








Modelarea piețelor financiare

Seminarul 20: Semnătura traiectoriilor și volatilitatea 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

Ajută forma trecutului la prognoza volatilității?

- Întrebările seminarului
  - **ajută forma traiectoriei recente a volatilității la prognoza variației realizate?**
  - **cum testăm corect acest lucru?**
- Livrabile
  - răspunsurile la A1–A3 și la aplicația independentă obligatorie A6
  - cîte o figură și un tabel pentru B1, B2, B3 și B4; B5 (opțional) transferă analiza B1 la alt activ
  - codul corectat din C1
  - o propoziție de interpretare pentru fiecare exercițiu
- Criterii de evaluare
  - derivări corecte
  - inferență cu erori standard robuste la heteroscedasticitate și autocorelație (HAC, *Heteroskedasticity and Autocorrelation Consistent*) și alternativa precizată
  - o interpretare pentru fiecare rezultat
  - critica răspunsului unui asistent de inteligență artificială (AI, *Artificial Intelligence*), documentată în `AI_ERRORS.md`
- Notebook-ul de azi: deschideți notebook-ul seminarului în Colab
  - fiecare cerință indică secțiunea din notebook
- Seminar: prof. Daniel Traian Pele, danpele@ase.ro; drd. Antoaneta Amza, antoaneta.amza@csie.ase.ro

Ce măsoară indicatorii realizați?

- VOLARE (*VOLatility Archive for Realized Estimates*) este o arhivă de indicatori zilnici construiți din prețuri intrazilnice (*intraday*); fiecare activ își păstrează calendarul de tranzacționare.
- JPM este simbolul bursier (NYSE) al JPMorgan Chase & Co.; C2 folosește toate cele 40 de acțiuni. ES este contractul futures E-mini S&P 500 (*Standard & Poor's*, simbolul CME), iar CL contractul futures pe petrol WTI (simbolul NYMEX). Aici ES este codul unui contract, nu Expected Shortfall.
- FX (*Foreign Exchange*) desemnează piața valutară. Codul unei perechi alipește două monede, de ex. EURUSD = EUR (moneda euro)/USD, în ordinea monedă de bază/monedă de cotare: EUR (euro), GBP (lira sterlină), AUD (dolarul australian), USD (dolarul american), CAD (dolarul canadian), JPY (yenul japonez).

Datele folosite

Setul de date	Active	Eșantion	Folosit în
Măsuri realizate VOLARE, acțiuni americane	JPM (toate cele 40 de acțiuni în C2)	2 ian. 2015 – 30 ian. 2026	A7, B1–B3, C2
Măsuri realizate VOLARE, futures	ES (E-mini S&P 500), CL (petrol)	28 sept. 2009 – 30 ian. 2026	B5–B7
Rezultatele Capitolului 20 (incluse în notebook)	50 de active	valori p (*p-values*) unilaterale ale testului Diebold–Mariano (DM)	B4, B8

- Variabilele din datele VOLARE Cipollini et al. (2026), cîte o valoare pe activ și zi de tranzacționare
  - Varianța realizată la cinci minute RV, semivarianțele RV_t^+, RV_t^- și cvarticitatea realizată RQ (definițiile)
  - Prețurile zilnice de deschidere și de închidere, pentru randamentul logaritmic r_t de la deschidere la închidere
- VOLARE este gratuit pentru cercetare la volare.unime.it: descărcați datele de varianță realizată (acțiuni, futures, FX) și setați calea către ele în prima celulă a notebook-ului

Pregătire: rulați și verificați

```
# 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))
```

- Rezultatul așteptat: 2 786 zile de tranzacționare, de la 2015-01-02 la 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

- O RV de $1{,}46 \times 10^{-4}$ înseamnă o volatilitate zilnică de $\sqrt{1{,}46 \times 10^{-4}} = 1{,}2\%$
- Dacă numărul de zile sau prima valoare diferă, verificați setul de date și activul

De ce contează scala timpului?

- Calculele manuale A1–A6 folosesc timpul în zile (canalul de timp și semnătura sînt definite în slide-urile următoare): fiecare pas zilnic contribuie cu o unitate în canalul de timp. Astfel, geometria exemplelor poate fi verificată ușor.
- Variabilele empirice din Capitolul 20 rescalează canalul de timp al fiecărei ferestre pe intervalul de la zero la unu. O fereastră cu 22 de observații are atunci 21 de pași de timp egali.
- Schimbarea scalei modifică fiecare termen al semnăturii care conține o coordonată de timp. Aria numerică din exercițiul manual nu se compară direct cu variabila empirică construită pe traiectoria rescalată.
- Păstrăm aceeași convenție cînd comparăm traiectorii sau calculăm distanțe pentru ponderi. Rescalăm înainte de calcularea semnăturii, fără să amestecăm variabile construite în convenții diferite.

O fereastră de date este o traiectorie

- Detalii (teoremele, specificația completă, rezultatele pe 50 de active): Capitolul 20
- Considerăm ultimele L zile din cîteva serii și unim punctele prin segmente: o traiectorie

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

  - L: lungimea ferestrei în zile (22 în exerciții); d: numărul de canale (coordonate), aici $d = 3$; $[a, b]$: intervalul de timp al ferestrei
  - RV_t: varianța realizată în ziua t; r_s: randamentul logaritmic de la deschidere la închidere al zilei s, deci al treilea canal cumulează randamente logaritmice intrazilnice, fără mișcările peste noapte (*overnight*)
- **Augmentarea cu timpul**: canalul 1 este chiar timpul
  - Fără el, o mișcare rapidă și una lentă, cu aceleași valori, nu pot fi distinse
- Incrementele: $\Delta_k = X_{t_k} - X_{t_{k-\mathbf{1}}} \in \mathbb{R}^d$, deplasarea tuturor canalelor pe segmentul k
  - Δ_k^i: componenta i a lui Δ_k; pentru canalul de timp $\Delta_k^{\mathbf{1}} = 1$ zi

Semnătura

- **Semnătura**: toate integralele iterate ale traiectoriei, grupate pe nivelul k Chevyrev & Kormilitzin (2016)

$$S^{i_1 \cdots i_k} = \int_{a<u_1<\cdots<u_k<b} dX^{i_1}_{u_1} \cdots dX^{i_k}_{u_k}$$

  - $(i_1, \ldots, i_k)$: un cuvînt, o listă ordonată de k canale din $\{1, \ldots, d\}$; $u_1 < \cdots < u_k$: momente ordonate; dX^i_u: mișcarea canalului i la momentul u

- Nivelul 1: incrementele totale; nivelul 2: matricea $d \times d$ a mișcărilor lui j ponderate cu cît s-a deplasat deja i

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

  - $S^{(1)}$, $S^{(2)}$: nivelul 1 (un vector) și nivelul 2 (o matrice); un singur segment cu incrementul Δ: $S^{(1)} = \Delta$, $S^{(2)} = \frac{1}{2}\Delta\Delta^\top$
  - Adîncimea N: se păstrează nivelurile $1, \ldots, N$ (semnătura trunchiată Sig^N); nivelul k are d^k termeni

- Termenii semnăturii trunchiate servesc drept variabile explicative într-o regresie

Identitatea lui Chen și identitatea shuffle

- **Identitatea lui Chen** Chen (1957): semnătura a două traiectorii unite este produsul tensorial al semnăturilor lor; la nivelul 2

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

  - $X * Y$: traiectoria X urmată de traiectoria Y (unite cap la cap); $S^{(\mathbf{1})}(X)\, S^{(\mathbf{1})}(Y)^{\top}$: matricea $d \times d$ a produselor, fiecare mișcare a lui X precede fiecare mișcare a lui Y

- Aplicată segment cu segment unei traiectorii cu incrementele $\Delta_{\mathbf{1}}, \ldots, \Delta_m$

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

  - Prima sumă: fiecare segment cu el însuși; a doua sumă: fiecare pereche ordonată de segmente, j înaintea lui k

- **Identitatea shuffle**: un produs de termeni de nivelul 1 este o sumă de termeni de nivelul 2

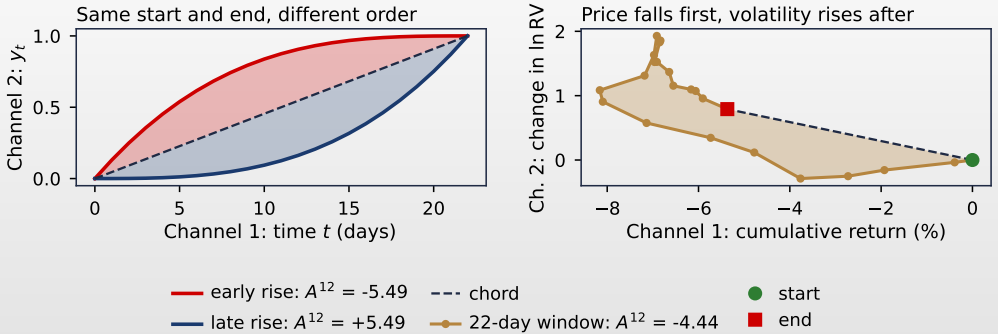
$$S^i S^j = S^{ij} + S^{ji}$$

  - Deci partea simetrică a nivelului 2 este $\frac{1}{2}\Delta X^i \Delta X^j$, cu $\Delta X^i = S^i$ schimbarea totală a canalului i: nu aduce informație nouă

Aria Lévy

- **Aria Lévy**: partea antisimetrică a nivelului 2
$$A^{ij} = \tfrac{1}{2}(S^{ij} - S^{ji})$$
  - Aria cu semn dintre traiectorie (proiectată pe canalele i, j) și coarda ei, dreapta de la început la sfîrșit; în sens trigonometric > 0, în sens orar < 0
  - Înregistrează **orientarea** traiectoriei; ordinea mișcărilor se citește din semn doar sub o ipoteză despre semnul asocierii
  - Dacă ambele canale cresc, $A^{ij} > 0$ cînd i crește primul; dacă i scade și apoi j crește, $A^{ij} < 0$, deși i se mișcă primul
- Un exemplu mic, canalele (timp, y), $y = (0, 1, 1)$: $\Delta_1 = (1, 1)$, $\Delta_2 = (1, 0)$
  - $S^{(1)} = (2, 1)$; $S^{(2)} = \frac{1}{2}\Delta_1\Delta_1^\top + \frac{1}{2}\Delta_2\Delta_2^\top + \Delta_1\Delta_2^\top = \left[\begin{smallmatrix} 2 & 0{,}5 \\ 1{,}5 & 0{,}5 \end{smallmatrix}\right]$; shuffle: $S^1S^2 = 2 = 0{,}5 + 1{,}5$
  - $A^{12} = (0{,}5 - 1{,}5)/2 = -0{,}5$: y a crescut devreme; pentru $y = (0, 0, 1)$ (creștere tîrzie) aceiași pași dau $A^{12} = +0{,}5$

