








Statistica piețelor financiare

Capitolul 2: Distribuții clasice și fapte stilizate



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 (1/2)

- **ACF**: Autocorrelation Function (funcția de autocorelație)
- **AI**: Artificial Intelligence (inteligență artificială)
- **AIC**: Akaike Information Criterion (criteriul informațional Akaike)
- **ARCH**: AutoRegressive Conditional Heteroskedasticity (heteroscedasticitate condiționată autoregresivă)
- **BET**: Bucharest Exchange Trading (index) — indicele de referință al Bursei de Valori București
- **BVB**: Bursa de Valori București
- **CDF**: Cumulative Distribution Function (funcția de repartiție)
- **CLT**: Central Limit Theorem (teorema limită centrală)
- **EGARCH**: Exponential GARCH (GARCH exponențial)
- **ES**: Expected Shortfall (pierderea așteptată în coadă)
- **ETF**: Exchange-Traded Fund (fond tranzacționat la bursă)

Acronime folosite în acest material (2/2)

- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity (heteroscedasticitate condiționată autoregresivă generalizată)
- **GJR-GARCH**: Glosten–Jagannathan–Runkle GARCH (GARCH asimetric Glosten–Jagannathan–Runkle)
- **JB**: Jarque–Bera (test) — testul de normalitate Jarque–Bera
- **LB**: Ljung–Box (test) — testul de autocorelație Ljung–Box
- **MLE**: Maximum Likelihood Estimation (estimarea prin metoda verosimilității maxime)
- **OMV**: Österreichische Mineralölverwaltung (Administrația austriacă a uleiurilor minerale, grupul OMV)
- **OUG**: Ordonanță de Urgență a Guvernului
- **QQ**: Quantile–Quantile (plot) — graficul cuantilă–cuantilă
- **SFE**: Statistics of Financial Markets, Exercises (Quantlet collection of the textbook) — colecția de Quantlet-uri a manualului Statistics of Financial Markets

Întrebarea de azi, traseul și rezultatele învățării

- **Întrebarea**: este distribuția Normală un model bun pentru randamentele zilnice și cum arată randamentele reale?
  - de răspuns depind măsurarea riscului, evaluarea opțiunilor și testarea ipotezelor
- **Traseul și rezultatele învățării**: la finalul capitolului veți putea
  - calcula probabilități și cuantile ale distribuțiilor Normală și lognormală
  - enunța teorema limită centrală (CLT, central limit theorem) și preciza cînd funcționează aproximarea Normală
  - estima asimetria și excesul de boltire și testa normalitatea cu testul Jarque–Bera
  - citi un QQ plot și estima o distribuție Student-t prin metoda verosimilității maxime
  - verifica șase fapte stilizate ale randamentelor Cont (2001) și explica ce înseamnă fiecare pentru un model
- Date reale, 2010–2026: S&P 500, DAX, BET, Bitcoin și patru acțiuni BVB (SNP, BRD, SNN, SNG)

Bibliografie și instrumente

- Manual: Franke, Härdle & Hafner (2019), *Statistics of Financial Markets*, ed. a 5-a, secț. 3.3 (Normală și lognormală) și secț. 11.2 (fapte stilizate)
  - exerciții rezolvate: Borak, Härdle & López-Cabrera (2013)
- Complementar: Tsay (2010), secț. 1.2 (proprietățile distribuției randamentelor); Cont (2001)
- Quantlet-urile Python ale capitolului: Quantlets/Ch_02
  - adaptate după Quantlet-urile SFE ale manualului (SFEDaxReturnDistribution, SFElognormal, SFEclt, SFENormalApprox1–4)
  - fiecare grafic din aceste slide-uri are un link către Quantlet-ul care îl generează
- Notebook-ul cursului: deschideți în Google Colab
- Curs video: Tukey's g- and h- transformations

O zi extremă în ipoteza distribuției Normale

- 16 martie 2020: S&P 500 a scăzut cu $-12{,}8\%$ (randament logaritmic) într-o singură zi
- Media și abaterea standard zilnică 2010–2026: 0,045% și 1,084%, deci scăderea a fost de $(-12{,}77 - 0{,}045)/1{,}084 = -11{,}8$ abateri standard
  - dacă randamentele ar urma distribuția Normală, probabilitatea unei asemenea zile ar fi de circa 2×10^{-32}
  - adică mai rar de o dată la 10^{29} ani de tranzacționare
- Totuși, asemenea zile au loc: 19 octombrie 1987, octombrie 2008, martie 2020, aprilie 2025
- Concluzia: greșit este modelul, nu piața

Întrebare pentru sală: cît de des apare o zi de 4 sigma?

- O **zi de** k **sigma**: un randament zilnic aflat la mai mult de k abateri standard de medie
- **Ce credeți?** Dacă randamentele zilnice ar fi Normale, cît de des ar apărea o zi de 4 sigma?
- **Ce credeți?** Cît de des apare în S&P 500 din 2010?
- **Răspuns**
  - distribuția Normală: o dată la 15 787 de zile de tranzacționare, cam o dată la 63 de ani
  - S&P 500: 28 de zile din 4 201, o dată la 150 de zile, cam 1,7 pe an

La ce folosește o distribuție

- **Risc**: probabilitatea unei pierderi mari
  - VaR 1% (Value at Risk, valoarea expusă la risc): pierderea depășită cu probabilitatea 1%; o cuantilă a distribuției (Capitolul 10)
- **Evaluare**: formula Black–Scholes presupune prețuri lognormale
  - cozile mai groase fac mai valoroase opțiunile mult în afara banilor (deep out-of-the-money)
- **Inferență**: testele t și intervalele de încredere presupun normalitate (aproximativă)
  - cozile groase fac erorile standard nesigure în eșantioane mici
- O distribuție greșită produce rezultate greșite, chiar și cu date perfecte

Modelul de referință al capitolului

- Prețul P_t, randamentul logaritmic $r_t = \ln(P_t/P_{t-1})$, deci $P_t = P_{t-1}e^{r_t} = P_0 e^{r_1+\cdots+r_t}$
  - notație: **i.i.d.** = independente și identic distribuite
- **Modelul de referință**: r_t i.i.d. $N(\mu, \sigma^2)$
  - $N(\mu, \sigma^2)$: distribuția Normală cu media μ și varianța σ^2 (media și varianța zilnice ale randamentelor logaritmice)
  - prețurile urmează un mers aleator geometric; P_t este lognormal Osborne (1959)
  - primul model de acest fel, pentru variațiile aritmetice ale prețurilor: Bachelier (1900)
- Capitolul verifică pe rînd ipotezele modelului de referință: normalitatea, distribuția identică și independența

Carl Friedrich Gauss și distribuția Normală

- De Moivre (1733) a obținut curba în formă de clopot ca limită a probabilităților binomiale
- Gauss (1809) a folosit-o pentru erorile observațiilor astronomice și a dedus din ea metoda celor mai mici pătrate
- Laplace (1810) a demonstrat o primă versiune a teoremei limită centrale
- De aici și celălalt nume: distribuția **gaussiană**
- În finanțe a devenit modelul implicit al randamentelor, pentru că sumele de variabile Normale rămîn Normale



Carl Friedrich Gauss (1777–1855)

Portret: Christian Albrecht Jensen (1840); domeniu public; Wikimedia Commons

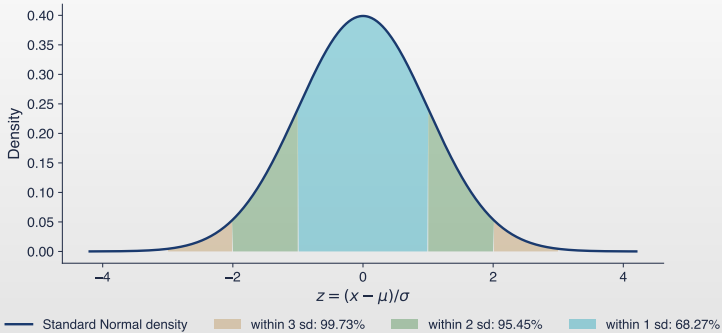
Distribuția Normală: definiție

- $X \sim N(\mu, \sigma^2)$ are **PDF** (probability density function, densitatea de probabilitate)
  - $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big), \quad x \in \mathbb{R}$
  - doi parametri: media μ (poziția) și varianța σ^2 (dispersia)
- **Distribuția Normală standard**: $Z = (X - \mu)/\sigma \sim N(0, 1)$
  - **CDF** (cumulative distribution function, funcția de repartiție) este $\Phi(z) = P(Z \leq z)$, fără formă închisă; se calculează cu tabele sau cu software
- Probabilitățile lui X se obțin din Z: $P(X \leq x) = \Phi\big((x - \mu)/\sigma\big)$
- Cuantila de ordin α: $x_\alpha = \mu + \sigma z_\alpha$, valoarea pentru care $P(X \leq x_\alpha) = \alpha$
  - $z_\alpha = \Phi^{-1}(\alpha)$: cuantila lui $N(0, 1)$; $z_{0,01} = -2{,}33$, $z_{0,05} = -1{,}64$

