








Statistica piețelor financiare

Seminarul 6: Selecția modelului și managementul riscului



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

Acronime folosite în acest material

- **AD**: Anderson–Darling (test) — testul Anderson–Darling
- **AI**: Artificial Intelligence (inteligență artificială)
- **AIC**: Akaike Information Criterion (criteriul informațional Akaike)
- **BET**: Bucharest Exchange Trading (index) — indicele de referință al Bursei de Valori București
- **BIC**: Bayesian Information Criterion (criteriul informațional bayesian)
- **EDF**: Empirical Distribution Function (funcția de repartiție empirică)
- **EODHD**: EOD Historical Data (market-data provider) — furnizor de date de piață
- **ES**: Expected Shortfall (pierderea așteptată în coadă)
- **GED**: Generalised Error Distribution (distribuția erorii generalizate)
- **KS**: Kolmogorov–Smirnov (test) — testul Kolmogorov–Smirnov
- **LR**: Likelihood Ratio (raportul de verosimilitate)
- **ML**: Maximum Likelihood (verosimilitate maximă)
- **NIG**: Normal Inverse Gaussian (distribution) — distribuția Normală invers Gaussiană
- **QQ**: Quantile–Quantile (plot) — graficul cuantilă–cuantilă
- **SFM**: Statistics of Financial Markets — Statistica piețelor financiare (cursul)

Întrebarea de azi și structura seminarului

- Întrebarea seminarului
  - **dintre mai multe distribuții candidate pentru randamentele zilnice, pe care o alegem și de unde știm că este suficient de bună pentru VaR 1%?**
- Rezultate de obținut
  - răspunsurile la exercițiile A
  - tabelul sau figura cerute la fiecare exercițiu B
  - un răspuns scurt și argumentat la C1; erorile găsite în răspunsul AI din C2 și răspunsul corectat
  - frazele de interpretare cerute la fiecare exercițiu
- Seminarul are rol de exercițiu: nu se predă nicio lucrare și nu se acordă note
- Notebook-ul de azi: deschideți notebook-ul seminarului în Google Colab
  - fiecare cerință indică secțiunea din notebook
- Exercițiile, tipurile și ordinea lor: harta exercițiilor (slide-ul următor)
- Contact: prof. Daniel Traian Pele, danpele@ase.ro

Harta exercițiilor

Cerința	Întrebarea	Tipul	Model
A1, A2	ce model are cel mai mic AIC și BIC și cît de puternice sînt dovezile?	Rezolvat, Propus	A1
A3, A4	este modelul mai general dintre două modele imbricate semnificativ mai bun?	Rezolvat, Propus	A3
A5, A6	trece un eșantion mic un test KS, cu parametri cunoscuți și cu parametri estimați?	Rezolvat, Propus	A5
A7, A8	VaR 1% pentru două modele; sînt depășirile prea multe sau prea puține?	Rezolvat, Propus	A7
B1, B2	care distribuție se potrivește cel mai bine randamentelor BET, DAX și Bitcoin?	Rezolvat, Propus	B1
B3, B4	trec cele mai bune modele testele de concordanță? sînt două zile din seria TLV erori de date?	Rezolvat, Propus	B3
B5, B6	ce model estimează corect VaR 1% în afara eșantionului?	Rezolvat, Propus	B5
C1, C2	se schimbă în timp cel mai bun model? ce este greșit într-un răspuns AI?	Propus	B1, A3

- **[Rezolvat]**: rezolvarea completă în slide-uri și în notebook, un model de urmat; **[Propus]**: îl rezolvați dumneavoastră, după model; soluția se discută la seminar
- Partea A: calcule pe hîrtie; Partea B: date reale, în notebook, fiecare cerință încheindu-se cu o întrebare de interpretare; Partea C: întrebări deschise și verificarea unui răspuns AI

Datele folosite

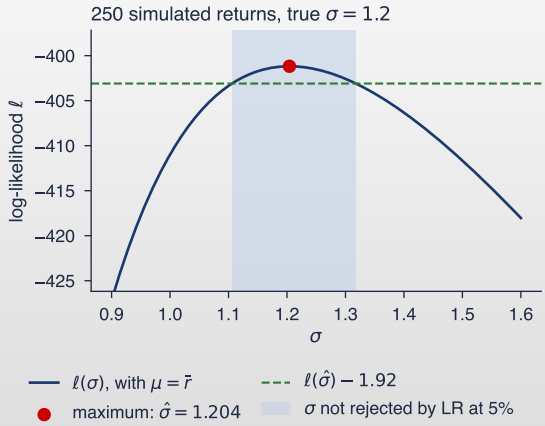
Seria	Sursa	Prețul folosit	Perioada
BET, S&P 500, DAX	EODHD	închidere	2000–2026
Bitcoin	EODHD	închidere, 7 zile pe săptămînă	2014–2026
Banca Transilvania (TLV)	EODHD	închidere ajustată	2010–2026

- Randamente logaritmice zilnice în %: $r_t = 100\,(\ln P_t - \ln P_{t-1})$; weekendurile și închiderile repetate din zilele libere se elimină (cu excepția Bitcoin)
- În notebook: `returns('bet')`, `fit_model('NIG', x)`; nu este nevoie de cont sau de cheie

Noțiuni necesare azi (1/17): log-verosimilitatea

- Randamentele $r_1, \ldots, r_n$ i.i.d. (independente și identic distribuite), cu densitatea $f(r; \theta)$
  - θ: parametrii modelului (de exemplu μ, σ); f: densitatea modelului
- **Log-verosimilitatea**: $\ell(\theta) = \sum_t \ln f(r_t; \theta)$
  - cît de plauzibile sînt randamentele observate pentru parametrii θ; valoarea mai mare este preferată
  - estimarea ML (maximum likelihood, verosimilitate maximă) $\hat{\theta}$ o maximizează; $\ell(\hat{\theta})$: valoarea maximă
- Modelul Normal: $\ell(\hat{\theta}) = -\frac{n}{2}[\ln(2\pi\hat{\sigma}^2) + 1]$
  - cu $\hat{\sigma}^2 = \frac{1}{n}\sum_t (r_t - \bar{r})^2$ și $\bar{r}$ media de selecție

Noțiuni necesare azi (2/17): log-verosimilitatea pe un grafic



- Modelul Normal cu μ fixat la $\bar{r}$: ℓ în funcție de σ

$$\ell(\sigma) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_t (r_t - \bar{r})^2$$

