









Modelarea piețelor financiare

Seminarul 9: Volatilitate realizată — aplicații



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

Academia de Studii Economice din București
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Academia Română, Institutul de Prognoză Economică
MSCA Digital Finance

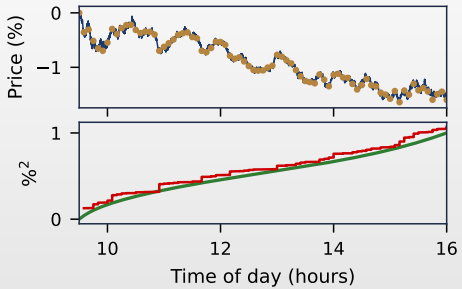
Îmbunătățește măsurarea intraday prognoza volatilității?

- Întrebarea seminarului
  - **cît de precis măsoară prețurile intraday dintr-o zi volatilitatea și îmbunătățește această măsurare o prognoză?**
- Două părți, în această ordine
  - Partea I, măsurare și diagnostic: A1, A2, B1, B2, A3
  - Partea a II-a, modele și prognoze: A4, B3, A5, A6, C1
- Livrabile
  - răspunsurile la exercițiile A
  - cîte o figură și un tabel pentru fiecare exercițiu B
  - tabelul de corecturi din C1
  - o frază de interpretare pentru fiecare exercițiu
- Notebook-ul de azi: deschideți notebook-ul seminarului în Colab; fiecare cerință indică secțiunea din notebook
- Materiale de sprijin: Capitolul 9 și Franke, Härdle & Hafner (2019)
- Seminar: prof. Daniel Traian Pele, danpele@ase.ro; drd. Antoaneta Amza, antoaneta.amza@csie.ase.ro

Ce calendar de piață și ce unități folosim?

- SPY (*SPDR S&P 500 ETF Trust*) este un fond tranzacționat la bursă (ETF, *Exchange-Traded Fund*) care urmărește indicele S&P 500 (*Standard & Poor's* 500)
- Prețurile Bitcoin sînt în dolari americani (USD)
- Etichetele temporale sînt în timpul universal coordonat (UTC, *Coordinated Universal Time*); barele SPY sînt convertite în ora New York
- Fișiere în data/market/intraday: `SPY.US_5m.csv` și `BTC-USD.CC_5m.csv`, cu `datetime_utc`
- Eticheta barei este începutul intervalului
- Axele graficelor intraday arată sfîrșitul intervalului
- Anualizăm cu 252 de zile pentru SPY și 365 pentru Bitcoin
- Randamentele sînt logaritmice, în procente; pătratele lor sînt în procente la pătrat
- C1 verifică răspunsul unui asistent de inteligență artificială (AI): pentru fiecare număr, precizați activul, intervalul, unitatea și momentul prognozei

Varianța realizată



- Detalii: Capitolul 9
- **Varianța integrată** (IV, *Integrated Variance*): varianța pe care ziua a produs-o efectiv

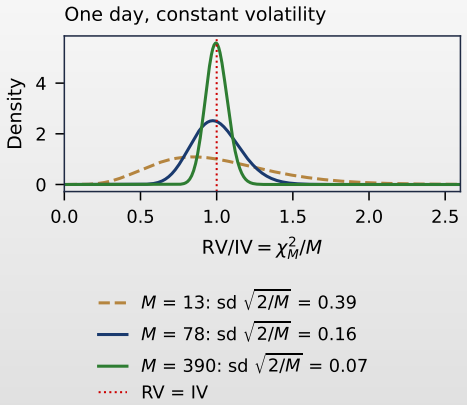
$$\mathrm{IV}_t = \int_{t-\mathbf{1}}^{t} \sigma_s^{\mathbf{2}}\, ds$$

  - σ_s: volatilitatea instantanee la momentul s din ziua t
- **Varianța realizată** (RV, *Realised Variance*) Andersen et al. (2001): suma celor M randamente intraday la pătrat

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

  - $r_{t,i} = 100\,(\ln P_{t,i} - \ln P_{t,i-\mathbf{1}})$: al i-lea randament din ziua t, în %; SPY: $M = 78$ de randamente de 5 minute
  - convergența cere un preț continuu, fără zgomot; volatilitatea anualizată $\sqrt{252\,\mathrm{RV}_t}$ (%)
  - graficul: o zi simulată, cu volatilitate intraday în formă de U; RV urmărește IV pas cu pas

Precizia RV dintr-o zi

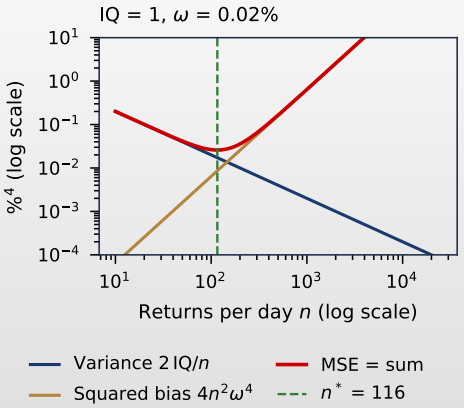


- Eroarea de măsurare a RV pentru M mare

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

  - IQ_t: cvarticitatea integrată; estimată prin cvarticitatea realizată $\mathrm{RQ}_t = \frac{M}{3}\sum_i r_{t,i}^4$
  - volatilitate constantă: $\mathrm{RV}/\mathrm{IV} = \chi^2_M/M$, o chi-pătrat cu M grade de libertate împărțită la M, abaterea standard relativă $\sqrt{2/M}$
- **Metoda delta** (*delta method*): pentru g derivabilă, eroarea standard $\mathrm{se}(g(X)) \approx |g'(X)|\,\mathrm{se}(X)$
  - $g = \ln$: $\mathrm{se}(\ln \mathrm{RV}) \approx \mathrm{se}(\mathrm{RV})/\mathrm{RV}$; un interval construit pe scara log și apoi exponențiat rămîne pozitiv
  - graficul: dispersia lui RV/IV pentru 13, 78 și 390 de randamente pe zi

Zgomotul de microstructură și frecvența de eșantionare



- Log-prețul observat = prețul eficient + zgomot

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

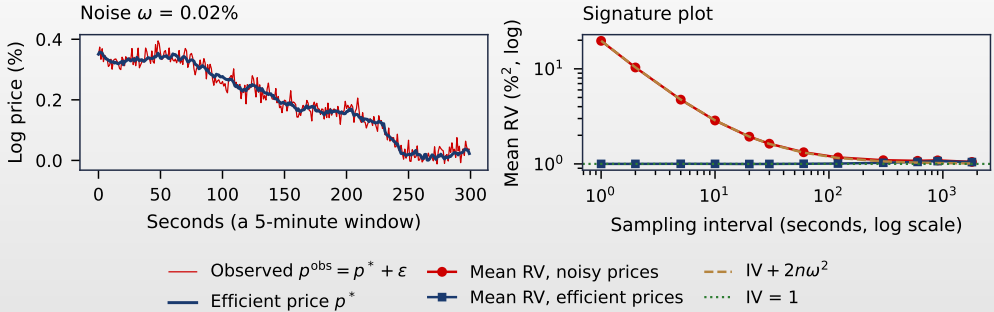
  - $\omega^{\mathbf{2}}$: dispersia zgomotului (oscilația între cotațiile de cumpărare și de vînzare, pasul de preț), ε independent de p^*
  - fiecare randament observat $r_i = r_i^* + \varepsilon_i - \varepsilon_{i-\mathbf{1}}$ conține doi termeni de zgomot

- Cu n randamente: deplasarea crește cu n, dispersia de eșantionare scade ca $1/n$ Bandi & Russell (2008)

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

  - MSE (*Mean Squared Error*, eroarea pătratică medie) = dispersia + pătratul deplasării; n^* optim echilibrează cei doi termeni (graficul: $\omega = 0{,}02\%$)
  - în practică, intervalul de 5 minute este greu de întrecut Liu, Patton & Sheppard (2015)

Zgomotul din prețuri și graficul de semnătură



- Simulare: 150 de zile, $\text{IV} = 1\,\%^2$, zgomot $\omega = 0{,}02\%$
  - stînga: de la o secundă la alta, prețul observat oscilează în jurul celui eficient
  - dreapta, **graficul de semnătură** (*signature plot*): RV medie în funcție de intervalul de eșantionare; prețurile cu zgomot urmează $\text{IV} + 2n\omega^2$, cele eficiente rămîn la IV
  - un grafic de semnătură aproximativ plat este compatibil cu o deplasare redusă din zgomot la intervalele analizate; nu exclude zgomotul

Grile rare, medierea și bootstrap pe blocuri

- **Grila rară** de $k \times 5$ minute care începe la decalajul $o \in \{0, \ldots, k-1\}$

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

  - media celor k grile decalate folosește toate prețurile și reduce dispersia de eșantionare, fără deplasare suplimentară din zgomot
  - raportul mediilor $\sum_t \mathrm{RV}_t^{(k)} / \sum_t \mathrm{RV}_t^{(\mathbf{5\ min})}$: aproape de 1 cînd deplasarea din zgomot la 5 minute este neglijabilă

- **Bootstrap pe blocuri suprapuse** Künsch (1989): un interval de încredere pentru o statistică a unor date dependente
  - alegem aleator zile de început și alăturăm blocuri de b zile consecutive (aici $b = 20$) pînă la lungimea inițială T
  - recalculăm statistica de B ori; percentilele de 2,5% și 97,5% dau intervalul de 95%
  - blocurile păstrează autocorelația din interiorul fiecărui bloc; reeșantionarea i.i.d. (independente și identic distribuite) ar distruge-o și ar da intervale prea înguste

Salturile și variația bipower

- Cu salturi J în timpul zilei, RV adună și pătratele lor

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

- **Variația bipower** (BV, *Bipower Variation*) Barndorff-Nielsen & Shephard (2004): produse de randamente absolute alăturate

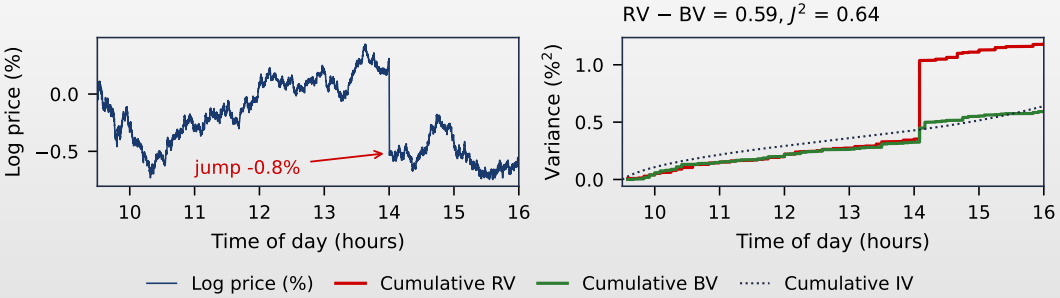
$$\mathrm{BV}_t = \mu_{\mathbf{1}}^{-\mathbf{2}} \frac{M}{M-1} \sum_{i=\mathbf{2}}^{M} |r_{t,i}|\,|r_{t,i-\mathbf{1}}| \xrightarrow{p} \mathrm{IV}_t, \qquad \mu_{\mathbf{1}} = \mathbb{E}|Z| = \sqrt{2/\pi}$$

  - $Z \sim N(0,1)$, deci $\mu_{\mathbf{1}}^{-\mathbf{2}} = \pi/2$; cu salturi de activitate finită și fără zgomot (cazul de bază), un salt intră în BV doar înmulțit cu un randament vecin mic; alte cadre cer condiții suplimentare
  - partea de salt $J_t = \mathrm{RV}_t - \mathrm{BV}_t$; ponderea salturilor în variația totală $\sum_t J_t / \sum_t \mathrm{RV}_t$
- **Cvarticitatea tripower** (TQ): o estimare a lui IQ_t robustă la salturi, folosită de testul de salt

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

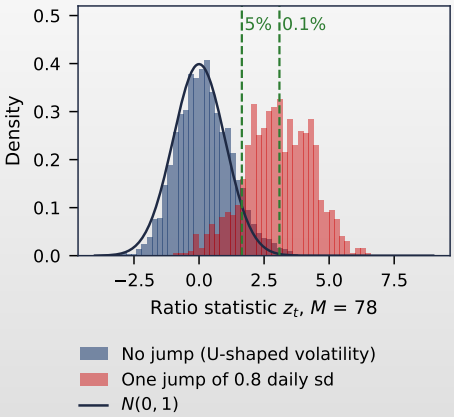
- Exemplu mic, $M = 4$, randamentele 0,1; 0,1; 0,1; 1,0 (%)
  - $\mathrm{RV} = 1{,}03$; $\mathrm{BV} = \frac{\pi}{\mathbf{2}} \cdot \frac{\mathbf{4}}{\mathbf{3}} (0{,}01 + 0{,}01 + 0{,}10) = 0{,}25$; partea de salt $\mathrm{RV} - \mathrm{BV} = 0{,}78$

Un salt văzut de RV și de BV



- Zi simulată, IV $= 0{,}64\,\%^2$, un salt de $-0{,}8\%$ la ora 14:00, 78 de randamente de 5 minute
  - RV cumulată crește brusc cu aproximativ $J^2 = 0{,}64$ la ora 14:00; BV cumulată continuă să urmeze IV
  - la închidere $\mathrm{RV} - \mathrm{BV} = 0{,}59$: aproape de pătratul saltului

Testul de salt al raportului



- H_0: nu există salt în ziua t; statistica raportului Huang–Tauchen Huang & Tauchen (2005)

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

  - $(\mathrm{RV}_t - \mathrm{BV}_t)/\mathrm{RV}_t$: măsura relativă a saltului, aproape de 0 sub H_0
  - $\theta = \pi^2/4 + \pi - 5 = 0{,}609$; $\mathrm{TQ}_t/\mathrm{BV}_t^2$ măsoară cît de neuniformă a fost volatilitatea în zi, cu pragul inferior 1

- Sub H_0, $z_t \approx N(0,1)$; test unilateral
  - respingem cînd $z_t > 1{,}645$; 2,326; 3,09 la 5%, 1%, 0,1%; valoarea p $p_t = 1 - \Phi(z_t)$, Φ: funcția de repartiție $N(0,1)$
  - N zile testate: aproximativ αN respingeri doar din întîmplare (testare multiplă, Capitolul 9)
  - graficul: cu volatilitate în formă de U și fără salt, 9% din zile depășesc 1,645 în loc de 5%

