








Modelarea piețelor financiare

Capitolul 8: Backtesting și evaluarea prognozelor de risc



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

Poate fi infirmată o prognoză de risc?

> *"The elicitability of a functional opens the possibility to compare competing forecasts and to rank them in terms of their realized scores."*
>
> — *Fissler & Ziegel (2016)*

- GARCH-EVT (volatilitate GARCH cu o coadă din teoria valorilor extreme): VaR (*Value-at-Risk*, valoarea expusă la risc) 1% este depășit în 1,15% din zile pe S&P 500 (*Standard & Poor's*) și în 0,97% pe BET (țința: 1%)
  - este suficientă o rată apropiată de 1% ca să numim modelul bun?
- Cu $T = 250$ de zile și o rată adevărată de depășire de 2%, testul POF al lui Kupiec (*proportion of failures*, proporția depășirilor) respinge doar în 24% din cazuri
  - o rată de depășire dublă nu este respinsă în trei ani din patru
- Testul DQ (*dynamic quantile*, cuantila dinamică) respinge toate modelele pe S&P 500, inclusiv simularea istorică filtrată (FHS)
- ES (Expected Shortfall) nu poate fi ordonat singur, dar perechea (VaR, ES) poate, cu scorul consistent FZ0 (funcția de pierdere Fissler–Ziegel omogenă de grad zero)
  - FHS are cea mai mică pierdere FZ0 medie pe S&P 500 și pe BET

▸ Graficul central: Puterea testului POF

Cum aflăm dacă o prognoză de risc este utilă?

- **Calibrare**
  - este o prognoză VaR (Value-at-Risk) sau ES (Expected Shortfall) corectă, zi de zi?
  - poate un backtest să o arate?
- **Comparație**
  - care dintre mai multe prognoze este mai bună?
  - este diferența semnificativă?
- **Decizie**: ce putem concluziona cînd coada conține doar cîteva zile pe an?

Exerciții rezolvate: Seminarul 8, A1 (Kupiec), A2 (Christoffersen), A4 (Z_2), B2 (FZ0 și DM); lecturi și demonstrații: ▸ Anexă

Rezultatele învățării

- derivați puterea locală a testului POF și minimul pierderii pinball
- calculați testele pentru numărul și momentul depășirilor VaR
  - cu mărimea și puterea fiecărui test pentru T-ul de care dispuneți
- calculați backtesting-ul ES cu reziduuri, cu Z_2 prin simulare și cu statistici construite pe PIT
  - PIT: probability integral transform, transformarea integrală a probabilității
- interpretați ierarhizarea prognozelor (VaR, ES) cu un scor consistent, cu teste DM și cu MCS
  - DM: Diebold–Mariano
  - MCS: model confidence set, mulțimea de încredere a modelelor
- interpretați ce garantează și ce nu garantează un VaR fără ipoteze de distribuție (conformal)
- Exemplele detaliate și subiectele avansate sînt în Aprofundări, înainte de Anexă
  - butonul din subsol duce la ele, iar butonul Înapoi revine la slide-ul de plecare
- Aprofundări și Anexa: lectură pentru familiarizare

Ce testăm? Pierderi și VaR

- **Backtesting**: compararea prognozelor de risc făcute cu informația $\mathcal{F}_{t-1}$ cu rezultatele care au urmat
  - $\mathcal{F}_{t-1}$: tot ce se cunoaște la închiderea zilei $t-1$
- Randamentul logaritmic zilnic: $X_t = r_t = \ln(P_t/P_{t-1})$, unde P_t este prețul de închidere din ziua t
- **Pierderea** zilnică: $L_t = -r_t$ (o valoare pozitivă indică o pierdere)
- **VaR** (valoarea expusă la risc): pierderea depășită cu probabilitatea α (Capitolul 7)
  - adică cuantila q_α a randamentului, cu semn schimbat
  - $\mathrm{VaR}_\alpha(X) = -\inf\{x : \ P(X \le x) > \alpha\} = -q_\alpha(X)$
  - echivalent, cuantila de ordin $1-\alpha$ a pierderii L_t
  - de exemplu, VaR $1\% = 3\%$: pierderea ar trebui să depășească 3% în aproximativ o zi din 100

Ce testăm? Expected Shortfall

- **ES** (pierderea așteptată în coadă): media cuantilelor q_u pe toate nivelurile cozii $u \leq \alpha$
  - adică pierderea medie dincolo de VaR
  - $\mathrm{ES}_\alpha(X) = -\frac{1}{\alpha}\int_0^\alpha q_u(X)\,du = E\left[L_t \mid L_t > \mathrm{VaR}_\alpha\right]$ (cazul continuu)
  - de exemplu, pierderi Student-t cu $\nu = 4$, abaterea standard 1% și $\alpha = 2{,}5\%$: VaR $= 1{,}96\%$, ES $= 2{,}82\%$
  - întotdeauna $\mathrm{ES} \geq \mathrm{VaR}$
- De ce aceste niveluri?
  - VaR 1% este nivelul backtesting-ului Basel (1996)
  - ES 2,5% este măsura de capital din FRTB (*Fundamental Review of the Trading Book*, 2019)
- Pentru distribuția Normală, ES $2{,}5\% = 2{,}338\,\sigma$ și VaR $1\% = 2{,}326\,\sigma$
  - ES 2,5% păstrează aproximativ aceeași prudență, dar ține cont și de adîncimea cozii

Depășiri și calibrarea condiționată

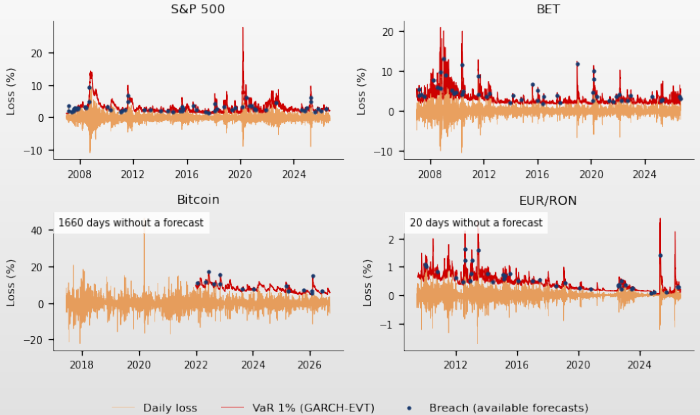
- O **depășire** (excepție): $I_t = 1\{L_t > \mathrm{VaR}_t\}$, unde VaR_t este prognoza făcută în ziua $t-1$ pentru ziua t
- Folosim prognoze la o zi: prognozele pe 10 zile s-ar suprapune și ar produce depășiri dependente
- **Calibrarea condiționată** (Christoffersen (1998)): $P(I_t = 1 \mid \mathcal{F}_{t-1}) = \alpha$ pentru orice t
  - deci I_t sînt Bernoulli(α) i.i.d.
- Două consecințe testabile
  - **Acoperire necondiționată**: aproximativ αT depășiri în T zile
  - **Independență**: depășirile nu se grupează

Șase metode de prognoză a pierderilor extreme

Model	VaR și ES la o zi din
HS	simularea istorică (historical simulation): cuantila empirică de ordin $1-\alpha$ a ultimelor 1 000 de pierderi
Normal	distribuția Normală cu media și abaterea standard din fereastră
Student-t	o distribuție Student-t estimată pe fereastră (locație, scală, grade de libertate ν)
GARCH-t	GARCH (heteroscedasticitate condiționată autoregresivă generalizată, Capitolul 5): $\sigma_t^2 = \omega + a\,\varepsilon_{t-1}^2 + b\,\sigma_{t-1}^2$, cu $\varepsilon_{t-1} = r_{t-1} - \mu$; a: reacția la șocul de ieri; b: persistența; șocuri Student-t
FHS	HS filtrată (filtered historical simulation): volatilitatea GARCH $\times$ cuantilele empirice ale lui $z_t = \varepsilon_t/\sigma_t$
GARCH-EVT	coadă EVT (extreme value theory, teoria valorilor extreme): volatilitatea GARCH $\times$ o coadă GPD (Pareto generalizată) estimată pe cele mai mari 10% pierderi standardizate (100 de excese din 1 000) (McNeil & Frey (2000))

- **Necondiționate** (HS, Normal, Student-t): o singură distribuție pentru întreaga fereastră de 1 000 de zile
  - toate zilele din fereastră au aceeași pondere, deci prognoza reacționează lent la un salt al volatilității
- **Condiționate** sau **filtrate** (GARCH-t, FHS, GARCH-EVT)
  - o scală σ_t actualizată zilnic prin GARCH $\times$ o coadă standardizată

VaR GARCH-EVT pe patru piețe (1/2)



MFM_ch8_four_assets ▶ Interactiv

VaR GARCH-EVT pe patru piețe (2/2)

- Pe grafic: pierderile zilnice (portocaliu), VaR 1% prognozat de GARCH-EVT (roșu) și depășirile (puncte albastre)
  - depășirile: zilele cu $L_t > \mathrm{VaR}_t$
- VaR urmează volatilitatea: se dublează sau se triplează în 2008, martie 2020 și la crahul cripto din 2022
- Rata de depășire a VaR 1% GARCH-EVT: 1,15% pe S&P 500 și 0,97% pe BET (țința: 1%)
- Pe Bitcoin și EUR (moneda euro)/RON (euro/leu), curba VaR începe după perioada cu prognoze GARCH necalculate (Aprofundări)
- EUR/RON are o scară mult mai mică: un curs cu flotare administrată (*managed float*), perioade lungi de calm, salturi rare
- Tipul de dovadă: descriere (arată ce s-a întîmplat în date, fără un test al cauzei)
- Întrebarea capitolului: este suficientă o rată apropiată de 1% pentru a considera un model adecvat?

MFM_ch8_four_assets

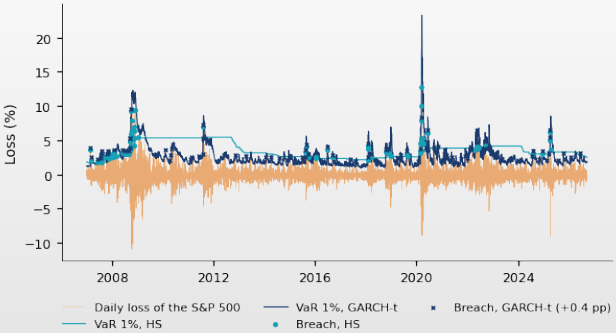
Ce poate și ce nu poate identifica un backtest

- Fiecare backtest verifică o consecință a calibrării condiționate
- Nerespingerea nu înseamnă validare
  - este informativă doar dacă testul are putere împotriva alternativelor relevante, la T disponibil
- H_0 este o ipoteză **comună** despre întreaga procedură de prognoză
  - distribuția, dinamica volatilității, parametrii estimați, lungimea ferestrei și regula de reestimare
- O respingere nu indică ce componentă este greșit specificată
- Un singur rezultat pe zi: dovezile se acumulează doar pe multe zile
- Limitele χ^2 din manuale tratează VaR_t ca fiind cunoscut
  - efectul estimării: backtesting sub riscul de estimare

Are VaR numărul corect de depășiri, la momentele corecte?

- Kupiec: numărul depășirilor
- Christoffersen: memoria de ordinul întîi
- Testele de durată: distanțele dintre depășiri
- DQ (Dynamic Quantile): orice predictor cunoscut ieri
- Pentru fiecare test: ipoteza nulă, statistica, distribuția ei sub H_0, mărimea și puterea la T disponibil
- Începem cu două prognoze S&P 500 care au aproape același număr de depășiri

S&P 500: două prognoze VaR, 2007–2026



- Pe grafic: VaR 1% al HS (turcoaz) și al GARCH-t (albastru)
  - pierderile zilnice (portocaliu) și depășirile fiecărui model (marcaje)
- HS este o funcție în trepte
  - se schimbă doar cînd zile extreme intră în fereastra de 1 000 de zile sau ies din ea
- GARCH-t crește din ziua următoare unui șoc de volatilitate
- Depășiri: 83 pentru HS și 88 pentru GARCH-t, în 4 959 de zile
- Depășirile HS apar în rafale (2008, 2020), deci trebuie testat și **cînd** apar

MFM_ch8_var_forecasts ▶ Interactiv

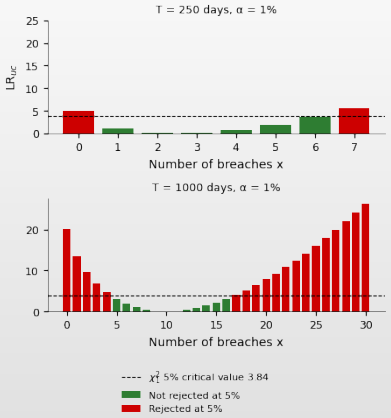
Kupiec (1995): testul POF (proportion of failures, proporția depășirilor)

- Ce testează: dacă rata depășirilor este egală cu α (acoperirea necondiționată), indiferent de momentul lor
- Date: $x = \sum_t I_t$ depășiri în T zile; rata estimată de depășire: $\hat{\pi} = x/T$
- H_0: $\pi = \alpha$, față de H_1: $\pi \neq \alpha$; π: probabilitatea reală de depășire
- Statistica raportului de verosimilitate (Kupiec (1995))
  - compară verosimilitatea Binomială a lui x sub α cu cea sub $\hat{\pi}$
  - $\mathrm{LR}_{uc} = -2 \ln \dfrac{(1-\alpha)^{T-x}\alpha^x}{(1-\hat{\pi})^{T-x}\hat{\pi}^x} \xrightarrow{d} \chi^2_1$
  - $\mathrm{LR}_{uc} \geq 0$ este egală cu 0 cînd $\hat{\pi} = \alpha$ și crește pe măsură ce $\hat{\pi}$ se îndepărtează de α

Testul POF: regula de decizie și un exemplu

- Respingem la 5% dacă $\mathrm{LR}_{uc} > 3{,}84$ ($= 1{,}96^2$, cuantila de 95% a lui χ^2_1)
- Testul este bilateral: și un număr prea mic de depășiri duce la respingere
- Pentru T mic folosim distribuția Binomială(T, α) exactă a lui x
- Mărimea efectivă: suma probabilităților Binomiale ale valorilor x respinse
  - cu x discret, nicio regulă nu are mărimea exact 5%
- Exemplu: S&P 500, GARCH-t, anul 2008: $x = 8$ depășiri în $T = 253$ de zile, față de 2,53 așteptate
- $\mathrm{LR}_{uc} = 7{,}60$, cu $p = 0{,}0058$ (aproximarea χ^2_1)
  - valoarea p (*p-value*) Binomială exactă, bilaterală (dublarea cozii mai mici): $2P(X \geq 8) = 0{,}009$
- Ambele resping la 5%: în 2008, VaR-ul GARCH-t a fost depășit prea des

Numărul acceptabil de depășiri (1/2)



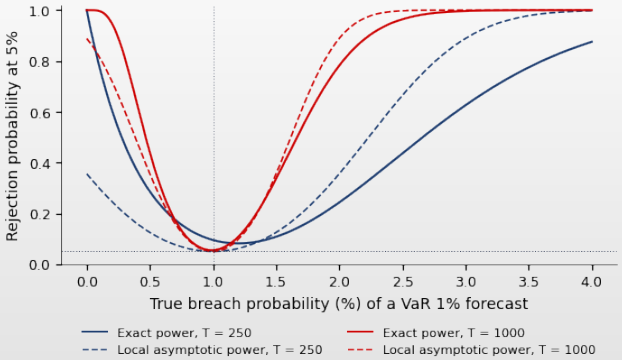
MFM_ch8_kupiec ▶ Interactiv

Numărul acceptabil de depășiri (2/2)

- LR_{uc} în funcție de numărul de depășiri x, un an (sus) și patru ani (jos) de VaR 1%
- Testul acceptă între 1 și 6 depășiri într-un an ($T = 250$) și între 5 și 16 în patru ani ($T = 1000$)
- Chiar și $x = 0$ este respins pe un an
  - $\mathrm{LR}_{uc} = -500 \ln 0{,}99 = 5{,}03$: un VaR niciodată depășit este prea conservator
- Un an conține aproximativ 2,5 zile informative pentru VaR 1%
  - folosim eșantioane lungi, două niveluri ($\alpha = 1\%$ și 2,5%) și mai multe teste

