








Modelarea piețelor financiare

Capitolul 14: Deep learning și modele fundaționale pentru serii de timp



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

Prognozează rețelele profunde mai bine decît modelele clasice?



Yann LeCun. **Foto: École polytechnique (2025); CC BY-SA 2.0; Wikimedia Commons**

"Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction."
— *LeCun, Bengio & Hinton (2015)*

- Semnalul este în volatilitate, nu în medie (S&P 500 din 2000, 6 718 zile)
  - Autocorelația de ordinul 1: $-0{,}099$ pentru randamente, 0,29 pentru randamentele absolute
- Randamente: niciun model nu depășește prognoza zero pe S&P 500 și Bitcoin
  - După corecția pentru teste multiple
- Volatilitate (SPY, 979 de zile): pierderea medie QLIKE (*quasi-likelihood*, de tip cvasi-verosimilitate) este 0,223 pentru Chronos-2 și 0,226 pentru log-HAR
  - Diferența nu este semnificativă (testul Diebold–Mariano, $p = 0{,}539$)
- Coada stîngă: Chronos-2 folosit direct depășește VaR (*Value-at-Risk*, valoarea la risc) 1% în 1,67% din zilele S&P 500
  - Pe S&P 500, simularea istorică filtrată (FHS) are cea mai mică pierdere a perechii (VaR, ES)

▸ **Graficul central: pierderile QLIKE ale celor opt modele**

Pot rețelele profunde prognoza mai bine decît modelele clasice?

- **Întrebarea**: prognozează rețelele profunde și **modelele fundaționale** pentru serii de timp mai bine decît modelele clasice?
  - Țintele: **randamentele**, **volatilitatea** și **riscul de coadă** (*tail risk*, coada stîngă)
  - Modelele clasice: cele din capitolele 5–9
  - **Deep learning** (învățare profundă): modele construite din multe straturi de unități neliniare simple, antrenate pe date
  - **Modelele fundaționale** (din 2024): rețele pre-antrenate pe miliarde de observații din alte serii, folosite fără nicio estimare pe seria-țintă
  - De ce merită testate: corpusurile lor de pre-antrenare sînt dominate de serii nefinanciare (energie, transport, vreme, web)
- Răspundem cu instrumentele de evaluare din Capitolul 8
  - Repere puternice, diferențe de pierdere, backtesting
- Partea I: rețele pentru secvențe, atenție, modele fundaționale, Testul 1 (randamente)
- Partea a II-a: Testul 2, Testul 3, studii de caz, AI pentru descoperire științifică
  - Testul 2: volatilitate realizată
  - Testul 3: riscul din coada stîngă (value-at-risk, VaR 1%, și expected shortfall, ES 2,5%)
  - AI: artificial intelligence, inteligența artificială

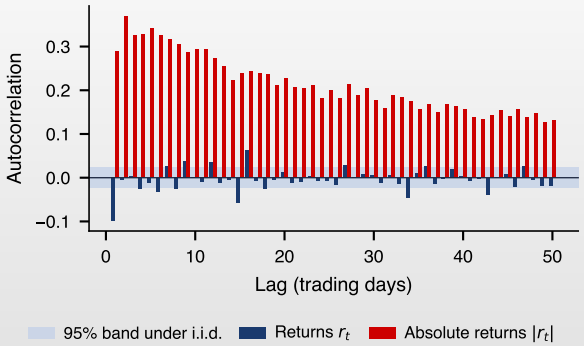
Rezultatele învățării

- Derivați dispariția gradientului din răspunsurile la impuls ale unui model vector autoregresiv (VAR)
- Derivați starea celulei LSTM ca un proces autoregresiv de ordinul 1, AR(1)
  - LSTM: long short-term memory, o rețea recurentă cu porți
- Calculați, dintr-o grilă de cuantile zero-shot, prognoze punctuale, de volatilitate și VaR/ES
  - Zero-shot: obținută fără nicio antrenare pe seria-țintă
- Interpretați comparația cu reperul potrivit fiecărei ținte
  - Prognoza zero pentru randamente
  - Log-HAR pentru volatilitate
  - FHS pentru cozi
- Interpretați testele unui model nou: imbricare, teste multiple, calibrare, contaminare și putere
- 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
- Aprofundări și Anexa: lectură pentru familiarizare

Trei ținte de prognoză

- **Randamentele** $r_t = 100 \ln(P_t / P_{t-1})$ (media condiționată)
  - P_t: prețul de închidere din ziua t
  - Raport semnal-zgomot foarte mic: tiparele predictibile dispar odată exploatate (Capitolul 2)
  - Reperul: prognoza zero
- **Volatilitatea** (varianța condiționată): persistentă, în episoade (capitolele 1 și 5)
  - Deci un semnal de învățat
  - Reperul: log-HAR, logaritmul varianței realizate regresat pe mediile lui zilnică, săptămînală și lunară (heterogeneous autoregressive, Testul 2)
- **Riscul de coadă**: forma cozii stîngi, nu doar centrul distribuției (Capitolul 7)
  - VaR 1% (value-at-risk): pierderea zilnică depășită cu probabilitatea de 1%
  - ES 2,5% (expected shortfall): pierderea medie în cele mai proaste 2,5% dintre zile
  - Reperul: FHS (filtered historical simulation): cuantila empirică a randamentelor standardizate cu volatilitatea GARCH (generalised autoregressive conditional heteroskedasticity, Capitolul 5), înmulțită cu volatilitatea prognozată

Semnalul: în medie sau în volatilitate



- Randamentele zilnice ale S&P 500 din 2000 (6 718 zile), lag-urile 1–50
  - Banda albastru deschis: $\pm 1{,}96/\sqrt{T} = \pm 0{,}024$ (T: numărul de zile), intervalul de 95% al unei autocorelații sub ipoteza i.i.d. (randamente independente și identic distribuite)
- Randamentele: autocorelația de ordinul 1 este $-0{,}099$
  - 15 din 50 de lag-uri depășesc banda, dar sînt mici și au semne diferite
- Randamentele absolute: autocorelația este 0,29 la lag-ul 1 și 0,13 la lag-ul 50
  - Toate cele 50 de lag-uri depășesc banda: gruparea volatilității (*volatility clustering*) este persistentă
  - Banda presupune $|r_t|$ i.i.d. cu $E[r_t^2] < \infty$; altfel se folosește o bandă robustă
- Modelele flexibile pot extrage semnalul din volatilitate, dar aproape nimic din medie

MFM_ch14_acf_signal ▶ Interactiv

De la perceptronul multistrat la deep learning (1/3)

- **Întrebarea secțiunii**: cum poate o rețea să folosească informația din zile îndepărtate?
  - Și de ce eșuează o rețea recurentă simplă?
- MLP (perceptronul multistrat, Capitolul 13)
  - Fiecare strat aplică o funcție neliniară unei combinații liniare a stratului anterior

$$h^{(\ell)} = \varphi\left(W^{(\ell)} h^{(\ell-\mathbf{1})} + b^{(\ell)}\right), \qquad h^{(\mathbf{0})} = x$$

  - x este vectorul de intrare, iar $h^{(\ell)}$ este ieșirea stratului ℓ
  - $W^{(\ell)}$, $b^{(\ell)}$: matricea de ponderi și vectorul de termeni liberi ai stratului ℓ, estimate din date
  - φ: funcția de activare, aplicată element cu element: ReLU (rectified linear unit) $\max(0, z)$, tanh sau sigmoida $\sigma(z) = 1/(1 + e^{-z}) \in (0, 1)$
- Un MLP primește o fereastră fixă de intrări
  - O rețea pentru secvențe trebuie să rețină informația de la un pas la altul

De la perceptronul multistrat la deep learning (2/3)

- Antrenarea: propagarea înapoi (Rumelhart, Hinton & Williams (1986))
  - Calculează gradientul pierderii în raport cu fiecare pondere
- Regula de actualizare a ponderilor: Adam (adaptive moment estimation, Kingma & Ba (2015))
  - Un pas de gradient normalizat prin mediile exponențiale ale gradienților și ale pătratelor lor
- Pierderea pentru prognoze punctuale: eroarea pătratică medie (MSE)
  - Minimizată de media condiționată
- Pierderea pentru o cuantilă de nivel u: pierderea pinball

$$\rho_u(y, q) = (u - 1\{y < q\})(y - q)$$

  - y: valoarea realizată; q: prognoza cuantilei
  - Pierderea este asimetrică: valorile peste q costă u pe unitate, cele sub q costă $1 - u$
  - Media pierderii este minimă cînd q este chiar cuantila reală de nivel u

De la perceptronul multistrat la deep learning (3/3)

- Aproximarea universală (Hornik, Stinchcombe & White (1989))
  - O rețea suficient de largă aproximează orice funcție continuă
  - Nu spune nimic despre mărimea eșantionului necesar, constrîngerea decisivă pentru o singură serie financiară
  - Regularizare: oprire timpurie (*early stopping*), penalizarea mărimii ponderilor (*weight decay*), medii pe inițializări aleatoare (*seeds*)
  - Blocul de validare trebuie să urmeze în timp blocului de antrenare
- Premiul Nobel pentru Fizică 2024: Hopfield și Hinton
  - Metodele: LeCun, Bengio & Hinton (2015)
  - În finanțe, Gu, Kelly & Xiu (2020) obțin cîștiguri pe o secțiune transversală de mii de acțiuni, nu pe un singur indice
  - Fischer & Krauss (2018) aplică LSTM pe acțiunile din S&P 500, prognozînd care dintre ele depășesc a doua zi mediana secțiunii transversale



Geoffrey Hinton, laureat Nobel 2024

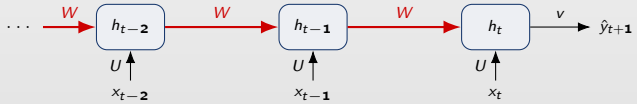
Foto: Arthur Petron (2024); CC BY-SA 4.0; Wikimedia Commons

RNN și LSTM (1/2)

- **Rețea neuronală recurentă** (RNN)
  - O stare ascunsă este actualizată la fiecare pas din valoarea ei anterioară și din noua intrare

$$h_t = \tanh\left(Wh_{t-1} + Ux_t + b\right), \qquad \hat{y}_{t+1} = v^\top h_t + c$$

  - h_t: starea ascunsă, adică memoria rețelei la momentul t, cu valori în $(-1, 1)$ datorită funcției tanh
  - x_t este intrarea de la momentul t (de exemplu randamentul de azi), iar $\hat{y}_{t+1}$ este prognoza
  - W, U, b: ponderile recurente, ponderile intrării și termenul liber
  - v, c: ponderile de ieșire și termenul liber al ieșirii
  - Aceleași ponderi la fiecare pas: desfășurată în timp, o rețea foarte adîncă, cu ponderi comune



RNN și LSTM (2/2)

- **Long short-term memory** (LSTM, Hochreiter & Schmidhuber (1997))
  - Porțile decid ce se uită, ce se scrie și ce se afișează

$$f_t, i_t, o_t = \sigma(W_\bullet x_t + U_\bullet h_{t-1} + b_\bullet), \quad g_t = \tanh(W_g x_t + U_g h_{t-1} + b_g)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g_t, \quad h_t = o_t \odot \tanh(c_t)$$

- Notațiile
  - $W_\bullet$ ponderează aici intrarea, iar $U_\bullet$ starea anterioară, cu literele inversate față de RNN; f_t, i_t, o_t: poarta de uitare, poarta de intrare și poarta de ieșire
  - Fiecare poartă are propriile ponderi $W_\bullet$, $U_\bullet$, $b_\bullet$, cu $\bullet \in \{f, i, o\}$
  - σ: sigmoida; o poartă apropiată de 0 blochează, una apropiată de 1 lasă să treacă
  - g_t: conținutul nou candidat
  - c_t: starea celulei, adică memoria pe termen lung
  - h_t: ieșirea celulei
  - $\odot$: produsul element cu element
- GRU (gated recurrent unit, unitate recurentă cu porți): o variantă mai ușoară, cu două porți

Dispariția gradientului: o majorare și un VAR(1) (1/2)

- Antrenarea are nevoie de efectul stărilor vechi asupra stării de azi
  - Propagarea înapoi în timp înmulțește cîte o matrice jacobiană la fiecare pas

$$\frac{\partial h_T}{\partial h_{T-k}} = \prod_{j=0}^{k-1} D_{T-j}\, W, \qquad D_t = \operatorname{diag}\left(\varphi'(W h_{t-1} + U x_t + b)\right)$$

  - $\partial h_T / \partial h_{T-k}$: matricea variațiilor lui h_T la o variație unitară a stării de acum k pași
  - D_t: matricea diagonală a pantelor φ' ale funcției de activare la pasul t
- Majorarea, din submultiplicativitate, $\|AB\|_2 \leq \|A\|_2 \|B\|_2$ (Pascanu, Mikolov & Bengio (2013))

$$\left\| \frac{\partial h_T}{\partial h_{T-k}} \right\|_2 \leq \left(\gamma_\varphi \|W\|_2\right)^k, \qquad \gamma_\varphi = \sup_z |\varphi'(z)|$$

  - $\|W\|_2$: norma spectrală, cel mai mare factor cu care W alungește un vector
  - γ_φ: cea mai mare pantă a funcției de activare ($= 1$ pentru tanh)
  - $\gamma_\varphi \|W\|_2 < 1$: gradientul scade geometric cu k, adică dispare
- Cazul scalar: cu $|w| = 0{,}9$ (Bengio, Simard & Frasconi (1994))
  - Gradientul scade la 0,12 după 20 de pași și la 0,005 după 50 de pași
- Remedii
  - Limitarea normei gradientului, împotriva exploziei
  - Porțile, ca în LSTM, împotriva dispariției

Dispariția gradientului: o majorare și un VAR(1) (2/2)

- Liniarizăm în jurul unui punct fix h^* al rețelei

$$h_t - h^* \approx A\,(h_{t-1} - h^*) + DUx_t, \qquad A = DW$$

  - Punctul fix: o stare care se reproduce cînd intrarea este zero
  - D: pantele funcției de activare în h^*
  - Starea ascunsă urmează deci un VAR(1) (autoregresie vectorială) cu matricea coeficienților A
- $\partial h_T / \partial h_{T-k} = A^k$ este răspunsul la impuls al acestui VAR la orizontul k
  - $\rho(A)$: raza spectrală, cea mai mare valoare proprie a lui A în valoare absolută
  - $\rho(A) < 1$: VAR-ul este stabil, iar $\|A^k\| \to 0$ geometric cînd $k \to \infty$
  - La orizont finit, $\|A^k\|$ poate crește temporar dacă A nu este normală ($AA^\top \neq A^\top A$)
  - $\rho(A) > 1$: gradienții explodează
- O RNN stabilă în jurul punctului fix este un model cu memorie scurtă, ca orice VAR staționar
  - Seminarul 14, exercițiul A4
  - Detalii în ▸ Anexă

Memoria LSTM: poarta de uitare

- Pe calea directă prin starea celulei, singurul factor multiplicativ este poarta de uitare
  - $\partial c_T / \partial c_{T-k} = \prod_{j=\mathbf{0}}^{k-\mathbf{1}} f_{T-j}$ pe drumul direct
  - $c_t = f_t \odot c_{t-\mathbf{1}} + i_t \odot g_t$: un AR(1) cu coeficient aleator
- Celulă scalară simplificată, cu poarta de uitare constantă f: un AR(1) cu coeficientul f
  - $f = \sigma(b_f)$ cînd intrările porții sînt nule; b_f: termenul ei liber
  - Cu intrarea $i_t g_t$ zgomot alb și inițializare staționară, $\mathrm{Corr}(c_t, c_{t-k}) = f^k$ scade geometric
  - Cu $b_f = 0$, poarta de uitare este $f = 0{,}5$, iar după 50 de pași rămîne $0{,}5^{\mathbf{50}} \approx 9 \times 10^{-\mathbf{16}}$ din memorie
  - Cu $b_f = 3$, poarta de uitare este $f = 0{,}953$, iar după 50 de pași rămîne $0{,}953^{\mathbf{50}} \approx 0{,}09$ din memorie
  - Timpul de înjumătățire este de 14 pași (k pentru care $f^k = \frac{\mathbf{1}}{\mathbf{2}}$, adică $k = \ln 0{,}5 / \ln f$)
- Memoria lungă scade hiperbolic: $\rho_k \sim C k^{\mathbf{2}d-\mathbf{1}}$
  - ρ_k: autocorelația de ordin k; $d \in (0, \frac{\mathbf{1}}{\mathbf{2}})$: parametrul de memorie; $C > 0$: o constantă
  - Randamentele absolute ale S&P 500: descreșterea geometrică ajustată dă $\rho_{\mathbf{150}} = 0{,}027$, față de 0,059 observată (▸ Anexă)



Sepp Hochreiter, co-inventator al LSTM

Foto: Eulenreich (2015); CC BY-SA 4.0; Wikimedia Commons

Atenția: queries, keys și values (1/2)

- **Întrebarea secțiunii**: cum poate o rețea să lege direct ziua de azi de o zi de acum 500 de zile?
  - Fără produsul de jacobiene care face gradientul unei RNN să dispară
- Fiecare pas de timp formează o interogare (query) (Vaswani et al. (2017))
  - O compară cu cheia (key) a fiecărui pas pe care are voie să-l vadă
  - Răspunsul este o medie ponderată a valorilor acestora

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \qquad Q = XW_Q,\ K = XW_K,\ V = XW_V$$

  - X: intrările, cîte un rînd pentru fiecare pas de timp
  - W_Q, W_K, W_V: matrice de ponderi învățate
  - Q, K, V: interogările, cheile și valorile

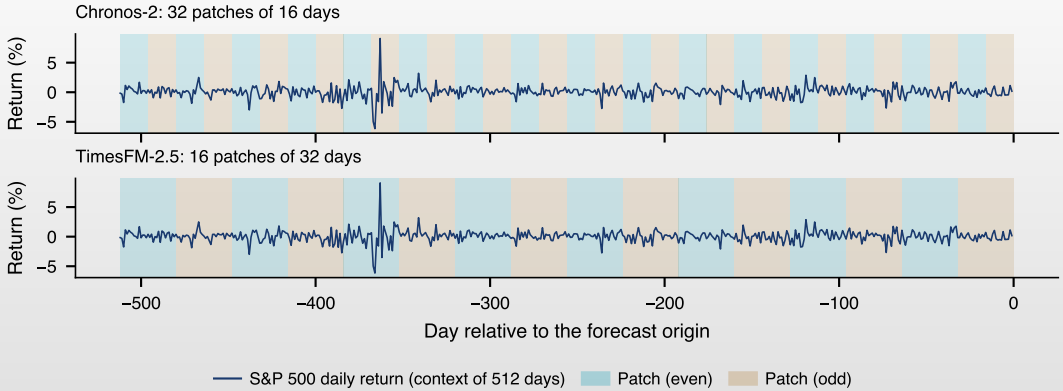
Atenția: queries, keys și values (2/2)

- $QK^\top$: scorul de similaritate pentru fiecare pereche de pași
- d_k este lungimea unui vector-cheie
  - Împărțirea la $\sqrt{d_k}$ împiedică saturarea funcției softmax
- softmax transformă fiecare rînd de scoruri $s_1, \ldots, s_n$ în ponderi pozitive cu suma 1

$$\text{softmax}(s)_j = \frac{e^{s_j}}{\sum_{i=1}^{n} e^{s_i}}$$

- Niciun produs de jacobiene în timp: orice pas trecut este legat direct de pasul curent
- Costul este $O(n^2)$: calculul crește cu pătratul numărului de intrări n

Împărțirea unui context de 512 zile în patch-uri



MFM_ch14_tokenization ▶ Interactiv

Poziții și patch-uri

- **Codificarea poziției**: atenția tratează intrările ca pe o mulțime, deci ordinea trebuie adăugată
  - Sinusoide în Transformerul original
  - Codificări rotative în Chronos-2
  - Codificările rotative (RoPE, rotary position embeddings): fiecare interogare și cheie este rotită cu un unghi proporțional cu poziția ei, deci scorul a doi pași depinde de distanța dintre ei, nu de locul lor în context
- **Împărțirea în patch-uri** (Nie et al. (2023)): un token este un bloc de observații consecutive
  - O funcție învățată transformă blocul într-un vector de lungime fixă (embedding)
  - 512 randamente zilnice împărțite în patch-uri de 16 zile dau 32 de tokeni și de 256 de ori mai puține scoruri de atenție (32^2 în loc de 512^2)
- Graficul: ultimele 512 randamente zilnice ale S&P 500
  - Împărțite așa cum le citesc Chronos-2 (16 zile) și TimesFM-2.5 (32 de zile)
  - Fiecare bloc colorat devine un token
- Modelele globale învață pe multe serii: DeepAR (Salinas et al. (2020)), N-BEATS (Oreshkin et al. (2020)), PatchTST (Nie et al. (2023))
  - Modelele fundaționale extind această idee la corpusuri foarte mari

