








Statistica piețelor financiare

Capitolul 5: Cozi groase și teoria valorilor extreme



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

- **AI**: Artificial Intelligence (inteligență artificială)
- **BET**: Bucharest Exchange Trading (index) — indicele de referință al Bursei de Valori București
- **BET-TR**: BET Total Return (index) — indicele BET cu dividendele reinvestite
- **BVB**: Bursa de Valori București
- **CI**: Confidence Interval (interval de încredere)
- **COVID-19**: COronaVIrus Disease 2019 (boala provocată de noul coronavirus, identificată în 2019)
- **EODHD**: EOD Historical Data (market-data provider) — furnizor de date de piață
- **ES**: Expected Shortfall (pierderea așteptată în coadă)
- **EVT**: Extreme Value Theory (teoria valorilor extreme)
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity (heteroscedasticitate condiționată autoregresivă generalizată)
- **GEV**: Generalised Extreme Value (distribution) — distribuția generalizată a valorilor extreme
- **GPD**: Generalised Pareto Distribution (distribuția Pareto generalizată)
- **OMV**: Österreichische Mineralölverwaltung (Administrația austriacă a uleiurilor minerale, grupul OMV)
- **POT**: Peaks Over Threshold (metoda vîrfurilor peste prag)
- **PP**: Probability–Probability (plot) — graficul probabilitate–probabilitate
- **QQ**: Quantile–Quantile (plot) — graficul cuantilă–cuantilă
- **SE**: Standard Error (eroare standard)
- **SFE**: Statistics of Financial Markets, Exercises (Quantlet collection of the textbook) — colecția de Quantlet-uri a manualului Statistics of Financial Markets

Întrebarea de azi și traseul

- **Întrebarea**: cît de mare poate fi o pierdere zilnică și cît de des ar trebui să ne așteptăm la ea?
  - Capitolele 2 și 3: randamentele au cozi mai groase decît distribuția Normală
  - managerii de risc și autoritățile de reglementare privesc coada distribuției, nu centrul ei
- **Traseul** capitolului
  - trei crahuri: 1987, 2008, 2020
  - cozi groase și cozi subțiri; variația regulată și tail index-ul α
  - estimatorul Hill și funcția mean excess
  - block maxima și distribuția GEV; peaks over threshold și distribuția GPD
  - VaR 1%, ES 2,5% și VaR 0,1% prin EVT, comparate cu simularea istorică și cu distribuția Normală

Rezultatele învățării

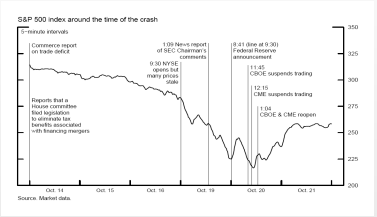
- Deosebiți o coadă groasă de una subțire și interpretați o coadă pe un grafic log-log
- Estimați tail index-ul cu estimatorul Hill, cu eroare standard și cu graficul Hill (Hill plot)
- Enunțați teoremele Fisher–Tippett–Gnedenko și Pickands–Balkema–de Haan în cuvinte și în formule
- Ajustați o distribuție GEV la maximele pe blocuri și o GPD la pierderile peste un prag; verificați ajustarea cu grafice PP și QQ
- Calculați return levels, VaR și ES dintr-o ajustare EVT, pas cu pas
- Judecați, pe date reale, cînd EVT aduce ceva în plus față de simularea istorică și de distribuția Normală

Bibliografie și instrumente

- Manual: Franke, Härdle & Hafner (2019), cap. 18 (statistica riscurilor extreme)
  - exerciții rezolvate: Borak, Härdle & López-Cabrera (2013), cap. 18
- Perspectiva managementului riscului: McNeil, Frey & Embrechts (2015), cap. 5 (teoria valorilor extreme)
- Cărți de referință: Embrechts, Klüppelberg & Mikosch (1997); Coles (2001)
- Quantlet-urile Python ale capitolului: Quantlets/Ch_05
  - portate din Quantlet-urile SFE SFElshill, SFEhillquantile, SFEMeanExcessFun, SFEevt1, SFEtailGEV, SFEtailGPareto, block_max și var_pot
- Notebook-ul cursului: deschideți în Google Colab
- Curs video: Measuring Statistical Risk

19 octombrie 1987: Lunea Neagră

- Pe 19 octombrie 1987, S&P 500 a scăzut cu aproximativ 20% într-o singură zi Carlson (2007)
  - ca randament logaritmic: $\ln(1 - 0{,}20) = -22{,}3\%$
  - mai mare decît orice pierdere zilnică din datele noastre, care încep în 1990 (cea mai mare: $-12{,}8\%$ pe 16 martie 2020)
- Evaluăm această zi cu distribuția Normală ajustată la S&P 500, 1990–2026
  - abaterea standard zilnică $\hat{\sigma} = 1{,}13\%$; $z = (r - \hat{\mu})/\hat{\sigma} = -19{,}7$ abateri standard
  - $P(Z \leq -19{,}7) \approx 10^{-86}$, cu $Z \sim N(0, 1)$: practic imposibil, și totuși s-a întîmplat
- Distribuția Normală nu spune nimic util despre astfel de zile: ne trebuie un model al cozii



S&P 500, 14–21 octombrie 1987, date la 5 minute (Federal Reserve)

Grafic: M. Carlson, Federal Reserve Board (2006); domeniu public; Wikimedia Commons

Septembrie 2008: Lehman Brothers

- Lehman Brothers a intrat în faliment pe 15 septembrie 2008; a urmat criza financiară globală
  - S&P 500: $-9{,}5\%$ pe 15 octombrie 2008 și $-9{,}4\%$ pe 1 decembrie 2008
  - BET: $-13{,}1\%$ pe 7 ianuarie 2009
- Pierderile mari au apărut grupat, nu izolat
  - volatility clustering (Capitolul 2) concentrează extremele în cîteva luni
- Un model VaR construit pe distribuția Normală atribuie acestor zile o probabilitate aproape nulă



Sediul Lehman Brothers, New York, 2007

Foto: David Shankbone (2007); CC BY-SA 3.0; Wikimedia Commons

Martie 2020: crahul COVID-19

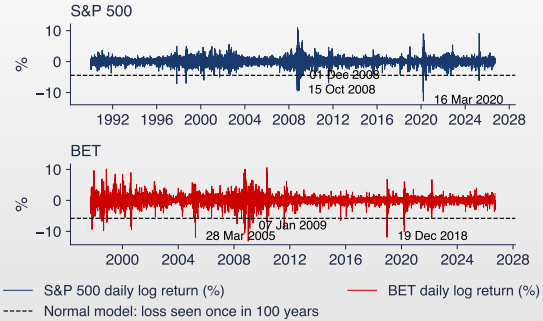
- Într-o singură săptămînă din martie 2020, trei dintre seriile noastre au înregistrat cea mai mare pierdere zilnică din întreaga perioadă
  - S&P 500: $-12{,}8\%$ pe 16 martie 2020
  - DAX: $-13{,}1\%$ pe 12 martie 2020
  - Bitcoin: $-46{,}5\%$ pe 12 martie 2020
- Datele acestui capitol
  - prețuri de închidere zilnice de la EODHD; S&P 500 și DAX din 1990, BET din 1997, Bitcoin din 2014



Districtul financiar din New York, pustiu, pe 29 martie 2020

Foto: Billie Grace Ward (2020); CC BY 2.0; Wikimedia Commons

Zilele de crah față de distribuția Normală



- Linia punctată: pierderea zilnică depășită o dată la 100 de ani de distribuția Normală ajustată la fiecare serie (S&P 500: $-4{,}4\%$; BET: $-5{,}8\%$)
- Observat: 35 de astfel de zile pentru S&P 500 din 1990 și 34 pentru BET din 1997

SFM_ch5_crashes

Cele mai mari pierderi zilnice în ipoteza distribuției Normale

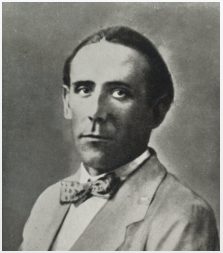
	cea mai proastă zi	pierdere (%)	z	distribuția Normală: o dată la (ani)	zile sub -4σ	așteptat
S&P 500	16 martie 2020	$-12{,}8$	$-11{,}3$	10^{27}	35	0,29
DAX	12 martie 2020	$-13{,}1$	$-9{,}5$	10^{19}	27	0,29
BET	7 ianuarie 2009	$-13{,}1$	$-8{,}9$	10^{16}	33	0,23
Bitcoin	12 martie 2020	$-46{,}5$	$-13{,}4$	10^{38}	18	0,14