Proprietăți folosite în finanțe

- Simetrică în jurul lui μ: media = mediana = modul
- Asimetria 0 și boltirea 3 (definite în secțiunea 5)
- **Închisă la adunare**: dacă $X \sim N(\mu_1, \sigma_1^2)$ și $Y \sim N(\mu_2, \sigma_2^2)$ sînt independente, atunci $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
  - randamentul logaritmic pe h zile în modelul de referință: $r_t(h) \sim N(h\mu, h\sigma^2)$
  - volatilitatea crește cu $\sqrt{h}$: regula rădăcinii pătrate a timpului din Capitolul 1
- Orice combinație liniară de randamente cu distribuție Normală multivariată este Normală: randamentele portofoliilor rămîn Normale
- Complet determinată de doi parametri: dacă știm μ și σ, putem calcula orice probabilitate

Regula 68–95–99,7



- La cel mult 1, 2 și 3 abateri standard de medie: 68,27%, 95,45% și 99,73% din probabilitate
- Dincolo de 3 abateri standard: o zi din 370; cozile distribuției Normale scad ca $e^{-z^2/2}$, foarte repede

SFM_ch2_normal_lognormal

Frecvența variațiilor mari în cazul distribuției Normale

k	$P(\|Z\| > k)$	o dată la ... zile de tranzacționare	o dată la ... ani
1	0,3173	3	0,013
2	0,0455	22	0,087
3	$2{,}7 \times 10^{-3}$	370	1,5
4	$6{,}3 \times 10^{-5}$	15 787	63
5	$5{,}7 \times 10^{-7}$	1 744 278	6 922
6	$2{,}0 \times 10^{-9}$	506 797 346	2 011 101

- $P(|Z| > k)$: probabilitatea ca o variabilă Normală standard să se afle la mai mult de k abateri standard de 0
- Anii sînt calculați cu 252 de zile de tranzacționare pe an
- Fiecare abatere standard în plus face o variație de 7 pînă la 291 de ori mai rară
- O zi de 6 sigma n-ar trebui să apară în toată istoria burselor

SFM_ch2_normal_lognormal

Exemplu lucrat: pierderile mari ale S&P 500

- S&P 500, 2010–2026: $n = 4\,201$ randamente logaritmice zilnice, media 0,045%, abaterea standard 1,08%
- O pierdere mai mare de 3%: $z = (-3 - 0{,}045)/1{,}08 = -2{,}81$
  - distribuția Normală: $\Phi(-2{,}81) = 0{,}25\%$, deci numărul așteptat de zile este 10,5
  - numărul observat: 49 de zile
- O pierdere mai mare de 5%: $z = -4{,}65$
  - distribuția Normală: $6{,}9 \times 10^{-3}$ zile așteptate; zile observate: 8
- BET: 13 zile sub -5%, față de $5{,}1 \times 10^{-3}$ așteptate
- Distribuția Normală subestimează frecvența pierderilor mari cu cîteva ordine de mărime

Recapitulare: distribuția Normală

- Doi parametri, simetrică, închisă la adunare, volatilitatea crește cu $\sqrt{h}$
- Cozile ei sînt subțiri: o zi de 4 sigma o dată la 63 de ani
- Randamentele reale au mult mai multe variații mari decît permite ea

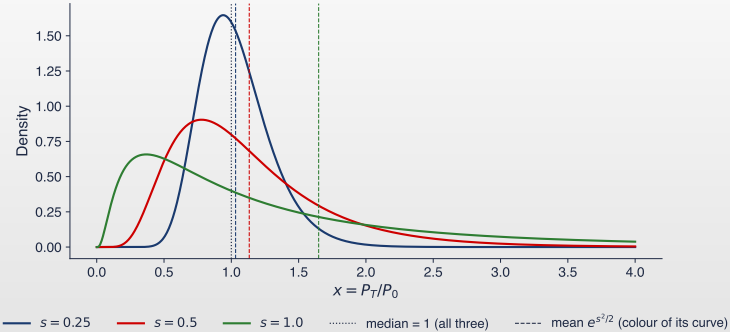
De la randamente logaritmice Normale la prețuri lognormale

- O variabilă pozitivă Y este **lognormală**, $Y \sim LN(m, s^2)$, dacă $\ln Y \sim N(m, s^2)$
  - m, s^2: media și varianța lui $\ln Y$, nu ale lui Y
- Densitatea: $f(y) = \dfrac{1}{ys\sqrt{2\pi}} \exp\Big(-\dfrac{(\ln y - m)^2}{2s^2}\Big), \quad y > 0$
- În modelul de referință, $\ln(P_T/P_0) = r_1 + \cdots + r_T \sim N(T\mu, T\sigma^2)$
  - T: orizontul, în zile; deci $P_T/P_0 \sim LN(T\mu, T\sigma^2)$ Osborne (1959)
- De ce prețuri lognormale și nu Normale?
  - prețurile nu pot fi negative; un preț cu distribuție Normală ar putea fi negativ
  - randamentul simplu $R = e^r - 1 > -1$: nu puteți pierde mai mult decît ați investit
- Modelul Black–Scholes se bazează pe ea

Media, mediana și modul

- Dacă $Y \sim LN(m, s^2)$:
- mediana e^m; modul e^{m-s^2}; media $E[Y] = e^{m+s^2/2}$
  - modul: valoarea cea mai probabilă; mediana: valoarea depășită cu probabilitatea 1/2
  - întotdeauna modul $<$ mediana $<$ media: o coadă dreaptă lungă (asimetrie pozitivă)
- Varianța: $\mathrm{Var}(Y) = (e^{s^2} - 1)\, e^{2m+s^2}$
- Media rezultă din $E[e^X] = e^{m+s^2/2}$ pentru $X \sim N(m, s^2)$ (Seminarul 2 o deduce)
  - termenul suplimentar $s^2/2$ este volatility drag din Capitolul 1, privit din perspectiva prețurilor
- Care dintre ele contează? Mediana este rezultatul unui investitor tipic; media este ridicată de cîteva traiectorii foarte favorabile

Densități lognormale (SFElognormal)



- $m = 0$: mediana este 1 pentru toate cele trei curbe; liniile punctate marchează mediile
- $s = 0{,}25$: media 1,03, modul 0,94, aproape simetrică; $s = 1$: media 1,65, modul 0,37, asimetria 6,18
- Orizonturi lungi sau volatilitate mare: distanța dintre medie și mediană crește

SFM_ch2_normal_lognormal

Exemplu lucrat: 100 de lei în BET

- BET, 2010–2026: randamentul logaritmic mediu anual $\mu = 11{,}6\%$ și volatilitatea $\sigma = 16{,}9\%$ (momentele zilnice $\times$ 250,6 și $\sqrt{250{,}6}$)
- După un an, $P_{\mathbf{1}}/P_{\mathbf{0}} \sim LN(\mu, \sigma^{\mathbf{2}})$
  - mediana $100e^{\mu} = 112{,}3$; media $100e^{\mu+\sigma^{\mathbf{2}}/\mathbf{2}} = 113{,}9$
  - probabilitatea unei pierderi $\Phi(-\mu/\sigma) = 24{,}8\%$; cuantila de 5% $100e^{\mu-\mathbf{1{,}645}\sigma} = 84{,}9$
  - $\Phi(-\mu/\sigma) = P(r < 0)$: probabilitatea unui randament logaritmic negativ; $-1{,}645 = z_{\mathbf{0{,}05}}$
- După 10 ani, $P_{\mathbf{10}}/P_{\mathbf{0}} \sim LN(10\mu, 10\sigma^{\mathbf{2}})$
  - mediana 317,7; media 366,8; probabilitatea unei pierderi 1,5%; cuantila de 5% 131,6
- Probabilitatea unei pierderi scade cu orizontul; după 10 ani, chiar și cuantila de 5% depășește 100, dar numai dacă media istorică se menține
- Atenție: aceste rezultate moștenesc toate slăbiciunile modelului Normal al randamentelor logaritmice

Recapitulare: distribuția lognormală

- Randamente logaritmice Normale $\Rightarrow$ prețuri lognormale: pozitive, asimetrice la dreapta
- Mediana e^m sub media $e^{m+s^2/2}$: diferența este volatility drag
- Raportați mediana și o cuantilă joasă, nu doar media

Tabla lui Galton

- Francis Galton (1889): bilele cad printre rînduri de cuie și sar la stînga sau la dreapta cu probabilitatea 1/2
- Poziția finală a unei bile este o **sumă** de multe șocuri mici independente
- Bilele acumulate formează o curbă în formă de clopot: probabilități binomiale apropiate de densitatea Normală
- Același argument a fost folosit pentru randamente: un randament zilnic este suma multor variații mici de preț din timpul zilei