MFM_ch8_kupiec

Puterea testului POF



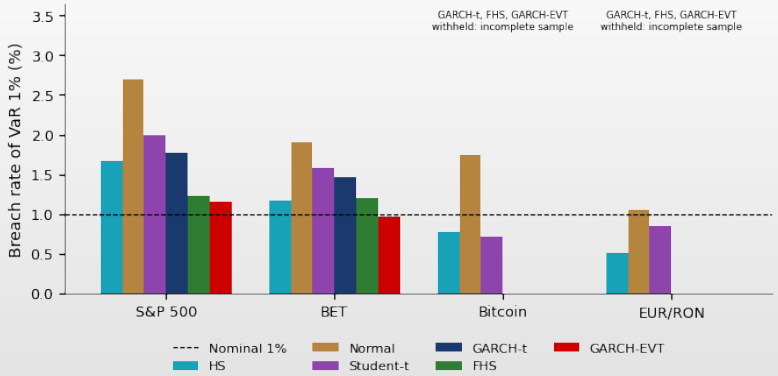
- Axe: rata reală de depășire (orizontal) și probabilitatea de respingere la 5% (vertical)
- Linia continuă: calculul Binomial exact; linia întreruptă: aproximarea locală $\chi^2_1(\lambda)$
  - $\lambda = T(\pi - \alpha)^2/(\alpha(1-\alpha))$
- $T = 250$, $\pi = 2\%$: puterea exactă este 24%, iar cea locală 36%
- Deci o rată de depășire dublă nu este respinsă în trei ani din patru
- Regula $\mathrm{LR}_{uc} > 3{,}84$ are mărimea efectivă 9,5% la $T = 250$, din care $P(x = 0) = 8{,}1\%$
- Pentru o putere de 50%: $\pi \approx 2{,}7\%$ la $T = 250$ și $\pi \approx 1{,}7\%$ la $T = 1000$

MFM_ch8_kupiec ▶ Interactiv

Puterea testului POF: alternative locale

- Alternative locale: rata reală se apropie de α cînd T crește, $\pi_T = \alpha + h/\sqrt{T}$, cu h o distanță fixă pe scara $\sqrt{T}$
- De ce: cu $\pi \neq \alpha$ fix, puterea tinde la 1
  - alternativele locale o țin între 0 și 1 și o fac calculabilă
- $\mathrm{LR}_{uc} = \dfrac{T(\hat{\pi}-\alpha)^2}{\alpha(1-\alpha)} + o_p(1) \xrightarrow{d} \chi^2_1(\lambda), \qquad \lambda = \dfrac{T(\pi-\alpha)^2}{\alpha(1-\alpha)}$
- λ: parametrul de necentralitate; crește cu T și cu $(\pi-\alpha)^2$
- Puterea $= P(\chi^2_1(\lambda) > 3{,}84)$ crește cu λ

Ratele de depășire pe modele și piețe



MFM_ch8_var_forecasts ▶ Interactiv

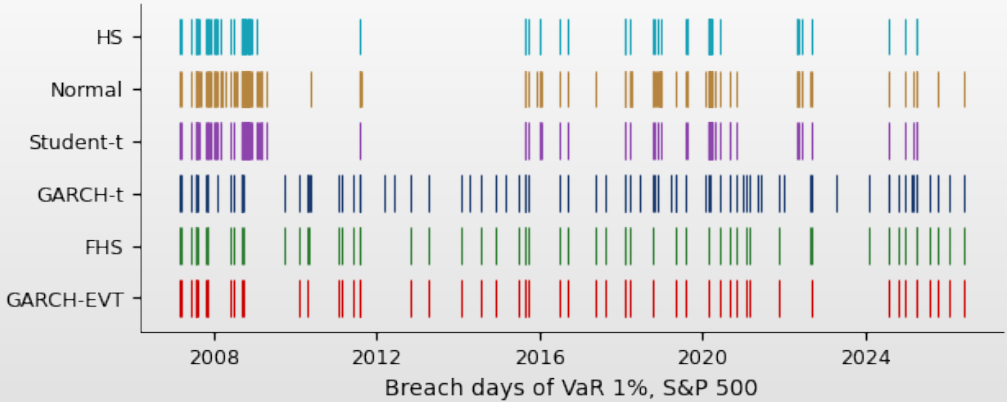
Interpretarea ratelor de depășire

- Pe grafic: ponderea zilelor cu $L_t > \mathrm{VaR}_t$, pentru VaR 1%
- Modelul Normal este respins pe S&P 500 (2,70%), BET (1,90%) și Bitcoin (1,74%), dar nu pe EUR/RON
- GARCH-t și FHS: aceeași volatilitate GARCH, altă coadă (t, respectiv empirică), pe S&P 500 și BET
  - GARCH-t are rate de 1,77%, respectiv 1,46%
  - FHS are 1,23%, respectiv 1,20%
- Datele pun sub semnul întrebării coada t simetrică
- HS este prea conservator pe EUR/RON (0,51%): o criză veche din fereastră îi menține VaR ridicat

Christoffersen (1998): testarea independenței

- Ce testează: dacă o depășire azi depinde de depășirea de ieri
  - testul Kupiec nu vede ordinea depășirilor
- Modelăm I_t ca lanț Markov de ordinul întîi, cu $\pi_{ij} = P(I_t = j \mid I_{t-1} = i)$
- n_{ij}: tranzițiile din starea i ieri în starea j azi
  - $\hat{\pi}_{01} = \frac{n_{01}}{n_{00}+n_{01}}$, $\hat{\pi}_{11} = \frac{n_{11}}{n_{10}+n_{11}}$
- H_0: $\pi_{01} = \pi_{11} = \pi$ (fără memorie), cu $\hat{\pi} = (n_{01} + n_{11})/\sum n_{ij}$
- $\mathrm{LR}_{ind} = -2 \ln \frac{(1-\hat{\pi})^{n_{00}+n_{10}} \hat{\pi}^{n_{01}+n_{11}}}{(1-\hat{\pi}_{01})^{n_{00}} \hat{\pi}_{01}^{n_{01}} (1-\hat{\pi}_{11})^{n_{10}} \hat{\pi}_{11}^{n_{11}}} \xrightarrow{d} \chi_1^2$
- Numărătorul: o singură probabilitate de depășire; numitorul: probabilități care depind de starea de ieri
  - respingem la 5% peste 3,84
- Acoperirea condiționată: $\mathrm{LR}_{cc} = \mathrm{LR}_{uc} + \mathrm{LR}_{ind} \xrightarrow{d} \chi_2^2$ testează ambele proprietăți
  - valoarea critică la 5%: 5,99

Momentele depășirilor (1/2)



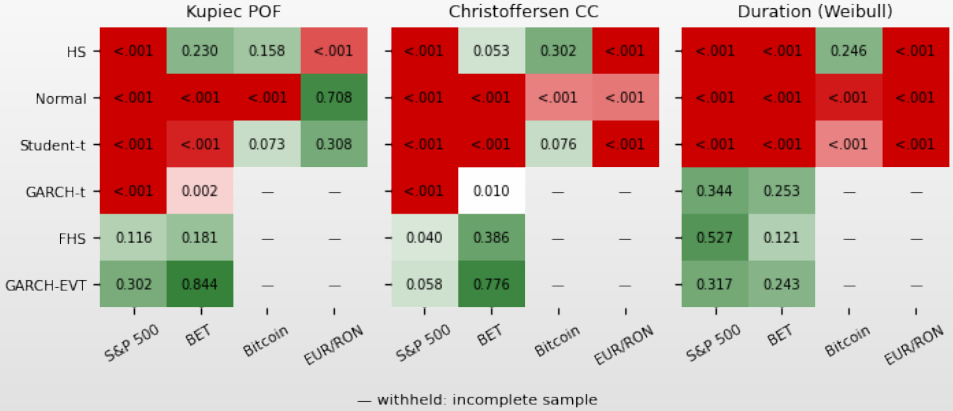
MFM_ch8_coverage_tests ▶ Interactiv

Momentele depășirilor (2/2)

- Fiecare linie verticală marchează o depășire a VaR 1% pe S&P 500; un rînd pentru fiecare model
- HS, Normal și Student-t: rafale de depășiri în 2008, 2015–2016, 2018, 2020, 2022
- Modelele filtrate (GARCH): fără rafale de mai multe luni, dar cu cîteva depășiri în zile consecutive
- Un test pe subeșantion poate rata gruparea: HS, 2007–2010, $n_{00} = 928$, $n_{01} = 38$, $n_{10} = 38$, $n_{11} = 3$
- $\hat{\pi}_{11} = 0{,}0732$, față de $\hat{\pi}_{01} = 0{,}0393$, totuși $\mathrm{LR}_{ind} = 0{,}95$, cu $p = 0{,}33$
- Motivul: $\hat{\pi}_{11}$ se sprijină doar pe 3 perechi de depășiri consecutive

MFM_ch8_coverage_tests

Sinteza testelor VaR (1/2)



MFM_ch8_coverage_tests ▶ Interactiv

Sinteza testelor VaR (2/2)

- Graficul: valorile p ale testelor Kupiec POF, de acoperire condiționată (CC) și de durată (Weibull), pentru VaR 1%
  - CC: conditional coverage
  - roșu: respingere, verde: nerespingere; „—”: prognoză GARCH necalculată
- Testul DQ (care nu apare în grafic; valori p asimptotice χ^2_6) respinge toate modelele pe S&P 500
  - inclusiv FHS (p <0,001) și GARCH-EVT (p <0,001)
  - pe BET nu respinge FHS ($p = 0{,}200$) și GARCH-EVT ($p = 0{,}279$)
  - depășirile sînt previzibile din nivelul VaR și din depășirile trecute, o abatere pe care simpla numărare nu o detectează
- Teste diferite examinează abateri diferite: interpretăm tabloul de ansamblu, nu o singură valoare p
  - modelele necondiționate sînt respinse de testul duratelor aproape peste tot
  - GARCH-EVT nu este respins de testele de acoperire condiționată sau de durată pe BET
- Tipul de dovadă: comparație între proceduri (arată care metodă funcționează mai bine pe aceleași date, nu de ce)

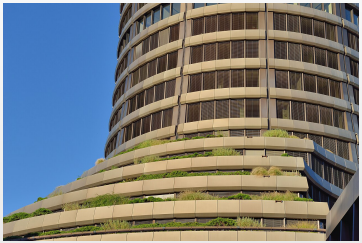
MFM_ch8_coverage_tests

Cum testează supraveghetorul? Semaforul Basel

Excepții	$P(X \leq x)$	Zona
0	8,11%	verde
2	54,32%	verde
4	89,22%	verde
5	95,88%	galben
7	99,60%	galben
9	99,97%	galben
10	99,99%	roșu

- Băncile cu modele interne compară zilnic VaR 1% cu P&L-ul efectiv din ultimele 250 de zile (BCBS (1996))
  - P&L: profitul și pierderea
- Ce testează: doar numărul excepțiilor, nu mărimea sau momentul lor
- $X \sim$ Binomială$(250; 0{,}01)$: excepțiile unui VaR 1% corect în 250 de zile
- $P(X \leq x) = \sum_{k=0}^{x} \binom{250}{k} 0{,}01^k \, 0{,}99^{250-k}$: coloana a doua a tabelului
- Zona galbenă începe unde $P(X \leq x) \geq 95\%$, cea roșie unde $P(X \leq x) \geq 99{,}99\%$
- De ce o regulă de numărare: simplă, automată, greu de manipulat

Regulile de backtesting din 1996 și 2019

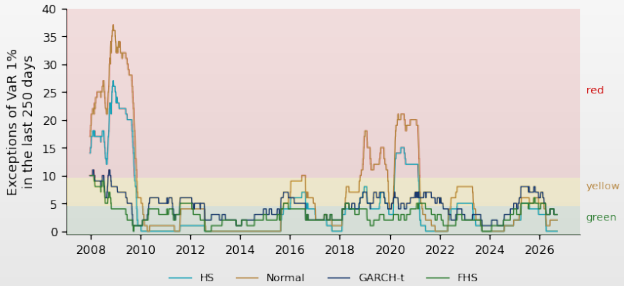


Turnul BIS (Bank for International Settlements), Basel: sediul secretariatului BCBS

Foto: Taxiarchos228 (2014); Free Art License; Wikimedia Commons

- BCBS (1996): capitalul pe VaR 1% la 10 zile, cu multiplicatorul 3
- În 1996, multiplicatorul crește cu 0,40–0,85 în zona galbenă și cu 1 în zona roșie
- BCBS (2019) (FRTB): capitalul pe ES 2,5%, dar backtesting-ul rămîne pe VaR 1% din ultimele 250 de zile
- Multiplicatorul FRTB: 1,50 în zona verde (0–4 excepții)
  - 1,70–1,92 în cea galbenă (5–9) și 2,00 în cea roșie (10 sau mai multe)
- Un birou de tranzacționare trece la abordarea standardizată dacă are
  - în 12 luni, peste 12 excepții ale VaR 1%
  - sau peste 30 ale VaR 2,5%
- Aprobarea IMA cere și un test de atribuire a P&L
  - IMA: internal models approach, abordarea pe modele interne
- Regula Basel de backtesting numără excepțiile VaR; ES nu este testat direct
  - faptul că ES nu poate fi ierarhizat singur (nu este elicitabil) nu împiedică backtesting-ul ES (Acerbi & Szekely (2014), secțiunea despre ES)
  - ea privește compararea prognozelor (secțiunea despre elicitabilitate)

Semaforul de-a lungul timpului



- Excepțiile VaR 1% în ultimele 250 de zile, S&P 500; benzi: zonele verde, galbenă, roșie
- Ponderea zilelor aflate în zona roșie: Normal 22,3%, HS 14,7%, GARCH-t 2,0%, FHS 0,8%
- GARCH-t se află în zona galbenă 46,5% din timp
- Maximul: 37 de excepții pentru Normal și 27 pentru HS, în cele 250 de zile încheiate la 01.12.2008
- Ferestrele se suprapun: penalizarea durează un an după criză, cînd capitalul este deja sub presiune
- Tipul de dovadă: comparație între proceduri

MFM_ch8_traffic_light ▶ Interactiv

Verificarea FRTB la nivel de birou, ultimele 250 de zile (limitele 12 și 30)

Model	S&P 500	BET	Bitcoin	EUR/RON
HS	0 / 2	4 / 7	3 / 4	1 / 5
Normal	2 / 2	6 / 9	3 / 5	1 / 4
Student-t	0 / 2	4 / 9	1 / 4	5 / 8
GARCH-t	3 / 6	4 / 9	3 / 5	4 / 5
FHS	3 / 4	4 / 6	4 / 5	3 / 4
GARCH-EVT	3 / 3	2 / 7	4 / 5	3 / 4

- Celule: excepții ale VaR 1% / VaR 2,5% în ultimele 250 de zile ale fiecărei serii, pînă la 18.09.2026
- În aceste zile, prognozele GARCH există pe toate cele patru piețe
- Toate modelele se încadrează confortabil în limitele 12 și 30
- Doar ilustrativ: FRTB aplică limitele P&L-ului birourilor (efectiv și ipotetic), nu randamentelor de indici
- Ultimul an a fost calm: o zonă verde în prezent spune puțin despre următoarea criză

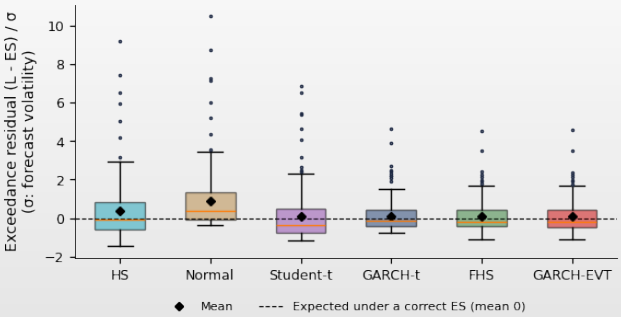
Este ES prognozat corect? Ce putem testa

- ES este media pierderilor dincolo de VaR: informația vine doar din aproximativ αT zile din coadă
- S&P 500, $\alpha = 2{,}5\%$: aproximativ 124 de zile din 4 959; într-un an, 6,25 zile
- Pe un an, eroarea standard a mediei din coadă este de ordinul $1/\sqrt{6{,}25} = 0{,}40$ abateri standard ale pierderilor din coadă
- Trei familii de teste
  - reziduurile de depășire (McNeil–Frey)
  - statisticile Z_1, Z_2 (Acerbi–Szekely)
  - statisticile pe PIT (Du–Escanciano, teste multinomiale)
- Toate cer mai mult decît prognoza ES: VaR, volatilitatea sau întreaga distribuție prognozată
  - un test doar pe ES: studiul de caz

