bh_count`, `periodicity_sd`
- **Tasks**
  1. Benjamini–Hochberg (BH) procedure Benjamini & Hochberg (1995): sort the p-values and reject the k smallest, k the largest index with $p_{(k)} \leq kq/N$; apply $q = 5\%$ and 1%, and Bonferroni ($p < 0.05/N$).
  2. Periodicity factor Boudt, Croux & Laurent (2011): scale returns by $\sqrt{\mathrm{BV}_t/M_t}$, take the root mean square of each 5-minute slot, normalise so that the mean of f_i^2 is 1; divide $r_{t,i}$ by f_i and repeat B2(2) and step 1.
- **Interpretation**: BH controls the FDR for valid p-values that are independent or positively regression dependent on the subset of true nulls (PRDS) Benjamini & Yekutieli (2001); which count would you use to build a jump regressor?
- **Report**: the number of rejections with BH at 5% and 1% and with Bonferroni, before and after the periodicity correction, and the count you would use for a jump regressor

Back Exercise list

B2 Extended [Solved]: Solution — Multiplicity and Intraday Seasonality



- BH at 5%: the largest k with $p_{(k)} \leq k \times 0.05/1\,489$ is $k = 43$ ($p_{(k)} = 0.00140$ against 0.00144): 43 jump days; at 1%: 14; Bonferroni: 7
- After dividing by f_i ($f = 1.78$ at the open, 0.73 at the minimum): 169 rejections at 5%, 21 at 0.1%, 17 under FDR 5%
- **Interpretation**: build the jump regressor from FDR-controlled, periodicity-corrected days

MFM_ch9_seminar

B4 [Proposed]: Jumps in Bitcoin — Task

- **Prerequisites and examples**: B2 [Solved]
- **Data**: Bitcoin price matrix `Pb`, 24-hour UTC days with at least 95% of the 288 bars, 659 days of 747; a day is labelled by its UTC date
- **Tasks**
  1. Apply the ratio test of B2 at 0.1% and build a daily table: date, M_t, RV, BV, z, weekend flag.
  2. For weekdays and weekends report the number of days, the number of rejections and their rate, and the mean BV/RV.
  3. Compute the jump share of total variation and a one-sided binomial test of the overall rate against 0.001.
- **Interpretation**: the binomial test assumes independent, correctly sized daily tests
  - what does a rejection establish?
  - what does it not establish?
  - which group rejects more often?
  - why might that group reject more often?
- **Report**: weekday/weekend table (days, rejections, rate, mean BV/RV), jump share and binomial p-value; chart of the daily jump statistic with weekends marked; four interpretation sentences

Back Exercise list

B5 [Proposed]: Returns Standardised by Realised Volatility — Task

- **Prerequisites and examples**: B1 [Solved] (block bootstrap)
- **Data**: SPY open-to-close returns (1 489 days) and Bitcoin 24-hour returns (659 days), each aligned with the same-day RV Andersen et al. (2001); one row: 7 April 2025, $r_t = 3.09\%$, $\mathrm{RV}_t = 45.1\,\%^2$, standardised return $z_t = r_t/\sqrt{\mathrm{RV}_t} = 0.46$
- **Tasks**
  1. Compute the Pearson kurtosis $\mathbb{E}(x - \bar{x})^4/s^4$ of $u_t = r_t/\mathrm{sd}(r)$ and of z_t, with 95% intervals from a paired moving-block bootstrap (20-day blocks, 2 000 resamples); these intervals are approximate: Normal-based inference on kurtosis needs moments up to order eight and weak dependence, and a finite fourth moment is not enough.
  2. Plot histograms and Normal quantile–quantile (QQ) plots of u_t and z_t; run Jarque–Bera (JB) on z_t; compute the share of $|z_t| > 3$ (Normal: 0.27%).
- **Interpretation**: where do the heavy tails come from? Optional benchmark: for i.i.d. Gaussian intraday returns z_t has kurtosis $3M/(M+2)$
- **Report**: the distribution summaries, requested test statistics and bootstrap intervals, with an interpretation of Normality