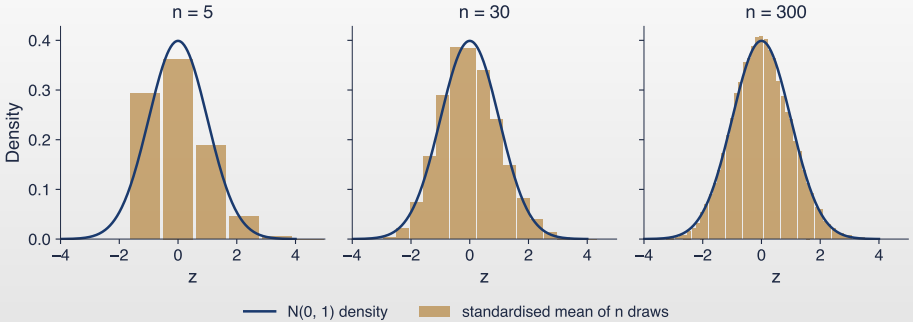
O tablă Galton după căderea bilelor

Foto: Klaus-Dieter Keller (2010); domeniu public; Wikimedia Commons

Teorema limită centrală

- **CLT** (Lindeberg–Lévy): $X_1, \ldots, X_n$ i.i.d. cu media μ și varianța finită σ^2
  - $\dfrac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$ cînd $n \to \infty$
  - $\bar{X}_n$: media celor n variabile; $\xrightarrow{d}$: convergență în distribuție, funcția de repartiție a membrului stîng se apropie de Φ
  - echivalent, suma $X_1 + \cdots + X_n$ este aproximativ $N(n\mu, n\sigma^2)$
- Trei condiții
  - independență; distribuție identică; varianță finită
- Cît de repede? Eroarea scade ca $1/\sqrt{n}$ și crește cu asimetria lui X (marginea Berry–Esseen)
- Explică de ce distribuția Normală apare peste tot și de ce intervalele de încredere pentru medie se construiesc cu distribuția Normală

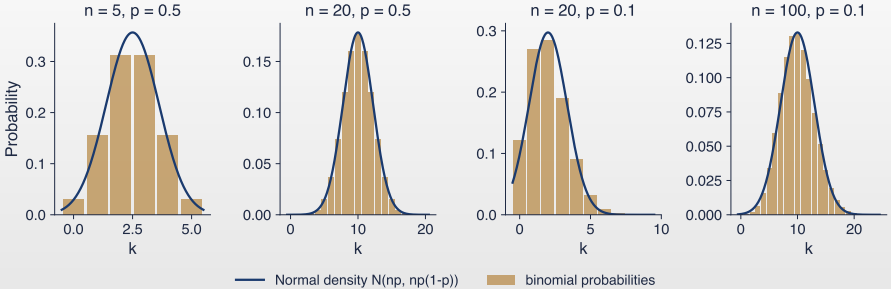
Ilustrarea CLT (SFEclt)



- Media standardizată a n extrageri Bernoulli(0,2) (o variabilă asimetrică), 20 000 de simulări
- Ponderea peste 1,96: 5,8% pentru $n = 5$, 2,7% pentru $n = 30$, 2,8% pentru $n = 300$; distribuția Normală: 2,5%
- Asimetria mediei, $(1 - 2p)/\sqrt{np(1-p)}$ cu $p = 0{,}2$: 0,67, 0,27, 0,09; scade ca $1/\sqrt{n}$

SFM_ch2_clt_normal_approx

Aproximarea Normală a distribuției binomiale (SFENormalApprox)



- $B(n,p)$: numărul de succese în n încercări independente, cu probabilitatea de succes p; $B(n,p) \approx N\left(np, np(1-p)\right)$
- Corecția de continuitate: $P(X \leq k) \approx \Phi\left((k+0{,}5-np)/\sqrt{np(1-p)}\right)$; termenul $+0{,}5$ distribuie fiecare întreg k pe intervalul $[k-0{,}5; k+0{,}5]$
- Cea mai mare eroare a CDF: 0,001 pentru $n=20, p=0{,}5$; 0,037 pentru $n=20, p=0{,}1$; 0,017 pentru $n=100, p=0{,}1$
- Regulă practică: aproximarea este bună cînd $np(1-p)$ este mare (cel puțin circa 9)

SFM_ch2_clt_normal_approx

Întrebare pentru sală: devin randamentele zilnice Normale datorită CLT?

- Un randament logaritmic zilnic este suma a mii de randamente logaritmice din timpul zilei
- **Ce credeți?** Ar trebui atunci randamentele zilnice să fie aproape Normale?
- **Răspuns**: nu neapărat; fiecare condiție a CLT poate fi încălcată
  - dependență: zilele calme și cele agitate apar grupat (secțiunea 8)
  - distribuție schimbătoare: volatilitatea diferă de la o zi la alta
  - varianță infinită: Mandelbrot (1963) a propus-o pentru prețurile bumbacului (Capitolul 3, distribuții α-stabile)
- Datele decid: secțiunile următoare măsoară cît de departe sînt randamentele de distribuția Normală

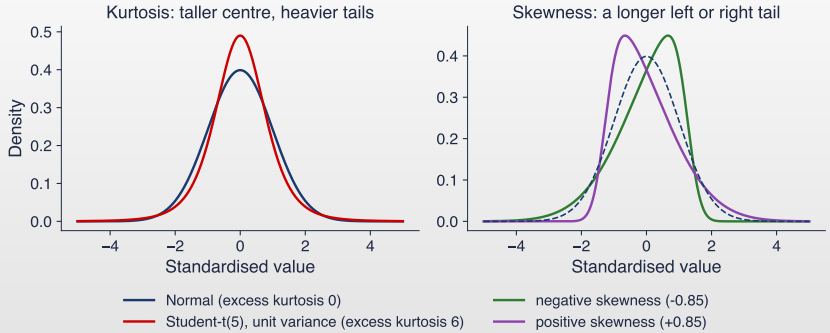
Recapitulare: teorema limită centrală

- Sumele multor termeni i.i.d. cu varianță finită sînt aproximativ Normale
- Convergența este mai lentă pentru variabile asimetrice; $np(1-p) \geq 9$ pentru distribuția binomială
- Randamentele pot încălca independența, distribuția identică sau varianța finită

Patru momente

- Media $\mu = E[X]$; varianța $\sigma^2 = E[(X-\mu)^2]$; $E[\cdot]$ este valoarea așteptată
- **Asimetria** (skewness): $\gamma_1 = E[(X-\mu)^3]/\sigma^3$
  - puterea a treia păstrează semnul fiecărei abateri; $\gamma_1 < 0$: o coadă stîngă mai lungă, pierderile mari mai frecvente decît cîștigurile mari
- **Boltirea** (kurtosis): $\gamma_2 = E[(X-\mu)^4]/\sigma^4$; distribuția Normală: $\gamma_2 = 3$
  - **excesul de boltire** $\gamma_2 - 3$: pozitiv înseamnă cozi groase și un centru înalt și îngust (distribuție **leptocurtică**)
- Ambele sînt adimensionale: descriu forma, nu poziția sau scala
- Puterile a patra fac boltirea foarte sensibilă la cîteva zile extreme

Asimetria și boltirea: ilustrare grafică



- Stînga: aceeași medie și variață; curba Student-$t(5)$ (secțiunea 7) are exces de boltire 6: mai multă masă în centru și în cozi, mai puțină în umeri
- Dreapta: densități skew-Normal cu asimetria $\pm 0{,}85$; curba punctată este distribuția Normală

SFM_ch2_moments_jarque_bera

Momentele de selecție

- Momentele centrate de selecție: $m_k = \frac{1}{n}\sum_{t=1}^{n}(r_t - \bar{r})^k$, media puterii k a abaterilor de la medie (m_2: varianța)
- Asimetria de selecție $S = m_3/m_2^{3/2}$; excesul de boltire de selecție $K = m_4/m_2^2 - 3$
- Dacă datele sînt i.i.d. Normale: $S \approx N(0, 6/n)$ și $K \approx N(0, 24/n)$
  - cu $n = 4\,201$: erorile standard $\sqrt{6/n} = 0{,}038$ și $\sqrt{24/n} = 0{,}076$
- Aceste erori standard sînt valabile doar în ipoteza de normalitate; pentru date cu cozi groase sînt mult prea mici (Seminarul 2, B1)