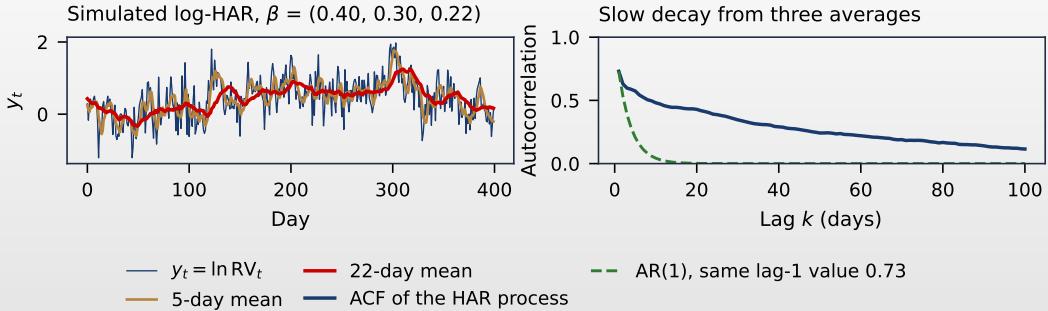
Modelul HAR

- Volatilitatea are persistență ridicată, compatibilă cu memoria lungă (Capitolul 1): funcția de autocorelație (ACF) a lui $|r_t|$ scade lent, iar prognoza unui model GARCH(1,1) (heteroscedasticitate condiționată autoregresivă generalizată, Capitolul 5) revine exponențial la medie
- HAR (*Heterogeneous Autoregressive*, autoregresiv heterogen) Corsi (2009): RV de mîine explicată prin trei medii ale RV trecute

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

  - β_d, β_w, β_m: ponderile zilei de ieri, ale ultimei săptămîni (5 zile) și ale ultimei luni (22 de zile); $\beta_{\mathbf{0}}$: constanta
  - HAR este un model autoregresiv AR(22) ale cărui 22 de coeficienți depind de trei pante: un HAR staționar are memorie scurtă
  - cu pante nenegative, persistența $\beta_d + \beta_w + \beta_m < 1$ este suficientă pentru staționaritate: șocurile se sting
  - în general (de exemplu, o pantă negativă), verificăm că rădăcinile polinomului AR(22) sînt în afara cercului unitate
- **log-HAR**: aceeași regresie pentru $y_t = \ln V_t$
  - prognoza nivelului $F_{t+\mathbf{1}} = \exp(\hat{y}_{t+\mathbf{1}} + s^{\mathbf{2}}/2)$, $s^{\mathbf{2}}$: dispersia reziduurilor (corecția lognormală)
  - prognoze pozitive, reziduuri mai apropiate de distribuția Normală
- Exemplu mic: $\hat{\beta} = (0{,}1;\ 0{,}4;\ 0{,}3;\ 0{,}2)$, $\mathrm{RV}_t = 2$, $\mathrm{RV}_t^{(\mathbf{5})} = 1{,}5$, $\mathrm{RV}_t^{(\mathbf{22})} = 1$
  - prognoza $0{,}1 + 0{,}8 + 0{,}45 + 0{,}2 = 1{,}55\,\%^{\mathbf{2}}$; persistența 0,9

Trei medii aproximează persistența pe orizonturi finite



- log-HAR simulat, $\beta_d = 0{,}40$, $\beta_w = 0{,}30$, $\beta_m = 0{,}22$ (persistența 0,92)
  - stînga: media pe 22 de zile se mișcă lent și ancorează prognoza; termenul zilnic reacționează rapid
  - dreapta: ACF este încă 0,24 la lagul 50; un AR(1) cu aceeași valoare la lagul 1 a uitat totul pînă la lagul 15

Inferența în regresii pe serii de timp

- Metoda celor mai mici pătrate (OLS): $\hat{\beta} = (X^\top X)^{-1} X^\top y$; pentru erori autocorelate și heteroscedastice, covarianța Newey–West, robustă la heteroscedasticitate și autocorelație (HAC)

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

  - $\hat{\Gamma}_k = \sum_t \hat{u}_t \hat{u}_{t-k}\, x_t x_{t-k}^\top$, $\hat{u}_t$: reziduurile, x_t: rîndul regresorilor; laguri $q = \lfloor 4(T/100)^{2/9} \rfloor$
- **Testul Wald** pentru q_R restricții liniare $H_0 : R\beta = r$

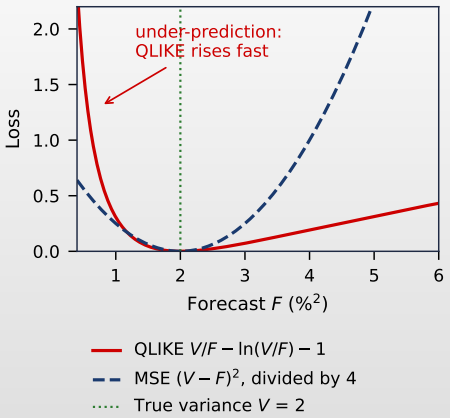
$$W = (R\hat{\beta} - r)^\top (R\hat{V}R^\top)^{-1} (R\hat{\beta} - r) \xrightarrow{d} \chi^2_{q_R}$$

  - R: un rînd pentru fiecare restricție; de ex. $\beta_w = \beta_m$ este rîndul $(0, 0, 1, -1)$, $r = 0$
- **Ljung–Box**: sînt primele h autocorelații nule împreună?

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

  - $\hat{\rho}_k$: autocorelația de selecție la lagul k; pe reziduuri, un Q mare = memorie neexplicată
  - pentru o serie i.i.d., $Q(h) \to \chi^2_h$; pe reziduurile unui model dinamic estimat (HAR) această referință nu este valabilă în general: calibrăm printr-un bootstrap care reestimează modelul
  - ajustarea: $R^2 = 1 - \sum_t \hat{u}_t^2 / \sum_t (y_t - \bar{y})^2$, ponderea dispersiei lui y explicată

Funcții de pierdere pentru prognozele de varianță

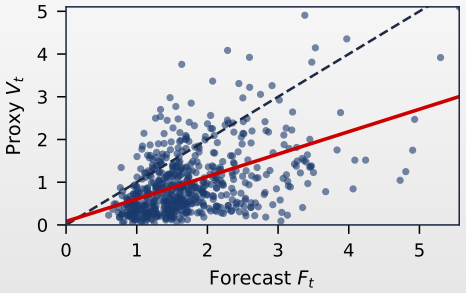


- Varianța adevărată nu se observă: comparăm F_t cu un indicator V_t nedeplasat condiționat, $\mathbb{E}[V_t \mid \mathcal{F}_{t-\mathbf{1}}] = \mathbb{E}[\mathrm{IV}_t \mid \mathcal{F}_{t-\mathbf{1}}]$
- Două funcții de pierdere care ordonează prognozele ca varianța adevărată Patton (2011)

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

  - ambele sînt 0 pentru $F = V$; QLIKE (*Quasi-Likelihood*, cvasi-verosimilitate) penalizează mai mult subestimarea
- **Diebold–Mariano** Diebold & Mariano (1995): $d_t = L(V_t, F_t^A) - L(V_t, F_t^B)$
  - $t = \bar{d}/\mathrm{se}_{\mathbf{NW}}(\bar{d}) \approx N(0,1)$ sub pierderi așteptate egale; $t < -1{,}96$: A semnificativ mai bun

Calibrarea: regresia Mincer–Zarnowitz



- Regresăm indicatorul pe prognoză

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

  - a: termenul liber, b: panta, u_t: eroarea; o prognoză calibrată satisface $a = 0$ și $b = 1$ în această regresie liniară; $\mathbb{E}[V_t - F_t] = 0$ singur nu implică acest lucru
  - $b < 1$: prognozele mari sînt urmate de medii realizate sub dreapta de 45° (graficul: $\hat{b} = 0{,}53$)

- Testul Wald comun pentru $(a, b) = (0, 1)$, cu $\delta = (\hat{a},\ \hat{b} - 1)^\top$

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

  - $\hat{V}$: covarianța 2×2 a lui $(\hat{a}, \hat{b})$, construită din erorile standard și corelația lor
  - elipsa de încredere de 95%: perechile (a, b) cu $W \le \chi^2_{2;0,95} = 5{,}99$
  - nerespingerea lui $(a, b) = (0, 1)$ nu dovedește calibrarea

Prognozele de referință, în afara eșantionului

- GARCH(1,1)-t pe randamente zilnice (Capitolul 5): o prognoză bazată pe un model

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

  - $\omega > 0$: constanta GARCH (alta decît ω, abaterea standard a zgomotului de microstructură); $a, b \geq 0$, $a + b < 1$
  - z_t: șocuri Student-t cu dispersia 1 și ν grade de libertate

- EWMA (medie mobilă ponderată exponențial, RiskMetrics): randamentele la pătrat cu ponderi exponențiale, $\lambda = 0{,}94$

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

  - un GARCH(1,1) cu $\omega = 0$ și $a + b = 1$: fără revenire la medie
  - RW (mers aleator, *random walk*): măsura de ieri ca prognoză pentru azi, $F_t = V_{t-1}$ (coloana `RW`)

- În afara eșantionului, cu fereastră extinsă
  - la fiecare origine $t-1$ fiecare model se reestimează pe datele pînă la $t-1$, apoi F_t se compară cu V_t
  - pierderile se mediază pe datele comune tuturor modelelor

Datele folosite

Seria	Conținut	Eșantion
SPY, bare de 5 minute	prețurile sesiunii regulate, ora UTC, fiecare bară etichetată cu începutul ei	2020-10-12 – 2026-09-18, 1 489 de zile
Bitcoin (USD), bare de 5 minute	zile UTC de 24 de ore	659 din 747 de zile păstrate
SPY și Bitcoin, prețuri zilnice	prețuri de închidere ajustate (SPY) și prețuri de închidere (Bitcoin), pentru GARCH	din 2000 (SPY), 2014 (Bitcoin)
Variațele zilnice SPY (calculate)	RV intraday, randamentul peste noapte și varianța totală V_t ($\%^2$, %)	un rînd pentru fiecare zi SPY
Prognoze pe o zi (calculate)	log-HAR, HAR, GARCH(1,1)-t, EWMA, RV de ieri (coloana `RW`) și ținta V_t	din 3 ian. 2022 (SPY), 1 mar. 2025 (Bitcoin)

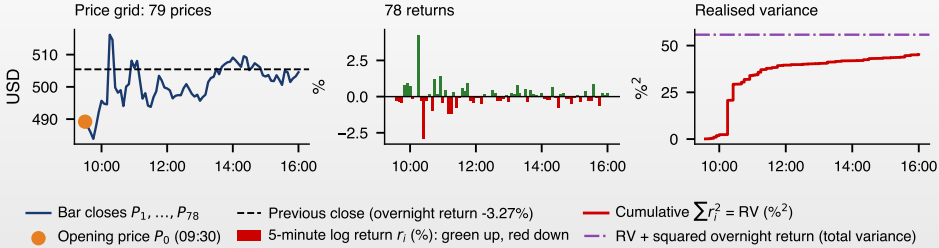
- SPY: ora UTC convertită în ora New York; barele 09:30–15:55; barele lipsă din timpul zilei preiau ultimul preț; 78 de randamente pe zi, 42 (09:30–13:00) în cele 12 zile cu program redus, în care bara etichetată 13:00 este după închidere și nu devine un interval
- Bitcoin: zilele cu mai puțin de 95% din cele 288 de bare sînt eliminate; golurile preiau ultimul preț, deci un interval lipsă se contopește cu următorul; prețul de la 00:00 este deschiderea primei bare sau închiderea barei de 23:55 din ziua precedentă; ziua se termină la ultimul preț observat
- Unități: randamente $r = 100\,\Delta \ln P$ în %, variațe în $\%^2$; anualizăm cu 252 (SPY) sau 365 (Bitcoin) de zile; toate datele se opresc la 18 septembrie 2026

Pregătire: rulați și verificați

```
# sectiunea "Date" din notebook defineste
# 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])
```

- Rezultatul așteptat
  - (1489, 79): zile SPY, de la 2020-10-12 la 2026-09-18
  - 1 477 de zile cu 78 de randamente, 12 cu 42
  - (659, 289): Bitcoin, de la 2024-09-01 la 2026-09-17
- Coloana 0 din P este prețul de deschidere al barei de 09:30; coloanele 1–78 sînt prețurile de închidere ale barelor
- Dacă numerele diferă, opriți-vă și verificați fusul orar, filtrul sesiunii și data de final

De la bare la randamente: o zi SPY



- 7 aprilie 2025: prețul de deschidere $P_0 = 489{,}19$, apoi închiderile barelor $P_1 = 487{,}60$, $P_2 = 485{,}88$, …, $P_{78} = 504{,}54$: 79 de prețuri dau 78 de randamente; doar din cele 78 de închideri am obține 77
- $r_1 = 100 \ln(P_1/P_0) = -0{,}33\%$, $r_2 = -0{,}35\%$; $\mathrm{RV} = \sum_i r_i^2 = 45{,}1\,\%^2$; randamentul peste noapte $-3{,}27\%$ este în afara barelor, varianța totală $V = \mathrm{RV} + r_{\text{on}}^2 = 55{,}8\,\%^2$
- Randamentul peste noapte: randamentul ajustat de la o închidere la alta minus randamentul de la deschidere la închidere, deci dividendele intră în partea de peste noapte

MFM_ch9_seminar

Notații folosite în toate exercițiile

- Ziua t are M randamente intraday $r_{t,i}$, $i = 1, \ldots, M$, în %
- Varianța zilnică totală $V_t = \mathrm{RV}_t + r_{\mathrm{on},t}^2$ pentru SPY ($r_{\mathrm{on},t}$: randamentul peste noapte); $V_t = \mathrm{RV}_t$ pentru Bitcoin (24 de ore)
- Prognoza F_t pentru V_t, făcută la închiderea zilei $t-1$; fiecare funcție de pierdere compară F_t cu V_t

Harta exercițiilor

Exercițiu	Întrebare	Tip	Exemplu de urmat
A1	cît de precisă este RV dintr-o zi?	Rezolvat	–
A2	cît de des eșantionăm un preț cu zgomot?	Rezolvat	–
B1	are SPY o deplasare din zgomot la 5 minute?	Rezolvat	A2
B2	ce zile SPY conțin un salt?	Rezolvat	–
A3	pot șase randamente să dezvăluie un salt?	Propus	B2, A1
A4	ce impune HAR unui AR(22)?	Rezolvat	–
B3	cît de persistentă este log-varianța SPY?	Rezolvat	A4
A5	ce funcție de pierdere ordonează corect prognozele de varianță?	Rezolvat	–
A6	este GARCH mai puțin precis, deplasat sau ambele?	Propus	A5; exemplul Wald
C1	ce este greșit într-un răspuns AI?	Propus	A2, B3
Pachete de extindere	A2–A5, B1–B3 extinse; A7, A8, B4–B9, C2, C3 (opționale, după seminar)	Rezolvat și Propus	indicat la fiecare cerință