Aria Lévy pe traiectorii simulate



- Stînga: două traiectorii de 22 de zile cu același început și același sfîrșit; regiunea hașurată este între traiectorie și coardă
  - creștere devreme: $A^{12} = -5{,}49$ (sens orar); creștere tîrzie: $A^{12} = +5{,}49$; nivelul 1 este identic, doar aria le deosebește
- Dreapta: canalele (randament cumulat, $\ln \mathrm{RV}$); prețul scade întîi, volatilitatea crește apoi: $A^{12} = -4{,}44$
  - tipul de efect de ordine pe care B1 îl testează ca predictor al volatilității viitoare

Sig-LK, variabilele explicative și ponderile nucleului

- Sig-LK (*Signature LASSO + kernel weights*): regresie LASSO (*Least Absolute Shrinkage and Selection Operator*, penalizată, cu selecția variabilelor) pe semnături, cu ponderi de nucleu, din Gu et al. (2024), aplicat la log RV pentru ziua (sau săptămîna, luna) următoare
- Modelul de referință: log-HAR, modelul HAR (Heterogeneous AutoRegressive) estimat în logaritmi
- Variabilele explicative ξ_τ ale zilei τ: cei trei termeni log-HAR și semnătura pînă la adîncimea 3 a traiectoriei de 22 de zile (timp, $\ln \mathrm{RV}$, r cumulat) care se încheie în ziua τ
  - τ: o zi trecută din fereastra de antrenare; t: originea prognozei, ultima zi observată
- Ponderile nucleului: zilele trecute a căror fereastră seamănă cu cea mai recentă primesc o pondere mai mare

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

  - δ_τ: pătratul distanței euclidiene dintre semnăturile trunchiate ale ferestrei care se termină în τ și ale ferestrei care se termină în t; s: parcurge toate zilele de antrenare, deci ponderile însumează 1
  - $\gamma \geq 0$: temperatura inversă (numită temperatură în articol); $c > 0$: o constantă fără unități; cu $c = 1$, o fereastră aflată la distanța mediană primește ponderea $e^{-\mathbf{1}} = 0{,}37$ față de una identică; $\gamma = 0$ dă ponderi egale

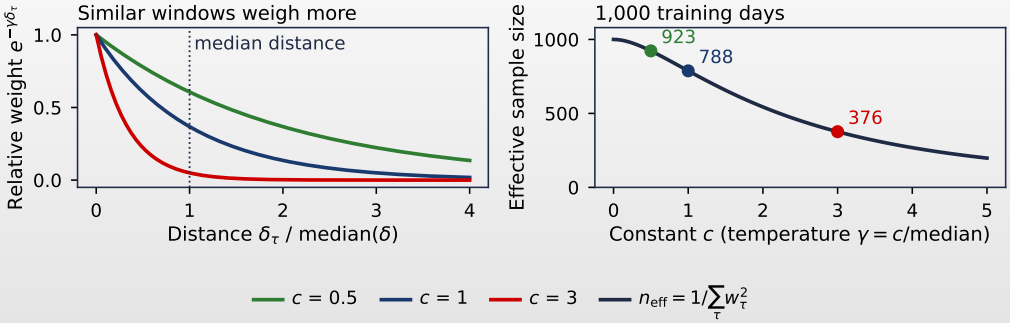
Sig-LK, mărimea efectivă a eșantionului și LASSO în doi pași

- Costul ponderilor inegale: mai puține observații efective (Kish (1965))

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

  - $n_{\text{eff}} = n$ pentru ponderi egale $1/n$; $n_{\text{eff}} = 1$ dacă o singură zi ia toată ponderea
- Doi pași de estimare pe eșantionul de antrenare ponderat
  - Pasul 1: un LASSO ponderat minimizează $\sum_\tau w_\tau (y_\tau - \xi_\tau^\top \theta)^2 + \lambda \sum_j |\theta_j|$; y_τ: ținta log RV, θ: coeficienții; penalizarea $\lambda \geq 0$ anulează exact unii coeficienți, adică selectează variabilele
  - λ se alege prin BIC (Bayesian Information Criterion, criteriul informațional bayesian: potrivirea plus o penalizare pentru fiecare coeficient păstrat)
  - Pasul 2: OLS (Ordinary Least Squares, cele mai mici pătrate) ponderat reestimează doar variabilele selectate, eliminînd contracția LASSO spre zero
- Specificația cu fereastră mobilă: ultimele 1 000 de perechi complete (variabile, țintă), reestimare la 5 zile; fiecare prognoză folosește doar datele pînă la momentul ei de origine (fără deplasarea de anticipare, *look-ahead bias*)

Ponderile nucleului și mărimea efectivă a eșantionului

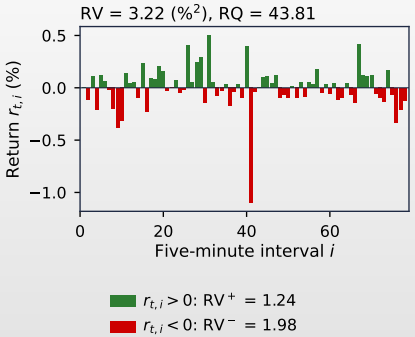


- Stînga: ponderea relativă $e^{-\gamma\delta_\tau}$ a unei ferestre trecute aflate la distanța δ_τ, cu $\gamma = c/\text{mediana}(\delta)$
  - la distanța mediană: $e^{-\mathbf{0{,}5}} = 0{,}61$, $e^{-\mathbf{1}} = 0{,}37$, $e^{-\mathbf{3}} = 0{,}05$ pentru $c = 0{,}5;\ 1;\ 3$
- Dreapta: 1 000 de distanțe simulate; $n_{\text{eff}} = 923,\ 788,\ 376$ pentru $c = 0{,}5;\ 1;\ 3$

Instrumentele din capitolele anterioare, exercițiu cu exercițiu

- B1, B2: HAR și log-HAR Corsi (2009), SHAR (*Semivariance HAR*, HAR cu semivarianțe) Patton & Sheppard (2015), HARQ (*HAR with Realised Quarticity*, HAR cu cvarticitatea realizată) Bollerslev, Patton & Quaedvlieg (2016), erori standard Newey–West (HAC): Capitolul 9
- A2, B3: pierderea QLIKE (*quasi-likelihood*, cvasi-verosimilitate) și motivul pentru care ierarhizează robust prognozele de varianță și cînd varianța realizată, folosită ca aproximare (*proxy*), are zgomot Patton (2011): Capitolul 9
- B3: testele Diebold–Mariano Diebold & Mariano (1995) și Giacomini–White (GW) condiționat Giacomini & White (2006): Capitolul 8; bootstrap staționar pentru un interval: Capitolul 5
- B4: corecțiile Holm (controlează FWER, probabilitatea unei respingeri false) și Benjamini–Hochberg (controlează FDR, proporția respingerilor false) Benjamini & Hochberg (1995): Capitolul 3

Varianța realizată, semivariațele și cvarticitatea



- Ziua t împărțită în M randamente intrazilnice $r_{t,i}$ ($M = 78$ de intervale de cinci minute)

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

  - RV_t estimează dispersia zilei t; $\sqrt{\mathrm{RV}_t}$: volatilitatea zilnică
  - $\mathrm{RV}_t^- = \mathrm{RV}_t - \mathrm{RV}_t^+$: volatilitatea „rea", din prețuri în scădere (SHAR)

- Cvarticitatea realizată: scala erorii de măsurare a lui RV_t (HARQ)

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

  - mare în zilele cu cîteva mișcări mari
  - fără salturi și fără zgomotul de microstructură (*microstructure noise*): dispersia erorii lui $\mathrm{RV}_t \approx 2\,\mathrm{RQ}_t / M$

- Graficul: o zi simulată cu un salt; $\mathrm{RV} = 3{,}22$, $\mathrm{RV}^- = 1{,}98 > \mathrm{RV}^+ = 1{,}24$
  - ilustrează descompunerea și efectul saltului asupra RQ

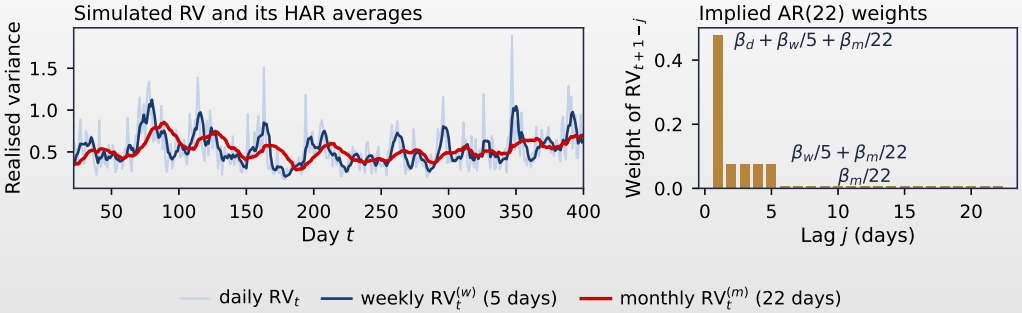
Reamintire, familia HAR (B1, B2)

- HAR Corsi (2009): RV din perioada următoare este o funcție liniară de mediile zilnică, săptămînală și lunară ale RV trecute

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

  - $\mathrm{RV}_t^{(w)}$, $\mathrm{RV}_t^{(m)}$: mediile RV pe ultimele 5 și 22 de zile; $\beta_d, \beta_w, \beta_m$: ponderile zilei de ieri, ale ultimei săptămîni, ale ultimei luni; $\varepsilon_{t+\mathbf{1}}$: eroarea; log-HAR: același model în logaritmi
- SHAR Patton & Sheppard (2015): termenul zilnic descompus în semivariațele RV_t^+ (din randamentele intrazilnice pozitive) și RV_t^- (din cele negative)
  - Coeficienții β^+, β^- în locul lui β_d; $\beta^- > \beta^+$: volatilitatea „rea" prezice mai multă volatilitate viitoare decît cea „bună"
- HARQ Bollerslev, Patton & Quaedvlieg (2016): ponderea lui RV_t scade în zilele în care RV este măsurată mai puțin precis
  - Adaugă la HAR termenul $\beta_Q \,\mathrm{RV}_t(\mathrm{RQ}_t^{\mathbf{1/2}} - \overline{\mathrm{RQ}^{\mathbf{1/2}}})$; RQ_t: cvarticitatea realizată; bara: media de selecție; se așteaptă $\beta_Q < 0$

Mediile HAR și ponderile lor pe laguri



- Stînga: o serie RV simulată cu mediile ei săptămînală și lunară; media lunară se mișcă lent
- Dreapta: HAR este un model autoregresiv de ordinul 22, AR(22) cu ponderi restricționate la trei trepte (A7); aici $\beta_d = 0{,}4$, $\beta_w = 0{,}35$, $\beta_m = 0{,}15$
  - lagul 1: 0,477, lagurile 2–5: cîte 0,077, lagurile 6–22: cîte 0,0068; 4 parametri în loc de 23

