








Modelling Financial Markets

Seminar 20: Path Signatures and 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 the Shape of the Past Help Forecast Volatility?

- Questions of the seminar
  - **does the shape of the recent volatility path help to forecast realised variance?**
  - **how do we test that honestly?**
- Deliverables
  - answers to A1–A3 and the required independent task A6
  - one figure and one table for each of B1, B2, B3 and B4; optional B5 transfers the B1 analysis to another asset
  - the corrected C1 code
  - one interpretation sentence per exercise
- Assessment criteria
  - correct derivations
  - inference with HAC (heteroskedasticity- and autocorrelation-consistent) standard errors and the alternative stated
  - one interpretation for every result
  - the AI (Artificial Intelligence) critique documented in `AI_ERRORS.md`
- Notebook for today: open the seminar notebook in Colab
  - each task names its notebook section
- Seminar: Prof. Daniel Traian Pele, danpele@ase.ro; Antoaneta Amza, PhD candidate, antoaneta.amza@csie.ase.ro

What Do the Realised Measures Measure?

- VOLARE (VOLatility Archive for Realized Estimates) provides daily measures constructed from intraday prices; each asset keeps its own trading calendar.
- JPM is the NYSE (New York Stock Exchange) ticker of JPMorgan Chase & Co.; C2 uses all 40 stocks. ES (E-mini S&P 500 futures, CME ticker; S&P: Standard & Poor's); CL (Crude oil, WTI light sweet futures, NYMEX ticker) denotes the WTI (West Texas Intermediate, crude oil benchmark) light sweet crude-oil futures (NYMEX ticker). Here ES is a contract code, not Expected Shortfall.
- FX (foreign exchange) denotes currency markets. A pair code joins two currency codes in base/quote order, e.g. EURUSD = EUR (euro)/USD: EUR (euro), GBP (British pound), AUD (Australian dollar), USD (US dollar), CAD (Canadian dollar), JPY (Japanese yen).

Data Used

Data set	Assets	Sample	Used in
VOLARE realised measures, US stocks	JPM (all 40 stocks in C2)	2 Jan 2015 – 30 Jan 2026	A7, B1–B3, C2
VOLARE realised measures, futures	ES (E-mini S&P 500), CL (crude oil)	28 Sep 2009 – 30 Jan 2026	B5–B7
Chapter 20 results (provided with the notebook)	50 assets	one-sided Diebold–Mariano (DM) p-values	B4, B8

- Variables of the VOLARE data Cipollini et al. (2026), one value per asset and trading day
  - Five-minute realised variance RV, its semivariances RV_t^+, RV_t^- and realised quarticity RQ (definitions)
  - The daily open and close prices, for the open-to-close log return r_t
- VOLARE is free for research at volare.unime.it: download the realised-variance data (stocks, futures, FX) and point the first notebook cell to them

Setup: Run and Verify

```
# notebook section "Signatures, models and tests"
# defines load_asset(kind, symbol): columns
# rv, rsp, rsn, rq and r = ln(close/open)
a = load_asset('stocks', 'JPM')
print(len(a), a.index[0].date(), a.index[-1].date())
print(pd.DataFrame({'RV x 1e4': 1e4 * a.rv,
                    'r (%)': 100 * a.r}).head(3))
```

- Expected output: 2 786 trading days, 2015-01-02 to 2026-01-30

date	RV $\times 10^4$	r (%)
2015-01-02	1.463	-0.22
2015-01-05	1.855	-2.56
2015-01-06	2.974	-2.77

- RV of 1.46×10^{-4} is a daily volatility of $\sqrt{1.46 \times 10^{-4}} = 1.2\%$
- If your count or first value differs, check the data set and the asset

Why Does the Time Scale Matter?

- The hand calculations A1–A6 use time in days (the time channel and the signature are defined on the next slides): each daily step contributes one unit to the time channel. This keeps the geometry of the examples easy to check.
- The empirical features of Chapter 20 instead rescale the time channel of every window to the interval from zero to one. A 22-observation window then has 21 equal time steps.
- Changing this scale changes every signature term containing a time coordinate. A numerical area from the hand exercise cannot be compared directly with an empirical feature built on the rescaled path.
- Keep the same convention when comparing paths or computing kernel distances. Rescale before computing the signature, not after mixing features from different conventions.

A Window of Data Is a Path

- Details (theorems, the full design, the 50-asset results): Chapter 20
- Take the last L days of a few series and join the points by straight lines: a path

$$X : [a, b] \to \mathbb{R}^d, \qquad X = \left(t,\ \ln \mathrm{RV}_t,\ \textstyle\sum_{s \le t} r_s\right)$$

  - L: window length in days (22 in the exercises); d: number of channels (coordinates), here $d = 3$; $[a, b]$: the time interval of the window
  - RV_t: realised variance of day t; r_s: open-to-close log return of day s, so the third channel cumulates intraday log returns, excluding overnight moves
- **Time augmentation**: channel 1 is time itself
  - Without it a fast move and a slow move with the same values look the same
- Increments: $\Delta_k = X_{t_k} - X_{t_{k-1}} \in \mathbb{R}^d$, the move of all channels over segment k
  - Δ_k^i: component i of Δ_k; for the time channel $\Delta_k^{\mathbf{1}} = 1$ day

The Signature

- **Signature**: all iterated integrals of the path, collected by level k Chevyrev & Kormilitzin (2016)

$$S^{i_{\mathbf{1}}\cdots i_k} = \int_{a<u_{\mathbf{1}}<\cdots<u_k<b} dX^{i_{\mathbf{1}}}_{u_{\mathbf{1}}} \cdots dX^{i_k}_{u_k}$$

  - $(i_{\mathbf{1}}, \ldots, i_k)$: a word, an ordered list of k channels from $\{1, \ldots, d\}$; $u_{\mathbf{1}} < \cdots < u_k$: ordered times; dX^i_u: move of channel i at time u

- Level 1: the total increments; level 2: the $d \times d$ matrix of moves of j weighted by how far i has already moved

$$S^i = X^i_b - X^i_a, \qquad S^{ij} = \int_a^b (X^i_u - X^i_a)\, dX^j_u$$

  - $S^{(\mathbf{1})}$, $S^{(\mathbf{2})}$: level 1 (a vector) and level 2 (a matrix); one straight segment with increment Δ: $S^{(\mathbf{1})} = \Delta$, $S^{(\mathbf{2})} = \frac{\mathbf{1}}{\mathbf{2}}\Delta\Delta^\top$
  - Depth N: levels $1, \ldots, N$ are kept (truncated signature Sig^N); level k has d^k terms

- The terms of the truncated signature serve as regression features

Chen's Identity and the Shuffle Identity

- **Chen's identity** Chen (1957): the signature of two joined paths is the tensor product of their signatures; at level 2

$$S^{(2)}(X * Y) = S^{(2)}(X) + S^{(1)}(X)\, S^{(1)}(Y)^\top + S^{(2)}(Y)$$

  - $X * Y$: path X followed by path Y (joined end to start); $S^{(1)}(X)\, S^{(1)}(Y)^\top$: the $d \times d$ matrix of products, every move of X comes before every move of Y

- Applied segment by segment to a path with increments $\Delta_1, \ldots, \Delta_m$

$$S^{(2)} = \sum_{k=1}^{m} \tfrac{1}{2} \Delta_k \Delta_k^\top + \sum_{j<k} \Delta_j \Delta_k^\top$$

  - First sum: each segment with itself; second sum: every ordered pair of segments, j before k

- **Shuffle identity**: a product of level-1 terms is a sum of level-2 terms

$$S^i S^j = S^{ij} + S^{ji}$$

  - So the symmetric part of level 2 is $\tfrac{1}{2}\Delta X^i \Delta X^j$, with $\Delta X^i = S^i$ the total change of channel i: it brings nothing new

The Lévy Area

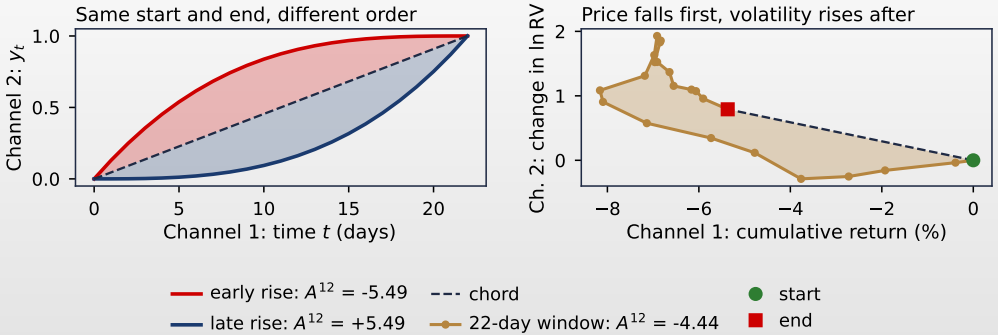
- **Lévy area**: the antisymmetric part of level 2

$$A^{ij} = \tfrac{1}{2}\left(S^{ij} - S^{ji}\right)$$

  - The signed area between the path (projected on channels i, j) and its chord, the straight line from start to end; anticlockwise > 0, clockwise < 0
  - It records the **orientation** of the path; the order of the moves follows from the sign only under an assumption on the sign of the association
  - If both channels rise, $A^{ij} > 0$ when i rises first; if i falls and then j rises, $A^{ij} < 0$, although i moves first