Back Exercise list

A4 Extended [Solved]: The Wald Test of the HAR Restrictions — Task

- **Prerequisites and examples**: A4 [Solved]
- **Tasks**
  1. Write the Wald statistic for $R\phi = 0$ with a Newey–West covariance $\hat{V}_{\mathrm{NW}}$ and give its limiting distribution under HAR.
  2. Explain why a HAC covariance is needed for the log-variance regression.
- **Report**: the Wald statistic, its limiting distribution and the reason for the HAC covariance

A4 Extended [Solved]: Solution — The Wald Test of the HAR Restrictions

- $W = (R\hat{\phi})^\top (R\hat{V}_{\text{NW}} R^\top)^{-1} R\hat{\phi} \to \chi^2_{19}$ under the 19 HAR restrictions
- The errors of the log-variance regression are heteroskedastic and may be autocorrelated: the usual OLS covariance would misstate $\text{Var}(R\hat{\phi})$
- The data test is B3 extended, step 1

What Changes in HAR Extensions?

- HARQ (HAR with realised quarticity) adds the realised quarticity RQ, which measures how precisely RV estimates IV
- HAR-CJ (HAR model with Continuous and Jump components) separates continuous variation (C) from jumps (J)
- SHAR (semivariance HAR) separates the variation from positive returns from that from negative returns
- The two semivariances add up to realised variance

B3 Extended [Solved]: The AR(22) Test and HARQ — Task

- **Prerequisites and examples**: A4 extended, B3 [Solved]
- **Data**: as in B3; realised quarticity RQ_t from the intraday returns (it does not cover the overnight return)
- **Tasks**
  1. Estimate the unrestricted AR(22) of $\ln V_t$ and test the 19 HAR restrictions with a Newey–West Wald test.
  2. Estimate HARQ in levels Bollerslev, Patton & Quaedvlieg (2016),
     $V_{t+\mathbf{1}} = \beta_{\mathbf{0}} + (\beta_d + \beta_Q \sqrt{\mathrm{RQ}_t}) V_t + \beta_w V_t^{(w)} + \beta_m V_t^{(m)} + \varepsilon$, on intraday RV and on total variance.
  3. Forecast from 3 January 2022 on an expanding window, replacing forecasts that fall outside the $[\min, \max]$ range of the in-sample values by the sample mean Bollerslev, Patton & Quaedvlieg (2016), and run the Diebold–Mariano (DM) test against HAR in levels under QLIKE.
- **Report**: W and its p-value, $\hat{\beta}_Q$ with standard errors on both targets, QLIKE for HARQ and HAR, and the DM test
- Notebook: section B3

B3 Extended [Solved]: Solution — The AR(22) Test and HARQ

- (1) $W = 18.2$, $p = 0.51$ (χ^2_{19}); R^2 0.477 (AR(22)) vs 0.470 (HAR): the restrictions are not rejected; HAR saves 19 parameters
- (2) $\hat{\beta}_Q = -0.0088$ (SE 0.0042) on intraday RV, significant at 5%; -0.0022 (SE 0.0024) on total variance, not significant
- (3) QLIKE 0.344 (HARQ) vs 0.326 (HAR), DM $t = 1.23$, $p = 0.22$: no significant difference
- **Interpretation**: the negative sign expected for β_Q (attenuation from measurement error, A8) appears, but with 5-minute RV the measurement error is small and the overnight return, not covered by RQ, dominates the noise in total variance

A8 [Proposed]: Attenuation and the Case for HARQ — Task

- **Prerequisites and examples**: A4 [Solved], B3 extended
- **Inputs**: $\mathrm{RV}_t = V_t^* + u_t$ (latent signal plus error), V_t^* the intraday integrated variance (IV), distinct from total variance V_t including overnight; $\mathbb{E}[u_t \mid \sigma] = 0$, $\mathrm{Var}(u_t \mid \sigma) = 2\,\mathrm{IQ}_t/M$ (asymptotic, A1); errors serially uncorrelated and uncorrelated with V^* and η; $V^*_{t+\mathbf{1}} = \beta_\mathbf{0} + \beta V_t^* + \eta_{t+\mathbf{1}}$, $\beta > 0$
- **Tasks**
  1. Using plim $\hat{\beta} = \mathrm{cov}(\mathrm{RV}_{t+\mathbf{1}}, \mathrm{RV}_t)/\mathrm{Var}(\mathrm{RV}_t)$, derive the attenuation of the OLS slope.
  2. Illustration (signal variance $\sigma^2_{V^*}$ assumed not to depend on IQ_t): derive the best linear weight on RV_t given IQ_t and show that it falls with IQ_t.
  3. Write an instrumental-variables estimator with instrument $\mathrm{RV}_{t-\mathbf{1}}$ and state its exclusion and relevance conditions.
- **Report**: the attenuation formula and its limiting cases, explaining the role of measurement uncertainty