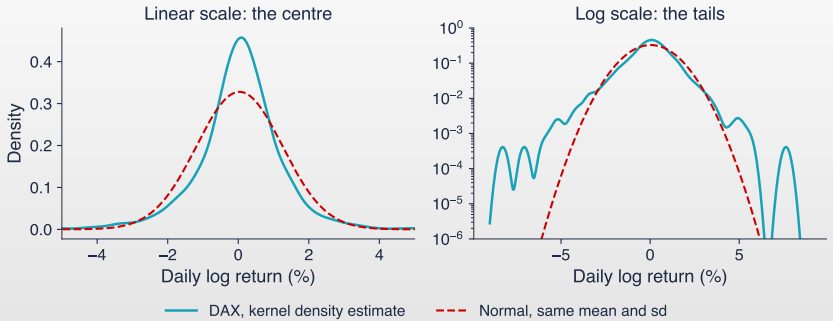
Randamente logaritmice zilnice: momente, 2010–2026

Seria	din	n	Media %	Ab. std. %	Asim.	Exces bolt.	Min %
S&P 500	2010	4 201	0,045	1,08	−0,61	13,5	−12,8
DAX	2010	4 244	0,034	1,22	−0,45	7,4	−13,1
BET	2010	4 186	0,046	1,07	−0,98	17,9	−11,9
Bitcoin	2014	4 384	0,118	3,49	−0,70	12,0	−46,5
OMV Petrom (SNP)	2010	3 815	0,076	1,62	−0,54	8,9	−15,8
BRD	2010	3 771	0,044	1,66	−0,38	7,3	−18,2
Nuclearelectrica (SNN)	2013	2 911	0,099	1,52	0,27	5,1	−9,4
Romgaz (SNG)	2013	2 909	0,090	1,40	−0,34	5,3	−11,0

- Toate seriile au exces de boltire între 5,1 (SNN) și 17,9 (BET): departe de 0
- Toate seriile, cu excepția SNN, au asimetrie negativă
- Cele mai slabe zile: S&P 500 pe 16 martie 2020, BET pe 19 decembrie 2018 (OUG 114, taxa pe activele bancare), Bitcoin pe 12 martie 2020

SFM_ch2_moments_jarque_bera

Randamentele DAX față de densitatea Normală (SFEDaxReturnDistribution)



- Estimatorul nucleu al densității (o histogramă netezită) față de densitatea Normală cu aceeași medie și abatere standard
- Centrul: vîrful densității 0,46 față de 0,33; 76,8% din zile în interiorul unei abateri standard, față de 68,3% pentru distribuția Normală
- Cozile (scară logaritmică): la -6% densitatea estimată este de 439 de ori mai mare decît densitatea Normală

SFM_ch2_moments_jarque_bera

Verificarea datelor înaintea estimării cozilor

- Banca Transilvania (TLV), 27–31 mai 2016: prețul de închidere scade de la 2,83 la 2,09 lei, data ex a unei distribuiri de acțiuni gratuite
- Prețul ajustat din datele noastre: 3,6014, apoi 3,0995, apoi 3,8060
  - randamente logaritmice $-15{,}0\%$ și $+20{,}5\%$ în două zile consecutive: ajustarea a fost aplicată cu o zi mai tîrziu
- Excesul de boltire al TLV, 2010–2026: 17,5 cu aceste două zile, 13,5 fără ele
- Două zile greșite din mii schimbă boltirea cu o treime: de aceea TLV nu apare în tabelele acestui capitol
- Regula din Capitolul 1: listați cele mai mari variații și verificați-le pe fiecare, mai ales perechile de salturi de semn opus

Recapitulare: momente

- Asimetria măsoară lipsa simetriei, excesul de boltire măsoară greutatea cozilor față de distribuția Normală
- Randamentele zilnice: în general asimetrie negativă și exces de boltire între 5 și 18
- Boltirea este determinată de cîteva zile: verificați-le înainte de a avea încredere în ea

Testul Jarque–Bera

- H_0: datele sînt Normale, deci asimetria $= 0$ și excesul de boltire $= 0$
- Statistica **JB** Jarque & Bera (1980), Jarque & Bera (1987): $\text{JB} = \frac{n}{6}\left(S^2 + \frac{K^2}{4}\right)$
  - suma a două statistici standardizate la pătrat: $(S/\sqrt{6/n})^2 + (K/\sqrt{24/n})^2$
- În ipoteza H_0 și pentru n mare: $\text{JB} \sim \chi^2(2)$; respingem la 5% dacă $\text{JB} > 5{,}99$
  - $\chi^2(2)$: distribuția hi-pătrat cu 2 grade de libertate, distribuția sumei pătratelor a două variabile $N(0,1)$ independente; 5,99 este cuantila ei de 95%
  - $\text{JB} = 0$ pentru o formă perfect Normală; valorile mari indică asimetrie sau exces de boltire departe de 0
- Este simplu, folosește doar două momente și este testul raportat de regulă în finanțele empirice

Exemplu lucrat: Jarque–Bera pentru S&P 500

- $n = 4\,201$, $S = -0{,}606$, $K = 13{,}49$
- $\mathrm{JB} = 700{,}17 \times (0{,}3673 + 45{,}497) = 32\,113$
- Valoarea critică 5,99: normalitatea este respinsă; p-value-ul este practic zero
- Termenul de boltire domină: aproape toată valoarea JB provine din $K^2/4$
- JB pentru celelalte serii: toate respinse la orice nivel uzual
  - indici și Bitcoin: BET 56 762, DAX 9 909, Bitcoin 26 482
  - acțiuni BVB: SNP 12 832, BRD 8 370, SNN 3 174, SNG 3 415

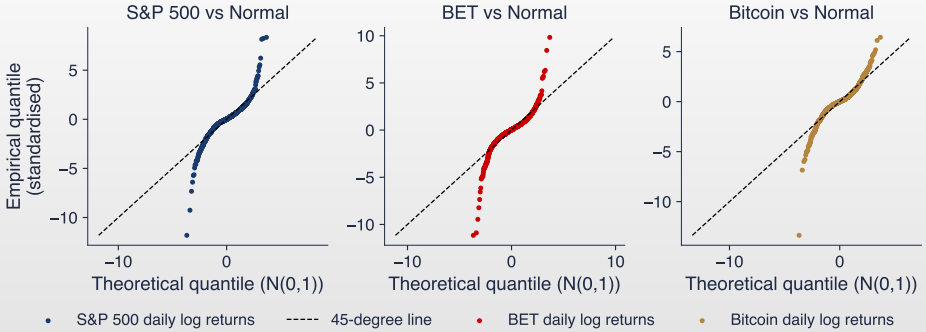
Limitele testului Jarque–Bera

- Cu mii de observații respinge chiar și abateri foarte mici
  - raportați mărimea lui S și K, nu doar p-value-ul
- Testul presupune date i.i.d.; cu volatility clustering, erorile standard ale lui S și K sînt prea mici
  - S&P 500: intervalul de 95% pentru K în ipoteza de normalitate: [13,34; 13,64]; intervalul bootstrap: [6,7; 20,6] (Seminarul 2, B1)
- Respingerea arată că datele nu sînt Normale, dar nu indică ce distribuție urmează
- Alte teste (Kolmogorov–Smirnov, Anderson–Darling) compară întreaga CDF: Capitolul 6

QQ plots

- Un **QQ plot** (quantile–quantile plot, graficul cuantilă–cuantilă) compară cuantilele empirice cu cele ale unui model Wilk & Gnanadesikan (1968)
  - ordonați datele: $r_{(1)} \leq r_{(2)} \leq \cdots \leq r_{(n)}$
  - reprezentați $r_{(i)}$ în funcție de cuantila modelului $F^{-1}\big((i-0{,}5)/n\big)$
  - $r_{(i)}$: al i-lea cel mai mic randament; F^{-1}: funcția cuantilă a modelului; $(i-0{,}5)/n$: proporția datelor aflate sub $r_{(i)}$
- Interpretare
  - puncte pe prima bisectoare: modelul se potrivește
  - o formă de S, capătul stîng sub dreaptă și cel drept deasupra: cozi mai groase decît ale modelului
  - un singur capăt se îndepărtează de dreaptă: asimetrie
- Spre deosebire de JB, arată **unde** greșește modelul: în centru, în umeri sau în cozi

QQ plots față de distribuția Normală



- Randamente standardizate față de cuantilele $N(0,1)$: toate trei se depărtează de dreaptă la ambele capete
- S&P 500: cel mai mic randament standardizat $-11{,}8$, unde cuantila Normală este $-3{,}7$; BET: $-11{,}2$ față de $-3{,}7$
- Coada stîngă se curbează mai mult decît cea dreaptă: asimetrie negativă

SFM_ch2_qq_plots

Recapitulare: testarea normalității

- JB combină asimetria și boltirea; respinge normalitatea pentru toate seriile
- Raportați mărimea abaterii; erorile standard din teoria Normală sînt prea mici
- QQ plots arată unde greșește modelul: aici, în ambele cozi