- Small example, channels (time, y), $y = (0, 1, 1)$: $\Delta_1 = (1, 1)$, $\Delta_2 = (1, 0)$
  - $S^{(1)} = (2, 1)$; $S^{(2)} = \tfrac{1}{2}\Delta_1\Delta_1^\top + \tfrac{1}{2}\Delta_2\Delta_2^\top + \Delta_1\Delta_2^\top = \left[\begin{smallmatrix} 2 & 0.5 \\ 1.5 & 0.5 \end{smallmatrix}\right]$; shuffle: $S^1S^2 = 2 = 0.5 + 1.5$
  - $A^{12} = (0.5 - 1.5)/2 = -0.5$: y rose early; for $y = (0, 0, 1)$ (late rise) the same steps give $A^{12} = +0.5$

The Lévy Area on Simulated Paths



- Left: two 22-day paths with the same start and end; the shaded region lies between path and chord
  - early rise: $A^{12} = -5.49$ (clockwise); late rise: $A^{12} = +5.49$; level 1 is identical, only the area tells them apart
- Right: channels (cumulative return, $\ln \mathrm{RV}$); the price falls first, volatility rises afterwards: $A^{12} = -4.44$
  - the kind of order effect that B1 tests as a predictor of future volatility

Sig-LK, Features and Kernel Weights

- Sig-LK (Signature LASSO + kernel weights): LASSO (Least Absolute Shrinkage and Selection Operator) on signatures, with the kernel weights of Gu et al. (2024), applied to next-day (or next-week, next-month) log RV
- Benchmark model: log-HAR, the HAR (Heterogeneous AutoRegressive) model estimated in logs
- Features ξ_τ of day τ: the three log-HAR terms and the signature up to depth 3 of the 22-day path (time, $\ln \mathrm{RV}$, cumulative r) ending at day τ
  - τ: a past day of the training window; t: the forecast origin, the last day observed
- Kernel weights: past days whose window looks like the latest one get more weight

$$w_\tau = \frac{e^{-\gamma \delta_\tau}}{\sum_s e^{-\gamma \delta_s}}, \qquad \gamma = \frac{c}{\mathrm{median}(\delta)}$$

  - δ_τ: squared Euclidean distance between the truncated signatures of the window ending at τ and of the window ending at t; s: runs over all training days, so the weights sum to 1
  - $\gamma \geq 0$: inverse temperature (called temperature in the paper); $c > 0$: a unit-free constant; with $c = 1$ a window at the median distance gets weight $e^{-1} = 0.37$ relative to an identical one; $\gamma = 0$ gives equal weights

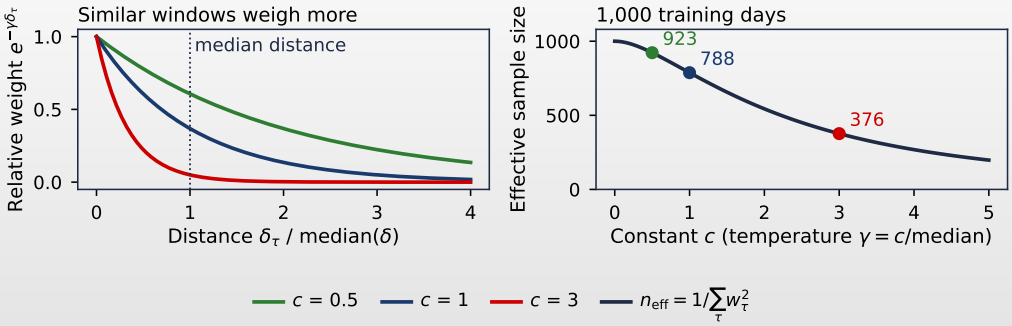
Sig-LK, Effective Sample Size and the Two-Step LASSO

- Cost of unequal weights: fewer effective observations (Kish (1965))

$$n_{\text{eff}} = \frac{1}{\sum_\tau w_\tau^2}$$

  - $n_{\text{eff}} = n$ for equal weights $1/n$; $n_{\text{eff}} = 1$ if one day takes all the weight
- Two estimation steps on the weighted training sample
  - Step 1: a weighted LASSO minimises $\sum_\tau w_\tau (y_\tau - \xi_\tau^\top \theta)^2 + \lambda \sum_j |\theta_j|$; y_τ: the log-RV target, θ: coefficients; the penalty $\lambda \geq 0$ sets some coefficients exactly to zero, i.e. selects the features
  - λ is chosen by BIC (Bayesian information criterion: fit plus a penalty per kept coefficient)
  - Step 2: weighted OLS (Ordinary Least Squares) refits the selected features only, removing the LASSO shrinkage towards zero
- Rolling design: the last 1000 complete pairs (features, target), re-estimated every 5 days; each forecast uses only data up to its origin (no look-ahead)

Kernel Weights and the Effective Sample Size

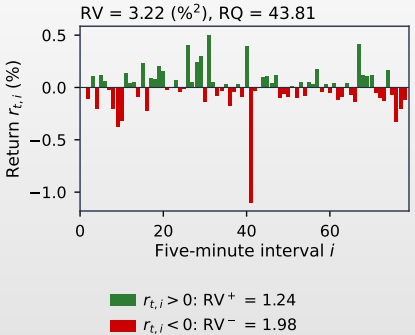


- Left: relative weight $e^{-\gamma\delta_\tau}$ of a past window at distance δ_τ, with $\gamma = c/\text{median}(\delta)$
  - at the median distance: $e^{-\mathbf{0.5}} = 0.61$, $e^{-\mathbf{1}} = 0.37$, $e^{-\mathbf{3}} = 0.05$ for $c = 0.5, 1, 3$
- Right: 1 000 simulated distances; $n_{\text{eff}} = 923$, 788, 376 for $c = 0.5, 1, 3$

Tools from Earlier Chapters, Task by Task

- B1, B2: HAR and log-HAR Corsi (2009), SHAR (Semivariance HAR) Patton & Sheppard (2015), HARQ (HAR with Realised Quarticity) Bollerslev, Patton & Quaedvlieg (2016), Newey–West (HAC) standard errors: Chapter 9
- A2, B3: QLIKE (quasi-likelihood) and why it ranks variance forecasts robustly to a noisy proxy Patton (2011): Chapter 9
- B3: Diebold–Mariano Diebold & Mariano (1995) and conditional Giacomini–White (GW) Giacomini & White (2006) tests: Chapter 8; stationary bootstrap for an interval: Chapter 5
- B4: Holm (controls the FWER, the probability of any false rejection) and Benjamini–Hochberg (controls the FDR, the share of false rejections) corrections Benjamini & Hochberg (1995): Chapter 3

Realised Variance, Semivariances and Quarticity



- Day t split into M intraday returns $r_{t,i}$ ($M = 78$ five-minute intervals)

$$\mathrm{RV}_t = \sum_{i=1}^{M} r_{t,i}^2, \qquad \mathrm{RV}_t^+ = \sum_{i=1}^{M} r_{t,i}^2 \, 1\{r_{t,i} > 0\}$$

  - RV_t estimates the variance of day t; $\sqrt{\mathrm{RV}_t}$: the daily volatility
  - $\mathrm{RV}_t^- = \mathrm{RV}_t - \mathrm{RV}_t^+$: "bad" volatility from falling prices (SHAR)
- Realised quarticity: the scale of the measurement error of RV_t (HARQ)

$$\mathrm{RQ}_t = \frac{M}{3} \sum_{i=1}^{M} r_{t,i}^4$$

  - large on days with a few big moves
  - without jumps and microstructure noise: error variance of $\mathrm{RV}_t \approx 2\,\mathrm{RQ}_t / M$
- Chart: one simulated day with one jump; $\mathrm{RV} = 3.22$, $\mathrm{RV}^- = 1.98 > \mathrm{RV}^+ = 1.24$
  - it shows the decomposition and the sensitivity of RQ to jumps, not the approximation above

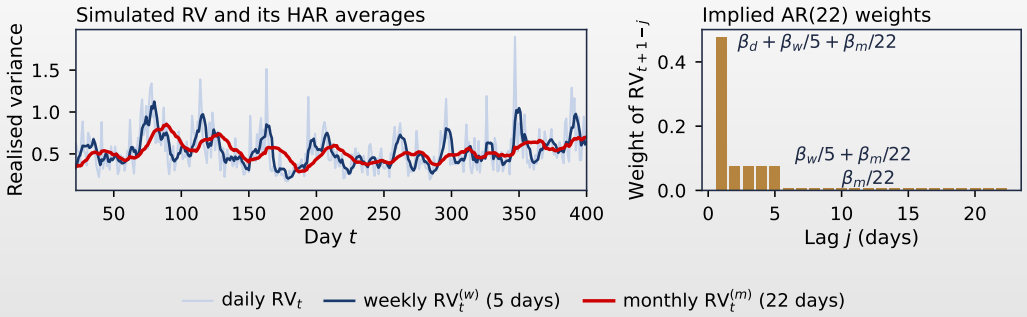
Recall, the HAR Family (B1, B2)

- HAR Corsi (2009): next-period RV is a linear function of the daily, weekly and monthly averages of past RV

$$\mathrm{RV}_{t+\mathbf{1}} = \beta_{\mathbf{0}} + \beta_d \,\mathrm{RV}_t + \beta_w \,\mathrm{RV}_t^{(w)} + \beta_m \,\mathrm{RV}_t^{(m)} + \varepsilon_{t+\mathbf{1}}$$