- **[Rezolvat]**: rezolvarea completă în slide-uri și în notebook, un model de urmat; **[Propus]**: îl rezolvați individual, iar rezolvarea se discută la seminar
- Exercițiile A: derivări pe hîrtie; exercițiile B: calcul și inferență pe SPY și Bitcoin, fiecare cu o întrebare de interpretare la final; exercițiile C: întrebări deschise și critica unui răspuns AI

A1 [Rezolvat]: precizia variației realizate — enunț

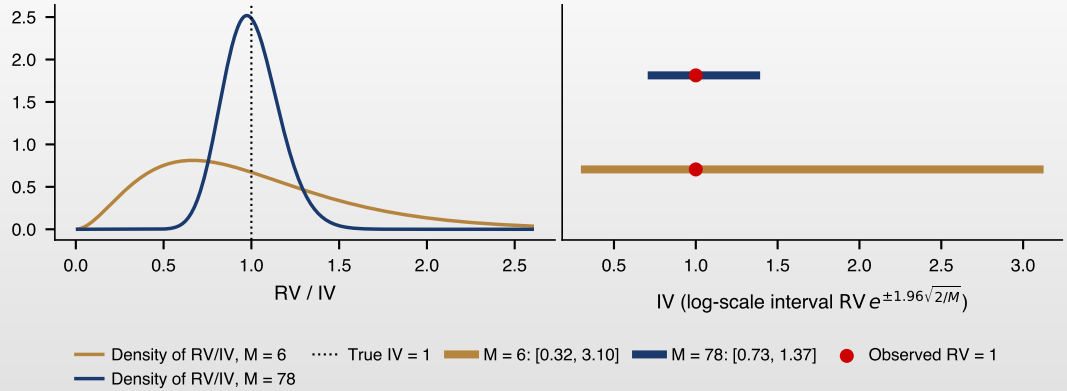
- **Întrebare**: cît de precis măsoară randamentele de 5 minute dintr-o zi varianța integrată?
- **Date de intrare**: volatilitate constantă, M intervale de lungime $\Delta = 1/M$, $r_i = \sigma\sqrt{\Delta}\, Z_i$ cu Z_i i.i.d. $N(0,1)$, $\mathrm{IV} = \sigma^2$; folosiți $\mathbb{E}Z^2 = 1$, $\mathrm{Var}(Z^2) = 2$, $\mathbb{E}Z^4 = 3$; χ^2_M: chi-pătrat cu M grade de libertate (suma a M pătrate de variabile $N(0,1)$ i.i.d.); fără fișier de date
- **Cerințe**
  1. Arătați că $\mathrm{RV} = \sigma^2 \chi^2_M / M$ și derivați $\mathbb{E}[\mathrm{RV}]$ și $\mathrm{Var}(\mathrm{RV})$.
  2. Arătați că $\mathrm{RQ} = \frac{M}{3}\sum_i r_i^4$ este nedeplasat pentru $\mathrm{IQ} = \sigma^4$ și scrieți eroarea standard estimată a RV.
  3. Folosiți metoda delta pentru $\ln \mathrm{RV}$ și scrieți intervalul de 95% pentru IV construit pe scara log.
  4. Calculați factorii intervalului $e^{\pm 1{,}96\sqrt{2/M}}$ pentru $M = 78$ și $M = 6$, precum și lățimea relativă pentru $M = 78$.
- **Raportați**: formulele, cele două perechi de factori și o propoziție despre cît de precisă este RV dintr-o zi
- Notebook: secțiunea A1

Înapoi Lista exercițiilor

A1 [Rezolvat]: rezolvare (1/3) — derivarea

- $\mathrm{RV} = \sum_i \sigma^2 \Delta Z_i^2 = \sigma^2 \frac{1}{M} \sum_i Z_i^2 = \sigma^2 \chi^2_M / M$, deci $\mathbb{E}[\mathrm{RV}] = \sigma^2$ (nedeplasată) și $\mathrm{Var}(\mathrm{RV}) = \sigma^4 \cdot 2M/M^2 = 2\sigma^4/M$
- $\mathbb{E}[r_i^4] = \sigma^4 \Delta^2 \mathbb{E} Z^4 = 3\sigma^4/M^2$, deci $\mathbb{E}[\mathrm{RQ}] = \frac{M}{3} \cdot M \cdot 3\sigma^4/M^2 = \sigma^4$; prin urmare $\widehat{\mathrm{Var}}(\mathrm{RV}) = 2\,\mathrm{RQ}/M$ și $\mathrm{se}(\mathrm{RV}) = \sqrt{\frac{2}{3} \sum_i r_i^4}$
- Metoda delta cu $g(x) = \ln x$, $g'(x) = 1/x$: $\mathrm{se}(\ln \mathrm{RV}) \approx \mathrm{se}(\mathrm{RV})/\mathrm{RV} = \sqrt{2\,\mathrm{RQ}/M}/\mathrm{RV}$
- Intervalul de 95%: $\mathrm{RV} \exp(\pm 1{,}96\, \mathrm{se}(\ln \mathrm{RV}))$, întotdeauna pozitiv și asimetric pe niveluri; cu volatilitate constantă $\mathrm{RQ}/\mathrm{RV}^2 \approx 1$, deci factorii sînt $e^{\pm 1{,}96\sqrt{2/M}}$
- Pentru volatilitatea $\sqrt{\mathrm{RV}}$, $g'(x) = 1/(2\sqrt{x})$ înjumătățește eroarea standard relativă: intervalul pentru volatilitate este mai îngust decît cel pentru varianță

A1 [Rezolvat]: rezolvare (2/3) — densitatea RV/IV și factorii intervalului



MFM_ch9_seminar

A1 [Rezolvat]: rezolvare (3/3) — acoperirea simulată și interpretarea

- $M = 78$: $\sqrt{2/78} = 0{,}160$, factorii $[0{,}73; 1{,}37]$: IV se află între $0{,}73\,\mathrm{RV}$ și $1{,}37\,\mathrm{RV}$, o lățime relativă totală de 0,64
- $M = 6$ (șase intervale egale în perioada observată): factorii $[0{,}32; 3{,}10]$: aproape neinformativ
- Acoperirea simulată a intervalului nominal de 95% cu RQ (200 000 de zile): 94,2% pentru $M = 78$, 86,7% pentru $M = 6$: nivelul de 95% este un rezultat asimptotic
- Cu volatilitate stocastică aceeași limită studentizată rămîne valabilă asimptotic, pentru un preț continuu și zgomot neglijabil, deoarece convergența este stabilă (Capitolul 9)
- **Interpretare**: o zi de date la 5 minute determină IV cu o precizie de aproximativ $-27\%/+37\%$

MFM_ch9_seminar

A2 [Rezolvat]: deplasarea din zgomot și frecvența optimă — enunț

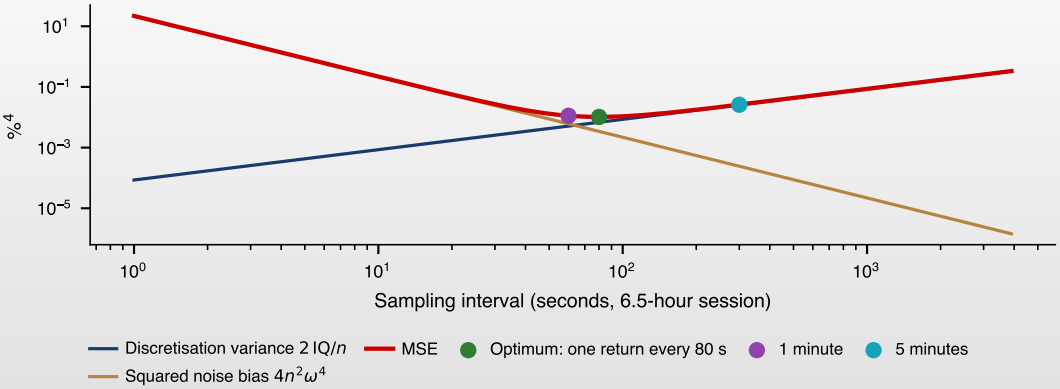
- **Întrebare**: cît de des eșantionăm un preț care conține zgomot?
- **Date de intrare**: modelul de zgomot din Zgomot și frecvență; drift nul, volatilitate constantă; $\mathrm{IV} = 1\,\%^2$, $\mathrm{IQ} = 1$, $\omega = 0{,}01\%$, deci $\omega^2 = 0{,}0001\,\%^2$; sesiune de 23 400 de secunde
- **Cerințe**
  1. Scrieți randamentul observat, dezvoltați $\mathbb{E}[r_i^2]$ și derivați $\mathbb{E}[\mathrm{RV}^{(n)}] = \mathrm{IV} + 2n\omega^2$ pentru n randamente.
  2. Derivați $\mathrm{cov}(r_i, r_{i-1}) = -\omega^2$.
  3. Din MSE(n) (slide-ul Zgomot și frecvență; Bandi & Russell (2008)), derivați n^* optim după MSE și intervalul de eșantionare corespunzător.
  4. Completați un tabel pentru 5 minute, 1 minut, 1 secundă și n^*: n, deplasarea $2n\omega^2$, $\mathbb{E}[\mathrm{RV}]$, MSE.
- **Raportați**: derivările, tabelul și ce interval ați folosi în practică, cu justificarea alegerii
- Notebook: secțiunea A2

◂ Înapoi ▸ Lista exercițiilor

A2 [Rezolvat]: rezolvare (1/3) — derivarea

- $r_i = p_i^{\text{obs}} - p_{i-1}^{\text{obs}} = r_i^* + \varepsilon_i - \varepsilon_{i-1}$
- $\mathbb{E}[r_i^2] = \mathbb{E}[r_i^{*2}] + \mathbb{E}[\varepsilon_i^2] + \mathbb{E}[\varepsilon_{i-1}^2] = \mathbb{E}[r_i^{*2}] + 2\omega^2$ (termenii încrucișați au media 0 prin independență); însumînd pe n randamente: $\mathbb{E}[\mathrm{RV}^{(n)}] = \mathrm{IV} + 2n\omega^2$
- Randamentele alăturate au în comun doar ε_{i-1}, cu semne opuse: $\operatorname{cov}(r_i, r_{i-1}) = \operatorname{cov}(-\varepsilon_{i-1}, \varepsilon_{i-1}) = -\omega^2$
- $\frac{d}{dn}\mathrm{MSE} = -2\,\mathrm{IQ}/n^2 + 8n\omega^4 = 0$ dă $n^3 = \mathrm{IQ}/(4\omega^4)$, $n^* = (\mathrm{IQ}/(4\omega^4))^{1/3}$; derivata a doua este pozitivă: un minim
- Cifre: $n^* = (1/(4 \times 10^{-8}))^{1/3} = 292$, un randament la fiecare $23\,400/292 = 80$ de secunde

A2 [Rezolvat]: rezolvare (2/3) — MSE în funcție de intervalul de eșantionare



MFM_ch9_seminar

A2 [Rezolvat]: rezolvare (3/3) — deplasarea și MSE la patru frecvențe

Interval	n	$2n\omega^2$	$\mathbb{E}[\mathrm{RV}]$	MSE
5 minute	78	0,0156	1,0156	0,0259
1 minut	390	0,0780	1,0780	0,0112
1 secundă	23 400	4,6800	5,6800	21,9025
n^*, la fiecare 80 s	292	0,0585	1,0585	0,0103

- $\mathrm{MSE} = 2/n + 4n^2 \times 10^{-8}$ în $\%^4$; la o secundă RV este de aproape șase ori IV
- În acest model 1 minut este aproape de optim; la 5 minute MSE este de 2,5 ori cea minimă
- **Interpretare**: alegerea uzuală de 5 minute se sprijină pe argumente din afara acestui model
  - ω este incert; zgomotul real nu este i.i.d.
  - RV pe 5 minute este greu de întrecut empiric Liu, Patton & Sheppard (2015)

MFM_ch9_seminar

B1 [Rezolvat]: alegerea frecvenței de eșantionare — enunț

- **Repere și exemple**: A2 [Rezolvat]
- **Întrebări**
  - are SPY o deplasare din zgomot măsurabilă la 5 minute?
  - ce grilă dă cea mai utilă măsură zilnică?
- **Date**: matricea de prețuri SPY `P` din pregătire (1 489 de zile); grilă de $k \times 5$ minute, $k = 1, 3, 6, 13$ (5, 15, 30, 65 de minute): prețurile din coloanele $o, o+k, o+2k, \ldots$ plus deschiderea și închiderea, pentru decalajele $o = 0, \ldots, k-1$ (funcția `sparse_points`)
- **Cerințe**
  1. Calculați RV zilnică pe o singură grilă ($o = 0$) și media pe cele k grile decalate și raportați $\sqrt{252\,\overline{\mathrm{RV}}}$ pentru fiecare interval.
  2. Pentru fiecare interval calculați raportul dintre media RV pe o grilă și media RV pe 5 minute și un interval de 95% prin bootstrap pe blocuri suprapuse (Grile și bootstrap): blocuri de 20 de zile ale seriilor zilnice pereche, 2 000 de reeșantionări.
  3. Calculați corelația dintre log RV de azi (o grilă și media grilelor) și log RV pe 5 minute de mîine.
- **Raportați**: un grafic și un tabel cu o linie pe interval: volatilitatea, raportul cu intervalul de încredere (CI) și corelația
- **Interpretare**
  - există o deplasare din zgomot măsurabilă la 5 minute?
  - de ce bootstrap pe blocuri și nu i.i.d.?
  - ce grilă este cea mai utilă pentru prognoză?
- Notebook: secțiunea B1

◂ Înapoi ▸ Lista exercițiilor

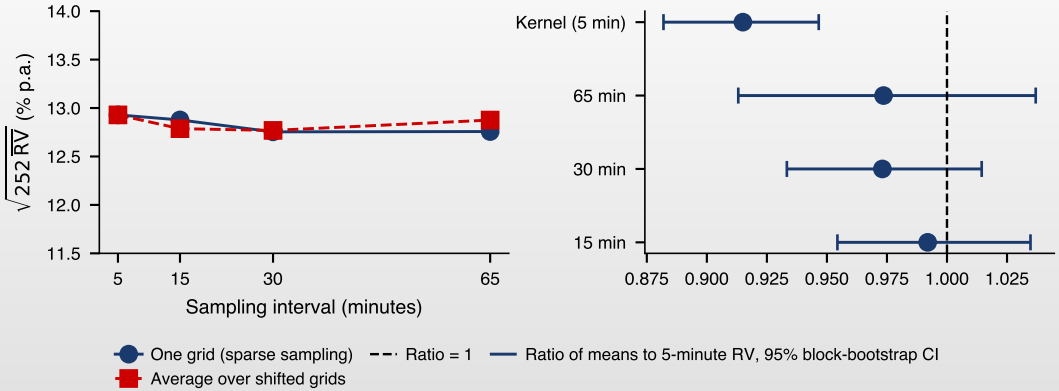
B1 [Rezolvat]: rezolvare (1/3) — procedura și tabelul

