








Modelarea piețelor financiare

Capitolul 13: Machine Learning în finanțe



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

Ce poate prezice machine learning în finanțe?



Campbell R. Harvey. **Foto: Xuthoria (2026); CC BY-SA 4.0; Wikimedia Commons**

"Hundreds of papers and factors attempt to explain the cross-section of expected returns. Given this extensive data mining, it does not make sense to use the usual criteria for establishing significance."

— *Harvey, Liu & Zhu (2016)*

- După sute de factori testați pe aceleași date, un factor nou are nevoie de $t > 3$, nu de $t > 2$
- Direcția S&P 500 (*Standard & Poor's*) pe 5 zile, $n = 6\,269$: niciun model nu depășește reperul „mereu în sus"
  - reperul are acuratețea 58,1%; logit-ul are 57,7% și prezice „sus" în 98,3% din zile
- Backtest pe ferestre succesive (*walk-forward*) 2010–2026: raportul Sharpe al strategiei GBM (*Gradient Boosting Machine*, model de boosting pe gradient) este 0,33
  - raportul Sharpe al buy-and-hold este 0,66
- Pe o țintă fără predictibilitate, K-fold amestecat raportează acuratețea 69,4%
  - K-fold cu purjare și embargo raportează 49,6%

▸ Graficul central: Compararea modelelor: mediile pe pliuri

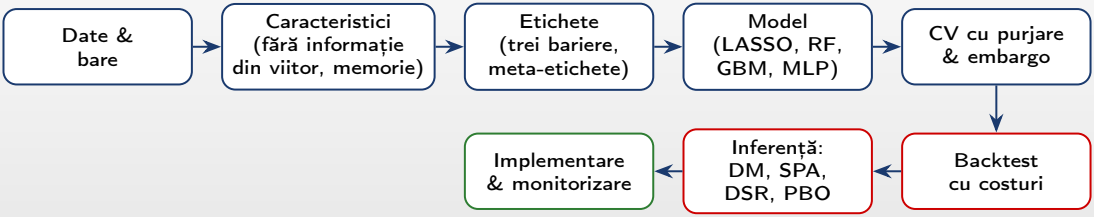
Ce leagă etapele analizei (1/2)

- Partea I: problema de predicție și metodele de bază
  - caracteristici și diferențiere fracționară
  - etichete
  - validarea fără informație din viitor
- Partea a II-a: un studiu de caz pe S&P 500 cu teste formale
  - selecția și supraajustarea (*overfitting*) backtest-ului
  - machine learning (ML) pentru inferență
  - un studiu de caz de cercetare
  - inteligența artificială (AI) pentru descoperire științifică
  - demonstrații și estimatori: Anexa
- Abrevierile din schemă
  - LASSO: regresie penalizată L_1
  - RF: pădure aleatoare (*random forest*)
  - GBM: gradient boosting machine
  - MLP: perceptron multistrat
  - CV: validare încrucișată

Ce leagă etapele analizei (2/2)

- Rezultatele învățării
  - calculați caracteristici și etichete fără informație din viitor
  - calculați scorul unui model validat cu purjare și embargo
  - derivați de ce FFD (*Fixed-width window Fractional Differentiation*, diferențiere fracționară cu fereastră fixă) trunchiată a unui preț $I(1)$ rămîne $I(1)$
  - interpretați testele unui clasificator și ale unui backtest față de un reper
  - Aprofundări și Anexa: lectură pentru familiarizare
- Testele folosite pentru această comparație
  - testele Pesaran–Timmermann (PT) și Diebold–Mariano (DM)
  - raportul Sharpe deflatat (DSR)
  - testul de abilitate predictivă superioară (SPA)
  - probabilitatea de supraajustare a backtest-ului (PBO)
- 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
- Majoritatea eșecurilor țin de proces, nu de algoritm (López de Prado (2018)); două erori domină
  - **deplasarea de anticipare** (*look-ahead bias*): folosirea unor informații indisponibile la momentul deciziei
  - **deplasarea de selecție** (*selection bias*): raportarea celei mai bune dintre multe încercări

Firul capitolului: de la date la o afirmație testată



- Caracteristicile folosesc doar informația disponibilă la data deciziei
- Validarea măsoară ce rămîne valabil pe observații nevăzute de model
- O direcție prezisă justifică o tranzacție doar după evaluarea incertitudinii și a costurilor

Machine learning în finanțe: promisiuni și dificultăți (1/2)

Promisiunea

- Forme funcționale flexibile, neliniare
  - interacțiuni descoperite automat
- Sute de predictori candidați
  - „factor zoo": Capitolul 3
- Date alternative: text, imagini, registre de ordine, date on-chain

Constrîngerile

- **Raport semnal-zgomot** foarte mic
  - o singură istorie, dependentă serial, care nu poate fi repetată
- **Nestaționaritate**: regimuri, rupturi structurale
  - piețele se adaptează strategiilor care le exploatează
- Etichetele trebuie construite explicit
  - erorile au costuri financiare reale
- Risc foarte mare de **supraajustare** (ajustarea zgomotului)

Machine learning în finanțe: promisiuni și dificultăți (2/2)

Ipoteza piețelor adaptive: Lo (2004)

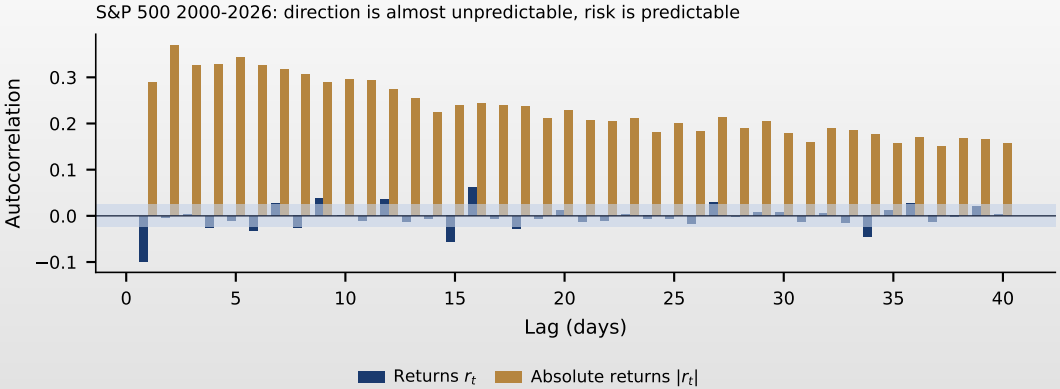
- Predictibilitatea apare, este eliminată prin arbitraj și reapare
  - un model care a funcționat ieri este o **ipoteză**, nu o lege
  - eficiența pieței: Capitolul 2
- Dificultatea nu constă în estimarea modelului, ci în **evitarea concluziilor false**
  - etichete corecte, validare corectă, backtest-uri oneste

Predictibilitatea direcției și a riscului (1/2)

- Întrebarea: ce parte a unei serii financiare este, în principiu, predictibilă?
- Autocorelația de ordin k: $\rho_k = \text{Corr}(x_t, x_{t-k})$
  - pentru randamentul logaritmic zilnic $r_t = \log P_t - \log P_{t-1}$ și pentru $|r_t|$
- Graficul următor: lagul k pe axa orizontală, $\hat{\rho}_k$ pe axa verticală
  - bare albastre pentru r_t, maro pentru $|r_t|$
- Banda albastru deschis, $\pm 1{,}96/\sqrt{T}$, este banda de 95% în absența autocorelației
  - T: numărul de observații
  - pentru $|r_t|$, banda presupune moment de ordin patru finit, pus sub semnul întrebării în Capitolul 1
- Autocorelațiile randamentelor r_t sînt mai mici decît cele ale lui $|r_t|$, dar 13 din cele 40 ies din bandă
  - la lagul 1, $\hat{\rho}_1 = -0{,}099$
  - banda nu este simultană pentru 40 de laguri și nu este robustă la heteroscedasticitate
  - depășirea benzii nu demonstrează profitabilitate după costuri
- Autocorelațiile lui $|r_t|$ sînt mari și persistente
- ML este cel mai util acolo unde datele au structură predictibilă
  - **volatilitate, risc, lichiditate, clasamente transversale**

MFM_ch13_model_comparison

Predictibilitatea direcției și a riscului (2/2)



MFM_ch13_model_comparison ▶ Interactiv

Dovezi și valoarea economică a unui R^2_{OOS} mic

- Gu, Kelly & Xiu (2020): randamente lunare ale acțiunilor americane individuale, 1957–2016
  - compară modele liniare, regresii penalizate, regresia pe componente principale (PCR)
  - și regresia prin cele mai mici pătrate parțiale (PLS), random forest, gradient boosting și rețele neuronale
- Măsura lor de potrivire: R^2 în afara eșantionului (*out-of-sample*, OOS); valoarea economică a unui R^2 mic: Aprofundări

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

  - cît din pătratul randamentului elimină prognoza, față de prognoza zero
  - r_t: randamentul realizat în luna t a perioadei de test; $\hat{r}_t$: prognoza făcută cu informația pînă la $t-1$
  - numitorul este eroarea prognozei zero, $\hat{r}_t = 0$: $R^2_{OOS} > 0$ dacă modelul o depășește, $R^2_{OOS} < 0$ dacă este mai slab
- Rezultate (Tabelul 1, R^2_{OOS} lunar)
  - rețeaua cu trei straturi 0,40%, gradient boosting 0,34%, random forest 0,33%
  - OLS (*Ordinary Least Squares*, metoda celor mai mici pătrate) cu 3 predictori 0,16%, OLS cu toți cei peste 900 de predictori −3,46%
- Avantajul arborilor și al rețelelor provine din **interacțiuni neliniare**
  - predictorii dominanți sînt momentum-ul, lichiditatea și volatilitatea
- Reperul contează
  - Welch & Goyal (2008) arată că majoritatea predictorilor publicați nu depășesc **media istorică**
  - Gu, Kelly & Xiu (2020) și Capitolul 14 folosesc o prognoză zero

Pierdere, penalizare și cinci familii de modele (1/5)

- Întrebarea secțiunii: de ce, cu un semnal slab, controlul variației contează mai mult decît flexibilitatea?
- Învățarea supervizată alege funcția care se potrivește cel mai bine datelor de antrenare

$$\hat{f} = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} L\big(y_i, f(\mathrm{x}_i)\big) + \lambda\, \Omega(f)$$

  - dintr-o familie de modele, după o penalizare a complexității
  - (y_i, x_i), $i = 1, \ldots, n$: ținta și vectorul de caracteristici ale observației i; $\mathcal{F}$: familia de modele (de ex. toți arborii de adîncime 3)
  - L: pierderea, eroarea pătratică $(y - f)^2$ sau, pentru o probabilitate f ca $y = 1$, log-loss-ul $-[y \log f + (1 - y) \log(1 - f)]$
  - $\Omega(f)$: penalizarea complexității, cu ponderea $\lambda \geq 0$; $\lambda = 0$: fără penalizare, λ mare: un model mai simplu
- λ se alege prin validare încrucișată (CV) în K pliuri $\mathcal{P}_1, \ldots, \mathcal{P}_K$:

$$CV(\lambda) = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i \in \mathcal{P}_k} L\big(y_i, \hat{f}_\lambda^{(-k)}(\mathrm{x}_i)\big), \qquad \hat{\lambda} = \arg\min_{\lambda} CV(\lambda)$$

  - n_k: numărul de observații din pliul k; $\hat{f}_\lambda^{(-k)}$: modelul estimat pe celelalte $K - 1$ pliuri

Pierdere, penalizare și cinci familii de modele (2/5)

- De ce penalizăm: eroarea de test așteptată într-un punct x_0 (Hastie, Tibshirani & Friedman (2009), cap. 7)

$$\mathbb{E}\left[(y_0 - \hat{f}(x_0))^2\right] = \underbrace{\sigma^2}_{\text{zgomot}} + \underbrace{\left(\mathbb{E}[\hat{f}(x_0)] - f(x_0)\right)^2}_{\text{deplasare}^2} + \underbrace{\text{Var}(\hat{f}(x_0))}_{\text{varianță}}$$

  - pentru $y = f(x) + \varepsilon$, cu $\mathbb{E}[\varepsilon] = 0$ și $\text{Var}(\varepsilon) = \sigma^2$
  - $\mathbb{E}$: media peste eșantioanele de antrenare posibile și peste zgomotul noii observații $y_0 = f(x_0) + \varepsilon_0$
  - deplasarea: eroarea sistematică a unei familii prea simple
  - varianța: sensibilitatea lui $\hat{f}$ la eșantionul de antrenare; crește cu flexibilitatea modelului
  - σ^2: varianța părții impredictibile a lui y, pe care niciun model nu o reduce
- Derivarea: adunăm și scădem $\mathbb{E}[\hat{f}(x_0)]$ în $y_0 - \hat{f}(x_0)$
  - termenii încrucișați au media zero, deoarece ε_0 este independent de eșantionul de antrenare
- În finanțe, prognozele de randament reduc foarte puțin eroarea
  - un R^2 în afara eșantionului sub 1%: modelul elimină mai puțin de 1% din pierderea pătratică a reperului
  - acest rezultat nu identifică σ^2: eroarea rămasă conține și misspecificare, variabile omise și eroare de estimare
  - deci penalizarea și medierea arborilor reduc varianța cu prețul unei deplasări mici

Pierdere, penalizare și cinci familii de modele (3/5)

- Regresiile penalizate, cu p caracteristici standardizate și coeficienții $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right)^2 + \lambda \left[\alpha \|\boldsymbol{\beta}\|_1 + \tfrac{1-\alpha}{2} \|\boldsymbol{\beta}\|_2^2\right]$$

  - $\|\boldsymbol{\beta}\|_1 = \sum_j |\beta_j|$, $\|\boldsymbol{\beta}\|_2^2 = \sum_j \beta_j^2$; $\alpha \in [0, 1]$: ponderea penalizării în valoare absolută
- Ridge, $\alpha = 0$ (Hoerl & Kennard (1970)): $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top \mathbf{X} + \frac{n\lambda}{2} \mathbf{I}\right)^{-1} \mathbf{X}^\top \mathbf{y}$
  - micșorează toți coeficienții, niciunul exact zero
  - stabilizează estimarea cînd predictorii sînt corelați
- LASSO (*least absolute shrinkage and selection operator*), $\alpha = 1$ (Tibshirani (1996))
  - penalizarea $|\beta_j|$ are un colț în zero, deci unii coeficienți devin exact zero
  - LASSO selectează caracteristici
- Elastic net, $0 < \alpha < 1$ (Zou & Hastie (2005)): păstrează împreună grupurile de predictori corelați
  - LASSO le reduce arbitrar la unul singur
- Reperul liniar al capitolului este logit-ul penalizat
  - $p = P(y = 1 \mid \mathbf{x}) = 1/\left(1 + e^{-\mathbf{x}^\top \boldsymbol{\beta}}\right)$
  - log-loss-ul în locul pierderii pătratice

Pierdere, penalizare și cinci familii de modele (4/5)

- Arbori: la fiecare nod v alegem caracteristica j și pragul s care reduc cel mai mult impuritatea Gini

$$(j^*, s^*) = \arg\min_{j,\,s} \left[\frac{n_L}{n_v} G(L) + \frac{n_R}{n_v} G(R) \right], \qquad L = \{x_j \le s\},\ R = \{x_j > s\}$$

  - impuritatea Gini: $G = 1 - \sum_c \hat{\pi}_c^2$
  - $\hat{\pi}_c$: proporția clasei c în nod; n_v, n_L, n_R: numărul de observații din nod și din cei doi fii
  - $G = 0$ pentru un nod pur; pentru două clase $G = 2\hat{\pi}(1 - \hat{\pi})$, maxim 0,5 la 50/50; o frunză prezice proporția clasei 1

- Random forest (RF, Breiman (2001)): $\hat{f}_{RF}(x) = \frac{1}{B} \sum_{b=1}^{B} T_b(x)$

$$\mathrm{Var}\big(\hat{f}_{RF}(x)\big) = \rho\,\sigma_T^2 + \frac{1-\rho}{B}\,\sigma_T^2$$

  - fiecare arbore T_b crescut pe un eșantion bootstrap (extras cu revenire din datele de antrenare), cu o submulțime aleatoare de caracteristici la fiecare împărțire
  - σ_T^2: varianța unui arbore; ρ: corelația dintre doi arbori (Hastie, Tibshirani & Friedman (2009), cap. 15)
  - submulțimile aleatoare de caracteristici reduc ρ, singurul termen pe care un număr mai mare de arbori nu îl elimină
  - eroarea out-of-bag (pe observațiile rămase în afara eșantionului bootstrap) este **optimistă** cînd etichetele se suprapun

Random Forests

Pierdere, penalizare și cinci familii de modele (5/5)

- Gradient boosting machine (GBM, Friedman (2001)): adăugăm, pas cu pas, cîte un arbore mic

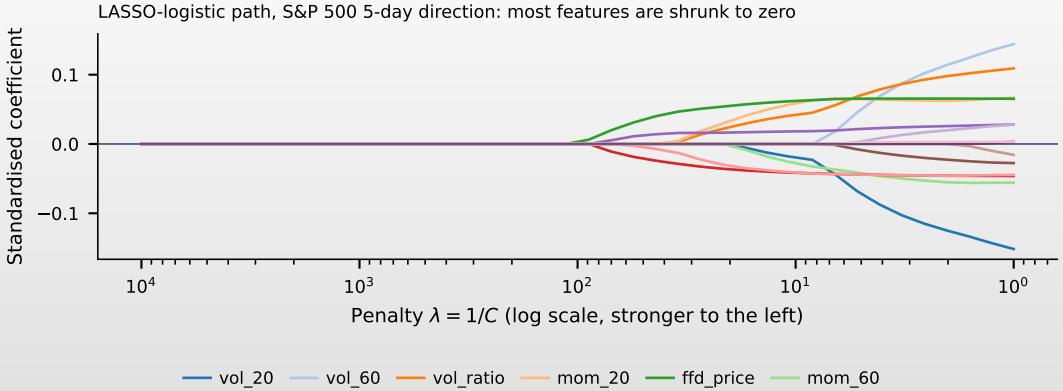
$$r_{im} = -\frac{\partial L(y_i, F)}{\partial F}\bigg|_{F=F_{m-\mathbf{1}}(\mathbf{x}_i)}, \qquad F_m(\mathrm{x}) = F_{m-\mathbf{1}}(\mathrm{x}) + \eta\, h_m(\mathrm{x})$$

  - fiecare arbore nou corectează erorile modelului curent
  - F_m: modelul după m pași; h_m: arborele ajustat pe pseudo-reziduurile r_{im}; pentru log-loss pe scorul logit, $r_{im} = y_i - \hat{p}_{m-\mathbf{1}}(\mathrm{x}_i)$
  - $\eta \in (0, 1]$: rata de învățare; un η mic învață lent și produce mai puțină supraajustare

- Perceptronul multistrat (MLP): straturi ascunse $\mathrm{h}^{(\ell)} = g\big(\mathrm{W}^{(\ell)}\mathrm{h}^{(\ell-\mathbf{1})} + \mathrm{b}^{(\ell)}\big)$, $\mathrm{h}^{(\mathbf{0})} = \mathrm{x}$

  - ieșirea: $\hat{p} = 1/\big(1 + e^{-(\mathbf{w}^\top \mathbf{h}^{(D)} + b)}\big)$
  - $\mathrm{W}^{(\ell)}$, $\mathrm{b}^{(\ell)}$: ponderile și termenii liberi ai stratului ℓ; D: numărul de straturi ascunse; $g(z) = \max(0, z)$: funcția ReLU (rectified linear unit), element cu element
  - ponderile minimizează log-loss-ul prin gradient descendent
  - oprirea timpurie: validare cronologică (ultimele 10% din antrenare, purjate); antrenarea se oprește cînd log-loss-ul pe validare nu mai scade
  - arhitecturile mai complexe sînt în Capitolul 14

Traiectoria de regularizare LASSO (1/2)



MFM_ch13_regularization ▶ Interactiv

Traiectoria de regularizare LASSO (2/2)

- Regresie logistică penalizată L_1 pentru direcția pe 5 zile a S&P 500, 14 caracteristici standardizate
- Etichetele graficului
  - vol_20, vol_60, vol_ratio: volatilitățile pe 20 și 60 de zile și raportul lor
  - mom_20, mom_60: randamentele logaritmice pe ultimele 20 și 60 de zile
  - ffd_price: log-prețul FFD
- Axa orizontală: penalizarea $\lambda = 1/C$, pe scară logaritmică, mai puternică spre stînga
  - axa verticală: coeficientul standardizat
- Penalizare puternică (stînga graficului): toți coeficienții sînt zero
  - prima caracteristică care intră în model este log-prețul FFD (diferențiere fracționară cu fereastră fixă)
  - acesta este un nivel de preț scalat (secțiunea următoare)
- Penalizare slabă: caracteristicile de volatilitate au cei mai mari coeficienți
- Toți coeficienții sînt mici în valoare absolută
  - semnal slab, în acord cu piețe aproape eficiente

MFM_ch13_regularization

Staționaritate, memorie și diferențierea fracționară (1/2)

- Întrebarea secțiunii: cum facem dintr-un preț o caracteristică staționară, fără să-i ștergem memoria?
- Log-prețurile $X_t = \log P_t$ se comportă ca $I(1)$
  - P_t: prețul de închidere din ziua t
  - randamentele $\Delta \log P_t = \log P_t - \log P_{t-1}$ sînt $I(0)$, dar **elimină memoria** nivelului prețului
- Ordinul de integrare: o serie X_t este $I(\delta)$ dacă $(1-B)^\delta X_t$ este $I(0)$
  - B: operatorul lag, $BX_t = X_{t-1}$
  - $I(0)$: serie staționară, cu memorie scurtă; $I(1)$: serie de tip mers aleator (*random walk*), staționară după o diferențiere
  - δ poate fi fracționar: seria este staționară doar pentru $\delta < \frac{1}{2}$, cu memorie lungă cînd $0 < \delta < \frac{1}{2}$ (Hosking (1981))
- Ideea diferențierii fracționare: diferențiem cu mai puțin de 1
  - eliminăm rădăcina unitară și păstrăm o parte din memorie

Staționaritate, memorie și diferențierea fracționară (2/2)

Definiție 1 (Diferența fracționară de ordin d)

- $\tilde{X}_t = (1-B)^d X_t = \sum_{k=0}^{\infty} w_k X_{t-k}$, cu $w_0 = 1$ și $w_k = -w_{k-1} \frac{d-k+1}{k}$

- Notațiile
  - d: ordinul de diferențiere ales de analist; k: lagul, în zile
  - w_k: ponderea valorii de acum k zile, din dezvoltarea binomială a lui $(1-B)^d$
- $d = 1$: $w = (1, -1, 0, \dots)$, randamentele obișnuite
  - $0 < d < 1$: ponderile scad hiperbolic, deci memoria se păstrează
- Un log-preț $I(1)$ diferențiat cu d devine $I(1-d)$
  - nestaționar pentru $d \leq \frac{1}{2}$, staționar cu memorie lungă pentru $\frac{1}{2} < d < 1$
- Ponderile infinite însumează $(1-1)^d = 0$: nivelul se anulează doar la limită
  - comportamentul asimptotic, în Anexă
- López de Prado (2018) propune cel mai mic d la care un test ADF (Augmented Dickey–Fuller) de rădăcină unitară respinge
  - ecuația testului și valoarea critică pe slide-ul ADF
  - această secțiune arată de ce regula eșuează

Fereastra fixă (FFD)

Definiție 2 (Diferențiere fracționară cu fereastră fixă (FFD))

- Trunchiem ponderile (López de Prado (2018)): păstrăm $w_0, \ldots, w_{l^*}$ cu $|w_k| \geq \tau$, renunțăm la rest
  - de exemplu $\tau = 10^{-4}$
- $\tilde{X}_t = \sum_{k=0}^{l^*} w_k X_{t-k}$ pentru orice $t > l^*$