Erorile standard Newey–West (HAC) (B1, B2, B3)

- Regresia $y_t = x_t^\top \beta + u_t$ cu erori heteroscedastice și autocorelate; covarianța de termen lung a lui $x_t u_t$

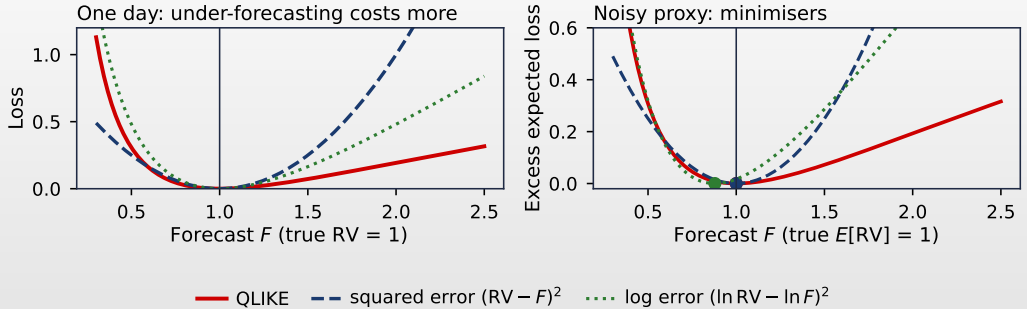
$$\hat{\Omega} = \hat{\Gamma}_{\mathbf{0}} + \sum_{l=\mathbf{1}}^{q} \left(1 - \frac{l}{q+1}\right)(\hat{\Gamma}_l + \hat{\Gamma}_l^\top), \qquad \hat{\Gamma}_l = \frac{1}{T} \sum_{t=l+\mathbf{1}}^{T} \hat{u}_t \hat{u}_{t-l}\, x_t x_{t-l}^\top$$

  - q: numărul de laguri (lățimea de bandă); $1 - l/(q+1)$: ponderile Bartlett, care păstrează $\hat{\Omega}$ pozitiv semidefinită
- Covarianța sandviș a estimărilor OLS și statisticile t

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

  - pentru diferența a doi coeficienți (B2): $\mathrm{SE}(\hat{\beta}^- - \hat{\beta}^+) = \sqrt{V_{--} + V_{++} - 2V_{+-}}$
  - țintele suprapuse pe h zile creează dependență în u_t pînă la lagul $h-1$: o structură de medie mobilă (MA, *Moving Average*) de ordinul $h-1$ dacă erorile de o zi nu sînt autocorelate
  - dependența poate continua dincolo de suprapunere, iar varianța naivă subestimează incertitudinea doar cînd suma autocovarianțelor relevante este pozitivă
- $R^{\mathbf{2}} = 1 - \sum_t \hat{u}_t^{\mathbf{2}} / \sum_t (y_t - \bar{y})^{\mathbf{2}}$: proporția din dispersia lui y explicată de regresori

QLIKE și un proxy cu zgomot al volatilității (A2, B3)



- $\mathrm{QLIKE}(\mathrm{RV}, F) = \mathrm{RV}/F - \ln(\mathrm{RV}/F) - 1$; stînga: cele trei pierderi într-o zi cu $\mathrm{RV} = 1$
- Dreapta: proxy-ul are zgomot, $\mathrm{RV} = \chi^2_8/8$ cu media 1; pierderea așteptată în exces, în funcție de prognoza constantă F
- **Ce credeți**: ce prognoză constantă minimizează eroarea logaritmică pentru un proxy cu zgomot?
  - **Răspuns**: $\exp(\mathbb{E}[\ln \mathrm{RV}]) = 0{,}88$, sub media 1

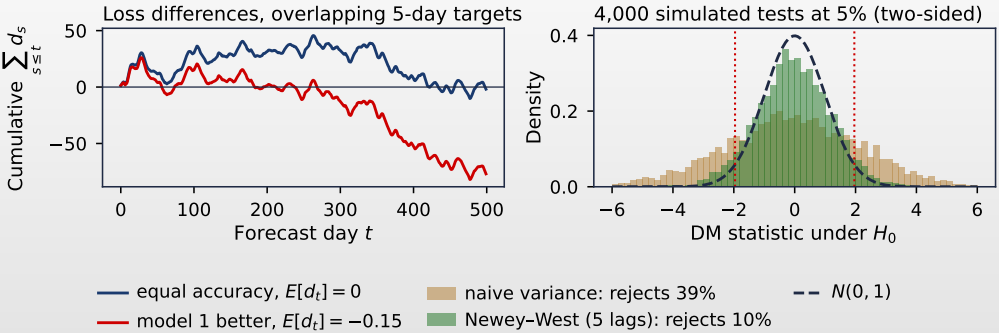
Reamintire, compararea a două prognoze (B3)

- Testul Diebold–Mariano Diebold & Mariano (1995): este diferența medie de pierdere zero?

$$d_t = L_t^{(1)} - L_t^{(2)}, \qquad \mathrm{DM} = \frac{\bar{d}}{\sqrt{\hat{\Omega}_{\mathrm{NW}}/T}} \approx N(0,1) \text{ sub } H_0 : \mathbb{E}[d_t] = 0$$

  - $L_t^{(1)}$, $L_t^{(2)}$: pierderile (de ex. QLIKE) modelelor 1 și 2 în ziua t; $\bar{d}$: media lui d_t pe cele T zile de prognoză; $\hat{\Omega}_{\mathrm{NW}}$: varianța de termen lung Newey–West (NW) a lui d_t
  - $\mathrm{DM} < -1{,}645$: modelul 1 este semnificativ mai precis la 5% (unilateral)
- Testul Giacomini–White condiționat Giacomini & White (2006): poate fi prezisă diferența de pierdere din informația trecută?
  - Pentru o țintă pe o zi, se regresează d_t pe instrumentele $(1, d_{t-1})$; statistica Wald pentru ambii coeficienți $= 0$ are aproximativ distribuția $\chi^2(2)$ sub H_0 ($2 =$ numărul de instrumente)
- Bootstrap staționar: se reeșantionează blocuri de zile consecutive cu lungime aleatoare (lungimea medie a blocului fixată de utilizator), pentru a păstra dependența serială
  - Cuantilele de 2,5% și 97,5% ale statisticii replicate dau un interval de încredere (CI, *Confidence Interval*) de 95%

Testul Diebold–Mariano cu ținte suprapuse



- Diferențe de pierdere d_t simulate pentru prognoze pe 5 zile: zilele vecine au în comun 4 din 5 zile, deci d_t este autocorelată
  - stînga: suma cumulată coboară sistematic doar cînd modelul 1 este într-adevăr mai bun
  - dreapta: sub H_0, varianța naivă respinge în 39% din cazuri în loc de 5%; Newey–West o aduce la 10%

Reamintire, multe teste simultan (B4)

- m teste (aici unul pe activ), cu valorile p ordonate crescător, $p_{(\mathbf{1})} \le \cdots \le p_{(m)}$
  - La 5% fiecare, se așteaptă aproximativ $0{,}05\,m$ respingeri din întîmplare, chiar dacă niciun model nu este mai bun (2,5 din 50)
- **Holm** controlează FWER (Family-Wise Error Rate): probabilitatea a cel puțin unei respingeri false este cel mult 5%

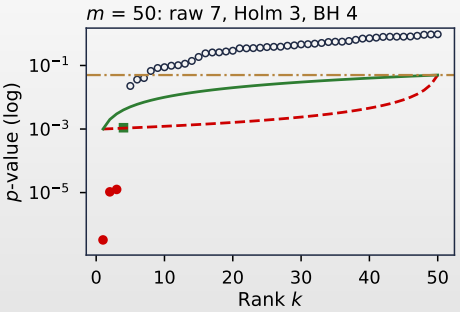
$$\text{respingem cît timp } p_{(k)} \le \frac{0{,}05}{m-k+1},\ k=1,2,\ldots;\ \text{ne oprim la primul eșec}$$

- **Benjamini–Hochberg (BH)** Benjamini & Hochberg (1995) controlează FDR (False Discovery Rate): proporția așteptată de respingeri false printre toate respingerile este cel mult 5%

$$k^* = \max\left\{k : p_{(k)} \le 0{,}05\,\frac{k}{m}\right\};\quad \text{respingem cele mai mici } k^* \text{ valori p}$$

  - Holm este valid sub orice dependență; BH cere valori p independente sau pozitiv dependente; BH respinge cel puțin la fel de multe ipoteze ca Holm

Holm și BH pe 50 de teste simulate



- 50 de teste unilaterale simulate, 5 cu un efect real
  - punctele: $p_{(k)}$ ordonate, colorate după decizia fiecărei proceduri
  - Holm (procedură descendentă, *step-down*) se oprește la primul rang care depășește pragul său
  - BH (procedură ascendentă, *step-up*) găsește ultimul rang sub pragul său și respinge toate rangurile pînă la el
- Respingerile: necorectat 7, Holm 3, BH 4
  - cu $m = 50$ primul prag Holm este $0{,}05/50 = 0{,}001$: rămîn doar valorile p foarte mici
  - dintre cele 7 respingeri necorectate, unele sînt false: din cele 45 de ipoteze nule adevărate se așteaptă aproximativ $0{,}05 \cdot 45 = 2{,}25$

Notații folosite în toate exercițiile

- RV_t: varianța realizată la cinci minute în ziua t; $r_t = \ln(\text{close}_t/\text{open}_t)$: randamentul logaritmic de la deschidere la închidere
- Mediile HAR: $\mathrm{RV}_t^{(w)} = \frac{1}{5}\sum_{j=0}^{4} \mathrm{RV}_{t-j}$, $\mathrm{RV}_t^{(m)} = \frac{1}{22}\sum_{j=0}^{21} \mathrm{RV}_{t-j}$; log-HAR regresează $\ln \mathrm{RV}_{t+1:t+h}$ pe $\ln \mathrm{RV}_t$, $\ln \mathrm{RV}_t^{(w)}$, $\ln \mathrm{RV}_t^{(m)}$, cu $\mathrm{RV}_{t+1:t+h}$ media RV în următoarele h zile
- Un exemplu: $\mathrm{RV} = 1$, $F = 2$: $\mathrm{QLIKE} = 0{,}5 - \ln 0{,}5 - 1 = 0{,}193$