- Procedura: pentru fiecare zi luăm prețurile grilei, calculăm $\mathrm{RV}^{(o)} = \sum (100\,\Delta \ln P)^2$ pe fiecare decalaj o și facem media după o; bootstrap-ul reeșantionează perechile $(\mathrm{RV}_t^{(k)}, \mathrm{RV}_t^{(\mathbf{5\ min})})$ în blocuri de 20 de zile și recalculează $\sum \mathrm{RV}^{(k)} / \sum \mathrm{RV}^{(\mathbf{5\ min})}$

Interval	Vol. (o grilă)	Vol. (media grilelor)	Raport [CI 95%]	Corel. ziua următoare	Corel. (medie)
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

- Bootstrap pe blocuri: RV zilnice sînt puternic autocorelate; reeșantionarea i.i.d. ar distruge această dependență și ar face intervalul prea îngust

B1 [Rezolvat]: rezolvare (2/3) — graficul de semnătură și rapoartele nucleului



MFM_ch9_seminar

B1 [Rezolvat]: rezolvare (3/3) — deplasarea din zgomot și corelația predictivă

- Toate intervalele conțin 1: grilele rare nu detectează o deplasare față de RV pe 5 minute; graficul de semnătură este plat
- Corelația predictivă scade de la 0,73 (5 minute) la 0,64 (65 de minute, o grilă): RV rară este mai zgomotoasă; media grilelor decalate recuperează o parte (0,70)
- **Interpretare**: RV pe 5 minute este măsura cea mai utilă aici
- Nucleul realizat (*realised kernel*), un estimator robust la zgomot (pachetul de extindere), dă în medie 0,915 din RV pe 5 minute (CI de 95% $[0{,}882; 0{,}947]$): o diferență între doi estimatori, nu neapărat o deplasare din zgomot

MFM_ch9_seminar

B2 [Rezolvat]: zile cu salt pentru SPY — enunț

- **Întrebări**
  - în ce zile SPY respinge testul raportului?
  - cîte respingeri ar produce doar întîmplarea?
- **Date**: matricea de randamente SPY `R` (1 489 de zile; $M_t = 78$ sau 42 în zilele cu program redus); testul raportului Huang & Tauchen (2005) (Variația bipower și Testul raportului); funcția `jump_test`
- **Cerințe**
  1. Pentru fiecare zi calculați RV, BV, TQ, z_t și valoarea p unilaterală $p_t = 1 - \Phi(z_t)$; calculați pas cu pas pentru ziua de 27 august 2021.
  2. Numărați zilele în care testul respinge la 5%, 1% și 0,1% și comparați fiecare număr cu αN.
  3. În zilele respinse la 0,1% găsiți intervalul cu cel mai mare $|r_{t,i}|$ și reprezentați histograma orei la care se încheie.
  4. Calculați ponderea salturilor $\sum_t J_t / \sum_t \mathrm{RV}_t$, cu $J_t = \mathrm{RV}_t - \mathrm{BV}_t$ în zilele respinse la 0,1% și 0 în rest.
- **Raportați**: ziua calculată pas cu pas, un tabel cu numărul de respingeri față de αN, histograma și ponderea salturilor
- **Interpretare**
  - sînt mai multe respingeri decît ar implica nivelul testului?
  - ce le-ar putea explica?
- Notebook: secțiunea B2

◂ Înapoi ▸ Lista exercițiilor

B2 [Rezolvat]: rezolvare (1/3) — o zi pas cu pas

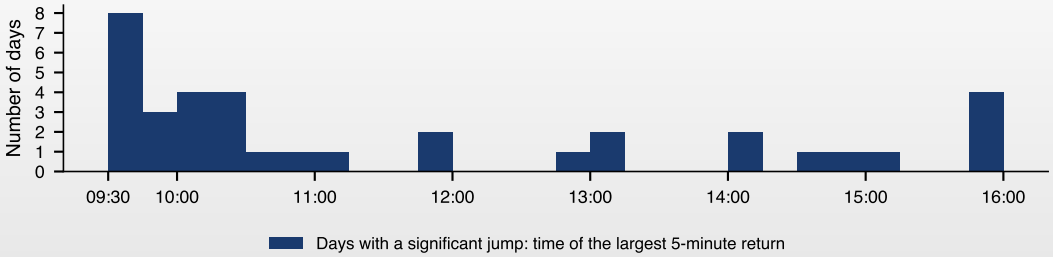
- 27 august 2021, $M = 78$: $\mathrm{RV} = 0{,}1354$, $\mathrm{BV} = 0{,}0687$, $\mathrm{TQ} = 0{,}00539$ (%^{2}, %^{2}, %^{4})
- $(\mathrm{RV} - \mathrm{BV})/\mathrm{RV} = 0{,}493$; $\mathrm{TQ}/\mathrm{BV}^2 = 1{,}143 > 1$, deci pragul nu este activ
- $z = 0{,}493/\sqrt{0{,}609 \times 1{,}143/78} = 5{,}22$, $p = 9{,}0 \times 10^{-8}$: respingem la 0,1%
- Cel mai mare randament, $+0{,}24\%$ în intervalul care se încheie la 10:05, reprezintă 44% din RV: la 10:00–10:05 a fost publicat discursul președintelui Rezervei Federale de la Jackson Hole
- Aceiași patru pași, vectorizați pe zile, sînt `jump_test(R)`

B2 [Rezolvat]: rezolvare (2/3) — numărul de respingeri și ponderea salturilor

Nivel	Respingeri	αN
5%	236	74,5
1%	110	14,9
0,1%	36	1,5

- z_t are media 0,50 și abaterea standard 1,15 în loc de 0 și 1: mult mai multe respingeri decît implică nivelul
- Ponderea salturilor în RV totală, cu zilele de la 0,1%: 2,8%
- Explicații: salturi reale și forma de U din cursul zilei, care deplasează BV în jos; pachetul de extindere aplică controlul ratei descoperirilor false (FDR) și corecția de periodicitate

B2 [Rezolvat]: rezolvare (3/3) — momentul celor mai mari randamente



- Ora celui mai mare $|r|$ în zilele semnalate: vîrful este după deschidere; atribuirea unei știri cere dovezi separate

MFM_ch9_seminar

Un randament mare și dovada unui salt

- **Ce credeți**: un singur randament mare, într-o zi calmă, dovedește un salt?
- **Răspuns**: nu; testul de salt întreabă dacă un preț continuu este neplauzibil, dată fiind frecvența observațiilor
- Statistica testului nu poate depăși $\sqrt{M/\theta}$: $\mathrm{BV} \geq 0$, iar maximul din numitor este cel puțin 1
- Comparați acest plafon cu valoarea critică înainte de a interpreta o nerespingere
- RV, BV și TQ folosesc aceeași unitate a randamentelor; statistica testului nu depinde de ea

A3 [Propus]: șase randamente și un salt — enunț

- **Repere și exemple**: B2 [Rezolvat] (o zi pas cu pas) și A1
- **Întrebare**: pot șase randamente de 5 minute să dezvăluie un salt?
- **Date de intrare**: randamentele din tabel (%); ultimul este un salt produs de o știre în **ultimul** interval; $\mathrm{TQ} = 0{,}0097$ (dată); $\theta = 0{,}609$; valoarea critică la 0,1%: 3,09; formulele: Variația bipower și Testul raportului
- **Cerințe**
  1. Completați tabelul și calculați RV și BV cu $M = 6$.
  2. Calculați partea de salt $\mathrm{RV} - \mathrm{BV}$ și ponderea ei în RV.
  3. Calculați $\mathrm{TQ}/\mathrm{BV}^2$, aplicați pragul $\max(1, \mathrm{TQ}/\mathrm{BV}^2)$ și calculați statistica raportului z.
  4. Decideți la 0,1% și comparați z cu cea mai mare valoare posibilă, $\sqrt{M/\theta}$.
- **Raportați**: tabelul completat, RV, BV, z, decizia și o propoziție de interpretare

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		

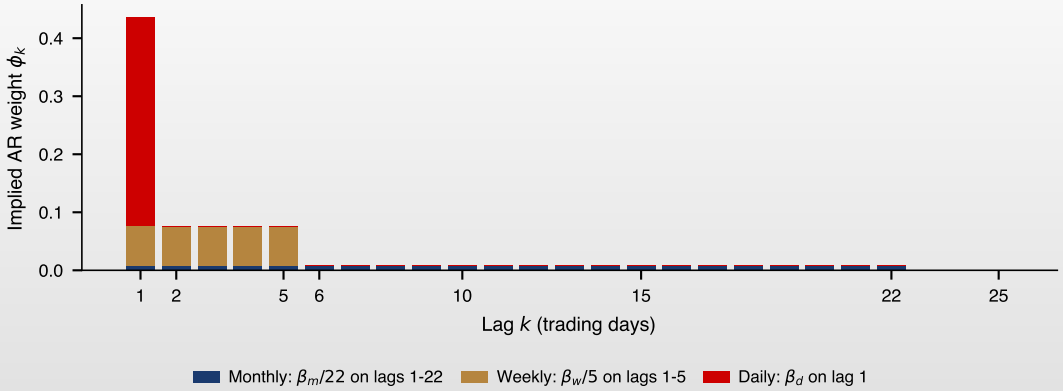
◂ Înapoi ▸ Lista exercițiilor

A4 [Rezolvat]: HAR ca AR(22) restricționat — enunț

- **Întrebare**: ce restricții impune HAR unui AR(22) pentru log-varianță?
- **Date de intrare**: log-HAR din Modelul HAR Corsi (2009), pentru $y_t = \ln V_t$: valoarea zilnică și mediile lui y pe 5 zile (săptămîna) și pe 22 de zile (luna); estimările SPY rotunjite $\hat{\beta} = (-0{,}08; 0{,}36; 0{,}34; 0{,}17)$
- **Cerințe**
  1. Dezvoltați cele două medii și scrieți coeficienții AR implicați ϕ_1, $\phi_2 = \ldots = \phi_5$ și $\phi_6 = \ldots = \phi_{22}$.
  2. Cu vectorul $\phi = (\phi_0, \phi_1, \ldots, \phi_{22})$, scrieți două rînduri ale matricei R din $R\phi = 0$ și numărați restricțiile.
  3. Calculați ϕ_1, $\phi_{2..5}$, $\phi_{6..22}$ și persistența $\sum_k \phi_k$ pentru $\hat{\beta}$ și desenați cele 22 de ponderi.
- **Raportați**: cele trei formule ale ponderilor, două rînduri din R, cifrele și ce implică persistența pentru șocuri
- Notebook: secțiunea A4

◂ Înapoi ▸ Lista exercițiilor

A4 [Rezolvat]: rezolvare (1/2) — ponderile AR(22) implicite



MFM_ch9_seminar

A4 [Rezolvat]: rezolvare (2/2) — restricțiile și persistența

- Lagul 1 apare în toate cele trei componente: $\phi_{\mathbf{1}} = \beta_d + \beta_w/5 + \beta_m/22$; lagurile 2–5 în cea săptămînală și în cea lunară: $\phi_k = \beta_w/5 + \beta_m/22$; lagurile 6–22 doar în cea lunară: $\phi_k = \beta_m/22$
- Rînduri din R (coloana 0: termenul liber): $(0, 0, 1, -1, 0, \ldots, 0)$ pentru $\phi_{\mathbf{2}} = \phi_{\mathbf{3}}$; $(0, \ldots, 0, 1, -1, 0, \ldots)$ în coloanele 6, 7 pentru $\phi_{\mathbf{6}} = \phi_{\mathbf{7}}$
- Numărul: 3 egalități între $\phi_{\mathbf{2}}, \ldots, \phi_{\mathbf{5}}$ plus 16 între $\phi_{\mathbf{6}}, \ldots, \phi_{\mathbf{22}}$: 19 rînduri, R are 19×23 (22 de laguri minus 3 pante libere)
- $\phi_{\mathbf{1}} = 0{,}36 + 0{,}34/5 + 0{,}17/22 = 0{,}436$; $\phi_{\mathbf{2..5}} = 0{,}076$; $\phi_{\mathbf{6..22}} = 0{,}008$; $\sum_k \phi_k = \beta_d + \beta_w + \beta_m = 0{,}87$
- **Interpretare**: persistența $0{,}87 < 1$: șocurile asupra log-variației se sting lent, dar dispar; testul pe date al celor 19 restricții este în pachetul de extindere

MFM_ch9_seminar

B3 [Rezolvat]: log-HAR pentru SPY cu inferență robustă — enunț

- **Repere și exemple**: A4 [Rezolvat]
- **Întrebări**
  - cît de persistentă este log-varianța totală a SPY?
  - schimbă erorile standard robuste vreo concluzie?
- **Date**: varianța zilnică totală SPY din pregătire, $V_t = \mathrm{RV}_t + r^2_{\mathrm{on},t}$, 1 489 de zile; regresorii datați $t-1$: $\ln V_{t-1}$ și mediile lui $\ln V$ pe zilele $t-5..t-1$ și $t-22..t-1$ (media logaritmilor); primele 22 de zile se pierd
- **Cerințe**
  1. Construiți matricea de design și verificați rîndul pentru 8 aprilie 2025 cu rezolvarea.
  2. Estimați log-HAR prin OLS; raportați erorile standard obișnuite și Newey–West (HAC) cu 7 laguri, după regula din Inferența în regresie.
  3. Testați $\beta_w = \beta_m$ cu un test Wald folosind covarianța Newey–West.
  4. Aplicați testul Ljung–Box cu 22 de laguri pe $\ln V_t$ și pe reziduuri, reprezentați ambele funcții de autocorelație și comparați R^2 cu un AR(1); pentru reziduuri, obțineți valoarea p dintr-un bootstrap care reestimează HAR.
- **Raportați**: tabelul coeficienților cu ambele erori standard, testele cu valorile p și graficul
- **Interpretare**
  - ce componentă contează cel mai mult?
  - schimbă erorile robuste vreo concluzie?
  - reușește HAR să explice persistența?
- Notebook: secțiunea B3

◂ Înapoi ▸ Lista exercițiilor

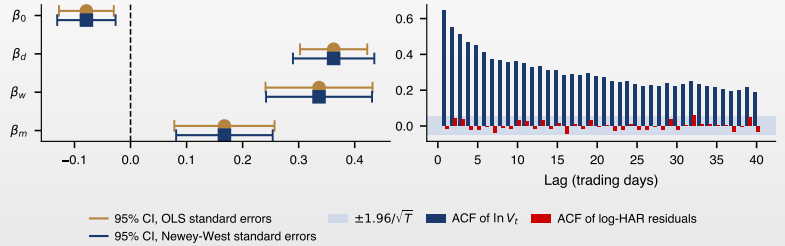
B3 [Rezolvat]: rezolvare (1/2) — un rînd din matricea de design și estimările