  - $\mathrm{RV}_t^{(w)}$, $\mathrm{RV}_t^{(m)}$: averages of RV over the last 5 and 22 days; $\beta_d, \beta_w, \beta_m$: weights of yesterday, last week, last month; $\varepsilon_{t+\mathbf{1}}$: error; log-HAR: the same in logs
- SHAR Patton & Sheppard (2015): the daily term split into the semivariances RV_t^+ (from positive intraday returns) and RV_t^- (from negative ones)
  - Coefficients β^+, β^- instead of β_d; $\beta^- > \beta^+$: bad volatility predicts more future volatility than good volatility
- HARQ Bollerslev, Patton & Quaedvlieg (2016): the weight of RV_t falls on days when RV is measured less precisely
  - Adds $\beta_Q \,\mathrm{RV}_t(\mathrm{RQ}_t^{\mathbf{1/2}} - \overline{\mathrm{RQ}^{\mathbf{1/2}}})$ to HAR; RQ_t: realised quarticity; bar: sample mean; expected $\beta_Q < 0$

The HAR Averages and Their Lag Weights



- Left: a simulated RV series with its weekly and monthly averages; the monthly one moves slowly
- Right: HAR is an autoregressive model of order 22, AR(22) with weights restricted to three steps (A7); here $\beta_d = 0.4$, $\beta_w = 0.35$, $\beta_m = 0.15$
  - lag 1: 0.477, lags 2–5: 0.077 each, lags 6–22: 0.0068 each; 4 parameters instead of 23

Newey–West (HAC) Standard Errors (B1, B2, B3)

- Regression $y_t = x_t^\top \beta + u_t$ with heteroskedastic, autocorrelated errors; long-run covariance of $x_t u_t$

$$\hat{\Omega} = \hat{\Gamma}_{\mathbf{0}} + \sum_{l=\mathbf{1}}^{q} \left(1 - \frac{l}{q+1}\right)\left(\hat{\Gamma}_l + \hat{\Gamma}_l^\top\right), \qquad \hat{\Gamma}_l = \frac{1}{T} \sum_{t=l+\mathbf{1}}^{T} \hat{u}_t \hat{u}_{t-l}\, x_t x_{t-l}^\top$$

  - q: number of lags (bandwidth); $1 - l/(q+1)$: Bartlett weights, which keep $\hat{\Omega}$ positive semidefinite

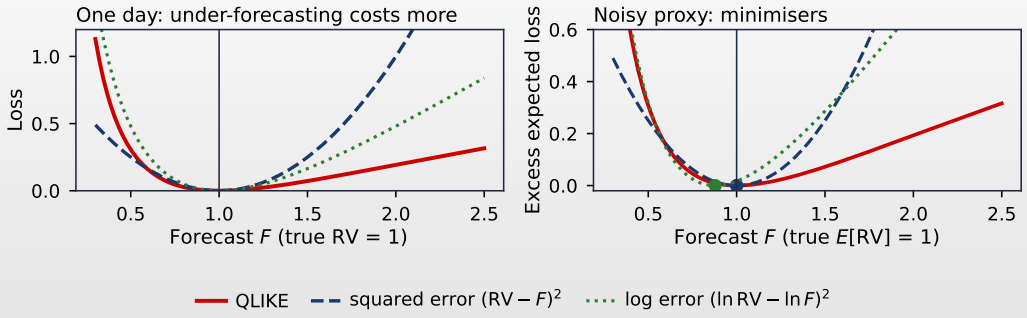
- The sandwich covariance of the OLS estimates and the t statistics

$$\widehat{\mathrm{Var}}(\hat{\beta}) = (X^\top X)^{-\mathbf{1}}\, T\hat{\Omega}\, (X^\top X)^{-\mathbf{1}}, \qquad t = \frac{\hat{\beta}_j}{\mathrm{SE}(\hat{\beta}_j)} \approx N(0,1)$$

  - for a difference of two coefficients (B2): $\mathrm{SE}(\hat{\beta}^- - \hat{\beta}^+) = \sqrt{V_{--} + V_{++} - 2V_{+-}}$
  - overlapping h-day targets create dependence in u_t up to lag $h-1$: a Moving Average (MA) structure of order $(h-1)$ if the one-day errors are not autocorrelated
  - dependence can extend beyond the overlap, and the naive variance understates the uncertainty only when the sum of the relevant autocovariances is positive

- $R^{\mathbf{2}} = 1 - \sum_t \hat{u}_t^{\mathbf{2}} / \sum_t (y_t - \bar{y})^{\mathbf{2}}$: share of the variance of y explained by the regressors

QLIKE and a Noisy Volatility Proxy (A2, B3)



- $\mathrm{QLIKE}(\mathrm{RV}, F) = \mathrm{RV}/F - \ln(\mathrm{RV}/F) - 1$; left: the three losses on one day with $\mathrm{RV} = 1$
- Right: the proxy is noisy, $\mathrm{RV} = \chi^2_8/8$ with mean 1; excess expected loss against the constant forecast F
- **What do you think**: which constant forecast minimises the log error for a noisy proxy?
  - **Answer**: $\exp(\mathbb{E}[\ln \mathrm{RV}]) = 0.88$, below the mean 1

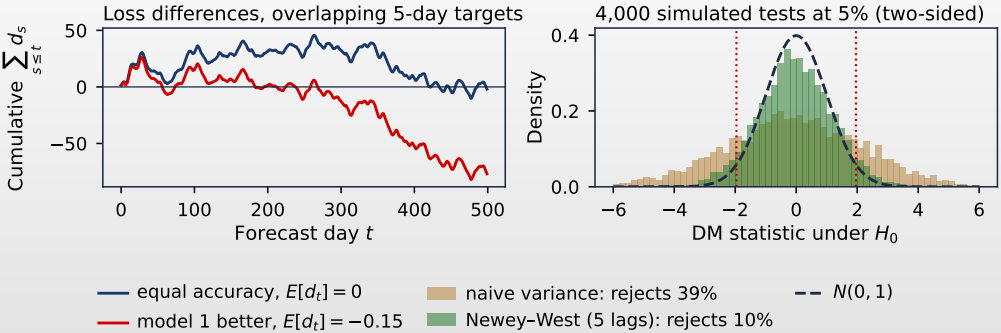
Recall, Comparing Two Forecasts (B3)

- Diebold–Mariano test Diebold & Mariano (1995): is the mean loss difference zero?

$$d_t = L_t^{(\mathbf{1})} - L_t^{(\mathbf{2})}, \qquad \mathrm{DM} = \frac{\bar{d}}{\sqrt{\hat{\Omega}_{\mathrm{NW}}/T}} \approx N(0,1) \text{ under } H_{\mathbf{0}}\colon \mathbb{E}[d_t] = 0$$

  - $L_t^{(\mathbf{1})}$, $L_t^{(\mathbf{2})}$: losses (e.g. QLIKE) of models 1 and 2 on day t; $\bar{d}$: mean of d_t over the T forecast days; $\hat{\Omega}_{\mathrm{NW}}$: Newey–West (NW) long-run variance of d_t
  - $\mathrm{DM} < -1.645$: model 1 significantly more accurate at 5% (one-sided)
- Conditional Giacomini–White test Giacomini & White (2006): can the loss difference be predicted from past information?
  - For a one-day target, regress d_t on instruments $(1, d_{t-\mathbf{1}})$; Wald statistic for both coefficients $= 0$, approximately $\chi^{\mathbf{2}}(2)$ under $H_{\mathbf{0}}$ (2 = number of instruments)
- Stationary bootstrap: resample blocks of consecutive days of random length (mean block length set by the user) to keep the serial dependence
  - The 2.5% and 97.5% quantiles of the replicated statistic give a 95% CI (confidence interval)

The Diebold–Mariano Test with Overlapping Targets



- Simulated loss differences d_t of 5-day forecasts: neighbouring days share 4 of 5 days, so d_t is autocorrelated
  - left: the cumulative sum drifts down only when model 1 is truly better
  - right: under H_0 the naive variance rejects 39% of the time instead of 5%; Newey–West brings it to 10%

Recall, Many Tests at Once (B4)

- m tests (here one per asset) with p-values sorted increasingly, $p_{(1)} \leq \cdots \leq p_{(m)}$
  - At 5% each, about $0.05\,m$ rejections are expected by chance even if no model is better (2.5 of 50)
- **Holm** controls the FWER (family-wise error rate): the probability of at least one false rejection is at most 5%

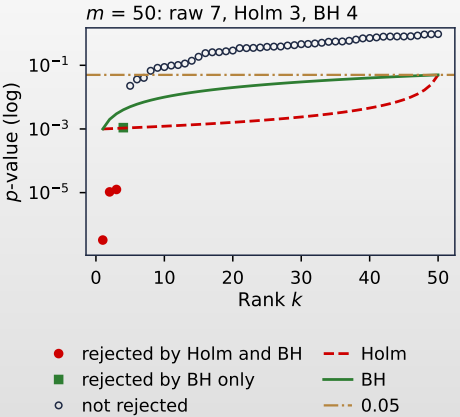
$$\text{reject while } p_{(k)} \leq \frac{0.05}{m-k+1},\ k = 1, 2, \ldots;\ \text{stop at the first failure}$$

- **Benjamini–Hochberg (BH)** Benjamini & Hochberg (1995) controls the FDR (false discovery rate): the expected share of false rejections among all rejections is at most 5%

$$k^* = \max\left\{k : p_{(k)} \leq 0.05\,\frac{k}{m}\right\};\quad \text{reject the } k^* \text{ smallest } p\text{-values}$$

  - Holm is valid under any dependence; BH needs independent or positively dependent p-values; BH rejects at least as many as Holm