McNeil și Frey (2000): reziduurile de depășire

- Ce testează: dacă pierderea medie în zilele de depășire este cea prognozată de ES, pe scala volatilității
- Reziduul de depășire: $e_t = (L_t - \mathrm{ES}_t)/\sigma_t$ în zilele cu $L_t > \mathrm{VaR}_t$ (VaR 2,5%)
  - $\sigma_t > 0$: volatilitatea prognozată la $t-1$
- Dacă VaR **și** ES sînt corecte: $E[e_t \mid L_t > \mathrm{VaR}_t, \mathcal{F}_{t-1}] = 0$
  - un VaR greșit schimbă evenimentul de condiționare
- H_0: $E[e_t] = 0$ față de H_1: $E[e_t] > 0$ (ES subestimat)
- Statistica: $t = \bar{e}/(s_e/\sqrt{N})$, cu $\bar{e}$ și s_e media și abaterea standard a celor N reziduuri
- valoarea p: bootstrap i.i.d. al reziduurilor centrate (5 000 de replicări) (McNeil & Frey (2000))
  - pentru că reziduurile sînt puține și au cozi groase
- Bootstrap-ul presupune reziduuri independente
  - cu reziduuri dependente se folosește un bootstrap pe blocuri

Reziduurile de depășire pe S&P 500



- Distribuția lui $(L_t - \mathrm{ES}_t)/\sigma_t$ în zilele de depășire a VaR 2,5% (romburile marchează mediile)
- Mediile reziduurilor: Normal 0,875, HS 0,363, Student-t 0,082, GARCH-t 0,100, FHS 0,104, GARCH-EVT 0,068
- valorile p bootstrap
  - Normal $<$0,001, HS $<$0,001, Student-t 0,173, GARCH-t 0,025
  - FHS 0,072, GARCH-EVT 0,157
- Modelul Normal subestimează media din coadă cu aproape o unitate de volatilitate
  - reziduurile extreme, de pînă la 10 volatilități, contează cel mai mult pentru capital

MFM_ch8_es_backtests ▶ Interactiv

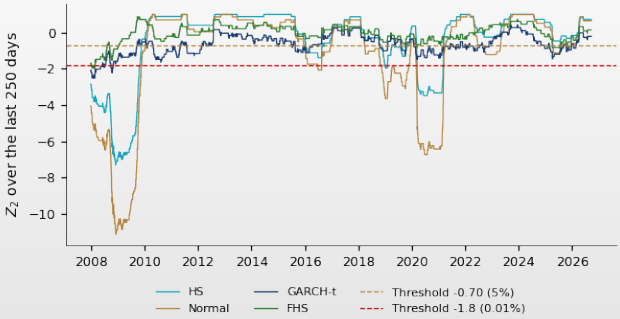
Acerbi și Szekely (2014): testarea directă a ES

- Indicatorul de depășire a VaR 2,5%: $I_t = 1\{L_t > \mathrm{VaR}_t^{\alpha}\}$, cu $\alpha = 0{,}025$
  - $N_T = \sum_t I_t$: numărul de depășiri în T zile
  - ES_t: prognoza ES pentru ziua t
- $Z_1 = 1 - \frac{1}{N_T} \sum_t \frac{L_t I_t}{\mathrm{ES}_t}$, condiționat de depășiri
  - este pierderea medie din coadă atît de mare cît s-a prognozat?
- $Z_2 = 1 - \frac{1}{T\alpha} \sum_t \frac{L_t I_t}{\mathrm{ES}_t}$: testează frecvența și mărimea împreună
- Sub H_0 (distribuția prognozată este corectă) $E[Z_2] = 0$; riscul subestimat dă $Z_2 < 0$ (Acerbi & Szekely (2014))
- Pentru $T = 250$ și multe distribuții, valorile critice de 5% și 0,01% sînt apropiate de $-0{,}70$ și $-1{,}8$
  - praguri de referință pentru un semafor ES

Calculul valorii p prin simulare

- Distribuția lui Z_2 sub H_0 depinde de distribuțiile prognozate F_t, deci valorile critice se simulează
  - F_t: funcția de repartiție a lui L_t, prognozată în ziua $t-1$
- Pasul 1, repetat de $M = 2\,000$ de ori: pentru fiecare zi extragem $U_t \sim \text{Uniform}(0,1)$
  - calculăm $L_t^* = F_t^{-1}(U_t)$ (F_t^{-1}: funcția cuantilă)
- Pasul 2: calculăm Z_2^* din pierderile simulate, cu aceleași prognoze VaR și ES
  - o calibrare cu distribuțiile prognozate fixate și U_t independente: riscul de estimare și reestimarea GARCH sînt ignorate
- valoarea p = ponderea valorilor $Z_2^* \leq Z_2$ observat
  - cu $M = 2\,000$, cea mai mică valoare p nenulă este 0,0005; nicio valoare simulată mai extremă înseamnă $p < 0{,}0005$, nu $p = 0$
- Contează doar extragerile din coadă
  - pentru HS și FHS păstrăm cele mai mari 25 de pierderi din fereastra fiecărei zile ($25 = 2{,}5\% \times 1\,000$)
- S&P 500, întreg eșantionul
  - $Z_2 = -0{,}664$ pentru GARCH-t ($p < 0{,}001$)
  - $Z_2 = -0{,}079$ pentru FHS ($p = 0{,}209$)
- Valoarea critică de 5% la $T = 4\,959$ este de circa $-0{,}150$, mai aproape de 0 decît $-0{,}70$ la $T = 250$

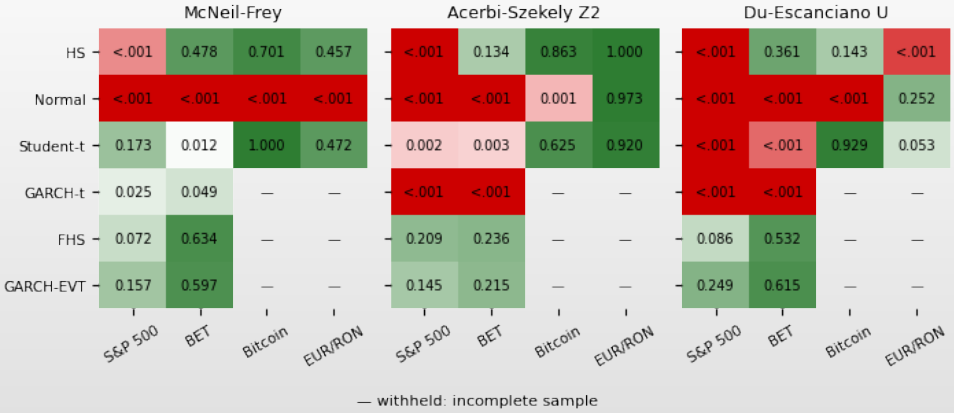
Z_2 pe ferestre mobile de 250 de zile



- Z_2 pentru ES 2,5% pe S&P 500, recalculat zilnic pe ultimele 250 de zile
  - liniile orizontale: pragurile $-0{,}70$ și $-1{,}8$
- Minim: Normal $-11{,}13$ (fereastra încheiată la 04.12.2008), HS $-7{,}30$ (01.12.2008)
- Minim: GARCH-t $-2{,}52$ (05.02.2008), FHS $-1{,}97$ (17.01.2008)
- Ferestre cu $Z_2 < -1{,}8$: Normal 21,9%, HS 15,8%, GARCH-t 3,0%, FHS 0,7%
- GARCH-t este sub $-0{,}70$ în 44,3% din ferestre
- Verificarea zilnică înseamnă testare repetată: alarmele false se acumulează
  - monitorizarea secvențială: ▸ Anexă

MFM_ch8_es_backtests ▶ Interactiv

Sinteza testelor ES (1/2)



MFM_ch8_es_backtests ▶ Interactiv

Sinteza testelor ES (2/2)

- La pragul de semnificație de 5%, FHS și GARCH-EVT nu sînt respinse pe S&P 500 și BET
  - nici de McNeil–Frey, nici de Z_2, nici de Du–Escanciano U, luate separat
- Ipoteze nule și date diferite dau decizii diferite
  - pentru Student-t pe S&P 500, McNeil–Frey dă $p = 0{,}173$, dar Z_2 dă $p = 0{,}002$
- Acestea sînt teste de calibrare, nu clasamente
- GARCH-t este respins de Z_2 pe S&P 500 ($p < 0{,}001$) și BET ($p < 0{,}001$)
- Semnul lui U contează: pe EUR/RON, HS are $U = -3{,}68$ ($p < 0{,}001$)
  - o respingere pentru că pierderile rămîn **mai puțin adînci** decît s-a prognozat
  - în acord cu Z_2 ($p = 1{,}000$) și cu rata de depășire a VaR 1% de 0,51%
- Tipul de dovadă: comparație între proceduri

MFM_ch8_es_backtests

Cum ierarhizăm prognozele? Scoruri consistente

- Un backtest întreabă dacă un model este **acceptabil**
  - mai multe modele pot fi acceptate sau toate pot fi respinse
- Backtesting-ul comparativ întreabă care model este **mai bun** (Nolde & Ziegel (2017))
- Ierarhizarea cere o **funcție de scor** $S(x, L)$
  - $S(x, L)$: pierderea asociată raportării lui x cînd se realizează L
  - comparăm $\bar{S} = \frac{1}{T}\sum_t S(x_t, L_t)$, mai mic = mai bun
- S este **consistent** pentru $\theta(F)$, iar θ este **elicitabilă**, dacă $\theta(F) = \arg\min_x E_F[S(x, L)]$ (Gneiting (2011))
- $\theta(F)$: funcționala-țintă (de exemplu VaR) a distribuției pierderii F; E_F: media cînd $L \sim F$
- Ierarhizarea cere ca valoarea adevărată să fie **unicul** punct de minim (consistență strictă)

Pierderea pinball și ES

- VaR este elicitabil, cu pierderea pinball $S(v, L) = (1\{L > v\} - \alpha)(L - v)$, unde v este prognoza VaR
- Dacă $E|L| < \infty$ și densitatea $f > 0$ în cuantilă:
  $$\frac{\partial}{\partial v} E_F[S(v, L)] = F(v) - (1 - \alpha) = 0 \;\Rightarrow\; v = F^{-1}(1 - \alpha) = \mathrm{VaR}_\alpha$$
- A doua derivată este $f(v) > 0$, deci minimul este unic
- ES **nu** este elicitabil singur (Gneiting (2011))
  - nicio funcție de scor care depinde doar de prognoza ES raportată e nu este strict consistentă pentru ES pe clasa relevantă de distribuții
- Motivul: mulțimile de nivel ale ES nu sînt convexe la mixturi de distribuții (Anexă)
- De aceea ES se ierarhizează împreună cu VaR (slide-ul următor)

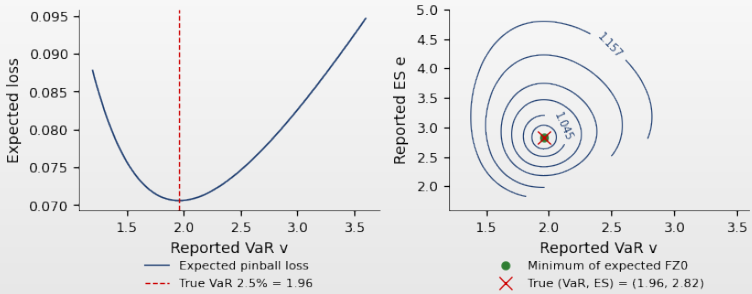
VaR și ES împreună: Fissler și Ziegel (2016)

- Perechea (VaR, ES) este elicitabilă împreună: Fissler & Ziegel (2016) caracterizează familia FZ (familia Fissler–Ziegel de funcții de scor pentru VaR și ES) a funcțiilor de scor consistente
- FZ0: membrul familiei FZ cu diferențe de scor omogene de grad zero (Patton, Ziegel & Chen (2019)), pentru pierderi pozitive, $0 < v \leq e$ și $E|L| < \infty$
- $\mathrm{FZ0}(v, e; L) = \dfrac{1\{L > v\}(L - v)}{\alpha\, e} + \dfrac{v}{e} + \ln e - 1$
- v, e: prognozele VaR și ES (pierderi pozitive); L: pierderea realizată
- Primul termen penalizează depășirile, scalate cu ES
- Ultimii termeni penalizează raportarea unui ES mare
- FZ0 mediu mai mic = prognoză comună VaR și ES mai bună
  - $\mathrm{FZ0}(cv, ce; cL) = \mathrm{FZ0}(v, e; L) + \ln c$: diferențele dintre modele nu depind de scala pierderilor

Pierderea FZ0: un exemplu pentru o zi

- $\alpha = 2{,}5\%$, prognozele $v = 3\%$ și $e = 3{,}9\%$, pierderi în procente
- Pierderea $L = 4{,}5\%$ (depășire): $\mathrm{FZ0} = 15{,}385 + 1{,}130 = 16{,}515$
- Pierderea $L = 1\%$ (fără depășire): rămîne doar $v/e + \ln e - 1 = 0{,}769 + 1{,}361 - 1 = 1{,}130$
- În zecimale ($v = 0{,}03$, $e = 0{,}039$), doar $\ln e$ se schimbă
  - fiecare valoare scade cu $\ln 100 = 4{,}605$
  - la $-3{,}475$ fără depășire și $11{,}910$ cu depășire
- De aceea pierderile FZ0 medii pe randamente zecimale sînt negative (S&P 500, FHS: $-3{,}551$)
  - diferențele dintre modele nu se schimbă

Elicitabilitatea, ilustrată grafic



- Pierderi Student-t cu $\nu = 4$ și dispersie unitară, $\alpha = 2{,}5\%$
- Stînga: pierderea pinball așteptată este minimă în VaR 2,5% adevărat, 1,96
- Dreapta: pierderea FZ0 așteptată este minimă în (1,96; 2,82)
  - adică în perechea (VaR, ES) adevărată, în limita grilei
- Un prognozator care cunoaște adevărul nu are motiv să raporteze altceva: de aceea ordonarea după scorul mediu are sens

MFM_ch8_elicitability ▶ Interactiv

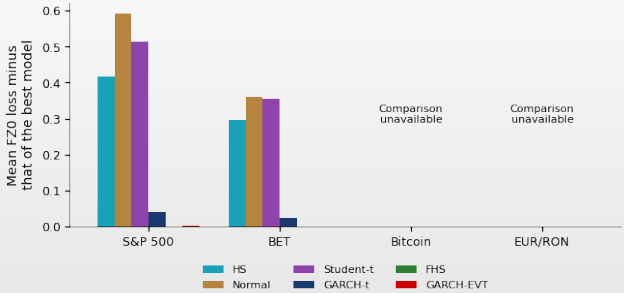
Este diferența dintre două prognoze semnificativă?

- Scorul mediu ierarhizează prognozele, dar diferențele mici pot fi zgomot de eșantionare
- Diebold–Mariano: au două prognoze aceeași pierdere medie?
- Giacomini–White: poate fi prezisă diferența de pierdere din informația de ieri?
- Nolde–Ziegel: este modelul intern mai bun decît un model standard?
- MCS: ce mulțime conține, cu o probabilitate dată, cele mai bune modele?
- Toate folosesc pierderea FZ0 a perechii (VaR 2,5%, ES 2,5%)

Testul Diebold–Mariano pe pierderile FZ0

- Ce testează: dacă două prognoze au aceeași pierdere medie
- $d_t = \mathrm{FZ0}_{A,t} - \mathrm{FZ0}_{B,t}$; H_0: $E[d_t] = 0$ (acuratețe predictivă egală), Diebold & Mariano (1995)
- $\mathrm{DM} = \dfrac{\bar{d}}{\sqrt{\widehat{\mathrm{LRV}}(d_t)/T}} \xrightarrow{d} N(0,1)$
- $\widehat{\mathrm{LRV}} = \hat{\gamma}_0 + 2\sum_{k=1}^{q}\left(1 - \frac{k}{q+1}\right)\hat{\gamma}_k$: dispersia pe termen lung, cu ponderi Newey–West (Newey & West (1987))
  - $\hat{\gamma}_k$: autocovarianța lui d_t la lagul k
- De ce: pierderile din coadă se grupează, deci d_t este autocorelat
- Regula lagurilor: $q = \lfloor 4(T/100)^{2/9} \rfloor = 9$ pentru $T = 4\,959$
- $\mathrm{DM} > 0$: modelul A are pierderea medie mai mare (mai slab)
  - $|\mathrm{DM}| > 1{,}96$: diferență semnificativă la 5%

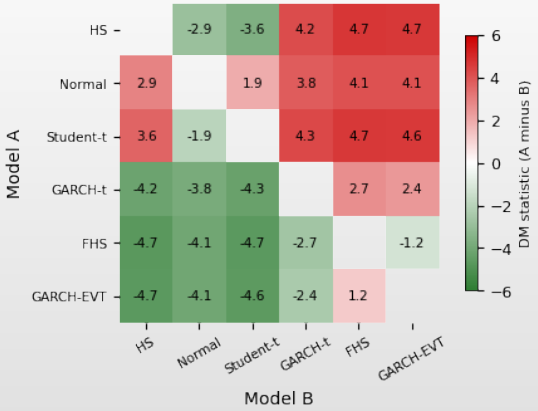
Pierderea FZ0 medie pe modele



- Pierderea FZ0 medie a prognozelor (VaR 2,5%, ES 2,5%) minus cea a celui mai bun model de pe fiecare piață
- Pe S&P 500 și BET, toate cele șase modele sînt comparate pe aceleași zile de evaluare
- Cea mai mică pierdere FZ0: FHS pe S&P 500 și BET
- Pierderea FZ0 medie a GARCH-t depășește pe cea a FHS cu 0,040 pe S&P 500
- Mărimea diferenței nu arată dacă datele separă modelele: o testăm

