








Modelarea piețelor financiare

Capitolul 9: Volatilitate realizată



Daniel Traian PELE

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

Cum măsurăm volatilitatea latentă?



Torben G. Andersen. **Foto: Soralum (2012); CC BY-SA 4.0; Wikimedia Commons**

"We provide a framework for integration of high-frequency intraday data into the measurement, modeling, and forecasting of daily and lower frequency return volatilities and return distributions."

— *Andersen, Bollerslev, Diebold & Labys (2003)*

- Varianța realizată (RV, realised variance), suma pătratelor randamentelor pe 5 minute, măsoară varianța fiecărei zile aproape fără model
  - volatilitatea anualizată calculată din RV medie: 12,9% pentru SPY (fondul tranzacționat la bursă pe S&P 500), 45,1% pentru Bitcoin
- Randamentele peste noapte aduc 38% din varianța zilnică a SPY, deși piața este închisă 17,5 ore
- Testele de salt pe 1 489 de zile SPY: 236 de respingeri la 5% fără corecție
  - 17 după corecția de periodicitate și controlul FDR (false discovery rate)
- În afara eșantionului (*out-of-sample*), SPY 2022–2026: modelul log-HAR (heterogeneous autoregressive) are o pierdere QLIKE (quasi-likelihood) de 0,310, față de 0,343 pentru GARCH(1,1)-t
  - testul Diebold–Mariano: $t = -2{,}30$, $p = 0{,}021$

▸ Graficul central: Prognozele în timp (1/2)

De la prețurile intraday la prognoza variației (1/2)

Întrebările capitolului

- Partea I, **măsurare și diagnostic**: varianța realizată și precizia ei, zgomotul de microstructură, salturile
- Partea a II-a, **modelare și prognoză**: fapte stilizate, modelul HAR, evaluarea prognozelor față de GARCH
  - HAR: modelul autoregresiv heterogen (heterogeneous autoregressive)
  - GARCH: generalised autoregressive conditional heteroskedasticity (Capitolul 5)
- Extensii: volatilitatea aspră (*rough volatility*), un studiu de caz de învățare automată și cercetarea asistată de AI
  - AI: artificial intelligence
- Anexa: teoria asimptotică, estimatori robuști la zgomot și salturi, extensii ale modelelor
- Exemplele detaliate și subiectele avansate sînt în Aprofundări, după Concluzii
  - butonul din subsol duce la ele, iar butonul Înapoi revine la slide-ul de plecare

De la prețurile intraday la prognoza variației (2/2)

Rezultatele învățării

- calculați varianța realizată zilnică, eroarea ei standard și un interval de încredere din prețurile intraday
- interpretați graficul de semnătură (*signature plot*) ca diagnostic al zgomotului de microstructură
- calculați testul raportului pentru salturi, cu controlul ratei falselor descoperiri (FDR, false discovery rate)
- calculați estimările log-HAR și interpretați persistența lor
- derivați funcția de pierdere de tip cvasi-verosimilitate (QLIKE, quasi-likelihood) din verosimilitatea Normală
- interpretați comparațiile dintre prognozele de variață: Diebold–Mariano, Mincer–Zarnowitz și mulțimea de încredere a modelelor
- Aprofundări și Anexa: lectură pentru familiarizare

Volatilitatea latentă și măsurarea ei (1/2)



Torben G. Andersen, 2012

Foto: Soralum (2012); CC BY-SA 4.0; Wikimedia Commons

- Volatilitatea determină riscul, prețurile opțiunilor și ponderile portofoliilor, dar nu este observată niciodată
  - riscul: Capitolele 7–8; prețurile opțiunilor: Capitolul 12; ponderile portofoliilor: Capitolul 4
  - Un randament zilnic este o singură observație dintr-o distribuție a cărei dispersie vrem să o estimăm
  - GARCH (Capitolul 5) **deduce** varianța de azi din randamentele zilnice trecute, printr-un model
- Ideea: suma pătratelor modificărilor de preț dintr-o zi **măsoară** varianța zilei aproape fără model Andersen & Bollerslev (1998)
  - într-o singură zi există sute de modificări de preț
  - O contribuție timpurie: Merton (1980) a arătat că varianța, spre deosebire de medie, se estimează mai bine cu date mai frecvente

Volatilitatea latentă și măsurarea ei (2/2)

- Două obiecte diferite
  - varianța realizată: o **măsurare ex post** a zilei t, cunoscută după închidere
  - varianța GARCH: o **prognoză condiționată** a zilei t, făcută cu informația din ziua $t-1$
- Întîi măsurăm, apoi modelăm
  - Andersen, Bollerslev, Diebold & Labys (2001), Andersen, Bollerslev, Diebold & Ebens (2001) studiază distribuția volatilității realizate
  - Andersen, Bollerslev, Diebold & Labys (2003) o prognozează direct
  - Prezentarea de manual a subiectului: Franke, Härdle & Hafner (2019)

Cum citim datele intraday?

- SPY: un fond tranzacționat la bursă (ETF, *exchange-traded fund*) care urmărește indicele bursier Standard & Poor's 500
  - ETF: un fond ale cărui unități se tranzacționează la bursă
- UTC înseamnă Coordinated Universal Time
  - zilele Bitcoin sînt delimitate de miezul nopții în acest fus orar
- VIX: indicele de volatilitate (*Volatility Index*) al Chicago Board Options Exchange (CBOE)
  - o măsură a volatilității anticipate extrasă din prețurile opțiunilor
- Toate seriile se opresc la 18 septembrie 2026
- Randamentele sînt $r_i = 100(\ln P_i - \ln P_{i-1})$, în %, unde P_i este prețul de la sfîrșitul barei i
- Graficele și orele citate în acest capitol marchează **sfîrșitul** fiecărui interval de 5 minute
  - 09:35 este primul randament SPY, 09:30–09:35

Un model pentru prețurile intraday (1/2)

- **Întrebarea**: cum transformăm randamentele de 5 minute ale unei zile într-o măsurare a variației acelei zile?
- Notații: ziua t are M intervale de lungime $\Delta = 1/M$, randamente $r_{t,i}$, randamentul zilnic $r_t = \sum_i r_{t,i}$
  - Variațele sînt exprimate în $\%^2$
  - Volatilitatea anualizată este $\sqrt{252 \times \text{varianța}}$ pentru SPY și $\sqrt{365 \times \text{varianța}}$ pentru Bitcoin
  - Randamentul SPY peste noapte (de la închidere la deschiderea următoare) nu face parte din bare

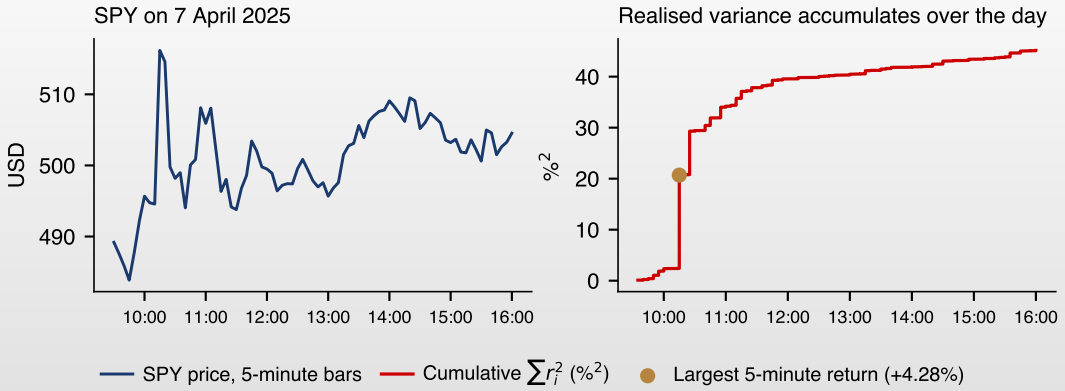
Un model pentru prețurile intraday (2/2)

- Logaritmul prețului în timp continuu: $dp(s) = \mu(s)\,ds + \sigma(s)\,dW(s)$, cu W o mișcare browniană (Capitolul 11)
  - $\sigma(s)$ este **volatilitatea instantanee**: se schimbă în cursul zilei și de la o zi la alta
- **Varianța integrată** a zilei t: mărimea pe care vrem să o măsurăm
  - $\mathrm{IV}_t = \int_{t-1}^{t} \sigma^2(s)\,ds$: varianța instantanee acumulată pe parcursul zilei; în acest capitol IV înseamnă mereu varianța integrată (integrated variance), nu volatilitatea implicită
- Dată fiind traiectoria lui σ, randamentul zilnic urmează distribuția Normală: $r_t \mid \sigma \sim N(\int \mu, \mathrm{IV}_t)$
  - ipoteză: σ este independent de W (fără efect de levier, *leverage effect*: scăderile de preț nu cresc volatilitatea)
  - ipoteză: drift-ul este nul (sau determinat de traiectoria lui σ)
  - Cu efect de levier (Anexă: semimartingale Itô) această normalitate condiționată nu mai are loc, dar teoria limită a varianței realizate (RV, slide-ul următor) rămîne valabilă

Definiție: varianța realizată

- **Varianța realizată**: suma pătratelor randamentelor intraday
  - $\mathrm{RV}_t = \sum_{i=1}^{M} r_{t,i}^2$, volatilitatea realizată $= \sqrt{\mathrm{RV}_t}$
- De ce pătrate: mediile condiționate ale pătratelor incrementelor aproximează $\int_{t-1}^{t} \sigma^2(s)\, ds$
  - pe un interval scurt $\mathbb{E}[r_{t,i}^2 \mid \sigma] \approx \sigma^2(s_i)\,\Delta$
  - suma observată a pătratelor converge la variația pătratică; nu este o sumă Riemann
- **Variația pătratică** a zilei t, cu $p = 100 \ln P$: $[p]_t - [p]_{t-1} = \operatorname{plim}_{M\to\infty} \sum_{i=1}^{M} r_{t,i}^2 = \mathrm{IV}_t$ pentru un preț fără salturi Barndorff-Nielsen & Shephard (2002)
  - Drift-ul aduce termeni de ordinul Δ pe zi, care dispar cînd $M \to \infty$; pentru SPY, media randamentelor de 5 minute este sub 1% din abaterea lor standard
- Fără parametri de estimat: RV_t este o măsurare **fără model** a varianței unei zile
- Randamentul zilnic la pătrat r_t^2 este cazul particular $M = 1$
  - nedeplasat (cu drift nul), dar extrem de zgomotos

Randamentele la cinci minute din 7 aprilie 2025 (1/2)



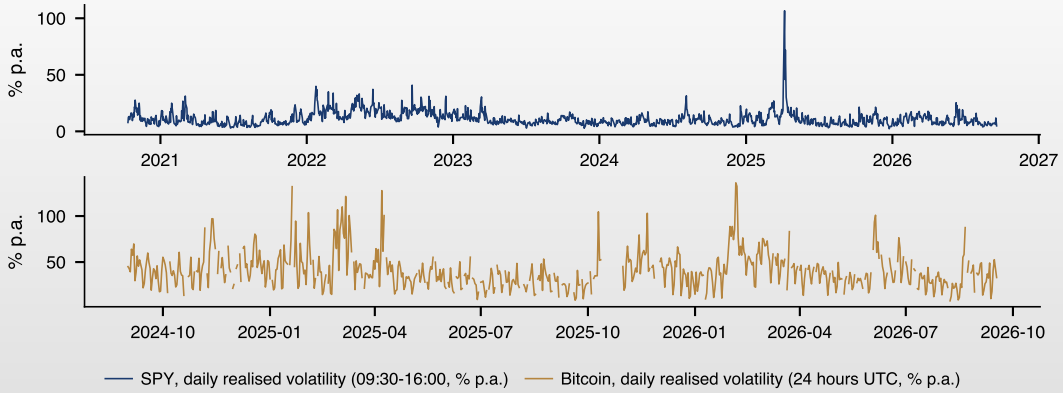
MFM_ch9_intraday_day ▶ Interactiv

Randamentele la cinci minute din 7 aprilie 2025 (2/2)

- Stînga: SPY pe bare de 5 minute în cea mai volatilă zi din eșantion
- Dreapta: suma cumulată a pătratelor randamentelor
- SPY a deschis cu un randament logaritmic peste noapte de $-3{,}27\%$, după anunțurile privind tarifele vamale
  - au urmat 78 de randamente de 5 minute
  - $\mathrm{RV} = 45{,}1\%^2$: volatilitatea realizată anualizată $\sqrt{252 \times 45{,}1} = 107\%$
- Un singur randament de 5 minute, de $+4{,}28\%$, a contribuit cu 41% din varianța realizată a zilei
  - intervalul care se încheie la 10:15
- De la deschidere la închidere ziua s-a încheiat cu $+3{,}09\%$
  - randamentul zilnic ascunde oscilațiile din timpul zilei pe care RV le înregistrează

MFM_ch9_intraday_day

Volatilitatea realizată zilnică: SPY și Bitcoin (1/2)



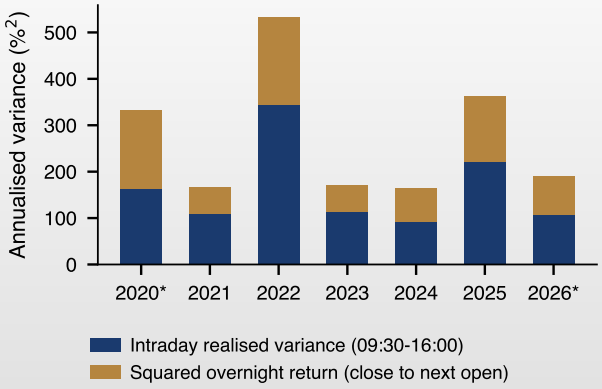
MFM_ch9_rv_series ▶ Interactiv

Volatilitatea realizată zilnică: SPY și Bitcoin (2/2)

- Volatilitate anualizată, calculată din randamente de 5 minute
  - SPY: pe sesiunea regulată din SUA (Statele Unite ale Americii)
  - Bitcoin: pe 24 de ore
  - golul din octombrie 2025 corespunde unor date lipsă
- SPY: volatilitatea calculată din RV medie este $\sqrt{252\,\overline{\mathrm{RV}}} = 12{,}9\%$ pe an
  - volatilitatea zilnică are mediana 9,7% și variază între 2,3% și 106,7%
  - Volatilitatea calculată din varianța medie depășește mediana: cîteva zile foarte volatile ridică media pătratică a volatilităților zilnice
- Bitcoin: volatilitatea calculată din RV medie este $\sqrt{365\,\overline{\mathrm{RV}}} = 45{,}1\%$ pe an
  - mediana zilnică este 37,8%
  - maximul, 136,1%, a fost atins pe 2026-02-05
- Ambele serii arată **gruparea volatilității** (*volatility clustering*), ca la GARCH (Capitolul 5)
  - dar fiecare punct este o măsurare, nu rezultatul unui model

MFM_ch9_rv_series

Riscul intraday și riscul peste noapte (1/2)



MFM_ch9_overnight ▶ Interactiv

Riscul intraday și riscul peste noapte (2/2)

- SPY, varianța zilnică medie pe ani, anualizată
  - varianța realizată intraday plus randamentul peste noapte la pătrat
  - 2020 și 2026 sînt ani incompleți
- Randamentele peste noapte aduc 38% din varianța zilnică SPY, deși piața este închisă 17,5 ore
  - Știrile (rezultate financiare, date macroeconomice, sesiunile din Asia și Europa) se acumulează și sînt încorporate în preț la deschidere
- Tipul de dovadă: descriere (arată ce s-a întîmplat în date, fără un test al cauzei)
- Volatilitatea anualizată a SPY
  - 16,6% de la închidere la închidere
  - 13,6% de la deschidere la închidere
  - 10,1% peste noapte
- Pentru riscul zilnic folosim varianța **totală**: $\mathrm{RV}_t + r^2_{\textbf{noapte},t}$
  - media anualizată: 16,4%, față de 12,9% doar din RV intraday
  - Alternative: scalarea RV intraday în sus sau ponderarea optimă a celor două părți Hansen & Lunde (2005)

MFM_ch9_overnight

Precizia varianței realizate

- Teorema limită centrală (CLT, *central limit theorem*) pentru RV, în forma fezabilă (calculabilă din date) Barndorff-Nielsen & Shephard (2002)
  - $\sqrt{M}\,(\mathrm{RV}_t - \mathrm{IV}_t)/\sqrt{2\,\mathrm{RQ}_t} \to N(0,1)$
  - $\mathrm{IQ}_t = \int_{t-1}^{t} \sigma^4(s)\,ds$: cvarticitatea integrată, care determină dispersia RV
  - IQ_t se estimează prin cvarticitatea realizată $\mathrm{RQ}_t = \frac{M}{3}\sum_i r_{t,i}^4$
- Eroarea standard: $\mathrm{se}(\mathrm{RV}_t) = \sqrt{\frac{2}{3}\sum_i r_{t,i}^4}$
- Intervalul de încredere (CI, confidence interval) de 95%: $\mathrm{RV}_t \exp\left(\pm 1{,}96\,\mathrm{se}(\mathrm{RV}_t)/\mathrm{RV}_t\right)$
  - construit pe scara log, deci întotdeauna pozitiv
  - Cu volatilitate constantă, $\mathrm{se}/\mathrm{RV} \approx \sqrt{2/M}$: 16% pentru $M = 78$
- Varianța limită $2\,\mathrm{IQ}_t$ este ea însăși aleatoare (o limită Normală mixtă)
  - înlocuirea ei cu $2\,\mathrm{RQ}_t$ este legitimă deoarece convergența este **stabilă**
  - enunțul formal în ▸ Anexă

De ce nu folosim fiecare tranzacție? (1/2)

- **Întrebarea**: de ce nu folosim fiecare tranzacție?
  - mai multe randamente micșorează eroarea relativă a RV ($\sqrt{2/M}$)
- Prețurile observate nu sînt prețul eficient: $p_i^{\mathbf{obs}} = p_i^* + \varepsilon_i$
  - deci $r_i = r_i^* + \varepsilon_i - \varepsilon_{i-\mathbf{1}}$
  - p_i^*: logaritmul prețului eficient (adevărat), ε_i: zgomotul din prețul observat, r_i^*: randamentul eficient
  - Sursele lui ε: oscilația între prețul de cumpărare (bid) și cel de vînzare (ask), pasul de cotare (discretizarea prețului), cotațiile neactualizate (*stale*) (Capitolul 10)

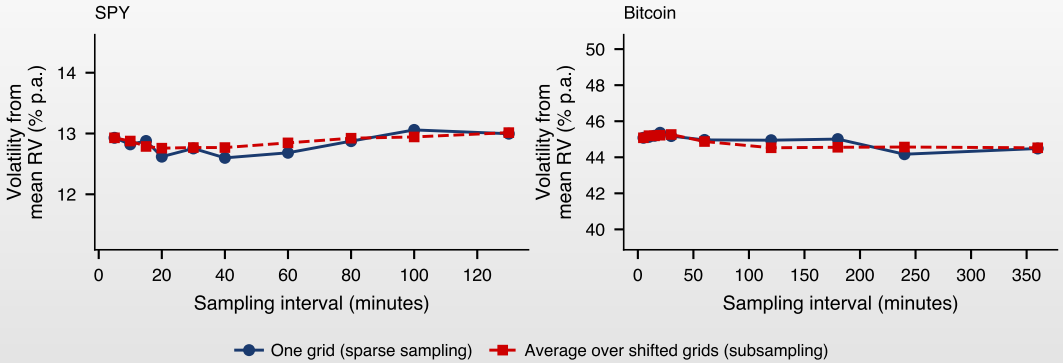
De ce nu folosim fiecare tranzacție? (2/2)

- Cu zgomot i.i.d. de variață ω^2 și n randamente pe zi Zhang, Mykland & Aït-Sahalia (2005), Hansen & Lunde (2006)
  - $\mathbb{E}[\mathrm{RV}_t^{(n)} \mid \sigma] = \mathrm{IV}_t + 2n\omega^2$ (drift nul, fără efect de levier): fiecare randament la pătrat conține $2\omega^2$ zgomot, deci deplasarea **crește** cu frecvența de eșantionare
  - Randamentele alăturate au în comun ε_{i-1}, cu semne opuse: $\mathrm{cov}(r_i, r_{i-1}) = -\omega^2$
- Compromis: $\mathbb{E}[(\mathrm{RV}^{(n)} - \mathrm{IV})^2] \approx 2\,\mathrm{IQ}/n + 4n^2\omega^4$
  - eroarea de discretizare plus deplasarea la pătrat
  - Numărul de randamente optim după MSE (eroarea pătratică medie), $n^* = \left(\mathrm{IQ}/(4\omega^4)\right)^{1/3}$ Bandi & Russell (2008): mai multe randamente reduc primul termen, dar îl cresc pe al doilea; derivarea în ▸ Anexă
- Exemplu numeric, cu $\mathrm{IV} = 1\%^2$ și $\omega = 0{,}01\%$, deci $\omega^2 = 10^{-4}\%^2$
  - volatilitate constantă în cursul zilei, deci $\mathrm{IQ} = \mathrm{IV}^2 = 1\ (\%^4)$
  - deplasarea $2n\omega^2$ este 0,016 ($\%^2$) la 5 minute ($n = 78$) și 4,680 la 1 secundă, mai mare decît IV însăși
- Numărul optim de randamente este $n^* = 292$, adică un randament la fiecare 80 secunde

De ce 5 minute?