- τ: cea mai mică pondere păstrată; l^*: ultimul lag păstrat, adică lungimea ferestrei
  - primele l^* zile nu au suficient istoric și sînt eliminate
- Fiecare $\tilde{X}_t$ folosește același număr de laguri și doar valori trecute
  - deci nu conține informație din viitor
- Cu cît d este mai mic, cu atît trebuie mai multe laguri
  - pentru $d = 0{,}25$ și $\tau = 10^{-4}$ fereastra are $l^* = 444$ de laguri
- Ponderile trunchiate nu mai au suma zero
  - suma lor $\sum_{k=0}^{l^*} w_k$ este 0,178 la $d = 0{,}25$ și 0,040 la $d = 0{,}5$

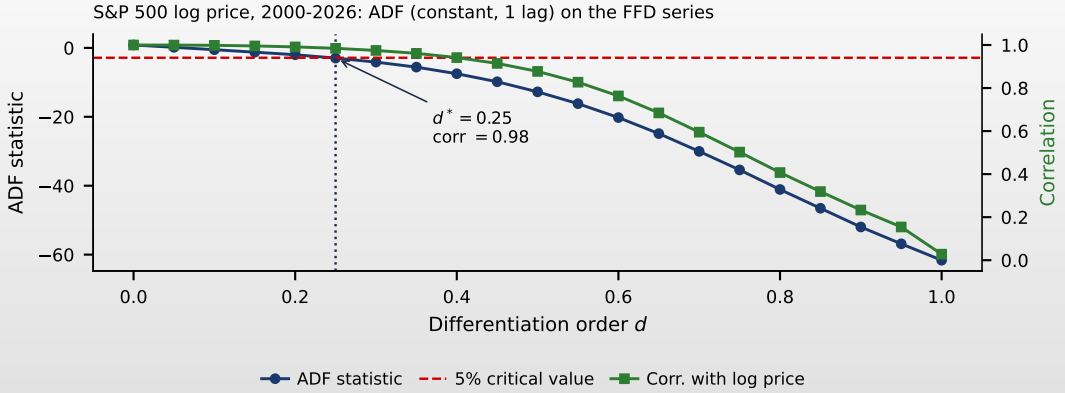
Statistica ADF în funcție de d: S&P 500 (1/3)

- Testul de rădăcină unitară Dickey–Fuller augmentat (ADF): statistica t a lui γ în

$$\Delta \tilde{X}_t = c + \gamma \, \tilde{X}_{t-1} + \sum_{i=1}^{p} \phi_i \, \Delta \tilde{X}_{t-i} + e_t$$

  - H_0: $\gamma = 0$, adică rădăcină unitară; p: numărul de laguri ale diferenței, care preiau autocorelația lui e_t
  - cu constantă, respingem H_0 la 5% sub $\approx -2{,}86$
- Regula lui López de Prado (2018): calculăm seria FFD pe o grilă de valori ale lui d între 0 și 1
  - alegem cel mai mic d la care ADF respinge
- Graficul următor: d pe axa orizontală
  - statistica ADF (albastru) pe axa din stînga
  - corelația seriei FFD cu log-prețul (verde) pe axa din dreapta
  - linia roșie punctată: valoarea critică de 5%; linia verticală: $d^* = 0{,}25$

Statistica ADF în funcție de d: S&P 500 (2/3)



MFM_ch13_ffd ▶ Interactiv

Statistica ADF în funcție de d: S&P 500 (3/3)

- Log-prețul FFD, S&P 500 2000–2026, constantă și 1 lag: ADF $= -2{,}96$
  - deci regula respinge de la $d = 0{,}25$
- La $d = 0{,}25$, $\mathrm{Corr}(\tilde{X}_t, \log P_t) = 0{,}98$
  - seria FFD este aproape log-prețul
- Aceeași serie, cu 34 de laguri: ADF $= -0{,}72$, nu respinge
  - lagurile sînt alese prin criteriul informațional Akaike (AIC)
- Verdictul depinde de numărul de laguri
- O corelație de 0,98 cu un preț $I(1)$ este un semnal de alarmă, nu o calitate
- Tipul de dovadă: comparație între proceduri (arată care metodă funcționează mai bine pe aceleași date, nu de ce)
  - același log-preț, două alegeri ale numărului de laguri, două verdicte

MFM_ch13_ffd

FFD trunchiată a unui preț $I(1)$ rămîne $I(1)$

- Descompunem seria FFD:

$$\tilde{X}_t = \sum_{k=0}^{l^*} w_k X_{t-k} = \Big(\sum_{k=0}^{l^*} w_k\Big) X_t - \sum_{k=1}^{l^*} w_k \left(X_t - X_{t-k}\right)$$

  - primul termen: nivelul de azi înmulțit cu suma ponderilor trunchiate
  - al doilea termen: o sumă ponderată a variațiilor pe k zile, $X_t - X_{t-k}$
- Dacă X_t este $I(1)$, fiecare variație $X_t - X_{t-k}$ este $I(0)$
- Primul termen este însă $0{,}178\, X_t$ la $d = 0{,}25$, adică un **mers aleator scalat**
- Componenta cu rădăcină unitară are varianță mică față de partea staționară
  - deci testul ADF cu puține laguri nu le poate distinge
  - Monte Carlo: mers aleator pur, $T = 6\,274$ de zile (lungimea seriei FFD a S&P 500), FFD cu $d = 0{,}25$
  - ADF (1 lag) respinge în **97,8%** din 500 de eșantioane, la un nivel nominal de 5%
- Un filtru finit cu suma ponderilor nenulă păstrează componenta de nivel
  - oricare ar fi valoarea p (*p-value*) a testului ADF
- Consecință: log-prețul FFD din studiul de caz este un nivel de preț scalat
  - nu o caracteristică staționară; memoria se măsoară direct în Aprofundări

MFM_ch13_inference

De la etichete cu orizont fix la etichete cu trei bariere (1/2)

- Întrebarea secțiunii: ce etichetă reproduce decizia reală a unui trader?
  - decizia are un prag de profit (*take-profit*), un prag de pierdere (*stop-loss*) și o durată maximă de deținere
- **Orizont fix**: eticheta este semnul randamentului logaritmic pe h zile

$$r_{t,t+h} = \log P_{t+h} - \log P_t, \qquad y_t = \begin{cases} +1 & \text{dacă } r_{t,t+h} > \kappa \\ -1 & \text{dacă } r_{t,t+h} < -\kappa \\ 0 & \text{altfel} \end{cases}$$

  - cu o zonă moartă în jurul lui zero
  - P_t: prețul de închidere din ziua t; h: orizontul, în zile; $\kappa \geq 0$: pragul, sub care variațiile primesc eticheta 0
- Trei probleme
  - un κ constant ignoră variația volatilității în timp
  - eticheta nu ține cont de traiectorie: un stop-loss ar fi putut fi atins înainte de $t + h$
  - clasele sînt dezechilibrate: pentru S&P 500, cu $h = 10$ și $\kappa = 0$, 60,5% dintre etichete sînt „sus" și 39,5% „jos"

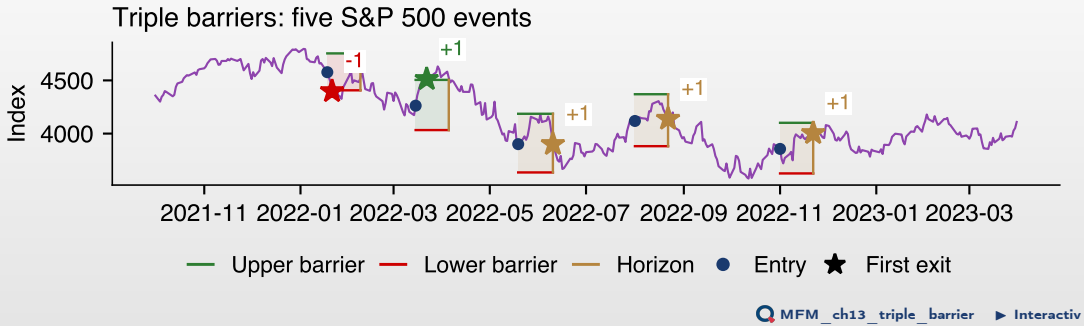
De la etichete cu orizont fix la etichete cu trei bariere (2/2)

Definiție 3 (Etichetă cu trei bariere, López de Prado (2018))

- La momentul evenimentului t_0, cu volatilitatea zilnică σ_{t_0} (EWMA: medie mobilă ponderată exponențial) și orizontul h
  - bariera superioară $\log P_t - \log P_{t_0} \geq \pi_u \sigma_{t_0} \sqrt{h}$ (profit)
  - bariera inferioară $\leq -\pi_l \sigma_{t_0} \sqrt{h}$ (stop-loss), bariera verticală $t = t_0 + h$
- Eticheta este dată de **prima barieră atinsă**, la momentul t_1
  - $+1$ pentru bariera superioară, -1 pentru cea inferioară
  - $\text{sign}(r_{t_0,t_1})$ (sau 0) pentru bariera verticală

- $\sigma_{t_0} \sqrt{h}$: variația tipică pe h zile
  - $\pi_u, \pi_l > 0$: lățimile barierelor, în unități de $\sigma_{t_0} \sqrt{h}$
- Fiecare etichetă are un **interval** $[t_0, t_1]$
  - datele de intrare pentru purjarea din validare

Trei bariere: S&P 500 în 2022 (1/2)

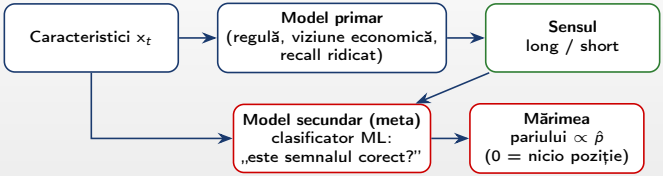


Trei bariere: S&P 500 în 2022 (2/2)

- Cinci evenimente în piața descendentă din 2022, $\pi_u = \pi_l = 1$, $h = 15$ zile
  - culoarea fiecărei stele este bariera atinsă prima
- Graficul: nivelul indicelui
  - linii verzi: bariera superioară, roșii: bariera inferioară, maro: orizontul
  - punctele: intrarea, stelele: prima barieră atinsă
- Evenimentul din ianuarie atinge stop-loss-ul (eticheta -1)
- Evenimentul din martie atinge bariera de profit (eticheta $+1$)
- Celelalte trei se termină la bariera verticală (stele galben-portocalii) peste nivelul de intrare
  - deci regula semnului le dă eticheta $+1$
  - o etichetă $+1$ nu înseamnă că bariera de profit a fost atinsă

MFM_ch13_triple_barrier

Meta-etichetarea: decizia de a acționa (1/2)



- Meta-eticheta $y_t^{meta} = 1$ dacă semnalul primar a fost profitabil, 0 altfel
  - profitabil: barieră atinsă în direcția lui
- Modelul secundar estimează $\hat{p} = P(y_t^{meta} = 1 \mid x_t)$
  - probabilitatea ca semnalul din ziua t să fie corect, date fiind caracteristicile x_t
- ML învață **cînd semnalul** unei reguli primare interpretabile este de încredere
  - nu direcția însăși

Meta-etichetarea: decizia de a acționa (2/2)

- Un filtru decide pe ce semnale acționăm; trei scoruri îl descriu

$$\text{precizia} = \frac{TP}{TP + FP}, \qquad \text{recall} = \frac{TP}{TP + FN}, \qquad F_1 = \frac{2 \cdot \text{precizie} \cdot \text{recall}}{\text{precizie} + \text{recall}}$$

  - TP: semnale profitabile urmate; FP: semnale neprofitabile urmate; FN (*False Negative*, fals negativ): semnale profitabile ignorate
  - dacă acționăm la orice semnal, recall-ul este 1; F_1: media lor armonică, între 0 și 1
- Exemplu (Seminarul 13, exercițiul B11): încrucișarea mediilor mobile (MA, *Moving Average*) pe 20 și 50 de zile pe Bitcoin
  - regula MA20 (*20-day Moving Average*, medie mobilă pe 20 de zile)/MA50, 1 660 de tranzacții de test
  - cu filtrul meta, precizia crește de la 0,475 la 0,517
  - recall-ul scade de la 1 la 0,541, iar F_1 scade de la 0,644 la 0,529
  - randamentul logaritmic brut mediu pe tranzacție, înainte de costuri, trece de la $-0{,}27\%$ la $+0{,}19\%$
- Meta-etichetarea crește precizia în detrimentul recall-ului
  - cînd tranzacțiile au costuri, contează precizia și randamentul pe tranzacție, nu F_1

Scurgerea de informație între antrenare și test

- Întrebarea secțiunii: de ce K-fold obișnuit supraestimează abilitatea unui model pe serii financiare?

Definiție 4 (Scurgerea de informație (*leakage*))

- Setul de antrenare conține informații prezente și în setul de test
  - scorul de test nu mai măsoară rezultatele pe date cu adevărat noi

- Caracteristici corelate serial ($x_t \approx x_{t+1}$) și etichete suprapuse ($y_t \approx y_{t+1}$)
  - un K-fold amestecat pune zile **aproape identice** de ambele părți
- Un model flexibil (RF, GBM) memorează observația vecină și obține un scor excelent în afara eșantionului
- Reglarea hiperparametrilor pe un CV cu scurgere de informație selectează modelul cu **cea mai mare supraajustare**
- Alte surse de scurgere de informație
  - transformări estimate pe tot eșantionul (scalare, analiza componentelor principale, alegerea lui d pentru FFD)
  - deplasarea de supraviețuire (*survivorship bias*), revizuiri ale datelor, marcaje de timp greșite

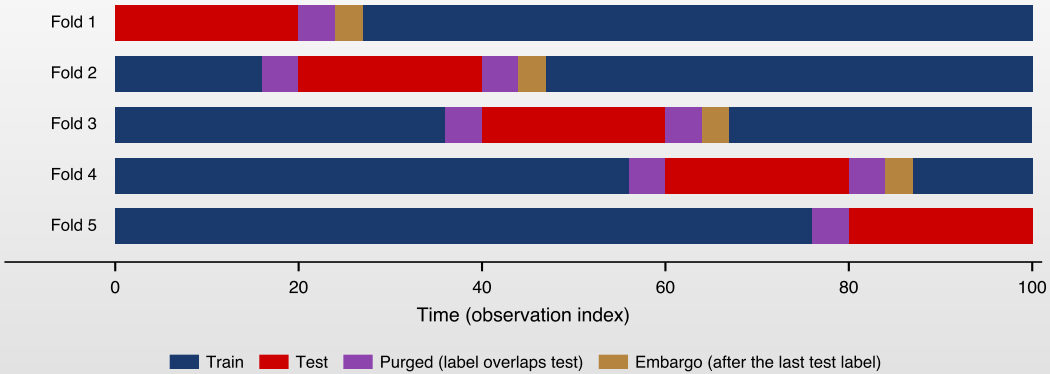
K-fold cu purjare și embargo (1/2)

- Pliurile sînt blocuri **contigue** în timp (López de Prado (2018))
  - pliul de test k conține observațiile $t_k^a, \ldots, t_k^b$
  - etichetele lor se termină cel tîrziu la $T_k = \max_{i \in \text{test}_k} t_{1,i}$
- Setul de antrenare păstrează doar observațiile care nu ating testul:

$$\mathcal{A}_k = \left\{ i \notin \text{test}_k : \ [t_{0,i}, t_{1,i}] \cap [t_k^a, T_k] = \emptyset \ \text{ și } \ t_{0,i} \notin (T_k, \ T_k + e] \right\}$$

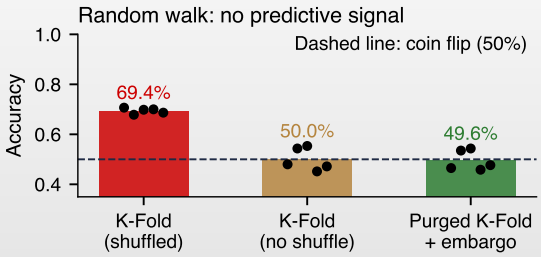
  - prima condiție este **purjarea**: elimină etichetele care se suprapun cu intervalul etichetelor de test
  - a doua este **embargoul**: e observații după sfîrșitul ultimei etichete de test, T_k (de exemplu 1%), necesar deoarece caracteristicile folosesc observații anterioare
- Scorul CV cu purjare este media celor K scoruri ale modelelor estimate pe $\mathcal{A}_k$ și evaluate pe pliul k
- Cu etichete de lungime h, o frontieră elimină h observații înainte de blocul de test și $h + e$ după el
  - un pliu interior pierde $2h + e$ observații
  - exemplu cu 10 observații: ▸ Anexă
- Graficul următor: cinci pliuri pe rînduri, indicele observației pe axa orizontală
  - albastru antrenare, roșu test, violet purjat, maro embargo

K-fold cu purjare și embargo (2/2)



MFM_ch13_purged_cv ▶ Interactiv

Scurgerea de informație pe o țintă fără predictibilitate



- Date simulate: prețul este un mers aleator pur
  - eticheta: semnul randamentului pe următoarele 20 de zile
  - 5 caracteristici: medii mobile pe 20 de zile ale unui zgomot independent
- Prin construcție, acuratețea reală este 50%
  - modelul: un random forest validat în 5 pliuri
- Acuratețea măsurată
  - **K-fold amestecat 69,4%**
  - K-fold contiguu 50,0%
  - K-fold cu purjare și embargo 49,6%
- Amestecarea eșuează pentru că etichetele se suprapun *și* caracteristicile sînt persistente
- Pliurile contigue elimină deja majoritatea scurgerii
  - doar frontierele dintre pliuri au randamente comune, iar purjarea le curăță
- Un rezultat neverosimil de bun indică de obicei o scurgere de informație
- Tipul de dovadă: comparație între proceduri

MFM_ch13_purged_cv ▶ Interactiv

Selecția modelului și evaluarea lui finală

- K-fold obișnuit este valid pentru un model pur autoregresiv cu erori necorelate serial (Bergmeir, Hyndman & Koo (2018))
  - cazurile în care eșuează: Anexă
- **Selecția modelului**: K-fold cu purjare sau CV combinatorică cu purjare (CPCV)
  - CPCV testează pe multe combinații de grupuri și dă o **distribuție** de rapoarte Sharpe în locul unui singur număr
- **Utilizarea**: evaluarea pe ferestre succesive (WF)
  - antrenare pe $[0, T_k]$, test pe $(T_k, T_{k+1}]$, apoi avansăm
  - T_k: sfîrșitul celei de-a k-a ferestre de antrenare
  - realist, dar o singură traiectorie, cu variantă mare
  - servește evaluării finale, nu alegerii (WF și CPCV: Anexă)

Studiu de caz: poate fi prezisă direcția S&P 500 pe 5 zile?

- Indicele S&P 500 și VIX, închideri zilnice din datele cursului, 2001–2026: $n = 6\,269$ de observații
  - eșantionul începe în octombrie 2001: primele 444 de zile din 2000–2001 alimentează fereastra FFD
  - VIX: indicele de volatilitate CBOE (*Chicago Board Options Exchange*, bursa de opțiuni din Chicago), volatilitatea implicită pe 30 de zile a opțiunilor pe S&P 500
- De ce S&P 500: cel mai lichid indice de acțiuni, cu o istorie VIX lungă
  - dacă direcția lui ar fi predictibilă, arbitrajul ar fi ieftin
- Țința: $y_t = 1\{\log P_{t+5} - \log P_t > 0\}$, egală cu 1 dacă indicele este mai sus peste 5 zile
  - eticheta cu orizont fix, cu $h = 5$ și $\kappa = 0$
- De ce 5 zile: un orizont săptămînal de tranzacționare
  - circa $6\,269/5 \approx 1\,250$ de etichete fără suprapunere
- Intervalul etichetei, $[t, t + 5]$, este folosit la purjare
- Validare: CV cu purjare în 5 pliuri și embargo de 1%, cu aceleași pliuri pentru toate modelele
  - fiecare pliu acoperă circa 5 ani, deci cel puțin un episod de criză

Studiu de caz: metricile, față de reperul „mereu în sus”

- Acuratețea: proporția zilelor a căror direcție este prezisă corect
- Aria de sub curba ROC (*receiver operating characteristic*), AUC
  - probabilitatea ca o zi de creștere aleasă la întîmplare să primească o probabilitate prezisă mai mare
  - decît o zi de scădere aleasă la întîmplare
  - AUC $= 0{,}5$ înseamnă nicio abilitate, iar AUC $= 1$ înseamnă ordonare perfectă
- Exemplu: zile „sus” cu probabilitățile 0,7 și 0,4, zile „jos” cu 0,5 și 0,2
  - 3 din cele 4 perechi sînt ordonate corect, deci AUC $= 0{,}75$
- De ce ambele: acuratețea depinde de pragul 0,5
  - AUC compară doar ordinea probabilităților
- **Media pe pliuri**: metrica calculată în fiecare dintre cele 5 pliuri de test, apoi mediată
- **Agregat**: metrica calculată o singură dată pe toate predicțiile din afara pliurilor de antrenare
  - pe aceasta se bazează testele formale

Studiu de caz: 14 caracteristici și patru modele

Caracteristici, informație pînă în ziua t

- **Momentum**: $\log P_t - \log P_{t-h}$, $h \in \{1, 5, 20, 60, 120\}$
  - randamentul logaritmic pe ultimele h zile
- **Volatilitate realizată** pe 20 și 60 de zile și raportul lor
  - abaterea standard anualizată a randamentelor logaritmice zilnice
- **Distanța față de media mobilă**
  - $\log P_t$ minus media lui $\log P$ pe 50 sau 200 de zile
- **RSI(14)**, indicele puterii relative: $100 - 100/(1 + RS)$
  - RS: cîștigul mediu netezit pe 14 zile raportat la pierderea medie
  - ia valori între 0 și 100 și este ridicat după creșteri
- **VIX**: nivelul logaritmic și variația pe 5 zile
- Log-prețul **FFD**, $d = 0{,}25$, ales pe tot eșantionul
  - testul fără această caracteristică, după compararea modelelor
  - în WF: $d = 0{,}15$, ales doar pe datele pînă în 2009

Modele, configurate deliberat conservator

- **Logit**: penalizat L2 (Ridge), caracteristici standardizate
- **RF**: 400 de arbori, frunza minimă 100
- **GBM**: pe histograme, adîncime 3, rată de învățare 0,03, 200 de arbori
- **MLP**: straturi ascunse de 32 și 16 unități, oprire timpurie pe o validare cronologică purjată
- Frunzele mari și arborii puțin adînci limitează varianța

Un clasificator care prognozează doar clasa majoritară (1/2)

	Prezis sus	**Prezis jos**
Real sus	TP $= 3\,575$	FN $= 65$
Real jos	FP $= 2\,589$	TN $= 40$

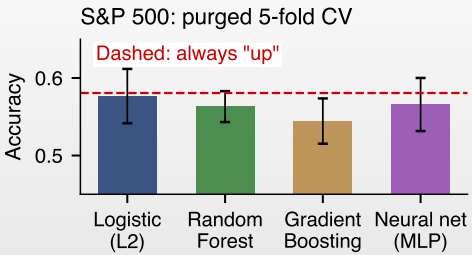
- Logit penalizat L2, $n = 6\,269$
  - predicții agregate în afara pliurilor din CV cu purjare în 5 pliuri
- TP, FN, FP, TN: rezultate adevărat pozitive, fals negative, fals pozitive, adevărat negative

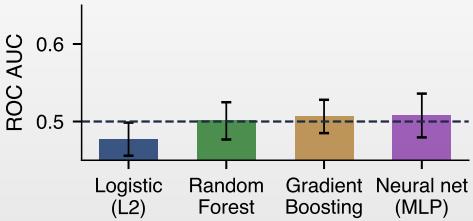
- Acuratețe $= (TP + TN)/n = 57{,}7\%$
- Precizie $= TP/(TP + FP) = 0{,}580$
- Recall $= TP/(TP + FN) = 0{,}982$
- $F_1 = 0{,}729$ (formula pe slide-ul despre meta-etichetare)
- Modelul prezice „sus” în 98,3% din zile
  - în esență, a învățat regula „mereu în sus”

Un clasificator care prognozează doar clasa majoritară (2/2)