- $z = (r - \hat{\mu})/\hat{\sigma}$, cu media și abaterea standard ale întregului eșantion
  - $p = \Phi(z)$: probabilitatea, în distribuția Normală, a unei zile cel puțin la fel de proaste; Φ: funcția de repartiție $N(0,1)$
  - o dată la $1/(p \times 252)$ ani (365 de zile pe an pentru Bitcoin)
- Timpii de așteptare implicați de distribuția Normală sînt absurzi: mult mai lungi decît istoria oricărei piețe
- Pierderi dincolo de 4σ: distribuția Normală prevede mai puțin de o zi pe serie; observăm între 18 și 35

SFM_ch5_crashes

Originile teoriei valorilor extreme

- **EVT** (extreme value theory, teoria valorilor extreme): statistica celor mai mari valori dintr-un eșantion
  - folosită întîi pentru inundații, furtuni și rezistența materialelor
  - Gumbel (1958): nivelul maxim anual al unui rîu determină înălțimea unui baraj
- În finanțe apar aceleași întrebări
  - cît de mare este pierderea depășită o dată la 10 ani?
  - de cît capital este nevoie pentru a acoperi cele mai proaste 1% dintre zile?
- EVT modelează doar coada și lasă datele din coadă să-i determine forma; primele aplicații în finanțe: Jansen & de Vries (1991), Longin (1996)



Emil Julius Gumbel (1891–1966)

Foto: autor necunoscut (c. 1931); domeniu public; Wikimedia Commons

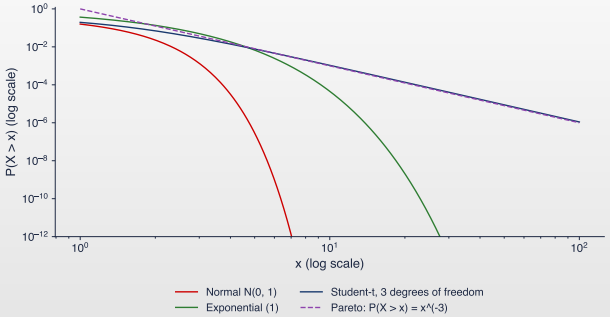
Rolul cozii în măsurarea riscului

- **VaR** (Value at Risk, valoarea expusă la risc) la nivelul p: pierderea depășită cu probabilitatea p
  - $\mathrm{VaR}_p = -q_p(r)$, cu q_p cuantila de ordin p a randamentului; de exemplu VaR 1%
- **ES** (expected shortfall, pierderea așteptată în coadă) la nivelul p: pierderea medie din cele mai proaste p dintre zile
  - $\mathrm{ES}_p = -E[r \mid r \leq q_p(r)]$; băncile raportează ES 2,5% Comitetul de la Basel (MAR33)
  - ES ține seama de întreaga coadă dincolo de VaR și este o măsură de risc coerentă Acerbi & Tasche (2002)
- Ambele sînt mărimi definite în coadă: un model care descrie greșit coada descrie greșit riscul
- **Întrebare pentru sală**: cu 250 de zile de tranzacționare pe an, de cîți ani de date este nevoie pentru circa 10 zile cu pierderi peste VaR 0,1%?
- **Răspuns**
  - $10/(0{,}001 \times 250) = 40$ de ani: doar datele istorice nu ajung; EVT extrapolează coada cu un model

Cozi subțiri și cozi groase

- Lucrăm cu pierderile $L = -r$ (în %): un L mare și pozitiv este o pierdere mare
  - **coada** lui L este descrisă de funcția de supraviețuire $\bar{F}(x) = P(L > x)$
- **Coadă subțire**: $\bar{F}(x)$ scade cel puțin la fel de repede ca o exponențială, e^{-cx}, pentru un $c > 0$
  - exemple: distribuția Normală ($\bar{F}(x) \approx e^{-x^2/2}$), distribuția exponențială (e^{-x}); toate momentele sînt finite
- **Coadă groasă (de tip putere)**: $\bar{F}(x)$ scade ca o putere, $x^{-\alpha}$
  - exemple: Pareto, Student-t, legile α-stabile din Capitolul 3
  - $\alpha > 0$ este **tail index-ul**: cu cît α este mai mic, cu atît coada este mai groasă

Cozile de tip putere și cozile subțiri



- Pe axe log-log, o lege de putere este o dreaptă cu panta $-\alpha$; cozile subțiri se curbează în jos și coboară rapid spre zero
- $P(X > 10)$: distribuția Normală $7{,}6 \times 10^{-24}$, distribuția exponențială $4{,}5 \times 10^{-5}$, Student-t cu 3 grade de libertate $1{,}1 \times 10^{-3}$

SFM_ch5_tails_loglog

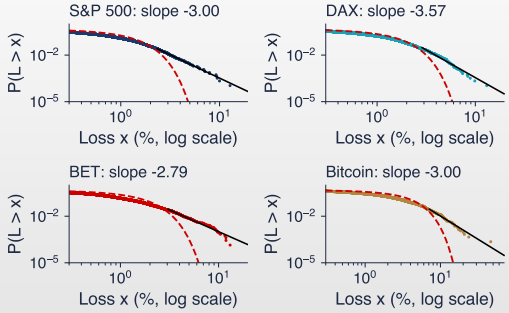
Variația regulată și tail index-ul

- L are o coadă **cu variație regulată**, de indice $\alpha > 0$, dacă $\bar{F}(x) = x^{-\alpha}\ell(x)$
  - în cuvinte: departe în coadă, $\bar{F}$ se comportă ca puterea $x^{-\alpha}$, pînă la un factor $\ell(x)$ care variază foarte lent
  - ℓ este **cu variație lentă**: $\ell(tx)/\ell(x) \to 1$ cînd $x \to \infty$, pentru orice $t > 0$ (de ex. o constantă, $\ln x$)
- Echivalent: $\dfrac{P(L > tx)}{P(L > x)} \to t^{-\alpha}$ pentru x mare
  - dublarea pierderii împarte probabilitatea ei la 2^{α}, oricare ar fi nivelul
  - Student-t cu 3 grade de libertate: $P(X > 20)/P(X > 10) = 0{,}128$, aproape de $2^{-3} = 0{,}125$
- Tail index-urile legilor uzuale
  - Pareto cu $\bar{F}(x) = (x/x_0)^{-\alpha}$: indicele α; Student-t cu ν grade de libertate: $\alpha = \nu$
  - α-stabilă cu $\alpha < 2$ (Capitolul 3): același α; distribuția Normală: fără coadă de tip putere ($\alpha = \infty$)

Tail index-ul și existența momentelor

- Pentru o coadă cu variație regulată: $E|L|^m < \infty$ dacă $m < \alpha$ și $E|L|^m = \infty$ dacă $m > \alpha$
  - $E|L|^m$: momentul de ordin m; este finit doar pentru ordine mai mici decît tail index-ul
  - $\alpha \leq 1$: nu există media; $\alpha \leq 2$: nu există varianța; $\alpha \leq 4$: nu există boltirea (kurtosis)
- „Legea cubică inversă" a randamentelor acțiunilor: $\alpha \approx 3$ Gopikrishnan et al. (1999)
  - varianța este finită, boltirea este infinită: boltirea de selecție din Capitolul 2 nu se stabilizează
- Legătura cu Capitolul 3
  - parametrul α al legii stabile ($\approx 1{,}5$) se estimează din întreaga distribuție; tail index-ul de aici folosește doar coada
  - un tail index în jur de 3 este argumentul împotriva varianței infinite, anunțat în Capitolul 3 Mandelbrot (1963)

Pierderi reale pe axe log-log



- Puncte: proporția zilelor cu o pierdere peste x; negru: dreapta celor mai mici pătrate prin cele mai mari 2,5% dintre pierderi; roșu: distribuția Normală
- Pante: S&P 500 $-3{,}00$, DAX $-3{,}57$, BET $-2{,}79$, Bitcoin $-3{,}00$: drepte, deci cozi de tip putere