Holm and BH on 50 Simulated Tests



- 50 simulated one-sided tests, 5 with a true effect
  - dots: sorted $p_{(k)}$, coloured by the decision of each procedure
  - Holm (step-down) stops at the first rank above its threshold
  - BH (step-up) finds the last rank below its threshold and rejects every rank up to it
- Rejections: unadjusted 7, Holm 3, BH 4
  - with $m = 50$ the first Holm threshold is $0.05/50 = 0.001$: only very small p-values survive
  - of the 7 unadjusted rejections, some are false: about $0.05 \cdot 45 = 2.25$ are expected from the 45 true nulls

Notation Used in Every Exercise

- RV_t: five-minute realised variance of day t; $r_t = \ln(\text{close}_t/\text{open}_t)$: open-to-close log return
- HAR averages: $\mathrm{RV}_t^{(w)} = \frac{1}{5}\sum_{j=0}^{4}\mathrm{RV}_{t-j}$, $\mathrm{RV}_t^{(m)} = \frac{1}{22}\sum_{j=0}^{21}\mathrm{RV}_{t-j}$; log-HAR regresses $\ln \mathrm{RV}_{t+1:t+h}$ on $\ln \mathrm{RV}_t$, $\ln \mathrm{RV}_t^{(w)}$, $\ln \mathrm{RV}_t^{(m)}$, with $\mathrm{RV}_{t+1:t+h}$ the mean RV over the next h days
- Worked example: $\mathrm{RV} = 1$, $F = 2$: $\mathrm{QLIKE} = 0.5 - \ln 0.5 - 1 = 0.193$

Exercise Map

Exercise	Question	Type	Model
A1	how does a signature record the order of moves?	Solved	–
A2	which forecast minimises QLIKE, which errors cost more?	Solved	–
A3	how does the inverse temperature trade similarity for sample size?	Solved	–
B1	does the Lévy area predict next month's RV of JPM?	Solved	–
B2	do the SHAR and HARQ effects hold for JPM?	Solved	–
B3	does Sig-LK beat log-HAR for JPM, one day ahead?	Solved	–
B4	how many wins survive Holm and BH over 50 assets?	Solved	–
C1	what is wrong in an artificial intelligence (AI) answer?	Proposed	A1, A2, B3
Extension packs	A4–A5, A7–A10, B5–B8, C2 (optional); A6 required	Solved and Proposed (marked on each task)	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
- Part A: derivations on paper; Part B: verification and inference on VOLARE data, each ending with an interpretation question; Part C: open questions and AI critique

A1 [Solved]: A Depth-2 Signature, Step by Step — Task

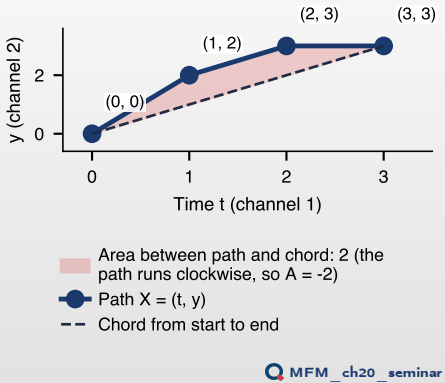
- **Question**: how does a depth-2 signature record *when* a series moved, not only how much?
- **Inputs**
  - A series takes the values $y = (0, 2, 3, 3)$ on days $t = 0, 1, 2, 3$; no data file
  - The time-augmented path $X_t = (t, y_t)$ is linear between consecutive days (channel 1 = time, channel 2 = y)
- **Tasks**
  (a) Write the three increments $\Delta_1, \Delta_2, \Delta_3 \in \mathbb{R}^2$ and compute level 1, $S^{(1)} = \sum_k \Delta_k$.
  (b) Use Chen's identity to derive the level-2 formula (squares plus ordered pairs), then compute this 2×2 matrix.
  (c) Check the result with the shuffle identity for $i = j = 1$ and for $i = 1, j = 2$.
  (d) Compute the Lévy area A^{12} and draw the path with its chord.
- **Report**: $S^{(1)}$, $S^{(2)}$ and A^{12}
- **Interpretation**: what does the sign of A^{12} say about when the series rose?
- Notebook: section A1

A1 [Solved]: Solution (1/2) — Levels 1 and 2

- (a) Increments: $\Delta_1 = (1, 2)$, $\Delta_2 = (1, 1)$, $\Delta_3 = (1, 0)$; level 1: $S^{(1)} = (3, 3)$, the total change in time and in y
- (b) One segment has $\mathrm{Sig} = (1, \Delta, \frac{1}{2}\Delta\Delta^\top, \dots)$; Chen multiplies the segments, and level 2 of the product collects every square term once and every ordered pair $j < k$ once
- Squares: $\frac{1}{2}\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}1.5 & 1.5\\ 1.5 & 2.5\end{bmatrix}$
- Ordered pairs: $\Delta_1\Delta_2^\top + \Delta_1\Delta_3^\top + \Delta_2\Delta_3^\top = \begin{bmatrix}1 & 1\\ 2 & 2\end{bmatrix} + \begin{bmatrix}1 & 0\\ 2 & 0\end{bmatrix} + \begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix} = \begin{bmatrix}3 & 1\\ 5 & 2\end{bmatrix}$
- Sum: $S^{(2)} = \begin{bmatrix}4.5 & 2.5\\ 6.5 & 4.5\end{bmatrix}$

A1 [Solved]: Solution (2/2) — Shuffle Check and Lévy Area

- (c) Shuffle, $i = j = 1$: $S^{\mathbf{1}}S^{\mathbf{1}} = 9 = 2S^{\mathbf{11}} = 2 \cdot 4.5$; $i = 1, j = 2$: $S^{\mathbf{1}}S^{\mathbf{2}} = 9 = S^{\mathbf{12}} + S^{\mathbf{21}} = 2.5 + 6.5$: both hold
- (d) Lévy area: $A^{\mathbf{12}} = (2.5 - 6.5)/2 = -2$
- Geometric check (chart): the path lies above its chord $y = t$; the shaded polygon has area 2 and is run clockwise, hence the minus sign
- Interpretation: with channels (time, y), a negative area means that y rose *early* in the window; the same net change spread evenly would give $A^{\mathbf{12}} = 0$

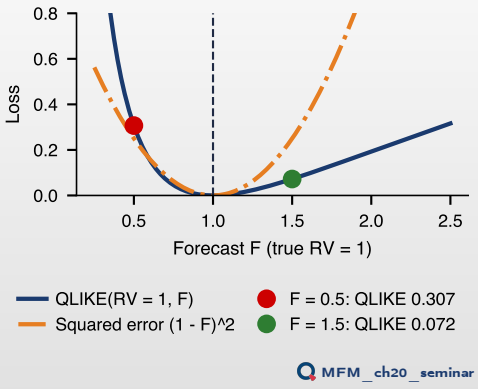


A2 [Solved]: QLIKE: Minimiser and Asymmetry — Task

- **Questions**
  - which forecast minimises expected QLIKE?
  - which errors does QLIKE punish more?
- **Inputs**: $\mathrm{QLIKE}(\mathrm{RV}, F) = \mathrm{RV}/F - \ln(\mathrm{RV}/F) - 1$, with realised proxy $\mathrm{RV} > 0$ and forecast $F > 0$; $\mathbb{E}_t$ = expectation given the information at the forecast origin
- **Tasks**
  (a) Differentiate $\mathbb{E}_t[\mathrm{QLIKE}(\mathrm{RV}, F)]$ with respect to F and solve the first-order condition.
  (b) Check the second-order condition.
  (c) For $\mathrm{RV} = 1$, compute QLIKE and the squared error for $F = 0.5$ and for $F = 1.5$, and plot both losses against F.
- **Report**: the minimiser, the four losses and the chart
- **Interpretation**
  - which error does QLIKE punish more?
  - why does this suit risk management?
- Notebook: section A2

A2 [Solved]: Solution

- (a) $\mathbb{E}_t[\text{QLIKE}] = \mathbb{E}_t[\text{RV}]/F + \ln F - \mathbb{E}_t[\ln \text{RV}] - 1$; $\partial/\partial F = -\mathbb{E}_t[\text{RV}]/F^2 + 1/F = 0 \Rightarrow F^* = \mathbb{E}_t[\text{RV}]$
- (b) $\partial^2/\partial F^2 = 2\mathbb{E}_t[\text{RV}]/F^3 - 1/F^2 = 1/F^{*2} > 0$ at F^*: a minimum; F^* depends only on $\mathbb{E}_t[\text{RV}]$, so with a conditionally unbiased proxy the *expected*-loss ranking is robust Patton (2011)
- (c) $\text{QLIKE}(1, 0.5) = 2 - \ln 2 - 1 = 0.307$; $\text{QLIKE}(1, 1.5) = 0.667 + \ln 1.5 - 1 = 0.072$; squared errors 0.25 and 0.25
- Interpretation: QLIKE punishes under-prediction of variance about four times more than the same over-prediction; under-predicted risk is the costly error (too little capital, breached VaR, value at risk)