„Student": William Sealy Gosset

- Chimist și berar la Guinness, Dublin, lucra cu eșantioane mici de orz și hamei
- Fabrica de bere nu își lăsa angajații să publice sub propriul nume: a semnat „Student" Student (1908)
- A dedus distribuția lui $(\bar{X} - \mu)/(s/\sqrt{n})$ pentru date Normale: distribuția t
- În finanțe este folosită din alt motiv: cozile ei sînt mai groase decît cele ale distribuției Normale Praetz (1972), Blattberg & Gonedes (1974)



William Sealy Gosset (1876–1937)

Foto: autor necunoscut (1908); domeniu public; Wikimedia Commons

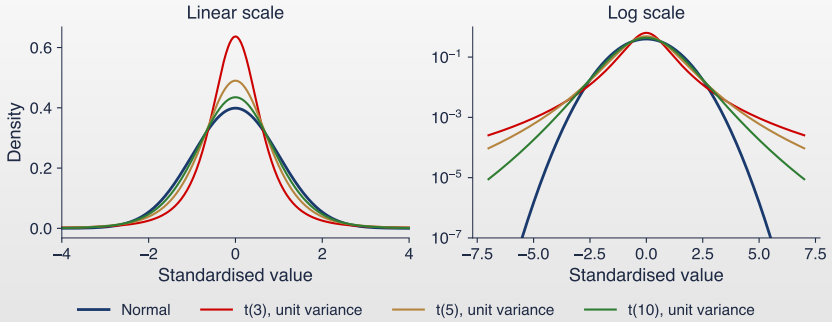
Distribuția Student-t: definiție

- $T = Z/\sqrt{V/\nu}$, cu $Z \sim N(0,1)$ independentă de $V \sim \chi^2(\nu)$: $T \sim t(\nu)$
  - $\nu > 0$: **gradele de libertate**, care controlează cozile
  - $\chi^2(\nu)$: distribuția sumei pătratelor a ν variabile $N(0,1)$ independente; Γ: funcția gamma, $\Gamma(k) = (k-1)!$ pentru k întreg
- Densitatea: $f(x) = \dfrac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \Big(1 + \dfrac{x^2}{\nu}\Big)^{-(\nu+1)/2}$
- Versiunea cu parametri de poziție și de scală pentru randamente: $r = m + s\,T$, cu poziția m, scala s și ν
- Cînd $\nu \to \infty$, $t(\nu) \to N(0,1)$: distribuția Normală este un caz limită

Momente și cozi

- Media 0 dacă $\nu > 1$; varianța $\nu/(\nu-2)$ dacă $\nu > 2$; excesul de boltire $6/(\nu-4)$ dacă $\nu > 4$
- Momentele de ordin ν și mai mare sînt infinite
  - $\nu \leq 4$: boltire infinită; $\nu \leq 2$: varianță infinită
- **Cozi de tip putere**: $P(|T| > x) \approx c\, x^{-\nu}$ pentru x mare
  - $c > 0$: o constantă; dublarea lui x împarte probabilitatea din coadă la aproximativ 2^{ν}
  - coada Normală scade ca $e^{-x^2/2}$, mult mai repede decît orice putere
  - ν este și **tail index**-ul, estimat direct în Capitolul 5
- Pentru comparația cu distribuția Normală la aceeași varianță, folosiți $T\sqrt{(\nu-2)/\nu}$, care are varianța 1

Densități Student-t cu varianța 1



- ν mai mic: centru mai înalt și cozi mai groase; pe scara logaritmică, cozile Normale scad ca o parabolă
- $P(|X| > 4)$ la varianța 1: $t(3)$ 0,62%, $t(5)$ 0,36%, $t(10)$ 0,12%, distribuția Normală 0,0063%
- Cu $\nu = 5$, o zi de 4 sigma este de 56 de ori mai probabilă decît în cazul distribuției Normale

SFM_ch2_student_t_fit

Estimarea unei distribuții Student-t prin verosimilitate maximă

- **MLE** (maximum likelihood estimation, estimarea prin verosimilitate maximă): alegem (ν, m, s) care maximizează log-verosimilitatea
  - $\ell(\nu, m, s) = \sum_{t=1}^{n} \ln f\big((r_t - m)/s; \nu\big) - n \ln s$
  - $f(\cdot; \nu)$: densitatea $t(\nu)$; termenul $-n \ln s$ corectează rescalarea randamentelor cu s
  - fără formă închisă: un optimizator numeric (`scipy.stats.t.fit` în notebook)
- **AIC** (Akaike information criterion, criteriul informațional Akaike) Akaike (1974): $\text{AIC} = 2k - 2\ell$
  - k = numărul de parametri (2 pentru distribuția Normală, 3 pentru t); se preferă modelul cu AIC mai mic
  - $\Delta\text{AIC} = \text{AIC}_{\text{Normal}} - \text{AIC}_t > 0$ favorizează t
- Primele rezultate empirice pentru prețurile acțiunilor: Praetz (1972), Blattberg & Gonedes (1974)

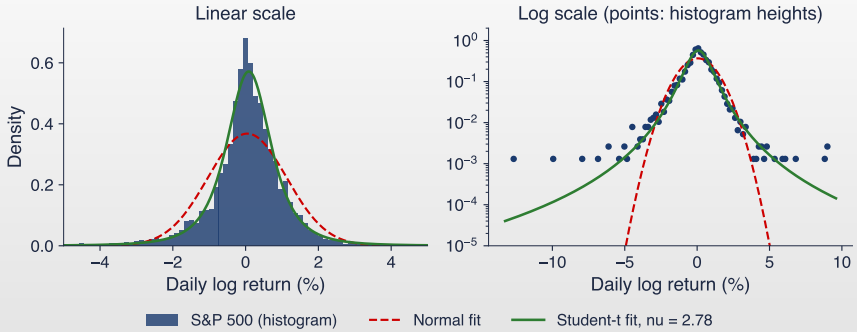
Distribuția Student-t estimată: gradele de libertate, 2010–2026

Seria	$\hat{\nu}$	ΔAIC
S&P 500	2,78	1213
DAX	3,31	739
BET	2,96	1454
Bitcoin	2,23	1335
OMV Petrom (SNP)	3,79	679
BRD	4,67	478
Nuclearelectrica (SNN)	3,08	541
Romgaz (SNG)	3,40	482

- Toate valorile $\hat{\nu}$ între 2,23 (Bitcoin) și 4,67 (BRD)
- $\hat{\nu} < 4$ pentru 7 din cele 8 serii: modelul estimat are boltire infinită
- ΔAIC este cel puțin 478: distribuția t este preferată distribuției Normale pentru toate seriile

SFM_ch2_student_t_fit

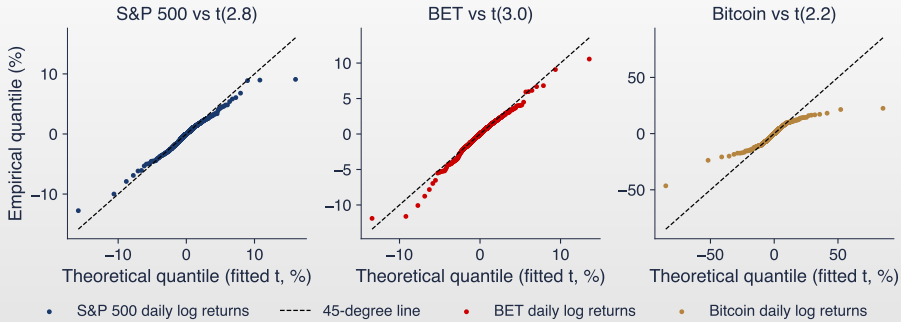
S&P 500: distribuția Normală și distribuția Student-t estimate



- Scara liniară: distribuția t ($\hat{\nu} = 2{,}78$) reproduce centrul înalt, pe care distribuția Normală nu îl reproduce
- Scara logaritmică: densitatea Normală scade abrupt dincolo de $\pm 4\%$; distribuția t urmează cozile observate

SFM_ch2_student_t_fit

QQ plots față de distribuția Student-t estimată



- Centrul și cozile moderate se află acum pe dreaptă
- Extremele: distribuția t estimată prevede minime mai mici decît cele observate: S&P 500 $-15{,}8\%$ față de $-12{,}8\%$; Bitcoin $-85{,}0\%$ față de $-46{,}5\%$
- Cu $\hat{\nu}$ apropiat de 2, modelul supraestimează cuantilele cele mai extreme

SFM_ch2_qq_plots

Limitele distribuției Student-t

- **Simetria**: t nu poate reproduce asimetria negativă
  - există distribuții t asimetrice, dar depășesc cadrul acestui capitol
- **Independența**: un model t i.i.d. nu are volatility clustering
  - modelele GARCH cu erori t le combină pe amîndouă Bollerslev (1987) (Capitolul 9)
- **Un singur tail index**: un singur ν pentru ambele cozi și pentru centru
  - teoria valorilor extreme modelează fiecare coadă separat (Capitolul 5)
- Rămîne cel mai simplu model cu cozi groase și o îmbunătățire mare față de distribuția Normală