MFM_ch8_comparative ▶ Interactiv

Teste Diebold–Mariano pe perechi, S&P 500 (1/2)



MFM_ch8_comparative ▶ Interactiv

Teste Diebold–Mariano pe perechi, S&P 500 (2/2)

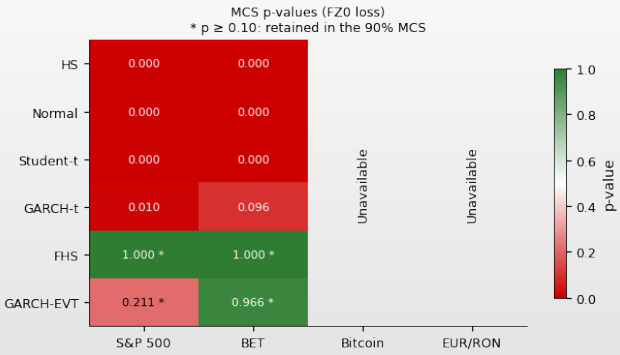
- DM pentru modelul A (rînd) minus B (coloană) pe pierderile FZ0
  - roșu: A mai slab, verde: A mai bun
- Față de FHS
  - HS $\mathrm{DM} = 4{,}69$, GARCH-t $\mathrm{DM} = 2{,}66$: ambele semnificativ mai slabe
  - GARCH-EVT $\mathrm{DM} = 1{,}24$: nu se deosebește
  - dar testul condiționat Giacomini–White îl respinge față de FHS ($p = 0{,}035$, Aprofundări)
- Normal minus HS: $\mathrm{DM} = 2{,}88$
  - o cuantilă empirică, deși se ajustează lent, este totuși superioară unei cozi subțiri
- Cincisprezece teste pe perechi: o problemă de testare multiplă
  - care modele rămîn **împreună**? (mulțimea de încredere a modelelor, MCS)

MFM_ch8_comparative

Mulțimea de încredere a modelelor: ideea

- MCS (Hansen, Lunde & Nason (2011)): mulțimea care conține cele mai bune modele cu o probabilitate asimptotică de cel puțin $1 - \alpha_{MCS}$
- Ce corectează: cele 15 teste DM pe perechi nu controlează eroarea pe întreaga familie
  - MCS compară toate modelele împreună
- Garanția cere condițiile de regularitate ale metodei
  - diferențe de pierdere staționare, cu dependență slabă
- Folosim $\alpha_{MCS} = 10\%$: păstrăm modelele cu valoarea p MCS $\geq 0{,}10$
- O mulțime MCS mare este un rezultat informativ: datele nu separă modelele

Modelele păstrate în MCS



- valorile p MCS pe pierderea FZ0 a perechii (VaR 2,5%, ES 2,5%)
  - modelele cu $p \geq 0{,}10$ formează MCS de 90%
- S&P 500 și BET: MCS conține FHS și GARCH-EVT
- GARCH-t este exclus pe S&P 500 ($p = 0{,}010$) și, la limită, pe BET ($p = 0{,}096$)
- HS, Normal și Student-t: $p \leq 0{,}001$ pe ambele piețe
- Bitcoin și EUR/RON: fără MCS, deoarece prognozele GARCH nu sînt calculate (slide-ul despre prognozele necalculate)
- Tipul de dovadă: comparație între proceduri

MFM_ch8_comparative ▶ Interactiv

Se poate garanta acoperirea fără ipoteze de distribuție?

- Modelele de pînă acum presupun o distribuție (Normală, t, GPD) sau un trecut stabil (HS)
- **Predicția conformală** (Lei et al. (2018), Angelopoulos & Bates (2021)) transformă orice prognoză într-o cuantilă cu garanție de acoperire
- Scorul: $s_t = L_t/\hat{\sigma}_t$, unde $\hat{\sigma}_t$ este abaterea standard a celor 250 de randamente anterioare
- Mulțimea de calibrare: ultimele $n = 250$ de scoruri (un an de tranzacționare), ordonate $s_{(1)} \leq \cdots \leq s_{(n)}$
- VaR conformal split: $\mathrm{VaR}_t = \hat{\sigma}_t\, s_{(k)}$, cu $k = \lceil (n+1)(1-\alpha) \rceil$ ($\lceil x \rceil$: cel mai mic întreg $\geq x$)
- Pentru $n = 250$ și $\alpha = 0{,}01$: $k = 249$, adică al doilea cel mai mare scor

Conformal split: garanția și limitele ei

- Garanția, dacă cele $n+1$ scoruri sînt **schimbabile**: $1-\alpha \le P(L_t \le \mathrm{VaR}_t) \le 1-\alpha+\frac{1}{n+1}$ (fără valori egale)
  - schimbabile: legea comună nu se schimbă la permutare
- Aici garanția este doar aproximativă, din două motive (extinderi: Anexă)
- Gruparea volatilității (*volatility clustering*) încalcă schimbabilitatea
- $\hat{\sigma}_t$ provine din ferestre mobile (*rolling*) suprapuse, deci scorurile nu sînt independente
- Conformal split autentic estimează scorul o singură dată, pe un eșantion separat

Inferența conformală adaptivă (ACI)

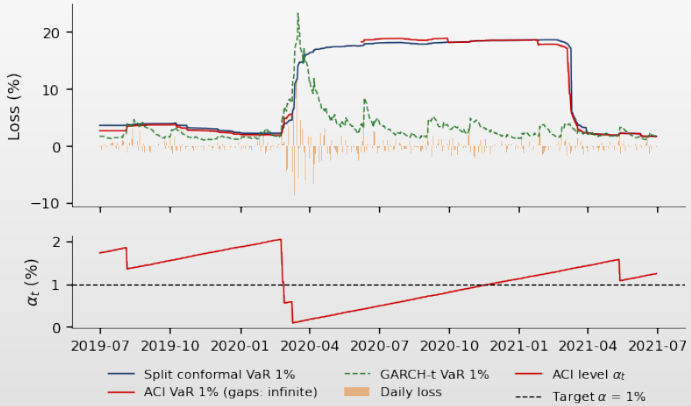


Emmanuel Candès, coautor al inferenței conformale adaptive

Foto: Cmichel67 (2026); CC BY-SA 4.0; Wikimedia Commons

- Gibbs & Candès (2021), ACI (*Adaptive Conformal Inference*): nivelul de lucru se actualizează după fiecare zi
  - $\alpha_{t+\mathbf{1}} = \alpha_t + \gamma\,(\alpha - \mathrm{err}_t)$, $\mathrm{err}_t = 1\{L_t > \mathrm{VaR}_t\}$
  - α_t: nivelul cuantilei conformale din ziua t; $\gamma > 0$: pasul; $\alpha_{\mathbf{1}}$: nivelul inițial
- După o depășire, α_t scade (VaR crește); după zile calme, α_t crește încet la loc
- Folosim $\gamma = 0{,}005$: un pas mic stabilizează nivelul
  - un pas mare reacționează mai repede, dar duce mai des la VaR infinit
- Garanția, pentru orice șir: $\left|\frac{1}{T}\sum_t \mathrm{err}_t - \alpha\right| \le \frac{\max(\alpha_{\mathbf{1}}, 1-\alpha_{\mathbf{1}}) + \gamma}{\gamma T}$
  - o **medie** pe termen lung, nu fiecare zi sau fiecare criză
- Dacă $\alpha_t < 1/(n+1)$, cuantila cerută nu există: VaR $= +\infty$
- Dacă $\alpha_t \ge 1$, pragul este $-\infty$ și depășirea este sigură; garanția cere ambele convenții

VaR conformal în perioada COVID-19 (1/2)



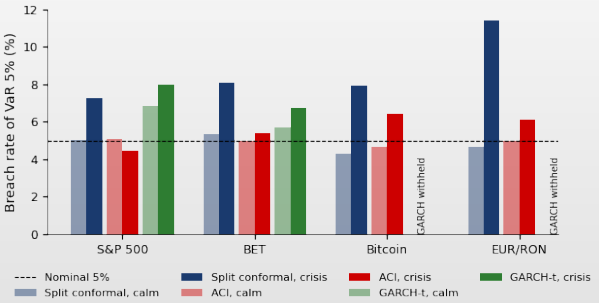
MFM_ch8_conformal ▶ Interactiv

VaR conformal în perioada COVID-19 (2/2)

- S&P 500, iulie 2019–iunie 2021, VaR 1%
  - sus: conformal split, ACI și GARCH-t
  - jos: nivelul ACI α_t
- Între 20.02 și 30.04.2020: 50 de zile, cu 0,5 depășiri așteptate
- Depășiri observate: conformal split 6, ACI 4 (8% din zile)
- ACI a stabilit VaR $= \infty$ în 37 din aceste zile: nicio depășire nouă, dar un VaR infinit nu aduce informație
- VaR 1% conformal split rămîne în jur de 18% timp de un an
  - $\hat{\sigma}_t$ și $s_{(\mathbf{249})}$ folosesc ultimele 250 de zile, deci păstrează crahul pînă în martie 2021
- GARCH-t se adaptează în cîteva zile: după zile calme, ponderea unui șoc trecut în σ_t^2 scade cu factorul b pe zi
  - cu parametrii estimați pe S&P 500, 1990–2026: $b = 0{,}892$, înjumătățire în circa 6 zile
  - prognoza variației revine în medie cu rata $a + b = 0{,}994$: înjumătățire în circa 124 de zile

MFM_ch8_conformal

Acoperirea în perioade calme și de criză



- Pe grafic: rata de depășire a VaR 5% pe piețe și metode, în zile calme și de criză
- Zi de criză, definită dinainte
  - volatilitatea pe 20 de zile depășește percentila 90 a celor 1 000 de zile anterioare
- De ce VaR 5%: în 713 zile de criză S&P 500, circa 36 de depășiri așteptate, față de 7 la VaR 1%
- Conformal split, toate zilele: S&P 500 5,34%, BET 5,63%, Bitcoin 4,58%, EUR/RON 5,06%
- Conformal split, zilele de criză: S&P 500 7,29%, BET 8,08%, Bitcoin 7,95%, EUR/RON 11,43%
- ACI, zilele de criză: S&P 500 4,49%, BET 5,38%
  - la VaR 1%, acest rezultat cere un VaR infinit în 13,4% din zilele de evaluare S&P 500
- Acoperirea marginală nu este acoperire în criză
- Tipul de dovadă: comparație între proceduri

MFM_ch8_conformal ▶ Interactiv

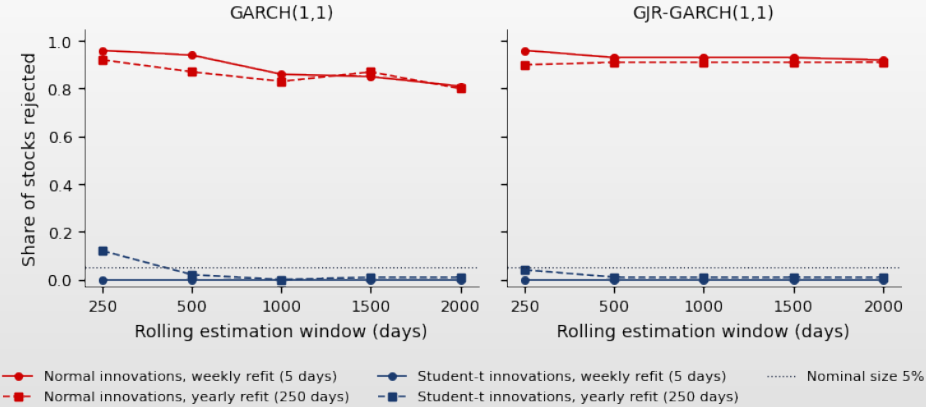
Studiu de caz (1/4): testarea ES doar pe baza prognozelor ES

- Bayer & Dimitriadis (2022): pot fi testate prognozele ES folosind doar randamentele observate și aceste prognoze?
- Relevanța: capitalul FRTB se calculează din ES 2,5% (perspectiva reglementatorului)
  - testele ES de mai sus cer și VaR, volatilitatea sau distribuția prognozată, pe care supraveghetorii nu le primesc neapărat
  - respingerea unui astfel de test comun nu arată care prognoză este greșită
- Ideea: o regresie Mincer–Zarnowitz a randamentelor pe prognozele ES
- H_0: termenul liber este 0 și panta este 1 (o prognoză corectă nu are nevoie de corecție)
- Regresia se poate estima pentru că perechea (VaR, ES) este elicitabilă în comun (Fissler & Ziegel (2016))
  - o regresie comună cuantilă–ES (Patton, Ziegel & Chen (2019))

Studiu de caz (2/4): regresia ES

- Convenția randamentelor: $\hat{e}_t = -\widehat{\mathrm{ES}}_t < 0$ este prognoza ES pe scala randamentelor, cu $\tau = \alpha = 2{,}5\%$
- Ecuația ES: $r_t = \gamma_{\mathbf{1}} + \gamma_{\mathbf{2}}\hat{e}_t + u_t^e$, cu $\mathrm{ES}_\tau(u_t^e \mid \mathcal{F}_{t-\mathbf{1}}) = 0$, deci $\gamma_{\mathbf{1}} + \gamma_{\mathbf{2}}\hat{e}_t$ este ES condiționat adevărat
- $\gamma_{\mathbf{1}}$: termenul liber; $\gamma_{\mathbf{2}}$: panta; prognoze corecte: $\gamma_{\mathbf{1}} = 0$, $\gamma_{\mathbf{2}} = 1$
- ESR (regresia expected shortfall) cere și o ecuație a cuantilei, deoarece ES se estimează împreună cu VaR
- **ESR auxiliar**: ecuația cuantilei pe $(1, \hat{v}_t)$, deci cere și prognoze VaR $\hat{v}_t$
- **ESR strict**: ecuația cuantilei pe $(1, \hat{e}_t)$, deci cere doar prognoze ES
- **ESR cu termen liber**: $r_t - \hat{e}_t = \gamma_{\mathbf{1}} + u_t^e$ (panta fixată la 1)
  - test t unilateral pentru $H_{\mathbf{0}}$: $\gamma_{\mathbf{1}} \geq 0$ față de $\gamma_{\mathbf{1}} < 0$ (ES subestimat)

Studiu de caz (3/4): cozile contează mai mult decît reestimarea



MFM_ch8_esr_case ▶ Interactiv

Studiu de caz (4/4): cozile contează mai mult decît reestimarea

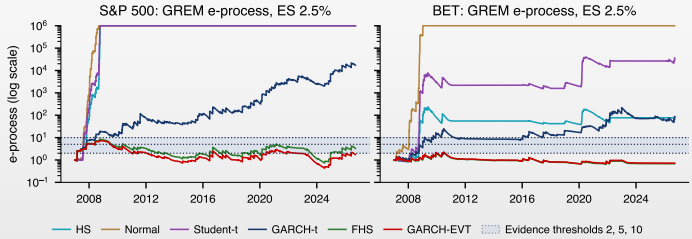
- Sursa: Bayer & Dimitriadis (2022), Tabelul 3: cele mai mari 200 de acțiuni din S&P 500, 2010–2019
- Axe: fereastra de estimare, 250–2 000 de zile (orizontal)
  - vertical: ponderea acțiunilor cu ES 2,5% respins de ESR cu termen liber, unilateral la 5%
- Linia continuă: reestimare la 5 zile; linia întreruptă: la 250 de zile
  - Tabelul 3 conține și 21, 62 și 125 de zile
- GJR-GARCH (Glosten–Jagannathan–Runkle): GARCH cu o reacție suplimentară la șocurile negative
- Inovații Normale: ponderea respinsă este între 0,80 și 0,96 în toate cele 50 de celule
  - 5 ferestre $\times$ 5 frecvențe $\times$ GARCH/GJR-GARCH
- Inovații Student-t: cel mult 0,12 cu ferestre de 250 de zile
  - cel mult 0,01 cu ferestre de cel puțin 1 000 de zile
- Contează cel mai mult cozile groase și fereastra lungă
  - frecvența reestimării contează doar la ferestre scurte
- Tipul de dovadă: comparație între proceduri
  - autorii arată că matricea de covarianță clasică respinge prea des în eșantioane mici