MFM_ch14_tokenization

Modelul fundațional: definiție

- **Întrebarea secțiunii**: poate un model pre-antrenat pe alte serii să prognozeze o serie financiară nouă?
  - Nouă: una pe care nu a văzut-o la antrenare
- Modelul fundațional: o rețea mare, pre-antrenată o singură dată pe un corpus foarte larg
  - Apoi folosită pentru sarcini noi
  - Idee preluată de la modelele mari de limbaj (LLM, large language models)
- Avantajul pentru finanțe: un singur model pentru mii de serii
  - Fără estimare separată pentru fiecare serie
- Trei familii de modele
  - **Chronos** (Ansari et al. (2024)): valorile devin tokeni ai unui model de limbaj
  - **TimesFM** (Das et al. (2024)): un Transformer doar cu decodor, care citește seria pe patch-uri
  - **Moirai** (Woo et al. (2024)): orice număr de variabile și o ieșire de tip amestec de distribuții (Anexă); nu îl testăm, pentru că are licență necomercială

Prognoza zero-shot: ce primește și ce returnează modelul

- Prognoza **zero-shot**: parametrii $\hat{\theta}$ sînt fixați la pre-antrenare
  - La momentul T, modelul primește doar istoricul recent al seriei (**contextul**)

$$\left(\hat{q}_{u_1,T+h},\ldots,\hat{q}_{u_K,T+h}\right)_{h=1,\ldots,H} = F_{\hat{\theta}}\left(y_{T-C+1},\ldots,y_T\right)$$

  - $F_{\hat{\theta}}$: rețeaua pre-antrenată, aceeași pentru orice serie
  - C: lungimea contextului (aici 512 zile), iar H: orizontul de prognoză
  - $\hat{q}_{u_k,T+h}$: cuantila de nivel u_k a valorii de la momentul $T+h$, deci rezultatul este o grilă de cuantile, nu doar un punct
- Fine-tuning-ul (antrenare suplimentară pe seria-țintă) este posibil, dar nu îl folosim
  - Testăm chiar promisiunea „fără estimare"
- Riscul specific: seria-țintă poate face parte din corpusul de pre-antrenare (slide-ul despre scurgerea de informație, *leakage*)

Chronos: seriile de timp ca limbaj

- Ansari et al. (2024): valorile devin tokeni
  - Un model de limbaj (T5, Text-to-Text Transfer Transformer) este antrenat cu entropia încrucișată
- De ce tokeni: arhitectura unui model de limbaj rămîne neschimbată
  - Distribuția pe tokeni nu impune o formă parametrică (poate fi asimetrică sau cu mai multe moduri)
- Scalarea prin medie și cuantizarea

$$\tilde{x}_t = \frac{x_t}{s}, \quad s = \frac{1}{C}\sum_{i=\mathbf{1}}^{C} |x_i|, \qquad z_t = \arg\min_j |\tilde{x}_t - c_j|, \quad c_j = -15 + \frac{30\,(j-1)}{4\,092}$$

  - s: media valorilor absolute ale celor C valori din context
  - c_j, $j = 1, \ldots, 4\,093$: centre echidistante în $[-15, 15]$, la distanța $30/4\,092 = 0{,}0073$, plus tokeni speciali
  - z_t: tokenul valorii x_t, adică indicele centrului cel mai apropiat
  - Prognoza: se eșantionează tokeni viitori z, se transformă înapoi în valori $s\,c_z$ și se citesc cuantilele din eșantioane
- Corpusul de pre-antrenare
  - Seturi publice de date: energie, transport, comerț, vreme, web, finanțe și economie
  - Serii sintetice TSMixup: combinații convexe aleatoare ale unor serii reale
  - Serii sintetice KernelSynth: traiectorii ale unor procese gaussiene cu nuclee de covarianță combinate aleator
- Chronos (T5): de la 8 la 710 milioane de parametri
  - Chronos-Bolt, testat aici: de la 9 la 205 milioane

Cele trei modele pe care le testăm

Model	Parametri	Patch	Niveluri de cuantile	Data publicării ponderilor
Chronos-Bolt (small)	48 M	16	10%–90%	25.11.2024
TimesFM-2.5	231 M	32	medie, 10%–90%	01.10.2025
Chronos-2	119 M	16	21: 1%, 5%, 10%, . . . , 95%, 99%	30.10.2025

- **Chronos-Bolt**: patch-uri și estimarea directă a cuantilelor, în locul eșantionării de tokeni
- **Chronos-2** (Ansari et al. (2025)): scalare robustă, atenție în timp și între grupuri de serii înrudite
- **TimesFM** (Das et al. (2024)): pre-antrenat pe circa 10^{11} puncte
- Parametrii sînt numărați pe modelele încărcate
  - TimesFM-2.5 este publicat ca timesfm-2.5-200m
- Toate rulează pe un CPU (procesor central obișnuit), cu un context de 512 zile
- Datele de publicare a ponderilor delimitează ce date ar fi putut vedea modelele la antrenare
- Alte modele și corpusurile lor sînt prezentate în ▸ Anexă

De la o grilă de cuantile la VaR și ES (1/2)

- Modelul furnizează cuantile q_u la nivelurile $u_1 < \cdots < u_K$: $P(Y \leq q_u) = u$
  - Y: valoarea viitoare; K: numărul de niveluri
- De ce o grilă: din aceleași cuantile obținem prognoza punctuală, volatilitatea și riscul din coadă
  - Deci același model servește toate cele trei teste
- Prognoze punctuale din grilă
  - mediana $q_{0,5}$ sau media $\int_0^1 q_u \, du$, media tuturor cuantilelor
  - pentru orice funcție g cu medie finită: $E[g(Y)] = \int_0^1 g(q_u) \, du$, deoarece Y are aceeași distribuție ca q_U, cu U uniformă (de exemplu $g = \exp$ cînd Y este logaritmul unei variante)
  - Dacă g este crescătoare, $g(q_u)$ este și cuantila de nivel u a lui $g(Y)$
  - Verificare: Y uniformă pe $(-1;\ 1)$, $q_u = 2u - 1$, $g(y) = y^2$: $\int_0^1 (2u-1)^2 \, du = 1/3 = E[Y^2]$, deși $(2u-1)^2$ nu este o cuantilă a lui Y^2 (scade, apoi crește în u)

De la o grilă de cuantile la VaR și ES (2/2)

- Riscul de coadă la nivelul α (Capitolul 7), pentru un randament Y

$$\mathrm{VaR}_\alpha = -q_\alpha, \qquad \mathrm{ES}_\alpha = -\frac{1}{\alpha}\int_0^\alpha q_u \, du$$

- Sub cel mai mic nivel, grila nu oferă informație
  - Păstrarea lui q_u la $q_{0,01}$ subestimează ES
  - Interpolarea poate produce erori în ambele sensuri
  - Seminarul 14, exercițiile A1 și A7
- Chronos-Bolt și TimesFM se opresc la 10%
  - O cerere de „VaR 1%" returnează cuantila de 10%
- Regula din capitol: interpolare liniară între niveluri și q_u constant sub cel mai mic nivel
  - Exemplul rezolvat este în Aprofundări

Scurgerea de informație și contaminarea

- Scurgerea de informație: perioada de test a fost văzută la antrenare, deci testul nu mai este în afara eșantionului (*out-of-sample*)
- Perioada de test poate fi „văzută" în patru moduri
  - (i) aceeași serie este în corpusul de pre-antrenare
  - (ii) serii corelate conțin aceleași șocuri
  - (iii) dezvoltatorii aleg versiunile după repere publice (*benchmarks*) care se suprapun cu datele de test
  - (iv) deplasarea de anticipare (*look-ahead bias*), prin folosirea datelor viitoare: scalări, ferestre, seed-uri sau contexte alese cu date de test
- Corpusul Chronos include serii financiare (de exemplu cursuri de schimb zilnice 1990–2016)
  - Deci nu putem exclude ca seriile noastre să fi fost văzute la pre-antrenare
- Remediul pentru (i)–(iii): o **fereastră de după publicare**, de la 3 noiembrie 2025
  - O dată-limită comună după publicarea tuturor celor trei seturi de ponderi
  - Scurtă: circa 220 de zile de tranzacționare (320 de zile calendaristice pentru Bitcoin), deci putere mică, mai ales pentru VaR 1%
  - Fără contaminare a ponderilor doar dacă checkpoint-ul folosit precedă fereastra: revizia și amprenta ponderilor sînt fixate în cod (anexa despre corpusuri)
- (iv) este sub controlul nostru: scalările, ferestrele și estimările folosesc doar date trecute
  - Alegerea retrospectivă a contextului sau a hiperparametrilor rămîne posibilă și după publicare
  - Protocol formal în ▸ Anexă

Protocolul testului pentru randamente

- **Întrebarea**: prognozează vreun model randamentul zilnic de mîine mai bine decît prognoza zero?
- Perioada de test: S&P 500 și BET de la 04.01.2016, Bitcoin de la 01.01.2019, pînă la 18.09.2026
  - BET: indicele Bursei de Valori București; 2 693 de zile (S&P 500), 2 675 (BET) și 2 818 (Bitcoin)
  - Circa zece ani (opt pentru Bitcoin), inclusiv crahul din martie 2020 și anul 2022
  - Estimarea folosește și datele dinainte: seriile încep în 1990 (S&P 500), 1997 (BET) și 2014 (Bitcoin)
- De ce trei piețe
  - S&P 500: piața de acțiuni cea mai lichidă, unde un semnal simplu ar fi fost deja exploatat
  - BET: o piață mai mică și mai puțin lichidă, unde prețurile neactualizate (*stale*) pot crea predictibilitate
  - Bitcoin: tranzacționat continuu, cu o abatere standard zilnică de aproape trei ori mai mare (3,21%, față de 1,13% pentru S&P 500)
- Modele: media istorică, AR(1), LSTM, Chronos-2, Chronos-Bolt și TimesFM-2.5
  - AR(1): $\hat{r}_t = \hat{a} + \hat{b}\, r_{t-1}$, estimat prin cele mai mici pătrate pe ultimele 1 000 de zile, reestimat zilnic
  - Modelele fundaționale: mediana grilei, o prognoză punctuală pe care valorile extreme nu o mută; mediile din grilă sînt raportate separat
- Fiecare prognoză folosește doar date pînă la închiderea precedentă

Evaluarea prognozelor de randament (1/2)

- Reperul este prognoza zero, iar R^2 în afara eșantionului se calculează ca

$$R^2 = 1 - \frac{\sum_t (r_t - \hat{r}_t)^2}{\sum_t r_t^2}$$

  - r_t este randamentul realizat, iar $\hat{r}_t$ este prognoza din datele pînă la $t-1$
  - $R^2 > 0$: modelul are erori pătratice mai mici decît prognoza zero
- De ce zero și nu media istorică: media estimată adaugă zgomot de estimare
  - Zero este reperul folosit de Gu, Kelly & Xiu (2020)
  - Media istorică este reperul lui Campbell & Thompson (2008)
- Mediile din grilă ($\int_0^1 q_u \, du$) dau rezultate și mai slabe decît medianele
  - Pe S&P 500, R^2 este $-0{,}77\%$ pentru Chronos-2 și $-3{,}42\%$ pentru TimesFM-2.5

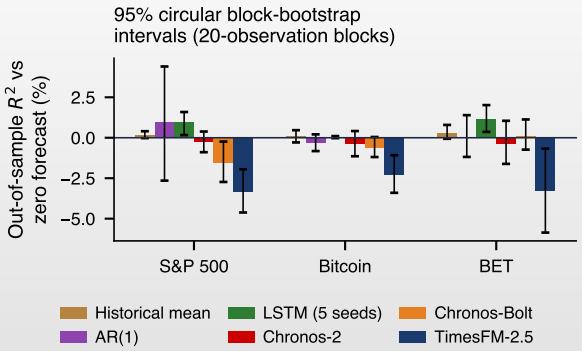
Evaluarea prognozelor de randament (2/2)

- Interval de încredere (CI, confidence interval) bootstrap pe blocuri pentru R^2
  - Reconstruim o perioadă de test de aceeași lungime din blocuri de 20 de zile consecutive, extrase cu întoarcere
  - De 1 000 de ori, apoi păstrăm percentilele de 2,5% și 97,5%
- Blocurile păstrează episoadele de volatilitate
- Testul Diebold–Mariano (DM, Diebold & Mariano (1995)) pe erorile pătratice

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{\hat{V}_{NW}/n}}, \qquad d_t = (r_t - \hat{r}_t)^2 - r_t^2$$

  - $\bar{d}$: media lui d_t pe cele n prognoze; $\mathrm{DM} > 0$: modelul are erori mai mari decît prognoza zero
  - $\hat{V}_{NW}$: varianța de termen lung a lui d_t, estimată Newey–West, un estimator HAC (heteroskedasticity and autocorrelation consistent, robust la heteroscedasticitate și autocorelație)
  - Sub H_0: $E[d_t] = 0$, DM are aproximativ distribuția Normală standard
  - Aproximarea cere ca d_t să aibă varianță de termen lung finită și nenulă și dependență slabă
  - Aici $d_t = \hat{r}_t^2 - 2 r_t \hat{r}_t$: contează momentele lui $r_t \hat{r}_t$, nu neapărat $E[r_t^4]$

R^2 în afara eșantionului pentru randamente



- R^2 față de prognoza zero (%), CI bootstrap de 95% (blocuri de 20 de zile)
  - Tabelul complet: ▸ Anexă
- S&P 500 și Bitcoin: niciun model nu bate prognoza zero (DM bilateral)
- Pe S&P 500, media istorică trece testul Clark–West pentru modele imbricate ($p = 0{,}007$)
  - Dar nu și testul comun SPA ($p = 0{,}130$); ambele teste sînt explicate mai jos
- TimesFM-2.5: semnificativ **mai slab** decît zero
  - S&P 500: R^2 −3,35%, DM $p = 0{,}001$
  - Abaterea standard a prognozei 0,141%, triplă față de LSTM
- BET: LSTM are R^2 +1,15% (DM $p = 0{,}014$, bilateral)
  - Un semnal mic, testat mai jos pentru imbricare, teste multiple și prețuri neactualizate

MFM_ch14_returns ▸ Interactiv

Modele imbricate: testul Clark–West față de DM

- **Problema**: prognoza zero este imbricată în media istorică și în AR(1)
  - Sub H_0, modelul mai mare doar adaugă zgomot de estimare, deci DM respinge prea rar
- Clark–West (Clark & West (2007)): adăugăm înapoi termenul de zgomot
  - Testăm media diferenței cu un test t HAC unilateral

$$\tilde{d}_t = e_{\mathbf{0},t}^2 - \left[e_{\mathbf{1},t}^2 - (\hat{r}_{\mathbf{0},t} - \hat{r}_{\mathbf{1},t})^2\right]$$

  - $\hat{r}_{\mathbf{0},t}$ și $\hat{r}_{\mathbf{1},t}$: prognoza zero (modelul mic) și prognoza modelului mare; $e_{\mathbf{0},t}$, $e_{\mathbf{1},t}$: erorile lor, $r_t - \hat{r}_{\cdot,t}$
  - $(\hat{r}_{\mathbf{0},t} - \hat{r}_{\mathbf{1},t})^2$: costul estimării unor coeficienți nuli sub H_0
  - H_0: $E[\tilde{d}_t] = 0$; un t mare și pozitiv arată că modelul mare adaugă informație
- Valori p (*p-values*) unilaterale, testul DM față de testul Clark–West
  - ambele teste cer variantă de termen lung finită și nenulă a diferențialului (pentru CW față de prognoza zero, $\tilde{d}_t = 2r_t\hat{r}_t$)
  - media istorică pe S&P 500: DM 0,048, Clark–West 0,007
  - media istorică pe BET: DM 0,041, Clark–West 0,003
  - AR(1) pe S&P 500: DM 0,320, Clark–West 0,055 (AR(1) pe BET și ferestrele mobile, *rolling*: ▸ Anexă)
- Interpretare: corecția reduce mult valoarea p a mediei istorice pe ambele piețe
  - Dar cîștigul rămîne mic (R^2 de 0,15% pe S&P 500 și 0,30% pe BET)

Teste multiple: Bonferroni și Holm

- 6 modele $\times$ 3 piețe $= 18$ teste
  - Cea mai mică valoare p DM unilaterală, 0,007, o are LSTM pe BET
- Problema: cel puțin un $p < 0{,}05$ ar apărea cu probabilitatea $1 - 0{,}95^{18} = 0{,}60$
  - Dacă niciun model nu ar avea semnal și testele ar fi independente
- Bonferroni: cu M teste, respingem H_0 doar dacă $p < 0{,}05/M$
  - Aici, pragul este 0,0028
- Holm: aceeași idee, aplicată pas cu pas, începînd cu cel mai mic p
- Ambele controlează probabilitatea de a face cel puțin o respingere falsă
- Valoarea p ajustată Holm pe cele 18 teste este 0,130: nesemnificativă

Teste multiple: Romano–Wolf și SPA

- Pentru cele șase modele pe BET: teste care folosesc corelația dintre modele
  - Romano–Wolf (Romano & Wolf (2005)): valoarea p ajustată prin procedura descendentă (step-down) a LSTM pe BET este 0,016
  - Testul SPA (superior predictive ability, Hansen (2005)), cu H_0: niciun model nu depășește prognoza zero
  - Valorile p SPA: BET 0,0185, S&P 500 0,130, Bitcoin 0,354
  - SPA și Romano–Wolf cer variante de termen lung finite și nenule ale diferențialelor și un bootstrap staționar valid
- Afirmația „un model depășește prognoza zero pe BET" este valabilă dacă BET a fost ales dinainte ca familie de teste
- Corectate Holm pentru cele trei piețe, valorile p SPA dau pentru BET $3 \times 0{,}0185 = 0{,}056$
  - Chiar peste pragul de 5%
- Afirmația „o anumită comparație dintre cele 18 este semnificativă" nu este valabilă
  - Stabiliți familia de teste înainte de a examina rezultatele
- Tipul de dovadă: comparație între proceduri (arată care metodă funcționează mai bine pe aceleași date, nu de ce)
  - Pe S&P 500, o singură lună, martie 2020, determină clasamentul (Aprofundări)
- Cum funcționează Holm, Romano–Wolf și SPA, cu un exemplu cu trei teste, în ▸ Anexă

Testul 2: întrebarea și metoda

- **Întrebarea**: prognozează modelele fundaționale zero-shot varianța zilnică a S&P 500 mai bine decît log-HAR?
  - Log-HAR: reperul econometric standard
- De ce SPY: ETF-ul (*Exchange-Traded Fund*, fond tranzacționat la bursă) pe S&P 500 este foarte lichid și are prețuri la 5 minute
  - Din ele se calculează varianța realizată
- De ce log-HAR: aproximează memoria lungă cu trei medii și doar patru coeficienți (Corsi (2009))
- De ce QLIKE (Patton (2011)): ordonează corect prognozele
  - Chiar cînd varianța adevărată este înlocuită de o aproximare zgomotoasă
- Orizontul: o zi, același cu al VaR-ului din Testul 3

Varianța realizată, modelul HAR și pierderea QLIKE (1/2)