- **Întrebarea**: ce frecvență alegem, cînd n^* depinde de ω^2 și de IQ_t?
  - ambele sînt necunoscute și diferite de la un activ la altul
- Frecvența optimă scade repede cînd zgomotul crește
  - $n^* = 292$ cu $\omega = 0{,}01\%$
  - $n^* = 14$ (un randament la circa 29 de minute) cu $\omega = 0{,}1\%$
- **Eșantionarea rară** la 5 minute ($n = 78$ într-o sesiune de 6,5 ore): un compromis robust Liu, Patton & Sheppard (2015)
  - deplasarea $2n\omega^2$ rămîne mică, iar eroarea relativă de discretizare este $\sqrt{2/78} = 16\%$
  - costul: renunțăm la majoritatea tranzacțiilor
- Barele din datele cursului au 5 minute, cea mai fină grilă disponibilă
  - graficul de semnătură (RV medie în funcție de intervalul de eșantionare) verifică pe date reale dacă 5 minute este suficient de rar

Graficul de semnătură pe date reale (1/2)



MFM_ch9_signature ▶ Interactiv

Graficul de semnătură pe date reale (2/2)

- Volatilitatea anualizată din RV medie în funcție de intervalul de eșantionare
  - o singură grilă față de media grilelor decalate în timp (subeșantionare)
- Volatilitatea rămîne aproximativ stabilă cînd intervalul de eșantionare crește: compatibil cu o deplasare redusă din zgomot, nu o dovadă că zgomotul lipsește
  - SPY: 12,9% la 5 minute, 12,8% la 30 de minute și 13,0% la 130 de minute
  - Bitcoin: 45,1% la 5 minute și 44,5% la 6 ore
  - Autocorelația de ordinul 1 a randamentelor de 5 minute este $-0{,}008$ pentru SPY și 0,005 pentru Bitcoin: zgomotul i.i.d. ar face-o negativă ($-\omega^2$), deci la 5 minute nu se detectează
- Limita acestor date: barele de 5 minute nu spun nimic despre zgomotul sub 5 minute
- Estimatori robuști la zgomot (Aprofundări), ca raport față de RV pe SPY: diferențe între estimatori, nu o deplasare din zgomot măsurată
  - RV pe două scări, pe grilele standard: 0,84; cu primul și ultimul preț al zilei în fiecare grilă: 0,97
  - nucleul realizat cu capete mediate (jittering): 0,92; cu prețurile brute de la capete: 0,95
  - cu 78 de bare, variantele teoretice pierd o parte din primul și din ultimul interval al zilei, cele mai volatile
- În acord cu Liu, Patton & Sheppard (2015): aproape 400 de estimatori pe 31 de active întrec rareori RV pe 5 minute
  - estimatorii robuști la zgomot aduc cîștig cînd zgomotul este mare față de semnal (date tick cu tick sau la nivel de secundă, active nelichide)
- Tipul de dovadă: descriere

MFM_ch9_signature

Salturile prețurilor: variația bipower (1/2)



Ole E. Barndorff-Nielsen, 2007

Foto: Thomas Steiner (2007); CC BY-SA 2.5; Wikimedia Commons

- **Întrebarea**: cît din varianța unei zile provine din salturi și cum separăm salturile de difuzie (partea continuă, $\sigma\, dW$)?
- Adăugăm salturi în model: $dp = \mu\, ds + \sigma\, dW + dJ$
  - J este un proces de salturi (știri, Capitolul 11)
  - Variația pătratică are acum două părți: $\mathrm{RV}_t \to \mathrm{IV}_t + \sum_j \kappa_j^2$, cu κ_j mărimile salturilor din ziua t
  - Motivul separării: salturile sînt mai puțin persistente decît volatilitatea difuzivă și sînt evaluate diferit (opțiuni, Capitolul 12)

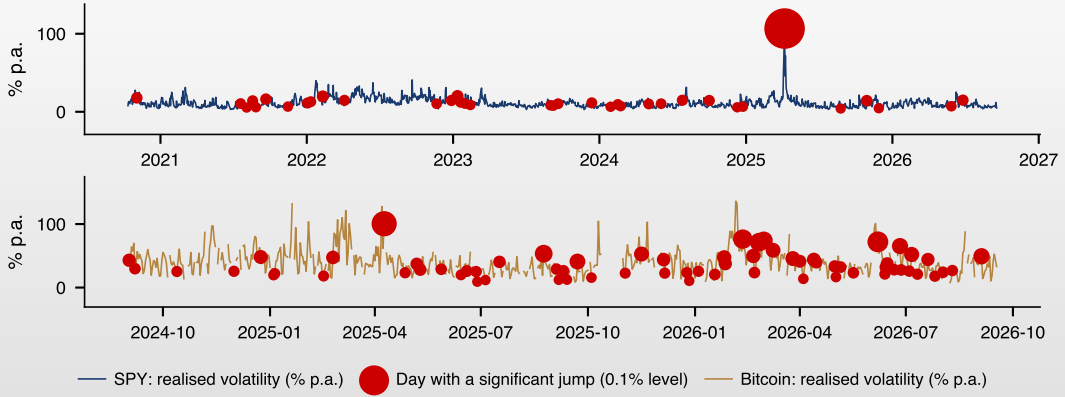
Salturile prețurilor: variația bipower (2/2)

- **Variația bipower** (BV) Barndorff-Nielsen & Shephard (2004): produse ale valorilor absolute ale randamentelor consecutive
  - $\mathrm{BV}_t = \mu_{\mathbf{1}}^{-\mathbf{2}} \frac{M}{M-\mathbf{1}} \sum_{i=\mathbf{2}}^{M} |r_{t,i}|\,|r_{t,i-\mathbf{1}}| \to \mathrm{IV}_t$
  - $\mu_{\mathbf{1}} = \mathbb{E}|Z| = \sqrt{2/\pi} = 0{,}798$ pentru $Z \sim N(0,1)$, deci $\mu_{\mathbf{1}}^{-\mathbf{2}} = \pi/2 = 1{,}571$; $\frac{M}{M-\mathbf{1}}$: normalizarea pentru eșantion finit
- Cazul de bază: salturi de activitate finită (un număr finit de salturi pe zi) și fără zgomot; alte cadre cer condiții suplimentare
- Atunci un salt intră într-un singur randament, înmulțit cu un vecin mic
  - efectul lui asupra BV dispare cînd M crește
- Partea de salt a zilei este $\max(\mathrm{RV}_t - \mathrm{BV}_t, 0)$
  - Exemplu, $M = 4$ randamente 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 0,78 (%$^{\mathbf{2}}$)

Un test pentru salturi

- Statistica raportului Huang & Tauchen (2005), Barndorff-Nielsen & Shephard (2006)
  - $z_t = \dfrac{(\mathrm{RV}_t - \mathrm{BV}_t)/\mathrm{RV}_t}{\sqrt{\theta\, M^{-1} \max\left(1, \mathrm{TQ}_t/\mathrm{BV}_t^2\right)}}$
  - $\theta = \pi^2/4 + \pi - 5 = 0{,}609$; cvarticitatea tripower $\mathrm{TQ}_t = M\mu_{4/3}^{-3} \frac{M}{M-2} \sum_i |r_{t,i}|^{4/3} |r_{t,i-1}|^{4/3} |r_{t,i-2}|^{4/3}$, $\mu_{4/3} = \mathbb{E}|Z|^{4/3}$
  - Numărătorul este ponderea relativă a saltului $(\mathrm{RV}_t - \mathrm{BV}_t)/\mathrm{RV}_t$, numitorul eroarea ei standard
  - θ este constanta din dispersia asimptotică, iar TQ, cvarticitatea tripower, estimează IQ_t robust la salturi
- Fără salturi $z_t \to N(0,1)$
  - un salt face $\mathrm{RV} > \mathrm{BV}$ și z_t mare și pozitiv
- Folosim $\alpha = 0{,}1\%$ (valoarea critică 3,09) pentru zilele cu salt din graficul următor
- Componenta de salt (estimează $\sum_j \kappa_j^2$ al procesului J): $J_t = \mathrm{RV}_t - \mathrm{BV}_t$ într-o zi cu salt semnificativ, altfel $J_t = 0$
  - partea continuă $C_t = \mathrm{RV}_t - J_t$

Zile cu salt: SPY și Bitcoin (1/2)



MFM_ch9_jumps ▶ Interactiv

Zile cu salt: SPY și Bitcoin (2/2)

- Punctele roșii: zile cu salt semnificativ la nivelul de 0,1%
  - mărimea punctului este proporțională cu partea de salt J_t
- SPY: 36 de zile cu salt din 1 489 (2,4%)
  - salturile reprezintă 2,8% din varianța realizată totală
  - într-o zi cu salt, saltul reprezintă în medie 36% din RV a zilei
  - cea mai mare statistică apare pe 2021-08-27 ($z = 5{,}2$), în intervalul 10:00–10:05, în timpul discursului președintelui Fed (Rezerva Federală a SUA) de la Jackson Hole
- Bitcoin: 67 de zile cu salt din 659 (10,2%), dar doar 1,8% din varianță
  - multe salturi mici în zile liniștite (weekenduri), în care o singură mișcare izolată domină ziua
- Zilele agitate din aprilie 2025 **nu** sînt în mare parte zile cu salt
  - multe randamente mari la rînd cresc și BV

MFM_ch9_jumps

Testarea multiplă și salturile detectate eronat

- Din 1 489 de teste zilnice la nivelul de 5% rezultă 236 de respingeri: care dintre ele sînt false?
  - față de 74 așteptate dacă nu ar exista salturi
- O parte din exces provine din sezonalitatea intraday: volatilitatea mai mare la deschidere și la închidere (Aprofundări)
  - împărțirea randamentelor la factorul de periodicitate f_i Boudt, Croux & Laurent (2011) reduce zilele respinse la 5% de la 236 la 169
- Benjamini & Hochberg (1995): ordonăm valorile p (*p-values*) și respingem cele mai mici k
  - k este cel mai mare indice pentru care $p_{(k)} \leq kq/N$, unde $p_{(k)}$ este al k-lea cea mai mică valoare p, N numărul de teste, iar q nivelul FDR țintă
  - aici valoarea p a fiecărei zile este $1 - \Phi(z_t)$, unilateral, pentru statistica raportului
  - FDR, $\mathbb{E}$[false/respingeri], rămîne $\leq q$ (false/respingeri $:= 0$ cînd nu se respinge nimic)
  - Garanția cere 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)
  - testele zilnice de salt îndeplinesc aceste condiții doar aproximativ

Numărul de salturi cu control pentru testarea multiplă

- SPY, 1 489 de zile, fără corecția de periodicitate: 43 de zile cu salt cu FDR la 5% și 14 cu FDR la 1%
- Cu corecția Bonferroni (fiecare zi testată la 5%/1 489, FWER 5%) rămîn 7
  - FWER: probabilitatea de cel puțin o respingere falsă
- Cu FDR la 5% după corecția de periodicitate rămîn 17
  - de la 236 de respingeri fără corecție la 17: scăderea nu arată care dintre respingeri erau false
- Tipul de dovadă: descriere
- Bajgrowicz, Scaillet & Treccani (2016), pe eșantionul și testele lor: fără corecție pentru testarea multiplă, majoritatea salturilor detectate sînt false
  - cele reale sînt rare și se concentrează în jurul anunțurilor
- Christensen, Oomen & Podolskij (2014), pe datele lor tick cu tick: majoritatea „salturilor" de 5 minute sînt creșteri bruște, dar continue, ale volatilității
- Un test zilnic răspunde la întrebarea „a conținut ziua t un salt?"
  - localizarea saltului în cursul zilei cere un test intraday (Lee–Mykland, Anexă)

Autoevaluare: măsurare, zgomot și salturi (1/2)

- De ce folosește eroarea standard a RV puterile a patra ale randamentelor?
  - Dispersia lui RV depinde de cvarticitatea integrată IQ_t, estimată prin $RQ_t = \frac{M}{3} \sum_i r_{t,i}^4$
- De ce este subestimat riscul zilnic al SPY dacă folosiți doar RV intraday?
  - Randamentele peste noapte aduc 38% din varianța zilnică a SPY
- De ce supraestimează varianța eșantionarea fiecărei tranzacții?
  - Fiecare randament la pătrat adaugă $2\omega^2$ zgomot, deci deplasarea $2n\omega^2$ crește cu numărul de randamente

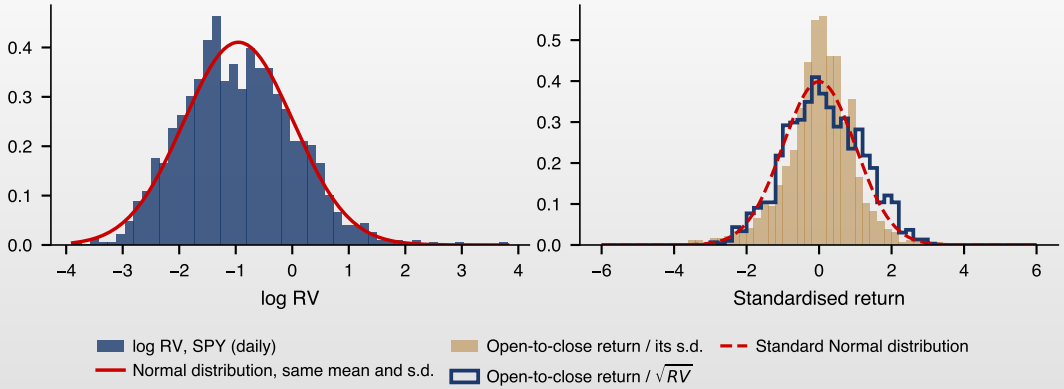
Autoevaluare: măsurare, zgomot și salturi (2/2)

- Ce vă spune un grafic de semnătură plat?
  - Este compatibil cu o deplasare redusă din zgomot la 5 minute și mai rar; nu exclude zgomotul
- Ce nu vă spune?
  - Nu spune nimic despre zgomotul sub 5 minute
- De ce rămîne variația bipower apropiată de IV într-o zi cu salt, iar RV nu?
  - Cu salturi de activitate finită și fără zgomot, saltul intră într-un singur randament: în RV la pătrat, în BV înmulțit cu un vecin mic
- Cu 1 489 de teste zilnice de salt, de ce „$p < 0{,}05$" nu este o dovadă de salt?
  - La 5%, 1 489 de zile fără salturi ar da circa 74 de respingeri false
- Ce controlează FDR?
  - FDR controlează proporția așteptată de respingeri false printre respingeri

Ce regularități are volatilitatea realizată?

- **Întrebarea**: ce proprietăți ale RV trebuie să reproducă un model de prognoză?
- Distribuția: este RV aproximativ lognormală? Dacă da, modelăm $\ln \mathrm{RV}$
- Randamentele împărțite la $\sqrt{\mathrm{RV}_t}$: cît din cozile grele se datorează volatilității variabile?
- Persistența: cît de lent scade autocorelația lui $\ln \mathrm{RV}$?
- Volatilitatea implicită: prognozează VIX volatilitatea realizată în luna următoare?