Recapitulare: Student-t

- Un parametru în plus, ν, controlează cozile; cozi de tip putere, distribuția Normală cînd $\nu \to \infty$
- $\hat{\nu}$ estimat între 2 și 5 pentru randamentele zilnice; AIC preferă clar t
- Nu surprinde asimetria și volatility clustering

Definiția faptului stilizat

- Un **fapt stilizat**: o proprietate statistică comună multor active, piețe și perioade Cont (2001)
- Calitativ, nu exact: „cozi groase", nu „$\nu = 3{,}2$"
- Rama Cont enumeră unsprezece; noi verificăm șase pe datele noastre
- Un model bun al randamentelor trebuie să le reproducă; modelul de referință i.i.d. Normal nu reproduce majoritatea lor
- Primele rezultate empirice: Mandelbrot (1963), Fama (1965)



Rama Cont

Foto: Renate Schmid (2012); CC BY-SA 2.0 DE; Wikimedia Commons

Șase fapte stilizate

- **1. Cozi groase**: variațiile mari sînt mult mai frecvente decît prevede distribuția Normală
- **2. Gaussianitate agregată** (aggregational Gaussianity): randamentele pe orizonturi mai lungi sînt mai apropiate de distribuția Normală
- **3. Absența autocorelației liniare**: randamentele trecute aproape nu au putere de predicție pentru cele viitoare
- **4. Volatility clustering**: variațiile mari sînt urmate de variații mari, de orice semn
- **5. Efectul de levier** (leverage effect): scăderile cresc volatilitatea viitoare mai mult decît creșterile
- **6. Asimetria cîștig/pierdere**: pierderile mari sînt mai mari și mai frecvente decît cîștigurile mari

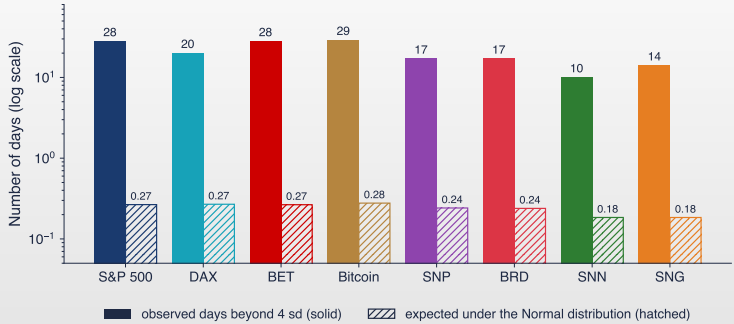
Faptul 1: cozile groase, numărul zilelor de k sigma

Seria	n	$> 3\sigma$	$> 4\sigma$	$> 5\sigma$	$> 6\sigma$	p-value (Poisson, 4σ)
S&P 500	4 201	62	28	12	8	2×10^{-46}
DAX	4 244	56	20	5	4	1×10^{-30}
BET	4 186	70	28	16	11	2×10^{-46}
Bitcoin	4 384	80	29	10	4	6×10^{-48}
Distribuția Normală, $n = 4\,201$		11,34	0,27	$2{,}4 \times 10^{-3}$	$8{,}3 \times 10^{-6}$	

- $> k\sigma$: numărul zilelor cu $|r_t - \bar{r}| > k\,s$; ultimul rînd: numărul așteptat conform distribuției Normale
- p-value: probabilitatea de a observa cel puțin atîtea zile de 4 sigma, dacă numărul lor urmează distribuția Poisson cu media $n \times 6{,}3 \times 10^{-5}$ (ipoteza de normalitate)
- S&P 500: de 105 de ori mai multe zile de 4 sigma decît permite distribuția Normală

SFM_ch2_sigma_days

Faptul 1: zile de 4 sigma pe toate piețele



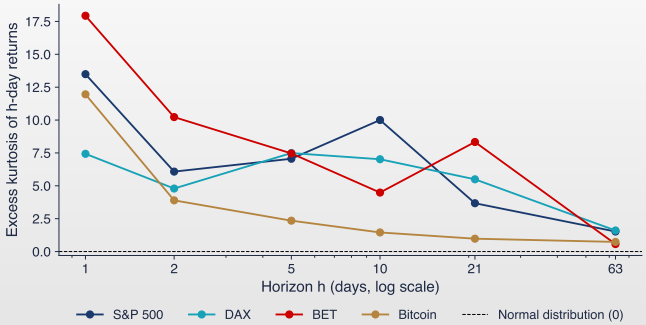
- Zilele observate dincolo de 4 abateri standard (scară logaritmică), față de numărul așteptat conform distribuției Normale, între 0,18 și 0,28 zile
- Indicii și Bitcoin: o zi de 4 sigma la fiecare 150–212 zile; acțiunile BVB: între 10 și 17 zile în total
- Cu o distribuție Student-t estimată, numerele așteptate sînt apropiate de cele observate: S&P 500 34,9, BET 28,1 (Seminarul 2, B4)

SFM_ch2_sigma_days

Faptul 2: gaussianitatea agregată

- Randamentul logaritmic pe h zile: $r_t(h) = r_t + r_{t-1} + \cdots + r_{t-h+1}$, calculat pe blocuri disjuncte
  - dacă se aplică CLT, distribuția lui $r_t(h)$ se apropie de distribuția Normală cînd h crește
- Faptul stilizat: forma se schimbă cu orizontul; nu este aceeași la toate scările de timp Cont (2001)
- Prețul plătit: mai puține observații
  - S&P 500, 17 ani: 4 201 randamente zilnice, 200 de randamente lunare și 66 de randamente trimestriale
  - estimările boltirii pe orizonturi lungi sînt foarte imprecise

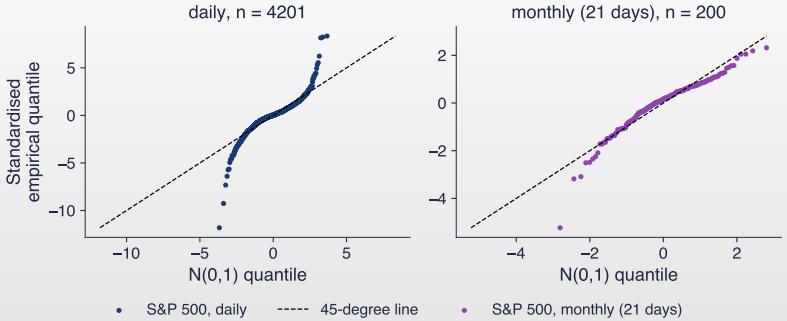
Faptul 2: boltirea scade cu orizontul



- Excesul de boltire, zilnic → lunar → trimestrial: S&P 500 13,5 → 3,7 → 1,5; BET 17,9 → 8,3 → 0,6; Bitcoin 12,0 → 1,0 → 0,7
- Scăderea este lentă și nemonotonă: o singură lună de crah (martie 2020) domină un eșantion de 200 de luni
- Randamentele trimestriale: JB nu mai respinge normalitatea pentru BET (p-value 0,07) și Bitcoin (p-value 0,22)

SFM_ch2_cont_stylised_facts

Faptul 2: S&P 500, QQ plots pentru randamentele zilnice și lunare



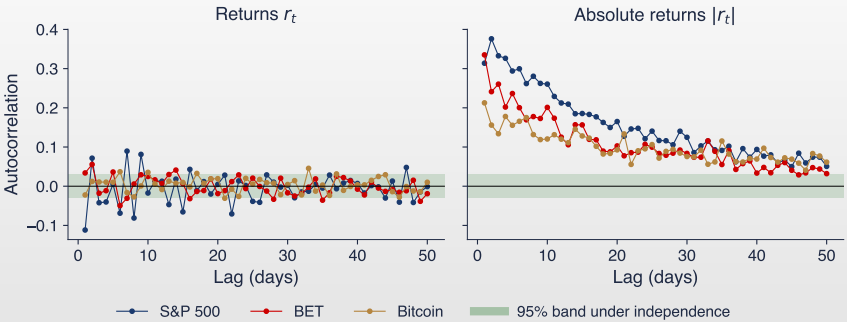
- Randamentele lunare sînt mult mai aproape de dreaptă decît cele zilnice
- Dar randamentele lunare nu sînt Normale: asimetria $-1{,}18$, excesul de boltire 3,7, $n = 200$; punctul cel mai de jos este martie 2020
- Cauza convergenței lente: volatility clustering face randamentele zilnice dependente, ceea ce încetinește convergența din CLT