- **Varianța realizată** (RV, Andersen et al. (2003))
  - Suma pătratelor randamentelor logaritmice la 5 minute din ziua de tranzacționare

$$\mathrm{RV}_t = \sum_{j=\mathbf{1}}^{\mathbf{78}} r_{t,j}^{\mathbf{2}}$$

  - $r_{t,j}$: randamentul logaritmic al intervalului j de 5 minute din ziua t
  - $r_{t,\mathbf{1}} = \ln(C_{t,\mathbf{1}}/O_{t,\mathbf{1}})$: deschiderea și închiderea barei de 09:30; apoi $r_{t,j} = \ln(C_{t,j}/C_{t,j-\mathbf{1}})$
  - Sesiunea regulată a SPY (09:30–16:00, ora New York-ului) are 78 de intervale de 5 minute; zilele fără grila completă sînt eliminate
  - Volatilitatea medie a SPY: 12,1% pe an
- **HAR** (*Heterogeneous AutoRegressive*, model autoregresiv heterogen) (Corsi (2009))
  - RV de azi regresată pe cea de ieri, pe media ultimei săptămîni și pe media ultimei luni

$$\mathrm{RV}_t = \beta_{\mathbf{0}} + \beta_d \mathrm{RV}_{t-\mathbf{1}} + \beta_w \overline{\mathrm{RV}}_{t-\mathbf{5}:t-\mathbf{1}} + \beta_m \overline{\mathrm{RV}}_{t-\mathbf{22}:t-\mathbf{1}} + \varepsilon_t$$

  - $\overline{\mathrm{RV}}_{t-\mathbf{5}:t-\mathbf{1}}$, $\overline{\mathrm{RV}}_{t-\mathbf{22}:t-\mathbf{1}}$: mediile pe ultimele 5 și 22 de zile de tranzacționare
  - β_d, β_w, β_m: ponderile zilnică, săptămînală și lunară

Varianța realizată, modelul HAR și pierderea QLIKE (2/2)

- log-HAR: aceeași regresie, estimată pe $\log \mathrm{RV}$, cu prognoza $\exp(\hat{y} + \hat{s}^2/2)$
  - $\hat{y}$: valoarea ajustată a log RV
  - $\hat{s}^2$: varianța reziduurilor din regresia pe logaritmi
  - Factorul $e^{\hat{s}^2/2}$ transformă prognoza logaritmică în media lui RV, dacă eroarea logaritmică are condiționat distribuția Normală
- Pierderea **QLIKE** a unei prognoze de varianță $\hat{h}_t$ (Patton (2011))
  - QLIKE: quasi-likelihood, de tip cvasi-verosimilitate

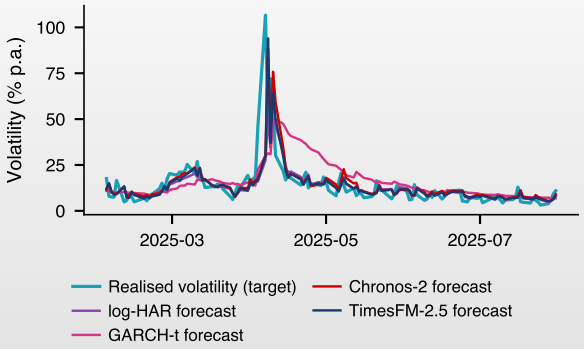
$$L = \frac{\mathrm{RV}_t}{\hat{h}_t} - \log \frac{\mathrm{RV}_t}{\hat{h}_t} - 1$$

  - $L \geq 0$, iar $L = 0$ cînd $\hat{h}_t = \mathrm{RV}_t$
  - Subestimările costă mai mult decît supraestimările de aceeași mărime
  - Media pierderii QLIKE este minimizată de media condiționată
  - QLIKE ordonează prognozele la fel ca varianța adevărată σ_t^2 cînd RV este o aproximare condiționat nedeplasată, $E[\mathrm{RV}_t \mid \mathcal{F}_{t-1}] = \sigma_t^2$, cu $\mathcal{F}_{t-1}$ informația pînă la $t-1$ (la fel și MSE)

Protocolul testului pentru volatilitate

- Varianța realizată a SPY, prognoze la o zi
  - De la 11.10.2022 pînă la 18.09.2026 (979 de zile)
  - HAR și log-HAR pe o fereastră mobilă de 500 de zile
  - Mersul aleator (*random walk*): $\hat{h}_t = \mathrm{RV}_{t-1}$
  - GARCH-t (GARCH(1,1) cu inovații Student-t) pe randamentele zilnice ale SPY, rescalat în unități RV prin raportul dintre media RV și media randamentelor la pătrat pe 500 de zile
  - LSTM pe ultimele 22 de valori ale log RV, reantrenat la fiecare 125 de zile pe 500 de zile (stare ascunsă 8, trei seed-uri)
- Modelele fundaționale pe $\log \mathrm{RV}$ (context 512): prognoza $\hat{h}_t = \int_0^1 \exp(q_u)\, du$ pe grila de cuantile
  - q_u sînt cuantilele log RV, $\exp(q_u)$ cele ale RV
  - Media lor este media RV
  - Pe grilă: regula trapezului pe nivelurile furnizate, cu q_u păstrat la valoarea de la cel mai mic nivel sub acesta și la valoarea de la cel mai mare nivel deasupra lui
  - Aceeași regulă pentru toate cele trei modele, fără ieșirea de medie a TimesFM
  - Fereastra de după publicare: 218 zile de la 3 noiembrie 2025 (S&P 500 are 220: ședințele scurtate din 28.11 și 24.12.2025 nu au grila completă de 5 minute)

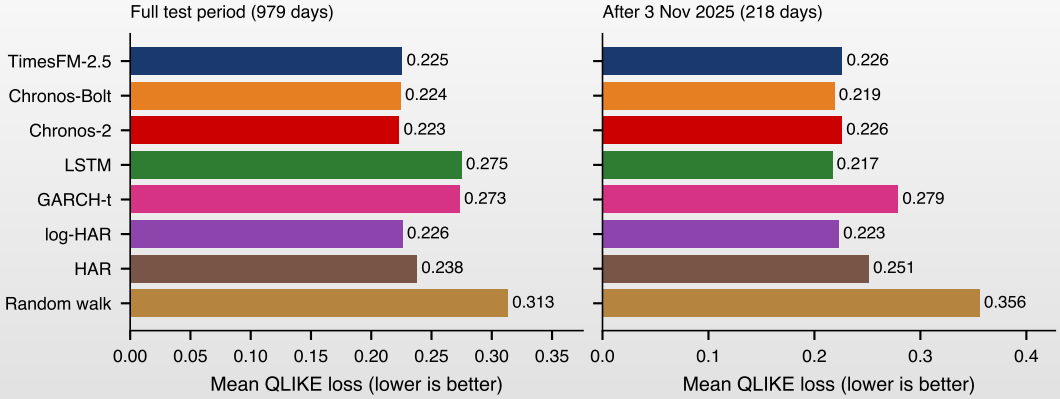
Prognozele în timpul șocului din aprilie 2025



- Volatilitatea realizată a SPY (anualizată) și prognozele la o zi, februarie–iulie 2025
  - Vîrful: 107% pe an la 07.04.2025, după anunțurile tarifare ale Statelor Unite
- Niciun model nu anticipează saltul din 04.04.2025 (46% pe an)
  - Prognozele pentru 07.04.2025 urcă la 27–30% pe an
  - Toate subestimează vîrful de 107%
  - Prognozele la o zi extrapolează trecutul, iar șocurile provin din știri
- Modelele fundaționale și log-HAR revin cu viteze asemănătoare
- GARCH-t reacționează la randamentele zilnice și revine mai lent

MFM_ch14_realised_vol ▶ Interactiv

Pierderile QLIKE ale celor opt modele de volatilitate (1/3)



MFM_ch14_realised_vol ▶ Interactiv

Pierderile QLIKE ale celor opt modele de volatilitate (2/3)

- Panoul din stînga arată toată perioada de test (979 de zile)
  - Cel din dreapta: doar zilele de după 3 noiembrie 2025 (218 zile)
  - Barele sînt pierderile QLIKE medii
- Pierderi apropiate: Chronos-2 (QLIKE 0,223), Chronos-Bolt, TimesFM-2.5 și log-HAR (QLIKE 0,226)
  - Testele DM nu arată diferențe semnificative ($p = 0{,}539$ pentru Chronos-2)
  - DM și MCS cer momente finite ale diferențialului $d_t = \mathrm{RV}_t(1/h_{A,t} - 1/h_{B,t}) + \ln(h_{A,t}/h_{B,t})$, deci și prognoze care nu se apropie de zero
- Mulțimea de încredere a modelelor de 90% (MCS, model confidence set, Hansen, Lunde & Nason (2011))
  - Elimină pe rînd cel mai slab model, pînă cînd egalitatea pierderilor medii nu mai este respinsă
- MCS păstrează șase din cele opt modele: sînt eliminate doar mersul aleator și GARCH-t
- HAR în niveluri supraestimează (media prognozei / media RV = 1,35)
- După publicare, LSTM-ul mic este cel mai bun (QLIKE 0,217): clasamentele sînt instabile
- Tabelul complet, cu valorile p DM și MCS, în ▸ Anexă
- Tipul de dovadă: comparație între proceduri

MFM_ch14_realised_vol

Pierderile QLIKE ale celor opt modele de volatilitate (3/3)

- Testul Giacomini–White (Giacomini & White (2006))
  - Întrebarea: poate fi prognozată diferența pierderilor d_t din informația de ieri?

$$\mathrm{GW} = n\,\bar{Z}^{\top}\hat{\Omega}^{-1}\bar{Z} \;\to\; \chi^2_2, \qquad Z_t = (1,\; d_{t-1})^{\top} d_t$$

  - d_t: pierderea QLIKE a modelului minus cea a log-HAR
  - Instrumentele 1 și d_{t-1}, cunoscute la $t-1$; H_0: $E[d_t] = 0$ și $E[d_{t-1}d_t] = 0$ simultan
  - $\bar{Z}$: media lui Z_t; n: numărul de prognoze
  - $\hat{\Omega}$: varianța de termen lung a lui Z_t (HAC, nucleu Bartlett, 6 lag-uri), pentru că $E[Z_t] = 0$ nu exclude autocorelația lui Z_t
- Rezultat: $p = 0{,}828$ pentru Chronos-2 și 0,812 pentru TimesFM-2.5
  - Deci nicio diferență condiționată
  - Testul cere ca Z_t să aibă varianță de termen lung finită și nesingulară
- Lipsa unei diferențe detectate nu dovedește egalitatea
  - Puterea testelor este discutată în secțiunea AI

Mediana față de medie sub QLIKE

Chronos-2 și TimesFM-2.5 pe log RV	QLIKE	Media prognozei / media RV
Chronos-2, $\int_0^1 \exp(q_u)\,du$	0,223	1,02
Chronos-2, $\exp(q_{0,5})$	0,249	0,80
TimesFM-2.5, $\int_0^1 \exp(q_u)\,du$	0,225	0,94
TimesFM-2.5, $\exp(q_{0,5})$	0,250	0,80

- Rezultat teoretic (Gneiting (2011)): QLIKE este consistentă pentru medie
  - Prognoza h care minimizează media pierderii, dată fiind informația $\mathcal{F}$, este media condiționată, $\arg\min_h E[L(\mathrm{RV}, h) \mid \mathcal{F}] = E[\mathrm{RV} \mid \mathcal{F}]$
  - Transformările monotone păstrează cuantilele: exp(mediana lui $\log \mathrm{RV}$) este mediana lui RV, elicitată de eroarea absolută (o minimizează în medie), nu de QLIKE
  - Dacă RV este lognormală, cu s^2 varianța lui log RV, atunci mediana/media $= e^{-s^2/2}$
  - Varianta cu mediana este semnificativ mai slabă (statistica DM +3,53 pentru Chronos-2) și mai slabă decît log-HAR (QLIKE 0,226)

Protocolul testului pentru riscul de coadă (1/2): întrebarea și țintele

- **Întrebarea**: dau modelele fundaționale prognoze VaR 1% și ES 2,5% mai bune decît FHS?
  - Fie direct, fie ca filtre de volatilitate
- Țintele: VaR_α și ES_α, calculate din cuantilele $q_u(r_{t+1} \mid \mathcal{F}_t)$
  - Slide-ul De la o grilă de cuantile la VaR și ES
  - $q_u(r_{t+1} \mid \mathcal{F}_t)$: cuantila de nivel u a randamentului de mîine, dată fiind informația de azi $\mathcal{F}_t$
  - VaR 1%: nivelul backtesting-ului din reglementarea Basel (teste de acoperire)
  - Perechea (VaR 2,5%, ES 2,5%): nivelul ES din standardul FRTB (Fundamental Review of the Trading Book)
- Modele clasice, fereastră de 1 000 de zile, reestimate la fiecare 20 de zile
  - Simularea istorică (HS), GARCH(1,1)-t, simularea istorică filtrată (FHS, McNeil & Frey (2000))
- De ce FHS ca reper: combină volatilitatea condiționată GARCH cu forma empirică a cozilor
  - Fără a presupune o distribuție

Protocolul testului pentru riscul de coadă (2/2): modelele fundaționale

- **Direct**: cuantilele Chronos-2 ale randamentului de mîine (context: 512 randamente)
- **Hibrid**: scala $\hat{\sigma}_t$ este prognoza lui $|r_t|$
  - Media din grilă pentru Chronos-2, mediana pentru TimesFM
  - Modelul fundațional citește seria $|r|$: contextul este format din cele 512 randamente absolute dinaintea zilei t
  - Apoi FHS pe $r_t/\hat{\sigma}_t$: cuantila empirică a randamentelor standardizate, înmulțită cu $\hat{\sigma}_t$
- Mediana sau media contează puțin aici, spre deosebire de Testul 2
  - FHS nu depinde de scală

$$c\hat{\sigma}_t \times q_\alpha\big(r/(c\hat{\sigma})\big) = \hat{\sigma}_t \times q_\alpha\big(r/\hat{\sigma}\big)$$

- Un factor constant în $\hat{\sigma}$ se anulează
  - De exemplu un raport stabil mediană/medie
  - Doar un factor variabil în timp schimbă VaR-ul

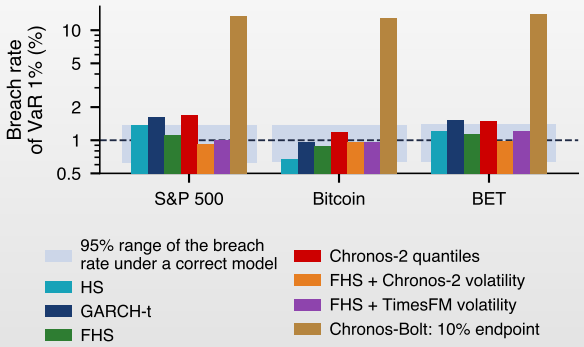
Evaluarea perechilor (VaR, ES): pierderea FZ0

- O caracteristică a distribuției este elicitabilă
  - Dacă o pierdere este minimizată, în medie, de valoarea ei adevărată
- ES singur nu este elicitabil, dar perechea (VaR, ES) este (Fissler & Ziegel (2016))
- Pierderea FZ0 a unei perechi de prognoze, cu pierderea zilnică $L_t = -r_t$
  - FZ0: funcția de pierdere Fissler–Ziegel, varianta omogenă de grad zero (Patton, Ziegel & Chen (2019))

$$\mathrm{FZ0} = \frac{1\{L_t > v\}(L_t - v)}{\alpha\, e} + \frac{v}{e} + \log e - 1$$

  - v și $e > 0$: prognozele VaR și ES
  - α: nivelul cozii
  - $1\{L_t > v\}$: egal cu 1 într-o zi cu depășire
  - O valoare mai mică este mai bună: primul termen penalizează depășirile, $\log e$ penalizează prognozele prea largi
- Modelele se compară prin teste DM pe diferențele FZ0 și prin MCS, ca în Testul 2

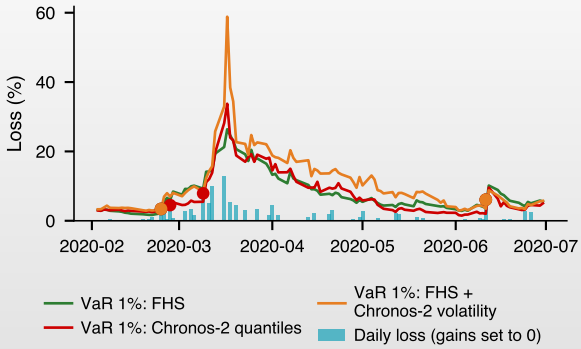
Ratele de depășire ale VaR 1%



- Ponderea zilelor cu o pierdere peste VaR 1% (scară logaritmică)
  - Banda albastru deschis este intervalul de 95% pentru un model corect
  - Kupiec (Capitolul 8): testul raportului de verosimilitate (LR) că rata depășirilor este α, cu distribuția χ^2_1 sub H_0
- Chronos-2 direct: prea multe depășiri pe S&P 500 (1,67%, p Kupiec = 0,001) și pe BET (1,50%)
- Chronos-Bolt, cea mai mică cuantilă (10%) folosită drept „VaR 1%": 13,5%
- Hibrizii: rate de depășire apropiate de 1% pe toate cele trei piețe
  - Doar acoperirea medie
  - Tabelul complet în ▸ Anexă

MFM_ch14_tail_risk ▶ Interactiv

Crahul COVID-19: VaR 1% pentru trei modele



- S&P 500, februarie–iunie 2020 (104 zile): pierderile zilnice și prognozele VaR 1%
  - Punctele marchează depășirile
- Numărul de depășiri: FHS 2, Chronos-2 direct 4, FHS cu volatilitatea Chronos-2 2, HS 12
- Toate modelele sînt depășite în primele zile ale crahului
- Chronos-2 reacționează rapid la context, deci cuantilele lui se lărgesc repede
- Hibridul reacționează excesiv: VaR 1% de 58,9% la 17.03.2020
  - Față de 26,4% pentru FHS și o pierdere maximă de 12,8% (Anexă)

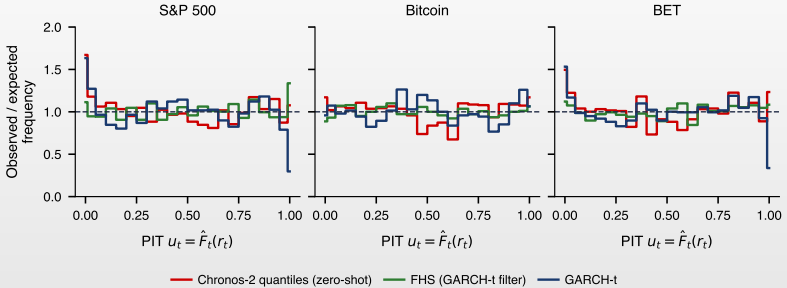
MFM_ch14_tail_risk ▶ Interactiv

Backtesting dincolo de Kupiec: gruparea depășirilor și testul DQ

p: indep. / DQ	HS	GARCH-t	FHS	Chronos-2 direct	FHS + Chronos-2	FHS + TimesFM
S&P 500	0,001 / <0,001	0,205 / 0,001	0,045 / 0,021	0,223 / <0,001	0,021 / 0,001	0,029 / <0,001
Bitcoin	0,611 / 0,760	0,261 / 0,463	0,220 / 0,092	0,404 / 0,034	0,026 / 0,004	0,261 / 0,182
BET	0,399 / <0,001	0,258 / 0,082	0,409 / 0,650	0,270 / 0,026	0,475 / <0,001	0,379 / <0,001

- Independența (Christoffersen (1998)): test LR, $\sim \chi^2_1$
  - O depășire azi nu schimbă șansa unei depășiri mîine (lanț Markov de ordinul 1)
- Acoperirea condiționată: Kupiec + independență, $\sim \chi^2_2$
- Cuantila dinamică (DQ, Engle & Manganelli (2004)): $\mathrm{Hit}_t = \beta_0 + \sum_{j=1}^{4} \beta_j \mathrm{Hit}_{t-j} + \beta_5 \mathrm{VaR}_t + u_t$
  - Cu $\mathrm{Hit}_t = I_t - \alpha$ și $I_t = 1$ într-o zi cu depășire
  - Statistica Wald testează dacă toți cei șase coeficienți sînt zero, $\sim \chi^2_6$
- S&P 500: hibrizii au rata corectă de depășire, dar sînt respinși de testele de independență și DQ
  - Testul DQ respinge toate modelele la 5%
  - BET: FHS trece ambele teste, iar hibrizii nu trec testul DQ
  - Calibrarea medie nu implică calibrarea condiționată

Transformarea integrală de probabilitate (PIT) și calibrarea (1/2)