- **Reperul clasei majoritare**, din același tabel
  - zilele reale „sus": $(3\,575 + 65)/6\,269 = 3\,640/6\,269 = 58{,}1\%$
  - acuratețea regulii care prezice „sus" în fiecare zi, fără niciun model
- Un model cu acuratețe de 57,7% este *mai slab* decît reperul trivial
- Cauza este dezechilibrul claselor
  - cu caracteristici slabe, probabilitățile prezise rămîn aproape de frecvența de bază 0,58, peste pragul 0,5
  - deci aproape fiecare zi este clasificată „sus"
  - studiul de caz păstrează intenționat eticheta cu orizont fix, pentru ca reperul să fie transparent
  - etichetele cu trei bariere schimbă echilibrul claselor văzut de clasificator (Seminarul 13, exercițiile B1 și B8)
- Alte repere: direcția de ieri, o regulă de medii mobile, buy-and-hold pentru tranzacționare
- Un model are valoare doar dacă depășește reperul **în afara eșantionului**
  - după costuri, cu mai mult decît eroarea de eșantionare

Compararea modelelor: mediile pe pliuri





Media pe pliuri	Logit	RF	GBM	MLP
Acuratețe	57,7%	56,3%	54,4%	56,6%
AUC	0,477	0,501	0,507	0,508

- Barele: media pe cele 5 pliuri purjate
- Mustățile: ± 1 abatere standard între pliuri
- Reperul „mereu în sus” are acuratețea **58,1%**
- Niciun model nu îl depășește

MFM_ch13_model_comparison ▶ Interactiv

▸ Aprofundări: testul fără FFD, Newey–West

Semnificația diferențelor: Pesaran–Timmermann și intervale pentru AUC (1/2)

- **Pesaran–Timmermann** (PT, Pesaran & Timmermann (1992)): sînt legate direcțiile prezise de cele realizate?

$$PT = \frac{\hat{P} - \hat{P}^*}{\widehat{\mathrm{SE}}} \xrightarrow{d} N(0,1), \qquad \hat{P}^* = \hat{p}_y \hat{p}_x + (1-\hat{p}_y)(1-\hat{p}_x)$$

  - $\hat{P}$: proporția zilelor cu semnul corect; $\hat{p}_y$, $\hat{p}_x$: proporțiile zilelor „sus" realizate și prezise; $\hat{P}^*$: proporția corectă așteptată dacă predicțiile ar fi independente de rezultate
  - H_0: predicțiile sînt independente de rezultate; $\widehat{\mathrm{SE}}$: eroarea standard a lui $\hat{P} - \hat{P}^*$ sub H_0 (Anexă)
  - $PT > 1{,}645$ indică abilitate la 5% (test unilateral); $PT < 0$ înseamnă mai puține potriviri decît la întîmplare
- PT presupune zile independente
  - cu etichete suprapuse folosim și statistica t a pantei în $y_t = a + b\,x_t + u_t$
  - erori Newey–West, $q = 4$ (slide-ul Newey–West)
  - x_t este direcția prezisă, y_t cea realizată, iar $b > 0$ înseamnă abilitate
- **AUC**: eroarea standard (SE) DeLong presupune zile independente (Anexă)
  - de aceea folosim un bootstrap circular pe blocuri de $L = 20$ de zile consecutive (Künsch (1989))
  - blocul are de patru ori lungimea etichetei și păstrează dependența serială

Semnificația diferențelor: Pesaran–Timmermann și intervale pentru AUC (2/2)

Agregat, predicții din afara pliurilor	Logit	RF	GBM	MLP
$\hat{P}$ / $\hat{P}^*$	0,577 / 0,578	0,563 / 0,570	0,544 / 0,554	0,566 / 0,569
PT / t HAC (lag 4)	−0,80 / −0,62	−2,27 / −1,67	−1,97 / −1,49	−0,83 / −0,60
AUC	0,480	0,495	0,501	0,497
CI 95% bootstrap pe blocuri	[0,456; 0,505]	[0,470; 0,520]	[0,478; 0,525]	[0,474; 0,520]

- Niciun model nu are o abilitate direcțională pozitivă semnificativă
  - fiecare interval bootstrap pentru AUC conține 0,5
- Cu erori standard HAC (*Heteroskedasticity and Autocorrelation Consistent*, robuste la heteroscedasticitate și autocorelație) cu 4 laguri, niciun model nu diferă semnificativ de întîmplare ($|t| \leq 1{,}67$)
  - PT, care tratează zilele ca independente, dă −2,27 pentru RF: etichetele suprapuse exagerează semnificația
- Pentru logit, intervalul DeLong de 95% este [0,466; 0,494] și nu conține 0,5
  - dar acest interval tratează zilele ca independente; intervalul bootstrap pe blocuri conține 0,5
- Rezultatul este în acord cu eficiența în formă slabă (Capitolul 2)
- Tipul de dovadă: comparație între proceduri
- Unde trebuie căutat semnalul
  - în secțiuni transversale (Gu, Kelly & Xiu (2020))
  - în ținte de risc precum volatilitatea (Capitolele 5 și 9)
  - în delistarea criptoactivelor (Będowska-Sójka, Wójcik & Pele (2026)) și în text (Capitolul 15)

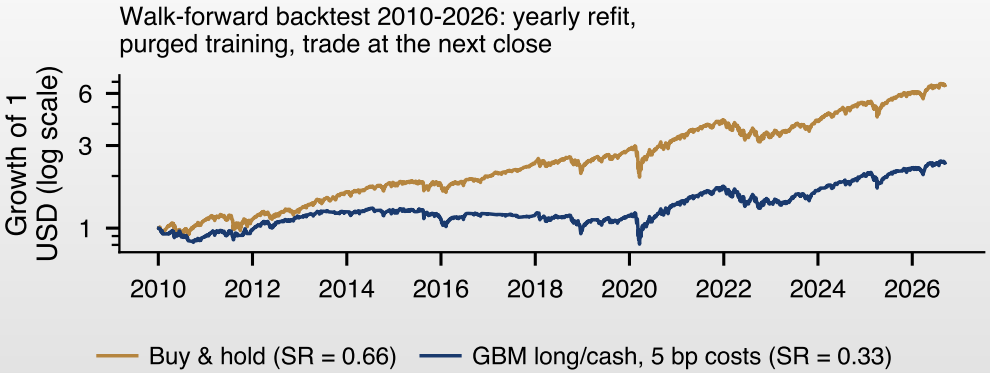
Backtest pe ferestre succesive: strategie și construcție

- Întrebarea: ar fi cîștigat un trader care urma modelul mai mult decît prin simpla deținere a indicelui?
  - comparația se face după costuri
- Poziția $w_t \in \{0, 1\}$: long pe S&P 500 dacă $\hat{p}_t > 0{,}5$, altfel numerar
  - decisă la închiderea zilei t, cu informații pînă la t; $\hat{p}_t$: probabilitatea de „sus" dată de model
  - tranzacția se execută la închiderea zilei $t + 1$
- Randamentul strategiei: randamentul indicelui cînd sîntem investiți, minus un cost la fiecare schimbare de poziție

$$r^{s}_{t+\mathbf{2}} = w_t \, r_{t+\mathbf{2}} - c \, |w_t - w_{t-\mathbf{1}}|$$

  - $r_{t+\mathbf{2}}$: randamentul logaritmic al indicelui între închiderile $t + 1$ și $t + 2$
  - turnover-ul $|w_t - w_{t-\mathbf{1}}|$ este 1 în zilele în care poziția se schimbă
  - $c = 5$ puncte de bază (bp), adică 0,05% pe unitatea tranzacționată
  - un cost prudent pentru un fond tranzacționat la bursă pe S&P 500
  - $r - c$ aproximează $\log\left((1 - c)\, e^{r}\right) = r + \log(1 - c)$, randamentul logaritmic după plata costului
- Expunerea $\bar{w} = \frac{1}{T} \sum_t w_t$: proporția celor T zile în care sîntem investiți
- Numerarul are randament 0
  - ambele rapoarte Sharpe folosesc randamente logaritmice zilnice cu rata fără risc 0, anualizate cu $\sqrt{252}$
  - deci dobînda din zilele în numerar este ignorată
- Model: GBM reestimat în fiecare ianuarie, 2010–2026, pe toate datele trecute
  - cu purjarea etichetelor care se suprapun cu anul de test; log-prețul FFD cu $d = 0{,}15$
- Reperul este buy-and-hold ($w_t = 1$: expunere 100%, fără turnover)

Backtest pe ferestre succesive: rezultate (1/2)



MFM_ch13_walk_forward ▶ Interactiv

Backtest pe ferestre succesive: rezultate (2/2)

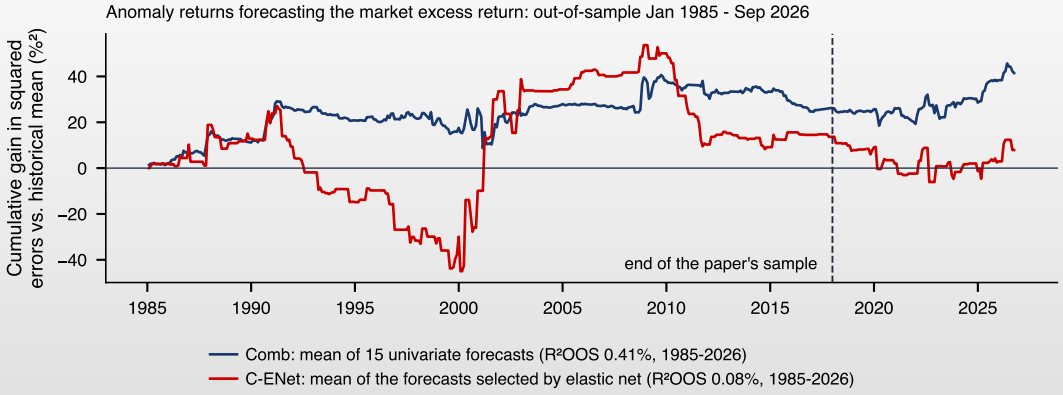
- 2010–2026: raportul Sharpe anualizat al strategiei este **0,33**, față de **0,66** pentru buy-and-hold
  - expunerea medie: 76%
  - modelul a rămas în principal long
  - ieșirile din piață (24% din zile) au redus raportul Sharpe
- Graficul: creșterea unui dolar investit, pe scară logaritmică
  - maro buy-and-hold, albastru strategia GBM
- Diferența dintre cele două rapoarte Sharpe, $-0{,}33$, are $p = 0{,}001$
  - test HAC Ledoit–Wolf, în Aprofundări
  - presupune moment de ordin patru finit, pus sub semnul întrebării în Capitolul 1
- Strategia este semnificativ mai slabă decît buy-and-hold; nu există dovezi de abilitate
- Tipul de dovadă: comparație între proceduri (strategia GBM față de buy-and-hold)
- Sensibilitatea raportului Sharpe al strategiei
  - 0,31 cu $d = 0{,}25$, ales pe tot eșantionul, deci cu informație din perioada de test
  - 0,48 fără log-prețul FFD
  - 0,43 cu tranzacția la aceeași închidere, o idealizare fără întîrziere de execuție

MFM_ch13_walk_forward

Studiu de caz: prezic anomaliile transversale piața?

- Dong, Li, Rapach & Zhou (2022): prezic randamentele long-short ale anomaliilor randamentul în exces al pieței din luna următoare?
  - anomalie: portofoliu long pe acțiunile cu randament mediu ridicat după o caracteristică și short pe cele cu randament mediu scăzut
  - exemple: valoarea (raportul book-to-market), momentum-ul, acumulările contabile (*accruals*), emisiunile nete de acțiuni
- Literatura transversală și predictibilitatea pieței se studiau separat
- Reperul este greu de bătut: Welch & Goyal (2008) arată că majoritatea predictorilor publicați nu depășesc media istorică
- Datele lucrării: 100 de anomalii americane, randamente lunare, ianuarie 1970 – decembrie 2017
  - prima prognoză în afara eșantionului are la dispoziție 180 de luni pentru 100 de predictori
  - regresia cu toți predictorii învață zgomotul
- Metoda: contracția coeficienților (*shrinkage*), adică elastic net, combinarea prognozelor și reducerea dimensiunii
- Afirmația principală: anomaliile au putere predictivă în afara eșantionului pentru piață, semnificativă statistic și economic
- Explicația propusă: limitele asimetrice ale arbitrajului și persistența corectării supraevaluării
  - supraevaluarea se arbitrează mai greu decît subevaluarea, deoarece vînzarea în lipsă costă
  - de aceea, componenta short (*short leg*) a anomaliilor conține cea mai mare parte a informației

Anomaliile și piața: cîștigul OOS față de media istorică



MFM_ch13_anomaly_market ▶ Interactiv

Anomaliile și piața: metoda și ce face replicarea diferit

- Regresia predictivă a anomaliei i, estimată pe o fereastră extinsă și reestimată lunar

$$r_{t+\mathbf{1}} = a_i + b_i\, x_{i,t} + \varepsilon_{i,t+\mathbf{1}}, \qquad \hat{r}^{(i)}_{t+\mathbf{1}} = \hat{a}_i + \hat{b}_i\, x_{i,t}$$

  - $r_{t+\mathbf{1}}$: randamentul în exces al pieței (Mkt-RF) în luna $t+1$
  - $x_{i,t}$: randamentul long-short al anomaliei i în luna t
- Comb (combinația simplă): media celor N prognoze $\hat{r}^{(i)}_{t+\mathbf{1}}$
- C-ENet (*combination elastic net*): elastic net în regresia lui r_t pe cele N prognoze, cu coeficienți nenegativi
  - prognoza este media prognozelor selectate
  - ENet: elastic net pe toți predictorii; OLS: regresia pe toți predictorii
- Perioadele lucrării: antrenare 1970–1979, validare 1980–1984, în afara eșantionului 1985–2017
- Replicarea, pe Kenneth French Data Library
  - 15 anomalii în loc de 100: SMB (*Small Minus Big*, factorul mărime: firme mici minus mari), HML (*High Minus Low*, factorul valoare: B/M mare minus mic), RMW (*Robust Minus Weak*, factorul profitabilitate: firme robuste minus slabe), CMA (Fama & French (2015)), momentum, inversarea (*reversal*) pe termen scurt și lung
  - și diferența dintre decilele extreme pentru acumulări contabile, emisiuni nete, beta, variantă, variantă reziduală, E/P, CF/P (fluxul de numerar, CF, raportat la preț), D/P
  - penalizarea elastic net ($\ell_{\mathbf{1}}$ și $\ell_{\mathbf{2}}$ în proporții egale) aleasă prin AIC corectat pentru eșantioane mici
  - perioada în afara eșantionului prelungită până în septembrie 2026

Anomaliile și piața: rezultatul și limitele

- Măsura: R^2_{OOS} față de media istorică $\bar{r}_t$, calculată pînă în luna $t-1$ (Campbell & Thompson (2008))

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

  - graficul cumulează diferența dintre erorile pătratice ale mediei istorice și ale modelului
  - curba urcă în lunile în care modelul bate media istorică
- Cifrele lucrării, versiunea publicată (Tabelul II), 1985–2017: C-ENet 2,81%, ENet 2,03%, Comb 0,89%, OLS −2 513,86%
- Replicarea, 1985–2017: Comb 0,35%, C-ENet 0,18%, ENet −4,75%, OLS −6,84%
- Replicarea, 1985–2026: Comb 0,41% (p unilateral Clark–West 0,045), C-ENet 0,08% ($p = 0{,}15$)
  - Clark & West (2007): diferența ajustată regresată pe constantă, erori Newey–West cu $\lfloor 4(T/100)^{2/9} \rfloor = 5$ laguri
  - p pentru Comb cu 3, 6 și 12 laguri: 0,063, 0,044 și 0,032; fără corecție HAC: 0,068
  - numărul de laguri este fixat de regulă, nu ales după semnificație
  - testul presupune moment de ordin patru finit, pus sub semnul întrebării în Capitolul 1
- După eșantionul lucrării, 2018–2026: Comb 0,61% ($p = 0{,}12$), C-ENet −0,23%
- Tipul de dovadă: comparație între proceduri (combinarea prognozelor se clasează înaintea regresiei pe toți predictorii)
  - nu susține amplitudinea din lucrare
- Limite: 15 anomalii nu reproduc cele 100
  - cu 15 prognoze, C-ENet nu selectează nicio prognoză în 57% din luni și folosește atunci media istorică

MFM_ch13_anomaly_market

Deplasarea de selecție și teorema strategiei false (1/2)

- Întrebarea secțiunii: cît de mare este cel mai bun raport Sharpe dintre N încercări fără abilitate?
- Un cercetător încearcă N strategii și o raportează pe cea mai bună
  - chiar dacă toate au Sharpe real zero, $\max_n \widehat{SR}_n$ crește cu N
  - acesta este **deplasarea de selecție**
- Harvey, Liu & Zhu (2016): față de sutele de factori publicați, un factor nou are nevoie de $t > 3$, nu de 2

Teoremă 1 (Teorema strategiei false, Bailey & López de Prado (2014))

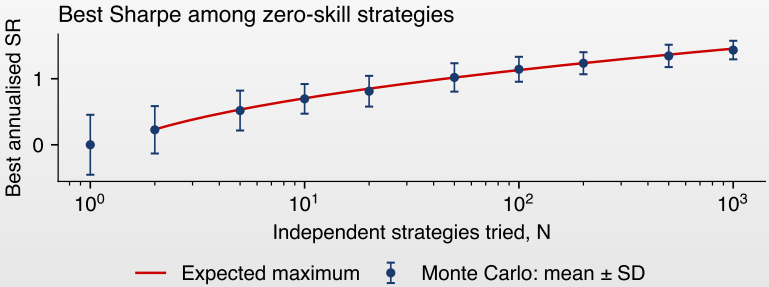
- Dacă $\widehat{SR}_n$, $n = 1, \ldots, N$, sînt independente și aproximativ $N\big(0, \mathrm{Var}[\widehat{SR}]\big)$ (fără abilitate), **raportul Sharpe maxim așteptat** este

$$SR_0 \equiv \mathbb{E}\Big[\max_n \widehat{SR}_n\Big] \approx \sqrt{\mathrm{Var}[\widehat{SR}]}\left[(1-\gamma_E)\,\Phi^{-1}\Big(1-\tfrac{1}{N}\Big) + \gamma_E\,\Phi^{-1}\Big(1-\tfrac{1}{Ne}\Big)\right]$$

Deplasarea de selecție și teorema strategiei false (2/2)

- Notațiile
  - $\widehat{SR}_n$: raportul Sharpe estimat al strategiei n
  - $\mathrm{Var}[\widehat{SR}]$: varianța de eșantionare a unui raport Sharpe (SR, *Sharpe Ratio*) estimat, $\approx 1/T$ pe perioadă pentru randamente i.i.d., T fiind numărul de perioade
  - Φ: funcția de repartiție a distribuției Normale standard; Φ^{-1}: funcția ei cuantilă; $\gamma_E \approx 0{,}5772$: constanta Euler–Mascheroni
- SR_0: raportul Sharpe așteptat de la cea mai norocoasă dintre N strategii fără valoare
  - crește cu N și cu zgomotul de eșantionare $\sqrt{\mathrm{Var}[\widehat{SR}]}$

Cea mai bună dintre N strategii fără abilitate



- Strategii cu randamente i.i.d. de medie zero, 5 ani de date zilnice ($\mathrm{Var}[\widehat{SR}]$ anualizat $\approx 1/5$)
- Din $N = 1000$ de variante rezultă un raport Sharpe maxim de $\approx 1{,}45$ — **fără nicio abilitate**
- Axa orizontală: numărul de strategii N (scară logaritmică); axa verticală: cel mai bun SR anualizat
  - linia roșie: teorema; punctele: media Monte Carlo $\pm$ o abatere standard, în acord cu teorema

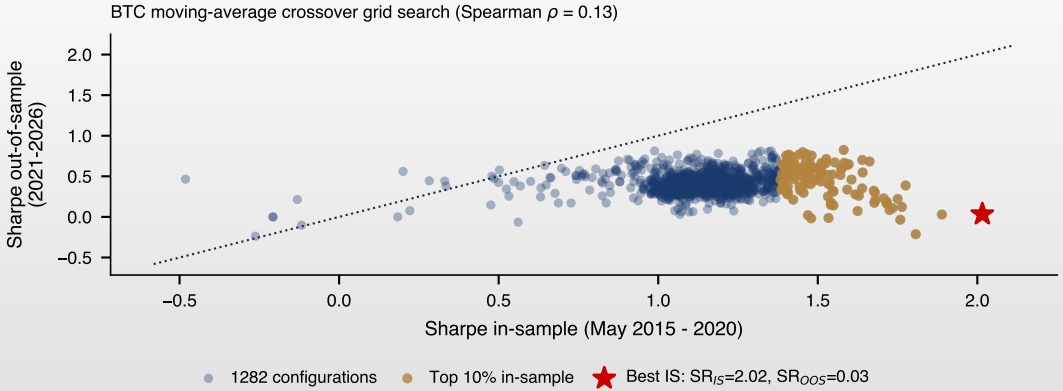
MFM_ch13_backtest_overfitting ▶ Interactiv

În eșantion (IS) și în afara lui (OOS): căutare pe grilă pe Bitcoin (1/2)

- Încrucișare de medii mobile (MA) pe Bitcoin
  - long cînd MA rapidă a log-prețului depășește MA lentă cu mai mult decît o bandă
- 1 284 de configurații: combinații de lungimi ale celor două MA și de lățimi ale benzii
  - în eșantion (*in-sample*, IS): 16 mai 2015 – 2020; în afara eșantionului (OOS): 2021–2026
  - familia testată, fixată doar din IS: cele 1 282 de reguli cu variantă nenulă în eșantion
  - 3 dintre ele stau în numerar tot OOS: SR OOS nedefinit, numărat ca 0
  - decizia la închiderea zilei t, tranzacția la închiderea $t+1$
- Graficul următor: SR IS pe axa orizontală, SR OOS pe axa verticală
  - punctele maro: cele mai bune 10% IS; steaua roșie: regula selectată; linia punctată: SR egale
- Regula selectată (MA de 11 și 20 de zile): SR **2,02** în eșantion și doar **0,03** în afara eșantionului
  - raportul Sharpe anualizat cu $\sqrt{365}$
- Corelația de rang Spearman dintre SR IS și SR OOS este $\rho = 0{,}13$
- SR median OOS: 0,50 pentru cele mai bune 10% dintre reguli IS
  - și 0,41 pentru toate regulile
- Rezultatul este compatibil cu supraoptimizarea regulii selectate
- Un SR pozitiv al familiei nu demonstrează informație predictivă incrementală
  - abilitatea față de buy-and-hold cere un test separat (SPA, în Aprofundări)
- Tipul de dovadă: comparație între proceduri (alegerea în eșantion, judecată în afara lui)

MFM_ch13_backtest_overfitting

În eșantion (IS) și în afara lui (OOS): căutare pe grilă pe Bitcoin (2/2)



MFM_ch13_backtest_overfitting ▶ Interactiv

Inferență pentru raportul Sharpe: eroarea standard și PSR (1/2)

- Întrebarea: cît de precis este un raport Sharpe estimat și cum îl comparăm cu un reper?
- Raportul Sharpe pe perioadă: $\widehat{SR} = \bar{r}/\hat{\sigma}$
  - randamente logaritmice zilnice, rata fără risc 0
- Forma distribuției, cu μ, σ media și abaterea standard a lui r
  - asimetria $\gamma_3 = \mathbb{E}[(r-\mu)^3]/\sigma^3$
  - kurtosis-ul obișnuit $\gamma_4 = \mathbb{E}[(r-\mu)^4]/\sigma^4$ (3 pentru distribuția Normală)
- Varianța de eșantionare a lui $\widehat{SR}$, randamente i.i.d., metoda delta (Lo (2002); Opdyke (2007)):

$$\mathrm{Var}(\widehat{SR}) \approx \frac{1}{T}\left(1 - \gamma_3\, SR + \frac{\gamma_4 - 1}{4}\, SR^2\right)$$

  - T: numărul de randamente; SR: raportul Sharpe real pe perioadă
  - asimetria negativă și cozile groase măresc incertitudinea
  - pentru distribuția Normală ($\gamma_3 = 0$, $\gamma_4 = 3$) varianța este $(1 + \frac{1}{2}SR^2)/T$
  - formula presupune moment de ordin patru finit, pus sub semnul întrebării în Capitolul 1

Inferență pentru raportul Sharpe: eroarea standard și PSR (2/2)