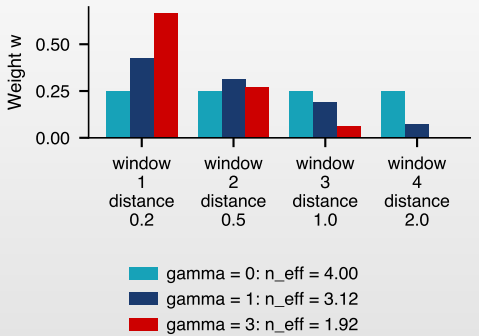


A3 [Solved]: Kernel Weights and Effective Sample Size — Task

- **Question**: how does the kernel inverse temperature γ trade similarity against the number of observations?
- **Inputs**
  - Four past windows with signature distances $\delta = (0.2, 0.5, 1.0, 2.0)$ to the current window
  - Kernel weights $w_\tau = e^{-\gamma\delta_\tau} / \sum_s e^{-\gamma\delta_s}$ Gu et al. (2024) and the effective sample size $n_{\text{eff}} = 1/\sum_\tau w_\tau^2$ Kish (1965)
- **Tasks**
  (a) Compute the four weights for $\gamma = 1$.
  (b) Compute the four weights for $\gamma = 3$.
  (c) Compute n_{eff} in both cases and compare with equal weights ($\gamma = 0$).
- **Report**: two weight vectors, two n_{eff}, a bar chart
- **Interpretation**
  - what is gained when γ increases?
  - what is lost?
- Notebook: section A3

A3 [Solved]: Solution

- (a) $e^{-\delta} = (0.819, 0.607, 0.368, 0.135)$, sum 1.928: $w = (0.425, 0.315, 0.191, 0.070)$
- (b) $e^{-\mathbf{3}\delta} = (0.549, 0.223, 0.050, 0.002)$, sum 0.824: $w = (0.666, 0.271, 0.060, 0.003)$
- (c) $n_{\text{eff}} = 4$ for $\gamma = 0$, 3.12 for $\gamma = 1$ and 1.92 for $\gamma = 3$
- Interpretation: a larger inverse temperature focuses on the most similar window (less bias from dissimilar periods)
- The cost: the information is halved (more variance)
- On the real data, the Sig-LK weights have on average $n_{\text{eff}} \approx 816$ of 1000: little information is lost



MFM_ch20_seminar

B1 [Solved]: Does the Lévy Area Predict Volatility? (JPM) — Task

- **Question**: does the order of price and volatility moves over the last month predict JPM's volatility next month, beyond log-HAR?
- **Inputs**
  - VOLARE Cipollini et al. (2026), JPM (JPMorgan Chase), 2 Jan 2015 – 30 Jan 2026, trading days only; five-minute RV, daily open and close prices
  - Regressor: A_t = Lévy area of the 22-day path (cumulative r, $\ln \mathrm{RV}$) ending at day t, divided by its full-sample standard deviation; channel order as written, no time channel
  - Target: $y_t = \ln \mathrm{RV}_{t+\mathbf{1}:t+\mathbf{22}} - \ln \mathrm{RV}_t^{(m)}$, next month's log RV minus the past month's (consecutive targets overlap in 21 of 22 days)
- **Tasks**
  (a) For every day, build the path of the last 22 days, compute its depth-2 signature and the Lévy area A_t, then standardise A_t.
  (b) Regress y_t by OLS on a constant, the three log-HAR terms and A_t, with Newey–West standard errors and 22 lags.
  (c) Test H_0: coefficient of $A_t = 0$ at 5%, two-sided, and compare the R^2 with and without A_t.
  (d) Draw a binned scatter of y_t against A_t, both residualised on the log-HAR terms.
- **Report**: coefficient, HAC SE (Standard Error), t, decision, both R^2, the chart
- **Interpretation**
  - why 22 lags?
  - is the gain economically relevant?
- Notebook: section B1

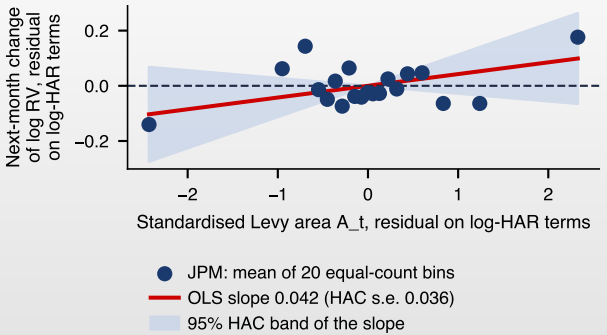
Back Exercise list

B1 [Solved]: Solution (1/2) — Steps

- (a) Signature of each 22-point window with channels (cumulative return, $\ln \mathrm{RV}$); A_t is negative in 46% of the windows (compatible with, but not identifying, leverage-type sequences)
- (b) 2 743 usable days; coefficient of $A_t = +0.042$, HAC SE 0.036
- (c) $t = +0.042/0.036 = +1.18$ (from unrounded estimates); $|t| < 1.96$: do not reject H_0 at 5%; R^2: 28.0% without A_t, 28.6% with A_t
- Why 22 lags: consecutive targets share 21 of their 22 days, so the errors are serially dependent even under H_0; persistent RV can extend the dependence, so a check with 44 lags is advisable

B1 [Solved]: Solution (2/2) — Chart and Interpretation

- (d) Chart: bins scatter around zero; the HAC band of the slope contains the horizontal line
- Size: half a point of in-sample R^2, next to the 28% explained by the log-HAR terms; economic value would need an out-of-sample evaluation
- Interpretation: for JPM the order of moves between price and volatility adds no reliable information about next month's volatility



MFM_ch20_seminar

B2 [Solved]: HAR, SHAR and HARQ for JPM — Task

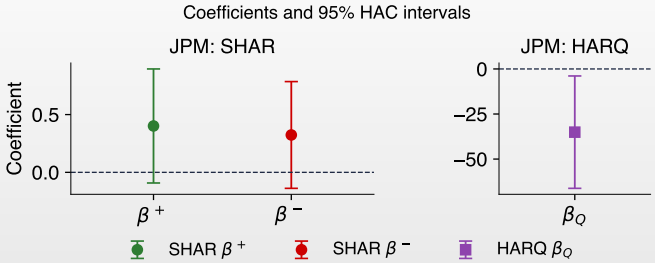
- **Question**: do bad volatility (SHAR) and the measurement error of RV (HARQ) change next-day forecasts for JPM?
- **Inputs**
  - VOLARE Cipollini et al. (2026), JPM (JPMorgan Chase), 2 Jan 2015 – 30 Jan 2026, trading days only; five-minute RV, its semivariances RV_t^+, RV_t^- and realised quarticity RQ; target: next-day RV in levels, full sample
  - SHAR Patton & Sheppard (2015): $\mathrm{RV}_{t+\mathbf{1}} = \beta_{\mathbf{0}} + \beta^+ \mathrm{RV}_t^+ + \beta^- \mathrm{RV}_t^- + \beta_w \mathrm{RV}_t^{(w)} + \beta_m \mathrm{RV}_t^{(m)} + \varepsilon_{t+\mathbf{1}}$, with β^+, β^- the weights of the positive and negative semivariance
  - HARQ Bollerslev, Patton & Quaedvlieg (2016): HAR plus the term $\beta_Q \, \mathrm{RV}_t (\mathrm{RQ}_t^{\mathbf{1/2}} - \overline{\mathrm{RQ}^{\mathbf{1/2}}})$
  - V: Newey–West (5 lags) covariance matrix of the estimates; V_{++}, V_{--}: variances of $\hat{\beta}^+$, $\hat{\beta}^-$; V_{+-}: their covariance
- **Tasks**
  (a) Estimate HAR Corsi (2009) by OLS and report the three slopes with HAC standard errors.
  (b) Estimate SHAR and test $H_{\mathbf{0}} : \beta^- = \beta^+$ against $\beta^- > \beta^+$ with $t = (\hat{\beta}^- - \hat{\beta}^+)/\mathrm{SE}(\hat{\beta}^- - \hat{\beta}^+)$, with $\mathrm{SE}(\hat{\beta}^- - \hat{\beta}^+) = \sqrt{V_{--} + V_{++} - 2V_{+-}}$.
  (c) Estimate HARQ and test $H_{\mathbf{0}} : \beta_Q = 0$ against $\beta_Q < 0$; plot β^+, β^- and β_Q with 95% HAC intervals.
- **Report**: estimates, t, p, decisions, the coefficient chart
- **Interpretation**: do the two effects of the literature hold for this stock?
- Notebook: section B2

Back Exercise list

B2 [Solved]: Solution (1/2) — Estimates and Tests

- (a) HAR (2 764 days): $\beta_d = 0.367$ (0.080), $\beta_w = 0.494$ (0.149), $\beta_m = -0.026$ (0.095)
- (b) SHAR: $\beta^+ = 0.401$ (0.252), $\beta^- = 0.324$ (0.236); difference -0.078 (from unrounded estimates), $t = -0.17$, one-sided $p = 0.566$: do not reject H_0
- (c) HARQ: $\hat{\beta}_Q = -35.062$ (15.904), $t = -2.20 < -1.645$, one-sided $p = 0.014$: reject H_0 at 5% one-sided

B2 [Solved]: Solution (2/2) — Chart and Interpretation



- Chart: the two SHAR intervals are wide and overlap (the semivariances are highly correlated for one stock); the HARQ interval lies below zero
- Interpretation: on noisy days HARQ trusts the day's RV less, and this holds for JPM; the good/bad asymmetry is a panel result Patton & Sheppard (2015), which cannot be detected on a single stock, JPM