Distribuții: log RV și randamente standardizate (1/2)



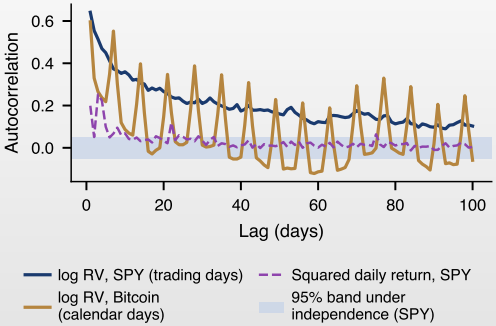
MFM_ch9_distributions ▶ Interactiv

Distribuții: log RV și randamente standardizate (2/2)

- Stînga: log RV zilnică pentru SPY și densitatea Normală
- Dreapta: randamentele deschidere-închidere scalate
  - cu abaterea lor standard (plin)
  - cu $\sqrt{\mathrm{RV}_t}$ (contur)
- Asimetria și kurtosis-ul pentru SPY (pentru distribuția Normală: 0 și 3)
  - RV: asimetria 20,0 și kurtosis-ul 550
  - log RV: asimetria 0,29 și kurtosis-ul 3,21
  - Log RV pentru Bitcoin are asimetria $-0{,}27$ și kurtosis-ul 3,33: de aceea modelăm volatilitatea pe scară logaritmică
- Randamentele deschidere–închidere: kurtosis-ul scade de la 15,73 la 2,66 după împărțirea la $\sqrt{\mathrm{RV}_t}$
  - kurtosis-ul empiric se poate calcula și cînd momentul de ordin patru al populației este infinit; atunci nu estimează un kurtosis finit al populației (Capitolul 1)
  - abaterea standard după împărțire: 1,05
  - Bitcoin: kurtosis-ul scade de la 7,56 la 2,77
- Cozile grele ale randamentelor zilnice (Capitolul 1) provin, ex post, în mare parte din volatilitatea variabilă
  - aproximativ un **amestec** de distribuții Normale cu variațele IV_t
  - exact doar fără efect de levier și fără salturi (secțiunea Varianța realizată)
- Tipul de dovadă: descriere

MFM_ch9_distributions

Persistența volatilității realizate (1/2)



- Autocorelații pînă la 100 de zile
  - log RV pentru SPY (varianța totală) și Bitcoin
  - randamentele zilnice SPY la pătrat
- Log RV pentru SPY: dependență **foarte persistentă**
  - 0,64 la lagul 1, 0,24 la o lună, încă 0,10 la 100 de zile
  - un model autoregresiv AR(1) ar scădea geometric: $0{,}64^{22} \approx 0{,}00005$
  - o funcție de autocorelație empirică este compatibilă cu memoria lungă (autocorelații care scad mai lent decît geometric), dar nu o poate dovedi

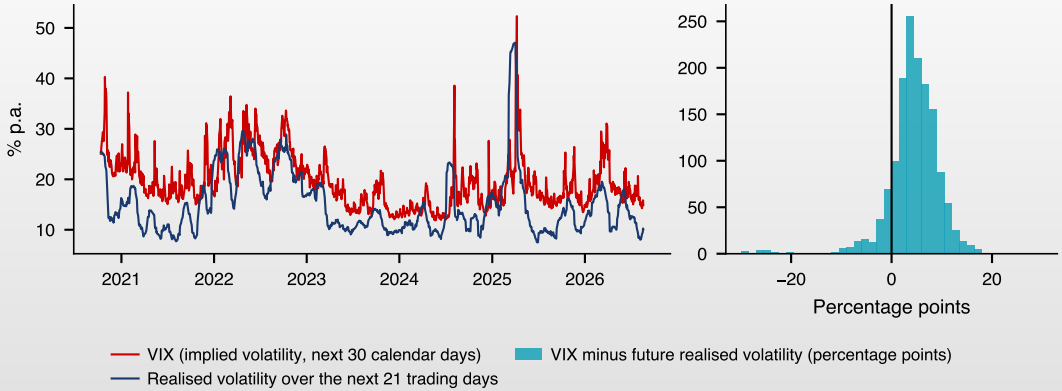
▶ Interactiv

Persistența volatilității realizate (2/2)

- Randamentele zilnice la pătrat arată aceeași persistență, ascunsă de zgomot
  - autocorelația la lagul 1: 0,20
  - autocorelația lui r_t^2 presupune moment de ordin patru finit, pus sub semnul întrebării în Capitolul 1
- Bitcoin: vîrfuri la 7, 14, 21, ... zile: un **ciclu săptămînal**
  - RV de weekend este doar 0,41 din RV din zilele lucrătoare (volatilitate anualizată de 27% sîmbăta, față de 48% miercurea)
  - fără știri bursiere și macroeconomice, cu mai puțini investitori instituționali

MFM_ch9_acf

Volatilitatea implicită față de cea realizată (1/2)



MFM_ch9_vix ▶ Interactiv

Volatilitatea implicită față de cea realizată (2/2)

- Stînga: VIX și volatilitatea realizată SPY în următoarele 21 de zile de tranzacționare
- Dreapta: diferența dintre ele
- VIX a avut media 19,52%, iar volatilitatea realizată ulterior a avut media 15,15%
- VIX a fost mai mare în 88% din 1 468 de zile, în medie cu 4,36 puncte procentuale: diferența medie ex post de volatilitate
- **Prima de risc a variației** (VRP) Carr & Wu (2009), Bollerslev, Tauchen & Zhou (2009)
  - $Q_t = \frac{252}{21} \sum_{j=1}^{21} V_{t+j}$: varianța totală anualizată în următoarele 21 de zile de tranzacționare
  - $\text{VRP}_t = \mathbb{E}_t^{\mathbb{Q}}[Q_t] - \mathbb{E}_t^{\mathbb{P}}[Q_t]$: așteptarea neutră la risc ($\mathbb{Q}$, din prețurile opțiunilor) minus cea reală ($\mathbb{P}$); VIX^2 aproximează primul termen (Capitolul 12)
  - $\text{VIX}_t^2 - Q_t \approx \text{VRP}_t + (\mathbb{E}_t^{\mathbb{P}}[Q_t] - Q_t)$: diferența ex post adaugă primei o eroare de prognoză
  - semnul diferenței realizate nu identifică semnul primei ex ante
- Tipul de dovadă: descriere

MFM_ch9_vix

Modelul HAR-RV (1/2)

- **Întrebarea**: cum prognozăm RV de mîine, cînd autocorelația lui $\ln \mathrm{RV}$ scade foarte lent?
  - autocorelația este încă 0,10 la 100 de zile
- Corsi (2009): $\mathrm{RV}_{t+1} = \beta_0 + \beta_d \mathrm{RV}_t + \beta_w \mathrm{RV}_t^{(w)} + \beta_m \mathrm{RV}_t^{(m)} + \varepsilon_{t+1}$
  - $\mathrm{RV}_t^{(w)} = \frac{1}{5} \sum_{j=0}^{4} \mathrm{RV}_{t-j}$ (săptămîna), $\mathrm{RV}_t^{(m)} = \frac{1}{22} \sum_{j=0}^{21} \mathrm{RV}_{t-j}$ (luna); Bitcoin se tranzacționează zilnic, deci folosim 7 și 30 de zile
  - β_d, β_w, β_m sînt ponderile RV de ieri și ale RV medii din ultima săptămînă și din ultima lună; β_0 este constanta, iar ε_{t+1} eroarea de prognoză
- Ipoteza pieței eterogene Müller et al. (1997): investitori cu orizonturi zilnice, săptămînale și lunare
  - de aici „H" (heterogen) din HAR-RV (*Heterogeneous AutoRegressive model of Realised Volatility*), modelul HAR al varianței realizate
  - cîteva componente cu persistență diferită aproximează persistența observată pe orizonturi finite
  - un HAR staționar este un AR finit: memorie scurtă în sens asimptotic

Modelul HAR-RV (2/2)

- Un AR(22) cu doar 3 pante libere
  - ponderea la lagul 1: $\beta_d + \beta_w/5 + \beta_m/22$
  - ponderea la lagurile 2–5: $\beta_w/5 + \beta_m/22$
  - ponderea la lagurile 6–22: $\beta_m/22$
- Estimare prin OLS (ordinary least squares, metoda celor mai mici pătrate)
  - erorile sînt heteroscedastice și autocorelate, deci folosim erori standard Newey–West
  - Newey–West este o eroare standard HAC (*heteroskedasticity and autocorrelation consistent*), adică robustă la heteroscedasticitate și autocorelare

Modelul log-HAR

- De ce logaritmi
  - $\ln \mathrm{RV}$ este aproape Normală, iar RV are cozi foarte grele (slide-ul *Distribuții*)
  - prognoza $\exp(\cdot)$ este întotdeauna pozitivă
  - zilele extreme nu mai domină estimarea OLS
- $\ln \mathrm{RV}_{t+1} = \beta_0 + \beta_d \ln \mathrm{RV}_t + \beta_w \overline{\ln \mathrm{RV}}_t^{(w)} + \beta_m \overline{\ln \mathrm{RV}}_t^{(m)} + \varepsilon_{t+1}$
  - $\overline{\ln \mathrm{RV}}_t^{(w)} = \frac{1}{5} \sum_{j=0}^{4} \ln \mathrm{RV}_{t-j}$ și $\overline{\ln \mathrm{RV}}_t^{(m)} = \frac{1}{22} \sum_{j=0}^{21} \ln \mathrm{RV}_{t-j}$
  - adică medii geometrice ale variațelor
- Prognoza variației: $\exp(\hat{y} + \hat{s}^2/2)$
  - $\hat{y}$: valoarea ajustată pentru $\ln \mathrm{RV}$; $\hat{s}^2$: variația reziduurilor
  - media condiționată dacă erorile sînt $N(0, s^2)$ (aproximare lognormală)
  - pentru erori care nu urmează distribuția Normală folosim factorul de corecție (*smearing factor*) $\frac{1}{T} \sum_t e^{\hat{\varepsilon}_t}$ Duan (1983)
  - cu erori heteroscedastice, ambele corecții sînt doar aproximări

Estimări HAR: SPY și Bitcoin

	SPY, HAR		Bitcoin, HAR	
	niveluri	log	niveluri	log
β_0	0,273	−0,079	1,496	0,223
	(0,064)	(0,027)	(0,412)	(0,094)
β_d	0,399	0,362	0,412	0,380
	(0,089)	(0,037)	(0,089)	(0,041)
β_w	0,222	0,336	0,409	0,465
	(0,096)	(0,048)	(0,122)	(0,080)
β_m	0,114	0,168	−0,065	−0,022
	(0,053)	(0,044)	(0,097)	(0,098)
R^2	0,31	0,47	0,34	0,36

- Eșantioanele complete
  - SPY: varianța zilnică totală (RV intraday plus randamentul peste noapte la pătrat), 1 467 de zile
  - Bitcoin: RV pe 24 de ore, 547 de zile în care sînt disponibili toți cei trei regresori HAR
- Erorile standard Newey–West în paranteze
  - erorile HAC cer regresori ortogonali pe eroare și o covarianță pe termen lung finită a lui $x_t\varepsilon_t$; pe niveluri, aceasta implică momentul de ordin patru al RV, nu al randamentelor (Capitolul 1)
  - HAC corectează autocorelarea și heteroscedasticitatea, nu endogenitatea

Interpretarea estimărilor HAR

- SPY: toate cele trei componente sînt semnificative
  - Persistența log-HAR $\hat{\beta}_d + \hat{\beta}_w + \hat{\beta}_m = 0{,}87$: șocurile se sting lent, ca la GARCH (Capitolul 5)
- $R^2 = 0{,}47$ pentru log-varianță și 0,31 pentru varianța pe niveluri
  - cele două valori nu sînt comparabile, deoarece variabilele dependente diferă
  - comparația se face prin pierderile în afara eșantionului pe o țintă comună (secțiunea Prognoze: HAR față de GARCH)
- Bitcoin: componentele zilnică și săptămînală contează
  - coeficientul lunar este aproape de 0 ($-0{,}022$, SE 0,098)
  - Eșantion scurt (547 de zile utilizabile din circa doi ani de date) și un ciclu săptămînal puternic: săptămîna preia cea mai mare parte a memoriei
- Tipul de dovadă: asociere (variabilele se mișcă împreună; nu dovedește o relație cauzală)
  - o ajustare predictivă în eșantion (*in-sample*); decid pierderile în afara eșantionului

HARQ și SHAR

- **HARQ** (*HAR model with Realised Quarticity*, model HAR cu cvarticitatea realizată) Bollerslev, Patton & Quaedvlieg (2016): RV de ieri este măsurată cu eroare, iar eroarea atenuează panta zilnică
  - $\mathrm{Var}(\mathrm{RV}_t - \mathrm{IV}_t) \approx 2\,\mathrm{IQ}_t / M$ (Anexă)
  - $\mathrm{RV}_{t+1} = \beta_0 + (\beta_d + \beta_Q \sqrt{\mathrm{RQ}_t})\,\mathrm{RV}_t + \beta_w \mathrm{RV}_t^{(w)} + \beta_m \mathrm{RV}_t^{(m)} + \varepsilon_{t+1}$
  - $\beta_Q < 0$: ponderea RV de ieri scade în zilele cu cvarticitate realizată mare, adică cu eroare de măsurare mare
- **SHAR** (*Semivariance HAR model*, model HAR cu semivarianțe) Patton & Sheppard (2015): semivarianțele realizate împart RV după semnul randamentului (Anexă)
  - $\mathrm{RS}_t^+ = \sum_i r_{t,i}^2 1\{r_{t,i} > 0\}$, $\mathrm{RS}_t^- = \sum_i r_{t,i}^2 1\{r_{t,i} < 0\}$, deci $\mathrm{RS}_t^+ + \mathrm{RS}_t^- = \mathrm{RV}_t$
  - $\mathrm{RV}_{t+1} = \beta_0 + \beta^+ \mathrm{RS}_t^+ + \beta^- \mathrm{RS}_t^- + \beta_w \mathrm{RV}_t^{(w)} + \beta_m \mathrm{RV}_t^{(m)} + \varepsilon_{t+1}$
  - motivul: volatilitatea „rea", din randamente negative, prezice mai multă volatilitate viitoare decît cea „bună", deci $\beta^- > \beta^+$

Care model prognozează mai bine varianța realizată?

- Ținta: varianța zilnică de mîine, prognozată doar cu datele disponibile în seara precedentă
  - SPY: varianța totală, de la 3 ianuarie 2022 pînă la 18 septembrie 2026 (1 180 de zile)
  - Bitcoin: de la 1 martie 2025 (422 de zile)
- Cinci modele
  - log-HAR și HAR pe varianța realizată din date intraday, reestimate zilnic pe o fereastră de estimare care se extinde în timp
  - GARCH(1,1) cu erori Student-t pe randamente zilnice din 2000 (SPY) sau 2014 (Bitcoin), reestimat la fiecare 21 de zile (GARCH: Bollerslev (1986); erori Student-t: Bollerslev (1987))
  - EWMA (media mobilă ponderată exponențial) pe randamente zilnice, cu factorul de atenuare $\lambda = 0{,}94$ (Capitolul 5)
  - „RV de ieri” (mers aleator, *random walk*: prognoza este ultima RV observată)
- Modelele, datele de intrare, lungimea istoricului și frecvența reestimării diferă: rezultatele compară aceste proceduri, așa cum sînt implementate, nu arhitectura HAR ca atare

Evaluarea față de o țintă zgomotoasă (1/2)

- Varianța adevărată nu este observată
  - comparăm prognozele F_t cu proxy-ul RV_t
  - O funcție de pierdere este **robustă** dacă ordinea după pierderea *așteptată* obținută cu un proxy nedeplasat condiționat (dată fiind informația prognozelor) coincide cu ordinea obținută cu varianța adevărată Patton (2011)
- MSE $= (\mathrm{RV}_t - F_t)^2$: robustă
  - măsoară erorile în unități de varianță
  - este dominată de zilele cele mai volatile
- QLIKE $= \dfrac{\mathrm{RV}_t}{F_t} - \ln \dfrac{\mathrm{RV}_t}{F_t} - 1 \geq 0$: robustă
  - depinde doar de raport
  - penalizează mai mult subestimarea
  - De unde vine QLIKE: pentru $r \sim N(0, F)$, $-2\ln$ densitatea este $\ln F + r^2/F$ plus o constantă; cu RV_t în locul lui r^2 și scăzînd minimul, atins la $F_t = \mathrm{RV}_t$, obținem formula de mai sus

Evaluarea față de o țintă zgomotoasă (2/2)

- Greșeală frecventă: MAE sau MSE pe volatilitate în loc de variață **nu** sînt robuste
  - MAE: eroarea absolută medie
  - Andersen & Bollerslev (1998): față de r_t^2, prognozele GARCH explicau cîteva procente din variație; față de RV aceleași prognoze explică mult mai mult: problema era zgomotul proxy-ului, nu modelul

Compararea a două prognoze: Diebold–Mariano

- Testul Diebold & Mariano (1995) (DM): $t = \bar{d}/\mathrm{se}_{\mathrm{NW}}(\bar{d}) \to N(0,1)$
  - diferența de pierdere: $d_t = L(\mathrm{RV}_t, F_t^A) - L(\mathrm{RV}_t, F_t^B)$
  - $L(\mathrm{RV}_t, F_t)$ este pierderea (QLIKE sau MSE) a prognozei F_t, $\bar{d}$ media lui d_t, iar $\mathrm{se}_{\mathrm{NW}}$ eroarea ei standard Newey–West
  - $t < 0$: modelul A are pierderea mai mică
  - Folosim o eroare standard HAC, deoarece d_t este autocorelat (Capitolul 8)
- Ce determină t: diferența medie **și** cît de stabilă este diferența de la o zi la alta
  - Verificare, ignorînd autocorelarea, $T = 100$ de zile: o diferență cu media 0,01 și abaterea standard 0,02 dă $t = 0{,}01/(0{,}02/\sqrt{100}) = 5$
  - o diferență cu media 0,03 și abaterea standard 0,30 dă $t = 1$
  - două prognoze care se mișcă împreună au o diferență stabilă: o diferență medie mică poate fi atunci foarte semnificativă
- Condiții: o teoremă limită centrală pentru d_t, cu varianța pe termen lung pozitivă
  - QLIKE: $d_t = \mathrm{RV}_t(1/F_t^A - 1/F_t^B) + \ln(F_t^A/F_t^B)$, deci momentele lui d_t depind și de prognozele de la numitor
  - modele imbricate (HAR în SHAR sau HARQ) cu parametri estimați: la acuratețe egală d_t poate degenera, iar $N(0,1)$ este doar o aproximare convențională (Clark & McCracken)
  - alternative valide: ferestre mobile (*rolling*) de lungime fixă cu testul Giacomini–White (GW) sau un bootstrap care reestimează modelele

Calibrarea: Mincer–Zarnowitz

- Mincer & Zarnowitz (1969) (MZ): $\mathrm{RV}_t = a + b\,F_t + u_t$
  - regresăm ținta realizată pe prognoză
  - a: termenul liber, b: panta, u_t: eroarea
  - O prognoză calibrată satisface $a = 0$ și $b = 1$ în această regresie liniară; $\mathbb{E}[\mathrm{RV}_t - F_t] = 0$ singur nu implică acest lucru
  - $b < 1$: prognozele mari sînt urmate de medii realizate sub dreapta de 45°
  - $b > 1$: prognozele mari sînt urmate de medii realizate peste dreapta de 45°
  - Verificare: o prognoză de 2 în zilele calme și de 4 în zilele agitate, cu medii realizate de 1,5 și 4,5, are $b = (4{,}5 - 1{,}5)/(4 - 2) = 1{,}5 > 1$ și $a = 1{,}5 - 1{,}5 \times 2 = -1{,}5$
  - Testul Wald comun folosește covarianța Newey–West: cere u_t ortogonal pe $(1, F_t)$ și o covarianță pe termen lung finită a lui $(1, F_t)\,u_t$
  - nerespingerea ipotezei $a = 0$, $b = 1$ nu dovedește calibrarea
  - R^2 măsoară cît din variație explică prognoza
  - R^2 este limitat de zgomotul din RV: nici varianța adevărată nu ar ajunge la 1 Andersen & Bollerslev (1998)

Rezultate: SPY, 2022–2026 (1/2)

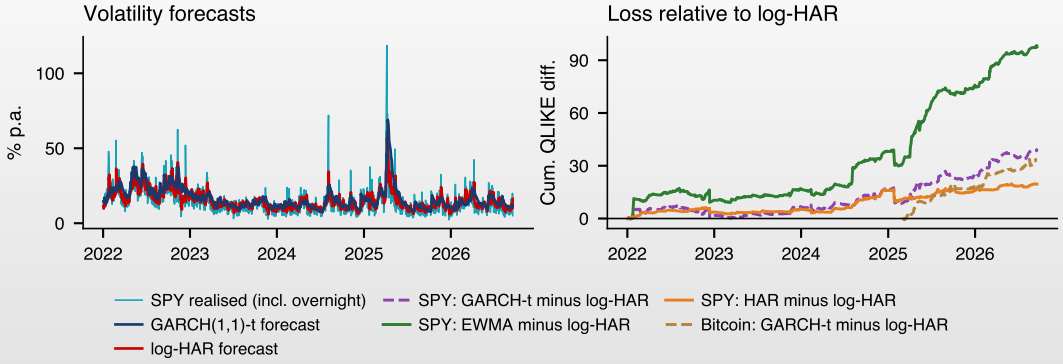
Model	QLIKE	DM_{QLIKE} t (p)	MSE ($\%^4$)	DM_{MSE} p	MZ b	MZ R^2
log-HAR	0,310	–	4,15	–	1,16	0,30
HAR	0,326	−1,95 (0,051)	4,48	0,117	0,75	0,26
GARCH(1,1)-t	0,343	−2,30 (0,021)	4,87	0,188	0,70	0,20
EWMA	0,393	−4,02 ($<$ 0,001)	5,14	0,050	0,74	0,14
RV de ieri	0,502	−5,95 ($<$ 0,001)	5,41	0,060	0,54	0,29