  - 250 de randamente simulate, cu $\sigma = 1{,}2$ adevărat
- Punctul de maxim (roșu) este estimarea ML
  - $\hat{\sigma} = 1{,}204$: valoarea pentru care datele sînt cele mai plauzibile
- Un vîrf ascuțit înseamnă o estimare precisă
  - valorile σ cu $\ell \geq \ell(\hat{\sigma}) - 1{,}92$ (zona colorată, între 1,11 și 1,32) formează un interval de 95%
  - $1{,}92 = 3{,}84/2$: jumătate din valoarea critică de 5% a lui $\chi^2(1)$ (testul LR urmează mai jos)

Noțiuni necesare azi (3/17): candidații

- Candidații și numărul lor de parametri k
  - Normală (2); Student-t (3: gradele de libertate ν, locație, scală)
  - skewed-t (4: adaugă asimetria λ; $\lambda = 0$ dă Student-t)
  - GED, generalised error distribution (distribuția erorii generalizate) (3: forma β; $\beta = 2$ dă distribuția Normală)
  - amestec Normal cu două componente (5); NIG, normal inverse Gaussian (4)
- Rolul parametrilor de formă
  - un ν mic sau un β mic: cozi mai groase; $\lambda < 0$: o coadă stîngă mai lungă
- Cu cît cozile datelor sînt mai groase, cu atît modelul Normal pierde mai mult în ℓ

Noțiuni necesare azi (4/17): densitățile candidaților

- **Student-t** cu locația m, scala s și ν grade de libertate

$$f(r) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\, s}\left(1 + \frac{(r-m)^2}{\nu s^2}\right)^{-\frac{\nu+1}{2}}$$

  - Γ: funcția gamma; densitatea scade ca $|r|^{-(\nu+1)}$, o coadă de tip putere cu $\alpha = \nu$
  - varianța $s^2\nu/(\nu-2)$ pentru $\nu > 2$; cînd $\nu \to \infty$, distribuția Normală

- **GED** cu forma β

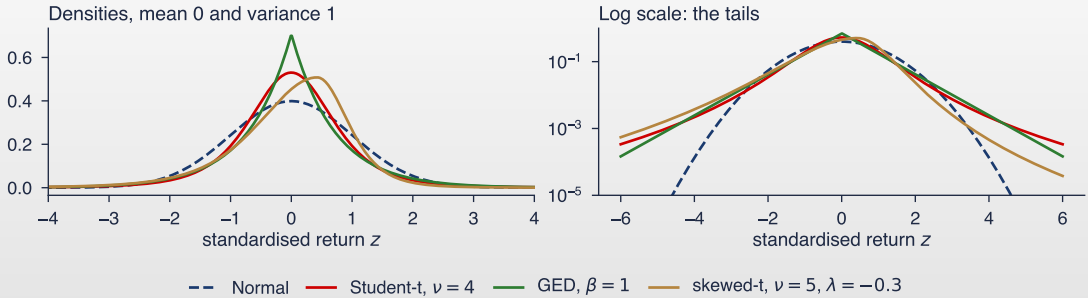
$$f(r) = \frac{\beta}{2s\,\Gamma(1/\beta)} \exp\left(-\left|\frac{r-m}{s}\right|^{\beta}\right)$$

  - $\beta = 2$: distribuția Normală; $\beta = 1$: Laplace, cu vîrf ascuțit; $\beta < 2$: cozi mai groase decît ale distribuției Normale

- **Amestecul Normal**: $f(r) = \pi\,\varphi(r; \mu_1, \sigma_1^2) + (1-\pi)\,\varphi(r; \mu_2, \sigma_2^2)$
  - $\varphi(\cdot\,; \mu, \sigma^2)$: densitatea Normală; cinci parametri $\pi, \mu_1, \sigma_1, \mu_2, \sigma_2$

- Skewed-t și NIG: un parametru suplimentar pentru asimetrie; densitățile lor se evaluează numeric

Noțiuni necesare azi (5/17): candidații pe un grafic



- Toate cele patru densități au media 0 și varianța 1: diferă prin vîrf și prin cozi
- Stînga: Student-t și GED au un vîrf mai înalt și mai ascuțit; skewed-t cu $\lambda < 0$ are coada stîngă mai lungă
- Dreapta (scară logaritmică): coada distribuției Normale scade ca o parabolă, celelalte mult mai lent

SFM_ch6_seminar

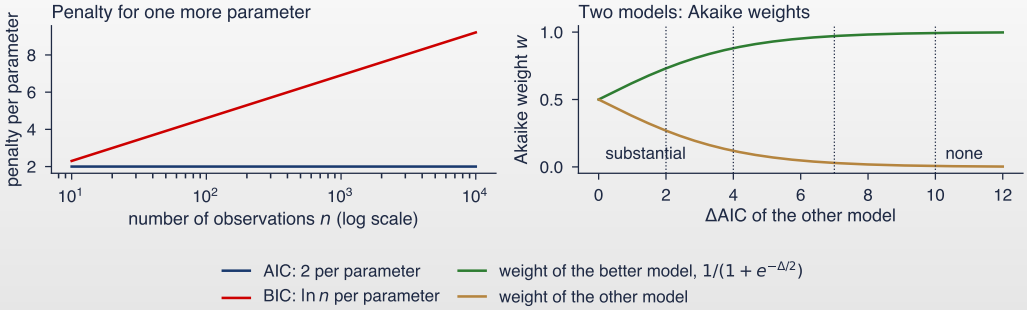
Noțiuni necesare azi (6/17): AIC și BIC

- $\mathrm{AIC} = -2\ell(\hat{\theta}) + 2k$ (Akaike information criterion, criteriul informațional Akaike)
  - ajustarea (-2ℓ) plus o penalizare de 2 pentru fiecare parametru
- $\mathrm{BIC} = -2\ell(\hat{\theta}) + k \ln n$ (Bayesian information criterion, criteriul informațional bayesian)
  - penalizarea $\ln n$ pe parametru; mai mare decît 2 cînd $\ln n > 2$, adică $n > 7$
- Interpretarea
  - valoarea mai mică este preferată; contează doar diferențele calculate pe aceleași date
  - AIC nu este un test: nu există p-value și nici „semnificație”

Noțiuni necesare azi (7/17): ΔAIC și ponderile Akaike

- $\Delta_i = \mathrm{AIC}_i - \min_j \mathrm{AIC}_j$
  - distanța modelului i față de cel mai bun model; $\Delta = 0$ pentru cîștigător
  - reguli practice Burnham & Anderson (2004): $\Delta \leq 2$ sprijin substanțial; 4–7 considerabil mai puțin; > 10 practic niciunul
- Ponderea Akaike: $w_i = e^{-\Delta_i/2} / \sum_j e^{-\Delta_j/2}$
  - ponderea dovezilor în favoarea modelului i; ponderile însumează 1
  - exemplu: $\Delta = 0$ și $\Delta = 2$ dau $w \approx 0{,}73$ și $0{,}27$
- VaR mediat: $\sum_i w_i \, \mathrm{VaR}_i$
  - VaR-ul fiecărui model, ponderat cu ponderea lui Akaike