MFM_ch8_esr_case

E-backtesting: monitorizarea continuă a ES

- Wang, Wang & Ziegel (2026), *Management Science*: un backtest al ES care poate fi verificat zilnic fără să-și piardă validitatea
- Problema testelor ES de mai sus: valoarea p este validă pentru un eșantion fixat dinainte
  - un supraveghetor care repetă testul în fiecare zi și se oprește la prima respingere are o eroare de tip I peste 5%
- **E-value**: o statistică nenegativă cu media cel mult 1 sub H_0
  - o valoare mare este o dovadă împotriva lui H_0, așa cum este o valoare p mică
- Afirmațiile lucrării
  - testul cere doar prognozele (VaR, ES) și pierderile realizate, nu modelul care le-a produs
  - validitatea nu cere aproximări asimptotice și se păstrează la orice moment de oprire (anytime-valid)
  - pentru pierdere și VaR fixate, un ES mai mare, deci mai prudent, dă o statistică e_t mai mică
- Rezultatul empiric al lucrării: indicele NASDAQ Composite (NASDAQ: *National Association of Securities Dealers Automated Quotations*, bursa americană), ES 2,5% din AR(1)-GARCH(1,1), din 3 ian. 2005
  - inovații Normale, e-procesul GREM (*growth rate with a mixture*; dovada cumulată împotriva lui H_0, definit mai jos): pragul 2 este atins după 540 de zile de tranzacționare, pragul 10 după 713
  - inovații skewed-t cu ES majorat cu 10%: niciun prag atins pînă la 31 dec. 2021

E-backtesting: e-procesele ES 2,5% pe S&P 500 și BET



- Orizontal: zilele de evaluare, ian. 2007 – sept. 2026 (S&P 500: 4 959 zile; BET: 4 937 zile)
  - vertical: e-procesul GREM al fiecărui model, pe scală logaritmică; pornește de la 1 (nicio dovadă)
- Banda: pragurile 2, 5 și 10; atingerea pragului $1/\delta$ respinge prognozele cu o eroare de tip I de cel mult δ (50%, 20%, 10%)
- Peste 10^6, limita superioară a axei: pe S&P 500, Normal, Student-t și HS (4 sept. 2008, 6 oct. 2008, 15 oct. 2008); pe BET, Normal (7 ian. 2009)
- FHS și GARCH-EVT rămîn sub 10 pe S&P 500 și sub 2,3 pe BET

MFM_ch8_ebacktest ▶ Interactiv

E-backtesting: statistica e și replicarea pe datele cursului

- Statistica e pentru ES 2,5%: $e_t = \dfrac{(L_t - v_t)_+}{0{,}025\,(s_t - v_t)}$, cu $(x)_+ = \max(x, 0)$
  - L_t: pierderea zilei t; v_t: prognoza VaR 2,5%; s_t: prognoza ES 2,5% (în lucrare: z_t, r_t)
  - H_0: pentru fiecare t, $v_t = \mathrm{VaR}_t$, $s_t \geq \mathrm{ES}_t$ și $s_t > v_t$, cu prognoze cunoscute la $t-1$
  - sub H_0, media condiționată a lui e_t este cel mult 1; un VaR greșit o poate ridica peste 1 chiar cu ES corect
- E-procesul: $M_t = \prod_{s \leq t}(1 - \lambda_s + \lambda_s e_s)$, cu $\lambda_s \in [0; 0{,}5]$ ales doar din datele de pînă la $s-1$
  - sub H_0, M_t este un supermartingal nenegativ, iar inegalitatea lui Ville dă $P(\sup_t M_t \geq 1/\delta) \leq \delta$
- λ_s maximizează creșterea logaritmică a lui M_t (aproximarea Taylor din lucrare)
  - GREE (growth rate with empirical e-statistics): din e-values ale ultimelor 500 de zile
  - GREL (growth rate with empirical losses): din pierderile ultimelor 500 de zile, la prognoza de azi
  - GREM (growth rate with a mixture): media celor două e-procese, varianta recomandată de autori
- Replicarea din curs, față de lucrare
  - prognozele capitolului: șase modele, fereastră mobilă de 1 000 de zile (lucrarea: AR(1)-GARCH(1,1), 500 de zile)
  - S&P 500 și BET, 2007–2026 (lucrarea: NASDAQ Composite, 2005–2021); Bitcoin și EUR/RON au prognoze GARCH necalculate
  - GREE pornește în ian. 2007 fără e-values anterioare; lucrarea avea cinci ani de prognoze înainte de 2005

E-backtesting: dovezile apar din 2007, cu excepția FHS și GARCH-EVT

Model	S&P 500: max. M_t	prima zi cu $M_t \geq 10$	BET: max. M_t	prima zi cu $M_t \geq 10$
HS	$1{,}5 \cdot 10^{9}$	19 oct. 2007	237	10 oct. 2008
Normal	$1{,}5 \cdot 10^{21}$	9 aug. 2007	$9{,}3 \cdot 10^{10}$	21 ian. 2008
Student-t	$4{,}7 \cdot 10^{11}$	15 aug. 2007	$4{,}1 \cdot 10^{4}$	6 oct. 2008
GARCH-t	$2{,}2 \cdot 10^{4}$	6 iun. 2008	218	17 feb. 2009
FHS	8,55	—	2,16	—
GARCH-EVT	7,47	—	2,28	—

- S&P 500: Normal atinge pragul 2 la 26 iul. 2007 și pragul 10 la 9 aug. 2007; HS la 19 oct. 2007; GARCH-t la 6 iun. 2008
  - media lui e_t: 1,89 pentru GARCH-t și 1,33 pentru FHS (valoarea așteptată pentru un ES corect: 1)
- FHS și GARCH-EVT pe S&P 500: maximele 8,55 și 7,47 (29 sept. 2008); pe BET: 2,16 și 2,28
- HS pe BET: Z_2 nu respinge pe întreg eșantionul ($p = 0{,}134$), dar e-procesul atinge 10 la 10 oct. 2008 (maximul 237)
- ES majorat cu 10%: FHS și GARCH-EVT pe S&P 500 rămîn sub 5 (maximele 3,38 și 3,46); GARCH-t pe BET rămîne sub 2 (1,79)
  - rezultat empiric: pariurile λ_s se estimează din trecut, deci e-procesul GREM nu este garantat monoton în ES
- Tipul de dovadă: un test al ES condiționat de ipoteza că VaR 2,5% este corect, valid în orice zi în care este citit
  - nu ordonează modelele: pentru asta servesc backtesting-ul comparativ și MCS
- Limite: după o perioadă prudentă, M_t scade și are nevoie de timp ca să crească din nou; pragurile 2 și 5 au eroarea de tip I mare (50%, 20%)

O întrebare deschisă: acoperirea condiționată în criză

- Garanțiile VaR fără ipoteze de distribuție sînt **marginale**
  - conformal split cere schimbabilitate
  - ACI controlează doar media pe termen lung
  - secțiunea „Predicția conformală pentru riscul de coadă (pierderi extreme)"
- Întrebarea formală: rămîne rata de depășire în criză egală cu α?
  - cu un indicator de criză $C_t \in \mathcal{F}_{t-1}$ cunoscut ex ante
  - $C_t = 1$ într-o zi de criză, 0 altfel
- H_0: $P(L_t > \mathrm{VaR}_t \mid C_t = 1) = \alpha$ față de H_1: $P(L_t > \mathrm{VaR}_t \mid C_t = 1) \neq \alpha$
- Testată prin momentul $E[C_t(I_t - \alpha)] = 0$, cu I_t depășirea din ziua t
  - inferență HAC (robustă la heteroscedasticitate și autocorelație) sau bootstrap pe blocuri, deoarece depășirile din criză sînt grupate; apoi corecția Holm
- Miza: capitalul este necesar exact în zilele în care garanțiile marginale spun cel mai puțin

Bucla de cercetare cu un asistent AI

- Asistentul AI: un LLM precum Claude, ChatGPT sau Gemini
  - AI: artificial intelligence, inteligență artificială
  - LLM: large language model, model lingvistic mare
- Literatura, cu DOI-uri (*Digital Object Identifier*, identificatorul digital al unei publicații) verificate pe Crossref
- Replicăm întîi o cifră cunoscută
- Trei explicații rivale: scală întîrziată, formă a cozii schimbătoare, artefact al selecției regimului
- Pre-înregistrăm testele
- Cerem asistentului un raport critic, în stilul unui referent (prompturile: ▸ Anexă)

Mini-studiu de caz: scala GARCH și acoperirea în criză

- Scorul alternativ: pierderea standardizată $s_t = (L_t + \mu_t)/\sigma_t$, cu μ_t, σ_t prognozele GARCH-t ale mediei și volatilității
- H_0: $E[C_t(I_t - 0{,}05)] = 0$, adică rata de depășire a VaR 5% în zilele de criză este 5%
  - test t cu variantă HAC (Newey–West), deoarece depășirile din criză sînt grupate
  - corecția Holm (ajustarea în trepte a valorilor p) separat în fiecare experiment: patru piețe, respectiv piețele cu prognoze GARCH complete (n/a: prognoze GARCH necalculate)

Piața	Conformal split, calendarul complet			Scală GARCH, calendarul comun		
	Zile de criză	Rata	p Holm	Zile de criză	Rata	p Holm
S&P 500	713	7,29%	0,187	613	2,77%	0,003
BET	520	8,08%	0,187	506	3,95%	0,233
Bitcoin	264	7,95%	0,187	n/a	n/a	n/a
EUR/RON	245	11,43%	0,119	n/a	n/a	n/a

ai_discovery_case.py

Mini-studiu de caz: interpretarea rezultatului

- Două calendare: conformal split pe calendarul complet al secțiunii conformale
  - scala GARCH cere 250 de zile de calibrare: S&P 500 are 4 709 de zile, dintre care 613 de criză
  - pe aceste zile, rata conformal split în criză este 7,01%
- Conformal split: rate de criză peste 5% pe toate piețele, între 7,29% și 11,43%
  - cu inferența HAC și corecția Holm, nicio respingere la 5% (cel mai mic p Holm: 0,119, EUR/RON)
- Scala GARCH pe S&P 500: 2,77% în criză, sub 5% (p Holm 0,003), deci prea conservatoare în criză
- Scala GARCH pe BET: 3,95%, diferența față de 5% nu este semnificativă (p Holm 0,233)
- Nerespingerea nu dovedește acoperirea în criză: eșantioanele de criză sînt mici, iar inferența HAC este asimptotică
- Tipul de dovadă: comparație între proceduri

Decizia practică



Banca Națională a României, București

Foto: Britchi Mirela (2018); CC BY-SA 4.0; Wikimedia Commons

- **Ce aleg**: FHS ca model principal, GARCH-EVT ca alternativă, reexaminate pe măsură ce se acumulează dovezi
  - FHS are cel mai mic FZ0 mediu pe S&P 500; pe BET, FHS și GARCH-EVT sînt egale la trei zecimale
- **În ce condiții**
  - FHS și GARCH-EVT nu sînt respinse de Z_2 la 5% și rămîn în MCS de 90% pe S&P 500 și BET (tabloul complet: ▸ Anexă)
  - Dovezile vin din eșantioane lungi: un an de date are putere redusă
- **Ce rezultat m-ar face să renunț**
  - O respingere de către Z_2 la 5% sau ieșirea din MCS de 90% la evaluări fixate dinainte, cu ajustare pentru repetare
  - În monitorizarea zilnică: e-procesul ES peste 20, adică nivelul 5%
  - O respingere a momentului de criză $E[C_t(I_t - \alpha)] = 0$ (inferență HAC, Holm)
  - Deja împotrivă: DQ respinge toate modelele pe S&P 500, inclusiv FHS
- Un portofoliu românesc (BET și EUR/RON): Seminarul 8, C2

Verificarea înțelegerii și capitolul 9

Întrebări de verificare

- De ce respinge regula „$\mathrm{LR}_{uc} > 3{,}84$” un VaR 1% corect în aproximativ un an din zece?
- De ce poate fi respins un model cu numărul corect de depășiri?
- De ce poate fi testat ES, dar nu poate fi ierarhizat singur?
- Ce garantează ACI?
- Ce nu garantează ACI?

Capitolul 9: volatilitate realizată

- Măsurăm dispersia zilnică din randamente intrazilnice
- Evaluăm prognozele de volatilitate cu scoruri robuste, precum QLIKE, și cu MCS
  - QLIKE: quasi-likelihood, pierderea de tip cvasi-verosimilitate

Verificări numerice

1. Kupiec: $x = 3$ depășiri ale VaR 1% în $T = 500$ de zile
   - calculați $\hat{\pi}$ și LR_{uc}, apoi decideți la 5%
2. FZ0: o zi, $\alpha = 5\%$, $v = 2\%$, $e = 2{,}5\%$
   - calculați scorul pentru $L = 3\%$ și pentru $L = 1\%$, în procente și în zecimale
3. Puterea locală: $T = 1000$, $\alpha = 1\%$, rata reală $\pi = 2\%$
   - calculați λ și puterea, apoi comparați cu curba întreruptă din graficul puterii
4. Z_1 și Z_2: $T = 250$, $\alpha = 2{,}5\%$, cinci depășiri cu $L_t/\mathrm{ES}_t = 1{,}1;\ 0{,}9;\ 1{,}3;\ 1{,}0;\ 1{,}2$
   - calculați ambele statistici și explicați semnele lor

Răspunsuri, cu pașii intermediari: Anexă. Exerciții rezolvate pe datele cursului: Seminarul 8, A1, A4, B2

Pe ce date evaluăm? Patru piețe

- S&P 500: piață de acțiuni dezvoltată și lichidă
- BET (indicele Bursei de Valori București): piață emergentă, mai puțin lichidă
- Bitcoin: se tranzacționează în fiecare zi și are cozi foarte groase
- EUR/RON: cursul de referință BNR (Banca Națională a României), un curs cu flotare administrată
- De ce acestea: patru regimuri diferite de volatilitate și de coadă
  - pe aceste regimuri, aceleași teste se pot comporta diferit
- S&P 500, BET și Bitcoin: închideri zilnice
  - fiecare serie își păstrează propriul calendar de tranzacționare

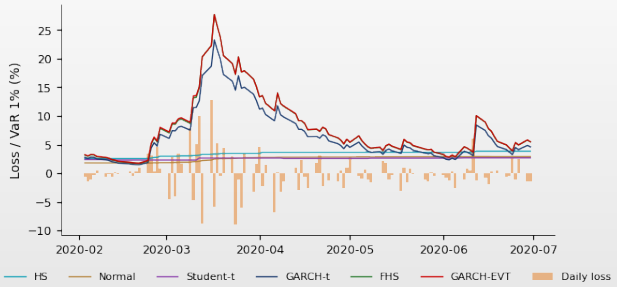
Cum evaluăm prognozele? Fereastra mobilă

- Fereastră mobilă de 1 000 de observații (circa patru ani de tranzacționare pe acțiuni)
  - destule date pentru GARCH și 100 de excese pentru coada GPD
- Fiecare prognoză folosește doar datele pînă în ziua precedentă (fără deplasarea de anticipare, *look-ahead bias*)
- Parametrii GARCH, Student-t și EVT (*Extreme Value Theory*, teoria valorilor extreme) se reestimează la fiecare 20 de zile (circa o lună de tranzacționare)
  - volatilitatea GARCH se actualizează zilnic
- Evaluarea începe la 01.01.2007 sau, la o serie mai scurtă, după primele 1 000 de observații
  - ultima zi: 18.09.2026
- S&P 500 și BET (din 2007): 4 959, respectiv 4 937 de zile de evaluare
- EUR/RON (curs din iulie 2005, deci evaluare din 2009): 4 352 de zile
- Bitcoin (date din septembrie 2014, deci evaluare din 2017): 3 384 de zile

Prognozele GARCH necalculate pe Bitcoin și EUR/RON

- O estimare GARCH(1,1)-t este **admisibilă** dacă optimizarea a convers și $\omega > 0$, $a, b \geq 0$, $\nu > 2$, $a + b < 1$
  - cu o marjă numerică de 10^{-8}
  - adică dispersia necondiționată $\omega/(1 - a - b)$ este finită
- La o reestimare neadmisibilă păstrăm ultimii parametri admisibili
  - volatilitatea se filtrează în continuare pe datele curente
- Fără nicio estimare admisibilă anterioară, prognozele GARCH-t, FHS și GARCH-EVT sînt **necalculate**
  - nu le înlocuim cu alte prognoze
- Bitcoin: primele 1 660 din 3 384 de zile de evaluare
  - primele 83 de reestimări au toate $a + b \geq 1 - 10^{-8}$
- EUR/RON: primele 20 din 4 352 de zile de evaluare (prima reestimare)
- Rezultatele GARCH pe eșantionul complet și MCS se raportează doar pe S&P 500 și BET
  - MCS cere toate cele șase modele
  - „n/a" în tabele, „—" pe grafice