- 1 180 de prognoze pentru ziua următoare ale variației zilnice totale
  - DM față de log-HAR: $d_t = L(\text{log-HAR}) - L(\text{model})$, cu L = QLIKE sau MSE, Newey–West
  - $t < 0$ favorizează log-HAR
  - condițiile pentru DM și MZ privesc diferențele de pierdere și erorile MZ (slide-ul Diebold–Mariano), nu momentele randamentelor
- Dintre aceste proceduri, log-HAR are cele mai mici QLIKE și MSE
  - QLIKE este cu 10% sub GARCH(1,1)-t
  - Față de EWMA și RV de ieri cîștigul este foarte semnificativ; față de HAR pe niveluri doar marginal
- Diferențele MSE nu sînt semnificative la 5% (EWMA la limită, $p = 0{,}050$)
  - cîteva zile de criză (aprilie 2025) domină erorile la pătrat

Rezultate: SPY, 2022–2026 (2/2)

- Mincer–Zarnowitz: ipoteza $a = 0, b = 1$ nu este respinsă pentru log-HAR ($p = 0{,}555$)
  - R^2 este mai mare decît pentru GARCH
  - log-HAR are $b > 1$: prognozele mari sînt ușor sub mediile realizate (0,66 erori standard de 1, SE 0,25)
  - GARCH are $b < 1$: prognozele mari sînt ușor peste mediile realizate, tot nesemnificativ ($b = 0{,}70$, SE 0,24; testul MZ comun: $p = 0{,}355$)
- RV de ieri are un R^2 ridicat, dar $b = 0{,}54$
  - direcția este corectă, dar prognozele mari sînt mult peste mediile realizate
- Tipul de dovadă: comparație între proceduri (arată care metodă funcționează mai bine pe aceleași date, nu de ce)
  - pentru Bitcoin (Aprofundări), cîștigul log-HAR față de HAR pe niveluri nu este semnificativ ($p = 0{,}223$)

Prognozele în timp (1/2)



MFM_ch9_forecasts ▶ Interactiv

Prognozele în timp (2/2)

- Stînga: volatilitatea realizată SPY (cu noaptea) și prognozele log-HAR și GARCH, anualizate
- Dreapta: pierderea QLIKE cumulată a fiecărui model minus log-HAR (crescătoare = log-HAR mai bun)
- După un șoc ambele prognoze cresc, dar GARCH rămîne ridicat mai mult
  - persistența GARCH este aproape de 1, iar log-HAR revine mai repede
  - aprilie 2025: GARCH peste 40% pe an pînă la sfîrșitul lunii
  - volatilitatea realizată sub 20% de la jumătatea lunii
- Diferența cumulată de pierdere crește constant, cu salturi în episoadele agitate
  - cîștigul nu provine dintr-o singură zi
- Dintre cele trei modele comparate în grafic cu log-HAR, EWMA pierde cel mai mult
  - un $\lambda = 0{,}94$ fix reacționează prea mult la zile izolate

MFM_ch9_forecasts

Mulțimea de încredere a modelelor

Model	QLIKE	valoarea p MCS	Eliminat la pasul	În MCS (90%)
log-HAR	0,310	1,000	–	✓
SHAR	0,320	0,223	–	✓
HAR	0,326	0,013	3	
HARQ	0,344	0,095	4–5	
GARCH(1,1)-t	0,343	0,095	4–5	
EWMA	0,393	0,012	2	
RV de ieri	0,502	$<$ 0,001	1	

- Mulțimea de încredere a modelelor (MCS, model confidence set) Hansen, Lunde & Nason (2011)
  - pe cele cinci modele plus SHAR și HARQ
  - testăm egalitatea capacității predictive, eliminăm cel mai slab model, repetăm
  - statistica amplitudinii $T_R = \max_{i,j} |\bar{d}_{ij}| / \widehat{\mathrm{se}}(\bar{d}_{ij})$, cu $\bar{d}_{ij}$ diferența medie de pierdere dintre modelele i și j rămase în mulțime
  - distribuția lui T_R: bootstrap pe blocuri de 20 de zile, 10 000 de replicări, deci cea mai mică valoare p raportabilă este de ordinul 10^{-4}
  - ca la DM, bootstrap-ul cere o teoremă limită centrală pentru diferențele de pierdere QLIKE
  - SHAR și HARQ includ HAR: pentru aceste perechi, distribuția de referință este o aproximare convențională
  - mulțimea rămasă conține cel mai bun model cu probabilitate asimptotică $\geq 90\%$
- SPY, 1 180 de zile din 2022: mulțimea de 90% este {log-HAR, SHAR}
- O procedură de comparații multiple
  - multe teste DM pe perechi, fără controlul probabilității erorilor de tip I, ar exagera tăria dovezilor (Capitolul 8)

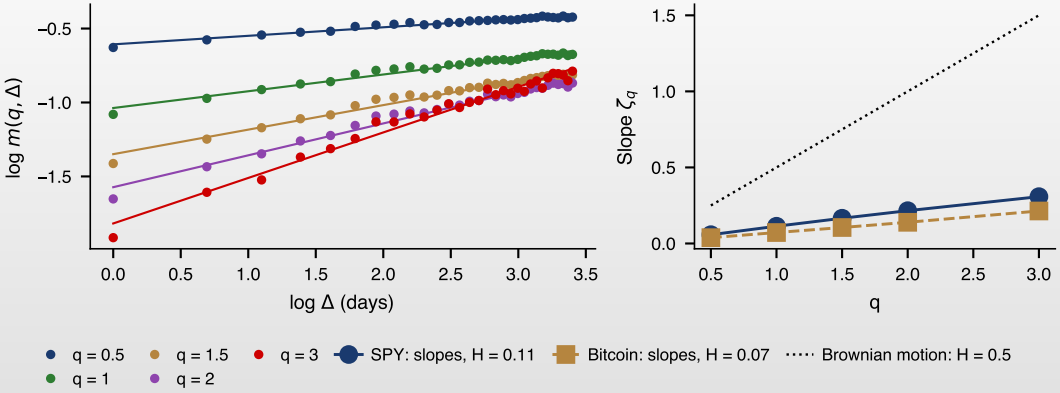
Este volatilitatea aspră?

- **Întrebarea**: sînt traiectoriile log-volatilității mai aspre decît cele ale unei mișcări browniene?
- Gatheral, Jaisson & Rosenbaum (2018): log-volatilitatea se comportă ca o **mișcare browniană fracționară**
  - exponentul Hurst este $H \approx 0{,}1$
  - $H = 0{,}5$: mișcarea browniană obișnuită
  - $H < 0{,}5$: incrementele sînt corelate negativ, iar traiectoriile sînt mai aspre
  - $H > 0{,}5$: traiectoriile sînt mai netede, cu tendință
- Modelele de volatilitate aspră reproduc forma zîmbetului volatilității implicite
  - pentru scadențe scurte (Capitolul 12)

Mișcarea browniană fracționară și legea de scalare

- Mișcarea browniană fracționară B_H, $0 < H < 1$: un proces gaussian cu $B_H(0) = 0$ și media 0
  - $\mathrm{Cov}(B_H(t), B_H(s)) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right)$
  - $H = 1/2$: covarianța este $\min(t, s)$, adică mișcarea browniană obișnuită
  - incrementele: $B_H(t+\Delta) - B_H(t) \sim N(0, \Delta^{2H})$ pentru orice t
- Modelul $\ln \sigma_t = \text{const} + \nu B_H(t)$
  - variația pe Δ zile are distribuția lui $\nu \Delta^H Z$, $Z \sim N(0,1)$
  - Legea de scalare: $m(q, \Delta) = \frac{1}{T}\sum_t |\ln \sigma_{t+\Delta} - \ln \sigma_t|^q$, cu $\sigma_t = \sqrt{\mathrm{RV}_t}$, estimează $\nu^q \, \mathbb{E}|Z|^q \, \Delta^{qH}$, deci panta lui $\ln m(q, \Delta)$ pe $\ln \Delta$ este $\zeta_q = qH$
- Estimare: pantele ζ_q pentru $q \in \{0{,}5; 1; 1{,}5; 2; 3\}$
  - apoi regresia lui ζ_q pe q prin origine: $\hat{H} = \sum_q q\,\zeta_q / \sum_q q^2$
- Verificare cu $q = 2$
  - pentru mișcarea browniană $m(2, \Delta) = \nu^2 \Delta$, deci dublarea lui Δ dublează m, iar $\zeta_2 = 1 = 2 \times 0{,}5$
  - cu $H = 0{,}1$, dublarea lui Δ înmulțește $m(2, \Delta)$ doar cu $2^{0{,}2} \approx 1{,}15$

Scalarea incrementelor log-volatilității (1/2)



MFM_ch9_rough ▶ Interactiv

Scalarea incrementelor log-volatilității (2/2)

- Stînga: $\ln m(q, \Delta)$ față de $\ln \Delta$ pentru SPY, $\Delta = 1, \ldots, 30$ de zile
- Dreapta: pantele ζ_q față de q
- Punctele pe scara log–log sînt apropiate de linii drepte, iar pantele sînt aproape liniare în q
  - compatibil cu legea de scalare, pe care graficul singur nu o poate stabili
  - $\hat{H} = 0{,}11$ pentru SPY (CI bootstrap pe blocuri de 95% $[0{,}08; 0{,}13]$) și 0,07 pentru Bitcoin: ambele mult sub 0,5
  - Pentru Bitcoin folosim doar perechile observate la fiecare decalaj calendaristic; dacă interpolăm zilele lipsă, obținem 0,08
- Problema: RV măsoară volatilitatea cu eroare
  - slide-urile următoare arată cum eroarea de măsurare coboară $\hat{H}$
- Tipul de dovadă: descriere

MFM_ch9_rough

Asprime sau zgomot de măsurare: problema de identificare

- Estimarea $H \approx 0{,}1$ provine din RV, un indicator zgomotos al volatilității latente σ_t
  - Cont & Das (2024): estimatorul de scalare aplicat volatilității realizate găsește asprime chiar cînd volatilitatea instantanee este netedă: un artefact de estimare
  - Bolko, Christensen, Pakkanen & Veliyev (2023) (un estimator GMM, metoda generalizată a momentelor, care modelează eroarea din RV) și Fukasawa, Takabatake & Westphal (2022) (un estimator consistent cînd frecvența de eșantionare crește) găsesc totuși $H < 0{,}5$
- Întrebarea formală: $\ln \mathrm{RV}_t^{1/2} = \ln \sigma_t + e_t$
  - $\ln \sigma_t$ este fracționar de indice H, iar e_t este eroarea de măsurare din RV
  - testăm H_0: $H = 1/2$ (volatilitate difuzivă clasică) față de H_1: $H < 1/2$
  - Problema de identificare: un H mic și un $\mathrm{Var}(e_t)$ mare aplatizează amîndouă $m(2, \Delta)$ la laguri mici (slide-ul următor)
- Asprimea pe termen scurt poate coexista cu memoria lungă Bennedsen, Lunde & Pakkanen (2022)
- Miza: panta zîmbetului la scadențe scurte și prognozele volatilității depind de H (Capitolul 12)

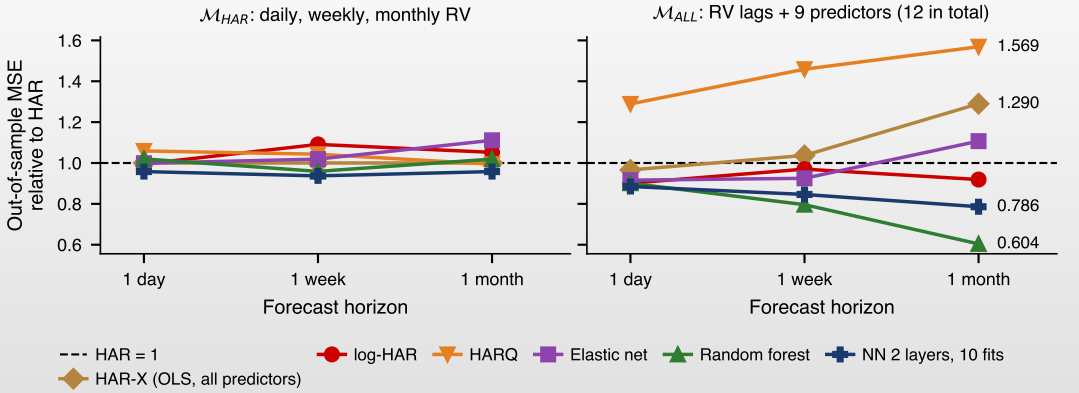
Cum aplatizează eroarea de măsurare legea de scalare

- Incrementul observat = incrementul latent + incrementul erorii
  - $(\ln \sigma_{t+\Delta} - \ln \sigma_t) + (e_{t+\Delta} - e_t)$
  - dacă e_t este i.i.d. și independent de σ, variațele se adună: $\mathrm{Var}(e_{t+\Delta} - e_t) = 2\,\mathrm{Var}(e_t)$ pentru orice $\Delta \geq 1$
  - de aici $m(2, \Delta) = c\,\Delta^{2H} + k$, $c = \nu^2$, $k = 2\,\mathrm{Var}(e_t)$: o constantă care nu crește cu Δ
- Exemplu: $H = 0{,}5$, $c = 1$, $k = 1$
  - $m(2,1) = 1 + 1 = 2$, $m(2,2) = 2 + 1 = 3$: panta pe scara log–log este $\ln(3/2)/\ln 2 = 0{,}58$, deci $\hat{H} = 0{,}58/2 = 0{,}29$ în loc de 0,5
  - la laguri mari constanta contează mai puțin: de la $\Delta = 20$ la 40, panta este $\ln(41/21)/\ln 2 = 0{,}97$, $\hat{H} = 0{,}48$
- Implicație testabilă: sub eroare i.i.d., $\hat{H}$ crește cînd renunțăm la lagurile mici
  - mini-studiul de caz din secțiunea AI pentru descoperire științifică face exact această verificare pe SPY

Studiu de caz: poate învățarea automată să întrecă HAR?

- Christensen, Siggaard & Veliyev (2023): oferă învățarea automată (ML) prognoze mai precise ale variantei realizate zilnice decît HAR?
  - metodele ML sînt folosite cu reglaje minime
  - dacă da, de ce?
  - motivul întrebării: HAR Corsi (2009) este liniar în trei medii ale RV, iar RV este asimetrică la dreapta și persistentă, deci o funcție neliniară sau alți predictori decît RV trecută pot îmbunătăți prognoza
- Date: 29 dintre cele 30 de acțiuni din Dow Jones Industrial Average (DJIA)
  - ianuarie 2001 – decembrie 2017, $T = 4\,257$ de zile
  - RV la 5 minute ($n = 78$)
- Mulțimi de informații: $\mathcal{M}_{\text{HAR}}$ și $\mathcal{M}_{\text{ALL}}$
  - $\mathcal{M}_{\text{HAR}}$: RV zilnică, săptămînală, lunară
  - $\mathcal{M}_{\text{ALL}}$ adaugă 9 predictori: volatilitatea implicită, VIX, volumul, momentum, indici macroeconomici și de incertitudine

Studiu de caz: cîștigurile pe orizonturi (1/2)



MFM_ch9_ml_vs_har ▶ Interactiv

Studiu de caz: cîștigurile pe orizonturi (2/2)

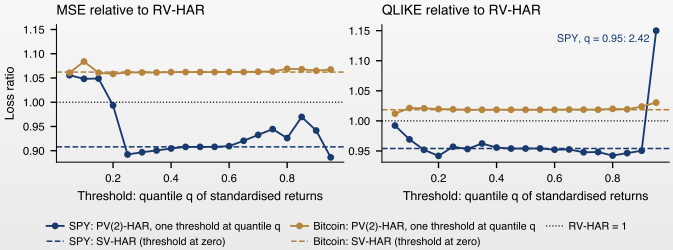
- Sursa: Christensen, Siggaard & Veliyev (2023)
  - MSE în afara eșantionului relativ la HAR, mediat pe 29 de acțiuni DJIA (setul de test: ultimele 20% din 2001–2017)
  - stînga: doar lagurile RV; dreapta: toți cei 12 predictori
  - HAR-X: OLS pe toți predictorii
- Cu doar lagurile RV, cele șase modele din grafic rămîn între 0,937 și 1,111 din MSE a HAR
  - pe toate cele trei orizonturi
  - trei medii ale RV lasă puțină informație suplimentară pentru un model flexibil
- Cu toți cei 12 predictori, cîștigul arborilor și al rețelelor crește cu orizontul
  - autorii îl atribuie persistenței mai mari a valorilor ajustate ML
  - random forest 0,901 $\to$ 0,796 $\to$ 0,604
  - rețeaua cu 2 straturi ascunse, media a 10 estimări (NN_2^{10}): 0,885 $\to$ 0,846 $\to$ 0,786
- Modelele liniare cu aceiași predictori se deteriorează
  - HAR-X de la 0,966 la 1,290
  - HARQ de la 1,289 la 1,569
  - Dintre modelele liniare, doar log-HAR rămîne sub 1 pe toate orizonturile (0,901; 0,970; 0,919)
- Tipul de dovadă: comparație între proceduri
  - medii ale rapoartelor MSE pe acțiuni, nu un test de semnificație al clasamentului (Aprofundări)
  - clasamentul depinde de împărțirea în eșantioane de antrenare și de test (Anexă)

MFM_ch9_ml_vs_har

Studiu de caz: este zero cel mai bun prag?