Definiție 5 (Raportul Sharpe probabilistic (PSR, probabilistic Sharpe ratio), Bailey & López de Prado (2012))

- Numit de autori probabilitatea ca SR real să depășească un reper SR^*
- Formal, Φ aplicat unei statistici z, cu varianța de mai sus la numitor:

$$\widehat{PSR}(SR^*) = \Phi\left(\frac{(\widehat{SR} - SR^*)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\,\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\,\widehat{SR}^2}}\right)$$

- SR^*: reperul, de exemplu 0 sau maximul așteptat SR_0
  - $\hat{\gamma}_3$, $\hat{\gamma}_4$: asimetria și kurtosis-ul empirice
- Interpretare: $\widehat{PSR} = 1 - p$, unde p este valoarea p unilaterală pentru $H_0 : SR \leq SR^*$
- $\widehat{PSR} > 0{,}95$ înseamnă respingerea lui H_0 la 5%
  - sub randamente i.i.d. și aproximarea prin metoda delta
  - nu o probabilitate de 95% pentru SR real

Raportul Sharpe deflatat: cea mai bună regulă pe Bitcoin

- **Raportul Sharpe deflatat** (DSR, Bailey & López de Prado (2014)): $DSR = \widehat{PSR}(SR_0)$
  - PSR cu reperul egal cu SR_0, maximul așteptat din teorema strategiei false
  - calculat cu varianța $\mathrm{Var}[\widehat{SR}_n]$ a încercărilor
- DSR corectează pentru selecție (N încercări), non-normalitate și lungimea eșantionului
  - $DSR \leq 0{,}95$: dovezi insuficiente la pragul ales, nu o dovadă că strategia nu are abilitate
  - raportăm întotdeauna N
- Exemplu rezolvat, cea mai bună regulă în eșantion: $T = 2\,057$ de zile, $N = 1\,282$ de încercări
  - $\widehat{SR} = 0{,}1055$ pe zi (2,02 anualizat), asimetrie $\hat{\gamma}_3 = 0{,}16$, kurtosis $\hat{\gamma}_4 = 12{,}23$
  - $\hat{\gamma}_4$ și DSR presupun moment de ordin patru finit, pus sub semnul întrebării în Capitolul 1
  - numitorul din PSR este $\sqrt{1 - 0{,}16 \cdot 0{,}1055 + \frac{12{,}23-1}{4} \cdot 0{,}1055^2} = 1{,}007$
  - $DSR = \Phi(z)$, cu $z = (\widehat{SR} - SR_0)\sqrt{T-1}/1{,}007$

$\mathrm{Var}[\widehat{SR}_n]$ din	SD $= \sqrt{\mathrm{Var}[\widehat{SR}_n]}$	SR_0 (pe zi)	z / DSR
dispersia celor 1 282 de SR	0,0109	0,036	3,11 / 0,999
încercări independente, $1/T$	0,0220	0,073	1,45 / 0,926

- Rîndul 1 folosește dispersia observată a celor 1 282 de SR, care include corelația dintre reguli
- Rîndul 2 tratează regulile ca independente ($\mathrm{Var} = 1/T$), varianta conservatoare
- Verdictul depinde de această alegere: SR de 0,03 OOS este în acord cu valoarea conservatoare (DSR = 0,926, sub 0,95)

Inferența după selecția variabilelor: selecția dublă (1/2)

- Întrebarea secțiunii: cum obținem un interval de încredere valid pentru efectul unei variabile?
  - după ce ML a ales controalele
- Traiectoria coeficienților LASSO răspunde la „ce caracteristică intră prima"
  - nu la „ce caracteristică contează"
  - statisticile t OLS calculate după LASSO ignoră pasul de selecție
- O variabilă de control corelată cu regresorul de interes d_t, dar doar slab cu y_t, este eliminată
  - efectul ei trece în coeficientul lui d_t (deplasare de variabilă omisă)
- Model parțial liniar: efectul lui d_t este liniar, controalele intră prin funcții necunoscute

$$y_t = \theta\, d_t + g(x_t) + \varepsilon_t, \qquad d_t = m(x_t) + v_t$$

  - y_t: variabila rezultat; d_t: regresorul de interes (nu ordinul d al FFD); θ: efectul lui d_t, parametrul de estimat
  - x_t: cele p controale; g: efectul lor asupra lui y_t
  - m: felul în care prezic pe d_t; ε_t, v_t: erori
  - ipoteza de raritate aproximativă (*sparsity*): cîteva controale explică cea mai mare parte din g și m

Inferența după selecția variabilelor: selecția dublă (2/2)

- Selecția dublă (*post-double selection*, Belloni, Chernozhukov & Hansen (2014)):
  1. LASSO pentru y_t pe $x_t \Rightarrow$ mulțimea selectată $\hat{S}_y$
  2. LASSO pentru d_t pe $x_t \Rightarrow$ mulțimea selectată $\hat{S}_d$
  3. OLS pentru y_t pe d_t și $x_{\hat{S}_y \cup \hat{S}_d}$, eroare standard robustă
- $\hat{S}_y$, $\hat{S}_d$: indicii controalelor păstrate de fiecare LASSO
  - $x_{\hat{S}_y \cup \hat{S}_d}$: controalele păstrate de oricare dintre cele două
- Un control este omis doar dacă nu contează în *niciuna* dintre ecuații
  - deci deplasarea de omisiune este de ordinul doi
- Procedura este validă pentru eșantionare independentă și modele rare
  - pentru serii de timp, erorile HAC sînt o aproximare
  - penalizarea și condițiile sînt în Anexă

Machine learning dublu (DML): scoruri ortogonale și cross-fitting

- Chernozhukov et al. (2018): rezidualizăm ambele variabile, apoi regresăm reziduul pe reziduu
  - cu *orice* algoritm (LASSO, RF, boosting)

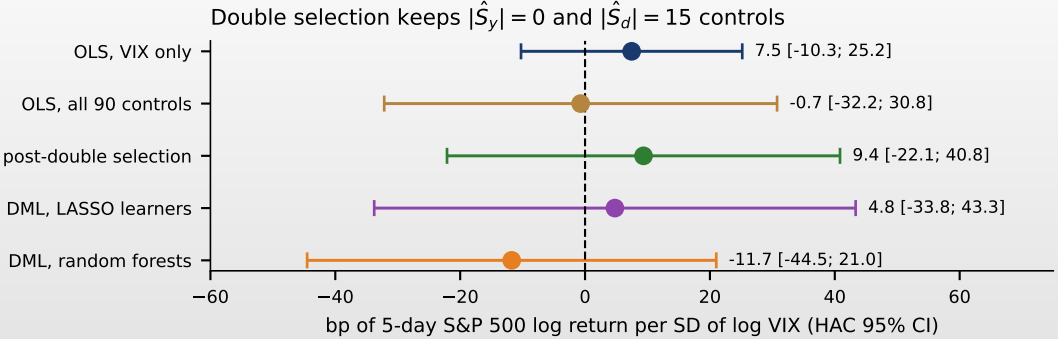
$$\tilde{y}_t = y_t - \hat{\ell}(\mathrm{x}_t), \quad \tilde{v}_t = d_t - \hat{m}(\mathrm{x}_t), \qquad \hat{\theta} = \frac{\sum_t \tilde{v}_t \tilde{y}_t}{\sum_t \tilde{v}_t^2}$$

  - $\hat{\ell}(\mathrm{x}) = \hat{\mathbb{E}}[y \mid \mathrm{x}]$ și $\hat{m}(\mathrm{x})$: predicțiile ML ale lui y_t și d_t din controale
  - $\tilde{y}_t$, $\tilde{v}_t$: părțile pe care controalele nu le explică; $\hat{\theta}$: panta OLS a lui $\tilde{y}_t$ pe $\tilde{v}_t$
- Scor Neyman-ortogonal $\psi = (\tilde{y} - \theta\tilde{v})\,\tilde{v}$
  - media lui este zero la valoarea reală a lui θ
  - erorile de ordinul întîi din $\hat{\ell}$ și $\hat{m}$ nu deplasează $\hat{\theta}$ (condițiile de rată în ▸ Anexă)
- Cross-fitting: estimăm $\hat{\ell}$, $\hat{m}$ pe $K-1$ pliuri și rezidualizăm pliul lăsat deoparte
  - repetăm pentru fiecare pliu
  - astfel eliminăm deplasarea de supraajustare pe propriile observații
- Teoria publicată presupune eșantionare independentă
  - cu etichete suprapuse pe 5 zile folosim pliuri contigue, purjate cu 5 zile în jurul pliului de test
  - și erori standard HAC, o adaptare în afara teoriei publicate

Exemplu rezolvat: puterea predictivă a VIX pentru randamentele S&P 500 (1/3)

- Întrebarea: prezice nivelul VIX randamentul S&P 500 pe următoarele 5 zile?
  - dat fiind un set mare de variabile de control
- y_t: randamentul logaritmic pe următoarele 5 zile, în puncte de bază (bp)
  - d_t: logaritmul VIX, standardizat (o abatere standard, SD, este 0,356)
- 90 de controale: 12 caracteristici tehnice, pătratele lor și produsele lor două cîte două
  - $n = 6\,269$ de zile; erori HAC cu 4 laguri
- LASSO-ul simplu elimină VIX și toate controalele
  - concluzia ar fi „niciun efect", fără nicio eroare standard
- Selecția dublă păstrează 0 controale în ecuația lui y_t ($|\hat{S}_y| = 0$)
  - și 15 în ecuația lui d_t ($|\hat{S}_d| = 15$)
- Graficul următor: pe fiecare rînd un estimator
  - punctul este estimarea lui θ (bp pe săptămînă la o creștere cu o SD a log VIX)
  - bara este CI (*Confidence Interval*, interval de încredere) HAC de 95%, linia punctată este efectul zero

Exemplu rezolvat: puterea predictivă a VIX pentru randamentele S&P 500 (2/3)



MFM_ch13_inference ▶ Interactiv

Exemplu rezolvat: puterea predictivă a VIX pentru randamentele S&P 500 (3/3)

- CI de 95% pentru θ prin selecție dublă: între $-22{,}1$ și $40{,}8$ bp pe săptămînă
  - la o creștere cu o abatere standard a log VIX
- OLS doar cu log VIX dă 7,5 $[-10{,}3;\ 25{,}2]$
  - OLS cu toate cele 90 de controale, selecția dublă și DML (*Double/Debiased Machine Learning*, machine learning dublu, fără deplasare, debiased) dau $-0{,}7$, 9,4, 4,8 (LASSO) și $-11{,}7$ (pădure aleatoare)
  - toate cu intervale de circa două ori mai largi
- Controalele care prezic pe d_t absorb o parte din variația log VIX
  - mai puțină variație proprie a lui d_t înseamnă un interval mai larg
  - prețul protecției împotriva deplasării de variabilă omisă
- Schimbarea de semn la DML-RF (DML cu păduri aleatoare) este zgomot
  - fiecare estimare punctuală se află în intervalele selecției duble și DML
- Intervalele corecte conțin zero și efecte mari de ambele semne
  - datele nu permit o concluzie
- Tipul de dovadă: asociere (variabilele se mișcă împreună; nu dovedește o relație cauzală)
  - o regresie predictivă: chiar un θ nenul nu ar arăta că VIX cauzează randamentele

MFM_ch13_inference

Virtutea complexității și o lectură critică [aprofundare]

- Kelly, Malamud & Zhou (2024): regresii ridge pe P caracteristici Fourier aleatoare ale cîtorva predictori
  - ferestre mobile (*rolling*) de doar $T = 12$ luni, P de ordinul miilor
  - caracteristici Fourier aleatoare: sin și cos ale unor combinații liniare aleatoare de predictori
  - teorie: pentru $P > T$, soluția de normă minimă (potrivirea exactă cu cel mai mic $\|\boldsymbol{\beta}\|_2$) induce implicit o contracție a coeficienților („double descent")
  - R^2_{OOS} poate fi negativ, în timp ce raportul Sharpe al sincronizării pieței **crește** cu complexitatea
  - complexitatea nu înseamnă automat supraajustare
- Nagel (2025): pentru $P \gg T$ prognoza este o medie ponderată după similaritate
  - media celor T randamente de antrenare
  - în ferestre scurte similaritatea înseamnă mai ales apropiere în timp
  - metoda devine o strategie de **momentum ajustat la volatilitate**
  - această strategie pierde pe date simulate cu reveniri la medie
- Avramov, Cheng & Metzker (2023): profiturile semnalelor deep learning vin din acțiuni greu de arbitrat
  - excluderea microcapitalizărilor, a firmelor în dificultate sau a episoadelor de volatilitate mare le atenuează
  - costurile le erodează în continuare
- Chen, Pelger & Zhu (2024): impun structură economică
  - condițiile de moment ale absenței arbitrajului devin funcția de pierdere a unei rețele adînci (slide-urile următoare)
- Sinteza domeniului: Kelly & Xiu (2023); manual: Dixon, Halperin & Bilokon (2020)

Studiu de caz: deep learning în evaluarea activelor, întrebare și metodă [aprofundare]

- Chen, Pelger & Zhu (2024): estimează factorul stocastic de actualizare (SDF) $M_{t+\mathbf{1}}$
  - SDF evaluează **toate acțiunile individuale de pe piața americană**
  - condiționat de 46 de caracteristici ale firmelor și de istoricul a 178 de serii macroeconomice
- Absența arbitrajului: SDF evaluează orice randament în exces (Secțiunea I.A, ec. 1)

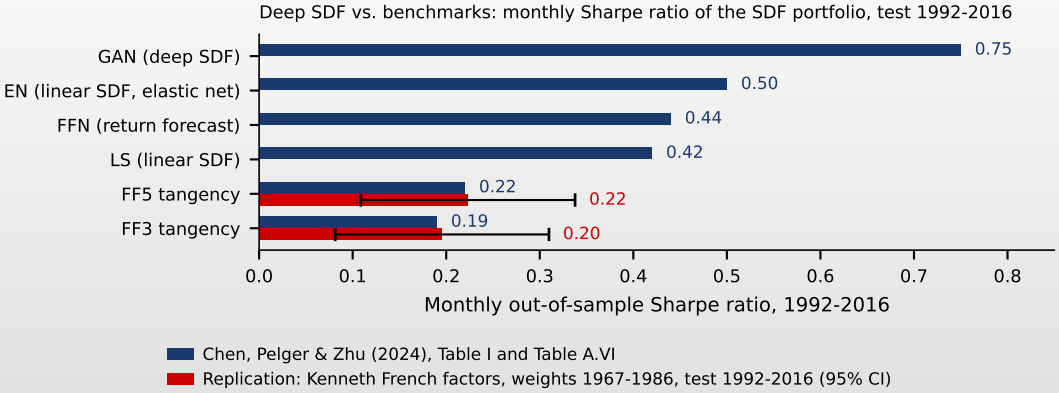
$$E_t[M_{t+\mathbf{1}} R^e_{t+\mathbf{1},i}] = 0, \qquad M_{t+\mathbf{1}} = 1 - \omega_t^\top R^e_{t+\mathbf{1}}$$

  - $R^e_{t+\mathbf{1},i}$: randamentul în exces al acțiunii i față de titlurile de stat pe termen scurt, din luna t în $t+1$
  - $R^e_{t+\mathbf{1}}$: vectorul randamentelor în exces ale tuturor acțiunilor
  - E_t: media condiționată de informația disponibilă la t
  - ω_t: ponderile unui portofoliu, funcție de caracteristici și de stările macro
- $F_{t+\mathbf{1}} = \omega_t^\top R^e_{t+\mathbf{1}}$, portofoliul SDF, este portofoliul tangent condiționat, iar

$$E_t[R^e_{t+\mathbf{1},i}] = \beta_{t,i}\, E_t[F_{t+\mathbf{1}}]$$

  - portofoliul tangent: raportul Sharpe maxim
  - $\beta_{t,i}$: expunerea acțiunii i la F
  - randamentul în exces așteptat este expunerea înmulțită cu prima lui F (derivarea în Aprofundări)

Deep learning în evaluarea activelor: rapoarte Sharpe OOS (1/2) [aprofundare]



MFM_ch13_sdf_replication ▶ Interactiv

Deep learning în evaluarea activelor: rapoarte Sharpe OOS (2/2) [aprofundare]

- Albastru: cifrele lucrării (versiunea publicată), SR lunar în afara eșantionului, 1992–2016 (Tabelul I)
  - GAN (rețeaua generativă adversarială a lucrării) 0,75, EN (SDF liniar cu elastic net) 0,50, FFN (rețeaua de prognoză *feedforward*) 0,44, LS (SDF liniar fără regularizare) 0,42
  - FF5 (*Fama–French five-factor model*, modelul Fama–French cu cinci factori) 0,22, FF3 0,19 (Tabelul A.VI)
  - modelul fără arbitraj regularizat întrece deja rețeaua de prognoză
- Roșu: replicarea, portofoliile tangente FF3 și FF5, $w \propto \hat{\Sigma}^{-1}\hat{\mu}$
  - pe factorii Kenneth French, cu ponderi fixate pe 1967–1986
  - SR lunar, cu CI de 95%: FF5 0,22 [0,11; 0,34], FF3 0,20 [0,08; 0,31]
  - eroarea standard: 0,058 (Lo (2002)), care presupune moment de ordin patru finit, pus sub semnul întrebării în Capitolul 1
  - $\hat{\mu}$, $\hat{\Sigma}$: vectorul mediilor și matricea de covariață empirică ale randamentelor factorilor
- Tipul de dovadă: comparație între proceduri (SR în afara eșantionului ale modelelor)
  - după 2016, aceleași ponderi fixe dau SR lunar 0,06 (FF5) și 0,03 (FF3): Aprofundări

MFM_ch13_sdf_replication

Descoperire sau căutare oportunistă în date la scară mare?

- Chen, Lopez-Lira & Zimmermann (2022): căutarea printre 29 000 de rapoarte contabile a celor cu $t > 2$
  - dă o predictibilitate transversală asemănătoare cu cea a predictorilor publicați în reviste cu evaluare inter pares
  - pentru ambele, circa jumătate rămîne după eșantioanele originale
- Novy-Marx & Velikov (2025): un flux automat cu un model lingvistic mare (LLM, *large language model*) caută peste 30 000 de semnale
  - păstrează 96 care trec un protocol fix
  - scrie sute de articole complete, cu motivații inventate și cîteva citări imaginare
  - HARKing (*hypothesising after the results are known*) la scară mare: ipoteze formulate după ce se cunosc rezultatele
- Întrebarea deschisă: cînd un sistem automat propune variabilele, modelul și interpretarea
  - care este numărul de încercări?
  - ce separă o descoperire de cea mai bună dintre multe extrageri?
- Bucla descoperirii
  - literatura → ipoteze concurente → date și cod → robustețe fixată dinainte
  - → critică de tip recenzent → raportarea fiecărui test
  - prompturi reutilizabile în ▸ Anexă
- Ce verifică cercetătorul
  - fiecare referință există (DOI, arXiv)
  - fiecare încercare este notată, deci numărul de teste M din corecție le numără pe toate
  - testele corelate primesc un reper comun (placebo sau bootstrap comun al scorurilor)
  - transformările rămîn în pliul de antrenare

Mini-studiu de caz: 102 teste, un singur efect cunoscut trece pragul placebo (1/2)

- O propunere tipică a unui asistent: „adaugă mai mulți indicatori”
  - „momentum, distanțe față de mediile mobile, rapoarte de volatilitate, RSI și variații ale VIX pe mai multe ferestre”
- 34 de variabile $\times$ 3 orizonturi (1, 5, 20 de zile) = 102 regresii predictive pe 2001–2015
  - statistici t Newey–West
  - doar ținte a căror fereastră se încheie pînă la 31 decembrie 2015: 1, 5 și 20 de zile eliminate la graniță
  - regresia j: $r_{t,t+h} = \alpha_j + \beta_j x_{j,t} + e_t$, cu ipoteza nulă $H_{0,j} : \beta_j = 0$
  - $x_{j,t}$: variabila j; $r_{t,t+h}$: randamentul pe următoarele h zile
- Corecții pentru $M = 102$ teste, cu $p_{(1)} \leq \cdots \leq p_{(M)}$ valorile p ordonate
  - Holm: respingem cît timp $p_{(j)} \leq 0{,}05/(M - j + 1)$
  - Holm controlează FWER (probabilitatea de cel puțin o respingere falsă)
  - Benjamini–Hochberg (BH, Benjamini & Hochberg (1995)): respingem cele mai mici j, pentru cel mai mare j cu $p_{(j)} \leq j\, q/M$
  - controlează rata falselor descoperiri (FDR) la nivelul q, proporția așteptată a respingerilor false printre toate respingerile
  - placebo prin deplasare: cuantila de 95% a lui $\max_j |t_j|$ cînd ținta este deplasată circular
  - deplasarea păstrează corelația dintre variabile, dar schimbă și relația lor cu volatilitatea condiționată: un reper euristic, nu un test exact

Mini-studiu de caz: 102 teste, un singur efect cunoscut trece pragul placebo (2/2)

- **22** din cele 102 teste au $|t| > 1{,}96$, față de 5,1 așteptate sub ipoteza nulă globală la 5%
- Numărul de teste semnificative după corecție
  - Holm, cu FWER de 5%: **0**
  - BH, cu FDR de 10%: **13**
  - pragul $|t| > 3$ (Harvey, Liu & Zhu (2016)): **4**
- Placebo prin deplasare pentru max-$|t|$: cuantila de 95% a lui $\max_j |t_j| = 3{,}32$
  - 500 de deplasări circulare ale țintei
  - un singur predictor depășește pragul, randamentul de azi pentru randamentul de mîine, $t = -3{,}38$
  - inversarea pe termen scurt (*short-term reversal*), un efect cunoscut
- În perioada de verificare 2016–2026, doar 3 din cele 22 își păstrează semnul cu $|t| > 1{,}96$
  - replicări nominale, toate variante de inversare pe termen scurt
  - cu Holm la 5% pe cele 22 de verificări, niciuna nu trece (p minim 0,019, pragul $0{,}05/22 = 0{,}0023$)
  - „descoperirile" bazate pe VIX nu se confirmă
- BH cere independență sau dependență pozitivă a valorilor p ale ipotezelor nule adevărate
  - predictorii corelați nu o asigură
- Tipul de dovadă: asociere
  - inversarea pe termen scurt este o regularitate predictivă, nu un mecanism identificat

MFM_ch13_ai_discovery

Un punct de plecare pentru proiect

- **Întrebările**
  - poate o căutare de variabile asistată de AI să găsească un predictor al randamentelor indicilor?
  - indicii: S&P 500, BET – indicele de referință al Bursei de Valori București, Bitcoin
  - rămîne acesta semnificativ după o corecție preînregistrată și pe o perioadă de verificare?
  - adaugă el informație dincolo de inversarea pe termen scurt?
- **Datele**: S&P 500 și Bitcoin sînt în datele cursului; BET nu este
  - echipa care alege BET numește sursa (de ex. Bursa de Valori București)
  - îngheață eșantionul și notează data descărcării, n, prima și ultima închidere
- **Livrabile**
  - o listă de variabile stabilită înainte de examinarea datelor
  - toate testele, corectate cu Holm, BH și ipoteza nulă comună (max-$|t|$)
  - o perioadă de verificare separată
  - DSR și PBO pentru orice strategie construită pe predictorii rămași
- **Jurnalul** (șabloane)
  - fiecare prompt care a influențat o decizie, în `AI_USE.md`
  - fiecare referință inventată, scurgere de informație sau încercare omisă, în `AI_ERRORS.md`, cu testul care a identificat-o

Concluzii-cheie

Patru lecții empirice

- O respingere ADF nu înseamnă staționaritate
  - FFD trunchiată a log-prețului S&P 500 rămîne $I(1)$
  - memoria trebuie estimată direct, cu local Whittle
- K-fold amestecat „găsește" abilitate pe zgomot pur
- Pe S&P 500 niciun model nu depășește „mereu în sus" (testele PT, DM, GW)
- Cel mai bun dintre multe backtest-uri trebuie deflatat sau testat prin bootstrap
  - pe grila Bitcoin, cea mai bună regulă în eșantion nu păstrează nimic în afara lui

Urmează

- Capitolul 14
  - deep learning și modele fundaționale pentru serii de timp
- Capitolul 15
  - modele lingvistice mari și analiza sentimentului

O listă de verificare înainte de orice afirmație

- ☐ Date disponibile la momentul deciziei, transformări estimate doar pe pliurile de antrenare
- ☐ Etichete scalate cu volatilitatea, cu intervalele înregistrate
- ☐ K-fold cu purjare sau CPCV cu embargo pentru selecție, WF pentru testul final
- ☐ Un reper naiv, testele PT, DM / CW (testul Clark–West pentru modele imbricate) / GW și intervale bootstrap pe blocuri
- ☐ Costuri incluse și fiecare încercare înregistrată
- ☐ DSR, SPA și PBO pentru strategia selectată
- ☐ „Selectat" sau „important" nu înseamnă „semnificativ": selecție dublă sau DML pentru inferență

Verificați-vă înțelegerea

1. De ce logit-ul cu acuratețea de 57,7% nu are abilitate, cînd 58,1% dintre zile sînt „sus”?
2. Ce spune o corelație de 0,98 între seria FFD ($d = 0{,}25$) și log-preț despre o respingere ADF?
3. De ce K-fold amestecat dă 69,4% pe o țintă fără predictibilitate, iar K-fold cu purjare și embargo 49,6%?
4. Cum se schimbă SR_0, maximul așteptat al celor N strategii fără abilitate, cînd N crește de la 10 la 1 000?
5. De ce cea mai bună regulă pe Bitcoin, cu SR de 2,02 în eșantion, are DSR 0,999 sau 0,926, după varianța aleasă?
6. De ce intervalul DeLong pentru AUC, [0,466; 0,494], și cel bootstrap pe blocuri, [0,456; 0,505], duc la concluzii diferite?
7. Ce întrebare despre VIX rezolvă selecția dublă și nu o rezolvă traiectoria LASSO?