- Rîndul pentru 8 aprilie 2025: $y_t = \ln V_t = 2{,}976$; regresorii $\ln V_{t-\mathbf{1}} = 4{,}022$ (7 aprilie), media săptămînală 2,068, media lunară 0,612
- Valoarea ajustată $\hat{y}_t = 2{,}177$: modelul anticipă o revenire parțială spre nivelul lunar
- Observații utilizabile: $1\,489 - 22 = 1\,467$; $R^2 = 0{,}470$ față de 0,413 pentru un AR(1) ($\hat{\phi} = 0{,}643$); persistența 0,866

	$\hat{\beta}$	SE OLS	SE NW	t_{NW}
β_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

- Erorile standard (SE) Newey–West sînt cu pînă la 20% mai mari decît cele OLS; toate pantele rămîn semnificative

B3 [Rezolvat]: rezolvare (2/2) — teste și grafic



- Wald $\beta_w = \beta_m$: 4,13, $p = 0{,}042$: respinsă la 5%, ponderea săptămînală este mai mare
- Ljung–Box(22): $\ln V_t$ 4 637, reziduuri 17,6: trei coeficienți absorb persistența; autocorelația de ordinul 1 a reziduurilor $-0{,}014$
- $Q(22)$ pe reziduuri: bootstrap wild (reziduuri cu semne aleatoare) care reestimează HAR (499 de eșantioane), $p = 0{,}59$; referința χ^2_{22} ($p = 0{,}73$) ignoră dinamica estimată și heteroscedasticitatea
- **Interpretare**: componentele săptămînală și zilnică domină; erorile robuste nu schimbă nicio concluzie; HAR nu lasă autocorelație semnificativă

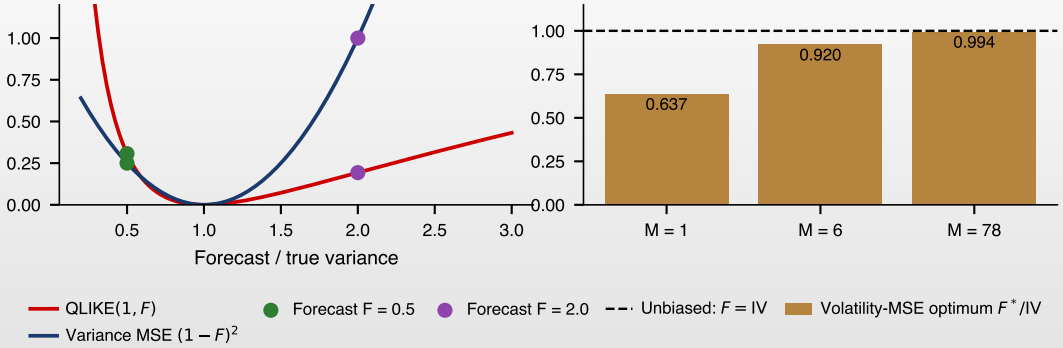
MFM_ch9_seminar

A5 [Rezolvat]: funcții de pierdere pentru ordonarea prognozelor de varianță — enunț

- **Întrebare**: ce funcție de pierdere ordonează corect prognozele de varianță când ținta este un indicator zgomotos?
- **Date de intrare**: indicatorul V_t nedeplasat condiționat și funcțiile QLIKE și MSE pe varianță din Pierderi de prognoză Patton (2011)
- **Cerințe**
  1. Minimizați $\mathbb{E}[\text{QLIKE}(V_t, F) \mid \mathcal{F}_{t-1}]$ în raport cu F și arătați că minimul se atinge în $\mathbb{E}[\text{IV}_t \mid \mathcal{F}_{t-1}]$.
  2. Pentru două prognoze F_A, F_B arătați că $\Delta L = V(1/F_A - 1/F_B) + \ln(F_A/F_B)$ este liniară în V, deci media ei condiționată este aceeași cu V_t sau cu IV_t: ordinea se păstrează.
  3. Cu varianța adevărată 1, calculați QLIKE și MSE pe varianță pentru $F = 0{,}5$ și $F = 2$ și comparați cele două ordonări.
- **Raportați**: condiția de ordinul întîi, argumentul pentru ordonare, cele patru valori ale pierderilor și ce funcție ați folosi
- Notebook: secțiunea A5

Înapoi Lista exercițiilor

A5 [Rezolvat]: rezolvare (1/2) — curbele QLIKE și MSE



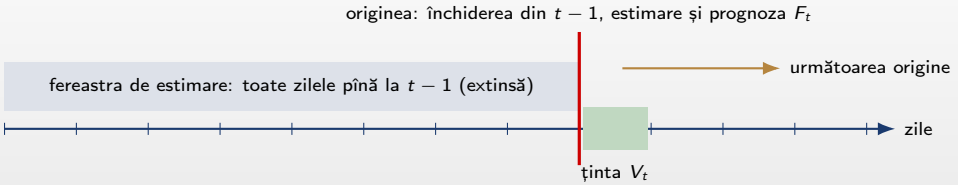
MFM_ch9_seminar

A5 [Rezolvat]: rezolvare (2/2) — condiția de ordinul întîi și ordonarea

- $\partial_F \mathbb{E}[V/F - \ln V + \ln F - 1] = -\mathbb{E}V/F^2 + 1/F = 0$ dă $F = \mathbb{E}[V_t \mid \mathcal{F}_{t-1}] = \mathbb{E}[\mathrm{IV}_t \mid \mathcal{F}_{t-1}]$; derivata a doua $1/F^2 > 0$
- $\mathrm{QLIKE}(V, F_A) - \mathrm{QLIKE}(V, F_B) = V(1/F_A - 1/F_B) + \ln(F_A/F_B)$: liniară în V, iar F_A, F_B sînt cunoscute în $t-1$, deci înlocuirea lui V_t cu IV_t, nedeplasat condiționat, nu schimbă $\mathbb{E}[\Delta L \mid \mathcal{F}_{t-1}]$
- QLIKE: $F = 0{,}5$ dă $2 - \ln 2 - 1 = 0{,}307$, $F = 2$ dă $0{,}5 + \ln 2 - 1 = 0{,}193$; MSE pe varianță: 0,25 și 1, ordinea inversă
- Graficul, dreapta: optimul MSE pe volatilitate, F^*/IV, pentru $M = 1, 6, 78$ (A5 extins)
- **Interpretare**: ambele funcții sînt robuste, dar QLIKE penalizează mai mult subestimarea, ceea ce contează pentru risc; MSE pe *volatilitate* nu este robustă (pachet de extindere)

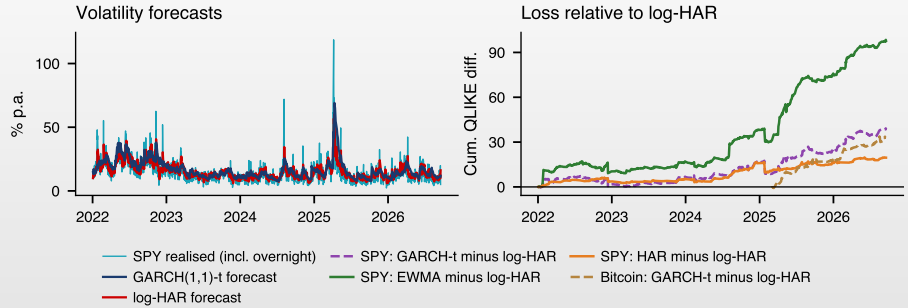
MFM_ch9_seminar

Evaluarea prognozelor: origine, țintă, fereastră



- Schema în afara eșantionului din Prognoze în afara eșantionului: nicio valoare din ziua t nu intră în prognoza F_t
- SPY: 1 180 de origini, de la 3 ianuarie 2022 la 18 septembrie 2026

Prognoze pe o zi și pierderi cumulate



- QLIKE medie: log-HAR 0,310, GARCH(1,1)-t 0,343; o linie cumulată crescătoare înseamnă că log-HAR are pierderea mai mică
- d_t = QLIKE(log-HAR) − QLIKE(GARCH) este punctul de plecare pentru A6: este diferența mai mare decît ar produce întîmplarea?

MFM_ch9_forecasts

Un test comun de calibrare: exemplu numeric

- Estimări ilustrative din regresia Mincer–Zarnowitz (MZ): $\hat{a} = 0{,}10$, $\mathrm{SE}(\hat{a}) = 0{,}05$, $\hat{b} = 0{,}90$, $\mathrm{SE}(\hat{b}) = 0{,}08$, corelația $\rho = -0{,}5$. Testăm $a = 0$, $b = 1$.
- Covarianța este $-0{,}5(0{,}05)(0{,}08) = -0{,}002$. Abaterile standardizate sînt $t_a = 2$, $t_b = -1{,}25$.
- $W = (t_a^2 + t_b^2 - 2\rho t_a t_b)/(1-\rho^2) = (4 + 1{,}5625 - 2{,}5)/0{,}75 = 4{,}0833$. Față de χ^2_2, $p = e^{-W/2} = 0{,}130$: nu respingem simultan restricțiile la 5%.
- Cînd $|\rho|$ este aproape de unu, matricea de covarianță este aproape singulară: rotunjirea poate schimba concluzia. În MZ, $b > 1$ înseamnă că varianța realizată variază mai mult decît prognozele în regresia de calibrare; aceștia nu sînt coeficienții GARCH.

A6 [Propus]: acuratețe și calibrare — enunț

- **Repere și exemple**: exemplul precedent de test comun de calibrare și B3 pasul 3
- **Întrebare**: este GARCH mai puțin precis decît log-HAR, deplasat sau ambele?
- **Date de intrare**: prognoze SPY pe o zi, ianuarie 2022 – septembrie 2026 (păstrați patru zecimale)
  - acuratețea: d_t = QLIKE(log-HAR) − QLIKE(GARCH) are media $\bar{d} = -0{,}0327$, SE Newey–West 0,0142
  - calibrarea, regresia MZ pentru GARCH, $V_t = a + bF_t + u_t$: $\hat{a} = 0{,}2418$ (SE 0,2310), $\hat{b} = 0{,}7046$ (SE 0,2382), $\operatorname{corr}(\hat{a}, \hat{b}) = -0{,}9750$
- **Cerințe**
  1. Acuratețea, H_0: pierdere așteptată egală. Calculați statistica t Diebold–Mariano, valoarea p bilaterală și decideți la 5%.
  2. Calibrarea, H_0: $a = 0$, $b = 1$. Calculați statisticile t separate pentru $a = 0$ și $b = 1$.
  3. Construiți matricea de covarianță 2×2, calculați statistica Wald comună W din Regresia de calibrare și valoarea p din χ^2_2.
  4. Desenați elipsa de încredere de 95%, marcați $(0, 1)$ și explicați cînd testele separate și cel comun nu coincid.
- **Raportați**: trei statistici, trei decizii și o propoziție care separă acuratețea de calibrare

◂ Înapoi ▸ Lista exercițiilor

C1 [Propus]: critica unui răspuns AI — enunț

- **Repere și exemple**: A2 și B3 [Rezolvat]; un exercițiu de citire a codului, nu sînt necesare date tick cu tick
- **Cadrul**: un student îi dă unui asistent AI o serie pandas `px` cu prețuri SPY curate, ordonate în timp, ora New York, 09:30–16:00, mai multe zile, în USD (deci randamentele sînt zecimale, nu în %)
- **Cerințe**
  1. Răspunsul de pe slide-ul următor conține trei erori conceptuale; restul este corect în acest cadru. Găsiți-le.
  2. Pentru fiecare eroare prezentați dovada: o derivare, o verificare a rezultatului sau o cifră din curs ori din acest seminar.
  3. Dați formula și linia de cod corecte. Pentru $\mathrm{RV}_t = 10^{-4}$, raportați volatilitatea anualizată afișată de codul inițial și de cel corectat.
  4. Consemnați fiecare eroare găsită efectiv în formatul din `AI_ERRORS.md`: cererea, răspunsul, eroarea, cum ați găsit-o, corecția.
- **Raportați**: un tabel cu cîte un rînd pentru fiecare eroare (ce, dovada, corecția) și cele două volatilități anualizate de la cerința 3
- Indicații: A2 despre $\mathbb{E}[\mathrm{RV}^{(n)}]$ cu zgomot; ce se întîmplă între 16:00 și 09:30; introduceți $\mathrm{RV}_t = 1{,}0 \times 10^{-4}$ ($1\%^2$)