- Bollerslev, Medeiros, Patton & Quaedvlieg (2022), *Journal of Econometrics*: este semnul randamentului, adică pragul 0 al semivariațelor, cea mai bună împărțire a RV de ieri pentru prognoză?
  - SHAR Patton & Sheppard (2015) împarte RV de ieri în semivariația negativă și cea pozitivă
  - pragul 0 are o motivație economică, aversiunea față de pierderi, dar statistic este arbitrar
- **Variația parțială** (PV, partial variance): partea din RV adusă de randamentele dintre două praguri
  - $\mathrm{PV}_t^{(g)} = \sum_i r_{t,i}^2 \, 1\{c_{g-1} < r_{t,i} \leq c_g\}$, cu $c_0 = -\infty$ și $c_G = +\infty$, deci $\sum_{g=1}^{G} \mathrm{PV}_t^{(g)} = \mathrm{RV}_t$
  - $G = 2$ și $c_1 = 0$: semivariațele; $G = 1$: RV
- **PV(G)-HAR**: în HAR, RV de ieri este înlocuită de cele G variațe parțiale
  - componentele săptămînală și lunară rămîn RV; în notația lucrării, SV-HAR (cu semivariațe realizate) este SHAR, iar RV-HAR (cu variația realizată) este HAR
- Afirmațiile lucrării, pe SPY și 28 de acțiuni din DJIA, 1997–2013
  - cu un singur prag, pragul 0 este greu de întrecut
  - două praguri, unul în coada stîngă și unul în dreapta, îmbunătățesc semnificativ prognoza
  - pentru matrice de covariață mari, pragul 0 rămîne alegerea robustă

Studiu de caz: pierderea în funcție de prag, SPY și Bitcoin



- Pierderea medie în afara eșantionului a PV(2)-HAR împărțită la cea a RV-HAR, pentru pragul q de pe axa orizontală
  - q: cuantila randamentelor de 5 minute împărțite la $\sqrt{\mathrm{RV}_t}$; sub 1, mai bine decît RV-HAR
  - linia întreruptă: SV-HAR, cu pragul exact 0
- SPY, 1 180 zile (ian. 2022 – sept. 2026): raportul MSE stă între 0,892 și 0,909 pentru q de la 0,25 la 0,60, în jurul valorii SV-HAR (0,908)
- Bitcoin, 422 zile (mar. 2025 – sept. 2026): orice prag dă un MSE între 1,058 și 1,084 din cel al RV-HAR

MFM_ch9_partial_variances ▶ Interactiv

Studiu de caz: metoda și replicarea pe datele cursului

- Pragul ca cuantilă: $c_{t,g} = \sqrt{\mathrm{RV}_t}\ \hat{Q}(z;\, q_g)$, unde $z = r_{t,i}/\sqrt{\mathrm{RV}_t}$
  - același q urmărește nivelul volatilității; $q = 0{,}5$ dă practic semivariațele
- Lucrarea: SPY, 4 202 zile, prognoze pe ferestre mobile de 1 000 de zile
  - PV(3)-HAR: cele două praguri, alese prin validare încrucișată (CV, cross-validation) în 5 blocuri, pe grila $q = 0{,}05; 0{,}10; \ldots; 0{,}95$
  - pierderi MSE și QLIKE; teste Diebold–Mariano (DM) unilaterale
- Replicarea: aceeași construcție pe barele de 5 minute ale cursului
  - SPY: ținta este RV intraday (09:30–16:00), ca în lucrare; slide-urile HAR ale capitolului adaugă randamentul peste noapte
  - Bitcoin: RV pe 24 de ore UTC, componente pe 7 și 30 de zile calendaristice
  - fereastră extinsă, reestimare zilnică, cuantile doar din zilele dinaintea prognozei
  - PV(3)-HAR: perechea de cuantile aleasă prin CV (MSE) la fiecare 21 de zile
  - o prognoză sub cea mai mică RV din fereastră este înlocuită cu aceasta (filtru de plauzibilitate, *insanity filter*)
- Diferențe: un eșantion de aproape 3 ori mai scurt, fără acțiunile individuale din DJIA, teste DM bilaterale

Studiu de caz: pragul zero și cîștigul concentrat în puține zile

Pierderea relativă la RV-HAR (MSE / QLIKE)	Lucrarea: SPY 1997–2013	SPY 2022–2026	Bitcoin 2025–2026
SV-HAR (pragul 0)	0,967 / 0,981	0,908 / 0,954	1,062 / 1,019
PV(3)-HAR (două praguri, CV)	0,950 / 0,971	0,815 / 1,013	1,145 / 1,061

- SPY: SV-HAR are QLIKE mai mic decît RV-HAR, DM $t = -1{,}79$ ($p = 0{,}073$); lucrarea (test unilateral): $p = 0{,}005$
  - pe tot eșantionul, cu ținta RV intraday: $\beta^- = 1{,}44$ (SE 0,63) și $\beta^+ = -0{,}34$ (SE 0,39); în Anexă, cu ținta varianța totală, alte valori
- PV(3)-HAR nu bate SV-HAR la QLIKE: DM $t = +1{,}66$ ($p = 0{,}10$)
  - în lucrare, PV(3)-HAR are pierderi mai mici decît SV-HAR: $p = 0{,}011$ (MSE) și 0,072 (QLIKE)
  - perechea aleasă se schimbă: cea mai frecventă, (0,05; 0,95), în 15 din 57 selecții; cuantilele medii alese: 0,23 și 0,74 (lucrarea: 0,14 și 0,70)
- Cele 10 zile cu cea mai mare RV aduc 89% din eroarea pătratică a RV-HAR pe SPY
  - fără ele, SV-HAR are MSE 1,240 și QLIKE 0,937 din cele ale RV-HAR: cîștigul MSE dispare

Studiu de caz: ce susține rezultatul

- Tipul de dovadă: comparație între proceduri de prognoză pe SPY
  - SV-HAR: cîștigul la QLIKE se reproduce, nesemnificativ la 5%
  - PV(3)-HAR: cîștigul la QLIKE nu se reproduce
  - PV(3)-HAR: MSE scade cu 18,5%, dar avantajul este concentrat în zilele extreme și nu este semnificativ în aproximația DM convențională ($t = -0{,}97$, $p = 0{,}33$)
- Inferența pentru modele imbricate: RV-HAR este SV-HAR sau PV(3)-HAR cu coeficienți zilnici egali
  - valorile p DM de mai sus (fereastră extinsă) sînt aproximația convențională
  - fereastră mobilă de lungime fixă, 287 de zile (Giacomini–White, GW): față de RV-HAR, MSE are $p = 0{,}34$ pentru SV-HAR și $p = 0{,}35$ pentru PV(3)-HAR
  - cu aceeași fereastră, QLIKE este mai mare decît la RV-HAR: raportul este 1,70 pentru SV-HAR și 3,23 pentru PV(3)-HAR
  - testele cer o teoremă limită centrală pentru diferențele de pierdere; la QLIKE, momentele lor depind și de prognozele de la numitor
- Limite: prognoze pe 4,7 ani pentru SPY și 1,5 ani pentru Bitcoin; MSE dominat de cîteva zile extreme

Bucla descoperirii aplicată asprimii volatilității (1/2)

- Întrebarea: este $\hat{H} \approx 0{,}1$ o proprietate a lui σ_t sau a erorii e_t din RV?
- O trecere prin buclă cu un asistent AI (prompturi reutilizabile în Anexă)
  - Literatura: tabelăm estimările publicate ale lui H (estimator, frecvență, tratamentul erorii din RV)
  - Verificăm fiecare DOI (digital object identifier) pe Crossref
  - Ipoteze concurente, fiecare cu forma ei pentru $m(2, \Delta)$: (i) log-volatilitate fracționară cu $H = 0{,}1$, (ii) $H = 0{,}5$ plus eroare i.i.d., (iii) un model cu doi factori, unul cu revenire rapidă la medie
  - Replicăm întîi $\hat{H} = 0{,}11$ din capitol, apoi extindem analiza
  - Fixăm dinainte ferestrele de laguri, măsurile și ajustarea cu termen constant de zgomot k

Bucla descoperirii aplicată asprimii volatilității (2/2)

- Verificări care revin cercetătorului
  - Eroarea indicatorului: simulăm $H = 0{,}5$ plus eroare de aceeași mărime; procedura nu trebuie să producă $H \approx 0{,}1$
  - Incrementele suprapuse invalidează erorile standard OLS ale lui ζ_q: folosim bootstrap pe blocuri
  - Raportăm toate ferestrele încercate
  - Asistenții citează „$H \approx 0{,}1$ pentru toate activele”: verificăm fiecare cifră în tabelul lucrării citate

Mini-studiu de caz: $\hat{H}$ ca artefact al erorii i.i.d. (1/2)

- O propunere tipică a asistentului: $\hat{H} \approx 0{,}1$ provine din eroarea de măsurare i.i.d. din RV
  - implicația testabilă: $\hat{H}$ **crește** cînd renunțăm la lagurile mici

	$\Delta \in [1, 30]$	$[5, 30]$	$[10, 50]$	Ajustare cu k, $[1, 50]$
$H = 0{,}5$ simulat plus eroare	0,18	0,24	0,26	0,39
SPY, RV totală	0,106	0,089	0,077	0,096
SPY, RV intraday	0,130	0,094	0,078	0,114

- Prima coloană replică valoarea din capitol
- Rîndul simulat arată mecanismul: $\hat{H}$ crește
  - ajustarea cu k recuperează mare parte din distanța pînă la 0,5
  - o singură traiectorie, nu un studiu de putere

Mini-studiu de caz: $\hat{H}$ ca artefact al erorii i.i.d. (2/2)

- SPY: $\hat{H}$ **scade** cu fereastra, iar k estimat este 0
  - ajustarea cu k dă $\hat{H} = 0{,}096$, cu CI bootstrap pe blocuri de 95% $[0{,}058; 0{,}268]$
- Interpretare: un diagnostic ilustrativ, incompatibil cu explicația prin eroare i.i.d.
  - o respingere formală cere un test calibrat prin simulare
  - mecanismul non-i.i.d. din Cont & Das (2024) nu este testat de acest design
- Tipul de dovadă: descriere

MFM_ch9_ai_discovery_case

Punct de plecare pentru proiect

- Întrebarea: rămîne volatilitatea SPY și Bitcoin aspră și cînd estimatorul modelează eroarea de măsurare din RV?
- Replicați mai întîi (Seminarul 9, B9)
  - $\hat{H} = 0{,}11$ (SPY) și $0{,}07$ (Bitcoin) din legea de scalare
  - $\hat{H} = 0{,}18$ pentru $H = 0{,}5$ simulat plus eroare
- Livrabile
  - O pre-înregistrare (ipoteze, estimatori, ferestre de laguri, teste, designul simulării) datată înainte de prima rulare
  - Estimatorul GMM din Bolko, Christensen, Pakkanen & Veliyev (2023) pe seriile RV ale cursului, cu mărimea și puterea Monte Carlo la $H \in \{0{,}1; 0{,}5\}$
  - Codul, un raport de două pagini cu toate specificațiile încercate și o critică, în stilul unui referent, a propriului rezultat
- Consemnați: github.com/danpele/MFM/tree/main/project
  - fiecare prompt care a influențat o decizie în `AI_USE.md`
  - fiecare eroare a asistentului în `AI_ERRORS.md`

Alegerea instrumentului

Situația	**Instrumentul sugerat**
Activ lichid, bare de 1–5 minute	RV la 5 minute, subeșantionată
Date tick cu tick sau o acțiune nelichidă (Bursa de Valori București, BVB)	Pre-averaging sau nucleu realizat flat-top ($n^{-1/4}$; Parzen $n^{-1/5}$; Anexă: ratele de convergență); verificați graficul de semnătură
Riscul zilnic al unui activ tranzacționat la bursă	Varianța totală: RV plus randamentul peste noapte la pătrat
Separarea salturilor din știri de difuzie	BV sau MedRV (varianța realizată bazată pe mediane; Anexă: estimatori robuști la salturi), testul raportului cu control FDR; Lee–Mykland pentru momentul saltului
Prognoza varianței zilnice sau săptămînale	log-HAR (plus salturi, weekenduri, RQ); reper GARCH
Compararea prognozelor	QLIKE și MSE pe varianță, Diebold–Mariano, Mincer–Zarnowitz, mulțimea de încredere a modelelor

- Rezultatul cel mai solid: în afara eșantionului, log-HAR are o QLIKE mai mică decît GARCH(1,1)-t pe ambele piețe
  - comparațiile după MSE nu sînt concludente
- Regula de 5 minute este justificată doar dacă graficul de semnătură este plat de la 5 minute în sus
- Numărul de salturi depinde de controlul FDR și de corecția de periodicitate

Recapitulare și capitolul următor: Capitolul 10

Întrebări de verificare

- De ce R^2 pentru log RV nu este comparabil cu R^2 pentru RV pe niveluri?
  - slide-ul *Interpretarea estimărilor HAR*
- De ce este QLIKE robustă la un indicator RV zgomotos, iar MSE pe volatilitate nu?
  - slide-ul *Evaluarea față de o țintă zgomotoasă*
- De ce MCS indică o mulțime de modele, nu un câștigător?
  - slide-ul *Mulțimea de încredere a modelelor*
- Cum poate eroarea de măsurare din RV să facă volatilitatea să pară aspră?
  - slide-ul *Cum aplatizează eroarea de măsurare legea de scalare*

Capitolul 10: microstructura pieței

- De unde provine zgomotul ε din acest capitol
  - registrul de ordine limită, spread-ul bid–ask, pasul de cotare
- Lichiditatea, impactul asupra pieței și tranzacționarea de înaltă frecvență, pe burse de acțiuni și cripto

Decizia practică

- **Ce aleg**
  - pentru riscul zilnic, varianța totală: RV pe 5 minute plus randamentul peste noapte la pătrat (noaptea aduce 38% din varianța zilnică SPY)
  - pentru prognoza variației de mîine, log-HAR: QLIKE 0,310 față de 0,343 pentru GARCH(1,1)-t pe SPY; rămîne în mulțimea de 90%, alături de SHAR
  - pentru zilele cu salt, testul raportului cu corecția de periodicitate și control FDR: 17 zile din 1 489 pentru SPY, nu 236
- **În ce condiții**
  - un activ lichid, cu graficul de semnătură plat de la 5 minute în sus; datele tick cu tick sau o acțiune nelichidă cer pre-averaging sau un nucleu realizat
  - clasamentul este valabil pentru aceste proceduri implementate și pentru pierderea QLIKE; comparațiile după MSE nu sînt concludente
  - valorile p DM și MCS cer o teoremă limită centrală pentru diferențele de pierdere QLIKE; pentru perechile imbricate (SHAR, HARQ față de HAR) sînt o aproximare convențională
- **Ce rezultat m-ar face să renunț**
  - la RV pe 5 minute: un grafic de semnătură care crește cînd intervalul scade spre 5 minute, amprenta zgomotului
  - la log-HAR: pe un eșantion nou, un test DM pe QLIKE care nu o mai favorizează față de GARCH (aici $t = -2{,}30$, $p = 0{,}021$) sau ieșirea ei din mulțimea de 90%

Datele: seriile intraday SPY și Bitcoin

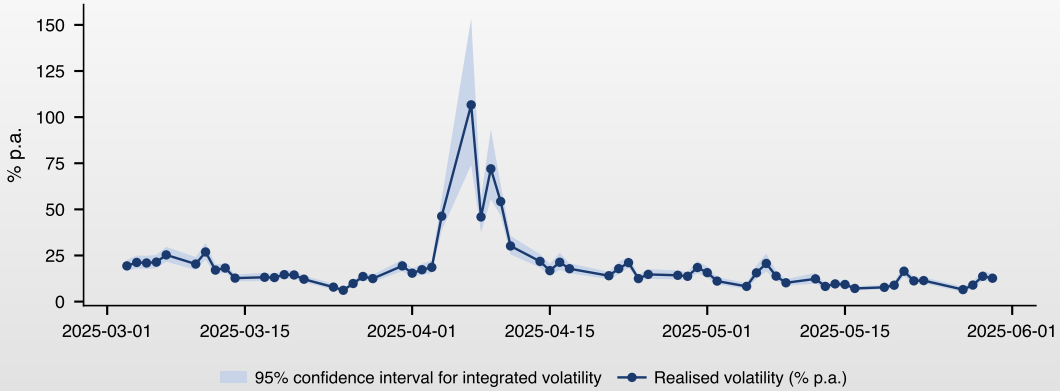
Seria	Frecvența	Perioada	Zile
SPY (ETF)	bare de 5 minute, 09:30–16:00 ora New York	2020–2026	1 489
Bitcoin	bare de 5 minute, zile de 24 de ore (UTC)	2024–2026	659
SPY, Bitcoin	prețuri zilnice de închidere	2000/2014–2026	pentru GARCH
VIX	valoare zilnică la închidere	2020–2026	volatilitate implicită

- SPY: 78 de randamente într-o zi completă și 42 (09:30–13:00) în cele 12 zile cu program redus
  - bara etichetată 13:00 începe după închidere: nu este un interval; primul ei preț este prețul de la 13:00 doar cînd bara de 12:55 lipsește
  - barele din afara grilei de 5 minute sînt eliminate
  - o bară lipsă din timpul zilei preia ultimul preț disponibil: 30 bare în 16 zile
- Bitcoin: 288 de randamente pe zi
  - o bară lipsă preia ultimul preț, deci intervalul ei se contopește cu următorul: 1 347 de bare în 245 din 659 de zile
  - ziua începe la prețul de la 00:00 (deschiderea primei bare sau închiderea barei de 23:55 din ziua precedentă) și se termină la ultimul preț observat
  - fără zilele incomplete, volatilitatea anualizată devine 12,97% pentru SPY (în loc de 12,93%) și 46,17% pentru Bitcoin (în loc de 45,08%)
  - zilele Bitcoin cu mai puțin de 95% din bare sînt eliminate: 88 din 747 de zile calendaristice, mai ales un gol de date în octombrie 2025

Cum se adună varianța de zi și cea de noapte

- SPY, variațe anualizate (%^{2} pe an)
  - RV intraday: 12,9%, varianța 167
  - varianța totală $\mathrm{RV}_t + r^2_{\text{noapte},t}$: 16,4%, varianța 270
  - partea de peste noapte: $270 - 167 = 103$, adică $\sqrt{103} \approx 10{,}1\%$; ponderea $103/270 = 38\%$
- De ce nu $13{,}6^2 + 10{,}1^2$?
  - valoarea 13,6% de la deschidere la închidere ridică la pătrat randamentul întregii sesiuni, nu fiecare randament de 5 minute
  - $(\sum_i r_{t,i})^2 = \mathrm{RV}_t + 2\sum_{i<j} r_{t,i}\, r_{t,j}$: produsele încrucișate sînt nule doar în medie teoretică; în acest eșantion adaugă aproximativ 17 ($13{,}6^2 \approx 184$)
  - volatilitatea de la închidere la închidere (16,6%) conține și termenul încrucișat $2\, r_{\text{noapte},t}\, r_{\text{deschidere-închidere},t}$

Intervale de încredere pentru volatilitatea zilnică (1/2)



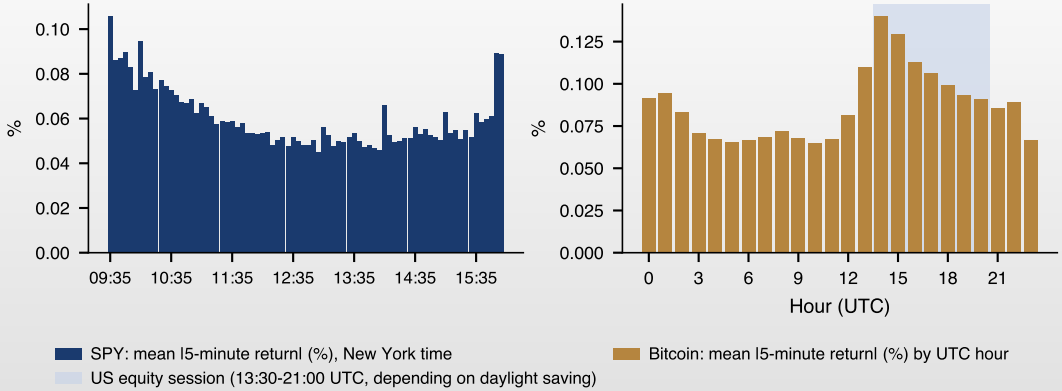
MFM_ch9_rv_ci

Intervale de încredere pentru volatilitatea zilnică (2/2)

- SPY, martie–mai 2025
  - banda: intervalul de încredere de 95% pentru $\sqrt{252\,\mathrm{IV}_t}$ din cvarticitatea realizată
- Banda pentru volatilitate ia rădăcina pătrată a capetelor intervalului pentru varianță
  - $\sqrt{\mathrm{RV}_t}\exp\left(\pm 0{,}98\,\mathrm{se}(\mathrm{RV}_t)/\mathrm{RV}_t\right)$: semi-lățimea pe scara log se înjumătățește
- SPY: volatilitatea unei zile este cunoscută cu aproximativ $\pm 20\%$
  - se/RV median $= 0{,}19$, deci factorii pentru volatilitate sînt $[0{,}83; 1{,}20]$
  - lățimea relativă mediană: 37%
- 9 aprilie 2025: volatilitatea a fost 72%, cu CI de 95% $[55\%, 93\%]$
  - ziua în care a fost anunțată amînarea tarifelor
  - Intervalul se lărgește cînd cîteva randamente domină suma: $\sum r^4$ crește mai repede decît $(\sum r^2)^2$
- Banda presupune un preț continuu și zgomot neglijabil
  - într-o zi cu salt, $\sum r^4$ păstrează termenul κ^4
  - deci banda nu are acoperirea nominală nici pentru variația totală, nici pentru IV_t (secțiunea Salturi)
- RV rămîne totuși mult mai precisă decît r_t^2
  - eroarea relativă a lui r_t^2 este de ordinul $\sqrt{2} \approx 141\%$ sub distribuția Normală

MFM_ch9_rv_ci

Sezonalitatea intraday (1/2)