Noțiuni necesare azi (8/17): penalizări și ponderi pe un grafic



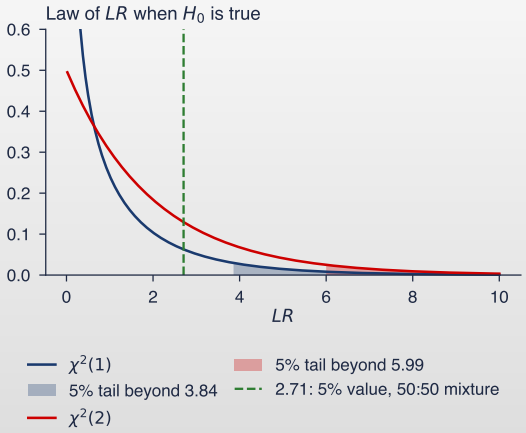
- Stînga: penalizarea pentru un parametru în funcție de n: 2 pentru AIC, $\ln n$ pentru BIC (6,9 pentru $n = 1\,000$)
- Dreapta: două modele ale căror AIC diferă cu Δ: $w_{\textbf{mai bun}} = 1/(1 + e^{-\Delta/\mathbf{2}})$
  - liniile punctate la $\Delta = 2, 4, 7, 10$: pragurile regulilor practice

SFM_ch6_seminar

Noțiuni necesare azi (9/17): testul raportului de verosimilitate

- Modelul 0 este **imbricat** în modelul 1 dacă este modelul 1 cu unii parametri fixați
  - exemplu: Student-t = skewed-t cu $\lambda = 0$
- $LR = 2[\ell_{\mathbf{1}}(\hat{\theta}_{\mathbf{1}}) - \ell_{\mathbf{0}}(\hat{\theta}_{\mathbf{0}})] \approx \chi^{\mathbf{2}}(k_{\mathbf{1}} - k_{\mathbf{0}})$ sub $H_{\mathbf{0}}$ (Wilks)
  - $H_{\mathbf{0}}$: modelul mai mic, 0, este adevărat; $k_{\mathbf{1}} - k_{\mathbf{0}}$: numărul parametrilor fixați
  - valori critice la 5%: $\chi^{\mathbf{2}}_{\mathbf{0,95}}(1) = 3{,}84$; $\chi^{\mathbf{2}}_{\mathbf{0,95}}(2) = 5{,}99$; respingem $H_{\mathbf{0}}$ peste ele
- **Frontiera**: distribuția Normală ca submodel al Student-t înseamnă $1/\nu = 0$, marginea spațiului parametrilor
  - LR urmează atunci un amestec 50:50 de 0 și $\chi^{\mathbf{2}}(1)$ Self & Liang (1987): valoarea critică la 5% 2,71
  - p-value-ul este jumătate din cel calculat cu $\chi^{\mathbf{2}}(1)$
- Perechi neimbricate (Student-t și GED): testul LR nu se aplică; folosiți AIC sau testul Vuong
  - Vuong: $V = \sqrt{n}\,\bar{m}/s_m$, $m_t = \ln f_{\mathbf{1}}(r_t) - \ln f_{\mathbf{2}}(r_t)$, cu $\bar{m}$, s_m media și abaterea standard a lui m_t; $|V| > 1{,}96$: un model este mai aproape

Noțiuni necesare azi (10/17): legea hi-pătrat a statisticii LR

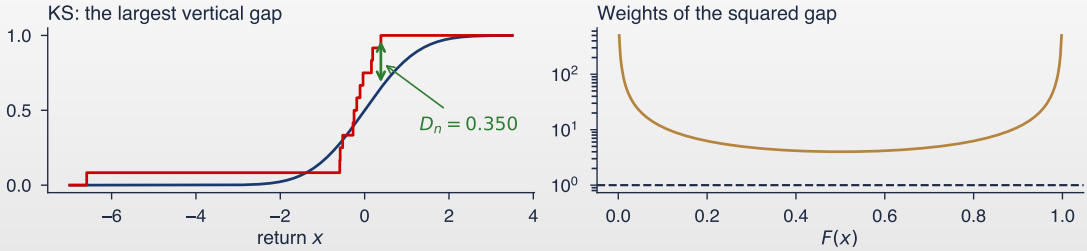


- $\chi^2(d)$, cu $d = k_1 - k_0$: legea lui LR sub H_0
  - legea unei sume de d pătrate ale unor variabile $N(0,1)$ independente; media ei este d
- Zonele colorate: cozile de 5% de dincolo de valorile critice de pe slide-ul anterior
- p-value-ul: aria din dreapta valorii LR observate
- Frontiera (verde): în jumătate din cazuri $LR = 0$, deci valoarea de 5% este cuantila de 90% a lui $\chi^2(1)$

Noțiuni necesare azi (11/17): teste de concordanță și valori critice

- EDF (empirical distribution function, funcția de repartiție empirică): $F_n(x) = \frac{1}{n}\#\{t : r_t \leq x\}$
  - proporția randamentelor cel mult egale cu x; se compară cu funcția de repartiție a modelului, F
- **KS** (Kolmogorov–Smirnov): D_n = cea mai mare distanță pe verticală dintre F_n și F
  - $D_n = \max_i \max\{i/n - u_{(i)},\ u_{(i)} - (i-1)/n\}$, cu $u_{(i)} = F(r_{(i)})$ și $r_{(i)}$ al i-lea cel mai mic randament
- **CvM** (Cramér–von Mises) și **AD** (Anderson–Darling) integrează pătratul distanței
  - AD îl ponderează cu $1/[F(1-F)]$, mare acolo unde F este aproape de 0 sau 1: pune accentul pe cozi
- Valori critice KS la 5%
  - F complet cunoscută: $n = 5$: 0,563; $n = 6$: 0,519; n mare: $1{,}358/\sqrt{n}$
  - media și varianța estimate din aceleași date (Lilliefors): $n = 5$: 0,354; $n = 6$: 0,333
  - alte modele: bootstrap parametric (simulăm din modelul estimat, reestimăm, recalculăm statistica)