Decizia practică

- **Ce aleg**
  - pentru direcția S&P 500 pe 5 zile: buy-and-hold, nu strategia GBM
  - pe ferestre succesive, 2010–2026, raportul Sharpe este 0,66 pentru buy-and-hold și 0,33 pentru GBM
  - pentru orice model: CV cu purjare și embargo la selecție, ferestre succesive la evaluarea finală
- **În ce condiții**
  - semnalul direcțional rămîne slab: niciun model nu depășește reperul „mereu în sus", de 58,1%
  - fiecare schimbare de poziție costă 5 bp
  - fiecare încercare este înregistrată, iar numărul de încercări N este raportat
- **Ce rezultat m-ar face să renunț**
  - un model care depășește buy-and-hold în afara eșantionului, după costuri
  - cu statistica PT peste 1,645
  - cu intervalul bootstrap pe blocuri pentru AUC în întregime peste 0,5
  - cu DSR peste 0,95 la varianța conservatoare $1/T$ (cea mai bună regulă pe Bitcoin are 0,926)

Dovezi și valoarea economică a unui R^2_{OOS} mic (1/2)

- Campbell & Thompson (2008): valoarea unui R^2 mic pentru un investitor

$$r_{t+1} = \mu + x_t + \varepsilon_{t+1}, \qquad R^2 = \frac{\mathrm{Var}(x_t)}{\mathrm{Var}(r_{t+1})}$$

  - randamentul are o parte predictibilă și una impredictibilă
  - μ: randamentul mediu; x_t: componenta predictibilă, cunoscută la t, de medie zero; ε_{t+1}: șocul impredictibil
  - R^2: proporția din varianța randamentului explicată de predictor
- Un investitor medie-variață fără restricții, care cunoaște x_t și se reoptimizează în fiecare perioadă

$$\mathbb{E}[SR_t^2] = \frac{SR_{BH}^2 + R^2}{1 - R^2}$$

  - SR_t: raportul Sharpe condiționat la t; fără costuri, fără erori de estimare
  - $\mathbb{E}[SR_t^2]$ nu este, în general, pătratul raportului Sharpe necondiționat al strategiei dinamice
  - utilitatea medie-variață optimă la t este $SR_t^2/(2\gamma)$, cu γ aversiunea la risc, deci cîștigul se citește prin utilitatea medie
  - SR_{BH}: SR al strategiei buy-and-hold, care ignoră x_t; pentru $R^2 = 0$ avem $\mathbb{E}[SR_t^2] = SR_{BH}^2$

Dovezi și valoarea economică a unui R^2_{OOS} mic (2/2)

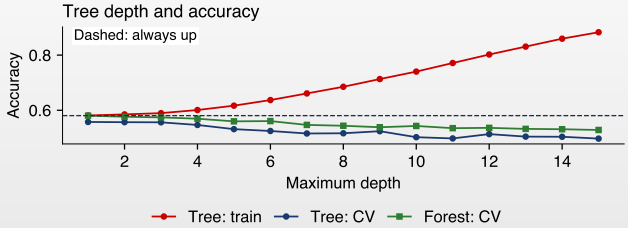
- S&P 500, randamente logaritmice lunare 2001–2026: buy-and-hold are $SR_{BH} = 0{,}128$ pe lună
  - adică 0,44 anualizat cu $\sqrt{12}$
  - un predictor cu $R^2 = 0{,}5\%$ pe lună dă $\sqrt{\mathbb{E}[SR_t^2]} = 0{,}51$ anualizat (cu 15% peste SR_{BH})
  - un predictor cu $R^2 = 1\%$ dă 0,57 (cu circa 27% peste SR_{BH})
- Pentru un **singur indice**, eșantionul efectiv este mult mai mic decît pentru mii de acțiuni
  - situația studiului de caz pe S&P 500

MFM_ch13_inference

De unde vine formula Campbell–Thompson

- Același model: $r_{t+1} = \mu + x_t + \varepsilon_{t+1}$, $\mathbb{E}[x_t] = 0$, ε_{t+1} independent de x_t
  - notăm $\sigma_r^2 = \mathrm{Var}(r_{t+1})$
- Pasul 1, riscul pe care îl suportă în continuare investitorul informat
  - $\mathrm{Var}(r_{t+1}) = \mathrm{Var}(x_t) + \sigma_\varepsilon^2$ și $R^2 = \mathrm{Var}(x_t)/\sigma_r^2$, deci
$$\sigma_\varepsilon^2 = (1 - R^2)\,\sigma_r^2$$
- Pasul 2, raportul Sharpe condiționat al investitorului la t
  - media $\mu + x_t$, varianța σ_ε^2
  - pătratul SR: $(\mu + x_t)^2/\sigma_\varepsilon^2$
- Pasul 3, media după t: $\mathbb{E}[(\mu + x_t)^2] = \mu^2 + \mathrm{Var}(x_t) = \mu^2 + R^2\sigma_r^2$
- Pasul 4, împărțim și folosim $SR_{BH} = \mu/\sigma_r$:
$$\mathbb{E}[SR_t^2] = \frac{\mu^2 + R^2\sigma_r^2}{(1 - R^2)\,\sigma_r^2} = \frac{SR_{BH}^2 + R^2}{1 - R^2}$$
- Deci formula dă media pătratului raportului Sharpe *condiționat*, nu SR necondiționat al strategiei
  - al unui investitor care se reoptimizează în fiecare lună
  - în unitățile datelor: aici, lunar
- Cifrele, pentru $R^2 = 0{,}005$
  - $\mathbb{E}[SR_t^2] = (0{,}128^2 + 0{,}005)/0{,}995 = 0{,}02149$, $\sqrt{\mathbb{E}[SR_t^2]} = 0{,}1466$ pe lună
  - $\times\sqrt{12} = 0{,}51$ pe an; $0{,}1466/0{,}128 = 1{,}15$

Deplasare–varianță în practică: adîncimea arborelui



- S&P 500, direcția pe 5 zile, cînd adîncimea crește
  - acuratețea pe antrenare urcă la $\approx 88\%$ la adîncimea 15
  - acuratețea CV cu purjare **scade** spre 50%
  - CV cu purjare: antrenarea exclude etichetele suprapuse cu testul (Secțiunea 5)
- Axa orizontală: adîncimea maximă a arborelui; axa verticală: acuratețea
  - roșu antrenarea, albastru CV al unui arbore, verde CV al pădurii
- Random forest se degradează mai lent, dar nu depășește niciodată reperul „mereu în sus” (linia punctată)

MFM_ch13_regularization

Ponderile diferențierii fracționare



- Axa orizontală: lagul k; axa verticală: ponderea w_k, pentru cinci valori ale lui d
- Pentru $d = 1$ contează doar primul lag
- Pentru d mic ponderile lagurilor îndepărtate rămîn nenule
- Toate ponderile după w_0 sînt negative
  - $\tilde{X}_t$ compară nivelul de azi cu o istorie ponderată

MFM_ch13_ffd

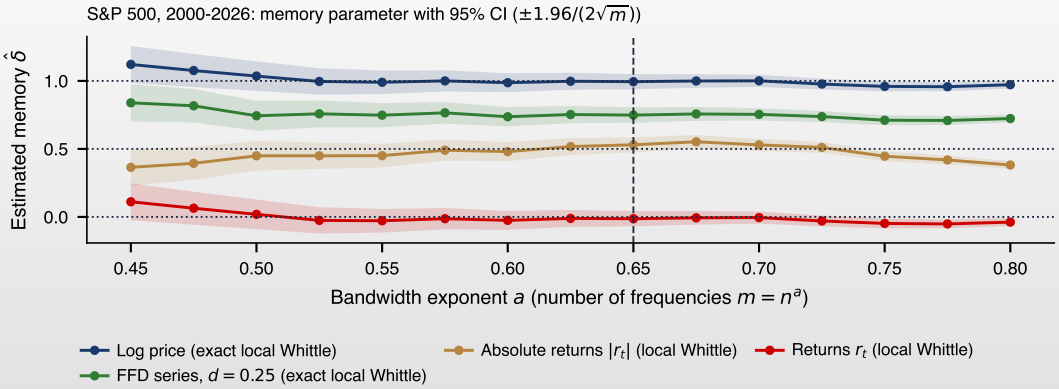
Memoria estimată direct: local Whittle (1/3)

- Ideea: o serie cu memorie lungă are cea mai mare parte a variației în ciclurile lente
  - panta spectrului lîngă frecvența 0 arată δ (aici λ este frecvența, nu penalizarea)

$$f(\lambda) \sim G\,\lambda^{-2\delta} \quad \text{pentru } \lambda \to 0$$

  - $f(\lambda)$: densitatea spectrală, variația adusă de ciclurile de frecvență λ (perioadă de $2\pi/\lambda$ zile); $G > 0$: o constantă de scală; $\sim$: raportul celor doi membri tinde la 1
  - $\delta > 0$: f explodează lîngă 0 (memorie lungă); $\delta = 0$: f este plat (memorie scurtă); $\delta < 0$: f tinde la 0
- Local Whittle (LW, Robinson (1995)): ajustează această lege doar pe frecvențele joase
  - primele $m = n^a$ frecvențe Fourier $\lambda_j = 2\pi j/n$, $j = 1, \ldots, m$
  - n: mărimea eșantionului; $a \in (0,1)$: exponentul lățimii de bandă; un a mai mare folosește mai multe frecvențe: variația scade, dar crește deplasarea din dinamica pe termen scurt
- Local Whittle exact (ELW, Shimotsu & Phillips (2005)): valid și pentru $\delta \geq \frac{1}{2}$ (serii nestaționare)
  - ambii estimatori sînt detaliați în ▸ Anexă
- Intervalul de încredere (CI, confidence interval) de 95% este $\hat{\delta} \pm 1{,}96/(2\sqrt{m})$
  - eroarea standard $1/(2\sqrt{m})$ depinde doar de m

Memoria estimată direct: local Whittle (2/3)



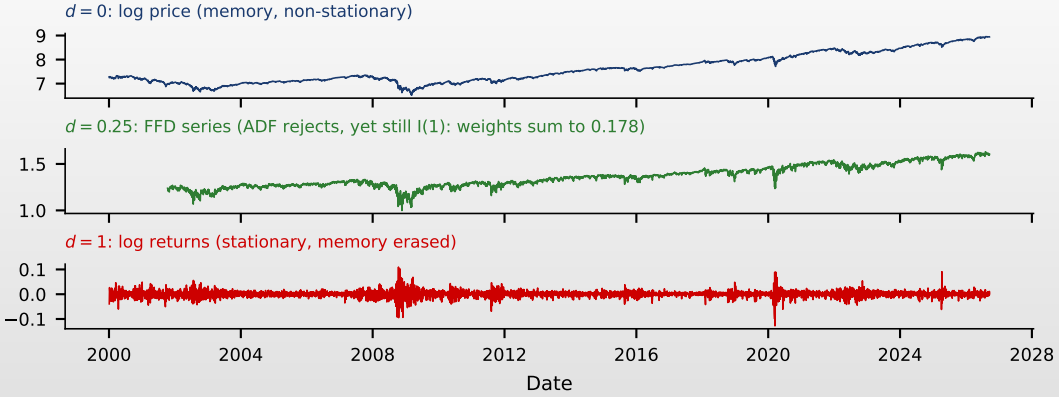
MFM_ch13_ffd

Memoria estimată direct: local Whittle (3/3)

- Graficul: exponentul a (cu $m = n^a$) pe axa orizontală, $\hat{\delta}$ pe axa verticală
  - banda CI de 95% în jurul estimării
  - linia verticală marchează $a = 0{,}65$, valoarea folosită mai jos (S&P 500, 2000–2026)
- Log-prețul (ELW): $\hat{\delta} = 0{,}99$, CI $[0{,}94;\ 1{,}05]$
- Seria FFD (ELW): $\hat{\delta} = 0{,}75$, CI $[0{,}69;\ 0{,}81]$
- Randamentele: $\hat{\delta} = -0{,}01$, CI $[-0{,}07;\ 0{,}04]$
- Randamentele absolute $|r_t|$: $\hat{\delta} = 0{,}53$, CI $[0{,}47;\ 0{,}59]$
- Pentru seria FFD estimarea rămîne aproape de $1 - d$ în banda de frecvențe folosită
  - deși ordinul de integrare este tot 1
- Memoria lungă este în **volatilitate** (Capitolele 5 și 9)
  - deci filtrele fracționare au sens pentru $|r_t|$, nu pentru preț

MFM_ch13_ffd

Prețuri, serie FFD și randamente (1/2)



MFM_ch13_ffd

Prețuri, serie FFD și randamente (2/2)

- Panoul de sus arată log-prețul S&P 500, o serie $I(1)$
- Panoul din mijloc arată seria FFD cu $d = 0{,}25$
  - trendul prețului rămîne vizibil, pentru că seria este $0{,}178 \log P_t$ plus o parte staționară
- Panoul de jos arată randamentele logaritmice, fără trend
  - se vede doar gruparea volatilității (*volatility clustering*)

MFM_ch13_ffd

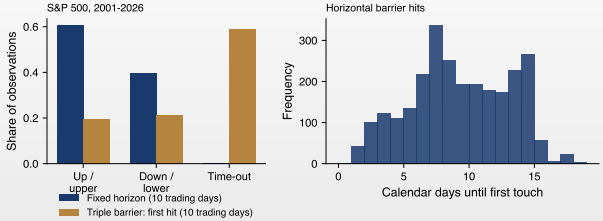
Volatilitatea EWMA din barierele etichetelor

- σ_{t_0}: rădăcina mediei mobile ponderate exponențial (EWMA), cu span 50

$$\sigma_t^2 \approx \alpha \sum_{k \geq 0} (1-\alpha)^k \left(r_{t-k} - \bar{r}_t\right)^2, \qquad \alpha = \frac{2}{\text{span} + 1} = \frac{2}{51} \approx 0{,}039$$

  - media se calculează pe abaterile pătratice ale randamentelor logaritmice zilnice
  - $\bar{r}_t$: media ponderată exponențial a randamentelor; span 50: o memorie efectivă de circa două luni de tranzacționare
- De ce bariere scalate cu volatilitatea
  - un prag fix ar fi atins aproape zilnic în 2008 sau în martie 2020 și rar în anii calmi

Distribuția etichetelor: orizont fix și trei bariere



- S&P 500, 2001–2026, $h = 10$, bariere la $\pm 1\,\sigma\sqrt{h}$: 19,6% sus, 21,3% jos, 59,0% expirare
  - etichetele, după regula semnului la bariera verticală: 59,1% $+1$ și 40,9% -1 (orizontul fix: 60,5% sus)
- Graficul numără **tipurile de ieșire**
  - expirare (time-out) = ieșire la bariera verticală
  - cu regula semnului, aceste 59,0% primesc tot ± 1 după semnul randamentului
  - cu convenția „0” formează o a treia clasă
- Panoul din dreapta: histograma zilelor calendaristice pînă la prima barieră orizontală atinsă
  - de obicei după 1–2 săptămîni

MFM_ch13_triple_barrier

Etichete suprapuse și unicitatea eșantioanelor (1/2)

- Eticheta i acoperă intervalul $[t_{0,i}, t_{1,i}]$
  - etichetele consecutive au randamente comune, deci eșantioanele **nu sînt i.i.d.**
- Etichetele active simultan: numărul de etichete al căror interval conține ziua t

$$c_t = \sum_i 1\{t_{0,i} \leq t \leq t_{1,i}\}$$

- Unicitatea medie a etichetei i

$$\bar{u}_i = \frac{1}{t_{1,i} - t_{0,i} + 1} \sum_{t=t_{0,i}}^{t_{1,i}} \frac{1}{c_t}$$

  - partea medie din fiecare zi a intervalului ei care îi aparține doar ei
  - $\bar{u}_i = 1$: eticheta i nu are niciun randament comun cu alte etichete
  - $\bar{u}_i = 1/5$: media armonică a lui c_t pe intervalul etichetei i este cinci

Etichete suprapuse și unicitatea eșantioanelor (2/2)

Consecințe

- Mărimea efectivă a eșantionului $\approx \sum_i \bar{u}_i \ll n$
  - n: numărul de etichete
  - o euristică a suprapunerii, nu o măsură bazată pe varianță
- Bagging-ul extrage observații aproape identice
  - scorurile CV sînt supraestimate

Remedii

- Ponderare cu $\bar{u}_i$ (și cu $|r|$)
- Bootstrap secvențial
- În RF, fracția de observații extrase pentru fiecare arbore
  - $\approx \bar{u}$, unicitatea medie

Transformări doar pe datele de antrenare: testul fără FFD

- „d minim" a fost ales pe 2000–2026 și apoi folosit în CV cu purjare
  - pliurile de test au influențat alegerea unei caracteristici
  - aceeași regulă ADF dă $d = 0{,}15$ pe datele pînă în 2009 și $d = 0{,}25$ pe tot eșantionul
- Studiul de caz fără log-prețul FFD (CV cu purjare în 5 pliuri, media pe pliuri)
  - acuratețe: logit 57,5%, RF 56,4%, GBM 55,6%, MLP 56,5%
  - toate sub 58,1%
  - AUC: logit 0,483, RF 0,520, GBM 0,514, MLP 0,526
- Concluzia negativă nu depinde de această caracteristică
- Protocolul corect
  - estimăm d (local Whittle, cu CI) doar în fiecare pliu de antrenare
  - preferăm caracteristici cu sens economic
  - de exemplu distanțe față de mediile mobile, filtre fracționare ale lui $|r_t|$

MFM_ch13_inference

Etichete suprapuse și varianța de lungă durată Newey–West

- Testele de semnificație ale studiului de caz (Pesaran–Timmermann) fac media unei serii zilnice z_t
  - de exemplu o potrivire, o diferență de pierderi d_t, un scor de regresie
  - cu etichete pe 5 zile, valorile z_t vecine au randamente comune, deci $\mathrm{Var}(\bar{z}) \neq \mathrm{Var}(z_t)/n$
- Varianța de lungă durată Newey–West (Newey & West (1987)), cu ponderi Bartlett

$$\hat{\Omega}_{NW} = \hat{\Gamma}_{\mathbf{0}} + 2\sum_{j=\mathbf{1}}^{q}\Big(1 - \frac{j}{q+1}\Big)\hat{\Gamma}_j, \qquad \hat{\Gamma}_j = \frac{1}{n}\sum_{t=j+\mathbf{1}}^{n}(z_t - \bar{z})(z_{t-j} - \bar{z})$$

  - robustă la heteroscedasticitate și autocorelație (HAC)
  - $\hat{\Gamma}_j$: autocovarianța empirică a lui z_t la lagul j; q: numărul de laguri („lag 4" în tabele); $\widehat{\mathrm{SE}}(\bar{z}) = \sqrt{\hat{\Omega}_{NW}/n}$
  - regula pentru laguri: etichetele pe h zile au randamente comune pînă la $h-1$ zile distanță, deci $q \geq h-1 = 4$ este minimul aici
- Exemplu: z_t = suma următoarelor 5 randamente zilnice i.i.d. de varianță $\sigma^{\mathbf{2}}$
  - $\Gamma_j = (5-j)\sigma^{\mathbf{2}}$ pentru $j \leq 4$, 0 după
  - formula i.i.d.: $\Gamma_{\mathbf{0}} = 5\sigma^{\mathbf{2}}$; varianța de lungă durată reală: $\Gamma_{\mathbf{0}} + 2(4+3+2+1)\sigma^{\mathbf{2}} = 25\sigma^{\mathbf{2}}$, deci un t i.i.d. este de $\sqrt{5} \approx 2{,}2$ ori prea mare
  - Bartlett cu $q=4$: $5 + 2(0{,}8\cdot 4 + 0{,}6\cdot 3 + 0{,}4\cdot 2 + 0{,}2\cdot 1) = 17\sigma^{\mathbf{2}}$; ponderile atenuează lagurile îndepărtate, deci $q = h-1$ este un minim, nu o alegere generoasă

Compararea prognozelor: Diebold–Mariano, Clark–West, Giacomini–White (1/2)

- Două prognoze de probabilitate pentru $y_t \in \{0, 1\}$
  - modelul 1 dă $\hat{p}_{1,t}$, reperul (modelul 0, de ex. o constantă) dă $\hat{p}_{0,t}$
- Pierderea unei prognoze $\hat{p}_t$
  - log-loss-ul de pe slide-ul despre pierdere și penalizare, cu $f = \hat{p}_t$, sau pierderea Brier $(y_t - \hat{p}_t)^2$
- Diferența pierderilor (alt d_t decît în FFD și DML): $d_t = L(y_t, \hat{p}_{1,t}) - L(y_t, \hat{p}_{0,t})$
  - $d_t > 0$ înseamnă că modelul 1 este mai slab în ziua t
- **Diebold–Mariano** (DM, Diebold & Mariano (1995)): este zero diferența medie a pierderilor?

$$DM = \frac{\bar{d}}{\sqrt{\hat{\Omega}_{NW}/n}} \xrightarrow{d} N(0,1) \quad \text{sub } H_0 : \mathbb{E}[d_t] = 0$$

  - $\bar{d}$: media lui d_t pe cele n zile de test
  - $\hat{\Omega}_{NW}$: varianța de lungă durată Newey–West a lui d_t (formula Newey–West, cu $z_t = d_t$)
  - $DM > 1{,}96$: modelul 1 este semnificativ mai slab la 5%; $DM < -1{,}96$: semnificativ mai bun

Compararea prognozelor: Diebold–Mariano, Clark–West, Giacomini–White (2/2)

- **Modele imbricate**: logit-ul cu 14 caracteristici include constanta, $\beta = 0$
  - sub H_0 modelul extins estimează cu zgomot coeficienți nuli, $\mathbb{E}[d_t] > 0$
  - DM rareori îl favorizează
- **Clark–West** (CW, Clark & West (2007)): test t HAC unilateral pentru $\mathbb{E}[f_t] = 0$ față de $\mathbb{E}[f_t] > 0$

$$f_t = (y_t - \hat{p}_{0,t})^2 - \left[(y_t - \hat{p}_{1,t})^2 - (\hat{p}_{0,t} - \hat{p}_{1,t})^2\right]$$

  - cîștigul de pierdere Brier al modelului 1, plus termenul $(\hat{p}_{0,t} - \hat{p}_{1,t})^2$
  - termenul elimină costul de zgomot al estimării unor coeficienți nuli
  - un t mare și pozitiv: caracteristicile adaugă informație
  - derivat pentru modele liniare imbricate estimate prin metoda celor mai mici pătrate (OLS)
  - pentru probabilitățile unui logit penalizat este o aproximare, nu o teoremă
- **Giacomini–White** (GW, Giacomini & White (2006)): testează *metoda* de prognoză așa cum este folosită
  - cu o fereastră de estimare mărginită
- Prin contrast, CW întreabă dacă modelul populației adaugă informație
  - statistica GW în Anexă