Harta exercițiilor

Exercițiu	Întrebare	Tip	Model
A1	cum înregistrează semnătura ordinea mișcărilor?	Rezolvat	–
A2	ce prognoză minimizează QLIKE și ce erori costă mai mult?	Rezolvat	–
A3	cum schimbă temperatura inversă asemănarea și mărimea eșantionului?	Rezolvat	–
B1	prognozează aria Lévy RV-ul lunii următoare (JPM)?	Rezolvat	–
B2	se confirmă efectele SHAR și HARQ pentru JPM?	Rezolvat	–
B3	este Sig-LK superior modelului log-HAR pentru JPM, la o zi?	Rezolvat	–
B4	cîte cîștiguri rămîn după Holm și BH, pe 50 de active?	Rezolvat	–
C1	ce este greșit într-un răspuns de inteligență artificială (AI)?	Propus	A1, A2, B3
Exerciții suplimentare	A4–A5, A7–A10, B5–B8, C2 (opționale); A6 obligatoriu	Ambele (indicat la fiecare)	indicat la fiecare cerință

- **[Rezolvat]**: rezolvarea completă în slide-uri și în notebook, un model de urmat; **[Propus]**: îl rezolvați dumneavoastră, rezolvarea se discută la seminar
- Partea A: derivări pe hîrtie; Partea B: verificare și inferență pe datele VOLARE, fiecare cu o întrebare de interpretare; Partea C: întrebări deschise și critica unui răspuns AI

A1 [Rezolvat]: O semnătură de adîncime 2, pas cu pas — Cerință

- **Întrebare**: cum înregistrează o semnătură de adîncime 2 *cînd* s-a mișcat o serie, nu doar cît?
- **Date de intrare**
  - O serie ia valorile $y = (0, 2, 3, 3)$ în zilele $t = 0, 1, 2, 3$; fără fișier de date
  - Traiectoria augmentată cu timpul $X_t = (t, y_t)$ este liniară între zile consecutive (canalul 1 = timpul, canalul 2 = y)
- **Cerințe**
  (a) Scrieți cele trei incremente $\Delta_1, \Delta_2, \Delta_3 \in \mathbb{R}^2$ și calculați nivelul 1, $S^{(1)} = \sum_k \Delta_k$.
  (b) Folosiți identitatea lui Chen pentru a deduce formula nivelului 2 (pătrate plus perechi ordonate), apoi calculați această matrice 2×2.
  (c) Verificați rezultatul cu identitatea shuffle pentru $i = j = 1$ și pentru $i = 1, j = 2$.
  (d) Calculați aria Lévy A^{12} și desenați traiectoria cu coarda ei.
- **Raportați**: $S^{(1)}$, $S^{(2)}$ și A^{12}
- **Interpretare**: ce spune semnul lui A^{12} despre momentul în care seria a crescut?
- Notebook: secțiunea A1

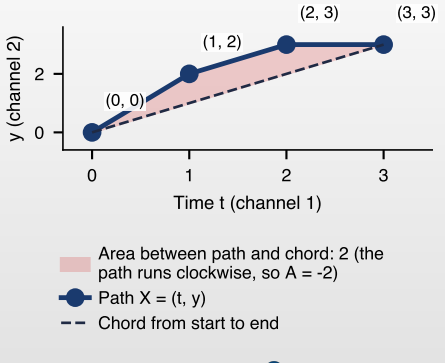
◂ Înapoi ▸ Lista exercițiilor

A1 [Rezolvat]: Rezolvare (1/2) — Nivelurile 1 și 2

- (a) Incremente: $\Delta_1 = (1,2)$, $\Delta_2 = (1,1)$, $\Delta_3 = (1,0)$; nivelul 1: $S^{(1)} = (3,3)$, schimbarea totală în timp și în y
- (b) Un segment are $\mathrm{Sig} = (1, \Delta, \frac{1}{2}\Delta\Delta^\top, \dots)$; Chen înmulțește segmentele, iar nivelul 2 al produsului adună fiecare termen pătratic o dată și fiecare pereche ordonată $j < k$ o dată
- Pătrate: $\frac{1}{2}\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}1{,}5 & 1{,}5\\ 1{,}5 & 2{,}5\end{bmatrix}$
- Perechi ordonate: $\Delta_1\Delta_2^\top + \Delta_1\Delta_3^\top + \Delta_2\Delta_3^\top = \begin{bmatrix}1 & 1\\ 2 & 2\end{bmatrix} + \begin{bmatrix}1 & 0\\ 2 & 0\end{bmatrix} + \begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix} = \begin{bmatrix}3 & 1\\ 5 & 2\end{bmatrix}$
- Suma: $S^{(2)} = \begin{bmatrix}4{,}5 & 2{,}5\\ 6{,}5 & 4{,}5\end{bmatrix}$

A1 [Rezolvat]: Rezolvare (2/2) — Verificarea shuffle și aria Lévy

- (c) Shuffle, $i = j = 1$: $S^{\mathbf{1}}S^{\mathbf{1}} = 9 = 2S^{\mathbf{11}} = 2 \cdot 4{,}5$; $i = 1, j = 2$: $S^{\mathbf{1}}S^{\mathbf{2}} = 9 = S^{\mathbf{12}} + S^{\mathbf{21}} = 2{,}5 + 6{,}5$: ambele sînt verificate
- (d) Aria Lévy: $A^{\mathbf{12}} = (2{,}5 - 6{,}5)/2 = -2$
- Verificare geometrică (graficul): traiectoria stă deasupra coardei $y = t$; poligonul hașurat are aria 2 și este parcurs în sens orar, de aici semnul minus
- Interpretare: cu canalele (timp, y), o arie negativă înseamnă că y a crescut *devreme* în fereastră; aceeași schimbare netă, distribuită uniform, ar da $A^{\mathbf{12}} = 0$



MFM_ch20_seminar

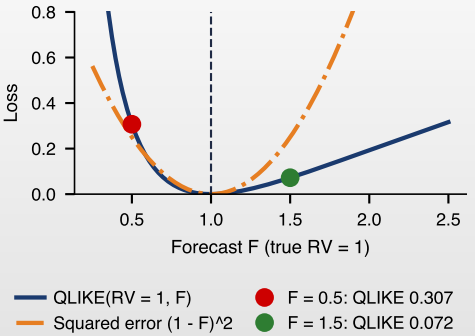
A2 [Rezolvat]: QLIKE: minimizator și asimetrie — Cerință

- **Întrebări**
  - ce prognoză minimizează QLIKE așteptată?
  - ce erori penalizează mai mult QLIKE?
- **Date de intrare**: $\mathrm{QLIKE}(\mathrm{RV}, F) = \mathrm{RV}/F - \ln(\mathrm{RV}/F) - 1$, cu proxy-ul realizat $\mathrm{RV} > 0$ și prognoza $F > 0$; $\mathbb{E}_t$ = media condiționată de informația de la originea prognozei
- **Cerințe**
  (a) Derivați $\mathbb{E}_t[\mathrm{QLIKE}(\mathrm{RV}, F)]$ în raport cu F și rezolvați condiția de ordinul întîi.
  (b) Verificați condiția de ordinul al doilea.
  (c) Pentru $\mathrm{RV} = 1$, calculați QLIKE și eroarea pătratică pentru $F = 0{,}5$ și pentru $F = 1{,}5$ și reprezentați ambele pierderi în funcție de F.
- **Raportați**: minimizatorul, cele patru pierderi și graficul
- **Interpretare**
  - ce eroare penalizează mai mult QLIKE?
  - de ce se potrivește aceasta managementului riscului?
- Notebook: secțiunea A2

◂ Înapoi ▸ Lista exercițiilor

A2 [Rezolvat]: Rezolvare

- (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) $\partial^2/\partial F^2 = 2\mathbb{E}_t[\text{RV}]/F^3 - 1/F^2 = 1/F^{*2} > 0$ în F^*: un minim; F^* depinde doar de $\mathbb{E}_t[\text{RV}]$, deci cu un proxy condiționat nedeplasat ordinea pierderilor *așteptate* este robustă Patton (2011)
- (c) $\text{QLIKE}(1;0{,}5) = 2 - \ln 2 - 1 = 0{,}307$; $\text{QLIKE}(1;1{,}5) = 0{,}667 + \ln 1{,}5 - 1 = 0{,}072$; erori pătratice 0,25 și 0,25
- Interpretare: QLIKE penalizează subestimarea variației de aproximativ patru ori mai mult decît aceeași supraestimare; riscul subestimat este eroarea costisitoare (capital prea mic, valoarea la risc, VaR (*Value at Risk*), depășită)



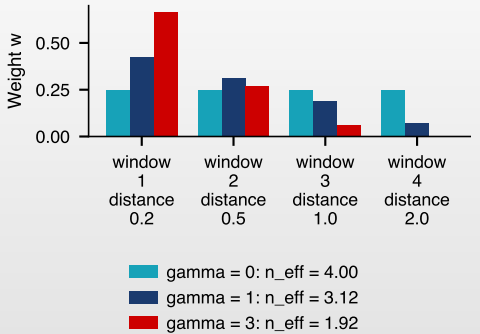
A3 [Rezolvat]: Ponderi de nucleu și mărimea efectivă a eșantionului — Cerință

- **Întrebare**: cum schimbă temperatura inversă a nucleului γ echilibrul dintre asemănare și numărul de observații?
- **Date de intrare**
  - Patru ferestre trecute cu distanțele de semnătură $\delta = (0{,}2; 0{,}5; 1{,}0; 2{,}0)$ față de fereastra curentă
  - Ponderile nucleului $w_\tau = e^{-\gamma\delta_\tau} / \sum_s e^{-\gamma\delta_s}$ Gu et al. (2024) și mărimea efectivă a eșantionului $n_{\text{eff}} = 1/\sum_\tau w_\tau^2$ Kish (1965)
- **Cerințe**
  (a) Calculați cele patru ponderi pentru $\gamma = 1$.
  (b) Calculați cele patru ponderi pentru $\gamma = 3$.
  (c) Calculați n_{eff} în ambele cazuri și comparați cu ponderile egale ($\gamma = 0$).
- **Raportați**: doi vectori de ponderi, două valori n_{eff}, un grafic cu bare
- **Interpretare**
  - ce se cîștigă cînd γ crește?
  - ce se pierde?
- Notebook: secțiunea A3

◂ Înapoi ▸ Lista exercițiilor

A3 [Rezolvat]: Rezolvare

- (a) $e^{-\delta} = (0{,}819; 0{,}607; 0{,}368; 0{,}135)$, suma 1,928: $w = (0{,}425; 0{,}315; 0{,}191; 0{,}070)$
- (b) $e^{-\mathbf{3}\delta} = (0{,}549; 0{,}223; 0{,}050; 0{,}002)$, suma 0,824: $w = (0{,}666; 0{,}271; 0{,}060; 0{,}003)$
- (c) $n_{\text{eff}} = 4$ pentru $\gamma = 0$, 3,12 pentru $\gamma = 1$ și 1,92 pentru $\gamma = 3$
- Interpretare: o temperatură inversă mai mare se concentrează pe fereastra cea mai asemănătoare (mai puțină deplasare din perioade diferite)
- Costul: informația se înjumătățește (mai multă varianță)
- Pe datele reale, ponderile Sig-LK au în medie $n_{\text{eff}} \approx 816$ din 1 000: pierderea de informație este mică