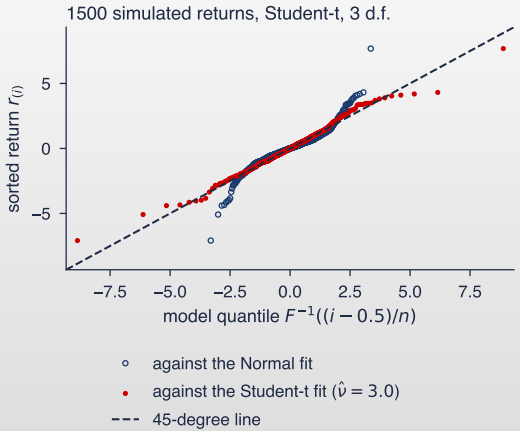
Noțiuni necesare azi (12/17): KS, CvM și AD pe un grafic



- Stînga: EDF a 12 randamente față de funcția de repartiție $N(0,1)$; $D_n = \sup_x |F_n(x) - F(x)|$ este cea mai mare distanță pe verticală (verde)
- Dreapta: CvM, $W^2 = n \int (F_n - F)^2 \, dF$, și AD, $A^2 = n \int \frac{(F_n - F)^2}{F(1-F)} \, dF$
  - pentru toate trei: valori mari înseamnă o ajustare slabă; respingem peste valoarea critică

SFM_ch6_seminar

Noțiuni necesare azi (13/17): citirea unui grafic QQ



- Graficul QQ al celor n randamente față de modelul ajustat F

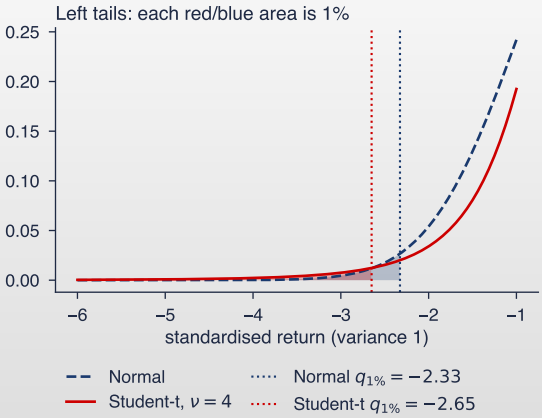
$$\left(F^{-1}\left(\frac{i-0{,}5}{n}\right),\ r_{(i)}\right), \quad i = 1, \ldots, n$$

  - $r_{(i)}$: al i-lea cel mai mic randament
- Albastru, ajustarea Normală: o formă de S; randamentele extreme depășesc cuantilele Normale: cozi mai groase
- Roșu, ajustarea Student-t: punctele rămîn pe dreaptă, cu excepția cîtorva dintre cele mai extreme

Noțiuni necesare azi (14/17): VaR, depășiri, în afara eșantionului

- $\mathrm{VaR}_{\mathbf{1\%}} = -q_{\mathbf{1\%}}$: pierderea zilnică depășită cu probabilitatea 1%
  - Normală: $\mathrm{VaR}_{\mathbf{1\%}} = -(\mu - 2{,}326\sigma)$; ES 2,5% (expected shortfall): pierderea medie dincolo de $q_{\mathbf{2{,}5\%}}$
  - ES 2,5% Normal: $-\mu + \sigma\varphi(z_{\mathbf{0{,}025}})/0{,}025$, cu φ densitatea $N(0,1)$
- **Depășire**: o zi cu $r_t < -\mathrm{VaR}$; model corect: $x \sim \mathrm{Binomial}(n; 0{,}01)$
  - testul Kupiec: $LR_{uc} = -2[\ell_{\mathbf{0}} - \ell_{\mathbf{1}}]$, comparat cu $\chi^{\mathbf{2}}(1)$
  - $\ell_{\mathbf{0}} = (n-x)\ln 0{,}99 + x \ln 0{,}01$ (rata 1%); $\ell_{\mathbf{1}} = (n-x)\ln(1-\hat{p}) + x\ln\hat{p}$ (rata observată $\hat{p} = x/n$)
- **În afara eșantionului**: estimăm pe datele de pînă la un moment dat și evaluăm pe datele ulterioare
  - **scorul logaritmic**: media lui $\ln f(r_t; \hat{\theta})$ pe zilele de test (valoarea mai mare este preferată); evaluează întreaga densitate
  - doar pentru VaR: pierderea pinball de pe ultimul slide al acestei secțiuni

Noțiuni necesare azi (15/17): VaR pentru două densități



- VaR pentru un model de locație și scală

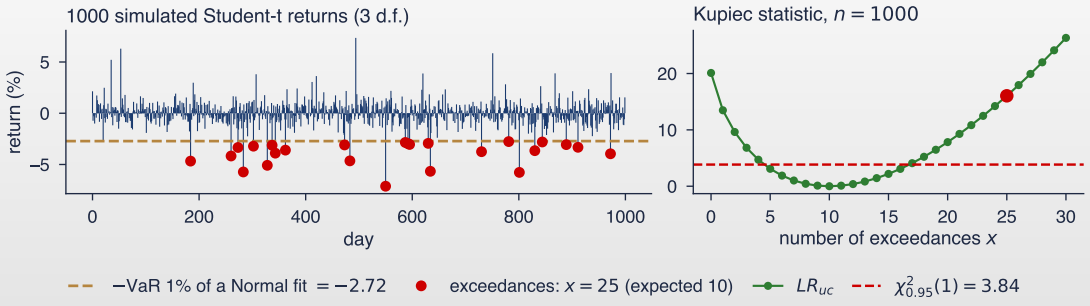
$$\mathrm{VaR}_{1\%} = -(\mu + \sigma\, q^*_{1\%})$$

  - σ: abaterea standard; $q^*_{1\%}$: cuantila de 1% a legii standardizate (varianța 1)

- Aceeași varianță, cozi diferite
  - distribuția Normală: $q^*_{1\%} = -2{,}33$; Student-t cu 4 grade de libertate: $q^*_{1\%} = -2{,}65$
  - pentru date Student-t, VaR Normal este cu circa 12% prea mic

SFM_ch6_seminar

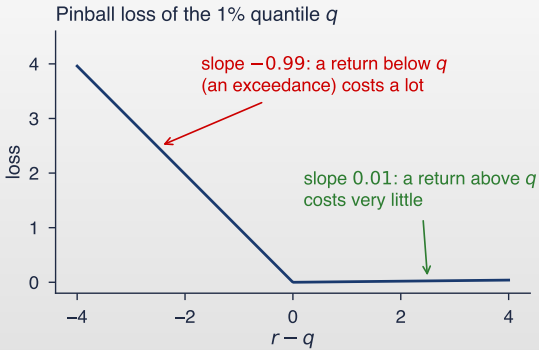
Noțiuni necesare azi (16/17): depășiri și testul Kupiec pe un grafic