Importanța caracteristicilor: MDI și MDA

- Întrebarea: pe ce caracteristici se sprijină modelul și care dintre ele ajută predicția pe date noi?
- Scăderea medie a impurității (MDI), doar pentru arbori, calculată în eșantion:

$$\mathrm{MDI}_j = \frac{1}{B}\sum_{b=\mathbf{1}}^{B} \sum_{v \in T_b:\ j(v)=j} \frac{n_v}{n}\Big[G_v - \frac{n_L}{n_v}\,G_L - \frac{n_R}{n_v}\,G_R\Big]$$

  - suma reducerilor de impuritate Gini ale nodurilor v care împart după caracteristica j
  - ponderate cu fracțiunea n_v/n a observațiilor care ajung în nod
  - favorizează caracteristicile continue, cu multe praguri posibile
- Scăderea medie a acurateței (MDA), pentru orice model, pe date de test:

$$\mathrm{MDA}_j = S\big(\hat{f};\ \mathcal{D}_{test}\big) - \frac{1}{R}\sum_{r=\mathbf{1}}^{R} S\big(\hat{f};\ \mathcal{D}_{test}^{\pi_r(j)}\big)$$

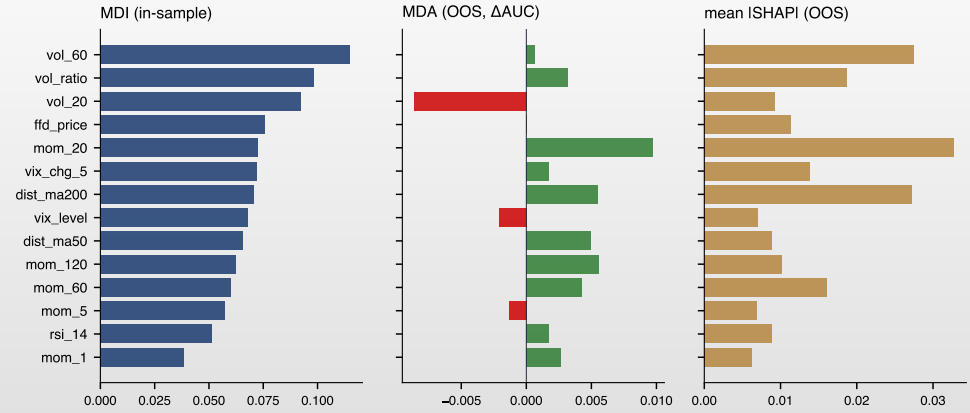
  - S: scorul (aici AUC); $\mathcal{D}_{test}^{\pi_r(j)}$: datele de test cu coloana j permutată aleator, în repetiția r
  - $\mathrm{MDA}_j < 0$: scorul a crescut după permutare; se citește împreună cu dispersia celor R permutări și cu incertitudinea estimării
  - un MDA negativ mic nu arată singur că o caracteristică dăunează

Importanța caracteristicilor: SHAP

- SHAP (Lundberg & Lee (2017)): valoarea Shapley $\phi_j(\mathrm{x})$
  - contribuția marginală medie a caracteristicii j la predicția pentru x
  - mediată pe toate coalițiile de alte caracteristici (formula coaliților: Anexă)
- Aditivă: $\hat{f}(\mathrm{x}) = \phi_0 + \sum_j \phi_j(\mathrm{x})$, ϕ_0 fiind predicția medie
- Importanța globală a caracteristicii j: $\frac{1}{n}\sum_i |\phi_j(\mathrm{x}_i)|$, calculată pe date de test
- MDI arată pe ce caracteristici s-a sprijinit modelul
- SHAP atribuie predicția modelului între caracteristici; nu folosește rezultatul realizat
- MDA măsoară schimbarea scorului pe setul evaluat cînd caracteristica este permutată
- Caracteristicile corelate își împart și își diluează importanța (substituție)
  - deci trebuie grupate întîi (López de Prado (2020))

Shapley Values

Importanța caracteristicilor în studiul de caz (1/2)



MFM_ch13_feature_importance

Importanța caracteristicilor în studiul de caz (2/2)

- Estimare pe primele 70%, evaluare pe ultimele 30% după purjare
  - **RF** dă MDI și MDA, **GBM** dă SHAP
  - deci diferențele țin și de model, nu doar de metodă
- Panourile: MDI (în eșantion), MDA și media lui $|\phi_j|$ (în afara eșantionului)
  - MDA: scăderea AUC în afara eșantionului; barele roșii sînt negative
- Etichetele noi față de traiectoria LASSO
  - mom_*h*: momentum-ul pe *h* zile; dist_ma50, dist_ma200: distanța față de media mobilă
  - rsi_14: indicele RSI(14); vix_level, vix_chg_5: nivelul VIX și variația lui pe 5 zile
- MDI pune volatilitatea pe primul loc
  - MDA și SHAP indică momentum-ul pe 20 de zile și distanța față de media mobilă pe 200 de zile
- Volatilitatea pe 20 de zile are MDI ridicat, dar MDA **negativ**
  - un semn de supraajustare, nu o dovadă
- Toate importanțele sînt foarte mici
  - permutarea oricărei caracteristici schimbă AUC cu mai puțin de 0,01 (interpretarea: Anexă)
- Caracteristici corelate: MDA și SHAP **extrapolează** (Hooker, Mentch & Zhou (2021))
  - permutarea creează puncte în afara varietății datelor

MFM_ch13_feature_importance

Ferestre succesive: prognozele modelelor față de o prognoză constantă

- Ferestre succesive 2010–2026: $n = 4\,198$ de zile, dintre care 60,2% „sus”
- Reperul: prognoza constantă egală cu proporția zilelor „sus” din antrenare
  - DM, Clark–West și GW folosesc Newey–West cu 4 laguri
- Logit (include constanta): log-loss 0,6824 față de 0,6756
  - DM $t = 2{,}50$ ($p = 0{,}012$): logit-ul estimat este semnificativ *mai slab*
  - Clark–West $t = -0{,}36$ (p unilateral $= 0{,}64$): nicio dovadă că informația caracteristicilor contează
  - Giacomini–White $= 6{,}29$ (χ^2_2, $p = 0{,}043$): logit-ul este mai slab și la 5%
  - testul GW este aici aproximativ, pentru că fereastra se extinde (fereastra mobilă: Anexă)
- GBM: log-loss 0,7027, semnificativ mai slab decît prognoza constantă
  - DM $t = 5{,}81$, GW $= 33{,}6$, ambele cu $p < 0{,}001$
  - statistica Pesaran–Timmermann este $-2{,}74$
- Concluzia: niciun model nu depășește „mereu în sus” sau o prognoză constantă

MFM_ch13_inference

Anualizarea raportului Sharpe și compararea a două strategii

- Anualizarea cu $\sqrt{A}$ cere absența autocorelației; altfel $SR_A = \eta(A)\, SR$ (Lo (2002))
  - A: numărul de perioade pe an (252 de zile de tranzacționare, 365 pentru Bitcoin)
  - $\eta(A) = A/\sqrt{A + 2\sum_{k=1}^{A-1}(A-k)\rho_k}$, cu ρ_k autocorelația randamentelor de ordin k
  - autocorelațiile pozitive, $\rho_k > 0$, fac ca $\eta(A) < \sqrt{A}$
- Exemplu: randamente lunare, $A = 12$, cu $\rho_1 = 0{,}2$ și celelalte autocorelații nule
  - $\eta = 12/\sqrt{12 + 2 \cdot 11 \cdot 0{,}2} = 12/\sqrt{16{,}4} = 2{,}96$, față de $\sqrt{12} = 3{,}46$
  - deci SR anual este cu circa 15% mai mic
- Două strategii pe aceleași zile: testăm $SR_a - SR_b$ direct
  - cu metoda delta HAC sau cu un bootstrap pe blocuri studentizat (Ledoit & Wolf (2008))
  - nu cu două intervale separate

Încercări corelate în DSR: aproximări alternative (1/2)

Două aproximări, cu ipoteze diferite

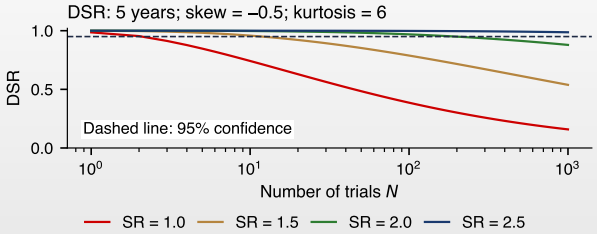
- (a) N brut, cu $\mathrm{Var}[\widehat{SR}_n]$ = dispersia SR între încercări
  - rîndul 1 din tabelul DSR pentru Bitcoin
- (b) $\mathrm{Var}[\widehat{SR}_n] = 1/T$, cu un număr efectiv de încercări independente $\hat{N} < N$
  - Seminarul 13, exercițiul B6
- Folosite împreună, dispersia *și* $\hat{N}$ pot corecta de două ori același efect al dependenței
  - modelul simplu de mai jos

- De ce funcționează (a), un model simplu: N încercări fără abilitate $\widehat{SR}_n = \sigma(\sqrt{\rho}\, Z + \sqrt{1-\rho}\, e_n)$
  - fiecare de varianță $\sigma^2 = 1/T$, cu $Z, e_1, \ldots, e_N$ independente Normal standard și corelația ρ între oricare două
  - șocul comun mișcă toate încercările împreună, deci doar $\sqrt{1-\rho}\,\sigma\, e_n$ creează selecție
  - $\mathbb{E}[\max_n \widehat{SR}_n] = \sqrt{1-\rho} \times$ valoarea pentru încercări independente
  - dispersia între încercări este $(1-\rho)\sigma^2$, deci (a) surprinde această scalare
  - cele două rînduri au raportul SD $0{,}0109/0{,}0220 = 0{,}50$, pe care modelul simplu îl citește ca $\rho \approx 1 - 0{,}50^2 \approx 0{,}75$

Încercări corelate în DSR: aproximări alternative (2/2)

- Dispersia conține și diferențe reale între reguli
  - pentru reguli reale, (a) este o aproximare
- Recomandare: raportăm rîndul conservator și una dintre variantele (a) sau (b), împreună cu N
  - numim o regulă descoperire doar dacă trece și de rîndul conservator

Raportul Sharpe deflatat în practică



- 5 ani de randamente zilnice, asimetrie $-0{,}5$, kurtosis 6
  - axa orizontală: numărul de încercări N (scară logaritmică), axa verticală: DSR
  - linia punctată: pragul de 0,95
- $N = 1$: fără selecție, $SR_0 = 0$, deci $DSR = \widehat{PSR}(0)$; pentru SR anualizat 1,0, DSR $= 0{,}986$
  - SR $= 1{,}0$ rămîne peste 0,95 la $N = 2$ (0,954) și scade sub prag de la $N = 3$ (0,913)
- SR $= 1{,}5$ rămîne semnificativ pînă la ≈ 10 încercări
  - SR $= 2{,}0$ pînă la ≈ 200, iar SR $= 2{,}5$ și la 1 000

MFM_ch13_backtest_overfitting

Căutarea repetată în aceleași date: grila Bitcoin (1/3)

- Întrebarea: depășește vreuna dintre cele N reguli reperul, după ce ținem cont că le-am încercat pe toate?
- Notațiile: $d_{k,t} = r_{k,t} - r_{0,t}$, cu media $\bar{d}_k$ pe cele n zile
  - $r_{k,t}$: randamentul regulii k în ziua t; $r_{0,t}$: randamentul reperului
- $H_0 : \max_k \mathbb{E}[d_{k,t}] \leq 0$: nicio regulă nu depășește reperul
- Distribuția sub ipoteza nulă vine dintr-un **bootstrap staționar** (Politis & Romano (1994))
  - reeșantionăm blocuri de zile întregi, de lungime aleatoare
  - $\bar{d}_k^*$: media în replica bootstrap
- Aceleași zile pentru toate regulile păstrează corelația dintre ele
  - deci nu mai este necesar un N „efectiv"
- Reality Check (RC, White (2000)): statistica $T^{RC} = \max_k \sqrt{n}\,\bar{d}_k$
  - valoarea p este proporția replicilor cu $\max_k \sqrt{n}\,(\bar{d}_k^* - \bar{d}_k) \geq T^{RC}$

Căutarea repetată în aceleași date: grila Bitcoin (2/3)

- Superior Predictive Ability (SPA, Hansen (2005)): versiunea studentizată, $T^{SPA} = \max\left(\max_k \sqrt{n}\,\bar{d}_k/\hat{\omega}_k,\ 0\right)$
  - $\hat{\omega}_k$: abaterea standard bootstrap a lui $\sqrt{n}\,\bar{d}_k$
  - replicile sînt recentrate cu $\bar{d}_k\, 1\{\sqrt{n}\,\bar{d}_k/\hat{\omega}_k \geq -\sqrt{2\log\log n}\}$
  - regulile evident slabe nu sînt recentrate, deci adăugarea unor reguli inutile nu reduce puterea testului
- StepM Romano–Wolf (Romano & Wolf (2005)): identifică *regulile* care depășesc reperul
  - cu controlul probabilității de cel puțin o eroare de tip I (FWER)
  - pasul 1: respingem $H_{0,k}$ pentru orice k cu $\sqrt{n}\,\bar{d}_k/\hat{\omega}_k > \hat{c}_1$
  - $\hat{c}_1$: cuantila de $1-\alpha$ a lui $\max_k \sqrt{n}\,(\bar{d}_k^* - \bar{d}_k)/\hat{\omega}_k$
  - pasul j: recalculăm cuantila doar pe regulile încă nerespinse
  - ne oprim cînd nu mai respingem nicio regulă

Căutarea repetată în aceleași date: grila Bitcoin (3/3)

- Configurația: 1 282 de reguli, $n = 2\,057$ de zile în eșantion
  - blocuri de 20 de zile în medie, 2 000 de replici bootstrap

Reper	cea mai mare medie studentizată	p RC	p SPA	reguli păstrate de StepM
numerar ($r_0 = 0$)	4,20	0,001	0,002	117
buy-and-hold	1,36 (regula 11/20)	0,60	0,50	0

- Față de numerar, regulile de trend au medii pozitive, care reflectă trendul propriu al Bitcoin
- Față de buy-and-hold, nicio regulă nu depășește deținerea Bitcoin, nici măcar în eșantion
  - în acord cu SR de 0,03 în afara eșantionului
- DSR are nevoie de N și de $\mathrm{Var}[\widehat{SR}]$, de unde cele două verdicte
- SPA ia ambele cantități direct din date

MFM_ch13_inference

Probabilitatea de supraajustare a backtest-ului (PBO)

Validare încrucișată combinatorică simetrică (CSCV): Bailey, Borwein, López de Prado & Zhu (2017)

1. Matricea de randamente: T perioade $\times$ N configurații; împărțim rîndurile în S blocuri contigue
2. Pentru fiecare dintre cele $\binom{S}{S/2}$ alegeri, jumătate din blocuri sînt în eșantion (IS), restul în afara lui (OOS)
3. Alegem cea mai bună configurație IS n^*
   - ordonăm crescător toate cele N configurații OOS și luăm $\bar{\omega} = \text{rang}(n^*)/(N+1) \in (0,1)$
4. Calculăm logit-ul $\lambda = \log \frac{\bar{\omega}}{1-\bar{\omega}}$ (alt λ decît penalizarea) și $PBO = \Pr[\lambda \leq 0]$
   - PBO: proporția împărțirilor în care învingătorul IS ajunge la sau sub mediana OOS

- $\bar{\omega}$: rangul relativ OOS al învingătorului IS, aproape de 1 dacă rămîne cel mai bun
  - $\lambda > 0$: peste mediana OOS
- $PBO \approx 0$: învingătorul IS rămîne sigur peste mediana OOS
  - $PBO \approx 0{,}5$ sau mai mult: a alege învingătorul IS nu este mai bun decît o alegere la întîmplare
- PBO evaluează **procedura de cercetare**, nu o singură strategie
  - nu este probabilitatea de a pierde bani
- Ca și DSR, cere înregistrarea **fiecărei** încercări (Seminarul 13, exercițiul B12)

De la condiția de evaluare la reprezentarea beta

- Trei pași (scriși necondiționat; în varianta condiționată înlocuim E cu E_t)
  1. Cu $M = 1 - F$: $E[(1-F)R_i^e] = 0$, deci $E[R_i^e] = E[R_i^e F]$ pentru orice acțiune
  2. F este el însuși un randament în exces; aplicăm pasul 1 pentru $R_i^e = F$: $E[F] = E[F^2]$
  3. Definim $\beta_i = E[R_i^e F]/E[F^2]$; atunci $E[R_i^e] = E[R_i^e F] = \beta_i E[F^2] = \beta_i E[F]$
- Acest β_i este definit cu momente de ordinul doi
  - cînd pasul 1 este satisfăcut, el coincide cu beta din regresie
  - deoarece $\mathrm{Cov}(R_i^e, F) = \beta_i E[F^2] - \beta_i E[F]^2 = \beta_i \mathrm{Var}(F)$
  - estimate pe date cu erori de evaluare, cele două pot diferi
  - folosim forma cu momente de ordinul doi, ca rețeaua pentru expuneri din lucrare (Anexă)
- Exemplu: $E[F] = E[F^2] = 0{,}01$ (1% pe lună), $E[R_i^e F] = 0{,}005$
  - $\Rightarrow \beta_i = 0{,}5$, $E[R_i^e] = 0{,}5\%$ pe lună
  - beta din regresie dă aceeași valoare, $(0{,}005 - 0{,}005 \cdot 0{,}01)/(0{,}01 - 0{,}01^2) = 0{,}5$
- De ce se schimbă funcția de pierdere
  - orice ω care evaluează toate activele trebuie să satisfacă pasul 1 pentru fiecare i
  - aceasta este o condiție de moment de ordinul întîi
  - pe aceasta o impune pierderea GMM (*Generalized Method of Moments*, metoda generalizată a momentelor) adversarială de pe slide-ul următor

Studiu de caz: deep learning în evaluarea activelor, întrebare și metodă (1/2)

- Motivația: prognoza ajustează un moment de ordinul doi acolo unde raportul semnal/zgomot este mic
  - prognoza: Gu, Kelly & Xiu (2020), rețeaua feedforward de referință, FFN
- Condiția de absență a arbitrajului țintește un moment de ordinul întîi
  - media în timp elimină o mare parte din zgomot
- Metodă: metoda generalizată a momentelor (GMM), adversarială (ec. 3, după Hansen & Jagannathan (1997))

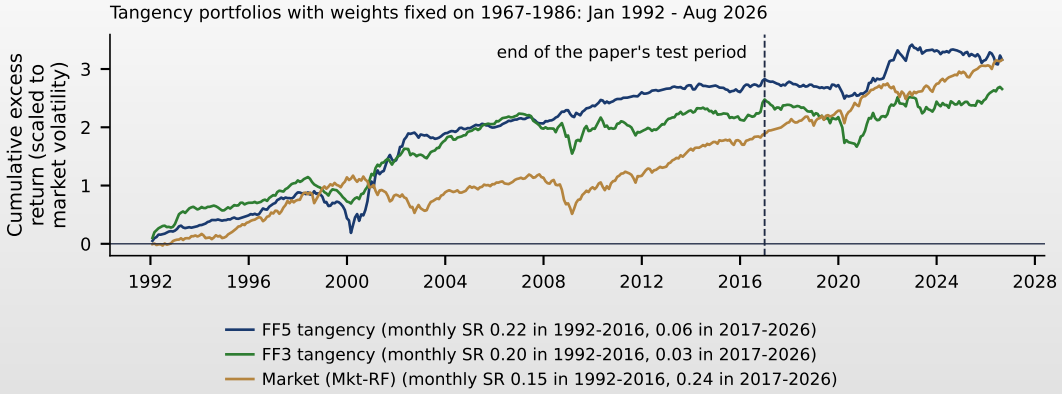
$$\min_{\omega} \max_{g} \frac{1}{N} \sum_{j=\mathbf{1}}^{N} \left(E\left[M_{t+\mathbf{1}} R^{e}_{t+\mathbf{1},j}\, g(I_t, I_{t,j}) \right] \right)^{\mathbf{2}}$$

  - I_t: informația macroeconomică; $I_{t,j}$: caracteristicile acțiunii j; N: numărul de acțiuni
  - g: funcția-instrument aleasă de un adversar astfel încît erorile de evaluare să fie cît mai mari
  - ω minimizează acest caz cel mai defavorabil

Studiu de caz: deep learning în evaluarea activelor, întrebare și metodă (2/2)

- Rețelele
  - o FFN pentru ponderile ω și un adversar g care construiește cele mai dificile active de test
  - o rețea LSTM (long short-term memory) care sintetizează istoricul macro în 4 stări
  - modelul complet se numește GAN (generative adversarial network)
- Împărțire cronologică: antrenare 1967–1986, validare 1987–1991, test 1992–2016
- Repere
  - SDF liniar fără regularizare (LS) sau cu elastic net (EN)
  - portofoliile tangente ale modelelor Fama–French cu trei și cinci factori (FF3, FF5, Fama & French (2015))
- Configurațiile sînt ordonate după raportul Sharpe pe validare
  - datele și reglarea în Anexă

Deep learning în evaluarea activelor: reperele după 2016 (1/2)



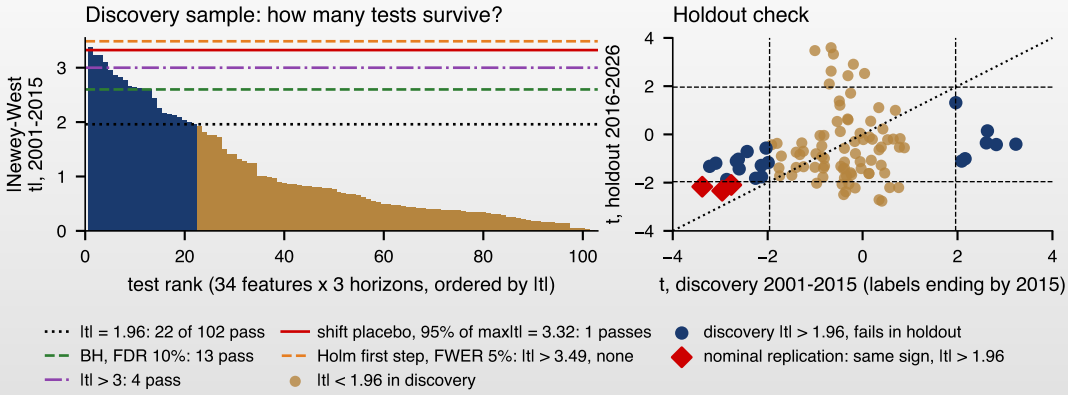
MFM_ch13_sdf_replication

Deep learning în evaluarea activelor: reperele după 2016 (2/2)

- Aceleași ponderi fixe, ianuarie 1992 – august 2026
  - fiecare serie scalată la volatilitatea randamentului în exces al pieței pe 1992–2016
  - SR lunar 2017–2026 ($T = 116$ luni, eroare standard $\approx 0{,}09$): FF5 0,06, FF3 0,03, piața 0,24
  - în grafic, Mkt-RF: randamentul în exces al pieței
  - pentru comparație, pe 1992–2016: FF5 0,22, FF3 0,20, piața 0,15
- Replicabile cu date publice: reperele Fama–French
  - nereplicabile: GAN, LS și EN la nivel de acțiune
  - panelul CRSP (Center for Research in Security Prices) nu este în datele cursului
  - scăderea de după 2016 este un semnal de alarmă, încă nu un test
  - aceeași întrebare se pune pentru un GAN fixat în 1991 (detalii în Anexă)

MFM_ch13_sdf_replication

Mini-studiu de caz: praguri și perioada de verificare (1/2)



MFM_ch13_ai_discovery

Mini-studiu de caz: praguri și perioada de verificare (2/2)