- PIT (*Probability Integral Transform*, transformarea integrală de probabilitate) $u_t = \hat{F}_t(r_t)$: probabilitatea prognozată a unui randament mai mic sau egal cu cel realizat (Diebold, Gunther & Tay (1998))
  - Dacă funcția de repartiție predictivă $\hat{F}_t$ este corectă, valorile u_t sînt i.i.d. uniforme pe $(0, 1)$
- Barele: frecvența observată raportată la cea așteptată în cele 22 de intervale ale grilei Chronos-2
  - O bară peste 1 arată prea multe rezultate în acea zonă

MFM_ch14_calibration ▶ Interactiv

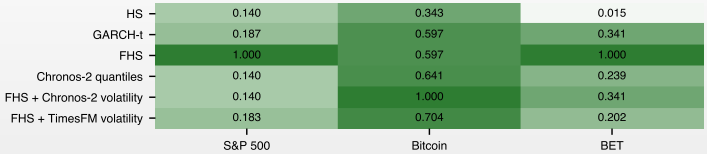
Transformarea integrală de probabilitate (PIT) și calibrarea (2/2)

- Testul Berkowitz (Berkowitz (2001)): transformăm $z_t = \Phi^{-1}(u_t)$
  - Variabila transformată are distribuția Normală standard dacă prognoza este corectă
- Testăm $\mu = 0$ și $\sigma = 1$ pentru z_t, cenzurat în afara intervalului $[1\%, 99\%]$
  - Φ^{-1}: funcția cuantilă a distribuției Normale standard
  - μ și σ: media și abaterea standard a lui z_t, iar un $\hat{\sigma} > 1$ indică cozi prea subțiri în prognoză
  - Chronos-2 este respins pe S&P 500 ($p = 0{,}019$, $\hat{\sigma} = 1{,}03$) și pe BET, iar FHS nu este respins pe nicio piață
- Scorul cuantilic (QS, quantile score): pierderea pinball medie pe cele K niveluri u_k ale grilei
  - Un analog pe grilă al CRPS (continuous ranked probability score, Gneiting & Raftery (2007))

$$\mathrm{QS} = \frac{2}{K} \sum_{k=1}^{K} \left(1\{r_t < q_{u_k}\} - u_k\right)\left(q_{u_k} - r_t\right)$$

  - q_{u_k}: cuantila prognozată la nivelul u_k, iar un QS mai mic este mai bun
  - Pe S&P 500 și BET, Chronos-2 este mai slab decît FHS atît în centrul distribuției, cît și în cozi
  - Statistica DM pe S&P 500: +2,51 în centru, +3,65 în cozi
  - Pe Bitcoin diferența nu este semnificativă (tabelul complet: Anexă)

Compararea prognozelor (VaR 2,5%, ES 2,5%): MCS



- Valorile p MCS de 90% (Hansen, Lunde & Nason (2011)) pentru pierderea FZ0 (verde mai închis: valori mai mari)
  - Singurul model eliminat ($p < 0{,}10$) este HS pe BET ($p = 0{,}015$, celula aproape albă)
  - Tabelul FZ0 complet este în ▸ Anexă
- S&P 500: FHS are cea mai mică pierdere FZ0
  - Hibrizii au rata corectă de depășire, dar un VaR mai mare
  - Media VaR 1% în cele 2 693 de zile de test: 3,20% pentru hibridul Chronos-2, față de 2,78% pentru FHS
  - Gruparea depășirilor nu îi deosebește: 2 depășiri consecutive pentru FHS și pentru fiecare hibrid
- Familia fixată dinainte: cele cinci teste DM față de FHS pe fiecare piață, cu corecția Holm
  - S&P 500: FHS este superior hibridului Chronos-2 (DM +2,65, p Holm = 0,032)
  - BET: DM +2,36, p Holm = 0,054, deci superioritatea nu este stabilită la 5%
  - Pe familia celor 15 teste (trei piețe): p Holm = 0,096, respectiv 0,180
- Tipul de dovadă: comparație între proceduri

MFM_ch14_tail_risk ▸ Interactiv

După publicare: puține depășiri, putere mică

Depășiri după 3.11.2025	HS	GARCH-t	FHS	Cuantile Chronos-2	FHS + vol. Chronos-2	FHS + vol. TimesFM
S&P 500 (220 de zile)	0	2	2	2	2	2
Bitcoin (320 de zile)	3	3	4	4	3	4
BET (217 zile)	4	4	4	5	1	2

- Depășiri așteptate: 2,2 (S&P 500), 3,2 (Bitcoin), 2,2 (BET)
  - Ponderile folosite au fost publicate înainte de 3.11.2025 (revizie Hugging Face fixată, amprentă SHA-256), dar testele au putere mică
  - Regula Kupiec cu pragul asimptotic χ^2_1 (nivel nominal 5%), cu puterea calculată exact din distribuția binomială, cînd rata reală de depășire este 1,67% în loc de 1%
  - Puterea (probabilitatea de a respinge modelul): 0,19 la $T = 220$ de zile, 0,50 la $T = 1\,000$ de zile, 0,87 la $T = 2\,693$ de zile
  - Nivelul efectiv al testului la $T = 220$ (rata de respingere cînd modelul este corect): 0,13
  - Puterea atinge 80% prima dată la circa 2 130 de zile (8,4 ani) și rămîne peste acest prag de la 2 320 de zile
  - Dacă rata reală este 2%, puterea la $T = 220$ este 0,29 (Seminarul 14, exercițiile A2 și A9)
- Putere mică nu înseamnă putere zero
  - HS are 0 depășiri pe S&P 500
  - Zero depășiri ($x = 0$) cad în regiunea de respingere Kupiec la $T = 220$, deci HS este respins
- O nerespingere nu dovedește calibrarea

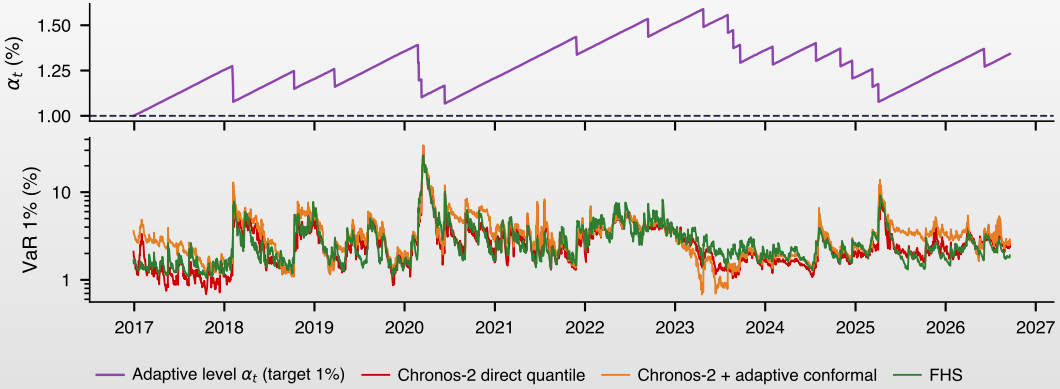
Recalibrarea cuantilelor: inferența conformală adaptivă

- Ideea: deplasăm cuantila modelului cu eroarea făcută recent
  - Și adaptăm nivelul după fiecare depășire
- Scorul este $E_t = \hat{q}_{\alpha,t} - r_t$, iar cuantila corectată este $\hat{q}_{\alpha,t} - \hat{Q}_{\mathbf{1}-\alpha_t}$
  - $\hat{q}_{\alpha,t}$: cuantila de nivel α a modelului pentru ziua t
  - $E_t > 0$: randamentul a scăzut sub această cuantilă
  - $\hat{Q}_{\mathbf{1}-\alpha_t}$: cuantila empirică de nivel $(1-\alpha_t)$ a ultimelor 250 de scoruri
- Nivelul adaptiv al inferenței conformale adaptive (ACI, *adaptive conformal inference*, Gibbs & Candès (2021))

$$\alpha_{t+\mathbf{1}} = \alpha_t + \gamma(\alpha - \mathrm{err}_t), \qquad \mathrm{err}_t = 1\{r_t < \hat{q}_{\alpha,t} - \hat{Q}_{\mathbf{1}-\alpha_t}\}, \qquad \alpha_{\mathbf{1}} = \alpha$$

  - α: nivelul-țintă (1%), iar $\mathrm{err}_t = 1$ într-o zi cu depășire
  - $\gamma > 0$: pasul de ajustare, aici 0,001
  - Cu $\gamma = 0$: calibrare pe fereastră mobilă cu nivel fix (corecția se recalculează zilnic din ultimele 250 de scoruri), nu split conformal standard
  - După o depășire α_t scade (următorul VaR este mai larg); după o zi calmă crește încet

VaR 1% conformal adaptiv pe S&P 500 (1/2)



MFM_ch14_calibration ▶ Interactiv

VaR 1% conformal adaptiv pe S&P 500 (2/2)

- Panoul de sus: nivelul adaptiv α_t (ținta 1%), care rămîne mai ales peste 1%
  - Deplasarea cu scorul singură ar lărgi prea mult VaR-ul
  - Cu $\gamma = 0$ (calibrare pe fereastră mobilă cu nivel fix): 16 depășiri, VaR 1% mediu 3,58%
  - Deci ACI relaxează nivelul
- Panoul de jos: VaR 1% (scară logaritmică)
  - Chronos-2 direct, Chronos-2 cu ACI, FHS
- ACI aduce depășirile aproape de 1%
  - S&P 500: 21 de depășiri, față de 24,4 așteptate
  - Acoperirea condiționată nu este respinsă pe nicio piață
  - Testul DQ respinge totuși pe S&P 500 și pe BET
- VaR-ul este mai larg, iar pierderea FZ0 este semnificativ mai mare
  - Media VaR 1% în cele 2 443 de zile de după cele 250 de zile de calibrare: 3,36%, față de 2,85% pentru FHS în aceleași zile
- Garanția privește media lui err_t, nu valoarea lui α_t
- Tipul de dovadă: comparație între proceduri (ACI față de Chronos-2 direct și FHS)
- Regulile de scor proprii recompensează concentrarea distribuției (*sharpness*) (Gneiting & Raftery (2007))
  - Sub constrîngerea calibrării
  - Calibrarea este necesară, nu suficientă

MFM_ch14_calibration

Studiu de caz: rolul complexității în prognoza randamentelor (1/2)

- **Întrebările** (Kelly, Malamud & Zhou (2024))
  - trebuie ca un model de prognoză a randamentelor să aibă mult mai puțini parametri P decît observații T?
  - pot modelele cu complexitatea $c = P/T > 1$ să ofere prognoze mai bune în afara eșantionului?
- Cadrul (ec. 1–2): randamentul lunii următoare este o funcție necunoscută de predictorii de azi
  - Aproximată printr-o rețea largă

$$R_{t+\mathbf{1}} = f(G_t) + \varepsilon_{t+\mathbf{1}}, \qquad f(G_t) \approx \sum_{i=\mathbf{1}}^{P} S_{i,t}\,\beta_i$$

  - $R_{t+\mathbf{1}}$: randamentul în exces al pieței
  - G_t: vectorul predictorilor cunoscuți la t
  - $\varepsilon_{t+\mathbf{1}}$: partea impredictibilă
  - $S_{i,t}$: a i-a trăsătură neliniară a lui G_t (o unitate ascunsă a unei rețele cu un strat ascuns), iar β_i este ponderea ei
  - Aproximarea se sprijină pe teorema de aproximare universală (Hornik, Stinchcombe & White (1989))
- **Relevanța**: un argument teoretic pentru modelele mari din acest capitol
  - Împotriva regulii parcimoniei din econometria clasică

Studiu de caz: rolul complexității în prognoza randamentelor (2/2)

- **Legătura cu Testul 1**: pe S&P 500 și Bitcoin, niciun model nu depășește prognoza zero
  - După corecția pentru teste multiple
  - Lucrarea întreabă dacă un R^2 negativ în afara eșantionului înseamnă lipsa valorii economice
  - Valoarea economică: raportul Sharpe al randamentului strategiei de ajustare a expunerii la piață (market timing) $\pi_t R_{t+1}$, cu poziția $\pi_t = \hat{\beta}' S_t$ (ec. 6): prognoza însăși este poziția
  - $\hat{\beta}$: ponderile estimate
  - S_t: vectorul celor P trăsături la t
- **Teoria** (Secțiunile III–IV): media lui R^2 în afara eșantionului și media raportului Sharpe al strategiei cresc cu c
  - În ipotezele lucrării, pentru un model greșit specificat, cu penalizarea ridge z (micșorează ponderile $\hat{\beta}$ spre zero) aleasă optim
  - Aproape de $c = 1$ (pragul de interpolare), metoda celor mai mici pătrate ajustează exact datele de antrenare și varianța explodează; dincolo de prag, rezultatele se îmbunătățesc din nou

Studiu de caz: date și metodă

- **Date** (Secțiunea V.A): randamentul lunar în exces al indicelui CRSP ponderat cu capitalizarea
  - CRSP: Center for Research in Security Prices
  - Cei 15 predictori din Welch & Goyal (2008) (14 variabile și randamentul întîrziat al pieței), perioada 1926–2020
  - Randamentele sînt scalate cu volatilitatea lor din ultimele 12 luni, predictorii cu abaterea standard calculată pe o fereastră cu început fix și lungime crescătoare
  - După 36 de luni de inițializare, eșantionul începe în 1930
- **Trăsăturile Fourier aleatoare** (RFF, random Fourier features; ec. 20, Rahimi & Recht (2007))
  - Fiecare trăsătură este un sinus sau un cosinus al unei combinații aleatoare a predictorilor

$$S_{i,t} = \big(\sin(\gamma\,\omega_i' G_t),\ \cos(\gamma\,\omega_i' G_t)\big)', \qquad \omega_i \sim N(0, I), \quad \gamma = 2$$

  - ω_i: vector aleator de ponderi, extras o singură dată din distribuția Normală standard (I: matricea identitate) și apoi fixat
  - γ: gradul de neliniaritate
  - Un γ mai mare dă trăsături care oscilează mai repede
  - O rețea cu două straturi și multe unități ascunse, cu ponderile primului strat aleatoare; doar ponderile de ieșire β sînt estimate
  - P numără coloanele: $P/2$ vectori ω_i, fiecare dînd un sinus și un cosinus (deci $P = 12\,000$ folosește $6\,000$ de vectori ω)

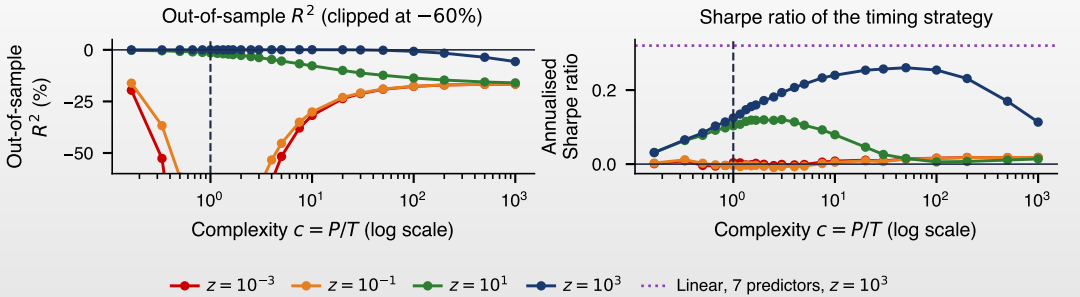
Studiu de caz: rezultatele principale

Model ($T = 12$)	R^2 (%)	Sharpe [t]	Raport inf. [t]	Pierdere max.	Asimetrie
Liniar, $z = 0^+$	< -100	$-0{,}11$ [$-1{,}0$]	$-0{,}16$ [$-1{,}6$]	98,5	$-0{,}9$
Liniar, $z = 10^3$	$-3{,}8$	0,46 [4,4]	0,33 [3,1]	2,4	$-0{,}1$
RFF, $c = 1\,000$, $z = 10^3$	0,6	0,47 [4,5]	0,31 [2,9]	1,2	2,5

Sursa: Tabelul I, Panoul A. Raport inf. față de piață; pierdere în abateri standard; t: media strategiei și alfa.

- Regresia pe toți predictorii (*kitchen sink*), fără penalizare ($P = 15 > T = 12$ luni): R^2 sub -100%, raportul Sharpe $-0{,}11$
  - Interpretare: la pragul de interpolare prognozele explodează, ceea ce reproduce verdictul negativ din Welch & Goyal (2008)
- Aceiași predictori cu $z = 10^3$: $R^2 = -3{,}8\%$, dar raportul Sharpe 0,46 ($t = 4{,}4$)
  - Interpretare: R^2 penalizează scala prognozelor, în timp ce profitul strategiei de market timing depinde de corelația lor cu randamentele (nota 5)
- RFF, $c = 1\,000$: raportul de informație față de cel mai bun model liniar este 0,26 ($t = 2{,}5$)
  - Cu $T = 60$ și 120 de luni, raportul de informație față de piață este 0,25, respectiv 0,24 ($t = 2{,}3$ și 2,2, Panourile B–C)
  - Interpretare: neliniaritatea extrage mai mult din aceiași predictori, cu o pierdere maximă mai mică și asimetrie pozitivă
- Pozițiile (Figura 10) sînt aproape mereu de cumpărare
  - Expunerea scade înainte de 14 din cele 15 recesiuni NBER (datate de National Bureau of Economic Research), în afara eșantionului
  - Interpretare: modelul învață singur restricția de semn din Campbell & Thompson (2008)
- Tipul de dovadă: comparație între proceduri (modele de complexitate diferită, în afara eșantionului)

Studiu de caz: complexitatea pe datele cursului (1/2)



MFM_ch14_virtue_complexity ▶ Interactiv

Studiu de caz: complexitatea pe datele cursului (2/2)

- Figurile 7–8 din Kelly, Malamud & Zhou (2024), refăcute pe datele cursului
  - Randamentele în exces ale SPY, 7 predictori, 354 de luni de test (februarie 1997–august 2026)
  - Parametrii: $T = 12$ observații complete, $\gamma = 2$, 1 000 de extrageri
  - Linia întreruptă marchează $c = 1$, iar linia punctată arată modelul liniar cu $z = 10^{3}$
  - Datele sînt descrise în ▸ Anexă
- Aproape fără penalizare ($z = 10^{-3}$): $R^2 = -1\,852\%$ la $c = 1$, apoi revine la $-16{,}6\%$ la $c = 1\,000$
  - Dubla descreștere (*double descent*) din Figura 7 (panoul din stînga limitat la -60%)
- Cu $z = 10^{3}$, raportul Sharpe al RFF este 0,12 la $c = 1$, 0,26 la $c = 50$ și 0,11 la $c = 1\,000$ ($t = 0{,}6$)
- Modelul liniar cu $z = 10^{3}$ are raportul Sharpe 0,32 ($t = 1{,}7$) și nu este depășit de RFF
  - t testează media randamentului strategiei ($\sqrt{n}\,\bar{R}/s_R$), nu raportul Sharpe în sine
  - Cu erori HAC (Newey–West, 3, 6 și 12 lag-uri): $t = 1{,}75$, 1,81 și 1,81
  - Raportul de informație al RFF este 0,07 față de modelul liniar ($t = 0{,}4$) și $-0{,}12$ față de SPY
- Interpretare: pe datele cursului, raportul Sharpe al RFF crește doar pînă la $c \approx 50$ și nu depășește modelul liniar
- Tipul de dovadă: comparație între proceduri

MFM_ch14_virtue_complexity

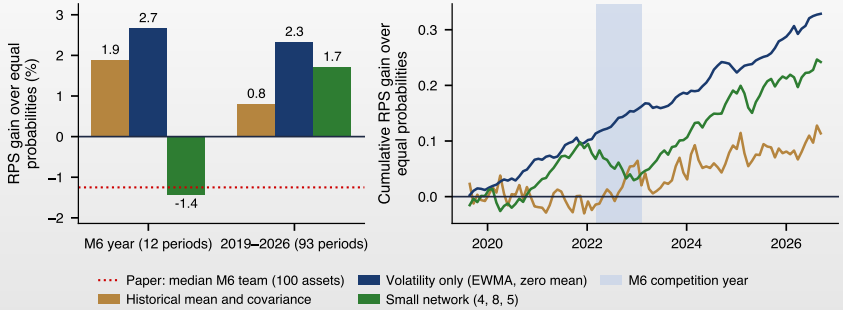
Studiu de caz: limite și o întrebare deschisă

- **Limitele** indicate de autori
  - Teoria acoperă modele ridge liniare pe trăsături aleatoare fixe și un singur activ; extinderea la un panel care ține cont de covariațele dintre active rămîne deschisă (Introducerea, nota 2)
  - Predictibilitatea este distribuită uniform între trăsături (Ipoteza 4)
  - Modelele cu selecție de variabile, de tip LASSO (Least Absolute Shrinkage and Selection Operator), sînt lăsate pentru cercetări viitoare (nota 22)
  - Nu există un grad optim de neliniaritate γ
  - Randamentele medii ale strategiei sînt de circa două ori mai mici în 1975–2020 decît în 1930–1974 (Secțiunea V.F)
- **Datele cursului** (Seminarul 14, exercițiul C3) au doar 7 din cei 15 predictori și circa 30 de ani
- Puterea unui test bilateral la 5% pentru un raport de informație de 0,31 este de doar circa 39%
  - $t \approx 0{,}31\sqrt{29{,}5} \approx 1{,}7$
- **Întrebare deschisă**: înlocuim trăsăturile aleatoare S_t cu trăsături învățate
  - De exemplu reprezentările interne (embeddings) ale unui model fundațional pre-antrenat precum Chronos-2
  - Mai este complexitatea o virtute pe un panel de piețe din datele cursului?