- Stînga: 1 000 de randamente Student-t cu VaR 1% al unei ajustări Normale: 25 de depășiri în loc de cele 10 așteptate
- Dreapta: LR_{uc} pentru fiecare număr posibil de depășiri x
  - 0 pentru $x = 10$; peste 3,84 pentru $x \leq 4$ sau $x \geq 17$; aici $LR_{uc} = 16{,}0$: VaR Normal este respins

SFM_ch6_seminar

Noțiuni necesare azi (17/17): pierderea pinball



- **Pierderea pinball** a unei prognoze q a cuantilei de 1%

$$\rho(r, q) = (0{,}01 - 1\{r < q\})(r - q)$$

  - $1\{\cdot\} = 1$ dacă respectiva condiție este îndeplinită, 0 altfel; se face media pe zilele de test; valoarea mai mică este preferată
- Un randament sub q costă $0{,}99\,|r - q|$; un randament peste q doar $0{,}01\,(r - q)$
- Cuantila adevărată de 1% minimizează pierderea așteptată: pierderea pinball recompensează un VaR corect, nu unul prudent

SFM_ch6_seminar

A1: patru modele pentru S&P 500 [Rezolvat]

Cerință

- S&P 500, ultimele $n = 1000$ de zile de tranzacționare (2022–2026); log-verosimilitățile maximizate:
- Normală: $\ell = -1400{,}9$, $k = 2$
- Student-t: $\ell = -1314{,}7$, $k = 3$
- skewed-t: $\ell = -1313{,}8$, $k = 4$
- GED: $\ell = -1322{,}6$, $k = 3$
- 1. Calculați AIC și BIC ($\ln 1000 = 6{,}908$) pentru fiecare model.
- 2. Calculați ΔAIC și ponderile Akaike.
- 3. Precizați ce model alege fiecare criteriu și cît de puternice sînt dovezile.
- Raportați: un tabel cu patru rînduri și două fraze.

Rezolvare

- 1. AIC: Normală 2805,8; Student-t 2635,4; skewed-t 2635,6; GED 2651,2. BIC, în aceeași ordine: 2815,6; 2650,1; 2655,2; 2665,9
- 2. ΔAIC: 170,4; 0,0; 0,2; 15,8. $e^{-\Delta/2}$: 0,000; 1,000; 0,905; 0,000. Ponderi: 0,000; 0,525; 0,475; 0,000
- 3. AIC: Student-t și skewed-t la egalitate ($\Delta = 0{,}2$); BIC: Student-t, cu un avans de 5,1 puncte; modelul Normal și GED nu au niciun sprijin.

◂ Înapoi ▸ Lista exercițiilor

A2: patru modele pentru Bitcoin [Propus]

Rezolvarea se discută la seminar

Cerință

- Bitcoin, ultimele $n = 1000$ de zile (2023–2026); log-verosimilitățile maximizate: Model: A1.
- Normală: $\ell = -2328{,}0$, $k = 2$
- Student-t: $\ell = -2262{,}0$, $k = 3$
- NIG: $\ell = -2258{,}4$, $k = 4$
- amestec Normal: $\ell = -2262{,}3$, $k = 5$
- 1. Calculați AIC, BIC, ΔAIC și ponderile Akaike.
- 2. Precizați dacă AIC și BIC aleg același model.
- 3. Explicați de ce amestecul Normal, cu cei mai mulți parametri, nu cîștigă.
- Raportați: un tabel cu patru rînduri și două fraze.

Înapoi Lista exercițiilor

A3: este necesară asimetria? Este suficientă distribuția Normală? [Rezolvat]

Cerință

- Folosiți log-verosimilitățile din A1 (S&P 500, $n = 1000$).
- 1. Testați Student-t (H_0: $\lambda = 0$) față de skewed-t cu un test al raportului de verosimilitate la 5%.
- 2. Testați distribuția Normală (H_0: $\beta = 2$) față de GED.
- 3. Comparați decizia LR de la pasul 1 cu verdictul AIC din A1.
- Raportați: două valori LR, două decizii și o frază.

Rezolvare

- 1. $LR = 2(-1313{,}8 - (-1314{,}7)) = 1{,}8 < 3{,}84$, $p = 0{,}18$: nu respingem $\lambda = 0$
- 2. $LR = 2(-1322{,}6 - (-1400{,}9)) = 156{,}6$, $p < 0{,}001$: respingem distribuția Normală
- 3. Rezultatele sînt concordante: după AIC, cele două modele t sînt la egalitate, iar testul LR nu arată un cîștig semnificativ adus de λ. Pentru perechi imbricate, AIC preferă modelul mai general exact atunci cînd $LR > 2$ pentru fiecare parametru în plus.

◂ Înapoi ▸ Lista exercițiilor

A4: un test la frontieră [Propus]

Rezolvarea se discută la seminar

Cerință

- Folosiți log-verosimilitățile din A2 (Bitcoin, $n = 1000$). Model: A3.
- 1. Calculați LR pentru distribuția Normală ca submodel al Student-t.
- 2. Explicați de ce distribuția statisticii sub H_0 nu este $\chi^2(1)$ și dați valoarea critică corectă la 5%.
- 3. Spuneți dacă un test LR poate compara Student-t cu NIG.
- Raportați: o valoare LR, o valoare critică și două fraze.

Înapoi Lista exercițiilor

A5: KS pe cinci randamente [Rezolvat]

Cerință

- Randamente zilnice (%): $-2{,}1$; $-0{,}4$; $0{,}3$; $0{,}9$; $1{,}6$.
- 1. Calculați D_5 față de $N(0,1)$ și comparați-l cu valoarea critică la 5% 0,563.
- 2. Estimați $\hat{\mu}$ și $\hat{\sigma}$ (numitorul n), calculați D_5 față de $N(\hat{\mu}, \hat{\sigma}^2)$ și comparați-l cu valoarea Lilliefors 0,354.
- 3. Explicați de ce valoarea critică Lilliefors este mai mică.
- Raportați: două valori D_5, două decizii, o frază.

Rezolvare

- 1. $u_{(i)} = \Phi(r_{(i)})$: 0,018; 0,345; 0,618; 0,816; 0,945; cea mai mare distanță $D_5 = 0{,}218 < 0{,}563$: nu respingem
- 2. $\hat{\mu} = 0{,}060$, $\hat{\sigma} = 1{,}266$; $z_{(i)}$: $-1{,}71$; $-0{,}36$; 0,19; 0,66; 1,22; $u_{(i)}$: 0,044; 0,358; 0,575; 0,747; 0,888; $D_5 = 0{,}175 < 0{,}354$: nu respingem
- 3. F estimată este apropiată de date, deci D_n este mai mic sub H_0; prin urmare, și valoarea critică trebuie să fie mai mică.