SFM_ch5_tails_loglog

Interpretarea graficului log-log al cozii

- Dacă $\bar{F}(x) \approx Cx^{-\alpha}$, atunci $\ln \bar{F}(x) \approx \ln C - \alpha \ln x$: o dreaptă cu panta $-\alpha$
  - **estimatorul celor mai mici pătrate**: regresia lui $\ln(i/n)$ pe $\ln L_{(i)}$ pentru cele mai mari k pierderi $L_{(1)} \geq \cdots \geq L_{(k)}$
  - $L_{(i)}$ este a i-a cea mai mare pierdere, o **statistică de ordine**
- Corpul distribuției (pierderile mici) se depărtează de dreaptă: legea de putere este valabilă doar în coadă
- Coada distribuției Normale coboară sub date încă de la pierderi de 3–4%, pentru indici
- Panta celor mai mici pătrate este simplă, dar punctele folosite nu sînt independente; estimatorul Hill este instrumentul standard

SFM_ch5_tails_loglog

Exemplu lucrat: o coadă Pareto

- Pierderile zilnice au o coadă Pareto cu $\alpha = 3$ și $P(L > 3\%) = 1\%$
  - $P(L > x) = 1\% \times (x/3)^{-3}$ pentru $x \geq 3\%$
- Pierdere peste 6%: $1\% \times 2^{-3} = 0{,}125\%$; peste 9%: $1\% \times 3^{-3} = 0{,}037\%$
  - dublarea pierderii împarte probabilitatea la $2^3 = 8$
- **Ce credeți?** Ce pierdere este depășită cu probabilitatea 0,1% (VaR 0,1%)?
- **Răspuns**
  - rezolvăm $1\% \times (x/3)^{-3} = 0{,}1\%$: $x = 3 \times 10^{1/3} = 3 \times 2{,}15 = 6{,}46\%$
  - o probabilitate de zece ori mai mică crește pierderea doar cu factorul $10^{1/\alpha}$

Estimatorul Hill

- Ordonăm pierderile: $L_{(1)} \geq L_{(2)} \geq \cdots \geq L_{(n)}$; păstrăm cele mai mari k
  - **Estimatorul Hill** Hill (1975): $\hat{\alpha}_k = \left[\frac{1}{k} \sum_{i=1}^{k} \ln \frac{L_{(i)}}{L_{(k+1)}} \right]^{-1}$
  - în cuvinte: inversul mediei logaritmilor rapoartelor dintre cele mai mari k pierderi și prag
  - pragul este a $(k+1)$-a cea mai mare pierdere, $u = L_{(k+1)}$
- De ce funcționează: dacă $P(L > x \mid L > u) = (x/u)^{-\alpha}$ (Pareto peste u), atunci $\ln(L/u)$, condiționat de $L > u$, are distribuție exponențială cu media $1/\alpha$
  - media celor k excese logaritmice estimează $1/\alpha$; inversa ei estimează α
  - Hill = estimatorul de verosimilitate maximă al unei cozi Pareto peste u
- Portat din Quantlet-ul SFE SFElshill (estimatorii celor mai mici pătrate și Hill ai tail index-ului) Franke, Härdle & Hafner (2019)

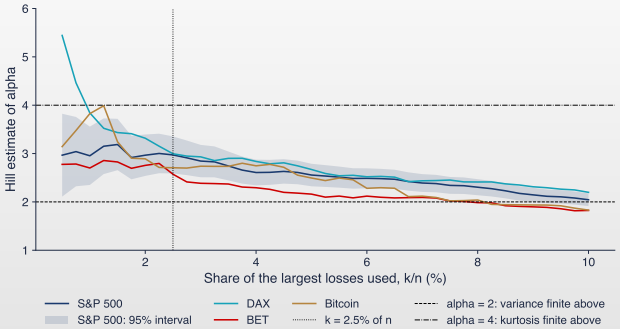
Exemplu lucrat: Hill pe șase pierderi

- Cele mai mari șase pierderi zilnice (%): 9,5; 7,2; 6,1; 5,4; 4,8; 4,0; luăm $k = 5$, deci $L_{(6)} = 4{,}0$
- Excesele logaritmice $\ln(L_{(i)}/4{,}0)$
  - 0,865; 0,588; 0,422; 0,300; 0,182; media 0,471
- $\hat{\alpha}_5 = 1/0{,}471 = 2{,}12$
- Eroarea standard în ipoteza de independență: $\mathrm{SE}(\hat{\alpha}) \approx \hat{\alpha}/\sqrt{k} = 0{,}95$
  - cu $k = 5$, estimarea este foarte imprecisă: aplicațiile reale folosesc sute de pierderi

Alegerea lui k: graficul Hill

- k **mic**: doar coada îndepărtată, puține puncte: variață mare
- k **mare**: intră corpul distribuției, unde legea de putere nu mai este valabilă: deplasare (bias)
- **Graficul Hill (Hill plot)**: $\hat{\alpha}_k$ în funcție de k; căutăm o zonă în care graficul este aproximativ orizontal Drees, Resnick & de Haan (2000)
  - aici k apare ca proporție din eșantion, k/n între 0,5% și 10%
- Fixăm regula înainte de a vedea rezultatul: aici $k = 2{,}5\%$ din n
  - și raportăm estimările pentru $k = 1\%$ și 5% din n, ca verificare

Grafice Hill ale pierderilor zilnice



- Între 1% și 5% din n, curbele sînt aproximativ orizontale, în mare parte între 2 și 3,5; banda: intervalul de 95% pentru S&P 500
- Peste 5%, estimările coboară: corpul distribuției introduce o deplasare în $\hat{\alpha}$

SFM_ch5_hill

Inferența pentru estimatorul Hill

- Pentru date i.i.d. cu coadă de tip Pareto: $\sqrt{k}\,(\hat{\alpha}_k - \alpha) \to N(0, \alpha^2)$, pentru $k \to \infty$, $k/n \to 0$
  - i.i.d.: independente și identic distribuite; pentru k mare, $\hat{\alpha}_k$ are aproximativ distribuție Normală în jurul lui α
  - $k/n \to 0$: numărul pierderilor din coadă crește, dar reprezintă o proporție tot mai mică din eșantion
  - $\mathrm{SE} = \hat{\alpha}/\sqrt{k}$; CI (confidence interval, intervalul de încredere) de 95%: $\hat{\alpha}(1 \pm 1{,}96/\sqrt{k})$
- Randamentele nu sînt independente: pierderile mari apar grupat (volatility clustering)
  - **bootstrap pe blocuri mobile**: reeșantionăm blocuri de 20 de zile consecutive, recalculăm $\hat{\alpha}$ și calculăm abaterea standard a estimărilor
  - blocurile păstrează episoadele grupate, deci eroarea standard ține seama de dependență
- Pentru fiecare serie: 500 de eșantioane bootstrap, cu k fixat pe slide-ul graficului Hill

Estimări Hill pentru patru piețe

	n	k	$\hat{\alpha}$	CI 95%	SE i.i.d.	SE blocuri	$k = 1\%$ / 5%	cîștiguri
S&P 500	9 240	231	2,97	[2,59; 3,35]	0,20	0,30	2,95 / 2,56	2,96
DAX	9 288	232	3,00	[2,61; 3,38]	0,20	0,20	3,84 / 2,67	3,22
BET	7 243	181	2,57	[2,20; 2,95]	0,19	0,23	2,70 / 2,16	2,82
Bitcoin	4 384	109	2,71	[2,20; 3,21]	0,26	0,23	3,83 / 2,49	3,37

- Pierderi: $\hat{\alpha}$ între 2,57 și 3,00
  - intervalele exclud 2 (varianță finită) și sînt sub 4 (boltire infinită)
- SE din bootstrap pe blocuri este apropiată de cea i.i.d. pentru DAX, BET și Bitcoin, dar mai mare pentru S&P 500 (0,30 față de 0,20); estimările variază cu k cu mai mult de o SE
- Ultima coloană: indicele Hill al cîștigurilor ($r_t > 0$), același k

SFM_ch5_hill

Estimatorul de cuantilă Hill (Weissman)

- **Cuantila Hill (Weissman)** Weissman (1978): $\hat{x}_p = L_{(k+1)} \left(\frac{k}{np} \right)^{1/\hat{\alpha}}$
  - $\hat{x}_p$: pierderea depășită cu probabilitatea p; k/n: proporția pierderilor de peste pragul $L_{(k+1)}$
  - extrapolată de la prag cu legea de putere; Quantlet-ul SFE SFEhillquantile
- **Exemplu lucrat**, S&P 500, VaR 0,1%: $n = 9\,240$, $k = 231$, $L_{(k+1)} = 2{,}34\%$, $\hat{\alpha} = 2{,}97$
  - $k/(np) = 231/9{,}24 = 25{,}0$; $25{,}0^{1/2{,}97} = 2{,}95$
  - $\hat{x}_{0{,}001} = 2{,}34 \times 2{,}95 = 6{,}92\%$; cuantila empirică de 0,1%: 6,94%
- Extrapolarea este doar atît de bună cît este $\hat{\alpha}$: o eroare de o SE în $\hat{\alpha}$ modifică vizibil $\hat{x}_p$