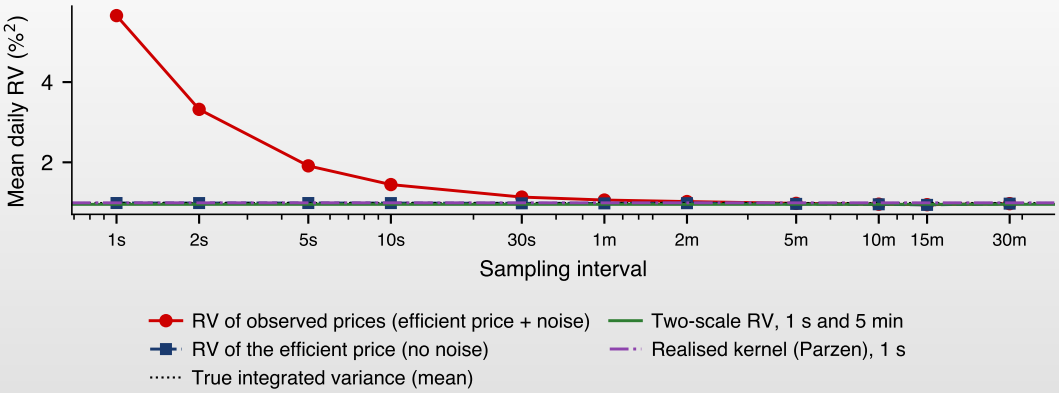
MFM_ch9_seasonality

Sezonalitatea intraday (2/2)

- Stînga: SPY, media $|r_{t,i}|$ pe intervale de 5 minute ale zilei
  - doar zile complete
  - etichetele marchează sfîrșitul intervalului, ora New York
- Dreapta: Bitcoin pe ore UTC
- SPY: o **formă de U** Andersen & Bollerslev (1997)
  - 0,106% în primele 5 minute, 0,050% la prînz, 0,089% în ultimele 5 minute
  - Deschiderea încorporează știrile de peste noapte; închiderea concentrează reechilibrările indicilor și ale fondurilor
- Bitcoin se tranzacționează non-stop, dar este cel mai volatil la ora 14:00 UTC
  - 14:00 UTC: în jurul deschiderii bursei din SUA (09:30 la New York: 13:30 UTC vara, 14:30 UTC iarna)
  - cel mai calm la 10:00 UTC, cînd volatilitatea este de 2,2 ori mai mică
- RV însumează pe toată ziua, deci sezonalitatea nu o deplasează
  - în schimb, deplasează variația bipower (BV, bipower variation, definită în secțiunea Salturi) și face ca testul de salturi să respingă prea des
  - remediul, împărțirea randamentelor la un factor de periodicitate f_i Boudt, Croux & Laurent (2011), este aplicat testului de salturi în secțiunea Salturi (Anexă)

MFM_ch9_seasonality

Graficul de semnătură: simulare (1/2)



MFM_ch9_noise_simulation

Graficul de semnătură: simulare (2/2)

- Un grafic de semnătură arată RV medie în funcție de intervalul de eșantionare
  - aici: 100 de zile simulate, un preț pe secundă
  - prețul eficient are volatilitate zilnică aleatoare, plus zgomot i.i.d., $\omega = 0{,}01\%$
- IV medie adevărată este 0,98, iar RV calculată din prețurile observate este
  - 5,66 la 1 secundă, 1,06 la 1 minut și 0,98 la 5 minute
  - Deplasarea corespunde lui $2n\omega^2$ (slide-ul *De ce nu folosim fiecare tranzacție?*)
- Fără zgomot, RV este constantă în raport cu intervalul de eșantionare
  - creșterea la frecvențe mari este amprenta zgomotului
- Estimatori robuști la zgomot pe date de 1 secundă: RV pe două scări 0,95, nucleul realizat 0,99
  - abaterea standard a erorii pe zile: RV pe 5 minute 0,187, pe două scări 0,150, nucleu 0,045
  - folosirea tuturor datelor aduce cîștig doar cu o corecție de zgomot

MFM_ch9_noise_simulation

Subeșantionare și RV pe două scări

- **Subeșantionarea**: calculăm RV pe K grile decalate și facem media celor K valori
  - decalajele sînt $0, 1, \ldots, K-1$; $\mathrm{RV}_t^{(k)}$ este RV pe grila cu decalajul k
  - $\mathrm{RV}_t^{\mathsf{avg}} = \frac{1}{K}\sum_{k=0}^{K-1} \mathrm{RV}_t^{(k)}$: aceeași deplasare ca o singură grilă rară, variață mai mică
- **RV pe două scări** (TSRV) Zhang, Mykland & Aït-Sahalia (2005): eliminăm deplasarea rămasă cu RV din toate datele
  - $\mathrm{TSRV}_t = \mathrm{RV}_t^{\mathsf{avg}} - \frac{\bar{n}}{n}\mathrm{RV}_t^{\mathsf{all}}$, $\bar{n} = K^{-1}\sum_k n_k$, n_k: randamentele grilei k; $\mathrm{RV}_t^{\mathsf{all}}$: RV pe toate cele n randamente, care estimează termenul de zgomot $2n\omega^2$
  - grilele standard (grila k folosește prețurile $k, k+K, \ldots$): $\bar{n} = (n-K+1)/K$, de exemplu $73/6 = 12{,}17$ pentru $n = 78$, $K = 6$
  - dacă adăugăm fiecărei grile primul și ultimul preț al zilei, $\bar{n} = 83/6 = 13{,}83$, dar zgomotul comun de la capete nu mai dispare: o aproximare pentru eșantion finit, nu TSRV consistent
  - În eșantioane mici, TSRV se împarte la $1 - \bar{n}/n$

Nuclee realizate



Neil Shephard, 2004

Foto: Renate Schmid (2004); CC BY-SA 2.0 de; Wikimedia Commons

- **Nucleul realizat** Barndorff-Nielsen, Hansen, Lunde & Shephard (2008): adăugăm autocovariațele ponderate ale randamentelor
  - $\mathrm{RK}_t = \gamma_0 + 2\sum_{h=1}^{H} k\left(\frac{h}{H+1}\right)\gamma_h$, $\gamma_h = \sum_i r_{t,i}\, r_{t,i-h}$
  - Funcția de ponderare Parzen k: 1 în 0, descrește continuu pînă la 0 în 1; păstrează RK (estimatorul realised kernel al variației, cu nucleu) pozitiv
- Zgomotul face γ_1 negativ
  - includerea acestui termen în estimatorul cu nucleu corectează deplasarea
- Lățimea de bandă H (aici nu exponentul Hurst) crește cu raportul zgomot/semnal
  - $H \approx 3{,}5\,\xi^{4/5} n^{3/5}$, cu $\xi^2 = \omega^2/\sqrt{\mathrm{IQ}}$ raportul zgomot/semnal
  - în codul cursului (pilot euristic): $\hat{\omega}^2 = \mathrm{RV}/(2n)$, iar RV pe 20 de minute ține locul lui $\sqrt{\mathrm{IQ}}$
- Jittering: primul și ultimul preț sînt înlocuite cu mediile primelor și ultimelor două prețuri, astfel încît zgomotul de la capete nu rămîne în RK

Regula Benjamini–Hochberg pe cinci teste

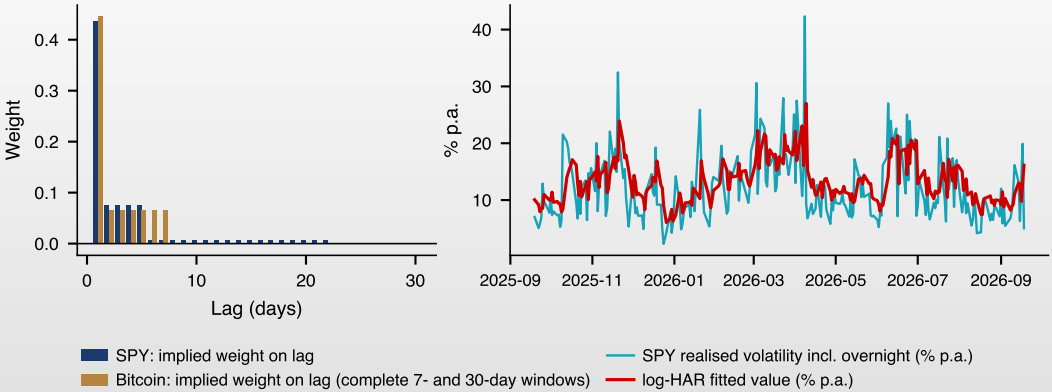
- Cinci zile, valori p ordonate $0{,}003; 0{,}012; 0{,}024; 0{,}045; 0{,}60$
  - cu $q = 0{,}05$ și $N = 5$, pragurile kq/N sînt $0{,}01; 0{,}02; 0{,}03; 0{,}04; 0{,}05$
- Primele trei trec, al patrulea nu: $k = 3$, respingem trei zile
  - $0{,}024 \leq 0{,}03$, dar $0{,}045 > 0{,}04$
- Pe aceleași date, numărul de respingeri
  - fără corecție ($p \leq 0{,}05$): 4
  - BH: 3
  - Bonferroni ($p \leq 0{,}05/5 = 0{,}01$): 1
- Procedura este step-up
  - cu valorile p $0{,}003; 0{,}012; 0{,}032; 0{,}038; 0{,}60$, al treilea nu trece de propriul prag ($0{,}032 > 0{,}03$)
  - dar al patrulea trece ($0{,}038 \leq 0{,}04$), deci $k = 4$ și respingem patru zile

Volatilitatea implicită ca prognoză a variației

- Ca prognoză: regresăm $\frac{252}{21}\mathrm{RV}_{t,t+21}$ pe VIX_t^2
  - $\frac{252}{21}\mathrm{RV}_{t,t+21}$: varianța totală, $\%^2$ anualizat
  - VIX_t^2: $\%^2$ anualizat; aproximări: 30 de zile calendaristice $\approx$ 21 de zile de tranzacționare, opțiuni pe indicele S&P 500 (SPX, Cboe) față de SPY
  - panta 0,56 $<$ 1, eroarea standard (SE, standard error) 0,06: Newey–West cu 20 de laguri, deoarece țintele pe 21 de zile se suprapun; 0,05 cu 7 laguri, 0,06 cu 42
  - panta sub 1 nu identifică o reacție excesivă a pieței și nici mecanismul primei
  - $R^2 = 0{,}23$: VIX este informativ
  - nivelul este o întrebare separată: diferența medie ex post este de 4,36 puncte procentuale în volatilitate și 138,2 în varianță ($\%^2$ anualizat)
  - niciuna nu identifică prima condiționată: ambele adaugă eroarea de prognoză $\mathbb{E}_t^{\mathbb{P}}[Q_t] - Q_t$
  - aceeași distincție între pantă și nivel revine în testul Mincer–Zarnowitz

MFM_ch9_vix

Ponderile implicite pe laguri și ajustarea în eșantion (1/2)



MFM_ch9_har

Ponderile implicite pe laguri și ajustarea în eșantion (2/2)

- Stînga: ponderile log-HAR pe fiecare zi trecută
  - Bitcoin: ferestre complete de 7 și 30 de zile
- Dreapta: SPY, ultimele 250 de zile, volatilitatea realizată și valorile ajustate log-HAR
- SPY, ponderile pe laguri
  - lagul 1: $\hat{\beta}_d + \hat{\beta}_w/5 + \hat{\beta}_m/22 = 0{,}362 + 0{,}067 + 0{,}008 = 0{,}437$
  - lagurile 2–5: cîte $0{,}067 + 0{,}008 = 0{,}075$
  - lagurile 6–22: cîte $0{,}008$
  - Trei coeficienți imită scăderea lentă a funcției de autocorelație
- Bitcoin: ziua de ieri 0,446, zilele 2–7 aproximativ 0,066 fiecare, aproape nimic după o săptămînă
- Valorile ajustate în eșantion urmăresc nivelul volatilității, nu zgomotul de la o zi la alta
- HAR este un AR(22) cu 19 restricții: $\phi_2 = \ldots = \phi_5$, $\phi_6 = \ldots = \phi_{22}$
  - testul Wald cu covarianță Newey–West pe SPY dă $W = 18{,}2$ și $p = 0{,}51$
  - deci restricțiile nu sînt respinse (derivarea în Anexă)

MFM_ch9_har

Extensii ale HAR

- **HAR-CJ** (*HAR model with Continuous and Jump components*, model HAR cu componentă continuă și componentă de salt) Andersen, Bollerslev & Diebold (2007): părțile continuă (C) și de salt (J) ca regresori separați
  - $\ln \mathrm{RV}_{t+1} = \beta_0 + \beta_{Cd} \ln C_t + \beta_{Cw} \ln C_t^{(w)} + \beta_{Cm} \ln C_t^{(m)}$
  - $\quad + \beta_{Jd} \ln(1 + J_t) + \beta_{Jw} \ln(1 + J_t^{(w)}) + \beta_{Jm} \ln(1 + J_t^{(m)}) + \varepsilon_{t+1}$
  - $C_t^{(w)}, J_t^{(w)}$ ($C_t^{(m)}, J_t^{(m)}$): medii aritmetice pe ultimele 5 (22) zile, ca $\mathrm{RV}_t^{(w)}$; logaritmul se aplică după mediere
- Variabile indicator pentru weekend sau sărbători
- Efectul de levier: randamentele negative cresc RV viitoare
- Volatilitatea implicită ca regresor suplimentar
- Pe mai mulți pași: regresăm direct $\sum_{j=1}^{h} \mathrm{RV}_{t+j}$ pe aceiași regresori

Rezultate: Bitcoin, 2025–2026 (1/2)

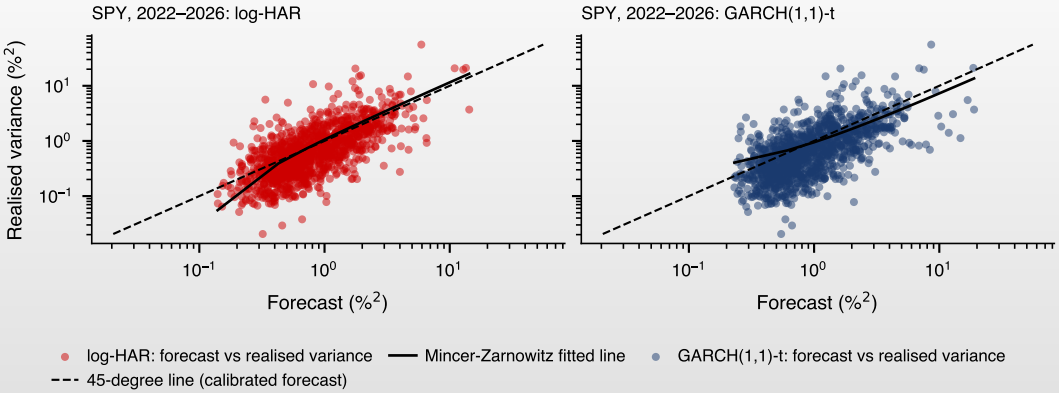
Model	QLIKE	$\mathrm{DM}_{\mathrm{QLIKE}}$ t (p)	MSE ($\%^4$)	$\mathrm{DM}_{\mathrm{MSE}}$ p	MZ b	MZ R^2
log-HAR	0,282	–	26,6	–	0,82	0,38
HAR	0,408	−1,22 (0,223)	27,6	0,431	0,89	0,34
GARCH(1,1)-t	0,360	−4,32 (< 0,001)	42,2	0,055	0,55	0,23
EWMA	0,346	−2,96 (0,003)	36,3	0,033	0,64	0,18
RV de ieri	0,466	−5,38 (< 0,001)	35,0	0,061	0,58	0,33

- 422 de prognoze pentru ziua următoare ale varianței realizate pe 24 de ore
  - aceleași modele și teste ca pentru SPY
- log-HAR are din nou cea mai mică QLIKE: cu 22% sub GARCH(1,1)-t
  - față de HAR pe niveluri diferența nu este semnificativă ($p = 0{,}223$)
  - GARCH pe date zilnice nu este calibrat: MZ $b = 0{,}55$ ($p < 0{,}001$): prognozele mari sînt urmate de medii realizate mai mici

Rezultate: Bitcoin, 2025–2026 (2/2)

- HAR pe niveluri este mai slab decît GARCH după QLIKE (0,408 față de 0,360)
  - după weekenduri calme prognozează o variață prea mică
  - QLIKE penalizează subestimarea
- După MSE, doar EWMA este semnificativ mai slab decît log-HAR la 5% ($p = 0{,}033$)
  - GARCH: $p = 0{,}055$
- MSE pentru Bitcoin și pentru SPY nu sînt comparabile
  - MSE se măsoară în $(\%^2)^2 = \%^4$ și crește cu pătratul nivelului variației
  - QLIKE depinde doar de RV_t / F_t și este comparabilă între piețe

Regresii Mincer–Zarnowitz (1/2)



MFM_ch9_forecasts

Regresii Mincer–Zarnowitz (2/2)

- SPY, 2022–2026: varianța realizată față de prognoză (scări log)
  - linie întreruptă: dreapta de 45 de grade
  - negru: regresia estimată pe niveluri
- Prognozele log-HAR acoperă un interval mai larg
  - separă mai bine zilele calme de cele agitate
- În zilele calme GARCH supraestimează varianța
  - prognozele GARCH coboară rar sub $0{,}26\%^2$ pe zi (1% din zile)
  - log-HAR coboară pînă la $0{,}14\%^2$
- Punctele sînt puternic dispersate în ambele grafice
  - varianța unei zile este greu de prognozat
  - RV însăși este măsurată cu eroare

MFM_ch9_forecasts

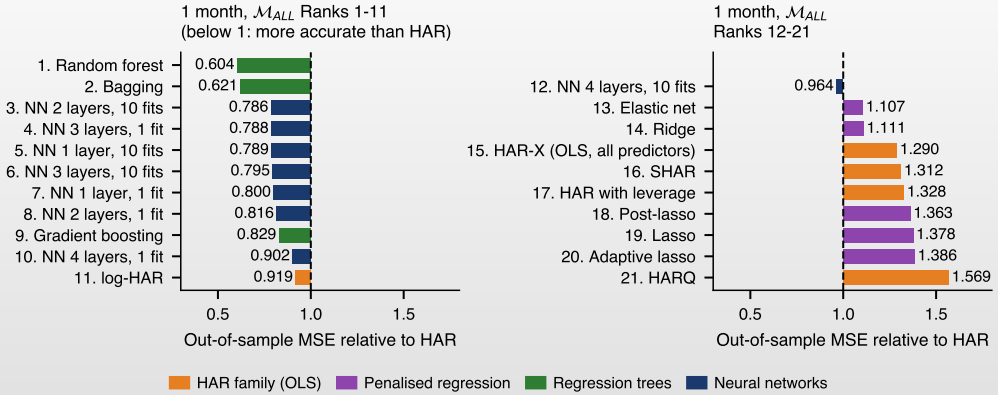
Interpretarea mulțimii de încredere a modelelor

- O valoare p MCS este un maxim cumulat
  - dacă un model este eliminat la pasul k, valoarea p MCS a modelului este $\max_{j \leq k} p_j$
  - p_j este valoarea p a testului de egalitate de la pasul j
  - de aceea valorile p MCS cresc în ordinea eliminării, pe care tabelul o citește din ele
  - valorile egale (GARCH, HARQ) înseamnă că valoarea p proprie al pasului ulterior nu a fost mai mare, deci ordinea lor relativă nu reiese din tabel
  - un model rămîne în mulțimea de 90% dacă valoarea p MCS a modelului depășește 0,10
- De ce este HAR (QLIKE 0,326) eliminat înaintea GARCH (0,343)?
  - fiecare pas elimină modelul cu cea mai mare diferență de pierdere **standardizată** față de ceilalți supraviețuitori
  - nu modelul cu cea mai mare pierdere medie
  - SHAR diferă de HAR doar prin termenul zilnic, deci pierderile lor se mișcă împreună, iar diferența este foarte stabilă: DM SHAR față de HAR $t = -3{,}39$, $p < 0{,}001$
  - GARCH pierde mai mult în medie, dar diferențele lui sînt mai zgomotoase: față de log-HAR $t = -2{,}30$, mai mic în valoare absolută decît diferența care separă HAR de SHAR (ca în verificarea de pe slide-ul Diebold–Mariano)
  - tabelul DM compară fiecare model doar cu log-HAR, deci nu arată perechea HAR–SHAR, motivul probabil al eliminării timpurii a HAR

Studiu de caz: poate învățarea automată să întreacă HAR?