Studiu de caz: competiția de prognoză M6

- **Întrebarea** (Makridakis et al. (2025)): poate o metodă de prognoză, oricît de complexă, să bată probabilitățile egale în prognoza performanței relative a activelor?
  - M6: competiție live, 2022–2023, 226 de echipe, 100 de active (50 de acțiuni din S&P 500 și 50 de ETF-uri internaționale)
  - Sarcina: probabilitatea ca randamentul pe patru săptămîni al fiecărui activ să cadă în fiecare dintre cele cinci chintile ale universului
- **Relevanța**: o evaluare în afara eșantionului în timp real, deci fără contaminare prin pre-antrenare
  - Echipele au folosit și machine learning, și modele de serii de timp, și judecata proprie
- **Afirmația principală**: probabilitățile egale sînt greu de bătut
  - Din cele 163 de echipe din clasamentul global, 38 (23,3%) au un RPS (ranked probability score) sub cel al reperului, 0,16; doar 3 îl bat în toate cele 12 luni
  - Cele 11 echipe care bat reperul și în prognoză, și în investiție reduc RPS cu cel mult 2,2%
  - Dintre primele 20% dintre echipe după RPS, 14 folosesc machine learning și 12 modele de serii de timp (Tabelul 7)
  - Corelația dintre RPS și raportul de informație (IR) al portofoliului echipei este 0,04 ($p = 0{,}601$), pe cele 138 de echipe cu prognoze diferite de reper

Studiu de caz: M6 pe datele cursului



- Stînga: reducerea RPS față de probabilitățile egale (%), în cele 12 perioade M6 și în 93 de perioade de patru săptămîni (26.07.2019–11.09.2026); linia punctată: echipa mediană din M6 (RPS 0,162, calculat în lucrare pe 162 din cele 163 de echipe; 100 de active), reper de context, nu prag comparabil
- Dreapta: suma reducerilor de RPS, perioadă cu perioadă; zona albastră marchează anul competiției

MFM_ch14_m6_competition ▶ Interactiv

Studiu de caz: RPS, datele și replicarea

- **RPS** (ec. 1): distanța dintre distribuția cumulată prognozată și cea realizată a chintilelor

$$\mathrm{RPS}_i = \frac{1}{5}\sum_{j=1}^{5}\Big(\sum_{k\le j} q_{ik} - \sum_{k\le j} f_{ik}\Big)^2$$

  - f_{ik}: probabilitatea prognozată ca activul i să cadă în chintila k
  - q_{ik} (notația din M6; aici nu este o cuantilă): în M6, 1 pentru chintila realizată și 0 în rest
  - Aici, cu 19 active: partea din chintila k acoperită de intervalul de rang $[(r-1)/19,\ r/19]$ al activului (ponderi fracționare)
  - Cu 100 de active, probabilitățile egale ($f_{ik} = 0{,}2$) dau RPS 0,16; cu ponderile fracționare de aici, 0,1516
- **Adaptarea**: cele 19 active M6 din datele cursului (AMZN, JPM, META, HYG, IEF, IWM, LQD, SHY, TLT și cele 10 ETF-uri sectoriale XL*), randamente totale din prețuri ajustate
  - Calendarul M6 de patru săptămîni (Tabelul 1), extins din 2019; cu 19 active, chintilele realizate sînt fracționare
  - Media istorică: randamente din distribuția Normală, cu media și covarianța din ultimele 250 de zile, 20 000 de simulări
  - Doar volatilitatea: media zero și covarianța EWMA (exponentially weighted moving average, $\lambda = 0{,}94$)
  - Rețeaua: 4 intrări (volatilitățile și randamentele pe 20 și 250 de zile), 8 neuroni, 5 probabilități; RPS ca pierdere, reestimată anual pe perioadele din 2004

Studiu de caz: rezultatul M6 pe datele cursului

- Modelul care folosește doar volatilitatea bate reperul: RPS mai mic cu 2,7% în anul M6 (11 din 12 perioade) și cu 2,3% din 2019 (DM $-6{,}35$, $p < 0{,}001$)
  - 2,7% aici și 2,2% în lucrare provin din universuri și ținte diferite, deci nu sînt direct comparabile
  - Mecanismul: în prima perioadă M6, META (volatilitate EWMA 76% pe an) are probabilitatea 0,82 de a cădea într-o chintilă extremă, HYG (7%) are 0,08, iar reperul dă 0,40
- Media istorică reduce RPS cu 1,9% în anul M6, dar doar cu 0,8% din 2019 ($p = 0{,}358$)
  - Cu aceeași covarianță pe 250 de zile, media zero are RPS mai mic cu 1,7% decît media estimată (DM $-1{,}75$, $p = 0{,}080$)
- Rețeaua crește RPS cu 1,4% în anul M6 și reduce RPS cu 1,7% din 2019 ($p = 0{,}016$)
  - Față de modelul doar cu volatilitate: RPS mai mare cu 0,6%, diferență nedetectată (DM 0,73, $p = 0{,}466$)
  - Fără intrările de randament, RPS crește cu doar 0,05% ($p = 0{,}70$)
- Tipul de dovadă: comparație între proceduri; informația utilă este în dispersia randamentelor, nu în semnul lor, ca în Testele 1 și 2
- Limite: 19 active, nu 100, și o singură piață
  - RPS este mărginit, deci testul DM nu cere ipoteze despre cozile distribuției
  - În lucrare: un singur an, iar doar 26 din cele 163 de echipe au trimis prognoze noi la toate cele 12 termene
  - Perioadele dinainte de 2022 folosesc retrospectiv universul ales pentru M6

MFM_ch14_m6_competition

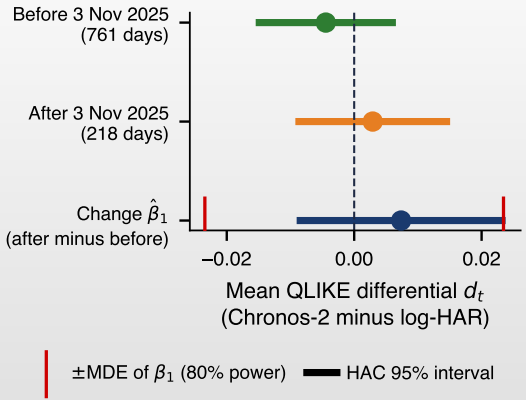
Întrebare deschisă: capacitate de prognoză sau memorarea datelor de antrenare?

- Testul 2: Chronos-2 zero-shot nu se distinge de log-HAR pe varianța realizată a SPY
- Rahimikia, Ni & Wang (2025): modelele fundaționale gata antrenate au rezultate slabe zero-shot și după fine-tuning
  - Pe randamentele zilnice în exces de pe piețele globale
  - Modelele pre-antrenate de la zero pe date financiare au rezultate mai bune
- Lopez-Lira, Tang & Zhu (2025): abilitatea de prognoză contrafactuală a unui model nu este identificată
  - Odată ce modelul a văzut valorile realizate
  - Memorarea detectată este doar o limită inferioară
  - Demonstrat pentru LLM-uri; logica se aplică oricărui model pre-antrenat
- Întrebări deschise
  - rezultatele comparabile reflectă învățarea structurii (memorie lungă, revenire la medie) sau memorarea perioadei de test?
  - poate o fereastră scurtă de după publicare să le distingă?
- Un test: $d_t = L(\hat{h}_t^{FM}, \mathrm{RV}_t) - L(\hat{h}_t^{\log \mathrm{HAR}}, \mathrm{RV}_t)$
  - Pierderea QLIKE L a modelului fundațional (FM) minus cea a log-HAR
  - Memorarea implică $E[d_t \mid \text{după}] > E[d_t \mid \text{înainte}]$: FM își pierde avantajul după publicarea ponderilor
  - H_0: cele două medii sînt egale, dar o respingere poate reflecta și o schimbare de regim

Bucla descoperirii în cinci pași

- **Literatura**: cereți unui asistent evaluări ale modelelor fundaționale pe date financiare din 2024 încoace
  - Apoi verificați fiecare identificator arXiv
  - Citiți direct documentația corpusului
- **Ipoteze**: predicții care separă memorarea de abilitate
  - Înainte și după publicare
  - Serii niciodată publice
  - Serii simulate cu aceeași dinamică
- **Date și cod**: din prognozele de volatilitate ale capitolului
  - Diferențialul QLIKE Chronos-2 minus log-HAR
  - Un test înainte/după cu erori HAC și efectul minim detectabil (MDE, minimum detectable effect)
- **Robustețea**, fixată dinainte
  - O dată de publicare placebo, cu un an mai devreme, nu trebuie să arate nicio ruptură
  - Seriile simulate nu pot fi memorate
- **Critica**: cereți asistentului să atace „nu există memorare, pentru că testul de după publicare este nesemnificativ”
  - Răspunsuri de așteptat: putere mică și regimuri diferite
  - Prompturi reutilizabile în Anexă

Mini-studiu de caz: un test fără putere (1/3)



MFM_ch14_ai_discovery ▶ Interactiv

Mini-studiu de caz: un test fără putere (2/3)

- Propunere tipică a unui asistent
  - „Paritatea Chronos-2 vine din date memorate, deci avantajul trebuie să dispară după publicare"
- Graficul: media diferențialului QLIKE d_t (Chronos-2 minus log-HAR) înainte (verde) și după (portocaliu) publicare
  - Cu intervalele HAC de 95%
- Rîndul de jos: schimbarea estimată $\hat{\beta}_1$, cu liniile roșii la $\pm$MDE
  - MDE: efectul minim pe care testul îl poate detecta
- Media lui d_t: $-0{,}0045$ înainte de 3 nov. 2025 (761 de zile) și $+0{,}0029$ după (218 zile)
- Schimbarea mediei: $+0{,}0074$, cu t HAC $= 0{,}88$ și $p = 0{,}38$
  - Testul cere variantă de termen lung finită a lui d_t, deci prognoze care nu se apropie de zero
- Tipul de dovadă: asociere (variabilele se mișcă împreună; nu dovedește o relație cauzală)
  - Chiar o respingere ar lega diferența doar de data publicării, nu de memorare

MFM_ch14_ai_discovery

Mini-studiu de caz: un test fără putere (3/3)

- Efectul minim detectabil: MDE $= (z_{0,975} + z_{0,8})\,\mathrm{SE}_{HAC}(\hat{\beta}_1) = 0{,}0234$
  - $z_{0,975} = 1{,}96$, $z_{0,8} = 0{,}84$: cuantile ale distribuției Normale standard, test bilateral la 5%, putere 80%
  - $\hat{\beta}_1$: deplasarea estimată a mediei lui d_t după publicare, iar SE_{HAC} este eroarea ei standard HAC
  - MDE este de 5,3 ori diferența medie dinainte de publicare și ar rămîne 0,0157 chiar cu o fereastră infinită de după publicare
- De ce există un prag: erorile HAC dau fiecărei perioade propria varianță de termen lung V

$$\mathrm{SE}^2 \approx \frac{V_{pre}}{761} + \frac{V_{post}}{218}$$

  - Mai multe zile de după publicare elimină doar al doilea termen, rămînînd $2{,}80\sqrt{V_{pre}/761}$
- Interpretare: „nu există memorare” nu este o concluzie
  - Puterea trebuie să vină din secțiunea transversală (multe serii testate împreună), nu din așteptarea mai multor zile de după publicare

MFM_ch14_ai_discovery

Un punct de plecare pentru proiect

- Întrebări (acolo unde memorarea este imposibilă)
  - se menține abilitatea zero-shot pentru varianța zilnică după publicare?
  - se menține ea pe serii simulate?
  - se menține ea pe piețe mai puțin acoperite?
- Replicați întîi: QLIKE din capitol pentru log-HAR (0,226) și Chronos-2 (0,223)
  - DM $p = 0{,}539$
  - Fereastra de 218 zile de după publicare
- Livrabile
  - Un calcul de putere preînregistrat
  - Un panel de serii din datele cursului, cu o aproximare a variației din amplitudinea zilnică și o pierdere robustă la aproximare (Patton (2011))
  - Teste înainte/după cu date de publicare placebo
  - MCS pentru toate modelele
  - O critică de tip recenzent și răspunsul echipei
- Notați fiecare prompt care a influențat o decizie în `AI_USE.md`
- Notați fiecare referință inventată, scurgere de informație sau cifră greșită identificată în `AI_ERRORS.md`
  - Împreună cu testul prin care a fost identificată (șabloanele de proiect)

Idei principale

- Porțile LSTM pot atenua dispariția gradientului
  - Cu termenul liber al porții de uitare egal cu 0, gradientul tot dispare
- Atenția conectează direct momente îndepărtate
- Modelele fundaționale prognozează zero-shot dintr-o fereastră de context și furnizează cuantile
- Cele trei teste
  - Randamente: după corecția pentru teste multiple, niciun model nu depășește prognoza zero pe S&P 500 și Bitcoin
  - Volatilitate: modelele fundaționale nu se disting de log-HAR
  - Cozi: FHS rămîne reperul; modelele fundaționale, cel mult ca filtre de volatilitate
- Evaluați orice model nou față de un reper puternic și după publicarea ponderilor lui
  - Cu inferența potrivită tipului de comparație
  - Clark–West pentru modele imbricate
  - SPA sau Romano–Wolf pentru multe modele
  - PIT, scoruri cuantilice și DQ pentru distribuții
- Lecturi suplimentare în Anexă

Autoevaluare și capitolul următor: Capitolul 15

Întrebări de verificare

- De ce uită o RNN stabilă în ritmul răspunsului la impuls A^k al unui VAR(1), nu neapărat ca $\rho(A)^k$?
- De ce este $\exp(q_{\mathbf{0},\mathbf{s}})$ din log RV o prognoză deplasată sub QLIKE?
- De ce poate un VaR hibrid să treacă testul Kupiec și totuși să piardă comparația FZ0?
- De ce „nesemnificativ după publicare" nu este o dovadă împotriva memorării?

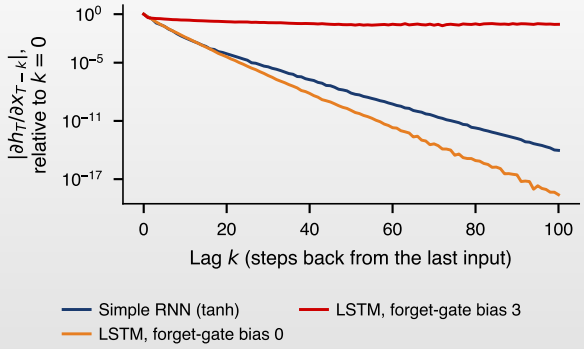
Urmează: LLM și analiza sentimentului

- Cursul următor: Capitolul 15
- Poate textul citit de un model mare de limbaj să prognozeze randamentele sau volatilitatea?
- Revin aceleași amenințări: deplasarea de anticipare, memorarea perioadei de test, repere slabe
- Aceleași instrumente: ferestre de după publicare, Clark–West, controlul testelor multiple

Decizia practică

- **Ce aleg**
  - Randamente zilnice ale indicilor: prognoza zero
  - Volatilitate realizată: log-HAR; un model fundațional zero-shot cînd trebuie prognozate mii de serii fără estimare
  - VaR 1% și ES 2,5%: FHS, cu un model fundațional cel mult ca filtru de volatilitate sau cu recalibrare conformală
- **În ce condiții**
  - Pe S&P 500 și Bitcoin, niciun model nu depășește prognoza zero după corecția pentru teste multiple
  - Pe BET, valoarea p SPA corectată Holm pentru cele trei piețe este 0,056
  - Pe SPY, Chronos-2 și log-HAR au QLIKE 0,223 și 0,226 (DM $p = 0{,}539$)
  - Pe S&P 500, FHS are cea mai mică pierdere FZ0 (DM +2,65 față de hibridul Chronos-2)
- **Ce rezultat m-ar face să renunț**
  - Un model care depășește prognoza zero după Holm sau SPA, într-o familie de teste fixată dinainte
  - Un model fundațional cu QLIKE semnificativ sub log-HAR în fereastra de după publicare
  - Un VaR hibrid care trece și testele de independență și DQ, nu doar testul Kupiec

Dispariția gradientului, măsurată numeric



- $|\partial h_T / \partial x_{T-k}|$ prin diferențiere automată, raportat la valoarea pentru $k = 0$, ca pe axa verticală
  - 16 unități ascunse, 50 de inițializări aleatoare, aceleași intrări din distribuția Normală
- RNN simplă: gradientul este $8{,}3 \times 10^{-5}$ la lag-ul 20 și $7{,}4 \times 10^{-9}$ la lag-ul 50
  - Nu se învață nimic din intrări mai vechi de cîteva săptămîni
- LSTM cu termenul liber al porții de uitare egal cu 0: gradientul este $1{,}2 \times 10^{-10}$ la lag-ul 50
  - Nu este mai bun decît RNN: $f \approx 0{,}5$ înjumătățește memoria la fiecare pas
- LSTM cu termenul liber egal cu 3: gradientul este 8,4% la lag-ul 50
  - Conform valorii teoretice $0{,}953^{50} \approx 0{,}09$
  - Inițializați termenul liber al porții de uitare cu o valoare pozitivă

MFM_ch14_vanishing_gradient

Atenția în cifre: trei pași și o mască cauzală (1/2)

- Trei pași de timp, $d_k = 1$
  - Interogările $q = (1;\ 0;\ 2)$
  - Cheile $k = (0{,}5;\ 1;\ -1)$
  - Valorile $v = (1;\ 2;\ 3)$
- Pasul 3: scorurile $q_3 k_j = (1;\ 2;\ -2)$
  - Ponderile softmax sînt $e^{s_j} / \sum_i e^{s_i} = (2{,}718;\ 7{,}389;\ 0{,}135)/10{,}243 = (0{,}265;\ 0{,}721;\ 0{,}013)$
  - Ieșirea: $0{,}265 \cdot 1 + 0{,}721 \cdot 2 + 0{,}013 \cdot 3 = 1{,}75$
  - Pasul 3 preia mai ales informația pasului 2, a cărui cheie se potrivește cel mai bine cu interogarea lui

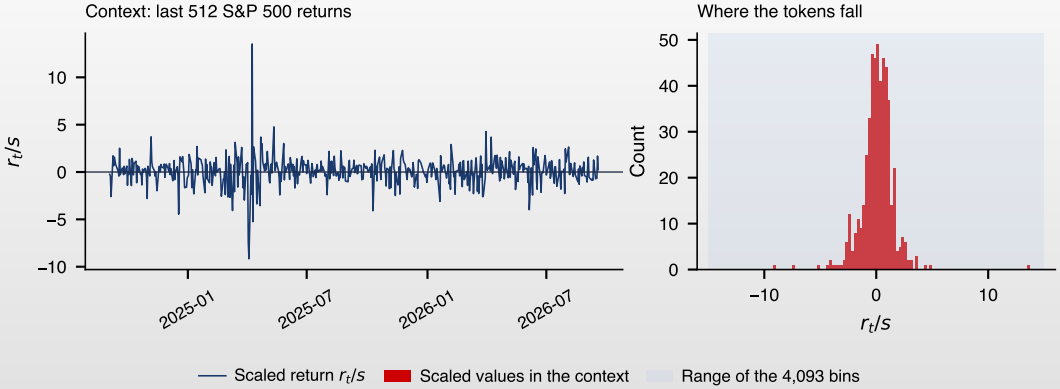
Atenția în cifre: trei pași și o mască cauzală (2/2)