Martie 2020: un test de stres neplanificat



- S&P 500, februarie–iunie 2020: pierderile zilnice (bare) și VaR 1% al celor șase modele
- Cea mai mare pierdere: 12,77%, la 16.03.2020
- Un VaR 1% corect are 1,04 depășiri așteptate în aceste 104 zile
- Depășiri, modele necondiționate: HS 12, Normal 17, Student-t 17
- Depășiri, modele condiționate: GARCH-t 3, FHS 2, GARCH-EVT 2
- Scala variabilă a contat mai mult decît coada: Student-t (coadă groasă, scală fixă) are 17 depășiri

MFM_ch8_var_forecasts

Puterea locală, calculată de mînă

- $\chi^2_1(\lambda)$ este legea lui $(Z+\sqrt{\lambda})^2$, cu $Z \sim N(0,1)$, iar $3{,}84 = 1{,}96^2$, deci
  - puterea $= P(|Z+\sqrt{\lambda}| > 1{,}96) = \Phi(\sqrt{\lambda} - 1{,}96) + \Phi(-\sqrt{\lambda} - 1{,}96)$, unde Φ este funcția de repartiție $N(0,1)$
- $T = 250$, $\pi = 2\%$: $\lambda = 2{,}525$, $\sqrt{\lambda} = 1{,}589$
  - puterea $= \Phi(-0{,}371) + \Phi(-3{,}549) = 0{,}355 + 0{,}0002 = 0{,}356$, adică cele 36% ale curbei întrerupte
- Puterea exactă este mai mică, 24%
  - testul acceptă între 1 și 6 depășiri, eveniment cu probabilitatea 75,7% cînd $\pi = 2\%$
  - cu 2,5 depășiri așteptate pe an, numărul este prea discret și prea asimetric pentru aproximarea Normală din spatele lui $\chi^2_1(\lambda)$

Independența pe întreg eșantionul S&P 500

- valorile p LR_{ind}, VaR 1%: HS, Normal și Student-t $<$0,001
  - GARCH-EVT 0,031; FHS 0,046; GARCH-t 0,095
- Independența este respinsă la 5% pentru toate modelele, cu excepția GARCH-t
- După o corecție Holm pentru cele șase teste, rămîn respinse doar HS, Normal și Student-t
- La GARCH-EVT, o depășire urmează unei depășiri cu probabilitatea 0,053, față de 0,011 după o zi fără depășire
- Testul comun CC nu respinge GARCH-EVT ($p = 0{,}058$): componenta de acoperire, corectă, diluează semnalul

MFM_ch8_coverage_tests

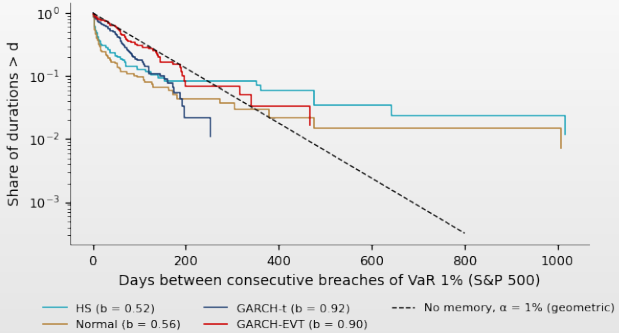
Teste bazate pe durate

- Ce testează: dacă distanțele dintre depășiri sînt cele ale unui proces fără memorie
  - folosește toate distanțele, nu doar perechile de zile consecutive
- Durata D_i: numărul de zile dintre depășirea $i-1$ și depășirea i
- Pentru un VaR corect, D_i urmează distribuția geometrică, cu media $1/\alpha$ (100 de zile pentru $\alpha = 1\%$)
- Christoffersen & Pelletier (2004): aproximăm duratele discrete prin densitatea Weibull
  $f(D; a, b) = a^b \, b \, D^{b-1} e^{-(aD)^b}$
- $a > 0$: rata (inversul scalei); $b > 0$: forma (alți a, b decît în GARCH)
- $b = 1$: cazul exponențial, fără memorie
- $b < 1$: duratele scurte sînt prea frecvente, adică depășirile se grupează

Testul de durată: statistica

- H_0: $b = 1$, cu a liber, față de H_1: $b \neq 1$
- $\mathrm{LR}_{dur} = 2\left[\ell(\hat{a}, \hat{b}) - \ell(\tilde{a}, 1)\right] \xrightarrow{d} \chi^2_1$, unde $\tilde{a}$ maximizează verosimilitatea cu $b = 1$
- $\ell(a, b) = \sum_i \left[C_i \ln S(D_i) + (1 - C_i) \ln f(D_i)\right]$, cu $S(D) = e^{-(aD)^b}$ funcția de supraviețuire
- $C_i = 1$ pentru prima și ultima durată, cenzurate: prima nu începe la o depășire, ultima nu se încheie la o depășire
- Testul nu verifică media $1/\alpha$: acoperirea rămîne în seama testului Kupiec
- La $T \geq 1\,000$ detectează gruparea mai bine decît testul de acoperire condiționată
  - la $T = 250$ nu (experimentul Monte Carlo de la sfîrșitul secțiunii)

Durata dintre depășiri



- Axe: durata d în zile (orizontal) și ponderea duratelor mai lungi decît d (vertical)
  - axa verticală are scală logaritmică
- Linia punctată: fără memorie (geometrică), cu $\alpha = 1\%$
- HS: $b = 0{,}52$, $p < 0{,}001$; durata mediană: 7 zile
- Normal: $b = 0{,}56$
- GARCH-t: $b = 0{,}92$, $p = 0{,}344$; durata mediană: 45 de zile
- GARCH-EVT: $b = 0{,}90$, $p = 0{,}317$
- Coada lungă a duratelor HS: ani fără depășiri după ce o criză lasă un VaR ridicat în fereastră

MFM_ch8_coverage_tests

Backtesting prin regresie: testul DQ

- Ce testează: dacă depășirile pot fi prezise din orice informație cunoscută ieri, nu doar din depășirea de ieri
- Sub calibrarea condiționată, $\mathrm{Hit}_t = I_t - \alpha$ este o diferență de martingal
  - $E[\mathrm{Hit}_t \mid \mathcal{F}_{t-1}] = 0$, $\mathrm{Var}(\mathrm{Hit}_t \mid \mathcal{F}_{t-1}) = \alpha(1-\alpha)$
- Regresăm $\mathrm{Hit}_t = X_t\beta + u_t$, cu $X_t = (1, \mathrm{Hit}_{t-1}, \ldots, \mathrm{Hit}_{t-4}, \mathrm{VaR}_t)$, și testăm $\beta = 0$ (Engle & Manganelli (2004))
- $\mathrm{DQ} = \dfrac{\hat{\beta}' X' X \hat{\beta}}{\alpha(1-\alpha)} = \dfrac{\mathrm{Hit}' X (X'X)^{-1} X' \mathrm{Hit}}{\alpha(1-\alpha)} \xrightarrow{d} \chi^2_6$
- X: matricea $T \times 6$ a regresorilor; Hit: vectorul depășirilor centrate
  - $\hat{\beta}$: estimatorul OLS (metoda celor mai mici pătrate)
  - u_t: eroarea
- DQ: suma pătratelor explicată de regresori, scalată cu dispersia sub H_0

Testul DQ: cazuri particulare și mărime

- Cu $X_t = 1$ obținem forma scor a testului Kupiec: $T(\hat{\pi} - \alpha)^2/(\alpha(1-\alpha))$
- Cu $X_t = (1, \mathrm{Hit}_{t-1})$ obținem un analog cu 2 grade de libertate al LR_{cc}
  - asimptotic echivalent, dar nu egal numeric
- Testarea doar a coeficientului lagului corespunde lui LR_{ind}
- Ce adaugă: putere împotriva memoriei dincolo de lagul 1 și a depășirilor care depind de nivelul VaR
- Limita: aproximarea χ^2_6 respinge prea des un model corect
  - experimentul Monte Carlo de la sfîrșitul secțiunii
- valorile p DQ din acest capitol sînt asimptotice; valorile critice trebuie simulate
  - sub procesul generator al datelor din experimentul Monte Carlo, cuantila de 95% a DQ pentru $T = 4\,959$ este 14,2, față de 12,59 pentru χ^2_6
  - FHS și GARCH-EVT pe S&P 500: frecvențe simulate 0,6% și 0,5%, față de valorile p asimptotice 0,014% și 0,011%
  - un reper sub acest proces didactic, nu valori p definitive pentru datele empirice

Mărimea și puterea testelor VaR: un experiment Monte Carlo

- DGP (procesul generator al datelor, adevărul simulat): GARCH(1,1)-t estimat pe S&P 500 1990–2026
  - $a = 0{,}102$, $b = 0{,}892$, $\nu = 6{,}1$
- Simulăm 1 000 de traiectorii și raportăm ratele de respingere (%) ale testelor la 5%, pentru VaR 1%

Prognoza	T	POF	Acop. cond.	Durate	DQ
Corect (mărime)	250	**9,7**	1,1	15,0	8,6
	1 000	6,4	3,3	8,0	11,0
	5 000	4,9	3,4	4,6	**8,8**
Normal-GARCH (putere)	250	**10,6**	4,8	14,8	18,8
	1 000	30,2	19,9	8,2	30,2
	5 000	83,7	77,2	7,3	70,7
HS, 1 000 de zile (putere)	250	51,5	17,0	**12,0**	32,5
	1 000	46,4	46,9	68,3	73,8
	5 000	57,1	86,6	100,0	99,5

MFM_ch8_size_power

Mărimea și puterea testelor VaR: cum citim tabelul

- Corect: parametrii și σ_t adevărați; rata de respingere este mărimea (țința: 5%)
- Normal-GARCH: σ_t adevărat, coadă Normală; rata de respingere este puterea
- HS, 1 000 de zile: fără dinamica volatilității; rata de respingere este puterea
- O valoare p nedefinită a testului de durată se numără ca nerespingere
  - nedefinit: sub 3 durate, în 28% din eșantioanele corecte la $T = 250$
- Eroarea standard Monte Carlo: 0,7 puncte procentuale la 5% și 1,6 la 50%
- În celula POF cu aldine (9,7%), eroarea standard este 0,9: mărimea depășește clar 5%

Mărimea și puterea testelor VaR: ce arată tabelul

- Într-un an, o coadă Normală pe volatilitatea corectă este aproape invizibilă
  - POF respinge în 10,6% din cazuri, cît mărimea (9,7%)
- Gruparea depășirilor HS este rar detectată de testul duratelor la $T = 250$ (12,0%)
- La $T = 1\,000$, testul duratelor (68,3%) și DQ (73,8%) detectează HS mult mai bine decît POF (46,4%)
- Testul de acoperire condiționată este conservator: mărimea 1,1% la $T = 250$
- DQ respinge prea des un model corect chiar la $T = 5\,000$ (8,8%): valorile lui critice ar trebui simulate
- La $T = 250$, niciun test nu are simultan mărime corectă și putere mare

Backtesting sub riscul de estimare

- Prognozele folosesc $\hat{\theta}$ estimat pe R zile, iar testul folosește P zile în afara eșantionului (*out-of-sample*)
  - depășirile $\hat{I}_t = 1\{L_t > \mathrm{VaR}_t(\hat{\theta})\}$ nu sînt Bernoulli(α) nici sub un model corect
- Corecția depinde de schema de prognoză (fixă, mobilă, recursivă) și crește cu P/R (Escanciano & Olmo (2010))
  - dezvoltarea în Anexă
- Schema fixă, 400 de traiectorii GARCH(1,1)-t, $R = 1000$
  - estimarea este admisibilă pe 324 dintre ele (rezultatele simulării)
- Capitolul folosește o fereastră mobilă ($R = 1000$, reestimare la fiecare 20 de zile)
  - un experiment cu schemă fixă nu măsoară mărimea acestor teste cu fereastră mobilă
- Remediul: reproducem estimarea în calibrare
  - de exemplu printr-un bootstrap parametric care reestimează modelul pe fiecare traiectorie simulată
  - Seminarul 8, B9, discută și selecția traiectoriilor admisibile

Du și Escanciano (2017): depășirile cumulate

- Ce testează: cît de adînc în coadă ajung pierderile, nu doar dacă depășesc VaR
  - folosește distribuția prognozată prin PIT
- PIT: $u_t = F_t(L_t)$, uniformă pe $(0,1)$ dacă distribuția prognozată F_t este corectă
- Depășirea cumulată $H_t = \frac{1}{\alpha}\,(u_t - (1-\alpha))\,1\{u_t > 1-\alpha\}$
  - sub H_0: $E[H_t] = \alpha/2$, $\mathrm{Var}(H_t) = \alpha(1/3 - \alpha/4)$
- Testul necondiționat: $U = \sqrt{T}\,(\bar{H} - \alpha/2)/\sqrt{\alpha(1/3-\alpha/4)} \xrightarrow{d} N(0,1)$, cu $\bar{H}$ media empirică a lui H_t
- Semnul lui U
  - $U > 0$: pierderi mai adînci decît s-a prognozat (risc subestimat)
  - $U < 0$: mai puțin adînci (risc supraestimat)
- Harta din sinteza testelor ES dă valoarea p bilaterală $p = 2\Phi(-|U|)$
  - testul unilateral la 5% respinge pentru $U > 1{,}645$
- Testul condiționat (Du & Escanciano (2017)): statistica Box–Pierce $C_5 = T\sum_{j=1}^{5}\hat{\rho}_j^2 \xrightarrow{d} \chi^2_5$, cu $\hat{\rho}_j$ autocorelațiile lui H_t

Teste VaR multinomiale: un test implicit pentru ES

- ES 2,5% este media cuantilelor dincolo de VaR 2,5%: testăm simultan mai multe niveluri VaR (Kratz, Lok & McNeil (2018))
- $N = 4$ niveluri $\alpha_j = 2{,}5\%\,(1-(j-1)/4)$: 2,5%, 1,875%, 1,25% și 0,625%
  - după PIT-ul u_t, fiecare zi cade într-una din 5 celule
- Statistica Pearson: $S = \sum_{j=0}^{4}(O_j - Tp_j)^2/(Tp_j) \xrightarrow{d} \chi^2_4$, cu O_j zilele observate și Tp_j zilele așteptate în celula j
- Probabilitățile celulelor: $p_0 = 97{,}5\%$, $p_1 = \cdots = p_4 = 0{,}625\%$
  - un S mare arată numere departe de cele așteptate

valoarea p	S&P 500	BET	Bitcoin	EUR/RON
HS	<0,001	0,725	0,564	0,004
Normal	<0,001	<0,001	<0,001	<0,001
Student-t	<0,001	0,024	0,015	0,172
GARCH-t	<0,001	<0,001	n/a	n/a
FHS	0,003	0,862	n/a	n/a
GARCH-EVT	0,378	0,853	n/a	n/a

- n/a: rezultat GARCH necalculat (slide-ul despre prognozele necalculate)

Frecvențe pe celule: S&P 500, GARCH-t

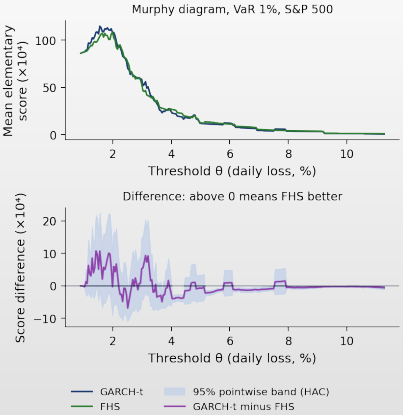
Interval PIT	Zile observate	Zile așteptate
[0; 0,975]	4 761	4 835,02
(0,975; 0,98125]	39	30,99
(0,98125; 0,9875]	46	30,99
(0,9875; 0,99375]	52	30,99
(0,99375; 1]	61	30,99

- Prima celulă este în afara cozii superioare a pierderilor, de 2,5%
- Fiecare dintre cele patru celule din coadă conține mai multe zile decît s-ar aștepta dacă modelul ar fi calibrat

Robustețea clasamentului la alegerea scorului: scoruri elementare

- Pentru o cuantilă există o infinitate de scoruri consistente
  - $S(v, L) = (1\{L \leq v\} - \tau)(g(v) - g(L))$, cu g crescătoare și $\tau = 1 - \alpha$
- Fiecare este o mixtură de **scoruri elementare** (Ehm, Gneiting, Jordan & Krüger (2016))
  - $S_\theta(v, L) = (1\{L < v\} - \tau)\,(1\{\theta < v\} - 1\{\theta < L\})$
  - $\theta \in \mathbb{R}$: un prag pe scala pierderii
- S_θ penalizează o prognoză doar cînd v și L cad de o parte și de alta a lui θ
- Pierderea pinball este $\int S_\theta \, d\theta$
- A este preferată lui B sub **orice** scor consistent dacă și numai dacă $\bar{S}_\theta(A) \leq \bar{S}_\theta(B)$ pentru orice θ
  - $\bar{S}_\theta$: scorul elementar mediu pe T zile
- **Diagrama Murphy**: $\bar{S}_\theta$ în funcție de θ
- Dacă curbele se intersectează, clasamentul depinde de scorul ales