SFM_ch5_hill

Patru acțiuni de la Bursa de Valori București

	n	$\hat{\alpha}$	CI 95%	SE blocuri	cîștiguri
OMV Petrom (SNP)	3 815	2,96	[2,36; 3,56]	0,39	3,71
BRD	3 771	3,22	[2,57; 3,87]	0,33	3,77
Banca Transilvania (TLV)	3 788	2,49	[1,98; 2,99]	0,28	3,55
Transgaz (TGN)	3 725	3,02	[2,41; 3,64]	0,31	3,56

- Randamente logaritmice zilnice, calculate din prețul ajustat, din 2010; $k = 2{,}5\%$ din n
- TLV: zilele de 30–31 mai 2016 sînt eliminate, din cauza unei ajustări înregistrate cu o zi întîrziere (Capitolul 2)
- Toate cele patru acțiuni: $\hat{\alpha}$ între 2,49 și 3,22, ca la indici; pierderile au cozi mai groase decît cîștigurile

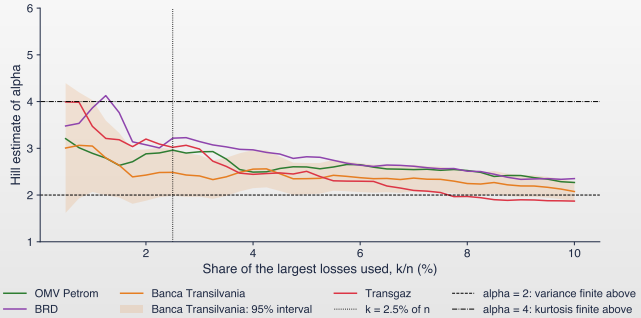


Palatul Bursei, București

Foto: Neoclassicism Enthusiast (2019); CC BY-SA 4.0; Wikimedia Commons

SFM_ch5_hill

Grafice Hill pentru acțiunile BVB



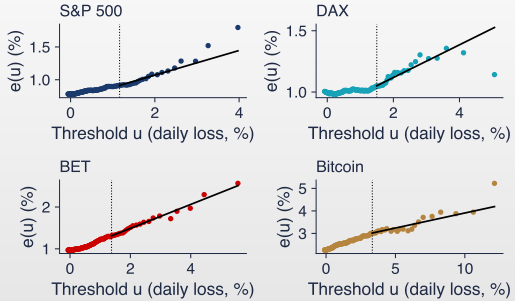
- Eșantioane mai scurte ($n \approx 3\,800$): benzi mai largi (afișată pentru TLV) și mai multe oscilații la k mic
- Nicio acțiune nu are $\hat{\alpha}$ sub 2 în zona plată: variațele sînt finite

SFM_ch5_hill

Funcția mean excess

- **Funcția mean excess** (excesul mediu): $e(u) = E[L - u \mid L > u]$, pierderea medie peste u, știind că u este depășit
  - versiunea empirică: media lui $L_i - u$ pentru pierderile cu $L_i > u$
- Forma ei arată tipul de coadă
  - distribuția exponențială: $e(u)$ constantă (lipsa memoriei)
  - Pareto cu $\alpha > 1$: $e(u) = u/(\alpha - 1)$, o dreaptă crescătoare
  - distribuția Normală: $e(u)$ scade spre 0 (aproximativ $1/u$ pentru distribuția Normală standard)
- **Exemplu**: coadă Pareto cu $\alpha = 3$: $e(u) = u/2$
  - peste o pierdere de 2%, pierderea suplimentară medie este 1,0%; peste 4% este 2,0%: cu cît pragul este mai mare, cu atît excesul mediu este mai mare

Mean excess empiric al pierderilor zilnice



- Toate cele patru funcții cresc peste cuantila de 90% u (punctat): o coadă groasă, de tip Pareto; negru: dreapta modelului de coadă ajustat peste u (Secțiunea 6)
- Ultimele puncte se bazează pe foarte puține pierderi și sînt instabile: nu le interpretați literal

SFM_ch5_mean_excess

Interpretarea graficului mean excess

- Căutăm pragul peste care $e(u)$ devine aproximativ o dreaptă
  - o dreaptă crescătoare înseamnă o coadă groasă; panta ei dă o primă idee despre forma cozii
- La cuantila de 90%: $e(u)$ este 0,92% pentru S&P 500 ($u = 1{,}16\%$) și 1,30% pentru BET ($u = 1{,}38\%$)
  - într-o zi cu o pierdere peste u, excesul peste u este, în medie, aproximativ egal cu u
- Funcția mean excess este unul dintre instrumentele pentru alegerea pragului în metoda peaks over threshold (Secțiunea 6)
- Portată din Quantlet-ul SFE SFEMeanExcessFun

SFM_ch5_mean_excess

Maxime pe blocuri (block maxima)

- Împărțim pierderile în blocuri de aceeași lungime și păstrăm cea mai mare pierdere din fiecare bloc
  - $M_n = \max(L_1, \ldots, L_n)$, aici cea mai mare pierdere zilnică din fiecare lună calendaristică ($n \approx 21$)
- Întrebarea: care este legea lui M_n pentru n mare?
  - analogul teoremei limită centrale, pentru maxime în loc de sume
- Fisher & Tippett (1928) au găsit cele trei legi limită posibile, studiind cel mai mare element al unui eșantion



Ronald A. Fisher (1890–1962), în 1913

Foto: autor necunoscut (1913); domeniu public; Wikimedia Commons

Teorema Fisher–Tippett–Gnedenko

- $L_1, L_2, \ldots$ i.i.d. Dacă există constante $a_n > 0$, b_n astfel încît $P\Big(\dfrac{M_n - b_n}{a_n} \leq x\Big) \to H(x)$, nedegenerată,
  - atunci H este o distribuție **GEV** (generalised extreme value, distribuția generalizată a valorilor extreme) Fisher & Tippett (1928), Gnedenko (1943)
- Notațiile
  - M_n: maximul a n pierderi; b_n îl centrează, a_n îl rescalează, așa cum se standardizează $\bar{X}_n$ în teorema limită centrală
  - nedegenerată: H este o distribuție propriu-zisă, nu o masă concentrată într-un singur punct
- Oricare ar fi legea pierderilor, maximul normalizat are o singură familie de limite posibile

Distribuția GEV

- **GEV** cu forma ξ, locația μ, scala $\sigma > 0$ (forma unificată a lui Jenkinson (1955))
  - $H(x) = \exp\left\{ -\left[1 + \xi \frac{x-\mu}{\sigma}\right]^{-1/\xi} \right\}$ pentru $1 + \xi(x-\mu)/\sigma > 0$
  - $\xi = 0$: $H(x) = \exp\{-e^{-(x-\mu)/\sigma}\}$, limita pentru $\xi \to 0$
- Rolul fiecărui parametru
  - μ: deplasează distribuția (mărimea tipică a maximului)
  - σ: o dilată (dispersia maximelor)
  - ξ: forma cozii drepte; cu cît ξ este mai mare, cu atît coada este mai groasă

Cele trei tipuri ale distribuției GEV

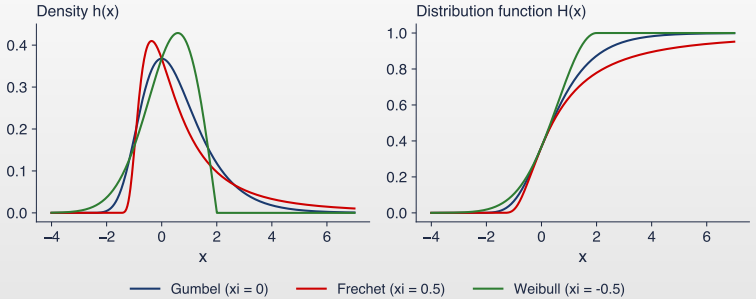
- **Fréchet** ($\xi > 0$): coadă groasă, $1 - H(x) \sim x^{-1/\xi}$
  - tail index-ul $\alpha = 1/\xi$; cazul pierderilor financiare
- **Gumbel** ($\xi = 0$): coadă subțire, exponențială
  - maximele variabilelor cu distribuție Normală, lognormală sau exponențială
- **Weibull** ($\xi < 0$): un capăt superior finit $\mu - \sigma/\xi$
  - maximele variabilelor mărginite (uniformă, Beta)
- Mulțimea legilor ale căror maxime converg către un H dat este **domeniul de atracție al maximului** său