Back Exercise list

B6 [Proposed]: Do Jumps Help? The HAR-CJ Model — Task

- **Prerequisites and examples**: B2 and B3 [Solved]
- **Data**: SPY as in B3; replication baseline: B2's 0.1% jump days, without corrections; $J_t = \mathrm{RV}_t - \mathrm{BV}_t$ on those days, 0 otherwise; $C_t = \mathrm{RV}_t - J_t + r^{\mathbf{2}}_{\mathbf{on},t}$: the intraday part classified as continuous, plus the overnight proxy, which may contain unidentified jumps (all in $\%^{\mathbf{2}}$) Andersen, Bollerslev & Diebold (2007)
- **Tasks**
  1. Regress $\ln V_{t+\mathbf{1}}$ on $\ln C_t$, $\ln C_t^{(\mathbf{5})}$, $\ln C_t^{(\mathbf{22})}$, $\ln(1+J_t)$, $\ln(1+J_t^{(\mathbf{5})})$, $\ln(1+J_t^{(\mathbf{22})})$, where (5), (22) are arithmetic means logged afterwards; report all coefficients with Newey–West SEs and test the three jump coefficients jointly.
  2. Forecast from 3 January 2022 on an expanding window, back-transform with $\exp(\hat{y} + \hat{s}^{\mathbf{2}}/2)$ as in B3, and compare QLIKE with log-HAR on common dates by DM.
  3. Sensitivity: repeat with the jump days of B2 extended (FDR control and periodicity correction), estimating the periodicity factors and running the BH selection at each forecast origin only on the data available then, or fixing both on a separate initial sample.
- **Interpretation**: HAR-CJ averages levels before logging, log-HAR averages logs: the two are not nested; why can a significant jump coefficient give no forecasting gain?
- **Report**: coefficient table, joint Wald test, forecast losses and DM comparison; state whether significance translates into forecasting gains

Back Exercise list

A5 Extended [Solved]: Why MSE on Volatility Is Not Robust — Task

- **Prerequisites and examples**: A5 [Solved]
- **Inputs**: constant IV and $\mathrm{RV} = \mathrm{IV}\,\chi^2_M/M$; supplied moment: $\mathbb{E}\sqrt{\chi^2_M} = \sqrt{2}\,\Gamma(\frac{M+1}{2})/\Gamma(\frac{M}{2})$, Γ: the gamma function
- **Tasks**
  1. Minimise $\mathbb{E}(\sqrt{\mathrm{RV}} - \sqrt{F})^2$ (MSE on volatility) over F and compare the minimiser with $\mathbb{E}\mathrm{RV}$ by Jensen's inequality.
  2. Compute F^*/IV for $M = 1$ (squared daily return), 6 and 78.
- **Report**: the optimal F^*, its comparison with $\mathbb{E}\mathrm{RV}$ and the three ratios F^*/IV

A5 Extended [Solved]: Solution — Why MSE on Volatility Is Not Robust

- $\mathbb{E}(\sqrt{\mathrm{RV}} - \sqrt{F})^2 = \mathbb{E}\mathrm{RV} - 2\sqrt{F}\,\mathbb{E}\sqrt{\mathrm{RV}} + F$, minimal at $\sqrt{F} = \mathbb{E}\sqrt{\mathrm{RV}}$; by Jensen $(\mathbb{E}\sqrt{\mathrm{RV}})^2 < \mathbb{E}\mathrm{RV}$: the optimal forecast is biased down
- $F^*/\mathrm{IV} = \frac{2}{M}\left(\Gamma(\frac{M+1}{2})/\Gamma(\frac{M}{2})\right)^2 = 0.637$ ($M = 1$), 0.920 ($M = 6$), 0.994 ($M = 78$); plotted in A5
- **Interpretation**: against squared daily returns, MSE on volatility rewards a forecast 36% too low; with 5-minute RV the distortion almost disappears