MFM_ch20_seminar

B3 [Solved]: Sig-LK vs log-HAR for JPM, One Day Ahead — Task

- **Question**: does Sig-LK forecast JPM's next-day RV better than log-HAR?
- **Inputs**
  - VOLARE Cipollini et al. (2026), JPM (JPMorgan Chase), 2 Jan 2015 – 30 Jan 2026, trading days only; five-minute RV, daily open and close prices
  - Sig-LK design: rolling window of 1000 days, re-estimation every 5 days, path (time, $\ln \mathrm{RV}$, cumulative r) over 22 days, depth 3, two-step LASSO with BIC, inverse-temperature scale $c = 0.5$
- **Tasks**
  (a) Produce one-day forecasts of Sig-LK and log-HAR from 27 Jan 2020 to 29 Jan 2026 (1 511 days) and compute the mean QLIKE of each.
  (b) Compute the ratio from unrounded mean losses and its 95% CI with the stationary bootstrap (mean block length 10, 1000 replications).
  (c) Run the Diebold–Mariano (DM) test (HAC variance, bandwidth $\max(h, \lceil T^{1/3} \rceil)$) and the conditional GW test with instruments $(1, d_{t-1})$ Giacomini & White (2006).
  (d) Plot the cumulative loss difference $\sum_{s \le t} d_s$.
- **Report**: losses, ratio and CI, t, p, decisions, the chart
- **Interpretation**: is there any evidence that the signature adds value for JPM?
- Notebook: section B3

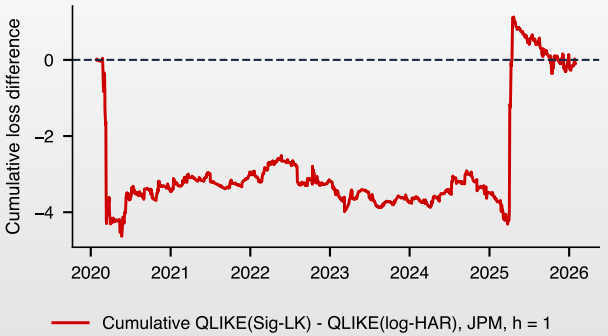
Back Exercise list

B3 [Solved]: Solution (1/2) — Losses and Tests

- (a) Mean QLIKE: Sig-LK 0.1726, log-HAR 0.1727; on average 4.2 regressors selected, effective sample size 851 of 1000
- (b) Ratio 1.000; bootstrap 95% CI $[0.956, 1.047]$ contains 1
- (c) DM: $d_t = \mathrm{QLIKE}_t^{\text{Sig-LK}} - \mathrm{QLIKE}_t^{\text{log-HAR}}$, $t = \bar{d}/\sqrt{\hat{\Omega}_{\mathrm{NW}}/T} = -0.02$, one-sided $p = 0.493$: do not reject equal accuracy
- GW: Wald statistic $2.85 \sim \chi^2(2)$, $p = 0.241$: the instruments $(1, d_{t-1})$ give no significant evidence that the loss difference is predictable

B3 [Solved]: Solution (2/2) — Cumulative Loss Difference

- (d) Upward segments: log-HAR better; downward segments: Sig-LK better; the largest moves sit in March 2020 and April 2025
- A handful of stress days decides the sign of the total: report the cumulative path, not only the mean
- Interpretation: we detect no robust advantage in the specification evaluated; close mean losses do not mean identical forecasts
- The CI for the ratio of mean losses rules out differences larger than about 5%



B4 [Solved]: Holm and BH, Step by Step — Task

- **Question**: after correcting for 50 tests, on how many assets does Sig-LK beat log-HAR one day ahead?
- **Inputs**
  - Per-asset DM p-values from Chapter 20 (provided with the notebook): one-sided DM p-values for "Sig-LK better than log-HAR" at $h = 1$ (QLIKE), one per asset; family = the 50 assets
  - The five FX pairs, sorted p-values: EURUSD (euro/dollar) 0.071, USDJPY (dollar/yen) 0.213, USDCAD (US/Canadian dollar) 0.336, AUDUSD (Australian/US dollar) 0.718, GBPUSD (pound/dollar) 0.817
  - m: number of tests; Holm (FWER 5%) and BH (FDR 5%) Benjamini & Hochberg (1995)
- **Tasks**
  (a) Apply Holm to the five FX p-values ($m = 5$).
  (b) Apply BH to the same five.
  (c) Repeat both procedures for all 50 assets, count the rejections and plot the sorted p-values against the three thresholds.
- **Report**: rejections raw, Holm, BH; the chart
- **Interpretation**
  - what may we claim?
  - about how many assets?
- Notebook: section B4

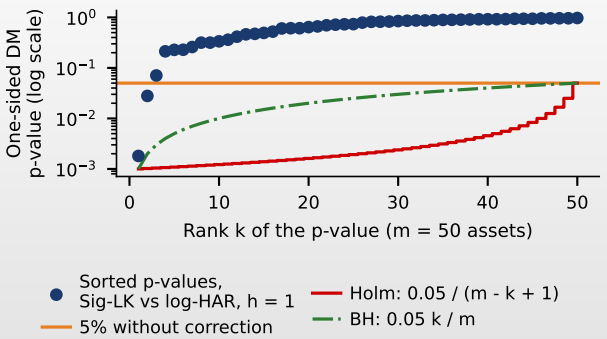
Back Exercise list

B4 [Solved]: Solution (1/2) — Steps

- (a) Holm, $m = 5$: $p_{(\mathbf{1})} = 0.071$ vs $0.05/5 = 0.01$: no rejection, stop; Holm-adjusted p: 0.354, 0.852, then 1 (from unrounded p-values)
- (b) BH: thresholds $0.01, 0.02, 0.03, 0.04, 0.05$; every $p_{(k)}$ is above its threshold: no rejection; the smallest BH-adjusted p is 0.354 (from unrounded p-values)
- (c) 50 assets: 2 raw rejections; Holm: the smallest, ES with $p = 0.002$, is above $0.05/50 = 0.001$: stop, 0 rejections; BH: $p_{(\mathbf{1})} > 0.001$ and $p_{(\mathbf{2})} = 0.028 > 0.002$, no k qualifies: 0

B4 [Solved]: Solution (2/2) — Chart and Interpretation

- Chart: only the lowest points cross the 5% line; none reaches the Holm staircase or the BH line
- Holm is valid under any dependence; BH needs independent or positively dependent *p*-values (PRDS: Positive Regression Dependence on a Subset) Benjamini & Yekutieli (2001)
- Interpretation: no asset-level improvement survives either correction; ES is at most a lead for a new, pre-registered study



MFM_ch20_seminar

C1: Data and Helper Convention

- Use the JPM RV series from the setup. The last 30% of the valid forecast origins (days with a complete 22-day window) form the evaluation split for both the candidate (the signature LASSO of the AI answer) and log-HAR; the earlier part is the training set.
- `sig_path(X, depth)` takes the rows of an array as successive points of a piecewise-linear path. It returns the flattened signature levels 1 through depth, without level 0 and without adding any channel.
- Assess both the program and its claimed conclusion: a comparison mentioned in prose must be supported by forecasts and losses computed on the same evaluation rows.

Back Exercise list

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

- **Prompt** sent to an AI assistant: "rv holds the daily 5-minute RV of JPM from the setup data. Build the signature LASSO of Gu et al. for next-day log RV and evaluate it with QLIKE on the last 30% of the sample."
- **Answer**: the code below and the conclusion

```
lrv = np.log(rv.values)
S = np.array([sig_path(lrv[t-21:t+1, None], depth=3)   # 22-day window
              for t in range(21, len(lrv) - 1)])
y = lrv[22:]                                            # next-day log RV
lam = LassoCV(cv=5).fit(S, y).alpha_                    # choose lambda
cut = int(0.7 * len(y))
m = Lasso(alpha=lam).fit(S[:cut], y[:cut])
f = np.exp(m.predict(S[cut:]))
qlike = lambda a, b: a / b - np.log(a / b) - 1
print(qlike(f, np.exp(y[cut:])).mean())
```

- **Conclusion of the AI**: "The signature LASSO has a lower QLIKE than HAR on the test set, so path signatures capture the shape of volatility and improve the forecast."

C1 [Proposed]: Critique an AI Answer — Task

Solution discussed in class

- **Question**: which errors in the answer would change the conclusion?
- **Model**: A1, A2, B3
- **Tasks**
  (a) Find at least three errors; for each, name the line and explain in one sentence why it is wrong.
  (b) For each error, describe a check that reveals it without knowing the answer (a unit test, a timing test, a known case).
  (c) Assess the exponential back-transform and whether the code is the two-step weighted method of Gu et al.
  (d) Write the corrected code and compare it with log-HAR on the same split, with a DM test.
  (e) Log each error as in `AI_ERRORS.md`: request, answer, error, how found, correction.
- **Report**: the error table (line, reason, diagnostic check), the corrected code, the entries in `AI_ERRORS.md`; on the same evaluation rows, both mean QLIKE losses, the DM statistic, its one-sided p-value and the direction of the alternative
- **Interpretation**: does the conclusion follow even if the code were correct?
- AI code usually runs without an error message: the mistakes are in the method, not in the syntax

C1: Design Choices

- ⊡ Choose the fit (fixed or rolling) and the weights (equal or kernel).
- ⊡ Select the penalty only with information available at each forecast.
- ⊡ State the channels of the path and justify every design choice in one sentence.