Maurice Fréchet (1878–1973)

Foto: autor necunoscut; domeniu public; Wikimedia Commons

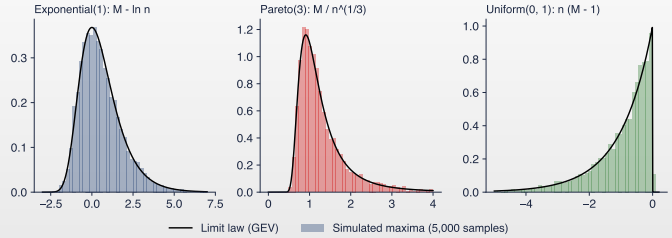
Cele trei legi ale valorilor extreme



- Parametri de locație și de scală standard ($\mu = 0$, $\sigma = 1$): Fréchet are cea mai lungă coadă dreaptă, Weibull se oprește la $x = 2$
- Portat din Quantlet-ul SFE SFEevt1

SFM_ch5_gev_block_maxima

Teorema într-o simulare



- 5 000 de eșantioane de $n = 250$; maximele exponențiale minus $\ln n \to$ Gumbel; maximele Pareto(3) împărțite la $n^{1/3} \to$ Fréchet cu $\xi = 1/3$; uniformă: $n(M_n - 1) \to$ Weibull
- **Întrebare pentru sală**: pierderile zilnice au un tail index în jur de 3; ce tip de distribuție au maximele lor lunare?
- **Răspuns**: Fréchet, cu $\xi = 1/\alpha \approx 0{,}3$

SFM_ch5_gev_block_maxima

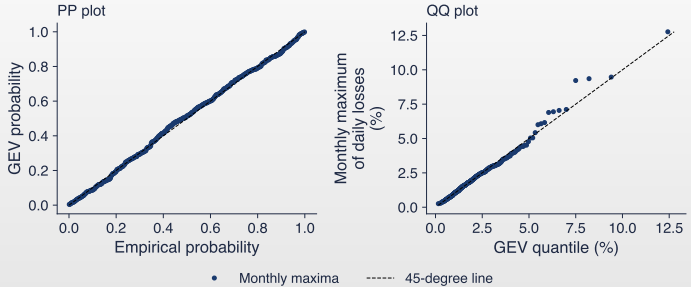
Ajustări GEV la maximele lunare

	luni	$\hat{\xi}$	CI 95%	$\hat{\mu}$ (%)	$\hat{\sigma}$ (%)	$1/\hat{\xi}$
S&P 500	440	0,21	[0,13; 0,29]	1,31	0,74	4,7
DAX	440	0,17	[0,09; 0,25]	1,69	0,91	5,9
BET	342	0,36	[0,26; 0,46]	1,47	0,92	2,7
Bitcoin	144	0,23	[0,08; 0,38]	5,19	2,85	4,3

- Verosimilitate maximă pe maximele lunare ale pierderilor zilnice; SE din curbura log-verosimilității
- $\hat{\xi} > 0$ pentru toate cele patru serii: tip Fréchet, cozi groase; intervalele exclud 0 (Gumbel)
- $1/\hat{\xi}$ este o a doua estimare a tail index-ului; cu o singură pierdere pe lună este mai puțin precisă decît estimatorul Hill

SFM_ch5_gev_block_maxima

Calitatea ajustării GEV: grafice PP și QQ



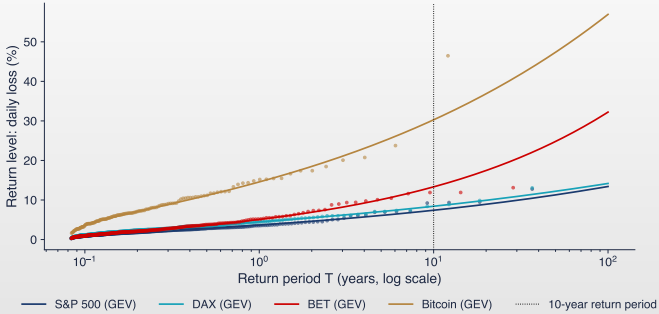
- **Graficul PP**: probabilitatea ajustată $H(m_{(i)})$ față de cea empirică $(i - 0{,}5)/n$; **graficul QQ**: maximele observate față de cuantilele ajustate; S&P 500
- Ambele grafice sînt aproape de prima bisectoare; în graficul QQ, cele cîteva luni extreme (2008, 2020) sînt puțin deasupra: este partea cea mai greu de ajustat

SFM_ch5_gev_block_maxima

Return levels (niveluri de revenire)

- **Return level** pentru o perioadă de revenire de T ani: nivelul pe care maximul lunar îl depășește, în medie, o dată la T ani
  - $z_T = H^{-1}(1 - 1/m)$ cu $m = 12T$ luni: $z_T = \mu + \frac{\sigma}{\xi}\left[(-\ln(1 - 1/m))^{-\xi} - 1\right]$
- **Exemplu lucrat**, S&P 500, $T = 10$ ani: $\hat{\xi} = 0{,}211$, $\hat{\mu} = 1{,}311$, $\hat{\sigma} = 0{,}737$
  - $m = 120$; $-\ln(1 - 1/120) = 0{,}00837$; ridicat la $-\hat{\xi}$: $2{,}747$
  - $z_{10} = 1{,}311 + (0{,}737/0{,}211)(2{,}747 - 1) = 7{,}4\%$
- O pierdere zilnică de circa 7,4% în S&P 500 este un „eveniment de 10 ani"; a fost depășită în 4 luni din 36,7 ani

Graficul return levels



- Linii: return levels GEV; puncte: maximele lunare observate, la perioadele lor empirice de revenire
- Dincolo de lungimea eșantionului (36,7 ani pentru S&P 500, 12 pentru Bitcoin), curbele sînt extrapolare

SFM_ch5_gev_block_maxima

Return levels ale pierderilor zilnice: 1, 10 și 50 de ani

	1 an	10 ani	interval 95%	50 de ani	cea mai mare observată	ani de date
S&P 500	3,7	7,4	[6,2; 8,9]	11,3	12,8	36,7
DAX	4,5	8,4	[7,1; 9,9]	12,2	13,1	36,7
BET	5,1	13,3	[10,3; 17,5]	24,8	13,1	28,5
Bitcoin	14,6	30,3	[21,2; 41,1]	47,4	46,5	12

- Pierderea zilnică (%) depășită în medie o dată la T ani; intervalul: bootstrap al maximelor lunare (200 de eșantioane)
- BET: cu $\hat{\xi} = 0{,}36$, nivelul de 50 de ani este 24,8%, aproape dublul celei mai mari pierderi observate
  - o mică modificare a lui $\hat{\xi}$ schimbă mult valorile din coada îndepărtată

SFM_ch5_gev_block_maxima

Peaks over threshold (vîrfuri peste prag)

- **POT** (peaks over threshold, vîrfuri peste prag): fixăm un prag înalt u și păstrăm fiecare pierdere de peste el
  - **excesul** peste u: $Y = L - u$, știind că $L > u$
  - distribuția excesului: $F_u(y) = P(L - u \leq y \mid L > u)$
- Folosește toate pierderile mari, nu una pe bloc: mai multe date în coadă
- Metoda care stă la baza majorității estimărilor EVT pentru VaR și ES McNeil & Frey (2000), McNeil, Frey & Embrechts (2015)



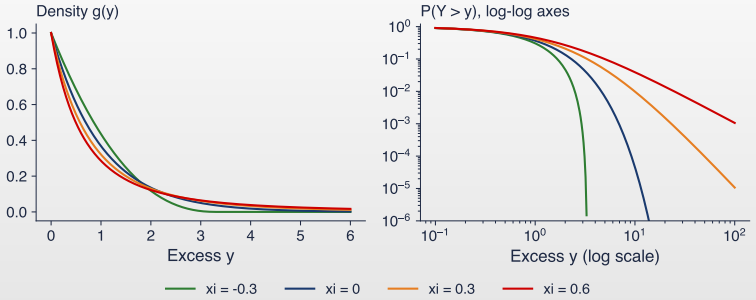
Laurens de Haan, Oberwolfach, 1987

Foto: Konrad Jacobs (1987), Oberwolfach Photo Collection; CC BY-SA 2.0 de; Wikimedia Commons