◂ Înapoi ▸ Lista exercițiilor

C1 [Propus]: critica unui răspuns AI — răspunsul

- **Cererea**: „Calculează din px varianța realizată zilnică, volatilitatea anualizată și regresorii HAR."
- **Răspunsul, text**
  - $\mathrm{RV}_t = \sum_i r_{t,i}^2$ cu $r_{t,i}$ randamentele log intraday din ziua t; folosiți fiecare tranzacție: mai multe randamente fac întotdeauna RV_t mai precisă
  - RV_t este varianța zilnică a SPY: comparați-o direct cu pătratul randamentelor de la o închidere la alta și cu varianța GARCH
  - HAR: regresia lui $\mathrm{RV}_{t+\mathbf{1}}$ pe RV_t și pe mediile lui pe 5 și pe 22 de zile 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()
```

Tema pentru acasă

Temă: alegeți una

- B4, salturi Bitcoin (model: B2), sau B7, prognoza săptămînii următoare (model: B3, A6)
- Scrieți-o ca folder Quantlet (cod, `Metainfo.txt`, graficul): formatul folosit în proiectul de echipă
- Folosirea AI este permisă; notați erorile depistate efectiv, în formatul din `AI_ERRORS.md`
- Pachetele de extindere sînt opționale

Criterii de evaluare

- corectitudinea calculelor și a inferenței
- reproductibilitatea codului
- interpretarea rezultatelor și a limitelor lor
- documentarea folosirii AI

Pachetele de extindere: conținut și cerințe prealabile

Pachet	Exerciții	Cunoștințe necesare
1. Zgomot	A2 extins [Rezolvat], A7, B1 extins [Rezolvat]	A2, B1
2. Salturi	A3 extins, B2 extins [Rezolvat], B4	A3, B2
3. Distribuții	B5	B1 (bootstrap pe blocuri)
4. HAR	A4 extins [Rezolvat], B3 extins [Rezolvat], A8, B6	A4, B2, B3
5. Prognoze	A5 extins [Rezolvat], B7, B7 extins, B8	A5, A6, evaluarea prognozelor
6. Asprime	B9	B1, Capitolul 9 (volatilitatea aspră)
7. Proiecte	C2, C3	B3, A5, B8, studiul de caz din curs

- Notebook-ul oferă exemplele rezolvate și funcțiile de calcul; extensiile cer și cod propriu. În cadrul unui pachet, rezolvați exercițiile în ordine
- C2 și C3 pot fi puncte de plecare pentru proiectul de echipă

Ce corectează estimarea pe două scări?

- Varianța realizată pe două scări (TSRV, *Two-Scale Realised Variance*) scade din varianța mediată pe grile rare o estimare a deplasării din zgomot
- TSRV anulează termenul principal de zgomot și păstrează informația multor observații
- Exemplu: grile rare 1,20, grila completă 1,50, ponderea corecției 0,20: $1{,}20 - 0{,}20 \times 1{,}50 = 0{,}90$
- Factorul pentru eșantion finit corectează atenuarea rămasă; precizați dacă l-ați aplicat

A2 extins [Rezolvat]: estimarea variației zgomotului — enunț

- **Întrebări**
  - cum poate fi estimat ω^2 din date?
  - în ce limită este consistent fiecare estimator?
- **Date de intrare**: modelul din A2, n randamente pe zi
- **Cerințe**
  1. Arătați că $\mathbb{E}[\mathrm{RV}^{\mathrm{all}}/(2n)] = \omega^2 + \mathrm{IV}/(2n)$ și precizați limita lui cînd $n \to \infty$.
  2. Propuneți un al doilea estimator din autocovarianța de ordinul întîi și precizați cînd este consistent.
- **Raportați**: cele două derivări, limita fiecărui estimator și condiția lui de consistență
- Notebook: secțiunea A2

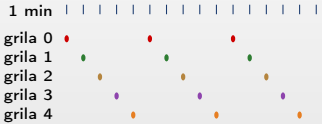
Înapoi Lista exercițiilor

A2 extins [Rezolvat]: rezolvare — estimarea variației zgomotului

- $\mathbb{E}[\mathrm{RV}^{\mathsf{all}} \mid \sigma] = \mathrm{IV} + 2n\omega^2$, deci $\mathbb{E}[\mathrm{RV}^{\mathsf{all}}/(2n)] = \omega^2 + \mathrm{IV}/(2n)$; legea numerelor mari pentru termenii de zgomot dă $\hat{\omega}^2 = \mathrm{RV}^{\mathsf{all}}/(2n) \xrightarrow{p} \omega^2$ cînd $n \to \infty$ Bandi & Russell (2008)
- $-\widehat{\mathrm{cov}}(r_i, r_{i-1})$ estimează ω^2 (A2), consistent cînd numărul de randamente crește, cu condiția ca zgomotul să fie i.i.d. și independent de preț
- În eșantioane finite primul estimator este contaminat de $\mathrm{IV}/(2n)$; A7 îl compară cu valoarea oracol

A7 [Propus]: RV pe două scări — enunț

- **Repere și exemple**: A2 [Rezolvat]
- **Date de intrare**: o acțiune nelichidă sintetică, cu o sesiune stabilită de 390 de minute: $\mathrm{IV} = \mathrm{IQ} = 1$, $\omega = 0{,}03\%$; într-o zi, randamentele de 1 minut ($n = 390$) dau $\mathrm{RV}^{\mathbf{all}} = 1{,}70$; $K = 5$ grile decalate de 5 minute dau $\mathrm{RV}^{\mathbf{avg}} = 1{,}10$



1. Calculați deplasarea din zgomot $2n\omega^2$ la 5 și la 1 minut, precum și n^* din A2 cu intervalul său.
2. Cu $\bar{n} = (n - K + 1)/K$, calculați TSRV $= \mathrm{RV}^{\mathbf{avg}} - \frac{\bar{n}}{n}\mathrm{RV}^{\mathbf{all}}$, cu și fără factorul $1/(1 - \bar{n}/n)$.
3. Estimați ω^2 de două ori: valoarea oracol $(\mathrm{RV}^{\mathbf{all}} - \mathrm{IV})/(2n)$ și estimatorul fezabil $\mathrm{RV}^{\mathbf{all}}/(2n)$ din A2 extins.
4. Arătați că TSRV elimină deplasarea din zgomot la ordinul întîi. Opțional: cu $\mathrm{Var}(\text{eroare}) \approx aK/n + bn/K^2$, găsiți K care echilibrează termenii și rata TSRV.

- **Raportați**: deplasările și n^*, cele două valori TSRV, cele două estimări ale lui ω^2 și o propoziție despre cînd merită TSRV

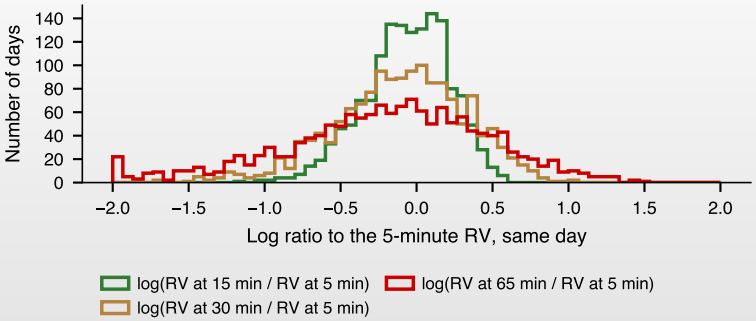
◂ Înapoi ▸ Lista exercițiilor

B1 extins [Rezolvat]: un nucleu realizat pe bare de 5 minute — enunț

- **Întrebare**: schimbă nucleul realizat, un estimator robust la zgomot, valoarea RV SPY pe 5 minute?
- **Date**: randamentele SPY `R` și prețurile `P`; funcția `realized_kernel` Barndorff-Nielsen et al. (2008): ponderi Parzen, lățimea de bandă $H = \lceil 3{,}5134\, \xi^{4/5} n^{3/5} \rceil$ pe fiecare zi ($\lceil \cdot \rceil$: rotunjire în sus), raportul zgomot/semnal $\xi^2 = \hat{\omega}^2 / \sqrt{\widehat{\mathrm{IQ}}}$ (în practică $\sqrt{\widehat{\mathrm{IQ}}} \approx \widehat{\mathrm{IV}}$), la capete, prețurile de deschidere și de închidere ale sesiunii
- **Cerințe**
  1. Calculați nucleul pentru fiecare zi și raportul dintre media lui și media RV, cu bootstrap-ul pe blocuri din B1.
  2. Reprezentați distribuția lui $\ln(\mathrm{RV}^{(k)} / \mathrm{RV}^{(5\,\mathrm{min})})$ pentru 15, 30 și 65 de minute.
- **Raportați**: raportul nucleu/RV cu CI-ul bootstrap, lățimea de bandă mediană H și graficul distribuțiilor
- Notebook: secțiunea B1

Înapoi Lista exercițiilor

B1 extins [Rezolvat]: rezolvare — un nucleu realizat pe bare de 5 minute



- Nucleu / RV $= 0{,}915$, CI $[0{,}882; 0{,}947]$ (lățimea mediană $H = 7$): nucleul adaugă autocorelațiile negative mici ($\hat{\rho}_1 = -0{,}008$) și mediază primele și ultimele două prețuri (jittering)
- Valorile pe grilele mai rare au o dispersie mare în jurul RV pe 5 minute: RV rară adaugă zgomot; nucleul realizat este ceva mai mic: o diferență între estimatori, nu o deplasare din zgomot măsurată

MFM_ch9_seminar

A3 extins [Propus]: robustețea BV la un salt — enunț

- **Repere și exemple**: A1 [Rezolvat]
- **Date de intrare**: un salt de mărime κ într-un interval **interior** j ($1 < j < M$): $r_j = \kappa + r_j^c$; celelalte randamente sînt difuzive, $O_p(\Delta^{1/2})$
- **Cerințe**
  1. Arătați că, la ordinul principal, saltul intră în BV doar prin $|r_{j-1}|\,|\kappa| + |\kappa|\,|r_{j+1}|$ și că aceste produse sînt $O_p(\Delta^{1/2})$.
  2. Deduceți că $\mathrm{BV} \to \mathrm{IV}$, iar $\mathrm{RV} \to \mathrm{IV} + \kappa^2$; prin ce diferă un salt în ultimul interval (ca în A3)?
- **Raportați**: derivarea, limitele BV și RV și comparația cu saltul din ultimul interval din A3

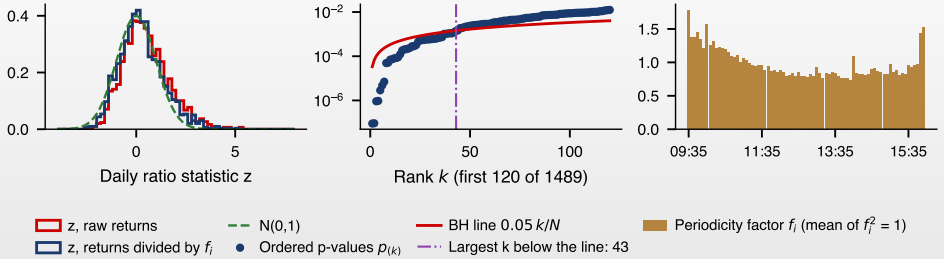
Înapoi Lista exercițiilor

B2 extins [Rezolvat]: multiplicitate și sezonalitate intraday — enunț

- **Întrebare**: cîte dintre respingerile din B2 rămîn după controlul ratei descoperirilor false și după corecția pentru forma de U din cursul zilei?
- **Date**: cele $N = 1\,489$ de valori p zilnice din B2; funcțiile `bh_count`, `periodicity_sd`
- **Cerințe**
  1. Procedura Benjamini–Hochberg (BH) Benjamini & Hochberg (1995): ordonați valorile p și respingeți cele mai mici k, cu k cel mai mare indice pentru care $p_{(k)} \leq kq/N$; aplicați $q = 5\%$ și 1%, precum și Bonferroni ($p < 0{,}05/N$).
  2. Factorul de periodicitate Boudt, Croux & Laurent (2011): scalați randamentele cu $\sqrt{\mathrm{BV}_t / M_t}$, luați rădăcina mediei pătratelor pe fiecare interval de 5 minute, normalizați astfel încît media lui f_i^2 să fie 1; împărțiți $r_{t,i}$ la f_i și repetați B2(2) și pasul 1.
- **Interpretare**: BH controlează FDR pentru valori p valide, independente sau cu dependență pozitivă prin regresie pe submulțimea ipotezelor nule adevărate (PRDS, *positive regression dependency on a subset*) Benjamini & Yekutieli (2001); ce număr ați folosi pentru a construi un regresor de salt?
- **Raportați**: numărul de respingeri cu BH la 5% și 1% și cu Bonferroni, înainte și după corecția de periodicitate, și numărul ales pentru regresorul de salt

◂ Înapoi ▸ Lista exercițiilor

B2 extins [Rezolvat]: rezolvare — multiplicitate și sezonalitate intraday



- BH la 5%: cel mai mare k cu $p_{(k)} \le k \times 0{,}05/1\,489$ este $k = 43$ ($p_{(k)} = 0{,}00140$ față de $0{,}00144$): 43 de zile cu salt; la 1%: 14; Bonferroni: 7
- După împărțirea la f_i ($f = 1{,}78$ la deschidere, $0{,}73$ la minim): 169 de respingeri la 5%, 21 la 0,1%, 17 sub FDR 5%
- **Interpretare**: construim regresorul de salt din zilele cu control FDR și corecție de periodicitate

MFM_ch9_seminar

B4 [Propus]: salturi în Bitcoin — enunț

- **Repere și exemple**: B2 [Rezolvat]
- **Date**: matricea de prețuri Bitcoin Pb, zile UTC de 24 de ore cu cel puțin 95% din cele 288 de bare, 659 de zile din 747; o zi este etichetată cu data ei UTC
- **Cerințe**
  1. Aplicați testul raportului din B2 la 0,1% și construiți un tabel zilnic: data, M_t, RV, BV, z, indicatorul de weekend.
  2. Pentru zilele lucrătoare și pentru weekend raportați numărul de zile, numărul de respingeri și rata lor, precum și media BV/RV.
  3. Calculați ponderea salturilor în variația totală și un test binomial unilateral al ratei totale față de 0,001.
- **Interpretare**: testul binomial presupune teste zilnice independente și cu mărimea corectă
  - ce stabilește o respingere?
  - ce nu stabilește?
  - în care grup sînt respingerile mai frecvente?
  - care ar putea fi motivul?
- **Raportați**: tabel zile lucrătoare/weekend (zile, respingeri, proporție, media BV/RV), contribuția salturilor și valoarea p a testului binomial; graficul statisticii zilnice de salt cu weekendurile marcate; patru propoziții de interpretare

◂ Înapoi ▸ Lista exercițiilor

B5 [Propus]: randamente standardizate cu volatilitatea realizată — enunț

- **Repere și exemple**: B1 [Rezolvat] (bootstrap pe blocuri)
- **Date**: randamentele SPY de la deschidere la închidere (1 489 de zile) și randamentele Bitcoin pe 24 de ore (659 de zile), fiecare aliniat cu RV din aceeași zi Andersen et al. (2001); un rînd: 7 aprilie 2025, $r_t = 3{,}09\%$, $\mathrm{RV}_t = 45{,}1\,\%^2$, randamentul standardizat $z_t = r_t/\sqrt{\mathrm{RV}_t} = 0{,}46$
- **Cerințe**
  1. Calculați kurtosis-ul Pearson $\mathbb{E}(x-\bar{x})^4/s^4$ pentru $u_t = r_t/\mathrm{sd}(r)$ și pentru z_t, cu intervale de 95% din bootstrap pe blocuri suprapuse, pe perechi (blocuri de 20 de zile, 2 000 de reeșantionări); aceste intervale sînt aproximative: inferența Normală pentru kurtosis cere momente pînă la ordinul opt și dependență slabă, iar un moment de ordin patru finit nu ajunge.
  2. Reprezentați histogramele și graficele cuantilă–cuantilă (QQ) față de distribuția Normală pentru u_t și z_t; aplicați Jarque–Bera (JB) pe z_t; calculați ponderea $|z_t| > 3$ (distribuția Normală: 0,27%).
- **Interpretare**: de unde provin cozile grele? Reper opțional: pentru randamente intraday i.i.d. cu distribuție Normală, z_t are kurtosis-ul $3M/(M+2)$
- **Raportați**: rezumatele distribuțiilor, statisticile cerute și intervalele bootstrap, cu interpretarea Normalității

înapoi Lista exercițiilor

A4 extins [Rezolvat]: testul Wald al restricțiilor HAR — enunț

- **Repere și exemple**: A4 [Rezolvat]
- **Cerințe**