Diagrama Murphy: GARCH-t față de FHS (1/2)



MFM_ch8_comparative

Diagrama Murphy: GARCH-t față de FHS (2/2)

- S&P 500, VaR 1%, 2007–2026, cu pragul θ între 0,9% și 11,3%
- Pierderea pinball medie ($\times 10^4$): GARCH-t 3,852, FHS 3,861, deci egalitate sub scorul obișnuit
- Sub $\theta = 5\%$ curbele se intersectează
  - FHS este mai bun pe 47% dintre praguri, iar banda conține zero pe 86%
  - peste $\theta = 5\%$, GARCH-t are scorul mai mic pe 89% dintre praguri
  - niciun model nu îl domină pe celălalt
- Benzile de 95% folosesc erori standard HAC și sînt punctuale, nu un test simultan de dominare
  - HAC: robuste la heteroscedasticitate și autocorelație
- Scorul se alege înainte de a vedea rezultatele

MFM_ch8_comparative

Giacomini–White: compararea metodelor de prognoză

- DM vizează acuratețea populațională a **modelelor**
- Cu ferestre mobile de 1 000 de zile și reestimări comparăm **metode de prognoză**
  - eroarea de estimare nu dispare cînd T crește
- Giacomini & White (2006): H_0: $E[d_t \mid \mathcal{F}_{t-1}] = 0$ pentru $d_t = \mathrm{FZ0}_{A,t} - \mathrm{FZ0}_{B,t}$
- Testul este valid și pentru modele imbricate, cu ferestre mobile de lungime finită (nu cu ferestre recursive)
- Ce adaugă: testează dacă diferența de pierdere poate fi prezisă din informația cunoscută la $t-1$
  - nu doar dacă diferența medie este zero

Giacomini–White: statistica și rezultatele

- Funcția test $h_{t-1} = (1, d_{t-1})'$, cunoscută la $t-1$; $Z_t = h_{t-1} d_t$
- $\mathrm{GW} = T\,\bar{Z}'\hat{\Omega}^{-1}\bar{Z} \xrightarrow{d} \chi^2_2$, $\hat{\Omega} = T^{-1}\sum_t Z_t Z_t'$ (prognoze la un pas)
- $\bar{Z}$: media lui Z_t; 2 grade de libertate, cîte unul pentru fiecare instrument
  - cu $h_{t-1} = 1$ obținem forma DM
- S&P 500, față de FHS
  - GARCH-EVT: $\mathrm{GW} = 6{,}71$ ($p = 0{,}035$)
  - GARCH-t: 7,60 ($p = 0{,}022$)
  - HS: 53,16 ($p < 0{,}001$)
- GARCH-EVT, la egalitate după DM, este respins condiționat
  - se respinge capacitatea predictivă condiționată egală pentru instrumentele h_{t-1}
- Respingerea, singură, nu arată cînd FHS are scorul mai mic
- Care instrument determină respingerea se vede doar testîndu-le separat

Backtesting comparativ în trei zone (Nolde și Ziegel)

- Comparăm modelul intern A cu un model standard S (aici HS) după pierderea FZ0 medie (Nolde & Ziegel (2017))
  - prin două teste unilaterale la 5%
- H_0^-: A este cel puțin la fel de bun ca S; H_0^+: A este cel mult la fel de bun
- **Verde**: respingem H_0^+ ($\mathrm{DM} < -1{,}645$); **roșu**: respingem H_0^- ($\mathrm{DM} > 1{,}645$); **galben**: niciuna
- De exemplu, $\mathrm{DM} = -1{,}8$ dă zona verde, deși $|\mathrm{DM}| < 1{,}96$
- Sarcina probei se inversează
  - banca trebuie să arate că modelul ei este mai bun, nu supraveghetorul că este mai slab

DM (A minus HS)	S&P 500	BET	Bitcoin	EUR/RON
GARCH-t	−4,21	−5,25	n/a	n/a
FHS	−4,69	−5,53	n/a	n/a
GARCH-EVT	−4,67	−5,57	n/a	n/a

- Culoarea indică zona statistică; n/a: rezultat GARCH necalculat (slide-ul despre prognozele necalculate)

Mulțimea de încredere a modelelor: statistica

- Pentru modelul i din mulțimea curentă $\mathcal{M}$: $d_{i\cdot} = \bar{L}_i - \frac{1}{|\mathcal{M}|}\sum_{j\in\mathcal{M}} \bar{L}_j$
  - pierderea medie față de media mulțimii
- $t_i = d_{i\cdot}/\widehat{\text{se}}(d_{i\cdot})$ și $T_{\max} = \max_{i\in\mathcal{M}} t_i$
  - H_0: toate modelele din $\mathcal{M}$ au aceeași pierdere așteptată
- $\widehat{\text{se}}$ și distribuția lui $T_{\max}$ sub H_0 provin dintr-un bootstrap circular pe blocuri de 10 zile
  - cu 1 000 de replicări
- Dacă respingem H_0, eliminăm modelul cu cel mai mare t_i și repetăm
- valoarea p MCS a unui model: cea mai mare valoare p a testelor pînă la eliminarea lui
- Cu 1 000 de replicări, cea mai mică valoare p rezolvabilă este de circa 0,001
  - un „0,000" pe grafic înseamnă $p \leq 0{,}001$

Studiu de caz (1/5): ce se estimează și sub ce ipoteze

- Estimator M: parametrii minimizează media empirică a unei pierderi de tip FZ0
- Teorie asimptotică sub specificare greșită (lucrarea, Teorema 1.4)
  - pachetele R esback și esreg
- Ipoteza 1, la ESR strict și ESR cu termen liber
  - ecuația cuantilei este corectă doar dacă VaR este o funcție afină de ES
  - de exemplu, modele de scală pură
- Estimatorul de covarianță folosit aici estimează anumite componente sub o ipoteză de locație și scală
  - covarianța sandwich sub specificare greșită nu cere, în general, această ipoteză
- Covarianța sandwich: $A^{-1}BA^{-1}$; A: curbura pierderii medii; B: dispersia gradientului ei
- Model de locație și scală: $r_t = \mu_t + \sigma_t z_t$, cu șocuri standardizate z_t i.i.d.

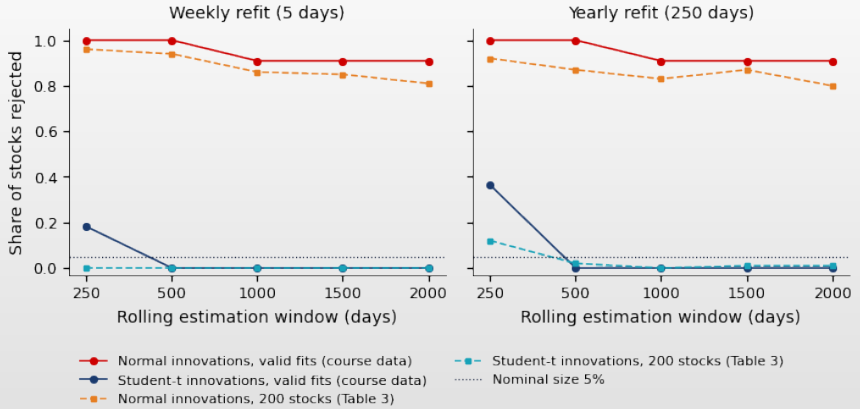
Studiu de caz (2/5): designul Monte Carlo al lucrării

- Lucrarea, Secț. 2: 10 000 de replicări, T de la 250 la 5 000
- EGARCH-t (GARCH exponențial, șocuri Student-t), calibrat pe S&P 500
- AR(1)-GARCH(1,1): medie autoregresivă de ordinul 1, dispersie GARCH
- Două modele GAS Student-t
  - GAS: generalised autoregressive score, scor autoregresiv generalizat
- Teste de comparație: McNeil & Frey (2000) și testele de calibrare din Nolde & Ziegel (2017)

Studiu de caz (3/5): mărimea și puterea în simulările lucrării

- Mărimea efectivă la nivelul nominal de 5%, EGARCH-t, $T = 250$ și $T = 1\,000$ (Tabelul 1)
  - ESR strict, covarianța robustă: 0,09, respectiv 0,05
  - ESR strict, covarianța clasică: 0,24, respectiv 0,11
  - testul simplu de calibrare condiționată: 0,29, respectiv 0,14
- Puterea (Fig. 3): testele reziduurilor de depășire aproape nu reacționează la două erori
  - o dinamică greșită a volatilității sau un nivel greșit $\tilde{\tau} \neq \tau$
  - ESR le detectează pe amîndouă
- O bancă ar putea raporta ES pentru un nivel mai puțin extrem fără a fi respinsă de testele mai vechi

Studiu de caz (4/5): Tabelul 3, replicat pe datele cursului



MFM_ch8_esr_case

Studiu de caz (5/5): Tabelul 3, replicat pe datele cursului

- Linia continuă: 11 acțiuni mari din S&P 500, ianuarie 2010 – august 2019 (2 432 de zile)
  - linia întreruptă: Bayer & Dimitriadis (2022), Tabelul 3
- GARCH(1,1) cu medie constantă; 25 de combinații fereastră–reestimare (grila 5×5)
  - ESR cu termen liber, cu esback
- Inovații Normale: ponderea respinsă este între 0,82 și 1,00 (în lucrare: 0,80–0,96)
- Student-t, ferestre de cel puțin 1 000 de zile: nicio acțiune respinsă (în lucrare: cel mult 0,01)
- Student-t, ferestre de 250 de zile: între 0,09 și 0,55 (în lucrare: 0,00–0,12)
- Cu 11 acțiuni, fiecare respingere adaugă $1/11 \approx 0{,}09$
  - eroarea standard a unei ponderi ajunge la $\sqrt{0{,}25/11} \approx 0{,}15$, un reper i.i.d.: acțiunile au șocuri de piață comune
- Rezultatul central se replică
  - diferența la fereastra de 250 de zile este mare chiar față de această eroare standard

MFM_ch8_esr_case

Lecturi și resurse pentru aprofundare

- Fissler & Ziegel (2016): VaR și ES sînt elicitabile împreună, baza pierderii FZ0
- Nolde & Ziegel (2017): backtesting comparativ și semaforul pentru VaR și ES
- Acerbi & Szekely (2014): teste ES fără ipoteze de model, inclusiv Z_2
- Prognoze VaR și ES bazate pe entropie: Pele, Lazar & Dufour (2017), Pele & Mazurencu-Marinescu-Pele (2019), Pele, Lazar & Mazurencu-Marinescu-Pele (2019); separabilitatea a două prognoze ES: Pele & Mazurencu-Marinescu-Pele (2026)
- Manuale: Franke, Härdle & Hafner (2019) (capitolele despre VaR și backtesting) și Härdle, Chen & Overbeck (2017)
- Quantinar: Machine learning in Financial Risk
- Cod și date: github.com/danpele/MFM/tree/main/Quantlets/Ch_08
  - S&P 500, BET, Bitcoin: închideri zilnice din datele cursului
  - EUR/RON: cursul de referință BNR

Înapoi

Riscul de estimare, dezvoltarea pentru schema fixă

- Schema fixă (Escanciano & Olmo (2010)): $\frac{1}{\sqrt{P}} \sum_t (\hat{I}_t - \alpha) = \frac{1}{\sqrt{P}} \sum_t (I_t - \alpha) + \sqrt{P/R}\, D' \sqrt{R}(\hat{\theta} - \theta_{\mathbf{0}}) + o_p(1)$
- Primul termen: depășirile centrate cu parametrii adevărați $\theta_{\mathbf{0}}$
  - al doilea: efectul erorii de estimare $\hat{\theta} - \theta_{\mathbf{0}}$
- $D = E\left[\partial P(L_t > \mathrm{VaR}_t(\theta) \mid \mathcal{F}_{t-\mathbf{1}})/\partial\theta\right]$ în $\theta_{\mathbf{0}}$: sensibilitatea probabilității de depășire la parametri
- Termenul de estimare adaugă o dispersie de ordinul P/R la $\sqrt{P}(\hat{\pi} - \alpha)$ și dispare doar dacă $P/R \to 0$
- Deci valorile critice $\chi^{\mathbf{2}}$ sînt prea mici cînd P/R este mare
- Ipoteze: $\mathrm{VaR}_t(\theta)$ netedă, $\hat{\theta}$ consistent cu rata $\sqrt{R}$, staționaritate
  - schemele cu fereastră mobilă și recursive schimbă forma corecției
- Remedii: dispersia corectată sau subeșantionarea din Escanciano & Olmo (2010)
  - ori un bootstrap parametric care reestimează modelul pe fiecare traiectorie simulată

Înapoi

Riscul de estimare: rezultatele simulării

- Schema fixă: 400 de traiectorii, $R = 1\,000$, testul Kupiec la 5%
- 324 de estimări admisibile
- 75 de estimări neidentificate numeric
- O estimare nu îndeplinește condiția de admisibilitate pentru dispersie finită
- Aici raportăm doar ratele cu parametrii adevărați, pe toate cele 400 de traiectorii
  - o rată pe cele 324 de traiectorii admisibile ar fi condiționată de succesul estimării
- Cu parametrii adevărați: 10,0% la $P = 250$, 5,2% la $P = 1\,000$, 4,0% la $P = 5\,000$
- Eroarea standard Monte Carlo la o probabilitate de respingere de 5%: circa 1,1 puncte procentuale

Înapoi

Principiul lui Osband și funcțiile de identificare

- Principiul lui Osband (Fissler & Ziegel (2016)): scorurile consistente se construiesc dintr-o funcție de identificare V
  - V are $E_F[V(x, L)] = 0$ doar în valoarea adevărată
- VaR: $V(v, L) = 1\{L \leq v\} - (1 - \alpha)$; pierderea pinball este integrala ei în v
- (VaR, ES): $V_1 = \alpha - 1\{L > v\}$ și $V_2 = e - v - \frac{1}{\alpha} 1\{L > v\}(L - v)$
  - mediile lor se anulează doar în perechea adevărată
- Mulțimile de nivel ale ES nu sînt convexe la mixturi
  - deci nu există o funcție de identificare doar în e (Seminarul 8, A8)
- FZ0 este membrul familiei FZ ale cărui diferențe nu depind de scala pierderilor (Patton, Ziegel & Chen (2019))

Înapoi

Monitorizarea secvențială a prognozelor de risc

- Reglementatorul verifică prognozele zilnic
  - repetarea unui test pe eșantion fix (de exemplu, Z_2 pe ferestre mobile) mărește probabilitatea a cel puțin unei alarme false
- Hoga & Demetrescu (2023): statistici de monitorizare pentru VaR și ES cu funcții-frontieră
  - funcțiile-frontieră țin sub nivelul nominal probabilitatea oricărei alarme false pe întregul orizont
- Față de semafor: dovezile se acumulează de la începutul monitorizării
  - nu dintr-o singură fereastră de 250 de zile
- Testul ES prin regresie (Bayer & Dimitriadis (2022)) din studiul de caz poate fi monitorizat la fel
  - ambele corectează limita semaforului, care evaluează un singur moment

Înapoi

Predicția conformală dincolo de schimbabilitate

- Barber, Candès, Ramdas & Tibshirani (2023): acoperirea fără schimbabilitate
  - abaterea acoperirii este mărginită de distanțe ponderate în variație totală
  - între vectorul scorurilor și același vector în care scorul de test este schimbat cu scorul de calibrare i
- Mărimea acestor distanțe în prezența grupării volatilității este o întrebare empirică
- Gibbs, Cherian & Candès (2025): acoperirea exactă condiționată pe o clasă de deplasări fixată dinainte este realizabilă
  - rămîne deschisă întrebarea ce clasă descrie o criză financiară
- Alegerea pasului ACI γ: Zaffran et al. (2022) agregă experți ACI pe o grilă de valori γ

Înapoi

Studiul de caz, alte rezultate principale

Rezultat	Sursa	Interpretare
Mărimile ESR strict și ESR auxiliar diferă cu cel mult 0,01 pentru fiecare proces și mărime a eșantionului, inclusiv AR(1)-GARCH cu $\phi = 0{,}5$	Tabelul 1	În aceste simulări, renunțarea la prognozele VaR aproape nu schimbă mărimea; puterea se compară separat (Fig. 2–3)
Student-t, fereastră de 250 de zile: ponderea de 0,12 provine din GARCH cu reestimare anuală; de la 500 de zile, GJR-GARCH-t și GARCH-t diferă cu cel mult 0,01	Tabelul 3	Efectul de levier din ecuația dispersiei aproape nu schimbă ponderile de respingere cînd fereastra este lungă

Înapoi

Studiul de caz, limite și extinderi

- Limită precizată de autori: ESR auxiliar și ESR strict sînt doar bilaterale (lucrarea, Secț. 1.3)
- Limită precizată de autori: covariança clasică respinge prea des în eșantioane mici
- Replicare (Seminarul 8, C3): ESR strict (bilateral) și ESR cu termen liber (unilateral)
  - aplicați-le prognozelor ES 2,5% ale celor șase modele pe cele patru piețe
  - comparați cu Z_2 și McNeil–Frey
- Extindere: replicarea Tabelului 3 cu GJR-GARCH, mai multe acțiuni și perioada de după august 2019
- Întrebare deschisă: mărimea ESR cu termen liber bilateral la $T = 250$ (Tabelul 1, EGARCH-t)
  - 0,04 cu covariança robustă
  - 0,15 cu cea clasică
- Mărimea unilaterală este raportată doar pentru $T = 2\,500$, 0,02 (Tabelul 2): care este ea pe un an?
- Poate un bootstrap al statisticii t să o mențină la 5% fără a pierde puterea din Fig. 4?