B7 [Proposed]: Forecasting the Next Week — Task

- **Prerequisites and examples**: B3 [Solved], A6, the evaluation workflow
- **Data**: SPY total variance V_t (B3) and SPY adjusted closes since 2000 for GARCH; target $Y_t = V_t + \ldots + V_{t+4}$, forecast at the close of $t-1$; 1 176 origins from 3 January 2022
- **Tasks**
  1. Direct log-HAR: regress $\ln Y$ on the B3 regressors, re-estimated at every origin using only targets fully observed by $t-1$ (the last four are not); forecast $\exp(\hat{y} + \hat{s}^2/2)$ (lognormal correction).
  2. GARCH(1,1)-t re-estimated every 21 days: with σ_t^2, ω, a, b as in Out-of-sample forecasts and the long-run variance $\bar{\sigma}^2 = \omega/(1-a-b)$, forecast $\sum_{h=0}^{4}[\bar{\sigma}^2 + (a+b)^h(\sigma_t^2 - \bar{\sigma}^2)]$.
  3. Compute QLIKE and MSE, DM on QLIKE with HAC errors, both MZ calibration coefficients (here a is the intercept and b the slope, distinct from the GARCH coefficients in step 2) with the joint test of $(a, b) = (0, 1)$, and plot the cumulative QLIKE difference.
- **Interpretation**
  - does the one-day advantage of log-HAR survive at a weekly horizon?
  - what do overlapping targets do to the DM test?
- **Report**: QLIKE/MSE table; DM statistic and p-value; MZ coefficients and joint p-value; cumulative QLIKE-difference chart and two interpretation sentences

Back Exercise list

From Conditional Losses to a Joint Test

- At one origin, instruments $Z = (1, 0.2)$ and loss difference $d = -0.1$ give $g = Zd = (-0.1, -0.02)$
- The second component tests whether past differences predict future ones
- Teaching example: $n = 100$, $\bar{g} = (0.02, -0.01)$, $\hat{\Gamma}_0 = \mathrm{diag}(0.02, 0.006)$, $\hat{\Gamma}_1 = \mathrm{diag}(0.01, 0.002)$, $\hat{\Gamma}_2 = \hat{\Gamma}_3 = \hat{\Gamma}_4 = 0$ ($\hat{\Gamma}_k$: the autocovariance of g at lag k)
- $\hat{\Omega} = \hat{\Gamma}_0 + \sum_{k=1}^{4}(\hat{\Gamma}_k + \hat{\Gamma}_k^\top) = \mathrm{diag}(0.04, 0.01)$, so $W = 100[0.02^2/0.04 + (-0.01)^2/0.01] = 2$
- Against χ^2_2, $p = e^{-1} \approx 0.368$: no rejection at 5%
- A non-rejection does not show that the methods are equally accurate in every state

B7 Extended [Proposed]: A Conditional Giacomini–White Test — Task

- **Prerequisites and examples**: the numerical example immediately before this task; B7 supplies the forecasting setup
- Sign as in A6: d_t = QLIKE(log-HAR) minus QLIKE(GARCH); a positive conditional mean of d_t favours GARCH. Under equal conditional predictive ability, the statistic of step 2 follows χ^2_2; report its p-value and the 5% decision.
- **Tasks**
  1. Re-estimate both methods on fixed-length rolling windows (log-HAR: last 250 targets; GARCH: last 1 000 returns), as the theory of Giacomini & White (2006) requires.
  2. With instruments $Z_t = (1, d_{t-5})$, known at origin $t-1$, and the loss difference d_t of the target Y_t, known after the close of $t+4$, compute $n\,\bar{g}^\top \hat{\Omega}^{-1} \bar{g}$ with $g_t = Z_t d_t$ (n: number of origins, $\bar{g}$: the mean of g, $\hat{\Omega}$: its long-run covariance); under the null g is uncorrelated beyond lag 4, so $\hat{\Omega}$ sums the unweighted autocovariances of lags 0–4.
- **Report**: the mean QLIKE losses, the Giacomini–White statistic, its p-value, the 5% decision and the unconditional test