B1 [Rezolvat]: Aria Lévy ca predictor al volatilității (JPM) — Cerință

- **Întrebare**: are ordinea mișcărilor prețului și ale volatilității din ultima lună putere predictivă pentru volatilitatea JPM din luna următoare, dincolo de log-HAR?
- **Date de intrare**
  - VOLARE Cipollini et al. (2026), JPM (JPMorgan Chase), 2 ian. 2015 – 30 ian. 2026, doar zilele de tranzacționare; RV la cinci minute, prețurile zilnice de deschidere și de închidere
  - Regresorul: A_t = aria Lévy a traiectoriei de 22 de zile (randament cumulat, $\ln \mathrm{RV}$) care se termină în ziua t, împărțită la abaterea standard din întregul eșantion; canalele în ordinea scrisă, fără timp
  - Ținta: $y_t = \ln \mathrm{RV}_{t+\mathbf{1}:t+\mathbf{22}} - \ln \mathrm{RV}_t^{(m)}$, log RV din luna următoare minus cea din luna trecută (ținte consecutive au în comun 21 din 22 de zile)
- **Cerințe**
  (a) Pentru fiecare zi, construiți traiectoria ultimelor 22 de zile, calculați semnătura ei de adîncime 2 și aria Lévy A_t, apoi standardizați A_t.
  (b) Regresați y_t prin OLS pe o constantă, cei trei termeni log-HAR și A_t, cu erori standard Newey–West și 22 de laguri.
  (c) Testați H_0: coeficientul lui $A_t = 0$ la 5%, bilateral, și comparați R^2 cu și fără A_t.
  (d) Reprezentați un grafic de dispersie pe grupe (*binned scatter*) al lui y_t față de A_t, ambele reziduale față de termenii log-HAR.
- **Raportați**: coeficientul, SE (*Standard Error*, eroare standard) HAC, t, decizia, ambele R^2, graficul
- **Interpretare**
  - de ce 22 de laguri?
  - are cîștigul importanță economică?
- Notebook: secțiunea B1

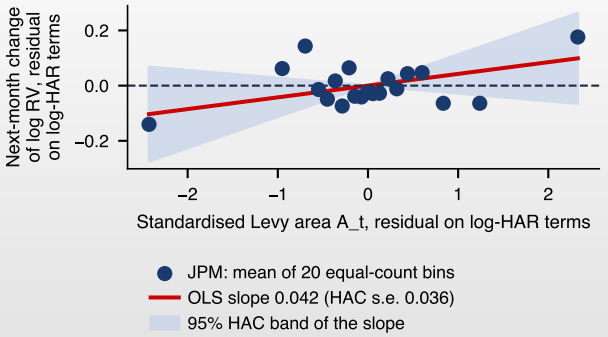
înapoi Lista exercițiilor

B1 [Rezolvat]: Rezolvare (1/2) — Pași

- (a) Semnătura fiecărei ferestre de 22 de puncte cu canalele (randament cumulat, $\ln \mathrm{RV}$); A_t este negativă în 46% din ferestre (compatibil cu secvențe de tip efect de levier, *leverage effect*, fără a le identifica)
- (b) 2 743 zile utilizabile; coeficientul lui $A_t = +0{,}042$, SE HAC 0,036
- (c) $t = +0{,}042/0{,}036 = +1{,}18$ (din estimările nerotunjite); $|t| < 1{,}96$: nu respingem H_0 la 5%; R^2: 28,0% fără A_t, 28,6% cu A_t
- Motivul celor 22 de laguri: țintele consecutive au în comun 21 din cele 22 de zile, deci erorile sînt dependente serial chiar sub H_0; RV persistentă poate prelungi dependența, deci este recomandată și o verificare cu 44 de laguri

B1 [Rezolvat]: Rezolvare (2/2) — Grafic și interpretare

- (d) Graficul: punctele grupelor sînt dispersate în jurul lui zero; banda HAC a pantei conține dreapta orizontală
- Mărimea: o jumătate de punct de R^2 în eșantion, față de cele 28% explicate de termenii log-HAR; utilitatea economică ar cere o evaluare în afara eșantionului
- Interpretare: pentru JPM, ordinea mișcărilor dintre preț și volatilitate nu aduce informație sigură despre volatilitatea lunii următoare



MFM_ch20_seminar

B2 [Rezolvat]: HAR, SHAR și HARQ pentru JPM — Cerință

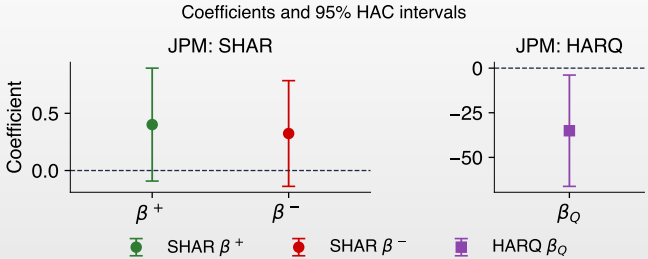
- **Întrebare**: schimbă volatilitatea „rea" (SHAR) și eroarea de măsurare a RV (HARQ) prognozele pentru ziua următoare la JPM?
- **Date de intrare**
  - VOLARE Cipollini et al. (2026), JPM (JPMorgan Chase), 2 ian. 2015 – 30 ian. 2026, doar zilele de tranzacționare; RV la cinci minute, semivarianțele RV_t^+, RV_t^- și cvarticitatea realizată RQ; ținta: RV din ziua următoare, în nivel, tot eșantionul
  - SHAR Patton & Sheppard (2015): $\mathrm{RV}_{t+\mathbf{1}} = \beta_{\mathbf{0}} + \beta^+ \mathrm{RV}_t^+ + \beta^- \mathrm{RV}_t^- + \beta_w \mathrm{RV}_t^{(w)} + \beta_m \mathrm{RV}_t^{(m)} + \varepsilon_{t+\mathbf{1}}$, cu β^+, β^- ponderile semivarianței pozitive și negative
  - HARQ Bollerslev, Patton & Quaedvlieg (2016): HAR plus termenul $\beta_Q \, \mathrm{RV}_t (\mathrm{RQ}_t^{\mathbf{1/2}} - \overline{\mathrm{RQ}^{\mathbf{1/2}}})$
  - V: matricea de covarianță Newey–West (5 laguri) a estimărilor; V_{++}, V_{--}: variațele lui $\hat{\beta}^+$, $\hat{\beta}^-$; V_{+-}: covarianța lor
- **Cerințe**
  (a) Estimați HAR Corsi (2009) prin OLS și raportați cele trei pante cu erori standard HAC.
  (b) Estimați SHAR și testați $H_0 : \beta^- = \beta^+$ față de $\beta^- > \beta^+$ cu $t = (\hat{\beta}^- - \hat{\beta}^+)/\mathrm{SE}(\hat{\beta}^- - \hat{\beta}^+)$, cu $\mathrm{SE}(\hat{\beta}^- - \hat{\beta}^+) = \sqrt{V_{--} + V_{++} - 2V_{+-}}$.
  (c) Estimați HARQ și testați $H_0 : \beta_Q = 0$ față de $\beta_Q < 0$; reprezentați β^+, β^- și β_Q cu intervale HAC de 95%.
- **Raportați**: estimările, t, p, deciziile, graficul coeficienților
- **Interpretare**: se confirmă cele două efecte din literatură pentru această acțiune?
- Notebook: secțiunea B2

◂ Înapoi ▸ Lista exercițiilor

B2 [Rezolvat]: Rezolvare (1/2) — Estimări și teste

- (a) HAR (2 764 zile): $\beta_d = 0{,}367$ (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); diferența $-0{,}078$ (din estimările nerotunjite), $t = -0{,}17$, unilateral $p = 0{,}566$: nu respingem H_0
- (c) HARQ: $\hat{\beta}_Q = -35{,}062$ (15,904), $t = -2{,}20 < -1{,}645$, p unilateral $= 0{,}014$: respingem H_0 la 5% unilateral

B2 [Rezolvat]: Rezolvare (2/2) — Grafic și interpretare



- Graficul: cele două intervale SHAR sînt largi și se suprapun (semivarianțele sînt puternic corelate pentru o singură acțiune); intervalul HARQ este sub zero
- Interpretare: în zilele zgomotoase HARQ reduce ponderea RV din acea zi, iar efectul se confirmă pentru JPM; asimetria dintre volatilitatea „bună" și cea „rea" este un rezultat pe panel Patton & Sheppard (2015), care pe o singură acțiune, JPM, nu se poate detecta

MFM_ch20_seminar

B3 [Rezolvat]: Sig-LK față de log-HAR pentru JPM, la o zi — Cerință

- **Întrebare**: prognozează metoda Sig-LK RV din ziua următoare pentru JPM mai bine decît log-HAR?
- **Date de intrare**
  - VOLARE Cipollini et al. (2026), JPM (JPMorgan Chase), 2 ian. 2015 – 30 ian. 2026, doar zilele de tranzacționare; RV la cinci minute, prețurile zilnice de deschidere și de închidere
  - Specificația Sig-LK: fereastră mobilă de 1 000 de zile, reestimare la 5 zile, traiectoria (timp, $\ln \mathrm{RV}$, r cumulat) pe 22 de zile, adîncimea 3, LASSO în doi pași cu BIC, scala temperaturii inverse $c = 0{,}5$
- **Cerințe**
  (a) Produceți prognoze la o zi Sig-LK și log-HAR de la 27 ian. 2020 la 29 ian. 2026 (1 511 zile) și calculați QLIKE medie pentru fiecare.
  (b) Calculați raportul din pierderile medii nerotunjite și CI de 95% prin bootstrap staționar (lungimea medie a blocului 10, 1 000 de replicări).
  (c) Rulați testul DM (variantă HAC, lățimea de bandă $\max(h, \lceil T^{1/3} \rceil)$) și testul GW condiționat cu instrumentele $(1, d_{t-1})$ Giacomini & White (2006).
  (d) Reprezentați diferența cumulată de pierdere $\sum_{s \le t} d_s$.
- **Raportați**: pierderile, raportul și CI, t, p, deciziile, graficul
- **Interpretare**: există vreo dovadă că semnătura aduce valoare pentru JPM?
- Notebook: secțiunea B3