Teorema Pickands–Balkema–de Haan

- Dacă maximele lui L converg către o GEV cu forma ξ, atunci pentru praguri înalte
  - $F_u(y) \approx G_{\xi,\beta(u)}(y)$, **GPD** (generalised Pareto distribution, distribuția Pareto generalizată) Pickands (1975), Balkema & de Haan (1974)
- $G_{\xi,\beta}(y) = 1 - \left(1 + \xi \frac{y}{\beta}\right)^{-1/\xi}$, $y \geq 0$ (pentru $\xi < 0$: $y \leq -\beta/\xi$); $\xi = 0$: $1 - e^{-y/\beta}$
  - y: excesul peste u; forma ξ: **aceeași** ca la GEV; scala $\beta > 0$ depinde de u
  - $\xi > 0$: coadă de tip Pareto cu $\alpha = 1/\xi$; $\xi = 0$: exponențială; $\xi < 0$: mărginită
- Momentele GPD: media $\beta/(1-\xi)$ dacă $\xi < 1$; varianța finită dacă $\xi < 1/2$
  - funcția mean excess a unei GPD: $e(v) = \dfrac{\beta + \xi(v-u)}{1-\xi}$, o dreaptă în $v > u$, ca în Secțiunea 4

Densități și cozi GPD



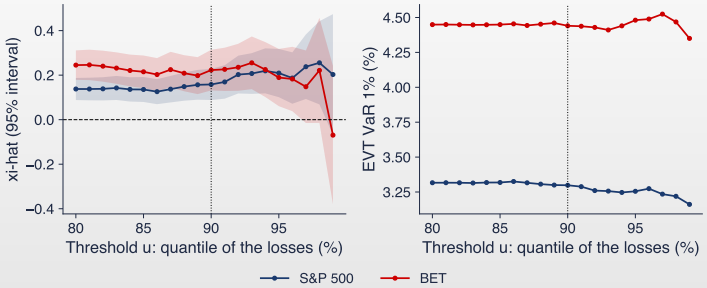
- $\beta = 1$; densitățile seamănă lîngă 0, cozile nu: pe axe log-log, doar $\xi > 0$ dă drepte
- $\xi = 0{,}3$ corespunde unui tail index $\alpha \approx 3{,}3$, apropiat de cel al randamentelor acțiunilor

SFM_ch5_pot_gpd

Alegerea pragului

- Același compromis ca la alegerea lui k pentru estimatorul Hill
  - u prea jos: aproximarea GPD este slabă (deplasare); u prea sus: puține excese (varianță)
- Instrumente grafice
  - graficul mean excess: aproximativ liniar peste u
  - graficul stabilității parametrilor: $\hat{\xi}$ nu ar trebui să se schimbe mult cînd u crește în continuare
- Regula folosită aici: $u =$ cuantila de 90% a pierderilor, adică cele mai mari 10% dintre pierderi
  - McNeil & Frey (2000) folosesc $k = 100$ de excese în ferestre de $n = 1000$ de zile, aceeași proporție

Stabilitatea parametrilor



- Stînga: $\hat{\xi}$ cu interval de 95% pentru u de la cuantila de 80% la cea de 99%; dreapta: VaR 1% EVT pentru aceleași praguri
- Între cuantilele de 85% și 95%: S&P 500 $\hat{\xi} \in [0{,}13; 0{,}22]$, VaR 1% $\in [3{,}25; 3{,}33]$%; VaR se modifică foarte puțin

SFM_ch5_pot_gpd

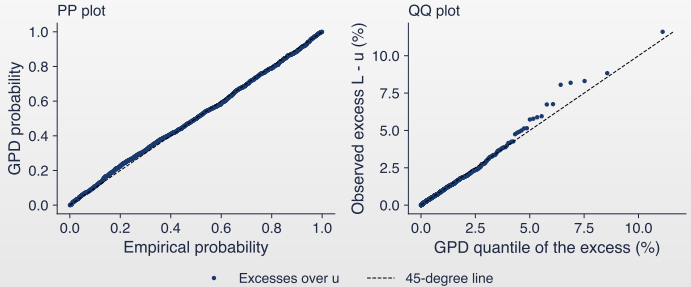
Ajustări GPD peste cuantila de 90%

	n	u (%)	N_u	$\hat{\xi}$	SE	$\hat{\beta}$	$1/\hat{\xi}$
S&P 500	9 240	1,16	924	0,158	0,04	0,769	6,3
DAX	9 288	1,50	929	0,118	0,04	0,927	8,5
BET	7 243	1,38	725	0,223	0,05	1,018	4,5
Bitcoin	4 384	3,33	439	0,118	0,05	2,663	8,5

- Verosimilitate maximă pe excesele $L - u$ Davison & Smith (1990); SE din curbura log-verosimilității
- $\hat{\xi}$ între 0,12 și 0,22: cozi groase cu varianță finită ($\xi < 0{,}5$) pentru toate cele patru serii
- $1/\hat{\xi}$ este mai mare decît $\hat{\alpha}$ Hill (2,5% din pierderi)
  - peste cuantila de 90%, o GPD cu scala liberă β are nevoie de un ξ mai mic
  - tail index-ul este greu de estimat precis: raportați mai mulți estimatori

SFM_ch5_pot_gpd

Calitatea ajustării GPD: grafice PP și QQ



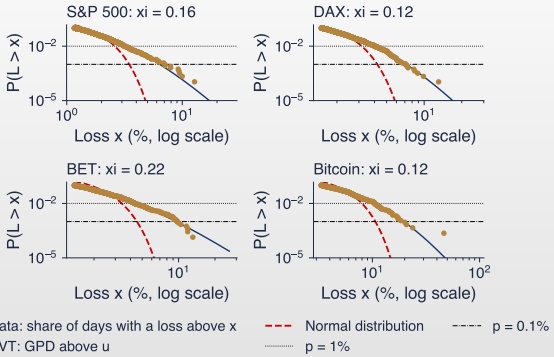
- Excesele S&P 500 peste u: punctele graficului PP stau pe dreaptă; cele ale graficului QQ, pînă la excese de circa 4%
- Cel mai mare exces (11,6%, martie 2020) față de o cuantilă ajustată de 11,1%; portat din Quantlet-urile SFE SFEtailGPareto_pp și _qq

SFM_ch5_pot_gpd

Estimatorul GPD al cozii pierderilor

- Pentru $x > u$: $P(L > x) = P(L > u)\, P(L > x \mid L > u)$
  - estimăm $P(L > u)$ prin N_u/n și $P(L > x \mid L > u)$ prin GPD
- **Estimatorul cozii**: $\hat{\bar{F}}(x) = \frac{N_u}{n}\left(1 + \hat{\xi}\,\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}$, $x > u$
  - N_u: numărul pierderilor de peste u; n: volumul eșantionului; paranteza este probabilitatea GPD de a depăși x
- În interiorul eșantionului folosim datele istorice; dincolo de el, o lege de putere ajustată
  - modelul se folosește doar acolo unde este justificat: peste u
- Inversarea lui $\hat{\bar{F}}(x) = p$ dă VaR; integrarea dincolo de el dă ES (Secțiunea 7)

Cozile ajustate



- Puncte: proporția zilelor cu o pierdere peste x (peste u); albastru: estimatorul EVT al cozii; roșu: distribuția Normală; linii punctate la $p = 1\%$ și 0,1%
- GPD urmează datele pînă la 10^{-4}; coada distribuției Normale se îndepărtează de ele înainte de $p = 1\%$

SFM_ch5_pot_gpd

Timpii de așteptare ai celor mai mari pierderi zilnice: distribuția Normală și EVT

	ziua	pierdere (%)	o dată la (ani) distribuția Normală	EVT
S&P 500	16 martie 2020	−12,8	10^{27}	89
DAX	12 martie 2020	−13,1	10^{19}	85
BET	7 ianuarie 2009	−13,1	10^{16}	12
Bitcoin	12 martie 2020	−46,5	10^{38}	237
S&P 500, Lunea Neagră	19 octombrie 1987	−22,3	10^{84}	1 602

- Timpul de așteptare EVT: $1/(\hat{\bar{F}}(x) \times 252)$ ani (365 pentru Bitcoin), cu ajustarea POT pe întregul eșantion
- Pentru EVT, zilele COVID-19 sînt evenimente rare, dar plauzibile, de tipul „o dată într-o viață"; pentru distribuția Normală sînt imposibile
- 1987 este în afara datelor noastre: EVT îl extrapolează la un timp de așteptare de 1 602 ani, cu o incertitudine mare