Comparable Forecasts before the MCS Test

- The model confidence set (MCS) keeps the forecasts that the joint comparison cannot separate
- SHAR is estimated on variance levels, on an expanding window, with the first forecast on 3 January 2022
- The intercept and the weekly and monthly terms are as in HAR; the positive and negative semivariances replace the daily term
- Use only outcomes observed before the forecast date
- A forecast outside the [min, max] range of the estimation-sample variance is replaced by the sample mean, as for HARQ
- `arch.bootstrap.MCS`, with the stationary bootstrap (random block lengths, mean 20): resample whole rows, jointly for all models
- Report the elimination order and both sets, 90% and 75%; explain why a model can stay in the first but not in the second

B8 [Proposed]: The Model Confidence Set — Task

- **Prerequisites and examples**: B3 extended, B7
- **Data**: SPY one-day forecasts (log-HAR, HAR, GARCH(1,1)-t, EWMA, yesterday's RV) plus HARQ (B3 extended) and SHAR Patton & Sheppard (2015): $\beta_d V_t$ replaced by $\beta^{+}\mathrm{RS}_t^{+} + \beta^{-}\mathrm{RS}_t^{-}$, $\mathrm{RS}_t^{\pm} = \sum_i r_{t,i}^{\mathbf{2}} 1\{r_{t,i} \gtrless 0\}$, overnight return counted as one more return; 1 180 common dates
- **Tasks**
  1. Build the date $\times$ model QLIKE matrix on common dates and check that every forecast is positive.
  2. Run the MCS Hansen, Lunde & Nason (2011) with the range statistic, 20-day blocks, 10 000 resamples, seed 42: `MCS(L, size=0.10, reps=10000, block_size=20, method='R', seed=42)`.
  3. Report the elimination order and every model's MCS p-value; compare the sets at 90% (size 0.10) and 75% (size 0.25).
- **Interpretation**: why is the MCS preferable to 21 pairwise DM tests?

Back Exercise list

B9 [Proposed]: How Rough Is Volatility? — Task

- **Prerequisites and examples**: B1 [Solved] (block bootstrap), Chapter 9 (Hurst exponent H; $H < 0.5$: rough volatility)
- **Data**: SPY daily series of four measures: total variance, intraday RV, BV, 30-minute subsampled RV; $x_t = \frac{1}{2}\ln(\text{measure}_t)$; lags $\Delta = 1, \ldots, 30$ trading days Gatheral, Jaisson & Rosenbaum (2018)
- **Tasks**
  1. Compute $m(q, \Delta)$, the mean of $|x_{t+\Delta} - x_t|^q$ over all valid pairs, for $q \in \{0.5, 1, 1.5, 2, 3\}$; check $m(2, 1)$ for total variance against the direct mean of squared first differences, using the same valid pairs.
  2. For each q regress $\ln m(q, \Delta)$ on $\ln \Delta$ with an intercept: slope ζ_q; then $H =$ slope of ζ_q on q through the origin.
  3. 95% CI: moving-block bootstrap, 60-day blocks, 300 resamples, increments computed only within each block.
  4. Simulate a path as long as the SPY total-variance series, using `numpy.random.default_rng(42)` and lags 1–30. Use Gaussian random-walk increments ($N(0, 0.1^2)$, seed 42), subtract `linspace(0,z[-1],n)` from the cumulative path z (still $H = 0.5$), estimate H, then draw i.i.d. $N(0, 0.25^2)$ noise from the same generator after the increments. Add it to the detrended path without detrending again; estimate H again.
- **Interpretation**
  - is H below 0.5 for every measure?
  - what does measurement error do?
- One path is an illustration, not a size or power study
- **Report**: H estimates and confidence intervals for every measure, scaling plots, and clean/noisy simulation comparison

Back Exercise list

C2 [Proposed]: Is Bitcoin's Volatility More Predictable Than the S&P 500's? — Task