  1. Scrieți statistica Wald pentru $R\phi = 0$ cu o covarianță Newey–West $\hat{V}_{\text{NW}}$ și dați distribuția ei limită sub HAR.
  2. Explicați de ce este necesară o covarianță HAC pentru regresia log-varianței.
- **Raportați**: statistica Wald, distribuția ei limită și motivul pentru covarianța HAC

Înapoi Lista exercițiilor

A4 extins [Rezolvat]: rezolvare — testul Wald al restricțiilor HAR

- $W = (R\hat{\phi})^\top (R\hat{V}_{\text{NW}} R^\top)^{-1} R\hat{\phi} \to \chi^2_{19}$ sub cele 19 restricții HAR
- Erorile regresiei pe log-varianță sînt heteroscedastice și pot fi autocorelate: covariança OLS obișnuită ar evalua greșit $\text{Var}(R\hat{\phi})$
- Testul pe date este în B3 extins, pasul 1

Ce schimbă extensiile HAR?

- HARQ (HAR cu cvarticitatea realizată) adaugă la HAR cvarticitatea realizată RQ, care măsoară incertitudinea cu care RV estimează IV
- HAR-CJ (HAR cu componentă continuă și componentă de salt, *Continuous and Jump*) separă variația continuă (C) de salturi (J)
- SHAR (HAR cu semivarianțe, *Semivariance HAR*) separă variația din randamentele pozitive de cea din randamentele negative
- Cele două semivarianțe însumate refac varianța realizată

B3 extins [Rezolvat]: testul AR(22) și HARQ — enunț

- **Repere și exemple**: A4 extins, B3 [Rezolvat]
- **Date**: ca la B3; cvarticitatea realizată RQ_t din randamentele intraday (nu acoperă randamentul peste noapte)
- **Cerințe**
  1. Estimați AR(22) nerestricționat pentru $\ln V_t$ și testați cele 19 restricții HAR cu un test Wald Newey–West.
  2. Estimați HARQ pe niveluri Bollerslev, Patton & Quaedvlieg (2016),
     $V_{t+\mathbf{1}} = \beta_{\mathbf{0}} + (\beta_d + \beta_Q \sqrt{\mathrm{RQ}_t}) V_t + \beta_w V_t^{(w)} + \beta_m V_t^{(m)} + \varepsilon$, pe RV intraday și pe varianța totală.
  3. Prognozați din 3 ianuarie 2022 pe o fereastră extinsă, înlocuind prognozele care ies din intervalul $[\min, \max]$ al valorilor din eșantion cu media eșantionului Bollerslev, Patton & Quaedvlieg (2016), și aplicați testul Diebold–Mariano (DM) față de HAR pe niveluri după QLIKE.
- **Raportați**: W și valoarea p, $\hat{\beta}_Q$ cu erorile standard pe cele două ținte, QLIKE pentru HARQ și HAR și testul DM
- Notebook: secțiunea B3

◂ Înapoi ▸ Lista exercițiilor

B3 extins [Rezolvat]: rezolvare — testul AR(22) și HARQ

- (1) $W = 18{,}2$, $p = 0{,}51$ (χ^2_{19}); R^2 0,477 (AR(22)) față de 0,470 (HAR): restricțiile nu sînt respinse; HAR economisește 19 parametri
- (2) $\hat{\beta}_Q = -0{,}0088$ (SE 0,0042) pe RV intraday, semnificativ la 5%; $-0{,}0022$ (SE 0,0024) pe varianța totală, nesemnificativ
- (3) QLIKE 0,344 (HARQ) față de 0,326 (HAR), DM $t = 1{,}23$, $p = 0{,}22$: nicio diferență semnificativă
- **Interpretare**: semnul negativ așteptat pentru β_Q (atenuarea din eroarea de măsurare, A8) apare, dar cu RV pe 5 minute eroarea de măsurare este mică, iar randamentul peste noapte, neacoperit de RQ, domină zgomotul din varianța totală

A8 [Propus]: atenuarea și argumentul pentru HARQ — enunț

- **Repere și exemple**: A4 [Rezolvat], B3 extins
- **Date de intrare**: $\mathrm{RV}_t = V_t^* + u_t$ (semnal latent plus eroare), V_t^* varianța integrată intraday (IV), diferită de varianța totală V_t cu randamentul peste noapte; $\mathbb{E}[u_t \mid \sigma] = 0$, $\mathrm{Var}(u_t \mid \sigma) = 2\,\mathrm{IQ}_t/M$ (asimptotic, A1); erori necorelate între zile și necorelate cu V^* și η; $V^*_{t+\mathbf{1}} = \beta_{\mathbf{0}} + \beta V_t^* + \eta_{t+\mathbf{1}}$, $\beta > 0$
- **Cerințe**
  1. Folosind plim $\hat{\beta} = \mathrm{cov}(\mathrm{RV}_{t+\mathbf{1}}, \mathrm{RV}_t)/\mathrm{Var}(\mathrm{RV}_t)$, derivați atenuarea pantei OLS.
  2. Ilustrare (varianța semnalului $\sigma^{\mathbf{2}}_{V^*}$ presupusă independentă de IQ_t): derivați ponderea liniară optimă pe RV_t dat fiind IQ_t și arătați că scade cu IQ_t.
  3. Scrieți un estimator prin variabile instrumentale cu instrumentul $\mathrm{RV}_{t-\mathbf{1}}$ și precizați condițiile de excludere și de relevanță.
- **Raportați**: formula atenuării și cazurile ei limită, cu rolul incertitudinii de măsurare

Înapoi Lista exercițiilor

B6 [Propus]: contribuția salturilor: modelul HAR-CJ — enunț

- **Repere și exemple**: B2 și B3 [Rezolvat]
- **Date**: SPY ca la B3; varianta de replicare: zilele cu salt de la 0,1% din B2, fără corecții; $J_t = \mathrm{RV}_t - \mathrm{BV}_t$ în acele zile, 0 în rest; $C_t = \mathrm{RV}_t - J_t + r^{\mathbf{2}}_{\mathbf{on},t}$: partea intraday clasificată continuă, plus pătratul randamentului peste noapte, care poate conține salturi neidentificate (toate în $\%^{\mathbf{2}}$) Andersen, Bollerslev & Diebold (2007)
- **Cerințe**
  1. Regresați $\ln V_{t+\mathbf{1}}$ pe $\ln C_t$, $\ln C_t^{(\mathbf{5})}$, $\ln C_t^{(\mathbf{22})}$, $\ln(1+J_t)$, $\ln(1+J_t^{(\mathbf{5})})$, $\ln(1+J_t^{(\mathbf{22})})$, unde (5), (22) sînt medii aritmetice logaritmate după mediere; raportați toți coeficienții cu SE Newey–West și testați comun cei trei coeficienți ai salturilor.
  2. Prognozați din 3 ianuarie 2022 pe o fereastră extinsă, reveniți la niveluri cu $\exp(\hat{y} + \hat{s}^{\mathbf{2}}/2)$ ca în B3 și comparați QLIKE cu log-HAR pe date comune, prin DM.
  3. Sensibilitate: repetați cu zilele cu salt din B2 extins (control FDR și corecție de periodicitate), estimînd factorii de periodicitate și aplicînd selecția BH la fiecare origine doar pe datele disponibile atunci sau fixîndu-le pe un eșantion inițial separat.
- **Interpretare**: HAR-CJ mediază nivelurile înainte de logaritmare, log-HAR mediază logaritmii: cele două modele nu sînt imbricate; de ce un coeficient al salturilor semnificativ poate să nu aducă cîștig de prognoză?
- **Raportați**: tabelul coeficienților, testul Wald comun, pierderile de prognoză și comparația DM; precizați dacă semnificația se traduce în cîștig de prognoză

◂ Înapoi ▸ Lista exercițiilor

A5 extins [Rezolvat]: lipsa de robustețe a MSE pe volatilitate — enunț

- **Repere și exemple**: A5 [Rezolvat]
- **Date de intrare**: IV constantă și $\mathrm{RV} = \mathrm{IV}\,\chi^2_M/M$; momentul dat: $\mathbb{E}\sqrt{\chi^2_M} = \sqrt{2}\,\Gamma(\frac{M+1}{2})/\Gamma(\frac{M}{2})$, Γ: funcția gamma
- **Cerințe**
  1. Minimizați $\mathbb{E}(\sqrt{\mathrm{RV}} - \sqrt{F})^2$ (MSE pe volatilitate) în raport cu F și comparați valoarea optimă cu $\mathbb{E}\mathrm{RV}$ prin inegalitatea lui Jensen.
  2. Calculați F^*/IV pentru $M = 1$ (randamentul zilnic la pătrat), 6 și 78.
- **Raportați**: valoarea optimă F^*, comparația ei cu $\mathbb{E}\mathrm{RV}$ și cele trei rapoarte F^*/IV

Înapoi Lista exercițiilor

A5 extins [Rezolvat]: rezolvare — lipsa de robustețe a MSE pe volatilitate

- $\mathbb{E}(\sqrt{\mathrm{RV}} - \sqrt{F})^2 = \mathbb{E}\mathrm{RV} - 2\sqrt{F}\,\mathbb{E}\sqrt{\mathrm{RV}} + F$, minim în $\sqrt{F} = \mathbb{E}\sqrt{\mathrm{RV}}$; din Jensen $(\mathbb{E}\sqrt{\mathrm{RV}})^2 < \mathbb{E}\mathrm{RV}$: prognoza optimă este deplasată în jos
- $F^*/\mathrm{IV} = \frac{2}{M}\left(\Gamma(\frac{M+1}{2})/\Gamma(\frac{M}{2})\right)^2 = 0{,}637$ ($M = 1$), $0{,}920$ ($M = 6$), $0{,}994$ ($M = 78$); reprezentat grafic în A5
- **Interpretare**: față de randamentele zilnice la pătrat, MSE pe volatilitate răsplătește o prognoză cu 36% prea mică; cu RV pe 5 minute distorsiunea aproape dispare

B7 [Propus]: prognoza săptămînii următoare — enunț

- **Repere și exemple**: B3 [Rezolvat], A6, evaluarea prognozelor
- **Date**: varianța totală SPY V_t (B3) și prețurile de închidere ajustate SPY din 2000, pentru GARCH; ținta $Y_t = V_t + \ldots + V_{t+4}$, prognozată la închiderea din $t-1$; 1 176 de origini din 3 ianuarie 2022
- **Cerințe**
  1. log-HAR direct: regresați $\ln Y$ pe regresorii din B3, reestimat la fiecare origine folosind doar țintele observate complet pînă în $t-1$ (ultimele patru nu sînt); prognoza $\exp(\hat{y} + \hat{s}^2/2)$ (corecție lognormală).
  2. GARCH(1,1)-t reestimat la fiecare 21 de zile: cu σ_t^2, ω, a, b ca în Prognoze în afara eșantionului și varianța de termen lung $\bar{\sigma}^2 = \omega/(1-a-b)$, prognoza $\sum_{h=0}^{4}[\bar{\sigma}^2 + (a+b)^h(\sigma_t^2 - \bar{\sigma}^2)]$.
  3. Calculați QLIKE și MSE, DM pe QLIKE cu erori HAC, ambii coeficienți de calibrare MZ (aici a este termenul liber și b panta, diferiți de coeficienții GARCH din pasul 2) cu testul comun pentru $(a, b) = (0, 1)$ și reprezentați diferența cumulată a QLIKE.
- **Interpretare**
  - se păstrează avantajul de la o zi al log-HAR la un orizont săptămînal?
  - ce efect au țintele suprapuse asupra testului DM?
- **Raportați**: tabel QLIKE/MSE; statistica DM și valoarea p; coeficienții MZ și valoarea p comună; graficul diferenței cumulate QLIKE și două propoziții de interpretare

înapoi Lista exercițiilor

De la pierderile condiționate la un test comun

- La o origine, instrumentele $Z = (1; 0{,}2)$ și diferența de pierdere $d = -0{,}1$ dau $g = Zd = (-0{,}1; -0{,}02)$
- A doua componentă testează dacă diferențele trecute le prezic pe cele viitoare
- Exemplu didactic: $n = 100$, $\bar{g} = (0{,}02; -0{,}01)$, $\hat{\Gamma}_0 = \mathrm{diag}(0{,}02; 0{,}006)$, $\hat{\Gamma}_1 = \mathrm{diag}(0{,}01; 0{,}002)$, $\hat{\Gamma}_2 = \hat{\Gamma}_3 = \hat{\Gamma}_4 = 0$ ($\hat{\Gamma}_k$: autocovarianța lui g la lagul k)
- $\hat{\Omega} = \hat{\Gamma}_0 + \sum_{k=1}^{4}(\hat{\Gamma}_k + \hat{\Gamma}_k^\top) = \mathrm{diag}(0{,}04; 0{,}01)$, deci $W = 100[0{,}02^2/0{,}04 + (-0{,}01)^2/0{,}01] = 2$
- Față de χ^2_2, $p = e^{-1} \approx 0{,}368$: nu respingem la 5%
- Nerespingerea nu arată că metodele sînt la fel de precise în orice stare

B7 extins [Propus]: un test Giacomini–White condiționat — enunț

- **Repere și exemple**: exemplul numeric precedent; B7 furnizează cadrul prognozelor
- Semnul, ca în A6: d_t = QLIKE(log-HAR) minus QLIKE(GARCH); o medie condiționată pozitivă a lui d_t favorizează GARCH. Sub abilitate predictivă condiționată egală, statistica de la pasul 2 urmează χ^2_2; raportați valoarea p și decizia la 5%.
- **Cerințe**
  1. Reestimați ambele metode pe ferestre mobile de lungime fixă (log-HAR: ultimele 250 ținte; GARCH: ultimele 1 000 de randamente), cum cere teoria din Giacomini & White (2006).
  2. Cu instrumentele $Z_t = (1, d_{t-\mathbf{5}})$, cunoscute la originea $t-1$, și diferența de pierdere d_t a țintei Y_t, cunoscută după închiderea din $t+4$, calculați $n\,\bar{g}^\top \hat{\Omega}^{-\mathbf{1}} \bar{g}$ cu $g_t = Z_t d_t$ (n: numărul de origini, $\bar{g}$: media lui g, $\hat{\Omega}$: covarianța ei de termen lung); sub ipoteza nulă g este necorelat dincolo de lagul 4, deci $\hat{\Omega}$ însumează autocovariațele neponderate de ordinele 0–4.
- **Raportați**: pierderile QLIKE medii, statistica Giacomini–White, valoarea ei p, decizia la 5% și testul necondiționat

Înapoi Lista exercițiilor

Prognoze comparabile înaintea testului MCS

- Mulțimea de încredere a modelelor (MCS, *Model Confidence Set*) păstrează prognozele pe care comparația comună nu le poate separa
- SHAR se estimează pe niveluri de varianță, pe fereastră extinsă, cu prima prognoză la 3 ianuarie 2022
- Termenul liber și termenii săptămînal și lunar sînt ca la HAR; semivariațele pozitivă și negativă înlocuiesc termenul zilnic
- Folosiți doar rezultate observate înaintea datei prognozate
- O prognoză în afara intervalului $[\min, \max]$ al variației din eșantionul de estimare se înlocuiește cu media eșantionului, ca la HARQ
- `arch.bootstrap.MCS`, cu bootstrap staționar (blocuri de lungime aleatoare, în medie 20): reeșantionați rînduri întregi, simultan pentru toate modelele
- Raportați ordinea eliminărilor și ambele mulțimi, de 90% și de 75%; explicați de ce un model poate rămîne în prima, dar nu în a doua

B8 [Propus]: mulțimea de încredere a modelelor — enunț

- **Repere și exemple**: B3 extins, B7