SFM_ch5_pot_gpd

Estimarea VaR și ES

- VaR și ES din Secțiunea 1, scrise pentru pierderile $L = -r$, cu funcția de repartiție F_L
  - $\mathrm{VaR}_p = F_L^{-1}(1-p) = -q_p(r)$
  - $\mathrm{ES}_p = \frac{1}{p}\int_0^p \mathrm{VaR}_s \, ds = E[L \mid L \geq \mathrm{VaR}_p]$ pentru distribuții continue
- Trei moduri de a le estima din pierderile zilnice
  - **simularea istorică**: cuantila empirică și media pierderilor de dincolo de ea
  - **distribuția Normală**: $\mathrm{VaR}_p = \hat{\mu}_L + \hat{\sigma} z_{1-p}$, $\mathrm{ES}_p = \hat{\mu}_L + \hat{\sigma}\, \varphi(z_{1-p})/p$
  - $\hat{\mu}_L$, $\hat{\sigma}$: media și abaterea standard a pierderilor; z_{1-p}: cuantila de ordin $1-p$ a distribuției $N(0,1)$; φ: densitatea distribuției Normale standard
  - **EVT (POT)**: inversăm estimatorul cozii
- Nivelurile folosite în practică: VaR 1% (zilnic, backtesting Basel), ES 2,5% Comitetul de la Basel (MAR33), VaR 0,1% (stres)

Formulele EVT pentru VaR și ES

- Rezolvăm $\frac{N_u}{n}\left(1+\xi\frac{x-u}{\beta}\right)^{-1/\xi} = p$ în raport cu x
  - $\mathrm{VaR}_p = u + \frac{\beta}{\xi}\left[\left(\frac{np}{N_u}\right)^{-\xi} - 1\right]$, valabil pentru $p < N_u/n$
  - pragul plus o cuantilă GPD a excesului; np/N_u: probabilitatea țintă raportată la proporția exceselor
- Excesul peste VaR_p are tot distribuție GPD, cu același ξ: media lui este $(\beta + \xi(\mathrm{VaR}_p - u))/(1-\xi)$
  - $\mathrm{ES}_p = \frac{\mathrm{VaR}_p}{1-\xi} + \frac{\beta - \xi u}{1-\xi}$, valabil pentru $\xi < 1$ McNeil & Frey (2000)
  - ES = VaR plus excesul mediu dincolo de VaR
- Cu $\xi > 0$, ES este un multiplu fix al VaR în coada îndepărtată: $\mathrm{ES}_p/\mathrm{VaR}_p \to 1/(1-\xi)$
- Portat din Quantlet-ul SFE var_pot (estimări VaR cu modelul POT)

Exemplu lucrat: S&P 500, pas cu pas

- Ajustarea: $n = 9\,240$, $u = 1{,}163\%$, $N_u = 924$ (10,0%), $\hat{\xi} = 0{,}158$, $\hat{\beta} = 0{,}769$, $\hat{\beta}/\hat{\xi} = 4{,}860$
- **VaR 1%**: $np/N_u = 0{,}100$; $0{,}100^{-\hat{\xi}} = 1{,}439$
  - $\mathrm{VaR}_{1\%} = 1{,}163 + 4{,}860\,(1{,}439 - 1) = 3{,}30\%$
- **ES 2,5%**: $np/N_u = 0{,}250$, $\mathrm{VaR}_{2{,}5\%} = 1{,}163 + 4{,}860\,(1{,}245 - 1) = 2{,}35\%$
  - $\mathrm{ES}_{2{,}5\%} = \dfrac{2{,}35 + 0{,}769 - 0{,}158 \times 1{,}163}{1 - 0{,}158} = 3{,}49\%$
- **VaR 0,1%**: $np/N_u = 0{,}010$, $0{,}010^{-\hat{\xi}} = 2{,}072$, $\mathrm{VaR}_{0{,}1\%} = 6{,}37\%$

VaR 1% și ES 2,5%: trei metode

	VaR 1% (%)			ES 2,5% (%)		
	EVT	istoric	Normal	EVT	istoric	Normal
S&P 500	3,30	3,15	2,60	3,49	3,48	2,62
DAX	3,95	4,02	3,16	4,13	4,15	3,18
BET	4,44	4,33	3,39	4,81	4,80	3,41
Bitcoin	10,38	10,51	7,99	10,90	10,84	8,03

- La aceste niveluri, EVT și simularea istorică diferă cu cel mult 0,14 puncte procentuale: există destule date la 1% și 2,5%
- Distribuția Normală subestimează riscul
  - VaR 1% EVT este cu 27% mai mare pentru S&P 500 și cu 31% mai mare pentru BET
  - ES 2,5%, cu o marjă asemănătoare

SFM_ch5_evt_var

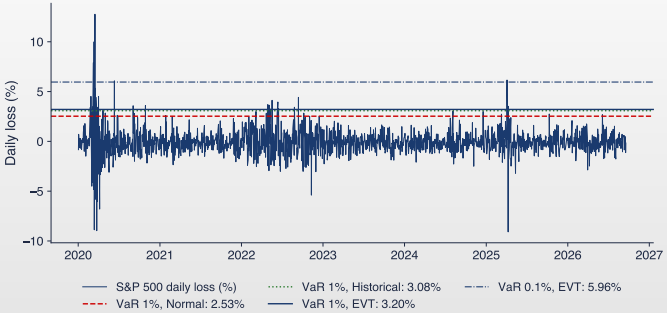
VaR la nivelul de 0,1%

	EVT (POT)	Hill	istoric	Normal	cea mai mare pierdere
S&P 500	6,37	6,92	6,94	3,47	12,8
DAX	7,17	8,37	6,97	4,21	13,1
BET	9,56	10,69	9,64	4,52	13,1
Bitcoin	19,61	23,30	18,05	10,65	46,5
OMV Petrom (SNP)	8,89	9,29	8,32	4,94	15,8
BRD	8,39	8,93	7,85	5,07	18,2
Banca Transilvania (TLV)	10,68	12,10	10,50	5,35	22,2
Transgaz (TGN)	8,29	9,18	8,23	4,68	10,1

- VaR 0,1% (%): pierderea depășită într-o zi din 1 000; acțiunile BVB din 2010
- VaR EVT este de circa 1,8 ori mai mare decît cel Normal pentru S&P 500 și de 2,1 ori pentru BET
- VaR 0,1% istoric se bazează doar pe cele mai mari 3–9 pierderi; EVT și Hill le netezesc cu un model al cozii

SFM_ch5_evt_var

În afara eșantionului: estimare pînă în 2019, test 2020–2026



- Pierderile zilnice ale S&P 500 după 2019, comparate cu VaR 1% al celor trei metode, toate estimate pe perioada 1990–2019
- Un VaR 1% bun este depășit în circa 1% din zilele de test, aici 16,9 zile

SFM_ch5_evt_var

Numărarea depășirilor

	zile de test	VaR 1% aşteptat	EVT	ist.	Normal	VaR 0,1% aşteptat	EVT	ist.	Normal
S&P 500	1 687	16,9	25	26	44	1,7	5	3	23
DAX	1 711	17,1	15	12	31	1,7	2	2	8
BET	1 675	16,8	4	5	13	1,7	0	1	4
Bitcoin	2 453	24,5	8	10	23	2,5	1	1	8

- VaR Normal este depășit mult prea des, mai ales la 0,1%: 23 de zile în loc de 1,7 pentru S&P 500
- VaR EVT și VaR istoric (ist.) sînt apropiate între ele; S&P 500 peste valoarea așteptată (episodul din 2020), BET sub ea (ani liniștiți)
- Un VaR estimat o singură dată ignoră variația volatilității în timp: EVT condiționată (GARCH + POT) McNeil & Frey (2000), Capitolele 9 și 10

SFM_ch5_evt_var

Posibilitățile și limitele EVT

- **Poate**: să ofere un model teoretic fundamentat al cozii și să extrapoleze dincolo de cea mai mare pierdere observată
  - VaR și ES stabile la 0,1%, acolo unde simularea istorică are doar cîteva puncte
- **Nu poate**: să elimine incertitudinea cozii îndepărtate
  - $\hat{\xi}$ are aici o eroare standard de 0,04–0,06; return levels dincolo de eșantion au intervale largi
- **Presupune**: pierderi i.i.d., sau măcar o coadă stabilă în timp
  - aplicați EVT pe randamente împărțite la o volatilitate GARCH (Capitolul 9) pentru riscul zilnic; pentru o alternativă, vedeți ES calculat pe baza entropiei Pele, Lazar & Mazurencu-Marinescu-Pele (2019)
- Sinteze despre EVT în finanțe: Rocco (2014); McNeil, Frey & Embrechts (2015), cap. 5