- **Masca cauzală**: pasul t poate folosi doar pașii $j \leq t$
  - Într-un model care prognozează fiecare pas următor
  - Echivalent, adăugăm $M_{tj} = 0$ pentru $j \leq t$ și $M_{tj} = -\infty$ pentru $j > t$

$$\text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}} + M\right) V$$

  - Pasul 1 își acordă atunci ponderea 1 (ieșirea 1)
  - Fără mască, scorurile lui (0,5; 1; -1) dau ponderile (0,35; 0,57; 0,08): 57% vine de la pasul 2, o zi ulterioară
  - Cu mască, pasul 2 ($q_2 = 0$, scoruri egale pe pașii 1 și 2) are ieșirea $(1 + 2)/2 = 1{,}5$
- Toate valorile unui context de 512 zile sînt deja observate la momentul prognozei
  - Deci un model care citește doar contextul poate folosi atenția în ambele sensuri
- Masca este necesară acolo unde modelul este antrenat să prognozeze pozițiile ulterioare din cele anterioare

Intrarea modelului Chronos: randamente scalate și tokeni (1/2)



MFM_ch14_tokenization

Intrarea modelului Chronos: randamente scalate și tokeni (2/2)

- Contextul: ultimele 512 randamente zilnice ale S&P 500, împărțite la scala $s = 0{,}672\%$
  - Randamentele scalate sînt între $-9{,}2$ și $13{,}5$
  - 96% dintre ele se află în intervalul ± 3
- Cele 512 randamente folosesc 336 de tokeni distincți din 4 093
  - Scalarea face modelul independent de unități, deci nivelul volatilității trebuie citit din context
- O zi de crah cu randamentul M (în %) intră în context, iar cel mai vechi randament x_{out} iese
  - s crește cu $(|M| - |x_{\text{out}}|)/C$ puncte procentuale, aproximativ $|M|/C$
- Creșterea relativă, circa $|M|/(Cs)$, este mare într-un context scurt și mică într-unul de 512 zile
  - Seminarul 14, exercițiile A3 și A6

MFM_ch14_tokenization

Transformerele pentru prognoză: blocul Transformer

- **Blocul Transformer** (Vaswani et al. (2017)): atenție cu mai multe capete, apoi un MLP mic
  - MLP-ul este aplicat separat fiecărei poziții
  - Fiecare parte are o conexiune reziduală și o normalizare

$$Z = \mathrm{LN}(X + \mathrm{MHA}(X)), \qquad H = \mathrm{LN}(Z + \mathrm{MLP}(Z))$$

  - $\mathrm{MHA}(X) = [\mathrm{head}_1, \ldots, \mathrm{head}_m]\, W_O$ (multi-head attention): m atenții în paralel, fiecare cu propriile W_Q, W_K, W_V, concatenate și proiectate cu W_O
  - LN (layer normalization): fiecare rînd este standardizat, apoi rescalat cu parametri învățați
  - $X + \mathrm{MHA}(X)$: conexiunea reziduală, care lasă gradientul să treacă direct prin bloc
  - Un model suprapune mai multe blocuri: H devine intrarea blocului următor

Transformerele pentru prognoză: arhitecturi și pierderi

- Trei moduri de a asambla blocurile
  - **Encoder** (codificator): atenție în ambele sensuri în interiorul contextului; Chronos-2 are doar encoder și furnizează grila de cuantile dintr-un singur pas (Ansari et al. (2025))
  - **Doar decoder** (decodor): mască cauzală, antrenat să prognozeze patch-ul următor la fiecare poziție: TimesFM (Das et al. (2024))
  - **Encoder–decoder** (T5, folosit de Chronos original): encoderul citește contextul, iar decoderul scrie prognoza token cu token
- **Entropia încrucișată** (cross-entropy, la Chronos original)
  - Prognoza tokenului următor este o clasificare în 4 093 de clase

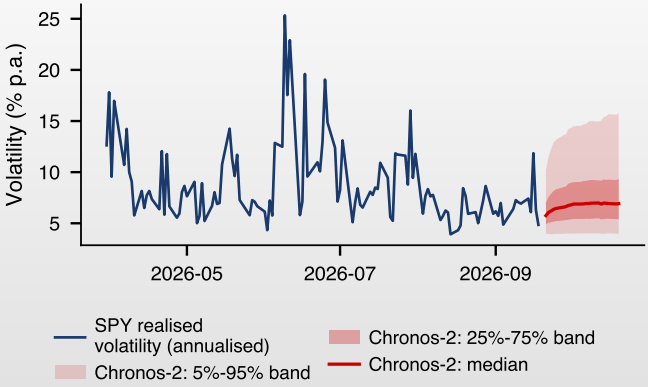
$$\mathcal{L}(\theta) = -\sum_{h=\mathbf{1}}^{H} \log \hat{p}_\theta \left(z_{C+h} \mid z_{\mathbf{1}}, \ldots, z_{C+h-\mathbf{1}} \right)$$

  - $\hat{p}_\theta(z \mid \cdot)$: probabilitatea acordată de model tokenului care s-a realizat, dați fiind tokenii anteriori
- **Scalarea robustă** (Chronos-2, Ansari et al. (2025)): $\tilde{x}_t = \sinh^{-\mathbf{1}}\left((x_t - \hat{\mu})/\hat{\sigma}\right)$
  - $\hat{\mu}$, $\hat{\sigma}$: media și abaterea standard a contextului
  - $\sinh^{-\mathbf{1}}$ crește ca un logaritm pentru valori mari și atenuează valorile extreme

De la o grilă de cuantile la VaR și ES

- Exemplu rezolvat, S&P 500, cuantilele pentru 18.09.2026 (context pînă la 17.09.2026): $q_{0,01} = -2{,}704\%$, $q_{0,05} = -1{,}400\%$
  - VaR 1% = 2,704%
  - interpolare liniară între cele două niveluri: $q_{0,025} = q_{0,01} + \frac{0{,}015}{0{,}04}(q_{0,05} - q_{0,01}) = -2{,}215\%$, deci VaR 2,5% = 2,215%
  - ES 2,5%: q_u constant sub 1%, regula trapezului între 1% și 2,5%
- $\mathrm{ES}_{2,5\%} = -\frac{1}{0{,}025}\left[0{,}01\, q_{0,01} + 0{,}015\, \frac{q_{0,01}+q_{0,025}}{2}\right] = 2{,}56\%$
  - În cifre: $40 \times [0{,}01 \times 2{,}704 + 0{,}015 \times 2{,}459] = 40 \times 0{,}06392 = 2{,}557$, adică 2,56%
- Interpretare: ES 2,5% este apropiat de VaR 1%
- ES/VaR 2,5% = 1,15 este sub valoarea Normală 1,19
  - Deși grila are o coadă stîngă mai grea decît distribuția Normală
  - Coada constantă sub 1% subestimează ES (Seminarul 14, exercițiul A7)

Un fan chart zero-shot pentru volatilitate (1/2)



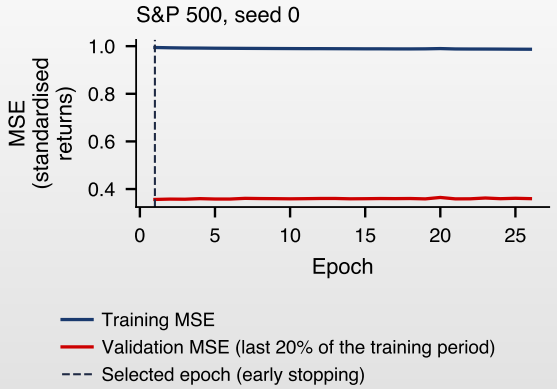
MFM_ch14_fan_chart

Un fan chart zero-shot pentru volatilitate (2/2)

- Chronos-2 pe logaritmul variației realizate a SPY (context de 512 zile)
  - SPY: fondul tranzacționat la bursă (ETF, exchange-traded fund) care replică S&P 500
  - 22 de zile de tranzacționare de la 18.09.2026
  - Benzi: cuantilele 5–95% și 25–75%
- Ultima volatilitate observată este 4,8% pe an
  - Media pe ultimele 20 de zile este 6,8%
  - Prognoza mediană la 22 de zile este 6,9%, cu banda de 90% între 4,0% și 15,8%
- Revenire la medie, ca la GARCH și HAR (capitolele 5, 9)
  - Fără nicio estimare pe SPY
  - Banda este asimetrică în sus parțial prin construcție: volatilitatea este $\sqrt{252\, e^{q}}$ dintr-o cuantilă q a log-variației, iar o funcție convexă întinde partea de sus
  - Pe scara log-volatilității (jumătate din log-variață), banda se află totuși la 0,56 sub mediană și la 0,82 deasupra ei, deci și modelul înclină spre salturi

MFM_ch14_fan_chart

LSTM-ul pentru randamente: curbele de antrenare



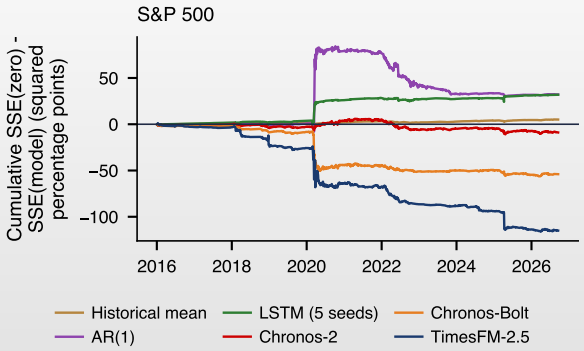
MFM_ch14_returns

LSTM-ul pentru randamente: arhitectura și early stopping

- Intrarea: ultimele 20 de randamente, standardizate pe perioada de antrenare
  - $h = 16$ unități ascunse, $d = 1$ intrare la fiecare pas, ieșire liniară
  - $4(hd + h^2 + 2h) + h + 1 = 1\,233$ de parametri (factorul 4: cele trei porți și conținutul candidat; $2h$: doi vectori de termeni liberi pe poartă)
- MSE, Adam, weight decay 10^{-4}
  - Antrenat o singură dată pe cel mult 4 000 de zile, dintre care ultimele 20% sînt blocul de validare
  - Early stopping: antrenarea se oprește după 25 de epoci fără un MSE de validare mai mic (patience), iar ponderile cele mai bune sînt păstrate
  - Prognoza = media a 5 seed-uri
- Graficul (S&P 500, seed-ul 0): MSE de antrenare și MSE de validare pe epoci
  - Pentru randamente standardizate
  - Linia verticală punctată marchează epoca păstrată
- Antrenarea se oprește după 26 de epoci, iar ponderile păstrate sînt cele din epoca 1
  - Rețeaua păstrată a văzut datele de antrenare o singură dată și rămîne aproape de inițializarea aleatoare: un motiv pentru care seed-ul contează
  - MSE de validare este mai mic decît MSE de antrenare, în principal pentru că ultimii ani ai perioadei de antrenare au fost mai calmi
- Semnal redus: abaterea standard a prognozei 0,047%, față de 1,13% pentru randamente

MFM_ch14_returns

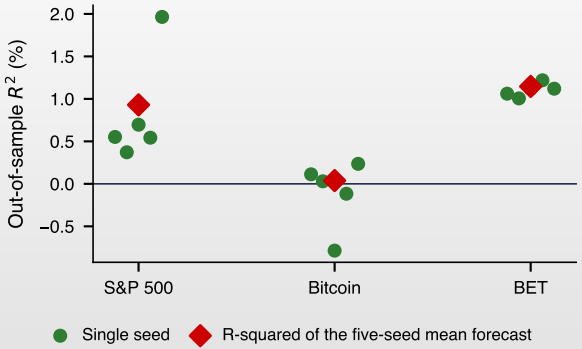
Reducerile cumulate ale erorii pătratice pe S&P 500



- Curba SSE(zero) − SSE(model) $= \sum_{s \le t} \left[r_s^2 - (r_s - \hat{r}_s)^2 \right]$
  - SSE: suma erorilor pătratice (sum of squared errors), în $\%^2$
  - Crește cînd modelul are o eroare mai mică decît prognoza zero și scade în caz contrar
- Toate curbele rămîn lîngă zero în anii calmi și au salturi în martie 2020
  - O singură lună determină clasamentul
- AR(1) cîștigă în 2020 datorită autocorelației negative din zilele de crah
  - Apoi pierde o parte din avantaj
  - TimesFM-2.5 are pierderi constante
- Raportați evoluția în timp, nu doar R^2 final

MFM_ch14_returns

Variabilitatea între seed-uri



- R^2 în afara eșantionului pentru cinci LSTM-uri care diferă doar prin seed-ul aleator și pentru media lor
- Bitcoin: R^2 al seed-urilor individuale este între $-0{,}79\%$ și $+0{,}24\%$
  - Pe S&P 500: între $+0{,}37\%$ și $+1{,}96\%$
- Variația dintre seed-uri este la fel de mare ca efectul măsurat
- BET: toate cele 5 seed-uri dau R^2 între $+1{,}01\%$ și $+1{,}22\%$
  - Rezultatul este robust la alegerea seed-ului
- Alegerea celui mai bun seed după examinarea rezultatelor de test este supraajustarea (*overfitting*) backtest-ului (Capitolul 13)

MFM_ch14_returns

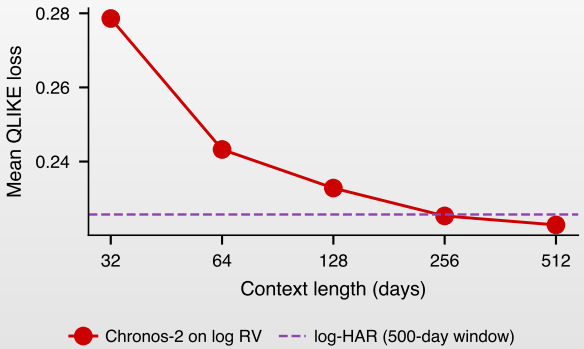
Predictibilitatea BET și tranzacționarea nesincronă

- Prețurile neactualizate induc autocorelație pozitivă în randamentele indicelui (Lo & MacKinlay (1990))
  - Prețurile neactualizate: ale acțiunilor tranzacționate rar
  - Chiar dacă fiecare preț este eficient
  - Autocorelația de ordinul 1 din 2016 încoace: 0,06 pentru BET și $-0{,}14$ pentru S&P 500
  - Prognoza LSTM are corelația 0,88 cu randamentul BET din ziua precedentă: a învățat în principal această autocorelație
- Verificare pe un instrument tranzacționabil: ETF-ul Patria-TVBETETF, care replică BET
  - 2 629 de zile comune, același semnal LSTM
  - Execuție în aceeași zi: $R^2 = 2{,}88\%$ (față de 1,26% pentru BET în aceleași 2 629 de zile)
  - Dar și ETF-ul se tranzacționează rar: autocorelația de ordinul 1 este 0,07, iar 4,2% din zile au randament zero
  - Execuție cu o zi întîrziere: $R^2 = 0{,}57\%$, p Clark–West $= 0{,}003$
- Interpretare: rezultatele sînt compatibile cu efectul prețurilor neactualizate
- Un semnal mic rămîne și cînd execuția este întîrziată cu o zi
  - Înainte de costurile de tranzacționare (Capitolul 2)

Calculul mediei din grilă: sensibilitatea la regula pentru cozi

- Grilă de exemplu: nivelurile 25%, 50%, 75%, cu log-cuantilele $q = -1;\ 0;\ 1$
  - Ținta este $E[Y] = \int_0^1 e^{q_u}\, du$
  - Ponderi egale pe cele trei puncte: $(e^{-1} + 1 + e)/3 = 1{,}36$
  - Regula din curs (trapez, constantă sub 25% și peste 75%): $\frac{1}{4}\left[e^{-1} + \frac{e^{-1}+1}{2} + \frac{1+e}{2} + e\right] = 1{,}41$
  - O distribuție Normală pentru log Y cu aceste cuartile ($s = 1/z_{0,75} = 1/0{,}674 = 1{,}48$, cu $z_{0,75}$ cuantila de 75% a distribuției Normale standard) are $E[Y] = e^{s^2/2} = 3{,}00$: cozile constante pierd mai mult de jumătate din ea
- Interpretare: exp este convexă și nemărginită
  - Deci partea de sus a distribuției contribuie mult la medie
- Trunchierea cozii superioare poate subestima puternic media; semnul erorii totale nu este garantat (fixarea cozii inferioare ridică integrala)
  - Testul 2: media prognozei / media RV este 1,02 pentru Chronos-2 (grilă pînă la 99%), 0,97 pentru Chronos-Bolt și 0,94 pentru TimesFM-2.5 (grile pînă la 90%)
  - Rezultatul este compatibil cu o coadă superioară tăiată, deși și prognozele în sine diferă
- Raportați regula de calcul împreună cu prognozele

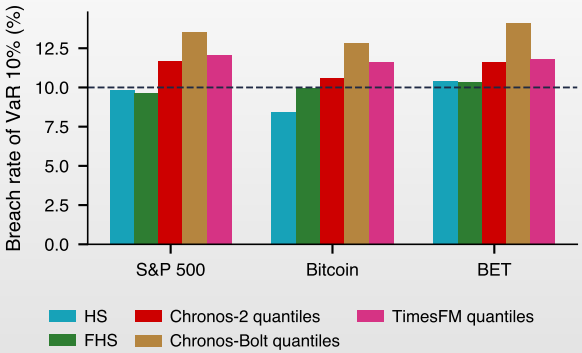
Lungimea contextului necesară pentru Chronos-2



- QLIKE pentru Chronos-2 pe log RV cu contexte de 32 pînă la 512 zile, în aceleași zile de prognoză
- QLIKE scade de la 0,279 (32 de zile) la 0,223 (512 zile)
- Sub circa 256 de zile, Chronos-2 este mai slab decît log-HAR (QLIKE 0,226)
- Memorie lungă: mai multe luni de istorie relevă nivelul lent variabil
  - HAR îl codifică explicit prin media pe 22 de zile
- Lungimea contextului este o alegere de modelare care trebuie raportată și testată
  - Ca lungimea unei ferestre

MFM_ch14_realised_vol

Suficiența decilelor: VaR 10%



- Rata de depășire a cuantilei de 10% (ținta 10%) pentru HS, FHS și cele trei modele fundaționale
- Ratele de depășire pe S&P 500
  - HS 9,8%, FHS 9,6%
  - Chronos-2 11,7%, Chronos-Bolt 13,5%, TimesFM-2.5 12,0%
- Kupiec respinge modelele fundaționale în opt din nouă cazuri
  - Coada stîngă este plasată prea sus, prin poziție, scală sau forma cozii
- Explicație plauzibilă: corpusurile de pre-antrenare sînt dominate de serii mai netede
- Folosite ca filtre de volatilitate în FHS, modelele lasă forma cozilor să fie dată de datele empirice

MFM_ch14_tail_risk

Inferența conformală adaptivă: garanția de acoperire și ES

- Garanție valabilă pentru orice secvență de randamente, fără ipoteze de model
  - Pe T zile rata de depășire se apropie de α

$$\left| \frac{1}{T} \sum_{t=1}^{T} \mathrm{err}_t - \alpha \right| \leq \frac{\max(\alpha_1, 1-\alpha_1) + \gamma}{\gamma T}$$

  - limita scade ca $1/T$: doar o afirmație pe termen lung, slabă pentru γ mic
- ES 2,5%: ES Chronos-2 deplasat cu corecția VaR 2,5%
  - O euristică fără garanție conformală
- Regulile și tabelul complet sînt în Anexă

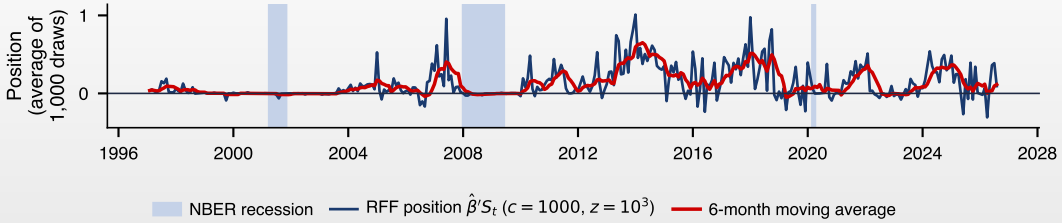
Studiu de caz: estimatorul ridge, planul experimentului și evaluarea