Extension Packs: What Each One Needs

Pack	Exercises	Prerequisites
1. Signature algebra	A4, A5 [Solved], A6 (required)	A1
2. HAR as an AR(22)	A7 [Solved], A8	B2, Wald tests
3. Losses and weights	A9, A10	A2, A3
4. Other assets	B5 (ES), B6 (CL), B7 (ES, $h = 5$)	B1, B2, B3
5. Multiple testing at one month	B8	B4
6. Project seed	C2, a level-aware kernel	B3, Chapter 20 (kernel weights)

- Each exercise has its own notebook section with starter code (an empty code cell for proposed ones); within a pack, work in the order given
- Proposed exercises name their model: the solved exercise that uses the same technique

A4 [Proposed]: Signature and Shuffle Identity — Task

Solution discussed in class

- **Questions**
  - what is the signature of a three-segment path in the plane?
  - what does signed area suggest about the order of moves?
- **Model**: A1
- **Inputs**: A path in $\mathbb{R}^2$ through $(0,0) \to (2,1) \to (3,-1) \to (4,2)$, linear between points
- **Tasks**
  (a) Write the increments and compute level 1.
  (b) Compute level 2 with the formula of A1 (squares plus ordered pairs).
  (c) Verify $S^1 S^2 = S^{12} + S^{21}$ and compute the Lévy area A^{12}.
- **Report**: $S^{(1)}$, $S^{(2)}$, A^{12}
- **Interpretation**
  - what does the sign of the area say about the tendency of channel 1 to move before channel 2?
  - why does it not imply strict precedence?
- Notebook: section A4

A5 [Solved]: Chen's Identity and Path Reversal — Task

- **Questions**
  - what is the signature of a path run there and back?
  - why does this matter for time augmentation?
- **Inputs**
  - Path $X : (0,0) \to (1,0) \to (1,2)$, with levels 1 and 2

$$S^{(1)} = (1,2), \qquad S^{(2)} = \begin{pmatrix} 1/2 & 2 \\ 0 & 2 \end{pmatrix}$$

  - $\overleftarrow{X}$ is the same path run backwards: $(1,2) \to (1,0) \to (0,0)$
- **Tasks**
  (a) Compute levels 1 and 2 of $\mathrm{Sig}(\overleftarrow{X})$ with the A1 formula.
  (b) Apply Chen's identity at level 2 to $\mathrm{Sig}(X)$ and $\mathrm{Sig}(\overleftarrow{X})$, to compute level 2 of $X * \overleftarrow{X}$ (X followed by $\overleftarrow{X}$).
  (c) State what the result implies for the uniqueness of the signature.
- **Report**: the two signatures and level 2 of the concatenation
- **Interpretation**: why do we add a time channel in practice?
- Notebook: section A5

Back Exercise list

A5 [Solved]: Solution

- (a) Reversed increments $(0,-2)$, $(-1,0)$: level 1 is $(-1,-2) = -S^{(1)}$; level 2 adds the squares and the ordered pair

$$\overleftarrow{S}^{(2)} = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 1/2 & 0 \\ 2 & 2 \end{pmatrix}$$

- (b) Chen, with the cross term $S^{(1)}(X)\, S^{(1)}(\overleftarrow{X})^{\top}$ in the middle:

$$\begin{pmatrix} 1/2 & 2 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} -1 & -2 \\ -2 & -4 \end{pmatrix} + \begin{pmatrix} 1/2 & 0 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

- Level 1 is $S^{(1)} - S^{(1)} = 0$
- (c) There and back is "tree-like": its signature is that of the constant path; the signature identifies a path only up to such pieces Hambly & Lyons (2010)
- Interpretation: a strictly increasing time channel cannot go back, so tree-like pieces are impossible and the signature determines the path up to its starting point: we time-augment and supply the level separately

A6 [Proposed]: Lévy Area of a Loop — Task

Solution discussed in class

- **Question**: what does level 2 of the signature see in a closed loop that level 1 misses?
- **Model**: A5
- **Inputs**: The unit square $(0,0) \to (1,0) \to (1,1) \to (0,1) \to (0,0)$, four linear segments
- **Tasks**
  (a) Compute level 1 and the Lévy area A^{12} for the square run anticlockwise.
  (b) Repeat for the square run clockwise, and for the square translated by $(5,5)$.
  (c) Explain why level 1 cannot distinguish the loop from a constant path while level 2 can.
- **Report**: three values of $S^{(1)}$ and A^{12}
- **Interpretation**: which invariance does the translation illustrate?
- Notebook: section A6

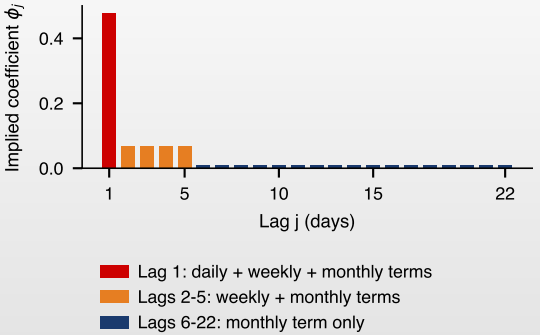
A7 [Solved]: HAR as a Restricted AR(22) — Task

- **Question**: which 22 lag coefficients does a HAR model for $\ln \mathrm{RV}$ imply for JPM?
- **Inputs**
  - VOLARE Cipollini et al. (2026), JPM (JPMorgan Chase), 2 Jan 2015 – 30 Jan 2026, trading days only
  - HAR with *averages of logs*: $\ell_t = \ln \mathrm{RV}_t$, $\ell_t^{(w)} = \frac{1}{5}\sum_{j=0}^{4} \ell_{t-j}$, $\ell_t^{(m)} = \frac{1}{22}\sum_{j=0}^{21} \ell_{t-j}$
  - The log-HAR of B1 and B3 uses logs of averages, $\ln \mathrm{RV}_t^{(w)} \neq \ell_t^{(w)}$ (Jensen), and has no exact AR form
  - Full sample (2 764 days), OLS, Newey–West (5 lags) SE: $\hat{\beta}_d = 0.408$ (0.027), $\hat{\beta}_w = 0.299$ (0.046), $\hat{\beta}_m = 0.196$ (0.040)
- **Tasks**
  (a) Substitute $\ell_t^{(w)}$ and $\ell_t^{(m)}$ into the HAR equation and collect the coefficient ϕ_j of each lag ℓ_{t-j+1}, $j = 1, \ldots, 22$.
  (b) Compute ϕ_1, $\phi_2 = \cdots = \phi_5$ and $\phi_6 = \cdots = \phi_{22}$ from the JPM estimates and plot them.
  (c) Compute the sum $\sum_{j=1}^{22} \phi_j$.
- **Report**: the three distinct values, the sum, the chart
- **Interpretation**
  - what does the sum measure?
  - is the process stationary?
- Notebook: section A7

Back Exercise list

A7 [Solved]: Solution

- (a) Using unrounded fitted coefficients:
  $\beta_w \ell_t^{(w)} = \sum_{j=1}^{5} (\beta_w/5)\ell_{t-j+1}$ and
  $\beta_m \ell_t^{(m)} = \sum_{j=1}^{22} (\beta_m/22)\ell_{t-j+1}$: lag 1 collects all three terms, lags 2–5 the weekly and monthly ones, lags 6–22 only the monthly one
- (b) $\phi_1 = 0.408 + 0.299/5 + 0.196/22 = 0.476$; $\phi_{2..5} = 0.0688$; $\phi_{6..22} = 0.0089$
- (c) $\sum_j \phi_j = \beta_d + \beta_w + \beta_m = 0.903$
- Interpretation: the sum is the persistence of log RV; all $\phi_j \geq 0$ and the sum is below 1, so the process is stationary, with a slow, step-wise decay that approximates long memory; HAR remains a short-memory process



MFM_ch20_seminar

A8 [Proposed]: Testing the HAR Restrictions — Task

Solution discussed in class

- **Question**: how do we test the restrictions that HAR imposes on an unrestricted AR(22)?
- **Model**: A7
- **Inputs**
  - The unrestricted AR(22) for $\ell_t = \ln \mathrm{RV}_t$ has an intercept and slopes $\phi_1, \ldots, \phi_{22}$; HAR with averages of logs (A7) is its restricted version
  - ϕ is the linear projection of ℓ_{t+1} on its 22 lags; weakly dependent data with finite fourth moments; errors may be heteroskedastic and serially correlated
- **Tasks**
  (a) Count the linear restrictions and write them as $R\phi = 0$.
  (b) Write the Wald statistic with an HAC covariance and its distribution under H_0.
  (c) State the 5% critical value.
- **Interpretation**: if the test rejects in a long sample, should we drop HAR for forecasting?
- Report the restriction matrix, Wald statistic, degrees of freedom and critical value.
- Notebook: section A8

Back Exercise list

A9 [Proposed]: QLIKE With Its Arguments Swapped — Task

Solution discussed in class

- **Question**: which forecasts does a QLIKE with swapped arguments reward?
- **Model**: A2
- **Inputs**: Assume $\mathrm{RV} > 0$ and $\mathbb{E}_t[1/\mathrm{RV}] < \infty$. A program computes $L(\mathrm{RV}, F) = F/\mathrm{RV} - \ln(F/\mathrm{RV}) - 1$, QLIKE with the arguments swapped
- **Tasks**
  (a) Find the forecast $F^\dagger$ that minimises $\mathbb{E}_t[L(\mathrm{RV}, F)]$.
  (b) Compare $F^\dagger$ with $\mathbb{E}_t[\mathrm{RV}]$ using Jensen's inequality.
  (c) Say which models the swapped loss favours.
- **Interpretation**: why can this bug pass elementary checks?
- Report the first-order condition, minimiser and comparison with the arithmetic-mean target.
- Notebook: section A9

A10 [Proposed]: Limits of the Temperature — Task

Solution discussed in class

- **Questions**
  - what happens to the kernel weights at the two extreme temperatures?
  - why scale by the median distance?
- **Model**: A3
- **Inputs**: n past windows, distances $\delta_\tau > 0$, inverse temperature $\gamma = c/\mathrm{median}(\delta)$
- **Tasks**
  (a) Find the limits of w and n_{eff} when $c \to 0$.
  (b) Find the limits when $c \to \infty$ (assume a unique nearest window).
  (c) Show that multiplying all distances by $k > 0$ leaves the weights unchanged.
- **Interpretation**: why is the median scaling needed to compare c across 50 assets?
- Report both weight-vector limits and effective sample sizes, with their interpretation.
- Notebook: section A10