Înapoi Lista exercițiilor

A6: KS cu o pierdere mare [Propus]

Rezolvarea se discută la seminar

Cerință

- Randamente zilnice (%): $-3{,}0$; $-0{,}6$; $-0{,}2$; $0{,}1$; $0{,}5$; $1{,}0$. Model: A5.
- 1. Calculați $D_{\mathbf{6}}$ față de $N(0,1)$; valoarea critică la 5% este 0,519.
- 2. Calculați $D_{\mathbf{6}}$ față de $N(\hat{\mu}, \hat{\sigma}^{\mathbf{2}})$; valoarea Lilliefors este 0,333.
- 3. Ziua de -3% are $\Phi(-3) = 0{,}001$; explicați de ce KS aproape nu reacționează la ea, în timp ce AD reacționează puternic.
- Raportați: două valori $D_{\mathbf{6}}$, două decizii, o frază.

Înapoi Lista exercițiilor

A7: VaR 1% pentru două modele [Rezolvat]

Cerință

- Randamente zilnice (%): Normală cu $\mu = 0{,}05$, $\sigma = 1{,}2$; Student-t cu $\nu = 4$, locația 0,06, scala 0,8; $t_{0{,}01}(4) = -3{,}747$.
- 1. Calculați VaR 1% pentru fiecare model.
- 2. O bancă folosește VaR Normal și observă 20 de depășiri în 1000 de zile; calculați statistica Kupiec.
- 3. Decideți la 5% și interpretați.
- Raportați: două valori VaR, o statistică, o decizie.

Rezolvare

- 1. Normală: $-(0{,}05 - 2{,}326 \times 1{,}2) = 2{,}742\%$; Student-t: $-(0{,}06 - 3{,}747 \times 0{,}8) = 2{,}938\%$
- 2. $\ell_0 = 980 \ln 0{,}99 + 20 \ln 0{,}01 = -101{,}95$; $\ell_1 = 980 \ln 0{,}98 + 20 \ln 0{,}02 = -98{,}04$; $LR_{uc} = -2(\ell_0 - \ell_1) = 7{,}83$
- 3. $7{,}83 > 3{,}84$ ($p = 0{,}0051$): respingem; numărul observat este dublul celor 10 depășiri așteptate: VaR Normal este prea mic, așa cum era de așteptat pentru o coadă subțire.

Înapoi Lista exercițiilor

A8: prea puține depășiri și ES [Propus]

Rezolvarea se discută la seminar

Cerință

- O altă bancă observă 4 depășiri ale VaR 1% în 1000 de zile. Model: A7.
- 1. Calculați statistica Kupiec și decideți la 5%.
- 2. Explicați de ce și prea puține depășiri înseamnă un eșec.
- 3. Calculați ES 2,5% al modelului Normal din A7 ($z_{0,025} = -1{,}960$, $\varphi(1{,}960) = 0{,}0584$).
- Raportați: o statistică, o decizie, o valoare ES și o frază.

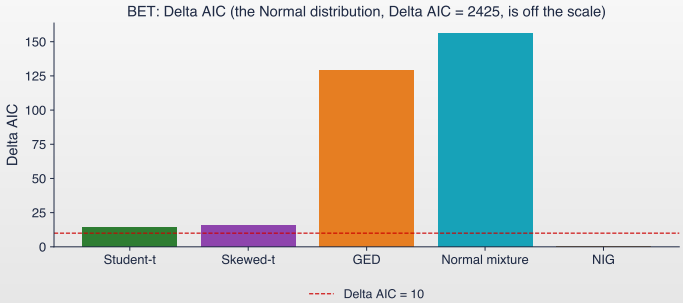
Înapoi Lista exercițiilor

B1: ce distribuție pentru BET? [Rezolvat]

- **Întrebarea**: care dintre cei șase candidați se potrivește cel mai bine randamentelor zilnice BET și este necesar parametrul de asimetrie?
- **Date**: închiderile BET, 2000–2026, randamente logaritmice zilnice în % ($n = 6\,670$)
- **Cerințe**
  1. Estimați cei șase candidați prin metoda verosimilității maxime și raportați $\ell(\hat{\theta})$ pentru fiecare.
  2. Calculați AIC, BIC, ΔAIC și ponderile Akaike.
  3. Testați Student-t față de skewed-t cu un test al raportului de verosimilitate.
  4. Interpretare: ce spun câștigătorul și testul LR despre cozile și asimetria BET?
- **Raportați**: un tabel cu șase rînduri, un test LR și două fraze
- Notebook: secțiunea B1

◂ Înapoi ▸ Lista exercițiilor

B1: rezolvare (1/2) [Rezolvat]



- Câștigător: NIG pe ambele criterii; Student-t $\hat{\nu} = 2{,}43$; skewed-t $\hat{\lambda} = -0{,}009$
  - $LR = 0{,}35$, $p = 0{,}555$: asimetria nu este necesară

SFM_ch6_seminar

B1: rezolvare (2/2) [Rezolvat]

	$\ell(\hat{\theta})$	AIC	BIC	ΔAIC	ΔBIC	w
Normală	−11 702	23 407	23 421	2425,2	2411,6	0,000
Student-t	−10 495	20 996	21 017	14,4	7,6	0,001
skewed-t	−10 495	20 998	21 025	16,0	16,0	0,000
GED	−10 553	21 111	21 132	129,1	122,3	0,000
amestec Normal	−10 564	21 138	21 172	155,8	162,7	0,000
NIG	−10 487	20 982	21 009	0,0	0,0	0,999

- Interpretare: cozi foarte groase (ν aproape de 2) și nicio asimetrie semnificativă; distribuția Normală este în urmă cu mii de puncte AIC

SFM_ch6_seminar

B2: DAX și Bitcoin [Propus]

- **Întrebarea**: aleg DAX și Bitcoin aceeași distribuție ca BET? Model: B1.
- **Date**: DAX din 2000, Bitcoin din 2014 (7 zile pe săptămînă), randamente logaritmice zilnice în %
- **Cerințe**
  1. Repetați pașii 1–3 din B1 pentru fiecare serie.
  2. Precizați dacă AIC și BIC aleg același model pentru fiecare serie.
  3. Interpretare: de ce ar putea diferi cel mai bun model între un indice bursier și un activ cripto?
- **Raportați**: două tabele și trei fraze
- Notebook: secțiunea B2