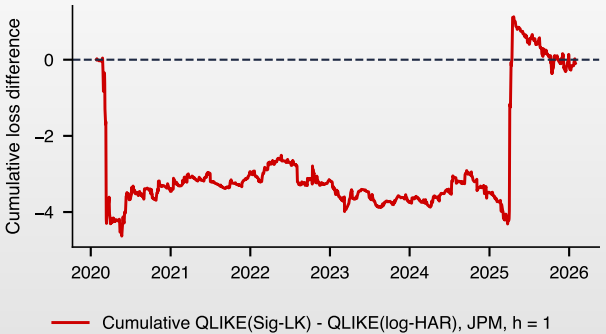
◂ Înapoi ▸ Lista exercițiilor

B3 [Rezolvat]: Rezolvare (1/2) — Pierderi și teste

- (a) QLIKE medie: Sig-LK 0,1726, log-HAR 0,1727; în medie 4,2 regresori selectați, mărimea efectivă a eșantionului 851 din 1 000
- (b) Raportul 1,000; CI 95% bootstrap [0,956; 1,047] conține 1
- (c) DM: $d_t = \mathrm{QLIKE}_t^{\text{Sig-LK}} - \mathrm{QLIKE}_t^{\text{log-HAR}}$, $t = \bar{d}/\sqrt{\hat{\Omega}_{\mathrm{NW}}/T} = -0{,}02$, unilateral $p = 0{,}493$: nu respingem acuratețea egală
- GW: statistica Wald 2,85 $\sim \chi^2(2)$, $p = 0{,}241$: instrumentele $(1, d_{t-1})$ nu dau dovezi semnificative că diferența de pierdere este previzibilă

B3 [Rezolvat]: Rezolvare (2/2) — Diferența cumulată de pierdere

- (d) Segmente în sus: log-HAR mai bun; segmente în jos: Sig-LK mai bun; cele mai mari variații apar în martie 2020 și aprilie 2025
- Cîteva zile de stres determină semnul sumei: raportați traiectoria cumulată, nu doar media
- Interpretare: nu detectăm un avantaj robust în specificația evaluată; pierderile medii sînt apropiate, ceea ce nu înseamnă prognoze identice
- CI pentru raportul pierderilor medii exclude diferențe mai mari de aproximativ 5%



B4 [Rezolvat]: Holm și BH, pas cu pas — Cerință

- **Întrebare**: după corecția pentru 50 de teste, pe cîte active este Sig-LK superior modelului log-HAR la o zi?
- **Date de intrare**
  - Valorile p DM pe active din Capitolul 20 (incluse în notebook): valori p DM unilaterale pentru „Sig-LK mai bun decît log-HAR" la $h = 1$ (QLIKE), una pe activ; familia = cele 50 de active
  - Cele cinci perechi FX, valorile p ordonate: EURUSD (euro/dolar) 0,071, USDJPY (dolar/yen) 0,213, USDCAD (dolar american/dolar canadian) 0,336, AUDUSD (dolar australian/dolar american) 0,718, GBPUSD (liră/dolar) 0,817
  - m: numărul de teste; Holm (FWER 5%) și BH (FDR 5%) Benjamini & Hochberg (1995)
- **Cerințe**
  (a) Aplicați Holm celor cinci valori p FX ($m = 5$).
  (b) Aplicați BH acelorași cinci.
  (c) Repetați ambele proceduri pentru toate cele 50 de active, numărați respingerile și reprezentați valorile p ordonate față de cele trei praguri.
- **Raportați**: respingerile necorectate, Holm, BH; graficul
- **Interpretare**
  - ce putem afirma?
  - despre cîte active?
- Notebook: secțiunea B4

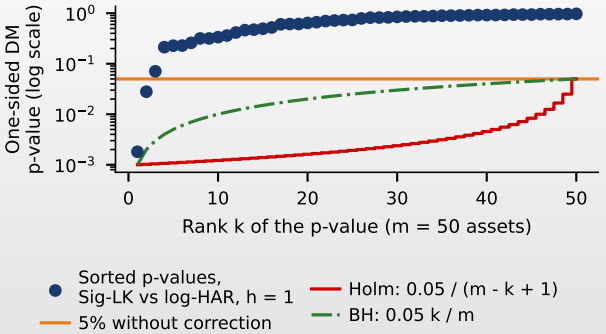
◂ Înapoi ▸ Lista exercițiilor

B4 [Rezolvat]: Rezolvare (1/2) — Pași

- (a) Holm, $m = 5$: $p_{(\mathbf{1})} = 0{,}071$ față de $0{,}05/5 = 0{,}01$: nicio respingere, ne oprim; valorile p ajustate Holm: 0,354, 0,852, apoi 1 (din valorile p nerotunjite)
- (b) BH: pragurile $0{,}01; 0{,}02; 0{,}03; 0{,}04; 0{,}05$; fiecare $p_{(k)}$ este peste pragul său: nicio respingere; cea mai mică valoare p ajustată BH este 0,354 (din valorile p nerotunjite)
- (c) 50 de active: 2 respingeri necorectate; Holm: cel mai mic, ES cu $p = 0{,}002$, este peste $0{,}05/50 = 0{,}001$: ne oprim, 0 respingeri; BH: $p_{(\mathbf{1})} > 0{,}001$ și $p_{(\mathbf{2})} = 0{,}028 > 0{,}002$, niciun k nu se califică: 0

B4 [Rezolvat]: Rezolvare (2/2) — Grafic și interpretare

- Graficul: doar punctele cele mai de jos trec sub linia de 5%; niciunul nu atinge scara Holm sau dreapta BH
- Holm este valid sub orice dependență; BH cere valori p independente sau pozitiv dependente în sensul PRDS (*Positive Regression Dependence on a Subset*, dependență pozitivă de regresie pe o submulțime) Benjamini & Yekutieli (2001)
- Interpretare: nicio îmbunătățire la nivel de activ nu rămîne semnificativă după corecție; ES este cel mult o pistă pentru un studiu nou, preînregistrat



MFM_ch20_seminar

C1: datele și convenția funcției

- Folosiți seria RV pentru JPM din pregătire. Ultimele 30% dintre originile de prognoză valide (zilele cu o fereastră completă de 22 de zile) formează eșantionul de evaluare pentru modelul candidat (LASSO-ul pe semnături din răspunsul AI) și log-HAR; partea anterioară este eșantionul de antrenare.
- `sig_path(X, depth)` primește rîndurile unui tablou ca puncte succesive ale unei traiectorii liniare pe segmente. Returnează nivelurile 1 pînă la depth, concatenate, fără nivelul 0 și fără să adauge vreun canal.
- Evaluați atît programul, cît și concluzia pretinsă: o comparație din text trebuie susținută de prognoze și pierderi calculate pe aceleași rînduri de evaluare.

Înapoi Lista exercițiilor

C1 [Propus]: critica unui răspuns AI — răspunsul

- **Cererea** trimisă unui asistent AI: „rv conține RV zilnică la 5 minute pentru JPM din datele de pregătire. Construiește LASSO-ul pe semnături al lui Gu et al. pentru log RV din ziua următoare și evaluează-l cu QLIKE pe ultimele 30% din eșantion.”
- **Răspunsul**: codul de mai jos și concluzia

```
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())
```

- **Concluzia AI**: „LASSO-ul pe semnături are o valoare QLIKE mai mică decît HAR pe setul de test, deci semnăturile surprind forma volatilității și îmbunătățesc prognoza.”

C1 [Propus]: Critica unui răspuns AI — cerință

Rezolvarea se discută la seminar

- **Întrebare**: ce erori din răspuns ar schimba concluzia?
- **Model**: A1, A2, B3
- **Cerințe**
  (a) Găsiți cel puțin trei erori; pentru fiecare, numiți rîndul și explicați într-o frază de ce este greșit.
  (b) Pentru fiecare eroare, descrieți o verificare care o dezvăluie fără să știți răspunsul (un test unitar, un test privind momentul informației, un caz cunoscut).
  (c) Evaluați retransformarea exponențială și dacă programul este metoda ponderată în doi pași a lui Gu et al.
  (d) Scrieți codul corectat și comparați-l cu log-HAR pe aceeași împărțire, cu un test DM.
  (e) Documentați fiecare eroare ca în `AI_ERRORS.md`: cererea, răspunsul, eroarea, cum ați găsit-o, corecția.
- **Raportați**: tabelul erorilor (rînd, motiv, verificare), codul corectat, înregistrările din `AI_ERRORS.md`; pe aceleași observații de evaluare, ambele medii QLIKE, statistica DM, valoarea p unilaterală și direcția alternativei
- **Interpretare**: ar fi justificată concluzia chiar dacă programul ar fi corect?
- Codul AI rulează de obicei fără mesaje de eroare: greșelile sînt în metodă, nu în sintaxă

C1: Alegerea metodei

- ⊡ Alegeți estimarea (fixă sau pe fereastră mobilă) și ponderile (egale sau de nucleu).
- ⊡ Selectați penalizarea doar cu informația disponibilă la fiecare prognoză.
- ⊡ Precizați canalele traiectoriei și justificați fiecare alegere într-o frază.

Exercițiile suplimentare: grupe și cunoștințe necesare

Grupa	Exerciții	Cunoștințe necesare
1. Algebra semnăturilor	A4, A5 [Rezolvat], A6 (obligatoriu)	A1
2. HAR ca AR(22)	A7 [Rezolvat], A8	B2, teste Wald
3. Pierderi și ponderi	A9, A10	A2, A3
4. Alte active	B5 (ES), B6 (CL), B7 (ES, $h = 5$)	B1, B2, B3
5. Testare multiplă la o lună	B8	B4
6. Idee de proiect	C2, un nucleu sensibil la nivel	B3, Capitolul 20 (ponderile nucleului)

- Fiecare exercițiu are secțiunea lui în notebook, cu cod de pornire (o celulă de cod goală pentru cele propuse); în cadrul unei grupe, lucrați în ordinea dată
- Exercițiile propuse își indică modelul: exercițiul rezolvat care folosește aceeași tehnică

A4 [Propus]: Semnătura și identitatea shuffle — Cerință

Rezolvarea se discută la seminar

- **Întrebări**
  - care este semnătura unei traiectorii cu trei segmente în plan?
  - ce sugerează aria cu semn despre ordinea mișcărilor?
- **Model**: A1
- **Date de intrare**: O traiectorie în $\mathbb{R}^2$ prin $(0,0) \to (2,1) \to (3,-1) \to (4,2)$, liniară între puncte
- **Cerințe**
  (a) Scrieți incrementele și calculați nivelul 1.
  (b) Calculați nivelul 2 cu formula din A1 (pătrate plus perechi ordonate).
  (c) Verificați $S^{\mathbf{1}}S^{\mathbf{2}} = S^{\mathbf{12}} + S^{\mathbf{21}}$ și calculați aria Lévy $A^{\mathbf{12}}$.
- **Raportați**: $S^{(\mathbf{1})}$, $S^{(\mathbf{2})}$, $A^{\mathbf{12}}$
- **Interpretare**
  - ce spune semnul ariei despre tendința canalului 1 de a se mișca înaintea canalului 2?
  - de ce nu implică o precedență strictă?
- Notebook: secțiunea A4