B5 [Proposed]: The Lévy Area for E-mini S&P 500 Futures — Task

Solution discussed in class

- **Question**: does the Lévy area predict next month's volatility of E-mini S&P 500 futures (ES)?
- **Model**: B1
- **Inputs**: VOLARE futures data, 28 Sep 2009 – 30 Jan 2026, ES; five-minute RV, open and close prices; target, regressors and 22 Newey–West lags as in B1
- **Tasks**
  (a) Repeat steps (a)–(d) of B1 for ES.
  (b) Test the coefficient of A_t at 1% and 5%, two-sided.
  (c) Compare the R^2 gain with the one for JPM.
- **Report**: the regression table, both significance decisions, the change in R^2 and the binned scatter.
- **Interpretation**
  - what does a positive coefficient say about windows with a negative area?
  - can the sign alone label them leverage episodes?
- Notebook: section B5

B6 [Proposed]: HAR, SHAR and HARQ for Crude Oil — Task

Solution discussed in class

- **Question**: do the SHAR asymmetry and the HARQ effect hold for crude oil futures (CL)?
- **Model**: B2
- **Inputs**: VOLARE futures data, 28 Sep 2009 – 30 Jan 2026, CL; five-minute RV, semivariances and realised quarticity; models, tests and Newey–West (5 lags) as in B2
- **Tasks**
  (a) Estimate SHAR and test $\beta^- > \beta^+$ as in B2(b).
  (b) Estimate HARQ and test $\beta_Q < 0$ as in B2(c); plot the coefficients with 95% HAC intervals.
  (c) Compare both results with JPM.
- **Report**: estimates with HAC SE, two one-sided tests and 5% decisions, the coefficient chart, and comparison with JPM.
- **Interpretation**: is the asymmetry stronger or weaker than for JPM, and what economic mechanism could explain either outcome?
- Notebook: section B6

Back Exercise list

B7 [Proposed]: Sig-LK vs log-HAR for ES, One Week Ahead — Task

Solution discussed in class

- **Question**: does Sig-LK beat log-HAR for ES when the target is the average RV of the next five days?
- **Model**: B3
- **Inputs**: VOLARE futures data, 28 Sep 2009 – 30 Jan 2026, ES; design of B3 with $h = 5$ and $c = 1.0$ (chosen in Chapter 20 on the futures validation block)
- **Tasks**
  (a) Produce forecasts at every daily origin of the five-day average RV, as in B3; evaluation starts after the last validation target: origins 7 Oct 2014 to 23 Jan 2026.
  (b) Compute the QLIKE ratio and its stationary-bootstrap 95% CI (mean block length 10).
  (c) Run the DM test (bandwidth at least $h = 5$) and the conditional GW test using instruments $(1, d_{t-\mathbf{5}})$; plot the cumulative loss difference.
- **Report**: both mean losses, their ratio and interval, the DM/GW statistics and p-values, and the cumulative-loss plot.
- **Interpretation**
  - is the CI asymmetric, and which features of the losses could explain its shape?
  - what does it imply for the power of DM?
- Notebook: section B7

Back Exercise list

B8 [Proposed]: Multiple Testing at One Month Against HAR — Task

Solution discussed in class

- **Question**: on how many assets does Sig-LK beat HAR one month ahead after correcting for 50 tests?
- **Model**: B4
- **Inputs**: Chapter 20 results: 50 one-sided DM p-values for "Sig-LK better than HAR" at $h = 22$ (QLIKE)
- **Tasks**
  (a) Count the rejections at 5% without correction.
  (b) Apply Holm and BH as in B4, count the rejections and name the asset with the smallest p-value.
  (c) Compare with the mean across assets of the per-asset QLIKE ratios of Sig-LK relative to HAR at $h = 22$ (1.334).
- **Report**: sorted p-values, both thresholds and decisions, three rejection counts, and the named asset.
- **Interpretation**: how can Sig-LK beat HAR on several assets and still lose on average?
- Notebook: section B8

C2 [Proposed]: A Level-Aware Signature Kernel — Task

Solution discussed in class

- **Question**: does a kernel that also sees the *level* of volatility make the weights useful in a crash?
- **Inputs**
  - The Sig-LK kernel is translation-invariant, so it cannot recognise a high-volatility regime; data: VOLARE, the 40 stocks, 2015–2026
  - Idea: prepend a **basepoint** Morrill et al. (2020): the path starts at the training mean of $\ln \mathrm{RV}$ and jumps to the first value of the window; by Chen, $\mathrm{Sig} = \exp_{\otimes}(v) \otimes \mathrm{Sig}(X)$ with v the standardised level jump and $\exp_{\otimes}(v) = (1, v, \frac{1}{2}v^{\otimes 2}, \dots)$ the signature of that single jump segment
- **Tasks**
  (a) Write a one-page pre-registration: 40 stocks, $h = 1$, the Sig-LK design of B3, $c = 0.5$, QLIKE, DM with Holm, the COVID-19 subperiod reported separately.
  (b) Implement the basepoint kernel (use the Chen update of the notebook) with the same features as Sig-LK, and forecast.
  (c) Compare with Sig-LK and log-HAR: mean QLIKE ratios (all days and COVID-19), wins over log-HAR raw and after Holm.
- **Interpretation**: if seeing the level does not help, where would you look next (features, target, horizon)?
- Notebook: section C2

Back Exercise list

C2: The Basepoint Jump

- Use channel order (time, log RV, cumulative return) and prepend $v = (0, (\ln \mathrm{RV}_{\mathrm{first}} - \mu_{\mathrm{train}})/s_{\mathrm{train}}, 0)$. Here the mean and standard deviation use log RV at the 1000 training origins, without future observations.
- Keep the ordinary path scaling from B3. This additional jump uses the standard deviation of levels, not of increments. Its time and return components are zero.
- Unlike the zero basepoint of Morrill et al. (2020), this basepoint is centred on the training mean. It measures whether the starting volatility level is unusually high relative to that training sample.
- Report the pre-registration, QLIKE ratio tables for all days and COVID-19, raw and Holm-adjusted win counts.

Further Study on Quantinar

Advanced courses on quantinar.com

- Statistics of Financial Markets
- Applied Time Series Analysis with Python

Code and data

- Seminar notebook (Colab); Quantlets: github.com/danpele/MFM/tree/main/Quantlets/Ch_20; data: volare.unime.it

Acronyms used in this material (1/2)

- **AI**: Artificial Intelligence
- **AR**: AutoRegressive (model)
- **AUD**: Australian Dollar (ISO 4217 code)
- **AUDUSD**: Australian dollar / US dollar exchange rate (FX pair code)
- **BH**: Benjamini–Hochberg (procedure)
- **BIC**: Bayesian Information Criterion
- **CAD**: Canadian Dollar (ISO 4217 code)
- **CI**: Confidence Interval
- **CL**: Crude oil (WTI light sweet) futures, NYMEX ticker
- **COVID-19**: COronaVIrus Disease 2019
- **DM**: Diebold–Mariano (test)
- **ES**: E-mini S&P 500 futures (CME ticker)
- **EURUSD**: Euro / US dollar exchange rate (FX pair code)
- **FDR**: False Discovery Rate
- **FWER**: Family-Wise Error Rate
- **FX**: Foreign eXchange
- **GBPUSD**: British pound / US dollar exchange rate (FX pair code)
- **GW**: Giacomini–White (test)
- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors)

Acronyms used in this material (2/2)

- **HAR**: Heterogeneous AutoRegressive (model)
- **HARQ**: HAR model with Realised Quarticity
- **JPM**: JPMorgan Chase & Co. (NYSE ticker)
- **LASSO**: Least Absolute Shrinkage and Selection Operator
- **MA**: Moving Average
- **NW**: Newey–West (standard errors)
- **OLS**: Ordinary Least Squares
- **PRDS**: Positive Regression Dependence on a Subset (of p-values)
- **QLIKE**: Quasi-LIKElihood (loss function)
- **RQ**: Realised Quarticity
- **RV**: Realised Variance
- **SHAR**: Semivariance HAR model
- **US**: United States
- **USDCAD**: US dollar / Canadian dollar exchange rate (FX pair code)
- **USDJPY**: US dollar / Japanese yen exchange rate (FX pair code)
- **VOLARE**: VOLatility Archive for Realized Estimates (database)

References (1/2)

- Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society 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.
- Chen, K.-T. (1957). Integration of paths, geometric invariants and a generalized Baker–Hausdorff formula. *Annals of Mathematics*, 65(1), 163–178.
- Chevyrev, I., Kormilitzin, A. (2016). A primer on the signature method in machine learning. arXiv:1603.03788.
- Cipollini, F., Cruciani, G., Gallo, G.M., Insana, A., Otranto, E., Spagnolo, F. (2026). VOLatility Archive for Realized Estimates (VOLARE). arXiv:2602.19732.
- 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)

- Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578.
- Gu, H., Guo, X., Jacobs, T.L., Kaminsky, P., Li, X. (2024). Transportation marketplace rate forecast using signature transform. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4997–5005.
- Hambly, B., Lyons, T. (2010). Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of Mathematics*, 171(1), 109–167.
- Kish, L. (1965). Survey Sampling. Wiley.
- Morrill, J., Fermanian, A., Kidger, P., Lyons, T. (2020). A generalised signature method for multivariate time series feature extraction. arXiv:2006.00873.
- 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.