- Modele, cu reglaj minim
  - familia HAR prin OLS și regresii penalizate
  - ansambluri de arbori (bagging, random forest, boosting) și rețele neuronale
  - catalogul complet în Anexă
- Evaluare: MSE în afara eșantionului raportată la MSE a HAR
  - calculată pe fiecare acțiune și apoi mediată
  - teste DM pe fiecare acțiune și MCS la 90%

Studiu de caz: 21 de modele, orizont de o lună (1/2)



MFM_ch9_ml_vs_har

Studiu de caz: 21 de modele, orizont de o lună (2/2)

- Sursa: Christensen, Siggaard & Veliyev (2023)
  - MSE relativă la HAR, orizont de o lună, toți cei 12 predictori, media pe 29 de acțiuni DJIA
  - NN_k^e: k straturi ascunse, media a e rețele
- Ansamblurile de arbori conduc: random forest 0,604, bagging 0,621
  - toate cele opt rețele sînt superioare HAR (0,786–0,964), dar adîncimea nu aduce cîștig
- Regresiile penalizate și extensiile liniare ale HAR sînt inferioare HAR (1,107–1,569)
- Fiecare bară este o medie a rapoartelor pe acțiuni, nu un test de semnificație al clasamentului

MFM_ch9_ml_vs_har

Semimartingale Itô și convergența stabilă

- Semimartingal Itô: $dp = \mu\, ds + \sigma\, dW + dJ$
  - μ, σ adaptate (cunoscute din trecut în fiecare moment)
  - σ càdlàg (continuă la dreapta, cu limite la stînga, deci poate avea salturi)
  - σ poate depinde de W (efect de levier)
  - Fără salturi: $\sqrt{M}(\mathrm{RV}_t - \mathrm{IV}_t) \xrightarrow{\mathcal{L}\text{-}s} \mathrm{MN}(0, 2\,\mathrm{IQ}_t)$, o distribuție Normală **mixtă**: varianța ei, $2\,\mathrm{IQ}_t$, este ea însăși aleatoare Barndorff-Nielsen & Shephard (2002)
  - $\mathrm{IQ}_t = \int \sigma^4(s)\, ds$ este cvarticitatea integrată
- Convergența **stabilă** ($\mathcal{L}$-s) Mykland & Zhang (2009)
  - convergența are loc împreună cu orice variabilă aleatoare măsurabilă în raport cu σ-algebra traiectoriei
  - De aceea $(\mathrm{RV}_t - \mathrm{IV}_t)/\sqrt{2\,\mathrm{RQ}_t/M} \to N(0,1)$: studentizarea cu $\sqrt{2\,\mathrm{RQ}_t/M}$, unde $\mathrm{RQ}_t \to \mathrm{IQ}_t$ este aleatoare, este legitimă doar sub convergența stabilă
- În eșantioane finite: statistica t pentru $\ln \mathrm{RV}_t$ are acoperire mai bună decît cea pe niveluri Barndorff-Nielsen & Shephard (2005)
  - intervalele de încredere din curs o folosesc

Înapoi

Derivarea deplasării din zgomot și a frecvenței optime

- Randamentul observat $r_i = r_i^* + \varepsilon_i - \varepsilon_{i-1}$, cu ε i.i.d. $(0, \omega^2)$, independent de p^*
  - Dată fiind traiectoria lui σ (drift nul, fără efect de levier: σ independent de W):
    $\mathbb{E}[\mathrm{RV}^{(n)} \mid \sigma] = \sum_i \mathbb{E}[r_i^{*2} \mid \sigma] + \sum_i \mathbb{E}[(\varepsilon_i - \varepsilon_{i-1})^2] = \mathrm{IV} + 2n\omega^2$
  - $\mathrm{cov}(r_i, r_{i-1}) = -\mathbb{E}[\varepsilon_{i-1}^2] = -\omega^2$
  - Din legea numerelor mari pentru termenii de zgomot, $\mathrm{RV}^{(n)}/(2n) \xrightarrow{P} \omega^2$ cînd $n \to \infty$: RV din toate tranzacțiile estimează zgomotul, nu IV Bandi & Russell (2008)
- MSE: $\mathbb{E}[(\mathrm{RV}^{(n)} - \mathrm{IV})^2] \approx \underbrace{2\,\mathrm{IQ}/n}_{\text{discretizare}} + \underbrace{4n^2\omega^4}_{\text{deplasarea la pătrat}} + O(n)$
  - Condiția de ordinul întîi $-2\,\mathrm{IQ}/n^2 + 8n\omega^4 = 0$ dă $n^* = \left(\mathrm{IQ}/(4\omega^4)\right)^{1/3}$ Bandi & Russell (2008)

Înapoi

Sezonalitatea intraday și testul de salturi

- Sezonalitatea deplasează BV (nu și RV)
  - $\mathbb{E}|r_i|\,|r_{i-\mathbf{1}}| \propto \sigma_i\sigma_{i-\mathbf{1}} \leq (\sigma_i^{\mathbf{2}} + \sigma_{i-\mathbf{1}}^{\mathbf{2}})/2$
  - Schimbările rapide de scară (deschiderea, închiderea) cresc $\mathrm{RV} - \mathrm{BV}$, deci testul raportului (secțiunea Salturi) respinge prea des
- Factorul de periodicitate f_i: estimatorul abaterii standard din Boudt, Croux & Laurent (2011)
  - scalăm fiecare randament cu $\sqrt{\mathrm{BV}_t/M_t}$
  - luăm rădăcina mediei pătratelor pe fiecare interval al zilei
  - normalizăm media lui $f_i^{\mathbf{2}}$ la 1
  - SPY: $f = 1{,}78$ la deschidere, 0,73 la minim, 1,52 la închidere
- Împărțind $r_{t,i}$ la f_i înainte de test, la nivelul de 0,1% rămîn 21 de respingeri SPY
  - pentru nivelul de 5%: textul principal

Înapoi

Ratele de convergență în prezența zgomotului

Estimator	Eroare de ordinul	Pozitiv	Sursa
RV pe toate tranzacțiile	inconsistent (deplasare $2n\omega^2$)	da	Bandi & Russell (2008)
RV rară, grilă fixă de 5 minute	$M^{-1/2}$, M fix; deplasarea rămîne	da	Liu, Patton & Sheppard (2015)
RV pe două scări	$n^{-1/6}$	nu	Zhang, Mykland & Aït-Sahalia (2005)
RV pe mai multe scări	$n^{-1/4}$	nu	Zhang (2006)
Nucleu realizat, flat-top / Parzen	$n^{-1/4}$ / $n^{-1/5}$	nu / da	Barndorff-Nielsen, Hansen, Lunde & Shephard (2008)
Pre-averaging	$n^{-1/4}$	nu	Jacod, Li, Mykland, Podolskij & Vetter (2009)

- $n^{-1/4}$ este rata optimă cu zgomot i.i.d.
  - prețul zgomotului este jumătate din exponentul obișnuit
- Cu $n = 23\,400$ de secunde: $n^{1/4} \approx 12{,}4$, față de $\sqrt{78} \approx 8{,}8$ pentru RV pe 5 minute
  - doar o euristică: ratele ignoră constantele, mărimea zgomotului și deplasarea RV rare
  - ratele sînt ale estimatorilor teoretici: grilele TSRV standard, nuclee cu capete mediate (jittering)
  - compatibilă cu cîștigurile modeste din Liu, Patton & Sheppard (2015)

înapoi

Pre-averaging (1/2)

- Medii locale ale unui număr de $k_n \approx \theta\sqrt{n}$ randamente, cu o pondere g pe $[0,1]$ Jacod, Li, Mykland, Podolskij & Vetter (2009)
  - θ: o constantă de reglaj, alta decît θ din testul de salturi
  - de ex. $g(x) = \min(x, 1-x)$
  - $\bar{r}_i = \sum_{j=1}^{k_n - 1} g(j/k_n)\, r_{i+j}$: varianța zgomotului scade ca $1/k_n$, semnalul păstrează ordinul k_n/n
- Estimatorul, cu $\psi_1 = \int_0^1 g'(x)^2 dx$ și $\psi_2 = \int_0^1 g(x)^2 dx$
  - $\psi_1 = 1$ și $\psi_2 = 1/12$ pentru g triunghiular; g': derivata lui g
  - $\widehat{\mathrm{IV}} = \frac{1}{\theta\, \psi_2 \sqrt{n}} \sum_i \bar{r}_i^2 - \frac{\psi_1}{2\,\theta^2 \psi_2\, n} \sum_i r_i^2$ (al doilea termen: corecția deplasării din zgomot)

înapoi

Pre-averaging (2/2)

- Rata $n^{-1/4}$
  - același procedeu dă estimatori robuști la salturi (bipower cu pre-averaging) și estimatori de covarianță
- Construită pentru date tick cu tick
  - pe bare de 5 minute rămîne puțin zgomot de mediat (graficul de semnătură empiric, secțiunea Zgomotul de microstructură)

înapoi

Covarianța realizată și tranzacționarea asincronă

- Covarianța realizată $\sum_i r^{(1)}_{t,i} r^{(2)}_{t,i} \to \int \sigma_{12}(s)\, ds$ cere prețurile ambelor active pe **aceeași** grilă
  - $r^{(1)}$, $r^{(2)}$: randamentele activelor 1 și 2, $\sigma_{12}(s)$: covarianța lor instantanee
  - Efectul Epps Epps (1979): corelația măsurată scade spre 0 cînd eșantionarea devine mai deasă, deoarece tranzacțiile sînt asincrone, iar ultimul preț este neactualizat
- Hayashi & Yoshida (2005): însumăm produsele **tuturor** perechilor de randamente ale căror intervale de timp se suprapun
  - $\sum_{i,j} r^{(1)}_i r^{(2)}_j \, 1\{(t_{i-1}, t_i] \cap (s_{j-1}, s_j] \neq \emptyset\}$: $(t_{i-1}, t_i]$, $(s_{j-1}, s_j]$: intervalele dintre tranzacțiile consecutive ale fiecărui activ
  - Estimatorul este nedeplasat fără sincronizare (în absența zgomotului)
- Cu zgomot în plus: versiuni cu pre-averaging sau nuclee multivariate
  - aplicație: riscul intraday al portofoliilor (Capitolul 4)

Înapoi

Localizarea salturilor în cursul zilei, Lee–Mykland (1/2)

- Lee & Mykland (2008): $L_{t,i} = r_{t,i}/\hat{\sigma}_{t,i}$
  - $\hat{\sigma}^2_{t,i}$: media bipower a celor $K = 270$ de randamente anterioare (valoarea lor pentru date de 5 minute)
  - Fără salturi, $\max_i |L_i|$ pe n randamente are o limită Gumbel: respingem dacă $(|L| - C_n)/S_n > -\ln(-\ln(1-\alpha))$
  - $C_n = \sqrt{2\ln n} - \frac{\ln\pi + \ln\ln n}{2\sqrt{2\ln n}}$ și $S_n = 1/\sqrt{2\ln n}$: constantele de centrare și de scalare ale limitei Gumbel
  - $\hat{\sigma}^2$ al nostru include factorul $\pi/2$, deci constanta $c = \sqrt{2/\pi}$ din Lee & Mykland (2008) devine 1
  - SPY, randamente împărțite la f_i, $\alpha = 1\%$, $n = 115\,441$ (întregul eșantion): pragul $|L| > 5{,}41$

Înapoi

Localizarea salturilor în cursul zilei, Lee–Mykland (2/2)

- Testul găsește 169 de randamente de 5 minute cu salt (nu zile) în 105 zile, dintre care 62 miercurea între 14:00 și 15:00
  - atunci se publică comunicatele FOMC (Federal Open Market Committee, comitetul de politică monetară al Fed) (14:00) și au loc conferințele de presă (14:30)
  - Cel mai mare salt: $L = -22{,}9$, pe 2025-10-29
  - Doar 10% din aceste zile sînt semnalate și de testul zilnic al raportului
- Cele două teste pun întrebări diferite
  - Lee–Mykland: un randament extrem față de volatilitatea locală
  - testul raportului: un decalaj zilnic între RV și BV

Înapoi

Estimatori ai variației integrate robuști la salturi (1/3)

- RV trunchiată Mancini (2009): $\mathrm{TRV}_t = \sum_i r_{t,i}^{\mathbf{2}} 1\{|r_{t,i}| \leq c\, \hat{\sigma}_t f_i\, \Delta^{\varpi}\}$, $\varpi \in (0, 1/2)$
  - Păstrăm doar randamentele la pătrat sub un prag: c: un multiplu, $\hat{\sigma}_t$: o volatilitate zilnică robustă la salturi, f_i: factorul de periodicitate, ϖ: exponentul lungimii intervalului Δ
  - Incrementele difuzive sînt $O_p(\Delta^{\mathbf{1/2}})$ (de ordinul $\sqrt{\Delta}$) și ajung sub prag; salturile sînt $O(1)$ și sînt eliminate

înapoi

Estimatori ai variației integrate robuști la salturi (2/3)

- MedRV (*Median Realised Variance*, varianța realizată mediană) Andersen, Dobrev & Schaumburg (2012):
  $\frac{\pi}{6-4\sqrt{3}+\pi}\frac{M}{M-2}\sum_i \mathrm{med}(|r_{i-1}|,|r_i|,|r_{i+1}|)^2$
  - med: mediana a trei randamente absolute consecutive; prima fracție face MedRV consistent pentru IV
  - Un randament izolat mai mare în valoare absolută decît ambii vecini nu intră în mediană (care trece la un vecin), deci un salt izolat mare este exclus, deși efectul lui nu dispare exact în eșantioane finite
  - MedRV este mai puțin sensibil decît BV la randamente nule izolate
  - Două salturi alăturate pot intra totuși în mediană

înapoi

Estimatori ai variației integrate robuști la salturi (3/3)

- Ponderea părții de salt în RV pentru SPY, însumată pe toate zilele, fără test
  - partea de salt: RV minus estimatorul robust
  - 6,3% cu BV, 7,2% cu MedRV și 12,5% cu TRV (*Truncated Realised Variance*, varianța realizată trunchiată)
  - pe zilele cu salt semnificativ: 2,8%
- Pentru TRV, cu $c = 3$ și $\varpi = 0{,}49$, trunchierea acționează în 44% din zile
  - c este o alegere de reglaj
- Aït-Sahalia & Jacod (2009): comparăm variațiile de putere $\sum |r|^p$, $p > 2$, la Δ și $k\Delta$
  - raportul tinde la 1 cu salturi și la $k^{p/2-1}$ fără
  - limite cunoscute sub ambele ipoteze

Înapoi

HAR ca AR(22) restricționat, testul Wald

- Nerestricționat: $\ln \mathrm{RV}_{t+\mathbf{1}} = \phi_{\mathbf{0}} + \sum_{k=\mathbf{1}}^{\mathbf{22}} \phi_k \ln \mathrm{RV}_{t+\mathbf{1}-k} + \varepsilon_{t+\mathbf{1}}$
  - HAR impune $\phi_{\mathbf{2}} = \ldots = \phi_{\mathbf{5}}$ (3 restricții) și $\phi_{\mathbf{6}} = \ldots = \phi_{\mathbf{22}}$ (16): $R\phi = 0$ cu 19 rînduri
- Wald: $W = (R\hat{\phi})^\top (R\hat{V}_{\mathbf{NW}} R^\top)^{-\mathbf{1}} R\hat{\phi} \to \chi^2_{\mathbf{19}}$ sub restricțiile HAR
  - R: matricea restricțiilor 19×23, $\hat{V}_{\mathbf{NW}}$: covariança Newey–West a lui $\hat{\phi}$
  - W mare = restricții respinse
- SPY, logaritmul variantei totale: $W = 18{,}2$ și $p = 0{,}51$
  - $R^{\mathbf{2}}$ este 0,477 pentru AR(22) și 0,470 pentru HAR
  - Restricțiile asupra ponderilor în trepte nu sînt respinse
  - HAR folosește cu 19 parametri mai puțini, cu o reducere minimă a calității ajustării în eșantion
  - zgomotul de estimare mai mic aduce un avantaj în afara eșantionului

Înapoi

Eroarea de măsurare din regresori și HARQ (1/2)

- $\mathrm{RV}_t = \mathrm{IV}_t + u_t$, cu $\mathbb{E}[u_t \mid \sigma] = 0$ și $\mathrm{Var}(u_t \mid \sigma) = 2\,\mathrm{IQ}_t/M$
  - CLT din secțiunea Varianța realizată
  - Dacă $\mathrm{IV}_{t+\mathbf{1}} = \beta_{\mathbf{0}} + \beta\,\mathrm{IV}_t + \eta_{t+\mathbf{1}}$, OLS pentru $\mathrm{RV}_{t+\mathbf{1}}$ pe RV_t dă plim $\hat{\beta} = \beta\,\dfrac{\mathrm{Var}(\mathrm{IV})}{\mathrm{Var}(\mathrm{IV}) + \mathbb{E}[2\,\mathrm{IQ}/M]} < \beta$ pentru $\beta > 0$: atenuare ($\eta_{t+\mathbf{1}}$: șocul asupra IV; plim: limita în probabilitate)
- Dacă varianța semnalului nu crește cu IQ_t, raportul semnal/zgomot al lui RV_t scade cînd IQ_t este mare
  - o aproximare de lucru
  - HARQ aproximează ponderea descrescătoare prin $\beta_d + \beta_Q\sqrt{\mathrm{RQ}_t}$, $\beta_Q < 0$ Bollerslev, Patton & Quaedvlieg (2016)

înapoi

Eroarea de măsurare din regresori și HARQ (2/2)

- SPY, regresie pe niveluri, cu erorile standard Newey–West în paranteze
  - $\hat{\beta}_Q = -0{,}0088$ (0,0042) pentru RV intraday
  - $\hat{\beta}_Q = -0{,}0022$ (0,0024) pentru varianța totală
  - Panta zilnică (RV intraday) 1,13 la mediana lui $\sqrt{\mathrm{RQ}}$, 1,09 la percentila 99: semnul prezis de teorie, un efect mic
- Citim $\hat{\beta}_d$ împreună cu celelalte pante
  - panta zilnică peste 1 (1,135) este parțial compensată de o pantă săptămînală negativă ($-0{,}120$)
  - cele trei medii ale RV se suprapun prin construcție (fiecare conține RV de ieri), deci împărțirea între ele este imprecisă (SE pentru $\hat{\beta}_d$: 0,494)
  - pantele însumează 1,04 (varianța totală: 0,80), deci regresia pe niveluri pentru RV intraday este un diagnostic în eșantion pentru semnul lui β_Q, nu un model de prognoză
- Comparație pe aceeași țintă, varianța totală: HAR $\hat{\beta}_d = 0{,}399$, HARQ $\hat{\beta}_d = 0{,}542$
  - dacă ponderea poate scădea cu RQ (*Realised Quarticity*, cvarticitatea realizată), panta zilnică de bază crește, în direcția prezisă de argumentul atenuării
  - diferența este sub o eroare standard