- **Data**: SPY and Bitcoin one-day forecasts, common calendar window 1 March 2025 – 18 September 2026 (567 calendar days: SPY 389 trading days, Bitcoin 422 days)
- **Tasks**
  1. Define three measures and compute them per asset: the calibration R^2 of $\ln V_t$ on $\ln F_t$; the QLIKE gain $1 - \overline{\text{QLIKE}}_{\textbf{log-HAR}} / \overline{\text{QLIKE}}_{\textbf{RV yest.}}$; the same gain over GARCH.
  2. Build a 95% CI for the R^2 difference (Bitcoin minus SPY): resample the same 28-calendar-day blocks for both assets (2 000 times), keeping each asset's own missing days.
  3. Add a dummy for a weekend target day to the Bitcoin log-HAR, use 7/30-day HAR averages and an expanding training window of all complete observations before each target date, starting on 1 March 2025; forecast again with the same lognormal correction and run DM under QLIKE with and without the dummy.
- **Report**: the table, the CI, the DM result and a two-sentence answer; project extension: Ethereum or a Bursa de Valori București (BVB) stock, and the entropy measure of Pele & Mazurencu-Marinescu-Pele (2019)

Back Exercise list

C3 [Proposed]: Does Machine Learning Beat HAR for SPY? — Project Brief (1/4)

- Context: Christensen, Siggaard & Veliyev (2023) (Chapter 9, case study)
  - random forests (RF) and neural networks beat HAR once predictors beyond past RV enter
  - the gain grows with the horizon
  - you test this on SPY, with the sample split and the model settings of the paper
- (a) Build the target series (the paper's Data description section).
  - daily open-to-close RV of SPY from the 5-minute bars of the regular session (78 returns a day, as in B2)
  - three targets: RV_{t+1}, the average RV over the next 5 days and over the next 22 days
- (b) Build two information sets dated t.
  - $\mathcal{M}_{\mathsf{HAR}}$: the daily, weekly and monthly RV
  - $\mathcal{M}_{\mathsf{ALL}}$ adds the VIX (VIX.INDX.csv, close), the squared daily log return of the Hang Seng (HSI.INDX.csv, close) and the 1-week SPY momentum
  - and the first differences of log dollar volume and of the 3-month T-bill rate (Federal Reserve Economic Data, FRED; series `DTB3`)
  - and the economic policy uncertainty index (FRED, `USEPUINXD`) and the Aruoba–Diebold–Scotti index (Philadelphia Fed)

Back Exercise list

C3 [Proposed]: Does Machine Learning Beat HAR for SPY? — Project Brief (2/4)

- (b, continued) Date each predictor by its release, not its reference date.
  - use only values published by the close of day t
  - the Aruoba–Diebold–Scotti index is revised: use the Philadelphia Fed's real-time vintages
  - disclose any series for which vintages are unavailable
  - state which predictors of the paper (Table 1) cannot be built for an index fund, and why: a data limitation, not a reason to change the method
- (c) Split the sample as in Appendix A.1 of the paper.
  - the first 1 000 usable days for training, the next 200 for validation, the rest for the test
  - a row belongs to a subset only if its whole target window lies inside it: drop the last $h - 1$ rows before each boundary
  - report the usable sample size per horizon
  - standardise every predictor with the training mean and variance only

C3 [Proposed]: Does Machine Learning Beat HAR for SPY? — Project Brief (3/4)

- (d) Estimate the linear models.
  - HAR, log-HAR and OLS on all predictors, on a rolling window over training and validation merged
  - log-HAR forecast: $\exp(\hat{f} + \hat{\sigma}^2/2)$, $\hat{f}$: fitted log RV, $\hat{\sigma}^2$: residual variance
  - elastic net on a rolling training window, without merging; λ and the mixing weight chosen on the validation set (paper, Section 2)
  - HAR keeps its three RV regressors under both information sets
- (d, continued) Estimate the machine-learning models (paper, Table A.6).
  - random forest: 500 trees on the merged window, $\lfloor J/3 \rfloor$ candidate predictors per split (J: number of predictors), minimum node size 5
  - NN_2^{10}: layers of 4 and 2 neurons, leaky ReLU (Rectified Linear Unit), Adam with learning rate 0.001, dropout 0.8 (probability of removing a neuron during training; all neurons active at prediction)
  - early stopping with patience 100; the average of the best 10 of 100 seeds; trained once on a fixed window
  - replace negative forecasts with the smallest in-sample RV

C3 [Proposed]: Does Machine Learning Beat HAR for SPY? — Project Brief (4/4)