- **Date**: prognozele SPY pe o zi (log-HAR, HAR, GARCH(1,1)-t, EWMA, RV de ieri) plus HARQ (B3 extins) și SHAR Patton & Sheppard (2015): $\beta_d V_t$ înlocuit cu $\beta^+ \mathrm{RS}_t^+ + \beta^- \mathrm{RS}_t^-$, $\mathrm{RS}_t^{\pm} = \sum_i r_{t,i}^2 1\{r_{t,i} \gtrless 0\}$, randamentul peste noapte socotit ca încă un randament; 1 180 de date comune
- **Cerințe**
  1. Construiți matricea QLIKE dată × model pe date comune și verificați că toate prognozele sînt pozitive.
  2. Aplicați MCS Hansen, Lunde & Nason (2011) cu statistica amplitudinii, blocuri de 20 de zile, 10 000 de reeșantionări, valoarea de inițializare 42: `MCS(L, size=0.10, reps=10000, block_size=20, method='R', seed=42)`.
  3. Raportați ordinea eliminării și valoarea p MCS a fiecărui model; comparați mulțimile de 90% (mărimea 0,10) și 75% (mărimea 0,25).
- **Interpretare**: de ce este MCS preferabilă celor 21 de teste DM pe perechi?

înapoi Lista exercițiilor

B9 [Propus]: estimarea exponentului Hurst al volatilității — enunț

- **Repere și exemple**: B1 [Rezolvat] (bootstrap pe blocuri), Capitolul 9 (exponentul Hurst H; $H < 0{,}5$: volatilitate aspră)
- **Date**: serii zilnice SPY pentru patru măsuri: varianța totală, RV intraday, BV, RV subeșantionată la 30 de minute; $x_t = \frac{1}{2}\ln(\text{măsura}_t)$; decalaje Δ între 1 și 30 de zile de tranzacționare Gatheral, Jaisson & Rosenbaum (2018)
- **Cerințe**
  1. Calculați $m(q, \Delta)$, media lui $|x_{t+\Delta} - x_t|^q$ pe toate perechile valide, pentru $q \in \{0{,}5; 1; 1{,}5; 2; 3\}$; verificați $m(2, 1)$ pentru varianța totală față de media directă a diferențelor de ordinul întîi la pătrat, pe aceleași perechi valide.
  2. Pentru fiecare q regresați $\ln m(q, \Delta)$ pe $\ln \Delta$ cu termen liber: panta ζ_q; apoi H = panta lui ζ_q pe q, prin origine.
  3. CI de 95%: bootstrap pe blocuri suprapuse, blocuri de 60 de zile, 300 de reeșantionări, incrementele calculate doar în interiorul fiecărui bloc.
  4. Simulați o traiectorie de lungimea seriei SPY de varianță totală, cu `numpy.random.default_rng(42)` și laguri 1–30. Folosiți incremente Normale ($N(0; 0{,}1^2)$, valoarea de inițializare 42), scădeți `linspace(0,z[-1],n)` din traiectoria cumulată z (tot $H = 0{,}5$), estimați H, apoi extrageți zgomot i.i.d. $N(0; 0{,}25^2)$ din același generator, după incremente. Adăugați-l traiectoriei din care s-a eliminat deja tendința, fără o nouă eliminare, și estimați din nou.
- **Interpretare**
  - este H sub 0,5 pentru fiecare măsură?
  - ce efect are eroarea de măsurare?
- O singură traiectorie este o ilustrare, nu un studiu de mărime sau putere
- **Raportați**: estimările H și intervalele pentru fiecare măsură, graficele de scalare și comparația simulării fără/cu zgomot

◂ Înapoi ▸ Lista exercițiilor

C2 [Propus]: previzibilitatea volatilității Bitcoin față de S&P 500 — enunț

- **Date**: prognozele pe o zi pentru SPY și Bitcoin, aceeași fereastră calendaristică 1 martie 2025 – 18 septembrie 2026 (567 de zile calendaristice: SPY 389 de zile de tranzacționare, Bitcoin 422 de zile)
- **Cerințe**
  1. Definiți trei măsuri și calculați-le pentru fiecare activ: R^2 de calibrare al lui $\ln V_t$ pe $\ln F_t$; cîștigul QLIKE $1 - \overline{\text{QLIKE}}_{\textbf{log-HAR}} / \overline{\text{QLIKE}}_{\textbf{RV ieri}}$; același cîștig față de GARCH.
  2. Construiți un CI de 95% pentru diferența de R^2 (Bitcoin minus SPY): reeșantionați aceleași blocuri de 28 de zile calendaristice pentru ambele active (de 2 000 de ori), păstrînd zilele lipsă ale fiecărui activ.
  3. Adăugați în log-HAR pentru Bitcoin o variabilă pentru ziua-țintă de weekend, folosiți mediile HAR pe 7/30 de zile și o fereastră de estimare extinsă cu toate observațiile complete dinaintea fiecărei date-țintă, începînd cu 1 martie 2025; păstrați corecția lognormală și aplicați DM după QLIKE cu și fără indicator.
- **Raportați**: tabelul, CI, rezultatul DM și un răspuns în două propoziții; extensie de proiect: Ethereum sau o acțiune de la Bursa de Valori București (BVB) și măsura de entropie din Pele & Mazurencu-Marinescu-Pele (2019)

înapoi Lista exercițiilor

C3 [Propus]: învățarea automată față de HAR pentru SPY — tema de proiect (1/4)

- Context: Christensen, Siggaard & Veliyev (2023) (Capitolul 9, studiul de caz)
  - pădurile aleatoare (RF, *random forest*) și rețelele neuronale întrec HAR cînd se adaugă predictori la valorile trecute ale RV
  - avantajul crește odată cu orizontul de prognoză
  - testați acest rezultat pe SPY, cu împărțirea eșantionului și setările modelelor din lucrare
- (a) Construiți seriile-țintă (secțiunea Data description a lucrării).
  - RV zilnică SPY de la deschidere la închidere, din barele de 5 minute ale sesiunii obișnuite (78 de randamente pe zi, ca la B2)
  - trei ținte: RV_{t+1}, media RV pe următoarele 5 zile și media pe următoarele 22 de zile
- (b) Construiți două mulțimi de predictori disponibili la momentul t.
  - $\mathcal{M}_{\mathsf{HAR}}$: RV zilnică, săptămînală și lunară
  - $\mathcal{M}_{\mathsf{ALL}}$ adaugă VIX (VIX.INDX.csv, close), pătratul randamentului log zilnic al indicelui Hang Seng (HSI.INDX.csv, close) și momentumul SPY pe o săptămînă
  - și diferențele de ordinul întîi ale logaritmului volumului în dolari și ale ratei titlurilor de stat la 3 luni (Federal Reserve Economic Data, FRED; seria `DTB3`)
  - și indicele incertitudinii politicii economice (FRED, `USEPUINXD`) și indicele Aruoba–Diebold–Scotti (Philadelphia Fed)

înapoi Lista exercițiilor

C3 [Propus]: învățarea automată față de HAR pentru SPY — tema de proiect (2/4)

- (b, continuare) Dataţi fiecare predictor după data publicării, nu după data la care se referă.
  - folosiți doar valorile publicate pînă la închiderea zilei t
  - indicele Aruoba–Diebold–Scotti este revizuit: folosiți versiunile în timp real ale Philadelphia Fed
  - precizați seriile pentru care aceste versiuni nu sînt disponibile
  - precizați care dintre predictorii lucrării (Tabelul 1) nu pot fi construiți pentru un fond indexat și de ce: o limitare a datelor, nu un motiv pentru a schimba metoda
- (c) Împărțiți eșantionul ca în Anexa A.1 a lucrării.
  - primele 1 000 de zile utilizabile pentru antrenare, următoarele 200 pentru validare, restul pentru test
  - un rînd aparține unui subset doar dacă întreaga fereastră a țintei se află în el: eliminați ultimele $h-1$ rînduri dinaintea fiecărei granițe
  - raportați numărul de observații utilizabile pentru fiecare orizont
  - standardizați fiecare predictor doar cu media și varianța din setul de antrenare

C3 [Propus]: învățarea automată față de HAR pentru SPY — tema de proiect (3/4)

- (d) Estimați modelele liniare.
  - HAR, log-HAR și OLS pe toți predictorii, pe o fereastră mobilă care reunește seturile de antrenare și validare
  - prognoza log-HAR: $\exp(\hat{f} + \hat{\sigma}^2/2)$, $\hat{f}$: valoarea ajustată pentru log RV, $\hat{\sigma}^2$: varianța reziduurilor
  - elastic net pe o fereastră mobilă de antrenare, fără reunire; λ și ponderea de combinare alese pe setul de validare (Secțiunea 2 a lucrării)
  - HAR își păstrează cei trei regresori RV în ambele mulțimi de informații
- (d, continuare) Estimați modelele de învățare automată (Tabelul A.6 din lucrare).
  - RF: 500 de arbori pe fereastra reunită, $\lfloor J/3 \rfloor$ predictori candidați la fiecare ramificare (J: numărul de predictori), cel puțin 5 observații într-un nod
  - NN_2^{10}: straturi de 4 și 2 neuroni, leaky ReLU (*Rectified Linear Unit*, funcția de activare liniară rectificată), Adam cu rata de învățare 0,001, dropout 0,8 (probabilitatea eliminării unui neuron la antrenare; toți neuronii activi la predicție)
  - oprire timpurie (*early stopping*) după 100 de epoci fără îmbunătățire; media celor mai bune 10 din 100 de inițializări; antrenată o singură dată pe o fereastră fixă
  - înlocuiți prognozele negative cu cea mai mică RV din eșantion

C3 [Propus]: învățarea automată față de HAR pentru SPY — tema de proiect (4/4)

- (e) Raportați comparația pe setul de test, ca la B8.
  - un tabel cu MSE și QLIKE relative la HAR, pentru cele trei orizonturi și ambele mulțimi de informații
  - valorile p ale testului DM unilateral față de HAR și MCS de 90%
  - țintele suprapuse de 5 și 22 de zile: erori HAC cu lățimea de bandă peste $h-1$ (ponderile Bartlett reduc contribuția lagului $h-1$) și o analiză de sensibilitate după lățimea de bandă
- Interpretare
  - se păstrează pe SPY clasamentul din Tabelele 2, 3 și 7 ale lucrării?
  - crește avantajul ML (*Machine Learning*, învățare automată) cu orizontul?
  - se păstrează avantajul ML după QLIKE?
  - cu un set de test de această lungime, ce vă poate spune un test DM nesemnificativ?
- Exemplu de urmat: B3, A5.
- Extensie de proiect
  - repetați pe RV Bitcoin pe 24 de ore
  - comparați între modele importanța RV și a VIX, măsurată prin efectele locale acumulate (Figura 7 a lucrării)

Acronime folosite în acest material (1/2)

- **ACF**: AutoCorrelation Function — *funcția de autocorelație*
- **AI**: Artificial Intelligence — *inteligență artificială*
- **AR**: AutoRegressive (model) — *model autoregresiv*
- **BH**: Benjamini–Hochberg (procedure) — *procedura Benjamini–Hochberg*
- **BV**: Bipower Variation — *variația bipower*
- **BVB**: Bursa de Valori București
- **CI**: Confidence Interval — *interval de încredere*
- **DM**: Diebold–Mariano (test) — *testul Diebold–Mariano de comparare a prognozelor*
- **ETF**: Exchange-Traded Fund — *fond tranzacționat la bursă*
- **EWMA**: Exponentially Weighted Moving Average — *medie mobilă ponderată exponențial*
- **FDR**: False Discovery Rate — *rata descoperirilor false*
- **FRED**: Federal Reserve Economic Data — *baza de date economice a Rezervei Federale (St. Louis)*
- **FWER**: Family-Wise Error Rate — *probabilitatea de cel puțin o eroare de tip I în familia de teste*
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity — *heteroscedasticitate condiționată autoregresivă generalizată*
- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors) — *erori standard robuste la heteroscedasticitate și autocorelație*
- **HAR**: Heterogeneous AutoRegressive (model) — *model autoregresiv heterogen*
- **HAR-CJ**: HAR model with Continuous and Jump components — *model HAR cu componentă continuă și componentă de salt*
- **HARQ**: HAR model with Realised Quarticity — *model HAR cu cvarticitatea realizată*

Acronime folosite în acest material (2/2)

- **IQ**: Integrated Quarticity — *cvarticitatea integrată*
- **JB**: Jarque–Bera (test) — *testul de normalitate Jarque–Bera*
- **MCS**: Model Confidence Set — *mulțimea de încredere a modelelor*
- **ML**: Machine Learning — *învățare automată*
- **MSE**: Mean Squared Error — *eroarea pătratică medie*
- **MZ**: Mincer–Zarnowitz (regression) — *regresia Mincer–Zarnowitz*
- **NW**: Newey–West (standard errors) — *erori standard Newey–West*
- **OLS**: Ordinary Least Squares — *metoda celor mai mici pătrate*
- **PRDS**: Positive Regression Dependence on a Subset (of p-values) — *dependența pozitivă de regresie pe o submulțime (a valorilor p)*
- **QLIKE**: Quasi-LIKElihood (loss function) — *funcția de pierdere de tip cvasi-verosimilitate*
- **QQ**: Quantile–Quantile (plot) — *graficul cuantilă–cuantilă*
- **RF**: Random Forest — *pădure aleatoare*
- **RQ**: Realised Quarticity — *cvarticitatea realizată*
- **RV**: Realised Variance — *varianța realizată*
- **SE**: Standard Error — *eroare standard*
- **SHAR**: Semivariance HAR model — *model HAR cu semivarianțe*
- **TQ**: Tripower Quarticity — *cvarticitatea tripower*
- **TSRV**: Two-Scale Realised Variance — *varianța realizată pe două scări*
- **UTC**: Coordinated Universal Time (Temps Universel Coordonné) — *timpul universal coordonat*
- **VIX**: CBOE Volatility IndeX — *indicele de volatilitate CBOE*

Bibliografie (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.

Bibliografie (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.