SFM_ch2_cont_stylised_facts

Faptul 3: absența autocorelației liniare

- **ACF** (autocorrelation function, funcția de autocorelație): $\rho(h) = \text{Corr}(r_t, r_{t-h})$, $h = 1, 2, \ldots$
  - versiunea de selecție $\hat{\rho}(h) = \sum_t (r_t - \bar{r})(r_{t-h} - \bar{r}) / \sum_t (r_t - \bar{r})^2$
  - h: lagul, în zile; $\rho(h) \in [-1; 1]$; pentru date i.i.d., $\hat{\rho}(h) \approx N(0, 1/n)$: banda de 95% $\pm 1{,}96/\sqrt{n}$
- Testul **LB** (Ljung–Box) Ljung & Box (1978) pentru H_0: $\rho(1) = \cdots = \rho(m) = 0$
  - $Q(m) = n(n+2) \sum_{h=1}^{m} \hat{\rho}(h)^2 / (n-h) \sim \chi^2(m)$; pentru $m = 10$, respingem dacă $Q > 18{,}3$
  - m: numărul de laguri testate împreună; Q adună pătratele autocorelațiilor, deci valorile mari arată că un $\rho(h) \neq 0$
- Explicația absenței: dacă randamentele ar fi previzibile, investitorii ar exploata acest lucru pînă cînd previzibilitatea dispare (piețe eficiente, Capitolul 7)

Faptele 3 și 4: ACF a randamentelor și a randamentelor absolute



- Stînga: randamentele; majoritatea autocorelațiilor sînt în bandă sau aproape de ea; S&P 500 $\hat{\rho}(1) = -0{,}11$, BET 0,03, Bitcoin $-0{,}02$
- Dreapta: randamentele absolute; toate pozitive, cu scădere lentă: S&P 500 0,31 la lagul 1, 0,26 la lagul 10, 0,05 la lagul 50

SFM_ch2_cont_stylised_facts

Faptul 3: autocorelația slabă a randamentelor

- Ljung–Box $Q(10)$ pentru randamente: S&P 500 199, BET 46, Bitcoin 20; valoarea critică 18,3
  - formal, unele autocorelații sînt semnificative
- Două motive pentru a nu exagera importanța rezultatului
  - valorile sînt mici din punct de vedere economic: $|\hat{\rho}(h)| \leq 0{,}11$, deci se explică cel mult 1,2% din varianță
  - banda $\pm 1{,}96/\sqrt{n}$ presupune randamente i.i.d.; cu volatility clustering este prea îngustă, deci „semnificativ" apare prea des
- Valoarea S&P 500 la lagul 1 provine mai ales din martie 2020, cînd scăderile mari și revenirile au alternat: fără 20 februarie–30 aprilie 2020, $\hat{\rho}(1) = -0{,}04$
- Teste robuste ale previzibilității (testele raportului variațelor): Capitolul 7

Faptul 4: volatility clustering

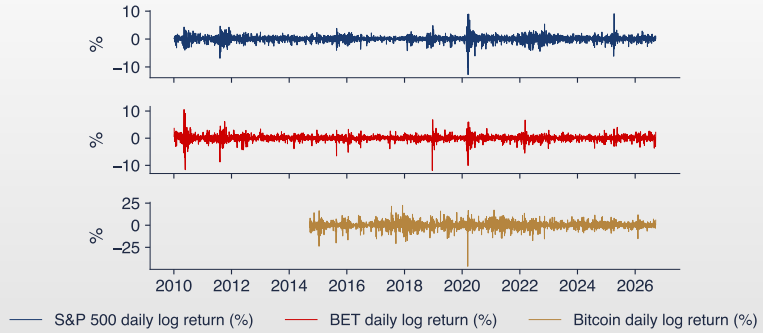
- „Variațiile mari tind să fie urmate de variații mari – de orice semn –, iar variațiile mici tind să fie urmate de variații mici" Mandelbrot (1963)
- Semnul randamentului de mîine este imprevizibil; mărimea lui nu este
- Dovezi: ACF a lui $|r_t|$ pozitivă pe 50 de laguri
  - Ljung–Box $Q(10)$ pentru $|r_t|$: S&P 500 3859, BET 2120, Bitcoin 1086
- Deci randamentele sînt necorelate, dar **nu independente**
- Modele: ARCH și GARCH (Capitolul 9)



Benoit Mandelbrot (1924–2010)

Foto: Rama (2007); CC BY-SA 2.0 fr; Wikimedia Commons

Faptul 4: perioade calme și perioade agitate



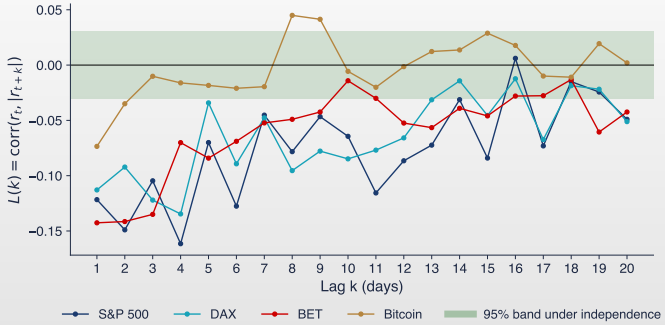
- Perioade agitate: 2010–2011 (criza datoriilor din zona euro), decembrie 2018 pentru BET, martie 2020 peste tot, aprilie 2025
- Între ele, ani calmi: volatilitatea unei zile depinde de perioada în care se află

SFM_ch2_cont_stylised_facts

Faptul 5: efectul de levier

- Măsura: $L(k) = \text{Corr}(r_t, |r_{t+k}|)$, $k = 1, 2, \ldots$
  - $L(k) < 0$: o scădere azi este urmată de variații mai mari decît o creștere de aceeași mărime
- Explicația prin efectul de levier financiar Christie (1982)
  - cînd prețul acțiunii scade, raportul datorii/capital propriu crește, deci capitalul propriu devine mai riscant
- O a doua explicație, volatility feedback: o volatilitate așteptată mai mare mărește randamentul cerut și reduce prețurile
- Modele cu răspuns asimetric (EGARCH, GJR-GARCH): Capitolul 9

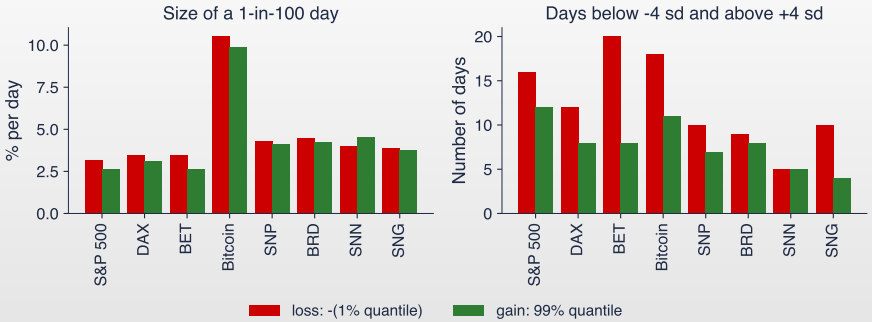
Faptul 5: corelațiile de levier



- Indicii bursieri: $L(k) < 0$ pentru majoritatea primelor 20 de laguri; media $L(1..5)$: S&P 500 $-0{,}12$, DAX $-0{,}10$, BET $-0{,}11$
- Bitcoin: $L(1) = -0{,}07$, apoi în interiorul benzii: în spatele Bitcoin nu există o firmă cu datorii, deci mecanismul de levier lipsește

SFM_ch2_cont_stylised_facts

Faptul 6: asimetria cîștig/pierdere



- Stînga: la indici, pierderea depășită într-o zi din 100 este mai mare decît cîștigul depășit într-o zi din 100: BET 3,45% față de 2,66%, S&P 500 3,15% față de 2,61%
- Dreapta: zile sub -4σ față de peste $+4\sigma$: BET 20 față de 8, S&P 500 16 față de 12, Bitcoin 18 față de 11
- Excepție: SNN, cu asimetrie pozitivă și cuantila din dreapta mai mare

SFM_ch2_sigma_days

Cele șase fapte stilizate în opt serii

Seria	Exces bolt.	Exces bolt., 21 de zile	$\hat{\rho}_r(1)$	$\hat{\rho}_{\lvert r\rvert}(10)$	$\bar{L}(1..5)$	$\lvert q_{0,01}\rvert / q_{0,99}$
S&P 500	13,5	3,7	−0,11	0,26	−0,12	1,21
DAX	7,4	5,5	0,01	0,16	−0,10	1,12
BET	17,9	8,3	0,03	0,20	−0,11	1,30
Bitcoin	12,0	1,0	−0,02	0,12	−0,03	1,07
OMV Petrom (SNP)	8,9	3,3	−0,02	0,13	−0,03	1,04
BRD	7,3	0,4	0,03	0,11	−0,05	1,07
Nuclearelectrica (SNN)	5,1	2,3	0,09	0,08	−0,03	0,88
Romgaz (SNG)	5,3	1,4	0,07	0,08	−0,04	1,02