O întrebare de cercetare deschisă

- **S-a schimbat coada pierderilor BET pe măsură ce piața din București a crescut?**
  - indicele Hill, $k = 2{,}5\%$ din n, 1997–2009: $\hat{\alpha} = 2{,}90$ (SE pe blocuri 0,33)
  - 2010–2026: $\hat{\alpha} = 2{,}16$ (SE pe blocuri 0,27)
  - $z = (\hat{\alpha}_1 - \hat{\alpha}_2)/\sqrt{\mathrm{SE}_1^2 + \mathrm{SE}_2^2} = 1{,}73$: o coadă mai groasă după 2010, dar nesemnificativă la 5% ($|z| < 1{,}96$)
- Motivul pentru care rămîne deschisă: pragul a scăzut de la 3,95% la 2,01% odată cu volatilitatea
  - schimbarea este în coadă sau în volatilitate?
- Instrumentele AI pot accelera un astfel de studiu; nu înlocuiesc verificarea rezultatelor Wang et al. (2023)

SFM_ch5_hill

Utilizări posibile ale instrumentelor AI

- **Literatura**: lista studiilor despre tail index-urile de pe piețele emergente și rezumatul metodelor lor
- **Cod**: o primă versiune a unei analize Hill și POT pe ferestre mobile, cu intervale bootstrap pe blocuri
- **Robustețe**: alte praguri, randamente standardizate cu GARCH, alți indici BVB
- Exemplu de prompt
  - `Write a Python function that computes the Hill tail index of daily BET losses with k = 2.5% of n on rolling 5-year windows, with moving-block bootstrap standard errors (blocks of 20 days).`

Verificări necesare

- Semnul: pierderile sînt $-r$; o estimare Hill pe randamente, nu pe pierderi, măsoară cîștigurile
- Parametrizarea: `genextreme` din scipy folosește $c = -\xi$; `genpareto` folosește $c = \xi$
- Regula pentru k sau u se fixează înainte de a vedea rezultatul, nu se ajustează pentru a obține o schimbare semnificativă
- Ferestrele mobile se suprapun: estimările lor nu sînt teste independente
- Referințele: fiecare lucrare citată trebuie să existe; verificați DOI-ul

Idee de proiect

- **Întrebarea**: este coada pierderilor BET mai subțire sau mai groasă azi decît înainte de 2010, după ce eliminăm volatilitatea?
  - date: închiderile zilnice BET din 1997 (EODHD); opțional BET-TR din 2014
- Pași
  - Hill și POT (u la cuantila de 90%) pentru cele două perioade, cu intervale bootstrap pe blocuri
  - repetați pe randamente împărțite la o volatilitate mobilă pe 60 de zile: se menține diferența?
  - comparați cu S&P 500 pe aceleași perioade
- Rezultat: un tabel, un grafic și un paragraf despre concluziile pe care datele le susțin și pe care nu le susțin
- Declarați orice folosire a AI și listați erorile AI pe care le-ați corectat

Idei principale

- Pierderile zilnice au cozi de tip putere cu tail index-ul $\alpha \approx 3$: varianță finită, boltire infinită
- Hill: tail index-ul din cele mai mari k pierderi; k ales cu ajutorul graficului Hill; SE prin bootstrap pe blocuri
- Block maxima $\rightarrow$ GEV; excesele peste un prag $\rightarrow$ GPD; aceeași formă $\xi = 1/\alpha$
- Verificați fiecare ajustare cu mean excess, stabilitatea parametrilor, grafice PP și QQ
- VaR și ES EVT: apropiate de cele istorice la 1% și 2,5%, mult peste valorile distribuției Normale la 0,1%
- Extrapolarea mult dincolo de date rămîne incertă: raportați intervale

Formule de reținut

Mărimea	**Formula**
Variație regulată	$P(L > x) = x^{-\alpha}\ell(x), \quad P(L > tx)/P(L > x) \to t^{-\alpha}$
Hill	$\hat{\alpha}_k = [\frac{\mathbf{1}}{k}\sum_{i=\mathbf{1}}^{k} \ln(L_{(i)}/L_{(k+\mathbf{1})})]^{-\mathbf{1}}, \quad \mathrm{SE} \approx \hat{\alpha}/\sqrt{k}$
Cuantila Hill	$\hat{x}_p = L_{(k+\mathbf{1})}(k/(np))^{\mathbf{1}/\hat{\alpha}}$
Mean excess	$e(u) = E[L - u \mid L > u]; \quad$ GPD: $(\beta + \xi(v-u))/(1-\xi)$
GEV	$H(x) = \exp\{-[1 + \xi(x-\mu)/\sigma]^{-\mathbf{1}/\xi}\}$
Return level	$z_T = \mu + \frac{\sigma}{\xi}[(-\ln(1 - \frac{\mathbf{1}}{\mathbf{12}T}))^{-\xi} - 1]$
GPD	$G(y) = 1 - (1 + \xi y/\beta)^{-\mathbf{1}/\xi}$
VaR (POT)	$u + \frac{\beta}{\xi}[(np/N_u)^{-\xi} - 1]$
ES (POT)	$(\mathrm{VaR}_p + \beta - \xi u)/(1-\xi)$

Autoevaluare

- **Întrebare**: o ajustare GPD dă $\hat{\xi} = 0{,}25$; ce momente ale pierderilor există?
  - **Răspuns**: $\alpha = 1/\xi = 4$: există media, varianța și asimetria; boltirea nu (doar $m < 4$)
- **Întrebare**: de ce folosim cele mai mari 2,5% dintre pierderi și nu toate pierderile pentru estimatorul Hill?
  - **Răspuns**: legea de putere este valabilă doar în coadă; dacă includem corpul distribuției, $\hat{\alpha}$ este deplasat
- **Întrebare**: este VaR 0,1% prin simulare istorică fiabil cu 10 ani de date?
  - **Răspuns**: nu: circa 2 500 de zile conțin doar 2–3 pierderi dincolo de el; EVT folosește întreaga coadă
- Urmează: Capitolul 6, selecția modelului și managementul riscului

Bibliografie (1/2)

- Acerbi, C., Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7), 1487–1503.
- Balkema, A.A., de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2(5), 792–804.
- Basel Committee on Banking Supervision. MAR33: Internal models approach – capital requirements calculation. *The Basel Framework*. Bank for International Settlements.
- Borak, S., Härdle, W.K., López-Cabrera, B. (2013). *Statistics of Financial Markets: Exercises and Solutions*, 2nd ed. Springer.
- Carlson, M. (2007). A brief history of the 1987 stock market crash with a discussion of the Federal Reserve response. *Finance and Economics Discussion Series* 2007-13, Board of Governors of the Federal Reserve System.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.
- Daníelsson, J., de Vries, C.G. (1997). Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, 4(2–3), 241–257.
- Davison, A.C., Smith, R.L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Series B*, 52(3), 393–425.
- Drees, H., Resnick, S., de Haan, L. (2000). How to make a Hill plot. *Annals of Statistics*, 28(1), 254–274.
- Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.
- Fisher, R.A., Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 180–190.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). *Statistics of Financial Markets: An Introduction*, 5th ed. Springer.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3), 423–453.
- Gopikrishnan, P., Plerou, V., Amaral, L.A.N., Meyer, M., Stanley, H.E. (1999). Scaling of the distribution of fluctuations of financial market indices. *Physical Review E*, 60(5), 5305–5316.

Bibliografie (2/2)

- Gumbel, E.J. (1958). *Statistics of Extremes*. Columbia University Press.
- Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5), 1163–1174.
- Jansen, D.W., de Vries, C.G. (1991). On the frequency of large stock returns: putting booms and busts into perspective. *Review of Economics and Statistics*, 73(1), 18–24.
- Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348), 158–171.
- Longin, F.M. (1996). The asymptotic distribution of extreme stock market returns. *Journal of Business*, 69(3), 383–408.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(4), 394–419.
- 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.
- McNeil, A.J., Frey, R., Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*, revised ed. Princeton University Press.
- 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.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1), 119–131.
- Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer.
- Rocco, M. (2014). Extreme value theory in finance: a survey. *Journal of Economic Surveys*, 28(1), 82–108.
- Wang, H., Fu, T., Du, Y., Gao, W., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620, 47–60.
- Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364), 812–815.