- Stînga: cele 102 valori $|t|$ din perioada de descoperire, în ordine descrescătoare
- Dreapta: statistica t din perioada de descoperire față de cea din perioada de verificare
  - cele trei replicări nominale (roșu) sînt toate inversare pe termen scurt; niciuna nu trece Holm la 5%
  - variabilele VIX semnificative cad spre zero

MFM_ch13_ai_discovery

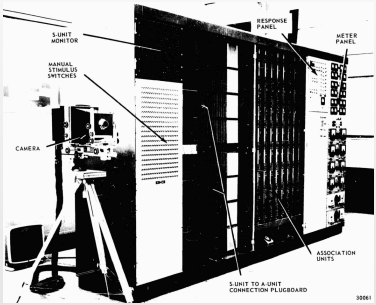
Pentru aprofundare pe Quantinar

Cursuri avansate pe quantinar.com

- Random Forests
- Shapley Values
- Shapley Curves: A Smoothing Perspective
- Machine Learning in Financial Risk
- Introduction to Reinforcement Learning
- Generative Adversarial Networks
- MSCA (*Marie Skłodowska-Curie Actions*) Digital: The Need for Explainable AI
- Business Analytics and Data Science
- Model Selection Criteria
- PAC (*Probably Approximately Correct*) Learning

înapoi

De la perceptron la Deep Blue



Mark I Perceptron, manualul operatorului (1960)

Figură: J.C. Hay, A.E. Murray (1960); domeniu public; Wikimedia Commons



IBM Deep Blue, Computer History Museum

Foto: Anton Chiang (2011); CC BY 2.0; Wikimedia Commons

- **Perceptronul** Rosenblatt (1958)
  - o regulă de învățare pentru un singur neuron artificial
  - Mark I: un dispozitiv construit pentru a recunoaște imagini simple
- 1997: Deep Blue (IBM, *International Business Machines*) îl învinge pe campionul mondial de șah Garry Kasparov (Campbell, Hoane & Hsu (2002))
- Deep Blue căuta cu reguli scrise de oameni
- ML-ul modern (acest capitol) *învață* regulile din date

înapoi

R^2 în eșantion al unui predictor fără abilitate

- Un R^2 lunar de 0,5% este mare economic
  - dar zgomotul de eșantionare are același ordin de mărime
- R^2 OLS în eșantion al unui predictor fără abilitate: $nR^2 \approx \chi^2_1$
  - termen liber, erori independente, $n \approx 300$ de luni
  - n: numărul de luni; χ^2_1: distribuția hi-pătrat cu 1 grad de libertate, cu media 1 și varianța 2
  - deci R^2 are media $1/n \approx 0{,}3\%$ și abaterea standard $\sqrt{2}/n \approx 0{,}5\%$
- R^2_{OOS} poate fi negativ
  - nu are această distribuție sub ipoteza nulă și cere un test propriu (Clark–West)

Înapoi

Exemplu rezolvat: agregarea predicțiilor arborilor

- Random forest-ul din studiul de caz, antrenat pe primele patru pliuri purjate
  - 400 de arbori, dimensiune minimă a frunzei 100
- Predicția pentru ultima observație (11 septembrie 2026)
  - P(creștere în 5 zile) dată de fiecare arbore
  - probabilitățile estimate de arborii individuali variază între 0,31 și 0,88 (abatere standard 0,11)
  - 81,5% dintre arbori votează „sus" ($p > 0{,}5$)
  - probabilitatea estimată de pădure: media celor 400 de probabilități ale arborilor $= 0{,}608$
- Predicțiile arborilor diferă mult
- Media lor este mult mai stabilă
- Stabilitatea nu înseamnă acuratețe
  - nici pădurea nu depășește „mereu în sus" (studiul de caz)

Înapoi

Eșantionarea datelor financiare: bare

Bare de timp și de tranzacții

- O **bară** = un rînd care rezumă un interval de eșantionare
  - deschidere, maxim, minim, închidere, volum
- Bare de timp: la interval fix (1 minut, 1 zi)
  - supraeșantionează perioadele liniștite și dau randamente cu cozi groase
- Bare de tranzacții (tick): o bară la fiecare N tranzacții

Bare de volum, valoare și informaționale

- O bară la fiecare V acțiuni sau la fiecare D dolari tranzacționați
  - eșantionare după **sosirea informației**
- Randamente mai apropiate de extrageri i.i.d. din distribuția Normală
  - Clark (1973); Ané & Geman (2000)
- Bare de dezechilibru / de serii
  - eșantionăm cînd fluxul ordinelor (*order flow*) devine unilateral

- Capitolul folosește bare zilnice de timp
  - date la nivel de tranzacție (platforme de tranzacționare cripto) în Capitolele 10 și 16

înapoi

Ponderile fracționare și comportamentul lor asimptotic

Definiție 6 (Operatorul de diferențiere fracționară)

- Pentru un ordin real d:

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k, \qquad \tilde{X}_t = \sum_{k=0}^{\infty} w_k X_{t-k}$$

- Ponderi: $w_0 = 1$, $w_k = -w_{k-1} \frac{d-k+1}{k}$
- Pentru $d \notin \{0, 1, 2, \dots\}$: $w_k = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$ și, pentru $k \to \infty$, $w_k \sim \frac{k^{-(1+d)}}{\Gamma(-d)}$

- $\binom{d}{k} = d(d-1)\cdots(d-k+1)/k!$: coeficientul binomial, definit pentru orice d real
- Γ: funcția gamma, cu $\Gamma(k+1) = k!$
- Ponderile scad ca $k^{-(1+d)}$, deci mai lent pentru un d mai mic
- Pentru $0 < d < 1$: $\sum_{k=0}^{\infty} w_k = (1-1)^d = 0$
  - rezultă din seria binomială în $B = 1$, convergentă pentru $d > 0$
  - $d = 0$: $w_0 = 1$ și restul zero, deci suma este 1; $d = 1$: $w = (1, -1)$, suma 0
  - pentru $d < 0$ seria ponderilor nu converge

Înapoi

Local Whittle și local Whittle exact (1/2)

- Periodograma: estimatorul empiric al densității spectrale la frecvența Fourier $\lambda_j = 2\pi j/n$

$$I(\lambda_j) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n} X_t\, e^{-\mathrm{i}\lambda_j t}\Big|^2$$

- Local Whittle (Robinson (1995)): alegem δ astfel încît $f(\lambda) \approx G\lambda^{-2\delta}$ să se potrivească pe primele m frecvențe

$$\hat{\delta} = \arg\min_{\delta}\ \log\Big(\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2\delta} I(\lambda_j)\Big) - \frac{2\delta}{m}\sum_{j=1}^{m}\log\lambda_j, \qquad \sqrt{m}\,(\hat{\delta}-\delta) \xrightarrow{d} N\big(0, \tfrac{1}{4}\big)$$

  - n: mărimea eșantionului; m: numărul de frecvențe folosite; i: unitatea imaginară
  - constanta G a fost eliminată prin optimizare
  - $\xrightarrow{d}$: convergență în distribuție, de unde eroarea standard $1/(2\sqrt{m})$

înapoi

Local Whittle și local Whittle exact (2/2)

- Local Whittle exact (ELW, Shimotsu & Phillips (2005); pentru media necunoscută, Shimotsu (2010))
  - folosește periodograma lui $(1-B)^{\delta}X_t$
  - este valid și pentru $\delta \geq \frac{1}{2}$ (serii nestaționare)
- De ce seria FFD pare să aibă $\hat{\delta} \approx 0{,}75$
  - peste frecvența $1/l^*$, filtrul trunchiat acționează ca $(1-B)^{0,25}$
  - l^*: lungimea ferestrei FFD (444 de laguri)
  - doar $n/(2\pi l^*) \approx 2$ din cele $m = 294$ de frecvențe sînt sub $1/l^*$, unde rădăcina unitară reapare
  - estimarea este memoria aparentă în banda de frecvențe, nu ordinul de integrare al seriei

MFM_ch13_ffd

înapoi

Exemplu rezolvat: purjare cu 10 observații

- Observații $t = 0, \ldots, 9$
- Eticheta lui t se termină la $t + 2$ (plafonat la 9)
- 5 pliuri contigue de cîte 2
- Embargo = 1 observație

Test	Eliminate	Antrenare
0, 1	2, 3, 4	5–9
2, 3	0, 1, 4, 5, 6	7–9
4, 5	2, 3, 6, 7, 8	0, 1, 9
6, 7	4, 5, 8, 9	0–3
8, 9	6, 7	0–5

- Pentru testul $\{4, 5\}$
  - etichetele observațiilor 2 și 3 se termină la 4 și 5, deci sînt purjate
  - observațiile 6 și 7 încep în ziua 7 sau înainte de ea, deci sînt purjate
  - ziua 7: ziua în care se termină ultima etichetă de test
  - embargoul elimină observația 8, prima care începe după ziua 7
- Cu etichete suprapuse și embargo, un pliu interior pierde $2h + e = 5$ din 10 observații

înapoi

Validitatea K-fold pentru serii de timp

- Bergmeir, Hyndman & Koo (2018): K-fold standard este valid pentru un model **pur autoregresiv**
  - $y_t = g(y_{t-1}, \ldots, y_{t-p}) + \varepsilon_t$, cu erori necorelate serial
  - g: o funcție oarecare, posibil neliniară, de ultimele p valori; ε_t: eroarea
  - erorile din pliul de test sînt necorelate cu erorile de antrenare
  - deci eroarea CV este nedeplasată pentru eroarea OOS
- Eșuează cînd **erorile** sînt corelate serial
  - etichete suprapuse: randamentele pe h zile înainte au în comun $h-1$ randamente zilnice
  - caracteristici persistente care nu fac parte din lagurile modelului (medii mobile, niveluri precum log-prețul FFD)
  - dinamică subspecificată (prea puține laguri)
- Experimentul de scurgere de informație izolează exact acest caz
  - etichete pe 20 de zile și medii mobile pe 20 de zile
- Verificați înainte de a alege: autocorelația reziduurilor CV
  - dacă este semnificativă, purjare și embargo

Înapoi

Ferestre succesive și CV combinatorică cu purjare

Ferestre succesive (WF)

- Imită implementarea în timp real, fără date viitoare în antrenare
- Dar o **singură traiectorie**
  - variantă mare și risc de supraajustare prin încercări repetate

CV combinatorică cu purjare (CPCV)

- N grupuri contigue
  - testăm pe toate cele $\binom{N}{k}$ combinații de k grupuri, cu purjare și embargo
- $\varphi = \frac{k}{N}\binom{N}{k}$: numărul de traiectorii complete de backtest
- De exemplu, $N = 6$ și $k = 2$ dau 15 împărțiri și 5 traiectorii
- O **distribuție** de rapoarte Sharpe în loc de un singur număr

- Fiecare grup este testat de φ ori, deci predicțiile de test formează φ istorii complete

înapoi

Testarea acurateței direcționale: Pesaran–Timmermann

- H_0: semnul prezis x_t este independent de semnul realizat y_t (Pesaran & Timmermann (1992))
- $\hat{P}$, $\hat{P}^*$, $\hat{p}_y$, $\hat{p}_x$ sînt definite pe slide-ul principal
  - n: numărul de prognoze

$$PT = \frac{\hat{P} - \hat{P}^*}{\sqrt{\widehat{V}(\hat{P}) - \widehat{V}(\hat{P}^*)}} \xrightarrow{d} N(0,1), \qquad \widehat{V}(\hat{P}) = \frac{\hat{P}^*(1-\hat{P}^*)}{n}$$

  - $\widehat{V}(\hat{P}^*) = \left[(2\hat{p}_y - 1)^2 \hat{p}_x(1-\hat{p}_x) + (2\hat{p}_x - 1)^2 \hat{p}_y(1-\hat{p}_y)\right]/n + 4\hat{p}_y\hat{p}_x(1-\hat{p}_y)(1-\hat{p}_x)/n^2$
  - $\widehat{V}(\cdot)$: varianța estimată sub H_0
  - $\hat{P}^*$ este și el estimat, de aici al doilea termen
- PT ține cont automat de frecvențele marginale
  - reperul nu trebuie specificat separat
- Logit-ul din studiul de caz
  - 58,1% din zile sînt „sus” ($\hat{p}_y = 0{,}581$)
  - modelul prezice „sus” în 98,3% din zile ($\hat{p}_x = 0{,}983$)
  - $PT = -0{,}80$, cu p unilateral $= 0{,}79$

MFM_ch13_inference ◂ Înapoi

AUC și incertitudinea ei de eșantionare

- Curba ROC (Receiver Operating Characteristic), pentru toate pragurile
  - rata rezultatelor adevărat pozitive $TP/(TP + FN)$
  - în funcție de rata rezultatelor fals pozitive $FP/(FP + TN)$
- Aria de sub curbă (AUC) $= P(\hat{p}_{\mathbf{poz}} > \hat{p}_{\mathbf{neg}}) + \frac{1}{2} P(\hat{p}_{\mathbf{poz}} = \hat{p}_{\mathbf{neg}}) = U/(n_1 n_0)$
  - $\hat{p}_{\mathbf{poz}}$, $\hat{p}_{\mathbf{neg}}$: probabilitățile prezise pentru o observație pozitivă (sus) și una negativă (jos)
  - cele două observații sînt extrase la întîmplare
  - n_1, n_0: numărul de observații pozitive și negative
  - U: statistica Mann–Whitney, numărul perechilor (pozitivă, negativă) ordonate corect
  - egalitățile sînt numărate cu $\frac{1}{2}$
- Eroarea standard DeLong (DeLong, DeLong & Clarke-Pearson (1988)) se calculează din valorile de plasare
  - valoarea de plasare: pentru fiecare observație, proporția observațiilor din cealaltă clasă pe care le depășește
  - presupune observații independente
- Logit, AUC agregat din predicțiile din afara pliurilor = 0,480
  - CI 95% DeLong, [0,466; 0,494], ar sugera un model „semnificativ mai slab decît întîmplarea”
  - CI 95% bootstrap pe blocuri, [0,456; 0,505], arată doar zgomot în jurul lui 0,5
- Un AUC sub 0,5 înseamnă ordonare inversă doar dacă trece un test care respectă dependența

MFM_ch13_inference
Înapoi

Giacomini–White, statistica și fereastra mobilă

- Giacomini & White (2006): ipoteza nulă privește **metoda** de prognoză
  - modelul, fereastra de estimare și regula de reestimare

$$H_0 : \mathbb{E}\big[d_{t+\tau} \mid \mathcal{F}_t\big] = 0 \;\Rightarrow\; \mathbb{E}\big[\mathrm{h}_t\, d_{t+\tau}\big] = 0, \qquad GW = n\,\bar{\mathrm{Z}}^{\top}\hat{\Omega}^{-1}\bar{\mathrm{Z}} \xrightarrow{d} \chi^2_q$$

- Notațiile
  - $d_{t+\tau}$: diferența pierderilor celor două prognoze făcute la t pentru orizontul $\tau = 5$ zile (aici τ este orizontul, nu un prag)
  - $\mathcal{F}_t$: informația disponibilă la t
  - h_t: funcții-test cunoscute la t, aici constanta 1 și diferența pierderilor întîrziată, deci $q = 2$
  - $\mathrm{Z}_t = \mathrm{h}_t d_{t+\tau}$; $\bar{\mathrm{Z}}$: media lui Z_t pe cele n prognoze; $\hat{\Omega}$: matricea de covarianță HAC a lui Z_t
- Interpretare: sub H_0, nicio informație de la t nu prezice care metodă va fi mai bună
- Teorema cere o fereastră de estimare mărginită (mobilă)
  - astfel încît zgomotul de estimare să nu dispară
  - o fereastră care se extinde iese din cadrul ei
- S&P 500, fereastră mobilă de 8 ani: GW $= 26{,}7$ ($p < 0{,}001$)
  - logit-ul prognozează mai slab decît constanta

MFM_ch13_inference

Înapoi

Contribuțiile Shapley: formula coalițiilor

Definiție 7 (Valoarea Shapley a caracteristicii j (SHAP, Lundberg & Lee (2017)))

- Contribuția marginală medie a lui j peste toate coalițiile $S \subseteq \{1, \ldots, p\} \setminus \{j\}$:

$$\phi_j = \sum_S \frac{|S|!\,(p - |S| - 1)!}{p!} \big[v(S \cup \{j\}) - v(S)\big], \qquad \hat{f}(\mathrm{x}) = \phi_{\mathbf{0}} + \sum_j \phi_j$$

- SHAP intervențional: $v(S) = \mathbb{E}[\hat{f}(\mathrm{x}_S, \mathrm{X}_{\bar{S}})]$, media lui $\hat{f}$ cu caracteristicile din afara lui S extrase dintr-un eșantion de referință
- $\phi_{\mathbf{0}} = v(\emptyset)$ este predicția medie
- Notații
  - p: numărul de caracteristici; $\hat{f}(\mathrm{x})$: predicția modelului pentru x
  - S: o coaliție, adică o submulțime a celorlalte caracteristici; $|S|$: mărimea ei; $\bar{S}$: caracteristicile din afara ei
  - $v(S)$: predicția cînd sînt cunoscute doar caracteristicile din S
- Ponderea este probabilitatea ca exact S să preceadă j într-o permutare aleatoare a celor p caracteristici
- SHAP condiționat folosește $v(S) = \mathbb{E}[\hat{f}(\mathrm{X}) \mid \mathrm{X}_S = \mathrm{x}_S]$
- Cele două variante diferă cînd caracteristicile sînt dependente

Shapley Values

înapoi

Interpretarea importanței caracteristicilor

- Soluții pentru caracteristicile corelate
  - permutați împreună coloanele unui grup, cu aceeași permutare a rîndurilor, sau reestimați fără caracteristică
  - permutările condiționate și cele pe blocuri temporale sînt variante separate
- Importanțele sînt estimări și cer o analiză proprie a incertitudinii
- Arborii „onești" dau *predicții* de pădure asimptotic normale și CI pentru ele (Wager & Athey (2018))
  - dar nu CI pentru importanțe
- **Backtest-ul nu este un instrument de cercetare**, analiza importanței este (López de Prado (2018))

înapoi

Selecția dublă și DML, penalizarea și condițiile (1/2)

- Model parțial liniar: $\mathbb{E}[\varepsilon_t \mid d_t, x_t] = 0$, $\mathbb{E}[v_t \mid x_t] = 0$, $\mathrm{Var}(v_t) > 0$
  - g și m sînt aproximativ rare (sparse)
  - erorile nu pot fi prezise din regresori
  - d_t are o variație proprie, dincolo de controale
- Selecția dublă: eroarea standard robustă la heteroscedasticitate este validă (Belloni, Chernozhukov & Hansen (2014))
  - pentru eșantionare independentă, condiții de regularitate și raritate $s^2 \log^2(p \vee n)/n \to 0$
  - s: numărul controalelor care contează; p: numărul total de controale
  - $p \vee n = \max(p, n)$
- Funcția obiectiv LASSO, cu ponderi de penalizare specifice fiecărei caracteristici

$$\frac{1}{n}\sum_t (y_t - x_t^\top \beta)^2 + \frac{\lambda}{n}\sum_j \hat{\psi}_j |\beta_j|, \qquad \lambda = 2c\sqrt{n}\,\Phi^{-1}\big(1 - \gamma/(2p)\big)$$

  - constantele plug-in: $c = 1{,}1$ și $\gamma = 0{,}1/\log n$
  - ponderile: $\hat{\psi}_j^2 = \frac{1}{n}\sum_t x_{tj}^2 \hat{e}_t^2$, cu $\hat{e}_t$ reziduurile pasului anterior
  - calculul se iterează

înapoi

Selecția dublă și DML, penalizarea și condițiile (2/2)

- DML (Chernozhukov et al. (2018)): cu $\ell(x) = \mathbb{E}[y \mid x]$, este suficient ca

$$\|\hat{m} - m\| \left(\|\hat{\ell} - \ell\| + \|\hat{m} - m\|\right) = o_P(n^{-1/2})$$

  - alături de consistență și condiții de regularitate
- Notațiile
  - $\|\hat{m} - m\|$: rădăcina erorii pătratice medii de estimare a algoritmului pentru m (analog pentru ℓ)
  - $o_P(n^{-1/2})$: tinde la zero în probabilitate mai repede decît $1/\sqrt{n}$
  - de exemplu, este suficient ca ambele erori să fie $o_P(n^{-1/4})$
  - fiecare algoritm poate converge mai lent decît un model parametric, deoarece contează doar produsul erorilor

înapoi

Machine learning pentru inferență în evaluarea activelor

- Feng, Giglio & Xiu (2020): adaugă un factor nou informație de evaluare, date fiind sutele de factori existenți?
  - selecție dublă a factorilor de control din „factor zoo", apoi un test t valid pentru prețul de risc al factorului nou
  - răspunsul inferențial la Harvey, Liu & Zhu (2016), care ridică pragul la $t > 3$
- Efecte eterogene: pădurile cauzale „oneste" (Wager & Athey (2018))
  - dau CI pentru efecte care variază cu covariatele
- Trei reguli
  - predicția și inferența au ținte diferite: un predictor bun poate da un $\hat{\theta}$ deplasat
  - ortogonalizați înainte de a testa și faceți cross-fitting cu pliuri care respectă timpul
  - raportați intervale de încredere: „selectat" sau „important" nu înseamnă „semnificativ"

Înapoi

Deep learning pentru secvențe

Rețele recurente (LSTM; GRU, gated recurrent unit)

- Starea ascunsă poartă memoria: $\mathrm{h}_t = g(\mathrm{h}_{t-1}, \mathrm{x}_t)$
  - g: o funcție neliniară cu ponderi antrenate; x_t: intrarea de la t
- Porțile controlează ce se păstrează sau se uită
  - atenuează dispariția gradienților
- Folosite pentru prognoza volatilității
  - și pentru prognoze bazate pe registrul de ordine și prognoze intrazilnice

Transformer

- Auto-atenție: $\mathrm{softmax}\left(QK^\top/\sqrt{d_k}\right)V$
  - matricele de interogare, cheie și valoare Q, K, V (transformări liniare ale intrărilor)
  - d_k: dimensiunea cheilor
  - softmax transformă fiecare rînd de scoruri în ponderi cu suma 1
- Dependență pe termen lung, antrenare paralelă
- Baza modelelor lingvistice mari (Capitolul 15)

Înapoi

Modele fundaționale pentru serii de timp

Modele de prognoză pre-antrenate

- Pre-antrenate pe colecții mari de serii, folosite **zero-shot** (fără reantrenare)
  - Chronos (Ansari et al., 2024), TimesFM (Das et al., 2024)
  - Moirai (Woo et al., 2024), Lag-Llama (Rasul et al., 2023)
- Repere utile pentru prognoza volatilității și a volumului
- Pentru randamente moștenesc aceleași limite de semnal/zgomot
- Datele de pre-antrenare se pot suprapune cu perioada de test
  - o nouă formă de scurgere de informație

- Tratate pe larg în Capitolul 14
  - cu instrumentele de evaluare a prognozelor de risc din Capitolul 8

Înapoi

Reinforcement learning, modele generative, AI explicabilă, risc (1/2)

Reinforcement learning (RL)

- O politică $\pi(a \mid s)$ alege acțiuni a în stări s
  - scopul: maximizarea recompensei cumulate
- Execuție optimă, activitatea formatorilor de piață (*market making*), deep hedging: Buehler et al. (2019)

Modele generative

- GAN, autoencodere variaționale (VAE), modele de difuzie
  - scenarii sintetice pentru teste de stres

AI explicabilă

- Cerută de autorități (EU AI Act, scoring de credit, riscul de model)
- SHAP, LIME, dependență parțială
  - LIME: explicații locale interpretabile, independente de model

înapoi

Reinforcement learning, modele generative, AI explicabilă, risc (2/2)

ML pentru risc

- Value-at-Risk (VaR): cuantile condiționate
  - păduri de regresie cuantilică, boosting cu pierdere cuantilică
- Expected Shortfall (ES)
  - media cuantilelor estimate din coadă sau o pierdere comună VaR–ES
- Volatilitatea: o țintă pe randamentele la pătrat (Capitolele 5 și 9)

- **Quantinar:** Reinforcement Learning GANs Explainable AI ML in Financial Risk

înapoi

Deep learning în evaluarea activelor, datele (Secțiunea III.A)