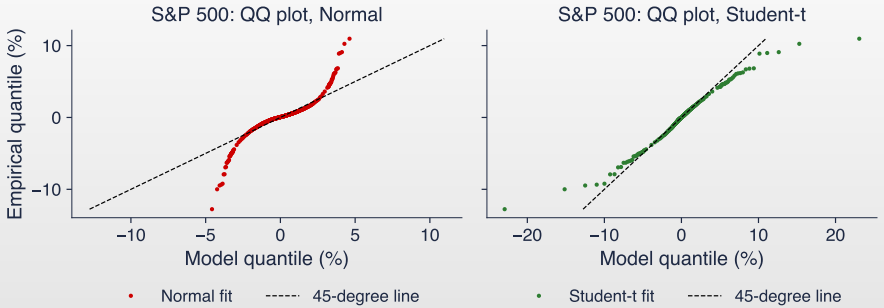
◂ Înapoi ▸ Lista exercițiilor

B3: trec cele mai bune modele testele? [Rezolvat]

- **Întrebarea**: trec modelele Normal și Student-t estimate pe S&P 500 testele KS, Cramér–von Mises și Anderson–Darling?
- **Date**: randamentele logaritmice zilnice S&P 500 din 2000 ($n = 6\,714$)
- **Cerințe**
  1. Estimați ambele modele și calculați cele trei statistici.
  2. Calculați p-value-ul KS naiv, care tratează parametrii ca fiind cunoscuți.
  3. Calculați p-value-uri prin bootstrap parametric cu $B = 99$ (simulare, reestimare, recalculare).
  4. Desenați graficul QQ al datelor față de ambele modele estimate.
  5. Interpretare: dacă ambele modele sînt respinse, la ce mai folosește testul?
- **Raportați**: un tabel cu șase statistici și p-value-urile lor, graficul QQ și două fraze
- Notebook: secțiunea B3

înapoi Lista exercițiilor

B3: rezolvare (1/2) [Rezolvat]



- Graficele QQ ale randamentelor S&P 500 față de ajustările Normală și Student-t: distribuția Normală eșuează în ambele cozi

SFM_ch6_seminar

B3: rezolvare (2/2) [Rezolvat]

	KS	CvM	AD	p-value KS naiv
Normală	0,093	22,84	127,35	$< 0{,}001$
valoare critică bootstrap 95%	0,011	0,12	0,75	
Student-t	0,018	0,57	4,69	0,021
valoare critică bootstrap 95%	0,009	0,08	0,65	

- p-value-ul bootstrap $= 0{,}01$ pentru fiecare statistică și ambele modele (cel mai mic posibil cu $B = 99$): ambele respinse
  - p-value-ul KS naiv al Student-t este mult prea mare
- Interpretare: cu $n = 6\,714$, orice model i.i.d. este respins
  - statisticile permit totuși ordonarea modelelor (AD Student-t 4,69 față de 127,35)
  - graficul QQ arată unde eșuează fiecare model

SFM_ch6_seminar

B4: crah sau eroare de date? Banca Transilvania [Propus]

- **Întrebarea**: care dintre cele mai mari randamente TLV sînt reale și schimbă eliminarea celor eronate modelul ales? Model: B1.
- **Date**: prețul ajustat TLV din 2010, randamente logaritmice zilnice în %; prețul de închidere (neajustat) pentru comparație; BET în aceleași zile
- **Cerințe**
  1. Identificați cele mai mari trei randamente în valoare absolută și datele lor.
  2. Pentru fiecare, comparați prețul de închidere și prețul ajustat în jurul datei, precum și randamentul BET din aceeași zi.
  3. Decideți care zile sînt erori de date și eliminați doar acele zile.
  4. Reestimați cinci candidați; comparați boltirea, $\hat{\nu}$ Student-t, statistica AD și cîștigătorul AIC înainte și după.
  5. Interpretare: ce s-ar fi întîmplat dacă ați fi eliminat toate cele trei zile?
- **Raportați**: un tabel cu trei zile și verdictul vostru, un tabel înainte/după și două fraze
- Notebook: secțiunea B4

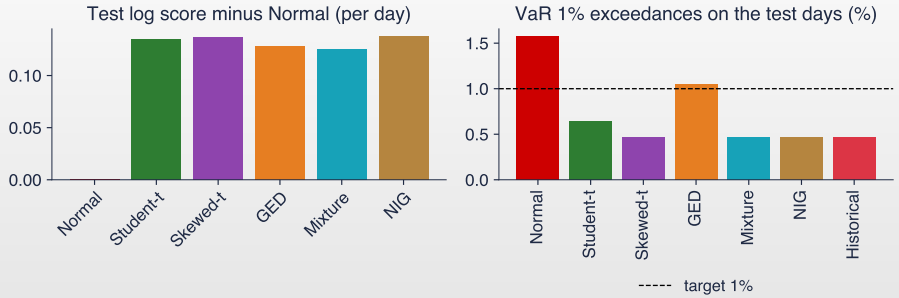
◂ Înapoi ▸ Lista exercițiilor

B5: VaR 1% în afara eșantionului pentru DAX [Rezolvat]

- **Întrebarea**: dă modelul ales de AIC pe 2000–2019 și cel mai bun VaR 1% pe 2020–2026?
- **Date**: randamentele logaritmice zilnice DAX: estimare 2000–2019 ($n = 5\,075$), test 2020–2026 ($n = 1\,711$)
- **Cerințe**
  1. Estimați cei șase candidați pe zilele de estimare și ordonați-i după AIC.
  2. Calculați scorul logaritmic mediu al fiecărui model pe zilele de test.
  3. Calculați VaR 1% al fiecărui model, numărați depășirile pe zilele de test și aplicați testul Kupiec.
  4. Adăugați simularea istorică (cuantila de 1% a zilelor de estimare) și VaR mediat cu ponderile Akaike.
  5. Interpretare: ce model ați alege pentru VaR 1% și de ce?
- **Raportați**: un tabel, graficul și două fraze
- Notebook: secțiunea B5

Înapoi Lista exercițiilor

B5: rezolvare (1/2) [Rezolvat]



- Stînga: scorul logaritmic de test minus cel al modelului Normal; dreapta: ratele de depășire ale VaR 1% pe zilele de test (ținta 1%)