- (e) Report the test-set comparison, as in B8.
  - a table of MSE and QLIKE relative to HAR, for the three horizons and both information sets
  - one-sided DM p-values against HAR and the 90% MCS
  - overlapping 5- and 22-day targets: HAC errors with a bandwidth above $h-1$ (Bartlett weights down-weight lag $h-1$) and a sensitivity check over bandwidths
- Interpretation
  - does the ranking of Tables 2, 3 and 7 of the paper survive on SPY?
  - does the ML (Machine Learning) advantage grow with the horizon?
  - does the ML advantage survive QLIKE?
  - with a test set of this length, what can a non-significant DM test tell you?
- Worked example to follow: B3, A5.
- Project extension
  - repeat on Bitcoin's 24-hour RV
  - compare across models the importance of RV and VIX from accumulated local effects (Figure 7 of the paper)

Acronyms used in this material (1/2)

- **ACF**: AutoCorrelation Function
- **AI**: Artificial Intelligence
- **AR**: AutoRegressive (model)
- **BH**: Benjamini–Hochberg (procedure)
- **BV**: Bipower Variation
- **BVB**: Bursa de Valori București — *the Bucharest Stock Exchange*
- **CI**: Confidence Interval
- **DM**: Diebold–Mariano (test)
- **ETF**: Exchange-Traded Fund
- **EWMA**: Exponentially Weighted Moving Average
- **FDR**: False Discovery Rate
- **FRED**: Federal Reserve Economic Data
- **FWER**: Family-Wise Error Rate
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors)
- **HAR**: Heterogeneous AutoRegressive (model)
- **HAR-CJ**: HAR model with Continuous and Jump components
- **HARQ**: HAR model with Realised Quarticity

Acronyms used in this material (2/2)

- **IQ**: Integrated Quarticity
- **JB**: Jarque–Bera (test)
- **MCS**: Model Confidence Set
- **ML**: Machine Learning
- **MSE**: Mean Squared Error
- **MZ**: Mincer–Zarnowitz (regression)
- **NW**: Newey–West (standard errors)
- **OLS**: Ordinary Least Squares
- **PRDS**: Positive Regression Dependence on a Subset (of p-values)
- **QLIKE**: Quasi-LIKElihood (loss function)
- **QQ**: Quantile–Quantile (plot)
- **RF**: Random Forest
- **RQ**: Realised Quarticity
- **RV**: Realised Variance
- **SE**: Standard Error
- **SHAR**: Semivariance HAR model
- **TQ**: Tripower Quarticity
- **TSRV**: Two-Scale Realised Variance
- **US**: United States
- **UTC**: Coordinated Universal Time (Temps Universel Coordonné)
- **VIX**: CBOE Volatility IndeX

References (1/2)

- Andersen, T.G., Bollerslev, T., Diebold, F.X. (2007). Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89(4), 701–720.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453), 42–55.
- Bandi, F.M., Russell, J.R. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, 75(2), 339–369.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6), 1481–1536.
- Barndorff-Nielsen, O.E., Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1), 1–37.
- Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1), 289–300.
- Benjamini, Y., Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4), 1165–1188.
- Bollerslev, T., Patton, A.J., Quaedvlieg, R. (2016). Exploiting the errors: a simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1), 1–18.
- Boudt, K., Croux, C., Laurent, S. (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance*, 18(2), 353–367.
- Christensen, K., Siggaard, M., Veliyev, B. (2023). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, 21(5), 1680–1727.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174–196.
- Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.

References (2/2)

- Franke, J., Härdle, W.K., Hafner, C.M. (2019). Statistics of Financial Markets: An Introduction, 5th ed. Springer.
- Gatheral, J., Jaisson, T., Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6), 933–949.
- Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578.
- Hansen, P.R., Lunde, A., Nason, J.M. (2011). The model confidence set. *Econometrica*, 79(2), 453–497.
- Huang, X., Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3(4), 456–499.
- Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3), 1217–1241.
- Liu, L.Y., Patton, A.J., Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187(1), 293–311.
- Patton, A.J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256.
- Patton, A.J., Sheppard, K. (2015). Good volatility, bad volatility: signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97(3), 683–697.
- Pele, D.T., Mazurencu-Marinescu-Pele, M. (2019). Using high-frequency entropy to forecast Bitcoin's daily value at risk. *Entropy*, 21(2), 102.
- Zhang, L., Mykland, P.A., Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472), 1394–1411.