Înapoi

Dincolo de HAR, semivarianțe și modelarea comună a randamentelor și a RV (1/2)

- Semivarianțele realizate intră în SHAR Patton & Sheppard (2015): $\beta_d \mathrm{RV}_t$ devine $\beta^+ \mathrm{RS}_t^+ + \beta^- \mathrm{RS}_t^-$
  - $\mathrm{RS}_t^{\pm} = \sum_i r_{t,i}^2 \, 1\{r_{t,i} \gtrless 0\}$, cu $\mathrm{RS}^+ + \mathrm{RS}^- = \mathrm{RV}$
  - SPY, ținta varianța totală, regresie pe niveluri (randamentul peste noapte socotit ca încă un randament al zilei): $\hat{\beta}^- = 0{,}56$ (0,23), $\hat{\beta}^+ = 0{,}21$ (0,18); studiul de caz (ținta RV intraday) dă 1,44 și $-0{,}34$
  - Testul Wald pentru $\beta^+ = \beta^-$ dă $p = 0{,}35$
  - Volatilitatea „rea" este mai persistentă, ca în Patton & Sheppard (2015), dar șase ani de date nu separă cele două pante
  - În comparația MCS, HARQ și SHAR sînt estimate pe niveluri cu filtrul din Bollerslev, Patton & Quaedvlieg (2016): prognozele care ies din intervalul $[\min, \max]$ al valorilor din eșantionul de estimare sînt înlocuite cu media eșantionului

Înapoi

Dincolo de HAR, semivarianțe și modelarea comună a randamentelor și a RV (2/2)

- Realized GARCH Hansen, Huang & Shek (2012): o ecuație GARCH plus o **ecuație de măsurare**
  - în ecuația GARCH intră măsura realizată x_t
  - $\ln h_t = \omega + \beta \ln h_{t-\mathbf{1}} + \gamma \ln x_{t-\mathbf{1}}$, $\ln x_t = \xi + \varphi \ln h_t + \tau(z_t) + u_t$, $r_t = \sqrt{h_t}\, z_t$: estimare comună prin cvasi-maximă verosimilitate
  - h_t este varianța condiționată, x_t măsura realizată (de ex. RV), z_t șocul standardizat, iar $\tau(z_t)$ funcția efectului de levier
  - ω este aici constanta ecuației GARCH (nu zgomotul); ξ și φ sînt coeficienții ecuației de măsurare, iar u_t eroarea ei
- Modelele HEAVY (high-frequency-based volatility) Shephard & Sheppard (2010): prognoze coerente pe mai mulți pași
  - o ecuație de tip GARCH pentru varianța zilnică, în care intră RV întîrziată
  - o ecuație pentru RV însăși

Înapoi

Datele studiului de caz, 29 de acțiuni DJIA și 12 predictori

- Date: 29 dintre cele 30 de acțiuni DJIA (fără Visa), 29 ianuarie 2001 – 31 decembrie 2017
  - $T = 4\,257$ de zile
  - RV calculată la 5 minute din prețurile tranzacțiilor de la New York Stock Exchange (NYSE), după filtrarea datelor
  - $n = 78$ de randamente pe zi
- Împărțire: antrenare 70% (2 964 de zile), validare 10% (424), test 20% (847)
  - $2\,964 + 424 + 847 = 4\,235$: primele 22 de zile servesc doar mediei lunare a RV
  - robustețe: antrenare 1 000 sau 2 000 de zile, validare 200
- $\mathcal{M}_{\mathbf{HAR}}$: RV zilnică, săptămînală, lunară
- $\mathcal{M}_{\mathbf{ALL}}$ adaugă 9 predictori
  - volatilitatea implicită, indicator pentru zilele în care se anunță rezultatele financiare, momentum pe o săptămînă, variația logaritmului volumului tranzacționat în dolari
  - VIX, randamentul Hang Seng la pătrat, indicele condițiilor de afaceri, variația ratei la 3 luni a titlurilor de stat, indicele incertitudinii politicii economice

Înapoi

Modelele studiului de caz, de la HAR la rețele neuronale

- Modele: HAR, log-HAR, HAR cu efect de levier, SHAR și HARQ
  - estimate prin OLS pe o fereastră mobilă
- Regresii penalizate: ridge, lasso, elastic net, lasso adaptiv și post-lasso
  - ridge: penalizare L_2; lasso: L_1; elastic net: o combinație
  - penalizarea este aleasă pe setul de validare
- Arbori: bagging, random forest, gradient boosting
  - random forest cu valorile implicite ale lui Breiman: 500 de arbori, $P/3$ predictori candidați la fiecare ramificare (P: numărul de predictori), cel puțin 5 observații într-un nod
- Rețele neuronale NN_k^e: $k = 1, \ldots, 4$ straturi ascunse (2; 4-2; 8-4-2; 16-8-4-2 neuroni)
  - leaky ReLU (*Rectified Linear Unit*, funcția de activare liniară rectificată), Adam, dropout, early stopping
  - $e = 1$: rețeaua cu cea mai mică eroare pe setul de validare dintre cele 100 de inițializări, iar $e = 10$: media celor mai bune 10
- Evaluare: raportul dintre MSE în afara eșantionului și MSE a modelului HAR
  - calculat pentru fiecare acțiune și apoi mediat
  - test DM unilateral pentru fiecare acțiune Diebold & Mariano (1995)
  - MCS la 90% Hansen, Lunde & Nason (2011)

Înapoi

Studiul de caz, rezultate suplimentare

- Prognoze pentru ziua următoare, doar cu valorile întîrziate ale RV
  - MSE relativ la HAR: log-HAR 0,995, HARQ 1,059, random forest 1,020, rețele neuronale 0,954–0,990
  - RV trecută conține puțină informație suplimentară
  - rețelele aduc cele mai clare cîștiguri (1–5%), log-HAR este superior HAR la limită
- Prognoze pentru ziua următoare, cu toți cei 12 predictori: elastic net 0,916
  - celelalte modele apar în graficul pe orizonturi al studiului de caz
  - Predictorii suplimentari aduc cel mai mult cu penalizare, neliniaritate sau transformarea logaritmică
  - HARQ supraajustează datele (*overfitting*)
- MCS la 90%, cu toți predictorii
  - rețelele rămîn în mulțime pentru aproximativ 80% din acțiuni
  - HAR rămîne pentru aproximativ 20–40% din acțiuni
  - Importanța variabilelor, măsurată prin efectele locale acumulate: RV zilnică, RV săptămînală și volatilitatea implicită domină în toate cele patru modele prezentate
  - Random forest și rețelele atribuie volatilității implicite o importanță mai mare decît OLS cu toți predictorii și elastic net

Înapoi

Studiul de caz, replicare, limite și extinderi (1/2)

- Replicare pe datele cursului: barele SPY de 5 minute
  - RV de la deschidere la închidere cu 78 de randamente pe zi
  - autorii raportează că SPY confirmă rezultatele pe acțiuni
  - Împărțirea de robustețe a autorilor: 1 000 de zile de antrenare, 200 de zile de validare, restul pînă la 18 septembrie 2026 ca set de test
  - Predictori: VIX (calculat din opțiunile pe indicele S&P 500; aproximează volatilitatea implicită a SPY, care ar cere opțiuni pe SPY), randamentul Hang Seng la pătrat, momentum și volum în dolari pentru SPY, rata titlurilor de stat și incertitudinea politicii economice (Federal Reserve Economic Data, FRED)
  - fără indicator de rezultate financiare pentru un fond
- Extindere: lucrarea compară prognozele de varianță doar prin MSE
  - se păstrează avantajul ML la evaluarea prin QLIKE, robustă la eroarea de măsurare din RV Patton (2011)?
  - se păstrează și pentru RV Bitcoin pe 24 de ore?

Înapoi

Studiul de caz, replicare, limite și extinderi (2/2)

- Limite precizate de autori
  - Reglarea mai atentă a hiperparametrilor ar putea îmbunătăți rezultatele modelelor ML
  - Rețelele sînt antrenate o singură dată și nu sînt reestimate în fereastra de test
  - Clasamentele depind de împărțire: cu 1 000 de zile de antrenare criza financiară intră în setul de test, iar extensiile HAR sînt superioare HAR
  - Modelele HAR cu o componentă de salt separată nu sînt incluse
  - Efectele locale acumulate descriu comportamentul mediu al predicțiilor și pot omite efectele evenimentelor rare, precum anunțarea rezultatelor financiare

Înapoi

Bucla descoperirii, prompturi reutilizabile (1/2)

- Instrumente: un asistent AI
  - un model lingvistic mare (LLM, large language model), precum Claude, ChatGPT, Gemini sau GitHub Copilot
  - căutarea literaturii: Semantic Scholar, Elicit, Google Scholar
- **1. Literatura**: sintetizăm rezultatele într-un tabel
  - apoi verificăm fiecare DOI pe Crossref (titlul trebuie să coincidă)
  - `Tabelează estimările lui H publicate din 2018: estimatorul, frecvența de eșantionare, tratamentul erorii de măsurare din RV, estimarea punctuală și eroarea standard, DOI.`
- **2. Ipoteze concurente**, fiecare cu forma implicată a lui $m(2, \Delta)$
  - `Derivează m(2, Delta) pentru trei procese care dau H în jur de 0,1 cu estimatorul de scalare: log-volatilitate fracționară cu H = 0,1; H = 0,5 plus eroare i.i.d.; un model cu doi factori, unul cu revenire rapidă la medie.`
- **3. Date și cod**: replicăm întîi cifra din capitol
  - `Scrie cod Python care construiește varianța zilnică totală din barele SPY de 5 minute ale cursului (RV la 5 minute plus pătratul randamentului peste noapte, de la închiderea precedentă la deschidere), calculează m(q, Delta) pentru q în {0,5; 1; 1,5; 2; 3}, Delta = 1..30, și panta H.` Ținta: $\hat{H} = 0{,}11$

Înapoi

Bucla descoperirii, prompturi reutilizabile (2/2)

- **4. Robustețe**, pre-înregistrată, cu o verificare a puterii
  - `Scrie un plan pre-înregistrat: ferestre de laguri [1,30], [5,30], [10,50]; RV, variația bipower și RV subeșantionată la 30 de minute; ajustarea m(2, Delta) = c Delta^(2H) + k; intervale bootstrap pe blocuri; mărimea și puterea prin simulare la H = 0,1 și 0,5.`
- **5. Critica**: cerem cel mai puternic contraargument
  - `Joacă rolul unui referent: ce structuri de eroare non-i.i.d. în log RV (autocorelate, dependente de nivel, provenite din randamentele peste noapte sau din salturi) nu ar fi eliminate de corecția cu k constant și cum ai detecta fiecare?`
- **6. Raportul**: toate ferestrele și măsurile, intervalele bootstrap, designul simulării, codul și jurnalul AI

Înapoi

Pentru aprofundare

- Manual și sinteze
  - Franke, Härdle & Hafner (2019): statistica piețelor financiare, modele de volatilitate
  - Andersen, Bollerslev, Diebold & Labys (2003): modelarea și prognoza volatilității realizate
  - Corsi (2009): modelul HAR
- Cripto-active
  - Informația intraday poate intra în prognozele de risc pentru Bitcoin; entropia randamentelor de înaltă frecvență a fost propusă ca un astfel de semnal Pele & Mazurencu-Marinescu-Pele (2019)
  - Cripto-activele formează o clasă de active distinctă după caracteristicile distribuțiilor randamentelor Pele et al. (2023)

Înapoi

Studiu suplimentar pe Quantinar

Cursuri avansate pe quantinar.com

- Nonparametric and Semiparametric Models
- Kalman Filter
- Applied Time Series Analysis with Python

Cod și date

- Quantlets: github.com/danpele/MFM/tree/main/Quantlets/Ch_09

Înapoi

Acronime folosite în acest material (1/3)

- **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*
- **CLT**: Central Limit Theorem — *teorema limită centrală*
- **CV**: Cross-Validation — *validare încrucișată*
- **DJIA**: Dow Jones Industrial Average — *indicele Dow Jones al marilor companii industriale americane*
- **DM**: Diebold–Mariano (test) — *testul Diebold–Mariano de comparare a prognozelor*
- **DOI**: Digital Object Identifier — *identificatorul permanent al unei publicații*
- **ETF**: Exchange-Traded Fund — *fond tranzacționat la bursă*
- **EWMA**: Exponentially Weighted Moving Average — *medie mobilă ponderată exponențial*
- **FDR**: False Discovery Rate — *rata descoperirilor false*
- **FOMC**: Federal Open Market Committee — *Comitetul Federal pentru Piața Deschisă al Rezervei Federale*
- **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*
- **GMM**: Generalized Method of Moments — *metoda generalizată a momentelor*

Acronime folosite în acest material (2/3)

- **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*
- **HAR-CJ**: HAR model with Continuous and Jump components — *model HAR cu componentă continuă și componentă de salt*
- **HAR-RV**: Heterogeneous AutoRegressive model of Realised Volatility — *model autoregresiv heterogen al volatilității realizate*
- **HAR-X**: Heterogeneous AutoRegressive model with additional regressors — *model HAR cu regresori suplimentari*
- **HARQ**: HAR model with Realised Quarticity — *model HAR cu cvarticitatea realizată*
- **HEAVY**: High-frEquency-bAsed VolatilitY (models) — *modele de volatilitate bazate pe date de înaltă frecvență*
- **LLM**: Large Language Model — *model lingvistic mare*
- **MAE**: Mean Absolute Error — *eroarea absolută medie*
- **MCS**: Model Confidence Set — *mulțimea de încredere a modelelor*
- **MedRV**: Median Realised Variance — *varianța realizată mediană*
- **ML**: Machine Learning — *învățare automată*
- **MSE**: Mean Squared Error — *eroarea pătratică medie*
- **MZ**: Mincer–Zarnowitz (regression) — *regresia Mincer–Zarnowitz*
- **NYSE**: New York Stock Exchange — *Bursa din New York*
- **OLS**: Ordinary Least Squares — *metoda celor mai mici pătrate*

Acronime folosite în acest material (3/3)

- **PRDS**: Positive Regression Dependence on a Subset (of p-values) — *dependența pozitivă de regresie pe o submulțime (a valorilor p)*
- **PV**: Partial Variance — *varianța parțială*
- **QLIKE**: Quasi-LIKElihood (loss function) — *funcția de pierdere de tip cvasi-verosimilitate*
- **ReLU**: Rectified Linear Unit — *unitatea liniară rectificată*, $\max(0, z)$
- **RK**: Realised Kernel — *estimatorul realised kernel al varianței (cu nucleu)*
- **RQ**: Realised Quarticity — *cvarticitatea realizată*
- **RS**: Realised Semivariance — *semivarianța realizată*
- **RV**: Realised Variance — *varianța realizată*
- **SE**: Standard Error — *eroare standard*
- **SHAR**: Semivariance HAR model — *model HAR cu semivarianțe*
- **SUA**: Statele Unite ale Americii
- **SV**: SemiVariance (SV-HAR = SHAR) — *semivarianța (SV-HAR = SHAR)*
- **TQ**: Tripower Quarticity — *cvarticitatea tripower*
- **TRV**: Truncated Realised Variance — *varianța realizată trunchiată*
- **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*
- **VRP**: Variance Risk Premium — *prima de risc a varianței*

Bibliografie (1/6)

- Aït-Sahalia, Y., Jacod, J. (2009). Testing for jumps in a discretely observed process. *Annals of Statistics*, 37(1), 184–222.
- Andersen, T.G., Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2–3), 115–158.
- Andersen, T.G., Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4), 885–905.
- 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., Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1), 43–76.
- 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.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), 579–625.
- Andersen, T.G., Dobrev, D., Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169(1), 75–93.
- Bajgrowicz, P., Scaillet, O., Treccani, A. (2016). Jumps in high-frequency data: spurious detections, dynamics, and news. *Management Science*, 62(8), 2198–2217.
- Bandi, F.M., Russell, J.R. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, 75(2), 339–369.

Bibliografie (2/6)

- 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. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, 64(2), 253–280.
- Barndorff-Nielsen, O.E., Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1), 1–37.
- Barndorff-Nielsen, O.E., Shephard, N. (2005). How accurate is the asymptotic approximation to the distribution of realised variance? In D.W.K. Andrews, J.H. Stock (Eds.), *Identification and Inference for Econometric Models*, Cambridge University Press, 306–331.
- Barndorff-Nielsen, O.E., Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1), 1–30.
- 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.
- Bennedsen, M., Lunde, A., Pakkanen, M.S. (2022). Decoupling the short- and long-term behavior of stochastic volatility. *Journal of Financial Econometrics*, 20(5), 961–1006.
- Bolko, A.E., Christensen, K., Pakkanen, M.S., Veliyev, B. (2023). A GMM approach to estimate the roughness of stochastic volatility. *Journal of Econometrics*, 235(2), 745–778.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.

Bibliografie (3/6)

- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69(3), 542–547.
- Bollerslev, T., Medeiros, M.C., Patton, A.J., Quaedvlieg, R. (2022). From zero to hero: realized partial (co)variances. *Journal of Econometrics*, 231(2), 348–360.
- 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.
- Bollerslev, T., Tauchen, G., Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11), 4463–4492.
- 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.
- Carr, P., Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 22(3), 1311–1341.
- Christensen, K., Oomen, R.C.A., Podolskij, M. (2014). Fact or friction: jumps at ultra high frequency. *Journal of Financial Economics*, 114(3), 576–599.
- Christensen, K., Siggaard, M., Veliyev, B. (2023). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, 21(5), 1680–1727.
- Cont, R., Das, P. (2024). Rough volatility: fact or artefact? *Sankhya B*, 86(1), 191–223.
- 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 (4/6)

- Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, 78(383), 605–610.
- Epps, T.W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366), 291–298.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). Statistics of Financial Markets: An Introduction, 5th ed. Springer.
- Fukasawa, M., Takabatake, T., Westphal, R. (2022). Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Mathematical Finance*, 32(4), 1086–1132.
- Gatheral, J., Jaisson, T., Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6), 933–949.
- Hansen, P.R., Huang, Z., Shek, H.H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6), 877–906.
- Hansen, P.R., Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7), 873–889.
- Hansen, P.R., Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2), 127–161.
- Hansen, P.R., Lunde, A., Nason, J.M. (2011). The model confidence set. *Econometrica*, 79(2), 453–497.
- Hayashi, T., Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2), 359–379.

Bibliografie (5/6)

- Huang, X., Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3(4), 456–499.
- Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and their Applications*, 119(7), 2249–2276.
- Lee, S.S., Mykland, P.A. (2008). Jumps in financial markets: a new nonparametric test and jump dynamics. *Review of Financial Studies*, 21(6), 2535–2563.
- 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.
- Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36(2), 270–296.
- Merton, R.C. (1980). On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics*, 8(4), 323–361.
- Mincer, J., Zarnowitz, V. (1969). The evaluation of economic forecasts. In J. Mincer (Ed.), *Economic Forecasts and Expectations*, NBER, 3–46.
- Müller, U.A., Dacorogna, M.M., Davé, R.D., Olsen, R.B., Pictet, O.V., von Weizsäcker, J.E. (1997). Volatilities of different time resolutions: analyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2–3), 213–239.
- Mykland, P.A., Zhang, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica*, 77(5), 1403–1445.
- Patton, A.J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256.

Bibliografie (6/6)

- 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.
- Pele, D.T., Wesselhöfft, N., Härdle, W.K., Kolossiatis, M., Yatracos, Y.G. (2023). Are cryptos becoming alternative assets? *European Journal of Finance*, 29(10), 1064–1105.
- Shephard, N., Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25(2), 197–231.
- Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli*, 12(6), 1019–1043.
- 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.