SFM_ch6_seminar

B5: rezolvare (2/2) [Rezolvat]

	scor log.	VaR 1%	depășiri	p Kupiec	pinball $\times 100$
Normală	−1,668	3,37	27	0,027	5,37
Student-t	−1,533	4,09	11	0,112	5,23
skewed-t	−1,532	4,38	8	0,013	5,36
GED	−1,539	3,91	18	0,830	5,21
amestec Normal	−1,542	4,39	8	0,013	5,37
NIG	−1,530	4,38	8	0,013	5,36
istorică	–	4,33	8	0,013	5,33
mediată	–	4,37	8	0,013	5,36

- AIC pe 2000–2019: NIG; cel mai bun scor logaritmic de test: NIG; depășiri așteptate 17,1; cea mai mică pierdere pinball: GED
- Interpretare: cîștigătorul AIC dă cea mai bună densitate, dar prea puține depășiri (prea prudent)
  - dacă scopul este doar VaR 1%, GED și Student-t au rate de depășire mai apropiate de 1%

SFM_ch6_seminar

B6: VaR 1% în afara eșantionului pentru Bitcoin [Propus]

- **Întrebarea**: se menține concluzia din B5 pentru Bitcoin? Model: B5.
- **Date**: Bitcoin: estimare 2014–2019 ($n = 1\,931$), test 2020–2026 ($n = 2\,453$)
- **Cerințe**
  1. Repetați pașii 1–4 din B5.
  2. Comparați VaR 1% empiric al zilelor de test cu cel al zilelor de estimare.
  3. Interpretare: de ce eșuează aici toate modelele cu cozi groase testul Kupiec și în ce direcție?
- **Raportați**: un tabel și două fraze
- Notebook: secțiunea B6

Înapoi Lista exercițiilor

C1: se schimbă în timp cel mai bun model? [Propus]

- **Întrebarea**: este cîștigătorul AIC pentru S&P 500 și BET același în fiecare fereastră de doi ani?
- **Date**: randamente logaritmice zilnice din 2000, ferestre de doi ani care nu se suprapun (2000–2001, 2002–2003, ...); candidați: Normală, Student-t, skewed-t, GED, NIG; modele: B1, A1
- **Cerințe**
  1. În fiecare fereastră, găsiți cîștigătorul AIC, ponderea lui Akaike și ΔAIC al celui de-al doilea.
  2. Numărați de cîte ori cîștigă fiecare model, pentru fiecare indice.
  3. Propuneți o explicație a schimbărilor și o cale de a o testa.
  4. Interpretare: ar trebui ca modelul de risc să fie ales din nou la fiecare doi ani?
- **Raportați**: un tabel, numărul de ferestre cîștigate de fiecare model și un plan de proiect
- Notebook: secțiunea C1

înapoi Lista exercițiilor

C2: verificarea unui răspuns generat de AI [Propus]

- **Un student a cerut unui asistent AI să aleagă o distribuție pentru randamentele logaritmice zilnice S&P 500 (%) din 2000 ($n = 6\,714$). Răspunsul:**
- `(a) Skewed-t are cel mai mic AIC, deci este semnificativ mai bun decît Student-t (Delta AIC = 17,6).`
- `(b) Testul KS al ajustării Student-t dă p = 0,021 cu scipy.stats.kstest, deci Student-t se potrivește la nivelul de 1%.`
- `(c) Testul LR al distribuției Normale față de Student-t folosește valoarea critică chi2(1) = 3,84.`
- `(d) BIC penalizează parametrii în plus mai puțin decît AIC, pentru că ln(n) este doar 8,81.`
- `(e) Testul Anderson-Darling este mai sensibil decît KS la calitatea ajustării în cozi.`
- `(f) Pentru că skewed-t are cea mai bună ajustare, VaR 1% calculat cu el este cel mai precis.`
- **Cerințe**
  - **1. Pentru fiecare afirmație, precizați dacă este corectă; dacă nu este, formulați afirmația corectă și, unde se poate, cifra corectă din notebook (secțiunea C2).**
  - **2. Raportați: o listă de șase verdicte, fiecare cu un rînd de justificare.**

Înapoi Lista exercițiilor

Idei de reținut

- AIC și BIC ordonează modelele; $\Delta \leq 2$ înseamnă egalitate; ponderile Akaike transformă Δ în dovezi
- Testele LR cer modele imbricate și o valoare a parametrului din interiorul spațiului parametrilor; la frontieră folosiți amestecul 50:50
- KS cu parametri estimați cere tabelele Lilliefors sau bootstrap; AD pune accentul pe cozi, KS nu
- Verificați cele mai mari randamente înainte de modelare: eliminați erorile, păstrați crahurile
- Cea mai bună densitate nu dă întotdeauna cel mai bun VaR: validați în afara eșantionului

După seminar

- Cursul 6 dezvoltă fiecare temă de azi: verosimilitatea, testele LR și Vuong, AIC și BIC, testele EDF, graficele QQ, VaR în afara eșantionului, incertitudinea de model
- C1 poate deveni un proiect de echipă: randamente standardizate, mai multe piețe, ferestre mobile
- Lectură: Burnham & Anderson (2004); Stephens (1974); Kupiec (1995)

Bibliografie

- **Akaike, H. (1974).** A new look at the statistical model identification. ***IEEE Transactions on Automatic Control***, **19(6), 716–723.**
- **Burnham, K.P., Anderson, D.R. (2004).** Multimodel inference: understanding AIC and BIC in model selection. ***Sociological Methods & Research***, **33(2), 261–304.**
- **Franke, J., Härdle, W.K., Hafner, C.M. (2019).** *Statistics of Financial Markets: An Introduction*, **5th ed. Springer.**
- **Hansen, B.E. (1994).** Autoregressive conditional density estimation. ***International Economic Review***, **35(3), 705–730.**
- **Kupiec, P.H. (1995).** Techniques for verifying the accuracy of risk measurement models. ***Journal of Derivatives***, **3(2), 73–84.**
- **Lilliefors, H.W. (1967).** On the Kolmogorov–Smirnov test for normality with mean and variance unknown. ***Journal of the American Statistical Association***, **62(318), 399–402.**
- **Schwarz, G. (1978).** Estimating the dimension of a model. ***Annals of Statistics***, **6(2), 461–464.**
- **Self, S.G., Liang, K.-Y. (1987).** Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. ***Journal of the American Statistical Association***, **82(398), 605–610.**
- **Stephens, M.A. (1974).** EDF statistics for goodness of fit and some comparisons. ***Journal of the American Statistical Association***, **69(347), 730–737.**
- **Wilks, S.S. (1938).** The large-sample distribution of the likelihood ratio for testing composite hypotheses. ***Annals of Mathematical Statistics***, **9(1), 60–62.**