Înapoi

Tabloul de bord pe patru piețe

- Fiecare celulă: rata de depășire a VaR 1% în % / valoarea p pentru Z_2 / valoarea p MCS (FZ0)
  - $\leq$0,001: la rezoluția celor 1 000 de replicări bootstrap

Model	S&P 500	BET	Bitcoin	EUR/RON
HS	1,67 / <0,001 / ≤0,001	1,17 / 0,134 / ≤0,001	0,77 / 0,863 / n/a	0,51 / 1,000 / n/a
Normal	2,70 / <0,001 / ≤0,001	1,90 / <0,001 / ≤0,001	1,74 / 0,001 / n/a	1,06 / 0,973 / n/a
Student-t	2,00 / 0,002 / ≤0,001	1,58 / 0,003 / ≤0,001	0,71 / 0,625 / n/a	0,85 / 0,920 / n/a
GARCH-t	1,77 / <0,001 / 0,010	1,46 / <0,001 / 0,096	n/a / n/a / n/a	n/a / n/a / n/a
FHS	1,23 / 0,209 / 1,000	1,20 / 0,236 / 1,000	n/a / n/a / n/a	n/a / n/a / n/a
GARCH-EVT	1,15 / 0,145 / 0,211	0,97 / 0,215 / 0,966	n/a / n/a / n/a	n/a / n/a / n/a

- n/a: prognoză GARCH necalculată și, de aceea, MCS necalculat (slide-ul despre prognozele necalculate)

înapoi

Răspunsuri la verificările numerice (1/2)

1. $\hat{\pi} = 0{,}006$; $\ln L(\alpha) = 497 \ln 0{,}99 + 3 \ln 0{,}01 = -18{,}81$, $\ln L(\hat{\pi}) = 497 \ln 0{,}994 + 3 \ln 0{,}006 = -18{,}34$
   - $\mathrm{LR}_{uc} = 2(-18{,}34 - (-18{,}81)) = 0{,}94 < 3{,}84$, $p = 0{,}33$: nu respingem; erau așteptate 5 depășiri
2. $L = 3\%$: $\frac{3-2}{0{,}05 \cdot 2{,}5} + \frac{2}{2{,}5} + \ln 2{,}5 - 1 = 8 + 0{,}8 + 0{,}916 - 1 = 8{,}716$; $L = 1\%$: $0{,}8 + 0{,}916 - 1 = 0{,}716$
   - în zecimale se schimbă doar $\ln e$, cu $\ln 0{,}01 = -4{,}605$: 4,111, respectiv $-3{,}889$
   - diferența dintre cele două zile rămîne aceeași

Verificări numerice

Răspunsuri la verificările numerice (2/2)

3. $\lambda = 1000 \cdot 0{,}01^{\mathbf{2}}/(0{,}01 \cdot 0{,}99) = 10{,}10$, $\sqrt{\lambda} = 3{,}178$
   - puterea $= \Phi(1{,}218) + \Phi(-5{,}138) = 0{,}888$ (al doilea termen este sub $10^{-\mathbf{6}}$): cele 88,8% ale curbei întrerupte pentru $T = 1000$
   - puterea exactă: 78%
4. $\sum_t L_t I_t / \mathrm{ES}_t = 5{,}5$, $N_T = 5$, $T\alpha = 6{,}25$
   - $Z_{\mathbf{1}} = 1 - 5{,}5/5 = -0{,}10$: în zilele de depășire, pierderile sînt în medie cu 10% mai adînci decît prognoza ES
   - $Z_{\mathbf{2}} = 1 - 5{,}5/6{,}25 = 0{,}12 > 0$: numărul de depășiri, mai mic decît cele 6,25 așteptate, compensează această adîncime
   - valoarea este departe de pragul de 5%, $-0{,}70$

◂ Verificări numerice

Bucla de cercetare (1/3): prompturi

- Prompturi (instrumente: Semantic Scholar, Elicit, Google Scholar; verificăm fiecare DOI pe Crossref)
  - `Enumeră lucrările evaluate de specialiști, din 2019 încoace, despre predicția conformală sub dependență serială, cu DOI, garanția exactă și ipotezele.`
  - `Dă trei explicații rivale pentru depășirile excesive în criză ale VaR conformal split și ce implică fiecare pentru ratele pe regimuri.`
  - `Joacă rolul unui referent critic: enumeră cele mai puternice motive pentru care acest efect de criză este un artefact și testul care ar tranșa fiecare motiv.`

Înapoi

Bucla de cercetare (2/3): verificările cercetătorului

- **Deplasarea de anticipare: informație din viitor**
  - modificăm observațiile de după ziua $t-1$ și confirmăm că VaR_t, scala și indicatorul de criză nu se schimbă
- **Depășiri dependente**
  - testele Christoffersen și de durată sau un bootstrap pe blocuri al șirului depășirilor
- **Puterea**
  - 613 de zile de criză dau 6,1 depășiri așteptate ale VaR 1%
  - o rată reală de 2% este detectată cu probabilitatea 57% de un test Binomial exact bilateral la 5% (cu depășiri presupuse i.i.d.)

Înapoi

Bucla de cercetare (3/3): alte verificări

- **Testarea multiplă**: 4 piețe $\times$ 2 niveluri $\times$ 3 metode $=$ 24 de teste
  - Holm controlează FWER (rata erorii pe familia de teste)
  - Benjamini–Hochberg controlează FDR (rata descoperirilor false) sub independență sau dependență pozitivă (PRDS: positive regression dependency on a subset)
  - pentru dependență arbitrară: Benjamini–Yekutieli
- **Teoreme enunțate greșit**
  - de exemplu, „ACI garantează acoperirea condiționată"
- **Cuantile decalate**
  - $k = \lceil (n+1)(1-\alpha) \rceil$ nu este implicit în `np.quantile`

înapoi

Punct de plecare pentru proiect, acoperirea în criză

- Întrebare: poate o prognoză VaR 1% și ES 2,5% fără ipoteze de distribuție să atingă acoperirea condiționată de criză?
- Întrebare: se menține această acoperire pe toate cele patru piețe, cu un test pre-înregistrat?
- Replicați mai întîi
  - ratele în criză ale VaR 5% conformal split (S&P 500 7,29%, EUR/RON 11,43%)
  - ponderea VaR 1% infinit la ACI (13,4% din zilele de evaluare S&P 500)
- Livrabil: o pre-înregistrare (ipoteze, teste, corecții, putere) datată înainte de prima rulare
- Livrabil: o metodă candidată, de exemplu Gibbs, Cherian & Candès (2025) cu regimul de volatilitate drept clasă de condiționare
  - teste VaR pe regimuri
  - ES testat cu testul ES prin regresie (Bayer & Dimitriadis (2022)) sau cu Z_2
- Livrabil: codul și un raport de două pagini cu toate specificațiile încercate
  - o critică, în stilul unui referent, a propriului rezultat
- Consemnați fiecare prompt care a influențat o decizie în `AI_USE.md`
  - fiecare eroare a asistentului în `AI_ERRORS.md`: github.com/danpele/MFM/tree/main/project

Înapoi

Acronime folosite în acest material (1/3)

- **ACI**: Adaptive Conformal Inference — *inferență conformală adaptivă*
- **AI**: Artificial Intelligence — *inteligență artificială*
- **AR**: AutoRegressive (model) — *model autoregresiv, ca în AR(1)-GARCH*
- **BCBS**: Basel Committee on Banking Supervision — *Comitetul de la Basel pentru Supraveghere Bancară*
- **BET**: Bucharest Exchange Trading (index) — *indicele de referință al Bursei de Valori București*
- **BIS**: Bank for International Settlements — *Banca Reglementelor Internaționale*
- **BNR**: Banca Națională a României
- **CC**: Conditional Coverage (test) — *testul de acoperire condiționată*
- **COVID-19**: COronaVIrus Disease 2019 — *boala provocată de coronavirus, 2019*
- **DGP**: Data-Generating Process — *procesul generator al datelor*
- **DM**: Diebold–Mariano (test) — *testul Diebold–Mariano de comparare a prognozelor*
- **DOI**: Digital Object Identifier — *identificatorul digital al unei publicații*
- **DQ**: Dynamic Quantile (test) — *testul cuantilei dinamice*
- **EGARCH**: Exponential GARCH — *GARCH exponențial*
- **ES**: Expected Shortfall — *pierderea așteptată în coadă*
- **ESR**: Expected Shortfall Regression (backtest) — *testul ES prin regresie*
- **EUR**: Euro — *moneda euro*
- **EVT**: Extreme Value Theory — *teoria valorilor extreme*
- **FDR**: False Discovery Rate — *rata descoperirilor false*
- **FHS**: Filtered Historical Simulation — *simulare istorică filtrată*

Acronime folosite în acest material (2/3)

- **FRTB**: Fundamental Review of the Trading Book — *revizuirea fundamentală a portofoliului de tranzacționare*
- **FWER**: Family-Wise Error Rate — *probabilitatea de cel puțin o eroare de tip I în familia de teste*
- **FZ**: Fissler–Ziegel (family of scoring functions for VaR and ES) — *familia Fissler–Ziegel de funcții de scor pentru VaR și ES*
- **FZ0**: Fissler–Ziegel loss with zero-homogeneous score differences — *funcția de pierdere Fissler–Ziegel cu diferențe de scor omogene de grad zero*
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity — *heteroscedasticitate condiționată autoregresivă generalizată*
- **GARCH-EVT**: GARCH filter combined with Extreme Value Theory for the tail — *filtru GARCH combinat cu teoria valorilor extreme pentru coadă*
- **GAS**: Generalised Autoregressive Score (model) — *model cu scor autoregresiv generalizat*
- **GJR-GARCH**: Glosten–Jagannathan–Runkle GARCH — *GARCH asimetric Glosten–Jagannathan–Runkle*
- **GPD**: Generalised Pareto Distribution — *distribuția Pareto generalizată*
- **GREE**: Growth Rate with Empirical e-statistics (e-backtest) — *rata de creștere cu e-statistici empirice (e-backtest)*
- **GREL**: Growth Rate with Empirical Losses (e-backtest) — *rata de creștere cu pierderi empirice (e-backtest)*
- **GREM**: Growth Rate with a Mixture (of GREE and GREL) — *rata de creștere cu o mixtură (media GREE și GREL)*
- **GW**: Giacomini–White (test) — *testul Giacomini–White al capacității predictive condiționate*

Acronime folosite în acest material (3/3)

- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors) — *erori standard robuste la heteroscedasticitate și autocorelație*
- **HS**: Historical Simulation — *simulare istorică*
- **IMA**: Internal Models Approach — *abordarea bazată pe modele interne*
- **LLM**: Large Language Model — *model lingvistic mare*
- **LR**: Likelihood Ratio — *raportul de verosimilitate*
- **LRV**: Long-Run Variance — *dispersia pe termen lung*
- **MCS**: Model Confidence Set — *mulțimea de încredere a modelelor*
- **NASDAQ**: National Association of Securities Dealers Automated Quotations — *bursa americană NASDAQ (cotații automate ale Asociației Naționale a Dealerilor de Valori Mobiliare)*
- **OLS**: Ordinary Least Squares — *metoda celor mai mici pătrate*
- **P&L**: Profit and Loss — *profitul și pierderea*
- **PIT**: Probability Integral Transform — *transformarea integrală a probabilității*
- **POF**: Proportion Of Failures (test) — *testul proporției de depășiri*
- **QLIKE**: Quasi-LIKElihood (loss function) — *funcția de pierdere de tip cvasi-verosimilitate*
- **RON**: Romanian leu — *leul românesc*
- **S&P 500**: Standard & Poor's 500 (index) — *indicele celor 500 de companii americane mari*
- **VaR**: Value-at-Risk — *valoarea expusă la risc*

Bibliografie (1/4)

- Acerbi, C., Szekely, B. (2014). Backtesting expected shortfall. MSCI Inc., December 2014.
- Angelopoulos, A.N., Bates, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. arXiv:2107.07511.
- Barber, R.F., Candès, E.J., Ramdas, A., Tibshirani, R.J. (2023). Conformal prediction beyond exchangeability. *Annals of Statistics*, 51(2).
- Basel Committee on Banking Supervision (1996). Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements. Bank for International Settlements.
- Basel Committee on Banking Supervision (2019). Minimum capital requirements for market risk (rev. February 2019). Bank for International Settlements.
- Bayer, S., Dimitriadis, T. (2022). Regression-based expected shortfall backtesting. *Journal of Financial Econometrics*, 20(3), 437–471.
- Christoffersen, P.F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862.
- Christoffersen, P., Pelletier, D. (2004). Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics*, 2(1), 84–108.
- Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.

Bibliografie (2/4)

- Du, Z., Escanciano, J.C. (2017). Backtesting expected shortfall: accounting for tail risk. *Management Science*, 63(4), 940–958.
- Ehm, W., Gneiting, T., Jordan, A., Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings. *Journal of the Royal Statistical Society, Series B*, 78(3), 505–562.
- Engle, R.F., Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4), 367–381.
- Escanciano, J.C., Olmo, J. (2010). Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics*, 28(1), 36–51.
- Fissler, T., Ziegel, J.F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, 44(4).
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). Statistics of Financial Markets: An Introduction, 5th ed. Springer.
- Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578.
- Gibbs, I., Candès, E. (2021). Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34; arXiv:2106.00170.
- Gibbs, I., Cherian, J.J., Candès, E.J. (2025). Conformal prediction with conditional guarantees. *Journal of the Royal Statistical Society, Series B*, 87(4), 1100–1126.

Bibliografie (3/4)

- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494), 746–762.
- Hansen, P.R., Lunde, A., Nason, J.M. (2011). The model confidence set. *Econometrica*, 79(2), 453–497.
- Härdle, W.K., Chen, C.Y.-H., Overbeck, L. (eds.) (2017). Applied Quantitative Finance, 3rd ed. Springer.
- Hoga, Y., Demetrescu, M. (2023). Monitoring value-at-risk and expected shortfall forecasts. *Management Science*, 69(5), 2954–2971.
- Kratz, M., Lok, Y.H., McNeil, A.J. (2018). Multinomial VaR backtests: a simple implicit approach to backtesting expected shortfall. *Journal of Banking & Finance*, 88, 393–407.
- Kupiec, P.H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), 73–84.
- Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R.J., Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523), 1094–1111.
- McNeil, A.J., Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3–4), 271–300.
- Newey, W.K., West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.

Bibliografie (4/4)

- Nolde, N., Ziegel, J.F. (2017). Elicitability and backtesting: perspectives for banking regulation. *Annals of Applied Statistics*, 11(4).
- Patton, A.J., Ziegel, J.F., Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2), 388–413.
- Pele, D.T., Lazar, E., Dufour, A. (2017). Information entropy and measures of market risk. *Entropy*, 19(5), 226.
- Pele, D.T., Lazar, E., Mazurencu-Marinescu-Pele, M. (2019). Modeling expected shortfall using tail entropy. *Entropy*, 21(12), 1204.
- Pele, D.T., Mazurencu-Marinescu-Pele, M. (2019). Using high-frequency entropy to forecast Bitcoin's daily value at risk. *Entropy*, 21(2), 102.
- Pele, D.T., Mazurencu-Marinescu-Pele, M. (2026). Finite-sample precision limits for expected shortfall forecast comparisons. *Mathematics*, 14(13), 2316.
- Wang, Q., Wang, R., Ziegel, J. (2026). E-backtesting. *Management Science*, 72(6), 4952–4973.
- Zaffran, M., Féron, O., Goude, Y., Josse, J., Dieuleveut, A. (2022). Adaptive conformal predictions for time series. *Proceedings of the 39th International Conference on Machine Learning*; arXiv:2202.07282.