Înapoi Lista exercițiilor

A5 [Rezolvat]: Identitatea lui Chen și inversarea traiectoriei — Cerință

- **Întrebări**
  - care este semnătura unei traiectorii parcurse dus-întors?
  - de ce contează acest lucru pentru augmentarea cu timpul?
- **Date de intrare**
  - Traiectoria $X : (0,0) \to (1,0) \to (1,2)$, cu nivelurile 1 și 2

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

  - $\overleftarrow{X}$ este aceeași traiectorie parcursă invers: $(1,2) \to (1,0) \to (0,0)$
- **Cerințe**
  - (a) Calculați nivelurile 1 și 2 ale $\mathrm{Sig}(\overleftarrow{X})$ cu formula din A1.
  - (b) Aplicați identitatea lui Chen la nivelul 2 semnăturilor $\mathrm{Sig}(X)$ și $\mathrm{Sig}(\overleftarrow{X})$, pentru a calcula nivelul 2 al lui $X * \overleftarrow{X}$ (X urmată de $\overleftarrow{X}$).
  - (c) Explicați ce implică rezultatul pentru unicitatea semnăturii.
- **Raportați**: cele două semnături și nivelul 2 al concatenării
- **Interpretare**: de ce adăugăm în practică un canal de timp?
- Notebook: secțiunea A5

Înapoi Lista exercițiilor

A5 [Rezolvat]: Rezolvare

- (a) Incrementele inversate $(0,-2)$, $(-1,0)$: nivelul 1 este $(-1,-2) = -S^{(1)}$; nivelul 2 adună pătratele și perechea ordonată

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

- (b) Chen, cu termenul mixt $S^{(1)}(X)\, S^{(1)}(\overleftarrow{X})^\top$ la mijloc:

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

- Nivelul 1 este $S^{(1)} - S^{(1)} = 0$
- (c) Dusul și întorsul sînt „de tip arbore": semnătura lor este cea a traiectoriei constante; semnătura identifică o traiectorie doar pînă la astfel de segmente Hambly & Lyons (2010)
- Interpretare: un canal de timp strict crescător nu se poate întoarce, deci segmentele de tip arbore sînt imposibile și semnătura determină traiectoria pînă la punctul de start: augmentăm cu timpul și furnizăm nivelul separat

A6 [Propus]: Aria Lévy a unei bucle — Cerință

Rezolvarea se discută la seminar

- **Întrebare**: ce informație despre o buclă închisă conține nivelul 2 al semnăturii, dar nu și nivelul 1?
- **Model**: A5
- **Date de intrare**: Pătratul unitate $(0,0) \to (1,0) \to (1,1) \to (0,1) \to (0,0)$, patru segmente liniare
- **Cerințe**
  (a) Calculați nivelul 1 și aria Lévy $A^{\mathbf{12}}$ pentru pătratul parcurs în sens trigonometric.
  (b) Repetați pentru pătratul parcurs în sens orar și pentru pătratul translatat cu $(5,5)$.
  (c) Explicați de ce nivelul 1 nu poate distinge bucla de o traiectorie constantă, iar nivelul 2 poate.
- **Raportați**: trei valori ale $S^{(\mathbf{1})}$ și $A^{\mathbf{12}}$
- **Interpretare**: ce invarianță ilustrează translația?
- Notebook: secțiunea A6

Înapoi Lista exercițiilor

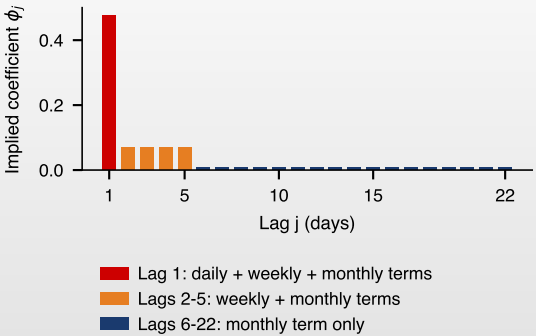
A7 [Rezolvat]: HAR ca AR(22) restricționat — Cerință

- **Întrebare**: ce coeficienți pentru cele 22 de laguri implică un model HAR pentru $\ln \mathrm{RV}$ la JPM?
- **Date de intrare**
  - VOLARE Cipollini et al. (2026), JPM (JPMorgan Chase), 2 ian. 2015 – 30 ian. 2026, doar zilele de tranzacționare
  - HAR cu *medii ale logaritmilor*: $\ell_t = \ln \mathrm{RV}_t$, $\ell_t^{(w)} = \frac{1}{5}\sum_{j=0}^{4} \ell_{t-j}$, $\ell_t^{(m)} = \frac{1}{22}\sum_{j=0}^{21} \ell_{t-j}$
  - Log-HAR din B1 și B3 folosește logaritmi ai mediilor, $\ln \mathrm{RV}_t^{(w)} \neq \ell_t^{(w)}$ (Jensen), și nu are o formă AR exactă
  - Tot eșantionul (2 764 zile), OLS, SE Newey–West (5 laguri): $\hat{\beta}_d = 0{,}408$ (0,027), $\hat{\beta}_w = 0{,}299$ (0,046), $\hat{\beta}_m = 0{,}196$ (0,040)
- **Cerințe**
  (a) Înlocuiți $\ell_t^{(w)}$ și $\ell_t^{(m)}$ în ecuația HAR și grupați coeficientul ϕ_j al fiecărui lag $\ell_{t-j+\mathbf{1}}$, $j = 1, \ldots, 22$.
  (b) Calculați $\phi_{\mathbf{1}}$, $\phi_{\mathbf{2}} = \cdots = \phi_{\mathbf{5}}$ și $\phi_{\mathbf{6}} = \cdots = \phi_{\mathbf{22}}$ din estimările JPM și reprezentați-i grafic.
  (c) Calculați suma $\sum_{j=\mathbf{1}}^{\mathbf{22}} \phi_j$.
- **Raportați**: cele trei valori distincte, suma, graficul
- **Interpretare**
  - ce măsoară suma?
  - este procesul staționar?
- Notebook: secțiunea A7

◂ Înapoi ▸ Lista exercițiilor

A7 [Rezolvat]: Rezolvare

- (a) Din coeficienții estimați nerotunjiți:
  $\beta_w \ell_t^{(w)} = \sum_{j=1}^{5} (\beta_w/5)\ell_{t-j+1}$ și
  $\beta_m \ell_t^{(m)} = \sum_{j=1}^{22} (\beta_m/22)\ell_{t-j+1}$: lagul 1 adună toți cei trei termeni, lagurile 2–5 pe cel săptămânal și pe cel lunar, lagurile 6–22 doar pe cel lunar
- (b) $\phi_1 = 0{,}408 + 0{,}299/5 + 0{,}196/22 = 0{,}476$; $\phi_{2..5} = 0{,}0688$; $\phi_{6..22} = 0{,}0089$
- (c) $\sum_j \phi_j = \beta_d + \beta_w + \beta_m = 0{,}903$
- Interpretare: suma este persistența log RV; toți $\phi_j \geq 0$ și suma este sub 1, deci procesul este staționar, cu o descreștere lentă, în trepte, care aproximează memoria lungă; HAR rămîne un proces cu memorie scurtă



MFM_ch20_seminar

A8 [Propus]: Testarea restricțiilor HAR — Cerință

Rezolvarea se discută la seminar

- **Întrebare**: cum testăm restricțiile pe care HAR le impune unui AR(22) nerestricționat?
- **Model**: A7
- **Date de intrare**
  - AR(22) nerestricționat pentru $\ell_t = \ln \mathrm{RV}_t$ are un termen liber și pantele $\phi_1, \ldots, \phi_{22}$; HAR cu medii ale logaritmilor (A7) este versiunea lui restricționată
  - ϕ este proiecția liniară a lui ℓ_{t+1} pe cele 22 de laguri; date slab dependente cu momente de ordinul patru finite; erorile pot fi heteroscedastice și autocorelate
- **Cerințe**
  (a) Numărați restricțiile liniare și scrieți-le ca $R\phi = 0$.
  (b) Scrieți statistica Wald cu o covariantă HAC și distribuția ei sub H_0.
  (c) Precizați valoarea critică la 5%.
- **Interpretare**: dacă testul respinge într-un eșantion lung, ar trebui să renunțăm la HAR pentru prognoză?
- Raportați matricea restricțiilor, statistica Wald, gradele de libertate și valoarea critică.
- Notebook: secțiunea A8

Înapoi Lista exercițiilor

A9 [Propus]: QLIKE cu argumentele inversate — Cerință

Rezolvarea se discută la seminar

- **Întrebare**: ce prognoze recompensează o funcție QLIKE cu argumentele inversate?
- **Model**: A2
- **Date de intrare**: Presupuneți $\mathrm{RV} > 0$ și $\mathbb{E}_t[1/\mathrm{RV}] < \infty$. Un program calculează $L(\mathrm{RV}, F) = F/\mathrm{RV} - \ln(F/\mathrm{RV}) - 1$, QLIKE cu argumentele inversate
- **Cerințe**
  (a) Găsiți prognoza $F^\dagger$ care minimizează $\mathbb{E}_t[L(\mathrm{RV}, F)]$.
  (b) Comparați $F^\dagger$ cu $\mathbb{E}_t[\mathrm{RV}]$ folosind inegalitatea lui Jensen.
  (c) Precizați ce modele favorizează pierderea inversată.
- **Interpretare**: de ce poate trece această eroare de verificările elementare?
- Raportați condiția de ordinul întîi, minimizatorul și comparația cu ținta medie aritmetică.
- Notebook: secțiunea A9

◂ Înapoi ▸ Lista exercițiilor

A10 [Propus]: Limitele temperaturii — Cerință

Rezolvarea se discută la seminar

- **Întrebări**
  - ce se întîmplă cu ponderile nucleului la cele două temperaturi extreme?
  - de ce scalăm cu distanța mediană?
- **Model**: A3
- **Date de intrare**: n ferestre trecute, distanțe $\delta_\tau > 0$, temperatura inversă $\gamma = c/\mathrm{mediana}(\delta)$
- **Cerințe**
  (a) Găsiți limitele lui w și n_{eff} cînd $c \to 0$.
  (b) Găsiți limitele cînd $c \to \infty$ (presupuneți că fereastra cea mai apropiată este unică).
  (c) Arătați că înmulțirea tuturor distanțelor cu $k > 0$ lasă ponderile neschimbate.
- **Interpretare**: de ce este necesară scalarea cu mediana pentru a compara c între 50 de active?
- Raportați ambele limite ale vectorului ponderilor și mărimile efective, cu interpretarea lor.
- Notebook: secțiunea A10