- **Ridge** (Secțiunea II.A), fără termen liber

$$\hat{\beta}(z) = \left(zI + \frac{1}{T}\sum_t S_t S_t'\right)^{-1} \frac{1}{T}\sum_t S_t R_{t+1}$$

  - $z > 0$: penalizarea ridge, care micșorează $\hat{\beta}$ spre zero
  - $z \to 0^+$: soluția fără penalizare (pseudo-inversa)
  - T: lungimea ferestrei
- **Designul** (Secțiunea V.C): ferestre mobile de $T = 12$, 60 sau 120 de luni
  - Cu P între 2 și 12 000 și $\log_{10} z = -3, \ldots, 3$
  - Rezultatele sînt mediate pe 1 000 de extrageri independente ale trăsăturilor
- **Evaluarea**: R^2 în afara eșantionului $= 1 - \mathrm{Var}(\text{eroare})/\mathrm{Var}(R)$ și raportul Sharpe anualizat
  - Nota 40: în teorie, MSE se împarte la $E[R^2]$, adică la eroarea prognozei zero din Testul 1
  - alfa și raportul de informație (alfa împărțit la volatilitatea reziduală) din regresia randamentului strategiei în funcție de randamentul pieței
  - Reperul: regresia liniară „kitchen sink", cu toți cei 15 predictori (Secțiunea V.D)

Studiu de caz: expunerea și recesiunile (1/2)



MFM_ch14_virtue_complexity

Studiu de caz: expunerea și recesiunile (2/2)

- Pe datele cursului, ca în Kelly, Malamud & Zhou (2024): poziția $\hat{\beta}' S_t$ pentru RFF cu $c = 1\,000$, $z = 10^3$
- Zonele umbrite marchează recesiunile NBER
  - Seria `USREC` din FRED, Federal Reserve Economic Data
  - Linia roșie: media mobilă a poziției pe 6 luni
- Poziții de cumpărare în 73% din luni, poziția medie 0,13
  - O primă de risc pozitivă învățată fără restricție de semn
- Poziția medie 0,05 în cele 6 luni dinaintea celor 3 recesiuni și $-0{,}01$ în timpul lor, față de 0,14 în rest
  - Șocul pandemic din 2020 nu este anticipat
- Trei recesiuni permit doar o descriere, nu un test
  - Lucrarea are 15 recesiuni și 90 de ani

MFM_ch14_virtue_complexity

Verificări metodologice pentru cercetător

- **Scurgerea de informație**: suprapunerea cu pre-antrenarea, scalarea pe tot eșantionul, medii HAR care includ ziua t
  - Seminarul 14, exercițiul B10 arată că acestea reduc artificial QLIKE cu un sfert
  - Verificare: fereastra de după publicare, serii simulate sau private
  - Recalculați prognoza pentru t din date trunchiate la $t-1$
- **Repere slabe**: comparați cu zero (randamente), log-HAR (volatilitate), FHS (cozi)
  - Nu cu un mers aleator
- **Funcționala greșită**: mediane folosite ca medii sau decile folosite drept cuantile de 1%
  - Media prognozei / media RV trebuie să fie aproape de 1; mediana Chronos-2: 0,80
- **Selecția**: seed-uri, contexte și modele alese după examinarea rezultatelor de test
  - Remediu: notați fiecare încercare și comparați întregul set de modele cu MCS
- **Inferența greșită**
  - DM pe modele imbricate
  - Fără control al testelor multiple
  - Acoperire fără independență
  - „Nesemnificativ" raportat fără MDE

Norme, rază spectrală și explozie

- $\gamma_\varphi \|W\|_2 < 1$ este suficient pentru dispariția geometrică (Pascanu, Mikolov & Bengio (2013))
  - Deci $\gamma_\varphi \|W\|_2 \geq 1$ este necesar pentru explozie
- Raza spectrală ρ (cea mai mare valoare proprie în valoare absolută)
  - Într-o RNN liniară, $\rho(W) > 1$ este necesar pentru explozia exponențială
  - $\rho(W) = 1$ cu un bloc Jordan dă totuși o creștere polinomială
  - Cu φ contează jacobienele locale $D_t W$: se poate $\rho(W) = 0$, dar $\rho(DW) > 1$
- VAR stabil: răspunsurile la impuls și gradienții scad asimptotic ca $\rho(A)^k$
  - La orizont finit, $\|A^k\|_2$ poate depăși 1: $A = \begin{pmatrix} 0{,}5 & 4 \\ 0 & 0{,}5 \end{pmatrix}$ are $\rho(A) = 0{,}5$, dar $\|A^2\|_2 \approx 4{,}0$
  - Pînă la un factor k^{m-1} dacă A are un bloc Jordan de dimensiune m
  - Bloc Jordan: o valoare proprie multiplă fără un set complet de vectori proprii
- DUx_t este o inovație doar dacă intrările x_t sînt zgomot alb centrat
- Intrările cu memorie lungă s-ar transmite mai departe prin rețea

Înapoi

Descreșterea geometrică și cea hiperbolică, ajustate pe date

- Cu porți constante: $\text{Corr}(c_t, c_{t-k}) = f^k$
  - Valori candidate g_t i.i.d. și inițializare în distribuția staționară
- Pentru un model ARFIMA cu $0 < d < 1/2$: $\rho_k \sim Ck^{2d-1}$
  - ARFIMA: autoregressive fractionally integrated moving average, autoregresiv fracționar integrat cu medie mobilă
- Randamentele absolute ale S&P 500 din 2000: funcția de autocorelație (ACF)
  - ACF este 0,289 la lag-ul 1 și 0,132 la lag-ul 50
  - Descreșterea geometrică cu amplitudine proprie, $\rho_k = \rho_1 f^{k-1}$ (un semnal AR(1) observat cu zgomot), reproduce autocorelațiile la ambele lag-uri pentru $f = 0{,}984$
  - Acestui f îi corespund $b_f = 4{,}1$ și un timp de înjumătățire de 43 de zile
  - Acest f implică o autocorelație de 0,027 la lag-ul 150
  - Datele arată 0,059, de peste două ori mai mult
- O poartă constantă $f < 1$ este un filtru cu memorie scurtă
- Un context lung sau medii lungi explicite, ca în HAR (Testul 2), doar aproximează memoria lungă
  - Pe un interval finit de lag-uri

Înapoi

Alte modele fundaționale și corpusurile lor (1/2)

- Chronos-2: corpus format din seturi selectate din Chronos și GIFT-Eval, plus serii multivariate sintetice
- Teste comparative (repere): GIFT-Eval (*General Time Series Forecasting Model Evaluation*, setul de referință GIFT-Eval pentru evaluarea modelelor generale de prognoză a seriilor de timp) (Aksu et al. (2024)) și fev-bench
- Corpusul TimesFM
  - Google Trends, afișări de pagini Wikipedia, serii sintetice
  - Seriile competiției de prognoză M4, electricitate, trafic, vreme
- Datele de publicare a ponderilor, din depozitele Hugging Face
  - TimesFM-2.5: versiunea actualizată a ponderilor a fost anunțată pe 15.09.2025 și publicată pe 01.10.2025
  - Chronos-2: anunțat pe 20.10.2025, publicat pe 30.10.2025
- Checkpoint-urile folosite: revizia fixată în cod și primele 8 caractere ale amprentei SHA-256 a fișierului de ponderi
  - Chronos-2: `ddcda3c7`, în depozit din 30.10.2025
  - Chronos-Bolt (small): `06a6a19b`, din 13.11.2024
  - TimesFM-2.5: `2f776efe`, din 01.10.2025

Înapoi

Alte modele fundaționale și corpusurile lor (2/2)

- **Moirai** (Woo et al. (2024)): un singur Transformer pentru orice frecvență și orice număr de variabile
- Ieșirea este un amestec de distribuții

$$p(y) = \sum_{k=1}^{4} w_k \, p_k(y), \qquad w_k \geq 0, \quad \sum_{k=1}^{4} w_k = 1$$

  - p_k: densitățile Student-t, log-normală, binomială negativă și Normală cu varianță mică
  - Ponderile w_k și parametrii componentelor sînt ieșiri ale rețelei, deci forma distribuției se adaptează seriei
- Versiunea Moirai 2.0: Liu et al. (2025)
- **Lag-Llama** (Rasul et al. (2023)): model doar cu decodor
  - Cu valori întîrziate ca variabile și ieșire Student-t

Înapoi

Un protocol formal pentru contaminare (1/2)

- Regresia înainte/după pe diferențialul de pierdere $d_t = L_t^{FM} - L_t^{reper}$, cu erori standard HAC

$$d_t = \beta_0 + \beta_1 1\{t \geq \tau\} + u_t$$

  - L_t^{FM}, L_t^{reper}: pierderile modelului fundațional și ale reperului în ziua t
  - τ: data publicării ponderilor (o dată-limită comună cînd se compară mai multe modele)
  - Memorarea ar implica $\beta_1 > 0$; ipoteza nulă H_0: $\beta_1 = 0$
- Raportați efectul minim detectabil (MDE) al lui β_1 (secțiunea AI)
  - Pragul de 0,0157 folosește varianța de termen lung a lui d_t în fiecare perioadă: zilele dinaintea publicării, care includ șocul din aprilie 2025, sînt de circa 2,9 ori mai variabile decît cele de după
  - Cu o variantă comună, pragul ar fi 0,0111

Înapoi

Un protocol formal pentru contaminare (2/2)

- Preînregistrarea: prognoze depuse înainte ca rezultatul să existe
  - Contaminarea devine imposibilă prin construcție
  - Repere evaluate în timp real: TS-Arena (serii de timp, în numele platformei TS-Arena) (Meyer et al. (2025)), Impermanent (Garza et al. (2026))
  - Fereastra noastră de după publicare este o versiune manuală a aceleiași idei
- Ablații: eliminarea componentei de model de limbaj din prognozele bazate pe LLM nu reduce precizia (Tan et al. (2024))
  - Adesea o îmbunătățește
- Prețurile energiei: Pan & Ezzat (2026) testează pe două seturi de date pentru a limita contaminarea
- Rezultatul: modelele fundaționale sînt competitive
  - Dar nu depășesc constant metodele specifice domeniului

Înapoi

Prognozele randamentelor, tabelul complet

R^2 (%) [DM p]	Media istorică	AR(1)	LSTM (5 seed-uri)	Chronos-2	Chronos-Bolt	TimesFM-2.5
S&P 500	+0,15 [0,095]	+0,94 [0,639]	+0,93 [0,090]	−0,26 [0,418]	−1,57 [0,078]	−3,35 [0,001]
Bitcoin	+0,07 [0,691]	−0,28 [0,326]	+0,04 [0,243]	−0,35 [0,413]	−0,61 [0,070]	−2,29 [0,001]
BET	+0,30 [0,081]	−0,00 [1,000]	+1,15 [0,014]	−0,39 [0,562]	+0,12 [0,831]	−3,27 [0,070]

- Prognoze punctuale: medianele din grilă (mediile din grilă: protocolul testului pentru randamente)
- DM: test bilateral pe erorile pătratice față de prognoza zero, variantă Newey–West

înapoi

Prognoze imbricate, AR(1) pe BET și ferestre mobile

- AR(1) pe BET ($R^2 = -0{,}00\%$, p DM $= 1{,}000$) nu este o prognoză degenerată
  - Abaterea standard a prognozei este 0,074%, iar panta medie este 0,046
  - Reducerile erorii pătratice din 2020 ($-12{,}7$) și din 2026 ($+14{,}7$) se anulează reciproc
  - Fără acești doi ani, $R^2 = -0{,}11\%$
- Modele pe ferestre mobile sau reantrenate: Giacomini & White (2006) testează metoda de prognoză
  - Inclusiv fereastra de estimare
  - Diebold (2015) despre folosirea și abuzul testului DM
- SPA rafinează Reality Check al lui White (2000)
  - Modelele foarte slabe (TimesFM) nu mai diluează testul
- Ambele teste folosesc un bootstrap staționar cu lungimea medie a blocului de 20 de zile

Înapoi

Testarea multiplă (1/2): Holm

- **Holm**: ordonăm cele M valori p, $p_{(\mathbf{1})} \leq \cdots \leq p_{(M)}$
  - Respingem cît timp $p_{(j)} < 0{,}05/(M - j + 1)$
  - Ne oprim la prima nerespingere
- Valoarea p ajustată este $\max_{i \leq j} \min\{1, (M - i + 1)\, p_{(i)}\}$
  - Exemplu, $M = 3$, $p = 0{,}01$; $0{,}02$; $0{,}04$: pragurile $0{,}05/3 = 0{,}0167$, $0{,}05/2 = 0{,}025$, $0{,}05$: toate trei sînt respinse
  - Bonferroni (pragul $0{,}0167$ pentru toate) o respinge doar pe prima
  - Valorile p ajustate: $0{,}03$; $0{,}04$; $0{,}04$
  - În capitol: LSTM pe BET are cel mai mic dintre cele 18 valori p DM unilaterale, $0{,}0072$, deci valoarea p ajustată este $18 \times 0{,}0072 = 0{,}130$

Înapoi

Testarea multiplă (2/2): Romano–Wolf și SPA

- **Romano–Wolf**: statisticile $t_k = \sqrt{T}\,\bar{d}_k/\hat{\omega}_k$, ordonate $t_{(\mathbf{1})} \geq \cdots \geq t_{(K)}$
  - $\bar{d}_k$: cîștigul mediu de pierdere al modelului k față de prognoza zero
  - $\hat{\omega}_k$: abaterea lui standard HAC

$$p^{RW}_{(j)} = \max_{i \leq j} P^*\Big(\max_{k \in \mathcal{A}_i} t^*_k \geq t_{(i)} \Big), \qquad t^*_k = \frac{\sqrt{T}\,(\bar{d}^*_k - \bar{d}_k)}{\hat{\omega}_k}$$

  - $\mathcal{A}_i$: ipotezele încă nerespinse la pasul i, adică $(i), \ldots, (K)$; P^*, $\bar{d}^*_k$: probabilitatea și media sub bootstrap-ul staționar
  - Pragul vine din maximul statisticilor corelate, deci procedura este mai puțin conservatoare decît Holm pentru prognoze ale aceleiași piețe
- **SPA**: un singur test, cu statistica $\max\left(0, \max_k t_k\right)$
  - H_0: niciun model nu are o pierdere așteptată mai mică decît reperul
  - Valoarea p vine dintr-un bootstrap staționar; modelele mult mai slabe ($t_k \leq -\sqrt{2 \ln \ln T}$) nu sînt recentrate la zero, deci nu diluează testul
- La ce întrebare răspunde fiecare
  - Holm și Romano–Wolf: *care* comparații sînt semnificative, cu o probabilitate de cel mult 5% a vreunei respingeri false
  - SPA: *dacă vreun* model depășește reperul

înapoi

QLIKE, DM și MCS, tabelul complet

Model	QLIKE	DM față de log-HAR	p	MCS p	QLIKE după 3.11.2025
Mers aleator	0,313	+4,44	<0,001	0,005	0,356
HAR	0,238	+1,17	0,241	0,33	0,251
log-HAR	0,226	–	–	0,95	0,223
GARCH-t	0,273	+3,45	<0,001	0,03	0,279
LSTM	0,275	+1,66	0,097	0,29	0,217
Chronos-2	0,223	−0,61	0,539	1,00	0,226
Chronos-Bolt	0,224	−0,20	0,842	0,95	0,219
TimesFM-2.5	0,225	−0,07	0,948	0,95	0,226

- Literatura: Brini (2026) compară nouă modele fundaționale zero-shot cu opt modele de tip HAR
  - Pe 50 de active
- Pierderile agregate favorizează modelele fundaționale din cauza cîtorva active atipice
  - Activ cu activ, doar un model mic depășește log-HAR la toate orizonturile, la limită
- Goel et al. (2025): TimesFM zero-shot este un reper rezonabil
  - Dar pentru a depăși modelele econometrice este necesar fine-tuning-ul incremental

Înapoi

Depășirile VaR 1% și testele Kupiec, tabelul complet

Depășiri [Kupiec p]	HS	GARCH-t	FHS	Cuantile Chronos-2	FHS + vol. Chronos-2	FHS + vol. TimesFM
S&P 500 (26,9)	37 [0,065]	44 [0,002]	30 [0,559]	45 [0,001]	25 [0,705]	27 [0,989]
Bitcoin (28,2)	19 [0,065]	27 [0,822]	25 [0,539]	33 [0,374]	27 [0,822]	27 [0,822]
BET (26,8)	32 [0,322]	41 [0,010]	30 [0,536]	40 [0,016]	26 [0,884]	32 [0,322]

- În paranteză după piață: numărul așteptat de depășiri, 1% din zile
- Kupiec: testul raportului de verosimilitate pentru o rată de depășire $\alpha = 1\%$
  - Cu x depășiri în T zile și $\hat{\pi} = x/T$

$$\mathrm{LR}_{uc} = -2 \ln \frac{(1-\alpha)^{T-x}\alpha^{x}}{(1-\hat{\pi})^{T-x}\hat{\pi}^{x}} \sim \chi^2_1$$

Înapoi

Crahul COVID-19, scalarea și reacția excesivă

- Chronos-2 standardizează contextul prin media și abaterea standard, apoi aplică $\sinh^{-1}$ (Ansari et al. (2025))
  - La 16.03.2020, crahul a mărit abaterea standard a contextului de 512 zile cu 9,7%
  - Scala din Chronos (media valorilor absolute) a crescut cu 3,2% (Seminarul 14, exercițiul A6)
  - Deci lărgirea combină schimbarea scalei cu citirea zilelor de crah
  - Doar o ablație le-ar putea separa
- Nu este o eroare de unități
  - După ziua crahului Chronos-2 a prognozat o medie a lui $|r|$ de 12,5% pentru ziua următoare
  - Apoi FHS o înmulțește cu cuantila de 1% a lui $r/\hat{\sigma}$ (4,7)
  - Filtrul extrapolează un salt de o zi ca pe un nou nivel: VaR de 4,6 ori pierderea maximă, un cost de capital ridicat pe care pierderea FZ0 îl penalizează

Înapoi

Testele Berkowitz și scorurile cuantilice, trei piețe

Piața	p Berkowitz: Chronos-2	FHS	Sub $q_{1\%}$ (C2)	QS: C2 / FHS	DM [p]
S&P 500	0,019	0,599	1,67%	0,510 / 0,506	+3,06 [0,002]
Bitcoin	0,595	0,306	1,17%	1,526 / 1,529	−0,90 [0,369]
BET	0,045	0,436	1,50%	0,465 / 0,460	+3,48 [<0,001]

- Testul Berkowitz evaluează funcția de repartiție interpolată liniar pe grilă
  - Și doar μ, σ, nu independența în timp a PIT
- QS ponderează egal cele 21 de niveluri
  - CRPS ar pondera fiecare nivel cu distanța dintre niveluri
- C2: Chronos-2
- Pe Bitcoin, nu detectăm o diferență între Chronos-2 și FHS

înapoi

Pierderile FZ0 și valorile p MCS, tabelul complet

FZ0 [MCS p]	HS	GARCH-t	FHS	Cuantile Chronos-2	FHS + vol. Chronos-2	FHS + vol. TimesFM
S&P 500	1,309 [0,14]	1,026 [0,19]	0,999 [1,00]	1,147 [0,14]	1,084 [0,14]	1,059 [0,18]
Bitcoin	2,237 [0,34]	2,192 [0,60]	2,192 [0,60]	2,181 [0,64]	2,151 [1,00]	2,160 [0,70]
BET	1,230 [0,015]	1,001 [0,34]	0,963 [1,00]	1,032 [0,24]	1,007 [0,34]	1,027 [0,20]

- Bitcoin: FHS hibrid cu volatilitatea Chronos-2 are cea mai mică pierdere FZ0
  - Dar diferențele nu sînt semnificative
- FZ0 pentru (VaR 2,5%, ES 2,5%)
- MCS folosește statistica $T_{\max}$ și un bootstrap pe blocuri