- Randamente lunare ale tuturor acțiunilor din CRSP, ianuarie 1967 – decembrie 2016
  - în exces față de dobînda la titlurile de stat pe o lună
- 46 de caracteristici ale firmelor, în șase grupe (Tabelul A.II)
  - randamente trecute, investiții, profitabilitate, active necorporale, valoare, fricțiuni de tranzacționare
  - sînt păstrate doar lunile cu toate cele 46 de caracteristici, adică circa 10 000 din 31 000 de acțiuni
  - caracteristicile devin ranguri lunare normalizate transversal
- 178 de serii macroeconomice, transformate în serii staționare
  - 124 de serii din FRED-MD, baza lunară de date macroeconomice a Rezervei Federale din St. Louis (McCracken & Ng (2016))
  - cele 46 de mediane lunare ale caracteristicilor
  - 8 predictori din Welch & Goyal (2008)

Înapoi

Deep learning în evaluarea activelor, metoda (Secțiunile I–II, Anexa C) (1/2)

- Obiectivul (ec. 3) și rețelele sînt pe slide-ul principal
  - adversarul g folosește 8 instrumente
- Antrenarea are trei pași (Secțiunea II.D)
  - SDF necondiționat, maximizare după g, apoi din nou minimizare
- Hiperparametri: 384 de configurații, ordonate după raportul Sharpe (SR) pe validare
  - configurația aleasă: 2 straturi de cîte 64 de unități, dropout, optimizatorul Adam
  - ansamblu de 9 estimări (Tabelul A.VIII)
  - coeficienții beta: o a doua FFN estimează $E_t[R^e_{t+\mathbf{1},i} F_{t+\mathbf{1}}] \propto \beta_{t,i}$
- Reperul liniar LS/EN: SDF $\omega_{t,i} = \theta^\top I_{t,i}$
  - pe componentele long și short ale fiecărei caracteristici
  - $\omega_{t,i}$: ponderea acțiunii i; $I_{t,i}$: caracteristicile ei
  - θ: cîte un coeficient pentru fiecare caracteristică, același pentru toate acțiunile și lunile

Înapoi

Deep learning în evaluarea activelor, metoda (Secțiunile I–II, Anexa C) (2/2)

- Măsuri (Secțiunea II.F): SR al lui F și două măsuri de potrivire
  - $\hat{\epsilon}_{t+1,i}$: eroarea de evaluare, partea din $R^e_{t+1,i}$ neexplicată de $\hat{\beta}_{t,i} F_{t+1}$
- Variația explicată (EV)

$$\text{EV} = 1 - \frac{\overline{\hat{\epsilon}^2}}{\overline{(R^e)^2}}$$

  - unu minus raportul dintre eroarea de evaluare pătratică medie și randamentul în exces pătratic mediu
  - bara de deasupra: media în timp a mediilor transversale lunare
  - EV $= 1$: factorul explică toată variația
- R^2 transversal: cît de bine explică modelul randamentele medii

$$R^2_{\text{transv}} = 1 - \frac{\frac{1}{N}\sum_i \frac{T_i}{T}\left(\bar{\hat{\epsilon}}_i\right)^2}{\frac{1}{N}\sum_i \frac{T_i}{T}\left(\bar{R}^e_i\right)^2}$$

  - $\bar{\hat{\epsilon}}_i$, $\bar{R}^e_i$: mediile în timp pentru acțiunea i
  - T_i: numărul de luni cu date pentru acțiunea i; T: numărul de luni din eșantion

Înapoi

Deep learning în evaluarea activelor, rezultate (1/2)

- Sursa: Chen, Pelger & Zhu (2024), *Management Science*
- SR al SDF în afara eșantionului (Tabelul I), anualizat: GAN 2,6, EN 1,7, FFN 1,5
  - valorile lunare pe slide-ul principal
- Acțiuni individuale (Tabelul I), GAN față de celelalte modele
  - variația explicată: 8% față de 3–4%
  - R^2 transversal: 23% față de 14–19%
  - Interpretare: criteriul bazat pe absența arbitrajului îmbunătățește ambii indicatori, deși nu îi maximizează direct
- Portofolii pe decilele coeficienților β
  - Tabelul II, perioada de test, valori anualizate
  - diferența de randament dintre decila superioară și cea inferioară: 0,39
  - alfa față de FF5: 0,39, cu $t = 9{,}96$
  - testul GRS respinge CAPM (modelul de evaluare a activelor financiare), FF3 și FF5 ($p = 0{,}00$)
  - testul GRS (Gibbons–Ross–Shanken): H_0 spune că toate alfa ale portofoliilor pe decile sînt zero
  - Interpretare: coeficienții beta au o relație liniară cu randamentele medii
  - $R^2 = 0{,}97$ al randamentului mediu pe β (Figura 8)
  - modelele factoriale nu le explică randamentele

Înapoi

Deep learning în evaluarea activelor, rezultate (2/2)

- 46 de seturi de decile sortate după caracteristici (Tabelul III)
  - R^2 transversal al GAN peste 90% pentru toate seturile
  - variația explicată a decilelor de inversare pe termen scurt: 0,70 pentru GAN, față de 0,43 pentru EN
  - variația explicată a decilelor de momentum: 0,54 față de 0,26
  - Interpretare: cîștigurile vin din decilele extreme și din interacțiuni
- Stări macro (Figura 6)
  - includerea directă a celor 178 de serii reduce puternic SR în afara eșantionului
  - SR al GAN scade cu circa 10% fără stările macro și cu circa 20% dacă folosește doar momente necondiționate
  - Interpretare: LSTM surprinde dinamica macroeconomică pe care ultima variație observată nu o poate reda
- Acțiuni cu capitalizare mare (Tabelul IV, SR lunar)
  - modele estimate pe toate acțiunile, evaluate pe subeșantion
  - cele mai mari circa 1 500 de acțiuni: GAN 0,41, FFN 0,30, EN 0,15
  - cele mai mari circa 550 de acțiuni: GAN 0,26, FFN 0,24, EN 0,23
  - modele reestimate doar pe cele 1 500 de acțiuni: GAN 0,42, FFN 0,28, EN 0,42
  - Interpretare: avantajul GAN scade puternic pentru acțiunile cu cea mai mare capitalizare
  - EN recuperează diferența cînd este reestimat pe aceste acțiuni

Înapoi

Deep learning în evaluarea activelor, replicare și extindere

- Datele: Kenneth French Data Library și FRED-MD
  - pe împărțirea publicată
- Replicare și extindere
  - Replicabile: ponderile tangente FF3 și FF5 estimate pe 1967–1986 (nota 23), SR lunar pe 1992–2016, alături de Tabelul A.VI
  - Nereplicabile cu datele cursului: LS și EN publicate (Tabelele I și III) folosesc componente long și short ponderate după rang pentru 46 de caracteristici și coeficienți beta la nivel de acțiune
  - Analog pe date de portofoliu, o extindere, nu o replicare
  - SDF liniar ($\hat{\theta}_{LS}$ în formă închisă, $\hat{\theta}_{EN}$ cu obiectivul din Secțiunea II.F) pe jumătățile (decilele 1–5, 6–10) celor 15 caracteristici pe care biblioteca le sortează în decile
  - nivelurile nu sînt comparabile cu Tabelele I și III

Înapoi

Deep learning în evaluarea activelor, limite și o extindere deschisă

- Limite declarate de autori
  - Pozitivitatea SDF nu este impusă (nota 9)
  - Sînt păstrate doar lunile cu toate caracteristicile, ceea ce elimină mai ales acțiunile foarte mici (Secțiunea III.A)
  - Fricțiuni (Secțiunea III.I, Figura 17): SR anual al GAN este 1,73 fără cele mai mici 40% dintre acțiuni, 2,07 fără cele 40% cu cele mai mari spread-uri bid–ask și 1,87 fără cele 40% cu turnover-ul cel mai mic (vezi și Avramov, Cheng & Metzker (2023))
  - Valorile indicatorilor sînt mai mari în eșantion decît în afara lui, semn de o anumită supraajustare (Secțiunea III.C)
  - Stările ascunse nu sînt comparabile între estimări, deci nu pot fi mediate (Secțiunea II.E)
- Extindere deschisă: păstrați specificația aleasă pe 1967–1991
  - evaluați-o pe 2017–2026, după eșantionul lucrării
  - mai întrece SDF-ul liniar fără arbitraj modelul FF5?
  - adaugă ceva stările LSTM extrase din FRED-MD?

Înapoi

Deep learning în evaluarea activelor, interpretarea celor două grafice

- GAN 0,75 față de FF5 0,22
  - cu erorile standard Lo 0,065 și 0,058, SE a diferenței este cel mult 0,12, oricare ar fi corelația
  - diferența depășește 4 SE (randamente i.i.d. cu distribuția Normală)
  - un test pe perechi, cu randamente aliniate (Ledoit & Wolf (2008)) este mai precis
- Ponderile fixe, scalate astfel încît $\sum |w| = 1$
  - FF5 se concentrează pe profitabilitate și investiții: factorul RMW (robust minus weak) are ponderea 0,42, iar CMA (conservative minus aggressive) 0,30
  - FF3 alocă 0,61 factorului HML (high minus low, valoare)
  - ambele combinații au fost alese pe 1967–1986

MFM_ch13_sdf_replication

înapoi

Bucla descoperirii, pas cu pas (1/2)

- **Literatura**
  - „Rezumă dovezile despre cît din predictibilitatea publicată a randamentelor rezistă out-of-sample și după publicare: DOI sau id arXiv, eșantion, estimarea principală a scăderii."
  - Verificați fiecare identificator
  - Pentru căutare: Semantic Scholar, Google Scholar, Elicit
- **Ipoteze**
  - „Pentru o variabilă care prognozează randamentele S&P 500 în 2001-2015, dă explicații concurente (primă de risc, reversal, microstructură, data mining) și ce prezice fiecare pentru 2016-2026 și pentru alți indici."
- **Date și cod**
  - „Scrie cod Python care construiește 34 de variabile tehnice din închiderile zilnice S&P 500 și VIX ale cursului doar cu informația pînă în ziua t și raportează statisticile t Newey-West pentru randamentele la 1, 5 și 20 de zile în 2001-2015."
  - Orice asistent (Claude, ChatGPT, Gemini, GitHub Copilot)

Înapoi

Bucla descoperirii, pas cu pas (2/2)

- **Robustețea**, fixată dinainte
  - „Scrie planul de analiză: lista completă de variabile și orizonturi, perioada de verificare, corecțiile (Holm, BH, max-t) și rezultatul care ar conta drept eșec."
  - Placebo: deplasarea circulară a țintei (cu 250–3 000 de zile) rupe alinierea cu variabilele
  - deplasarea păstrează dependența proprie a fiecărei serii
  - staționaritatea țintei nu ajunge pentru nulul comun al tuturor pantelor: deplasarea schimbă și relația variabilelor cu volatilitatea condiționată
  - deci un reper euristic; o alternativă: bootstrap comun al scorurilor, centrat sub nul, cu ipoteze explicite
- **Critica**
  - „Joacă rolul unui recenzent: care dintre aceste variabile semnificative sînt un singur efect numărat de mai multe ori și care sînt anomalii cunoscute, nu noi?"
- **Raportul**
  - toate cele M teste, toate încercările, inclusiv cele abandonate
  - DSR și PBO pentru orice strategie construită pe ele

Înapoi

Acronime folosite în acest material (1/5)

- **ADF**: Augmented Dickey–Fuller (test) — *testul Dickey–Fuller augmentat*
- **AI**: Artificial Intelligence — *inteligență artificială*
- **AIC**: Akaike Information Criterion — *criteriul informațional Akaike*
- **AUC**: Area Under the (ROC) Curve — *aria de sub curba ROC*
- **BET**: Bucharest Exchange Trading (index) — *indicele de referință al Bursei de Valori București*
- **BH**: Benjamini–Hochberg (procedure) — *procedura Benjamini–Hochberg*
- **C-ENet**: Combination Elastic Net — *media prognozelor univariate selectate prin elastic net*
- **CAPM**: Capital Asset Pricing Model — *modelul de evaluare a activelor financiare*
- **CBOE**: Chicago Board Options Exchange — *bursa de opțiuni din Chicago*
- **CF**: Cash Flow (in CF/P: cash flow to price) — *fluxul de numerar (în CF/P: flux de numerar raportat la preț)*
- **CI**: Confidence Interval — *interval de încredere*
- **CMA**: Conservative Minus Aggressive (investment factor) — *factorul investiții: firme conservatoare minus agresive*
- **CPCV**: Combinatorial Purged Cross-Validation — *validare încrucișată combinatorie cu purjare*
- **CRSP**: Center for Research in Security Prices — *Centrul de cercetare a prețurilor titlurilor (Universitatea din Chicago)*
- **CSCV**: Combinatorially Symmetric Cross-Validation — *validare încrucișată combinatorie simetrică*
- **CV**: Cross-Validation — *validare încrucișată*
- **CW**: Clark–West (test for nested models) — *testul Clark–West pentru modele imbricate*

Acronime folosite în acest material (2/5)

- **DM**: Diebold–Mariano (test) — *testul Diebold–Mariano de comparare a prognozelor*
- **DML**: Double/Debiased Machine Learning — *machine learning dublu, fără deplasare (debiased)*
- **DSR**: Deflated Sharpe Ratio — *raportul Sharpe deflatat*
- **ELW**: Exact Local Whittle (estimator) — *estimatorul local Whittle exact*
- **ENet**: Elastic Net — *regresie cu penalizare combinată L1 și L2*
- **ES**: Expected Shortfall — *pierderea așteptată în coadă*
- **EU**: European Union — *Uniunea Europeană*
- **EV**: Explained Variation — *variația explicată*
- **EWMA**: Exponentially Weighted Moving Average — *medie mobilă ponderată exponențial*
- **FDR**: False Discovery Rate — *rata descoperirilor false*
- **FF3**: Fama–French three-factor model — *modelul Fama–French cu trei factori*
- **FF5**: Fama–French five-factor model — *modelul Fama–French cu cinci factori*
- **FFD**: Fixed-width window Fractional Differentiation — *diferențiere fracționară cu fereastră fixă*
- **FFN**: FeedForward Network — *rețea neuronală feedforward (fără bucle)*
- **FN**: False Negative — *fals negativ*
- **FP**: False Positive — *fals pozitiv*
- **FRED-MD**: Federal Reserve Economic Data, Monthly Database (McCracken–Ng) — *baza lunară de date macroeconomice a Rezervei Federale din St. Louis*
- **FWER**: Family-Wise Error Rate — *probabilitatea de cel puțin o eroare de tip I în familia de teste*

Acronime folosite în acest material (3/5)

- **GAN**: Generative Adversarial Network — *rețea generativă adversarială*
- **GB**: Gradient Boosting — *boosting pe gradient; notația seminarului pentru GBM*
- **GBM**: Gradient Boosting Machine — *model de boosting pe gradient (în acest capitol nu înseamnă mișcarea browniană geometrică)*
- **GMM**: Generalized Method of Moments — *metoda generalizată a momentelor*
- **GRS**: Gibbons–Ross–Shanken (test) — *testul Gibbons–Ross–Shanken*
- **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*
- **HARKing**: Hypothesising After the Results are Known — *formularea ipotezelor după ce se cunosc rezultatele*
- **HML**: High Minus Low (value factor) — *factorul valoare: B/M mare minus mic*
- **IBM**: International Business Machines (Corporation) — *compania International Business Machines*
- **IS**: In-Sample — *pe datele folosite la estimare*
- **LASSO**: Least Absolute Shrinkage and Selection Operator — *regresie penalizată L1 cu selecția variabilelor*
- **LIME**: Local Interpretable Model-agnostic Explanations — *explicații locale interpretabile, independente de model*
- **LLM**: Large Language Model — *model lingvistic mare*
- **LS**: Linear SDF estimated by unregularised Least Squares — *SDF liniar estimat prin cele mai mici pătrate, fără regularizare*

Acronime folosite în acest material (4/5)

- **LSTM**: Long Short-Term Memory (network) — *rețea cu memorie pe termen lung și scurt*
- **LW**: Local Whittle (estimator) — *estimatorul local Whittle*
- **MA**: Moving Average — *medie mobilă*
- **MA20**: 20-day Moving Average — *medie mobilă pe 20 de zile*
- **MDA**: Mean Decrease Accuracy — *scăderea medie a acurateței (importanța prin permutare)*
- **MDI**: Mean Decrease Impurity — *scăderea medie a impurității*
- **ML**: Machine Learning — *învățare automată*
- **MLP**: MultiLayer Perceptron — *perceptron multistrat*
- **MSCA**: Marie Skłodowska-Curie Actions — *programul european Acțiunile Marie Skłodowska-Curie*
- **OLS**: Ordinary Least Squares — *metoda celor mai mici pătrate*
- **OOS**: Out-Of-Sample — *pe date nefolosite la estimare*
- **PAC**: Probably Approximately Correct (learning) — *învățare probabil aproximativ corectă*
- **PBO**: Probability of Backtest Overfitting — *probabilitatea de supraajustare a backtest-ului*
- **PCR**: Principal Component Regression — *regresia pe componente principale*
- **PLS**: Partial Least Squares — *regresia prin cele mai mici pătrate parțiale*
- **PSR**: Probabilistic Sharpe Ratio — *raportul Sharpe probabilistic*
- **PT**: Pesaran–Timmermann (test of directional accuracy) — *testul Pesaran–Timmermann al acurateței direcționale*

Acronime folosite în acest material (5/5)

- **RC**: Reality Check (White's test for data snooping) — *testul Reality Check al lui White pentru căutarea repetată în aceleași date*
- **ReLU**: Rectified Linear Unit — *funcția* $\max(0, z)$, *unitatea liniară rectificată*
- **RF**: Random Forest — *pădure aleatoare*
- **RL**: Reinforcement Learning
- **RMW**: Robust Minus Weak (profitability factor) — *factorul profitabilitate: firme robuste minus slabe*
- **ROC**: Receiver Operating Characteristic (curve) — *curba caracteristică de operare a receptorului*
- **RSI**: Relative Strength Index — *indicele puterii relative*
- **SD**: Standard Deviation — *abatere standard*
- **SDF**: Stochastic Discount Factor — *factorul stocastic de actualizare*
- **SE**: Standard Error — *eroare standard*
- **SHAP**: SHapley Additive exPlanations — *explicații aditive Shapley*
- **SMB**: Small Minus Big (size factor) — *factorul mărime: firme mici minus mari*
- **SPA**: Superior Predictive Ability (test) — *testul abilității predictive superioare*
- **SR**: Sharpe Ratio — *raportul Sharpe*
- **TN**: True Negative — *adevărat negativ*
- **TP**: True Positive — *adevărat pozitiv*
- **VAE**: Variational AutoEncoder — *autoencoder variațional*
- **VaR**: Value-at-Risk — *valoarea la risc*
- **VIX**: CBOE Volatility IndeX — *indicele de volatilitate CBOE*
- **WF**: Walk-Forward (validation) — *evaluare pe ferestre succesive*

Bibliografie (1/9)

- Ané, T., Geman, H. (2000). Order flow, transaction clock, and normality of asset returns. *Journal of Finance*, 55(5), 2259–2284.
- Ansari, A.F. et al. (2024). Chronos: learning the language of time series. *arXiv:2403.07815*.
- Avramov, D., Cheng, S., Metzker, L. (2023). Machine learning vs. economic restrictions: evidence from stock return predictability. *Management Science*, 69(5), 2587–2619.
- Bailey, D.H., Borwein, J., López de Prado, M., Zhu, Q.J. (2017). The probability of backtest overfitting. *Journal of Computational Finance*, 20(4), 39–69.
- Bailey, D.H., López de Prado, M. (2012). The Sharpe ratio efficient frontier. *Journal of Risk*, 15(2), 3–44.
- Bailey, D.H., López de Prado, M. (2014). The deflated Sharpe ratio: correcting for selection bias, backtest overfitting and non-normality. *Journal of Portfolio Management*, 40(5), 94–107.
- 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.

Bibliografie (2/9)

- Belloni, A., Chernozhukov, V., Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2), 608–650.
- Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57(1), 289–300.
- Bergmeir, C., Hyndman, R.J., Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Buehler, H., Gonon, L., Teichmann, J., Wood, B. (2019). Deep hedging. *Quantitative Finance*, 19(8), 1271–1291.
- 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.
- Campbell, M., Hoane, A.J., Hsu, F. (2002). Deep Blue. *Artificial Intelligence*, 134(1–2), 57–83.

Bibliografie (3/9)

- Chen, A.Y., Lopez-Lira, A., Zimmermann, T. (2022). Does peer-reviewed research help predict stock returns? *arXiv:2212.10317*.
- Chen, L., Pelger, M., Zhu, J. (2024). Deep learning in asset pricing. *Management Science*, 70(2), 714–750.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1), C1–C68.
- Clark, P.K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1), 135–155.
- Clark, T.E., West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291–311.
- Das, A., Kong, W., Sen, R., Zhou, Y. (2024). A decoder-only foundation model for time-series forecasting. *Proceedings of the 41st International Conference on Machine Learning (ICML)*, PMLR 235.
- DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3), 837–845.

Bibliografie (4/9)

- Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Dixon, M.F., Halperin, I., Bilokon, P. (2020). *Machine Learning in Finance: From Theory to Practice*. Springer.
- Dong, X., Li, Y., Rapach, D.E., Zhou, G. (2022). Anomalies and the expected market return. *Journal of Finance*, 77(1), 639–681.
- Fama, E.F., French, K.R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22.
- Feng, G., Giglio, S., Xiu, D. (2020). Taming the factor zoo: a test of new factors. *Journal of Finance*, 75(3), 1327–1370.
- Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
- Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578.
- Gu, S., Kelly, B., Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223–2273.

Bibliografie (5/9)

- Hansen, L.P., Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance*, 52(2), 557–590.
- Hansen, P.R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4), 365–380.
- Harvey, C.R., Liu, Y., Zhu, H. (2016). . . . and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5–68.
- Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning*. 2nd ed. Springer.
- Hoerl, A.E., Kennard, R.W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.
- Hooker, G., Mentch, L., Zhou, S. (2021). Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. *Statistics and Computing*, 31(6), 82.
- Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, 68(1), 165–176.
- Kelly, B., Malamud, S., Zhou, K. (2024). The virtue of complexity in return prediction. *Journal of Finance*, 79(1), 459–503.

Bibliografie (6/9)

- Kelly, B., Xiu, D. (2023). Financial machine learning. *Foundations and Trends in Finance*, 13(3–4), 205–363.
- Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3), 1217–1241.
- Ledoit, O., Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15(5), 850–859.
- Lo, A.W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4), 36–52.
- Lo, A.W. (2004). The adaptive markets hypothesis. *Journal of Portfolio Management*, 30(5), 15–29.
- López de Prado, M. (2018). *Advances in Financial Machine Learning*. Wiley.
- López de Prado, M. (2020). *Machine Learning for Asset Managers*. Cambridge University Press.

Bibliografie (7/9)

- Lundberg, S.M., Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- McCracken, M.W., Ng, S. (2016). FRED-MD: a monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4), 574–589.
- Nagel, S. (2025). Seemingly virtuous complexity in return prediction. NBER Working Paper 34104.
- Newey, W.K., West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Novy-Marx, R., Velikov, M. (2025). AI-powered (finance) scholarship. NBER Working Paper 33363.
- Opdyke, J.D. (2007). Comparing Sharpe ratios: so where are the p-values? *Journal of Asset Management*, 8(5), 308–336.
- Pesaran, M.H., Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4), 461–465.
- Politis, D.N., Romano, J.P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428), 1303–1313.

Bibliografie (8/9)

- Rasul, K. et al. (2023). Lag-Llama: towards foundation models for probabilistic time series forecasting. *arXiv:2310.08278*.
- Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23(5), 1630–1661.
- Romano, J.P., Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4), 1237–1282.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408.
- Shimotsu, K. (2010). Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory*, 26(2), 501–540.
- Shimotsu, K., Phillips, P.C.B. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics*, 33(4), 1890–1933.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1), 267–288.

Bibliografie (9/9)

- Wager, S., Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228–1242.
- 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. et al. (2024). Unified training of universal time series forecasting transformers. *arXiv:2402.02592*.
- Zou, H., Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B*, 67(2), 301–320.
- **Cod și date:** github.com/danpele/MFM/tree/main/Quantlets/Ch_13