- Coloanele: excesul de boltire al randamentelor zilnice și pe 21 de zile; ACF a randamentelor la lagul 1 și a lui $|r_t|$ la lagul 10; media $L(k)$ pe lagurile 1–5; cuantila de 1% împărțită la cea de 99%
- Faptul 1 este valabil pentru toate seriile, faptul 2 cu estimări imprecise; faptul 3 doar aproximativ, $|\hat{\rho}_r(1)|$ pînă la 0,11 (S&P 500; SNN 0,09); faptul 4 peste tot, cel mai slab pentru SNN și SNG (0,08)
- Faptul 5: clar pentru indici (între $-0{,}12$ și $-0{,}10$), la fel de slab pentru Bitcoin, SNP, SNN și SNG (între $-0{,}04$ și $-0{,}03$); faptul 6 nu este valabil pentru SNN (0,88), aproape 1 pentru SNP și SNG
- O comparație între distribuțiile randamentelor criptomonedelor și cele ale activelor tradiționale: Pele et al. (2023)

SFM_ch2_cont_stylised_facts

Implicațiile faptelor stilizate pentru modele

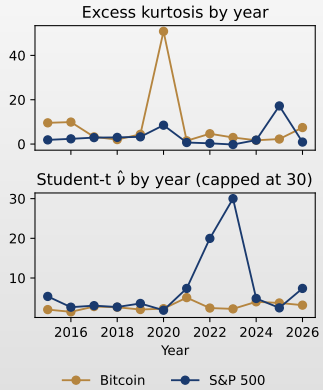
Model	1	2	3	4	5	6
i.i.d. Normal	–	trivial	✓	–	–	–
i.i.d. Student-t	✓	✓	✓	–	–	–
GARCH cu erori t (cap. 9)	✓	✓	✓	✓	–	–
EGARCH, GJR-GARCH (cap. 9)	✓	✓	✓	✓	✓	(✓)

- Coloanele: cele șase fapte în ordinea din această secțiune; (✓): doar pentru randamente pe mai multe zile, dacă erorile nu sînt asimetrice
- Măsurile de risc (VaR 1%, ES 2,5%, Capitolul 10) moștenesc calitatea modelului: un VaR Normal este prea mic în crize
- O măsură înrudită a riscului de coadă, entropia informațională: Pele, Lazar & Dufour (2017)

Recapitulare: fapte stilizate

- Cozi groase: o zi de 4 sigma la fiecare cîteva luni, în loc de una la cîteva decenii
- Randamentele sînt aproape necorelate, dar mărimea lor este puternic autocorelată
- Pe orizonturi mai lungi, randamentele se apropie lent de distribuția Normală
- Indicii bursieri: scăderile cresc volatilitatea, iar pierderile sînt mai mari decît cîștigurile

O întrebare de cercetare deschisă: evoluția cozilor Bitcoin



- Pe măsură ce o piață se maturizează (mai mulți participanți, futures, ETF-uri), cozile ei s-ar putea subția
- Bitcoin, 2015–2026: $\hat{\nu}$ anual între 1,5 și 5,1; corelația rangurilor cu anul 0,63 (p-value 0,03)
- De ce rămîne deschisă: 12 estimări anuale imprecise, un an (2020, exces de boltire 51) dominat de o singură zi
- Instrumentele AI pot accelera un astfel de studiu; nu înlocuiesc verificarea lui Wang et al. (2023)

SFM_ch2_tails_by_year

Utilizări posibile ale instrumentelor AI

- **Literatura**: găsirea studiilor despre tail index-urile activelor cripto și rezumarea metodelor lor
- **Cod**: o primă versiune a estimărilor pe ferestre mobile pentru ν și excesul de boltire
- **Robustețe**: propunerea altor ferestre, altor măsuri ale cozilor (Capitolul 5), altor active cripto
- Exemplu de prompt
  - `Write a Python function that fits a Student-t distribution to daily Bitcoin log returns on rolling 365-day windows and returns the degrees of freedom with a bootstrap 95% interval.`

Verificări necesare

- Datele: aceeași coloană de preț și același calendar pentru fiecare an; erori printre cele mai mari variații
- Estimarea: a convers optimizatorul? $\hat{\nu}$ aproape de 2 sau peste 30 trebuie reverificat
- Inferența: o singură estimare $\hat{\nu}$ pe an este imprecisă; raportați intervale, nu doar estimări punctuale
- Volatility clustering: cozi anuale mai subțiri pot însemna doar ani mai calmi (Capitolul 9)
- Referințele: fiecare lucrare citată trebuie să existe; verificați DOI-ul

Idee de proiect

- **Întrebarea**: se subțiază cozile activelor cripto pe măsură ce piețele lor se maturizează?
  - date: Bitcoin, Ethereum și Solana din datele cursului; S&P 500 ca reper
- Pași
  - estimați $\hat{\nu}$ și excesul de boltire pe ferestre mobile de un an, cu intervale bootstrap
  - testați existența unui trend; repetați pe randamente împărțite la o volatilitate mobilă
  - comparați numărul zilelor de 4 sigma pe an cu S&P 500
- Livrabile: un tabel, un grafic și un paragraf despre ce pot și ce nu pot arăta datele
- Declarați orice folosire a AI și listați erorile AI pe care le-ați corectat

Formule de reținut

Mărimea	**Formula**
Densitatea Normală	$f(x) = (\sigma\sqrt{2\pi})^{-1} \exp\big(-(x-\mu)^2/(2\sigma^2)\big)$
Momentele lognormale	mediana e^m, media $e^{m+s^2/2}$
CLT	$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$
Asimetria, excesul de boltire	$S = m_3/m_2^{3/2}, \quad K = m_4/m_2^2 - 3$
Jarque–Bera	$\mathrm{JB} = \frac{n}{6}(S^2 + K^2/4) \sim \chi^2(2)$
Student-$t(\nu)$	varianța $\nu/(\nu-2)$, excesul de boltire $6/(\nu-4)$
AIC	$2k - 2\ell$
ACF, Ljung–Box	$\hat{\rho}(h), \quad Q(m) = n(n+2)\sum_{h=1}^{m} \hat{\rho}(h)^2/(n-h)$
Levier	$L(k) = \mathrm{Corr}(r_t, \lvert r_{t+k}\rvert)$

Autoevaluare

- **Întrebare**: randamentele zilnice sînt $N(0{,}05\%, 1\%^2)$; la cîte abateri standard se află o zi de -3%?
  - **Răspuns**: $z = (-3 - 0{,}05)/1 = -3{,}05$
- **Întrebare**: $n = 2500$, $S = -0{,}5$, $K = 6$; cît este JB?
  - **Răspuns**: $\frac{2500}{6}(0{,}25 + 9) = 3854$; normalitatea este respinsă
- **Întrebare**: un randament logaritmic anual este $N(5\%, 20\%^2)$; care sînt mediana și media valorii unei investiții de 100 după un an?
  - **Răspuns**: mediana $100e^{0{,}05} = 105{,}1$, media $100e^{0{,}05+0{,}02} = 107{,}3$
- Urmează: Capitolul 3, distribuții α-stabile

Bibliografie (1/2)

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Bachelier, L. (1900). Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure*, 17, 21–86.
- Blattberg, R.C., Gonedes, N.J. (1974). A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business*, 47(2), 244–280.
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69(3), 542–547.
- Borak, S., Härdle, W.K., López-Cabrera, B. (2013). *Statistics of Financial Markets: Exercises and Solutions*, 2nd ed. Springer.
- Christie, A.A. (1982). The stochastic behavior of common stock variances: value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4), 407–432.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.
- Fama, E.F. (1965). The behavior of stock-market prices. *Journal of Business*, 38(1), 34–105.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). *Statistics of Financial Markets: An Introduction*, 5th ed. Springer.
- Jarque, C.M., Bera, A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255–259.
- Jarque, C.M., Bera, A.K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172.

Bibliografie (2/2)

- Ljung, G.M., Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(4), 394–419.
- Osborne, M.F.M. (1959). Brownian motion in the stock market. *Operations Research*, 7(2), 145–173.
- Pele, D.T., Lazar, E., Dufour, A. (2017). Information entropy and measures of market risk. *Entropy*, 19(5), 226.
- Pele, D.T., Wesselhöfft, N., Härdle, W.K., Kolossiatis, M., Yatracos, Y.G. (2023). Are cryptos becoming alternative assets? *European Journal of Finance*, 29(10), 1064–1105.
- Praetz, P.D. (1972). The distribution of share price changes. *Journal of Business*, 45(1), 49–55.
- Student (1908). The probable error of a mean. *Biometrika*, 6(1), 1–25.
- Tsay, R.S. (2010). *Analysis of Financial Time Series*, 3rd ed. Wiley.
- Wang, H., Fu, T., Du, Y., Gao, W., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620, 47–60.
- Wilk, M.B., Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55(1), 1–17.