înapoi

Recalibrarea conformală, reguli și rezultate

După 250 de zile de calibrare: depășiri [p cc / p DQ]	FHS	Chronos-2 direct	FHS + vol. C2	Chronos-2 + ACI
S&P 500 (24,4)	28 [0,097 / 0,014]	42 [0,002 / <0,001]	23 [0,058 / <0,001]	21 [0,307 / 0,049]
Bitcoin (25,7)	24 [0,449 / 0,108]	31 [0,410 / 0,026]	26 [0,087 / 0,004]	24 [0,449 / 0,191]
BET (24,2)	26 [0,709 / 0,727]	36 [0,048 / 0,184]	23 [0,776 / 0,011]	25 [0,762 / 0,032]
FZ0, S&P 500 [p DM]	1,013	1,142 [0,009]	1,087 [0,022]	1,205 [0,001]

- $\hat{Q}$: al $\lceil (1-\alpha_t)(250+1) \rceil$-lea cel mai mic scor trecut ($\lceil \cdot \rceil$: rotunjirea în sus)
- $\hat{Q} = +\infty$ dacă rangul depășește 250 sau $\alpha_t \leq 0$, iar $\hat{Q} = -\infty$ dacă $\alpha_t \geq 1$
- Cu $\gamma = 0{,}005$, VaR-ul este infinit în 24 de zile pe S&P 500, 100 pe Bitcoin și 270 pe BET
- Garanția teoretică pentru $T = 2\,443$ și $\gamma = 0{,}001$: limita este 0,41, neinformativă pentru o rată de 1%
- Metoda conformală ponderată limitează abaterea produsă de drift (Barber et al. (2023))
  - Fără interschimbabilitate
- ES se poate evalua doar împreună cu VaR
  - Prin pierderi consistente pentru pereche, ca FZ0 (Nolde & Ziegel (2017))
  - Comparațiile ES au limite de precizie în eșantioane finite (Pele & Mazurencu-Marinescu-Pele (2026))
- O coadă modelată cu distribuția Pareto generalizată (GPD, McNeil & Frey (2000))
  - Este o alternativă pentru ES

înapoi

Studiul de caz, replicarea cu datele cursului

- Randamentele totale lunare ale SPY din 1993 (datele cursului)
  - În exces față de randamentul titlurilor de stat americane la 3 luni (FRED `TB3MS`)
  - Predictorii din Welch & Goyal (2008) disponibili în datele cursului sînt 7 din cei 15
  - Din FRED: dobînda la titlurile de stat pe termen scurt (`tbl`), randamentul pe termen lung (`lty`), panta curbei (`tms`), spread-ul de credit (`dfy`), inflația (`infl`)
  - Varianța acțiunilor (`svar`), calculată din randamentele zilnice ale S&P 500
  - Randamentul în exces întîrziat
  - Limitarea datelor: rapoartele de evaluare (dividend/preț `dp`, randamentul dividendului `dy`, cîștig/preț `ep`, rata de distribuire `de`, valoare contabilă/de piață `b/m`), emisiunile nete de acțiuni (`ntis`) și randamentele obligațiunilor (pe termen lung `ltr`, spread-ul randamentelor corporative `dfr`) nu sînt în datele cursului
  - Eșantionul acoperă circa 30 de ani, în loc de 90
  - Metoda rămîne aceeași: $\gamma = 2$, $T = 12$, $P \leq 12\,000$, $\log_{\mathbf{10}} z = -3, \ldots, 3$, $1\,000$ de extrageri
- Rezultatele pe datele cursului
  - kitchen sink liniar, $z = 0^+$ ($c = 7/12$): $R^{\mathbf{2}} = -847\%$, raportul Sharpe $-0{,}03$
  - liniar, $z = 10^{\mathbf{3}}$: $R^{\mathbf{2}} = 0{,}26\%$, raportul de informație față de SPY $0{,}25$ ($t = 1{,}3$)
  - RFF, $c = 1\,000$, $z = 10^{\mathbf{3}}$: $R^{\mathbf{2}} = -5{,}7\%$, raportul de informație față de SPY $-0{,}12$ ($t = -0{,}6$) și față de modelul liniar $0{,}07$ ($t = 0{,}4$)

Înapoi

Alinierea lunară și inflația lipsă

- Fiecare lună a predictorilor prognozează randamentul lunii calendaristice următoare
- Un predictor lipsă nu mută ținta într-o lună ulterioară
  - Lag-urile și țintele păstrează calendarul lunar
- Inflația este cea cunoscută la sfîrșitul fiecărei luni: vintage-ul ALFRED (*ArchivaL Federal Reserve Economic Data*, arhiva versiunilor istorice ale seriilor FRED) al seriei CPIAUCSL (*FRED series: Consumer Price Index for All Urban Consumers, all items*, seria FRED a indicelui prețurilor de consum din SUA (Statele Unite ale Americii)) din acea zi, fără revizii ulterioare
  - Seria revizuită de azi ar introduce deplasarea de anticipare (corelația lunară cu seria în timp real: 0,94)
  - Octombrie 2025 nu a fost publicat: în decembrie 2025 inflația din noiembrie nu se poate calcula și nu se imputează
- Țintele de test merg din februarie 1997 pînă în august 2026: 354 de prognoze
  - Fără ianuarie 2026
- Scalarea volatilității folosește 12 luni calendaristice, înainte de eliminarea rîndurilor incomplete
- Regresia folosește ultimele 12 observații complete
  - În jurul datelor lipsă, fereastra acoperă mai mult de 12 luni

Înapoi

Bucla descoperirii, prompturi reutilizabile (1/2)

- **Literatura**: `"Găsește evaluări ale modelelor fundaționale pentru serii de timp pe randamente sau volatilitate financiară din 2024 încoace: modele, date, pierdere, reper și dacă perioada de test se suprapune cu corpusul de pre-antrenare."`
  - Semantic Scholar, Google Scholar, Elicit ajută la căutare
- **Ipoteze**: `"Enumeră predicțiile care separă memorarea de abilitate pentru un model zero-shot de volatilitate: înainte și după publicare, serii niciodată publice, serii simulate cu aceeași dinamică."`
  - Diagnostice, nu dovezi: o traiectorie log-HAR simulată testează transferul la acel proces
  - O diferență înainte/după poate reflecta și o schimbare de regim
- **Date și cod**: `"Din prognozele zilnice ale variantei realizate SPY date de Chronos-2 și log-HAR, calculează diferențialul QLIKE Chronos-2 minus log-HAR, testează diferența dintre perioada de dinainte și cea de după 3 noiembrie 2025 cu erori HAC și raportează efectul minim detectabil la o putere de 80%."`
  - Orice asistent (Claude, ChatGPT, Gemini, GitHub Copilot)

Înapoi

Bucla descoperirii, prompturi reutilizabile (2/2)

- **Robustețea**: `"Proiectează un test preînregistrat cu putere mai mare: ce secțiune transversală de serii, ce aproximare a variației, ce test DM pe panel și cîte zile de după publicare."`
- **Critica**: `"Atacă afirmația: nu există memorare, pentru că testul DM de după publicare este nesemnificativ. Ce este greșit în inferență?"`
  - De așteptat: putere mică, absența dovezii și regimuri diferite: șocul tarifar din aprilie 2025 cade în perioada dinaintea publicării, fereastra de după publicare este mai calmă
- **Raportul**: versiunile modelelor și datele publicării ponderilor
  - Lungimea contextului, media sau mediana
  - Fiecare seed și fereastră încercată

Înapoi

Lecturi suplimentare

- Gu, Kelly & Xiu (2020): machine learning pentru secțiunea transversală a randamentelor așteptate
- Kelly, Malamud & Zhou (2024): de ce modelele foarte mari pot prognoza mai bine randamentele
- Ansari et al. (2024): Chronos, seriile de timp citite ca un limbaj
- Aplicații pe criptoactive
  - Prognoza lichidității Agakishiev, Härdle & Pele (2025)
  - Avertizare timpurie pentru criptoactivele „zombie" Będowska-Sójka, Wójcik & Pele (2026)
  - Clasificarea statistică a criptoactivelor Pele et al. (2023)
- Manuale: Franke, Härdle & Hafner (2019) și Härdle, Chen & Overbeck (2017)
  - Pentru statistica piețelor financiare din spatele modelelor de referință

Înapoi

Acronime folosite în acest material (1/4)

- **ACF**: AutoCorrelation Function — *funcția de autocorelație*
- **ACI**: Adaptive Conformal Inference — *inferență conformală adaptivă*
- **AI**: Artificial Intelligence — *inteligență artificială*
- **AR**: AutoRegressive model — *model autoregresiv*
- **ARFIMA**: AutoRegressive Fractionally Integrated Moving Average (model) — *model autoregresiv cu medie mobilă, integrat fracționar*
- **BET**: Bucharest Exchange Trading (index) — *indicele de referință al Bursei de Valori București*
- **CI**: Confidence Interval — *interval de încredere*
- **COVID-19**: COronaVIrus Disease 2019 — *boala provocată de coronavirus, 2019*
- **CPU**: Central Processing Unit — *procesorul central al calculatorului*
- **CRPS**: Continuous Ranked Probability Score — *scorul de probabilitate ordonat continuu*
- **CRSP**: Center for Research in Security Prices — *Centrul de cercetare a prețurilor titlurilor (Universitatea din Chicago)*
- **DM**: Diebold–Mariano (test) — *testul Diebold–Mariano de comparare a prognozelor*
- **DQ**: Dynamic Quantile (test) — *testul cuantilei dinamice*
- **ES**: Expected Shortfall — *pierderea așteptată în coadă*
- **ETF**: Exchange-Traded Fund — *fond tranzacționat la bursă*
- **EWMA**: Exponentially Weighted Moving Average — *medie mobilă ponderată exponențial*

Acronime folosite în acest material (2/4)

- **FHS**: Filtered Historical Simulation — *simulare istorică filtrată*
- **FM**: Foundation Model — *model fundațional*
- **FRED**: Federal Reserve Economic Data — *baza de date economice a Rezervei Federale (St. Louis)*
- **FRTB**: Fundamental Review of the Trading Book — *standardul Basel pentru riscul de piață al portofoliului de tranzacționare*
- **FZ0**: Fissler–Ziegel loss, zero-homogeneous version — *funcția de pierdere Fissler–Ziegel, varianta omogenă de grad zero*
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity — *heteroscedasticitate condiționată autoregresivă generalizată*
- **GIFT**: General Time Series Forecasting Model Evaluation (benchmark GIFT-Eval) — *setul de referință GIFT-Eval pentru evaluarea modelelor generale de prognoză a seriilor de timp*
- **GPD**: Generalised Pareto Distribution — *distribuția Pareto generalizată*
- **GRU**: Gated Recurrent Unit — *unitate recurentă cu porți*
- **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*

Acronime folosite în acest material (3/4)

- **HS**: Historical Simulation — *simulare istorică*
- **i.i.d.**: independent and identically distributed — *independente și identic distribuite*
- **IR**: Information Ratio — *raportul de informație*
- **LASSO**: Least Absolute Shrinkage and Selection Operator — *regresie penalizată L1 cu selecția variabilelor*
- **LLM**: Large Language Model — *model lingvistic mare*
- **LN**: Layer Normalisation — *normalizarea pe strat*
- **LR**: Likelihood Ratio (test) — *testul raportului de verosimilitate*
- **LSTM**: Long Short-Term Memory (network) — *rețea cu memorie pe termen lung și scurt*
- **M4**: the fourth Makridakis forecasting competition — *a patra competiție de prognoză Makridakis*
- **M6**: the sixth Makridakis forecasting competition — *a șasea competiție de prognoză Makridakis*
- **MCS**: Model Confidence Set — *mulțimea de încredere a modelelor*
- **MDE**: Minimum Detectable Effect — *efectul minim detectabil*
- **MHA**: Multi-Head Attention — *atenția cu mai multe capete*
- **MLP**: MultiLayer Perceptron — *perceptron multistrat*
- **MSE**: Mean Squared Error — *eroarea pătratică medie*
- **NBER**: National Bureau of Economic Research — *Biroul Național de Cercetări Economice (SUA)*

Acronime folosite în acest material (4/4)

- **PIT**: Probability Integral Transform — *transformarea integrală de probabilitate*
- **QLIKE**: Quasi-LIKElihood (loss function) — *funcția de pierdere de tip cvasi-verosimilitate*
- **QS**: Quantile Score — *scorul cuantilic*
- **ReLU**: Rectified Linear Unit — *unitatea liniară rectificată*, $\max(0, z)$
- **RFF**: Random Fourier Features — *trăsături Fourier aleatoare*
- **RNN**: Recurrent Neural Network — *rețea neuronală recurentă*
- **RoPE**: Rotary Position Embeddings — *codificări rotative ale poziției*
- **RPS**: Ranked Probability Score — *scorul de probabilitate pe clase ordonate*
- **RV**: Realised Variance — *varianța realizată*
- **SPA**: Superior Predictive Ability (test) — *testul abilității predictive superioare*
- **SSE**: Sum of Squared Errors — *suma pătratelor erorilor*
- **SUA**: Statele Unite ale Americii
- **T5**: Text-to-Text Transfer Transformer — *modelul de limbaj de tip Transformer folosit de Chronos*
- **TS**: Time Series (in the name TS-Arena) — *serii de timp (în numele platformei TS-Arena)*
- **VAR**: Vector AutoRegression — *model vector autoregresiv*
- **VaR**: Value-at-Risk — *valoarea la risc*

Bibliografie (1/7)

- Agakishiev, I., Härdle, W.K., Pele, D.T. (2025). Cryptocurrency liquidity forecasting. In *Springer Handbooks of Computational Statistics*, 41–60. Springer.
- Aksu, T., Woo, G., Liu, J., Liu, X., Liu, C., Savarese, S., Xiong, C., Sahoo, D. (2024). GIFT-Eval: a benchmark for general time series forecasting model evaluation. arXiv:2410.10393.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), 579–625.
- Ansari, A.F., Shchur, O., Küken, J., Auer, A., et al. (2025). Chronos-2: from univariate to universal forecasting. arXiv:2510.15821.
- Ansari, A.F., Stella, L., Turkmen, C., Zhang, X., et al. (2024). Chronos: learning the language of time series. *Transactions on Machine Learning Research*; arXiv:2403.07815.
- Barber, R.F., Candès, E.J., Ramdas, A., Tibshirani, R.J. (2023). Conformal prediction beyond exchangeability. *Annals of Statistics*, 51(2), 816–845.
- Będowska-Sójka, B., Wójcik, P., Pele, D.T. (2026). Early warning systems for cryptocurrency markets: predicting 'zombie' assets using machine learning. *North American Journal of Economics and Finance*, 81, 102543.
- Bengio, Y., Simard, P., Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166.
- Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4), 465–474.

Bibliografie (2/7)

- Brini, A. (2026). Forecasting realized volatility with time series foundation models: a comparison with econometric benchmarks. arXiv:2607.05291.
- Campbell, J.Y., Thompson, S.B. (2008). Predicting excess stock returns out of sample: can anything beat the historical average? *Review of Financial Studies*, 21(4), 1509–1531.
- Christoffersen, P.F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862.
- Clark, T.E., West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291–311.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174–196.
- Das, A., Kong, W., Sen, R., Zhou, Y. (2024). A decoder-only foundation model for time-series forecasting. *Proceedings of ICML*, PMLR 235; arXiv:2310.10688.
- Diebold, F.X. (2015). Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics*, 33(1), 1.
- Diebold, F.X., Gunther, T.A., Tay, A.S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4), 863–883.
- Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.

Bibliografie (3/7)

- Engle, R.F., Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4), 367–381.
- Fischer, T., Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669.
- Fissler, T., Ziegel, J.F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, 44(4), 1680–1707.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). Statistics of Financial Markets: An Introduction, 5th ed. Springer.
- Garza, A., Rosillo, R., Mendoza-Smith, R., Salinas, D., Williams, A.R., Ashok, A. (2026). Impermanent: a live benchmark for temporal generalization in time series forecasting. arXiv:2603.08707.
- Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578.
- Gibbs, I., Candès, E. (2021). Adaptive conformal inference under distribution shift. *NeurIPS*; arXiv:2106.00170.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494), 746–762.
- Gneiting, T., Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.

Bibliografie (4/7)

- Goel, A., Pasricha, P., Magris, M., Kanniainen, J. (2025). Foundation time-series AI model for realized volatility forecasting. arXiv:2505.11163.
- Gu, S., Kelly, B., Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223–2273.
- Hansen, P.R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4), 365–380.
- Hansen, P.R., Lunde, A., Nason, J.M. (2011). The model confidence set. *Econometrica*, 79(2), 453–497.
- Härdle, W.K., Chen, C.Y.-H., Overbeck, L. (eds.) (2017). Applied Quantitative Finance, 3rd ed. Springer.
- Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.
- Kelly, B., Malamud, S., Zhou, K. (2024). The virtue of complexity in return prediction. *Journal of Finance*, 79(1), 459–503.
- Kingma, D.P., Ba, J. (2015). Adam: a method for stochastic optimization. *ICLR*; arXiv:1412.6980.

Bibliografie (5/7)

- LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Liu, C., Aksu, T., Liu, J., Liu, X., et al. (2025). Moirai 2.0: when less is more for time series forecasting. arXiv:2511.11698.
- Lo, A.W., MacKinlay, A.C. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, 45(1–2), 181–211.
- Lopez-Lira, A., Tang, Y., Zhu, M. (2025). The memorization problem: can we trust LLMs' economic forecasts? arXiv:2504.14765.
- Makridakis, S., Spiliotis, E., Hollyman, R., Petropoulos, F., Swanson, N., Gaba, A. (2025). The M6 forecasting competition: bridging the gap between forecasting and investment decisions. *International Journal of Forecasting*, 41(4), 1315–1354.
- McNeil, A.J., Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3–4), 271–300.
- Meyer, M., Kaltenpoth, S., Albers, H., Zalipski, K., Müller, O. (2025). TS-Arena: a live forecast pre-registration platform. arXiv:2512.20761.
- Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J. (2023). A time series is worth 64 words: long-term forecasting with Transformers. *ICLR*; arXiv:2211.14730.
- Nolde, N., Ziegel, J.F. (2017). Elicitability and backtesting: perspectives for banking regulation. *Annals of Applied Statistics*, 11(4), 1833–1874.
- Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y. (2020). N-BEATS: neural basis expansion analysis for interpretable time series forecasting. *ICLR*; arXiv:1905.10437.

Bibliografie (6/7)

- Pan, Z., Ezzat, A.A. (2026). Evaluating time series foundation models for electricity price forecasting: contamination risk, distributional shifts, and covariate dependence. arXiv:2607.02623.
- Pascanu, R., Mikolov, T., Bengio, Y. (2013). On the difficulty of training recurrent neural networks. *ICML*; arXiv:1211.5063.
- Patton, A.J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256.
- Patton, A.J., Ziegel, J.F., Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2), 388–413.
- Pele, D.T., Mazurencu-Marinescu-Pele, M. (2026). Finite-sample precision limits for expected shortfall forecast comparisons. *Mathematics*, 14(13), 2316.
- Pele, D.T., Wesselhöfft, N., Härdle, W.K., Kolossiatis, M., Yatracos, Y.G. (2023). Are cryptos becoming alternative assets? *The European Journal of Finance*, 29(10), 1064–1105.
- Rahimi, A., Recht, B. (2007). Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, 20.
- Rahimikia, E., Ni, H., Wang, W. (2025). Re(Visiting) time series foundation models in finance. arXiv:2511.18578.
- Rasul, K., Ashok, A., Williams, A.R., Ghonia, H., et al. (2023). Lag-Llama: towards foundation models for probabilistic time series forecasting. arXiv:2310.08278.

Bibliografie (7/7)

- Romano, J.P., Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4), 1237–1282.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.
- Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T. (2020). DeepAR: probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181–1191.
- Tan, M., Merrill, M.A., Gupta, V., Althoff, T., Hartvigsen, T. (2024). Are language models actually useful for time series forecasting? *NeurIPS*; arXiv:2406.16964.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., et al. (2017). Attention is all you need. *NeurIPS*; arXiv:1706.03762.
- Welch, I., Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455–1508.
- White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5), 1097–1126.
- Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., Sahoo, D. (2024). Unified training of universal time series forecasting Transformers. *ICML*; arXiv:2402.02592.