Înapoi Lista exercițiilor

B5 [Propus]: Aria Lévy pentru futures E-mini S&P 500 — Cerință

Rezolvarea se discută la seminar

- **Întrebare**: are aria Lévy putere predictivă pentru volatilitatea din luna următoare a contractelor futures E-mini S&P 500 (ES)?
- **Model**: B1
- **Date de intrare**: Datele VOLARE pentru futures, 28 sept. 2009 – 30 ian. 2026, ES; RV la cinci minute, prețurile de deschidere și de închidere; ținta, regresorii și cele 22 de laguri Newey–West ca în B1
- **Cerințe**
  (a) Repetați pașii (a)–(d) din B1 pentru ES.
  (b) Testați coeficientul lui A_t la 1% și la 5%, bilateral.
  (c) Comparați cîștigul de R^2 cu cel pentru JPM.
- **Raportați**: tabelul regresiei, deciziile la ambele praguri, diferența de R^2 și graficul pe grupe.
- **Interpretare**
  - ce spune un coeficient pozitiv despre ferestrele cu arie negativă?
  - poate semnul singur să le eticheteze drept episoade de efect de levier?
- Notebook: secțiunea B5

Înapoi Lista exercițiilor

B6 [Propus]: HAR, SHAR și HARQ pentru petrol — Cerință

Rezolvarea se discută la seminar

- **Întrebare**: se confirmă asimetria SHAR și efectul HARQ pentru futures pe petrol (CL)?
- **Model**: B2
- **Date de intrare**: Datele VOLARE pentru futures, 28 sept. 2009 – 30 ian. 2026, CL; RV la cinci minute, semivariațele și cvarticitatea realizată; modelele, testele și Newey–West (5 laguri) ca în B2
- **Cerințe**
  (a) Estimați SHAR și testați $\beta^- > \beta^+$ ca în B2(b).
  (b) Estimați HARQ și testați $\beta_Q < 0$ ca în B2(c); reprezentați coeficienții cu intervale HAC de 95%.
  (c) Comparați ambele rezultate cu JPM.
- **Raportați**: estimările cu SE HAC, cele două teste unilaterale și deciziile la 5%, graficul coeficienților și comparația cu JPM.
- **Interpretare**: este asimetria mai puternică sau mai slabă decît pentru JPM și ce mecanism economic ar putea explica fiecare rezultat?
- Notebook: secțiunea B6

◂ Înapoi ▸ Lista exercițiilor

B7 [Propus]: Sig-LK față de log-HAR pentru ES, la o săptămînă — Cerință

Rezolvarea se discută la seminar

- **Întrebare**: este Sig-LK superior modelului log-HAR pentru ES cînd ținta este RV medie din următoarele cinci zile?
- **Model**: B3
- **Date de intrare**: Datele VOLARE pentru futures, 28 sept. 2009 – 30 ian. 2026, ES; specificația din B3 cu $h = 5$ și $c = 1{,}0$ (ales în Capitolul 20 pe blocul de validare al contractelor futures)
- **Cerințe**
  (a) Produceți prognoze la fiecare origine zilnică pentru RV medie pe cinci zile, ca în B3; evaluarea începe după ultima țintă de validare: originile 7 oct. 2014 – 23 ian. 2026.
  (b) Calculați raportul QLIKE și CI 95% prin bootstrap staționar (lungimea medie a blocului 10).
  (c) Rulați testul DM (lățime de bandă cel puțin $h = 5$) și testul GW condiționat cu instrumentele $(1, d_{t-5})$; reprezentați diferența cumulată de pierdere.
- **Raportați**: cele două pierderi medii, raportul și intervalul, statisticile și valorile p DM/GW, plus graficul diferenței cumulate.
- **Interpretare**
  - este CI asimetric și ce proprietăți ale pierderilor i-ar putea explica forma?
  - ce implică aceasta pentru puterea testului DM?
- Notebook: secțiunea B7

Înapoi Lista exercițiilor

B8 [Propus]: Testare multiplă la o lună față de HAR — Cerință

Rezolvarea se discută la seminar

- **Întrebare**: pe cîte active este Sig-LK superior modelului HAR la o lună, după corecția pentru 50 de teste?
- **Model**: B4
- **Date de intrare**: Rezultatele Capitolului 20: 50 de valori p DM unilaterale pentru „Sig-LK mai bun decît HAR" la $h = 22$ (QLIKE)
- **Cerințe**
  (a) Numărați respingerile la 5% fără corecție.
  (b) Aplicați Holm și BH ca în B4, numărați respingerile și numiți activul cu cea mai mică valoare p.
  (c) Comparați cu media pe active a rapoartelor QLIKE individuale ale Sig-LK față de HAR la $h = 22$ (1,334).
- **Raportați**: valorile p ordonate, ambele praguri și deciziile, cele trei numere de respingeri și activul identificat.
- **Interpretare**: cum poate Sig-LK să fie superior HAR pe mai multe active și totuși inferior în medie?
- Notebook: secțiunea B8

◂ Înapoi ▸ Lista exercițiilor

C2 [Propus]: Un nucleu de semnătură sensibil la nivel — Cerință

Rezolvarea se discută la seminar

- **Întrebare**: devin ponderile utile într-un crah dacă nucleul ține cont și de *nivelul* volatilității?
- **Date de intrare**
  - Nucleul Sig-LK este invariant la translație, deci nu poate recunoaște un regim de volatilitate ridicată; date: VOLARE, cele 40 de acțiuni, 2015–2026
  - Ideea: adăugăm un **punct de bază** Morrill et al. (2020): traiectoria pornește din media de antrenare a $\ln \mathrm{RV}$ și sare la prima valoare a ferestrei; prin Chen, $\mathrm{Sig} = \exp_{\otimes}(v) \otimes \mathrm{Sig}(X)$, cu v saltul de nivel standardizat și $\exp_{\otimes}(v) = (1, v, \frac{1}{2}v^{\otimes 2}, \dots)$ semnătura acestui unic segment de salt
- **Cerințe**
  (a) Scrieți o preînregistrare de o pagină: 40 de acțiuni, $h = 1$, specificația Sig-LK din B3, $c = 0{,}5$, QLIKE, DM cu Holm, subperioada COVID-19 raportată separat.
  (b) Implementați nucleul cu punct de bază (folosiți actualizarea Chen din notebook) cu aceleași trăsături ca Sig-LK și prognozați.
  (c) Comparați cu Sig-LK și log-HAR: rapoartele QLIKE medii (toate zilele și COVID-19), numărul activelor cu rezultate semnificativ mai bune față de log-HAR, fără corecție și după Holm.
- **Interpretare**: dacă sensibilitatea la nivel nu ajută, unde ați căuta în continuare (trăsături, țintă, orizont)?
- Notebook: secțiunea C2

Înapoi Lista exercițiilor

C2: saltul din punctul de bază

- Folosiți ordinea (timp, log RV, randament cumulat) și adăugați la început $v = (0, (\ln \mathrm{RV}_{\mathrm{first}} - \mu_{\mathrm{train}})/s_{\mathrm{train}}, 0)$. Media și abaterea standard folosesc log RV la cele 1 000 de origini de antrenare, fără observații viitoare.
- Păstrați scalarea traiectoriei din B3. Saltul suplimentar folosește abaterea standard a nivelurilor, nu a incrementelor. Componentele sale de timp și randament sînt zero.
- Spre deosebire de punctul de bază zero al lui Morrill et al. (2020), aici referința este media de antrenare. Măsurăm dacă nivelul inițial al volatilității este neobișnuit de ridicat față de acel eșantion.
- Raportați preînregistrarea, tabelele rapoartelor QLIKE pentru toate zilele și COVID-19, numărul cîștigurilor brute și după Holm.

Studiu suplimentar pe Quantinar

Cursuri avansate pe quantinar.com

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

Cod și date

- Notebook-ul seminarului (Colab); Quantlets: github.com/danpele/MFM/tree/main/Quantlets/Ch_20; date: volare.unime.it

Acronime folosite în acest material (1/3)

- **AI**: Artificial Intelligence — *inteligență artificială*
- **AR**: AutoRegressive (model) — *model autoregresiv*
- **AUD**: Australian Dollar (ISO 4217 code) — *dolarul australian (cod ISO 4217)*
- **AUDUSD**: Australian dollar / US dollar exchange rate (FX pair code) — *cursul dolar australian / dolar american (codul perechii FX)*
- **BH**: Benjamini–Hochberg (procedure) — *procedura Benjamini–Hochberg*
- **BIC**: Bayesian Information Criterion — *criteriul informațional bayesian*
- **CAD**: Canadian Dollar (ISO 4217 code) — *dolarul canadian (cod ISO 4217)*
- **CI**: Confidence Interval — *interval de încredere*
- **CL**: Crude oil (WTI light sweet) futures, NYMEX ticker — *contractul futures pe petrol (WTI), simbolul NYMEX*
- **COVID-19**: COronaVIrus Disease 2019 — *boala provocată de coronavirus, 2019*
- **DM**: Diebold–Mariano (test) — *testul Diebold–Mariano de comparare a prognozelor*
- **ES**: E-mini S&P 500 futures (CME ticker) — *contractul futures E-mini S&P 500 (simbolul CME)*

Acronime folosite în acest material (2/3)

- **EURUSD**: Euro / US dollar exchange rate (FX pair code) — *cursul euro / dolar american (codul perechii FX)*
- **FDR**: False Discovery Rate — *rata descoperirilor false*
- **FWER**: Family-Wise Error Rate — *probabilitatea de cel puțin o eroare de tip I în familia de teste*
- **FX**: Foreign eXchange — *piața valutară*
- **GBPUSD**: British pound / US dollar exchange rate (FX pair code) — *cursul liră sterlină / dolar american (codul perechii FX)*
- **GW**: Giacomini–White (test) — *testul Giacomini–White*
- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors) — *erori standard robuste la heteroscedasticitate și autocorelație*
- **HAR**: Heterogeneous AutoRegressive (model) — *model autoregresiv heterogen*
- **HARQ**: HAR model with Realised Quarticity — *model HAR cu cvarticitatea realizată*
- **JPM**: JPMorgan Chase & Co. (NYSE ticker) — *JPMorgan Chase & Co. (simbolul bursier NYSE)*
- **LASSO**: Least Absolute Shrinkage and Selection Operator — *regresie penalizată L1 cu selecția variabilelor*
- **MA**: Moving Average — *medie mobilă*
- **NW**: Newey–West (standard errors) — *erori standard Newey–West*

Acronime folosite în acest material (3/3)

- **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*
- **RQ**: Realised Quarticity — *cvarticitatea realizată*
- **RV**: Realised Variance — *varianța realizată*
- **SHAR**: Semivariance HAR model — *model HAR cu semivarianțe*
- **USDCAD**: US dollar / Canadian dollar exchange rate (FX pair code) — *cursul dolar american / dolar canadian (codul perechii FX)*
- **USDJPY**: US dollar / Japanese yen exchange rate (FX pair code) — *cursul dolar american / yen japonez (codul perechii FX)*
- **VOLARE**: VOLatility Archive for Realized Estimates (database) — *arhiva de estimatori realizați ai volatilității (bază de date)*

Bibliografie (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.

Bibliografie (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.