








Modelarea piețelor financiare

Capitolul 3: Modele factoriale și evaluarea activelor



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

De ce au activele randamente medii diferite?



Eugene F. Fama. **Foto: Bengt Nyman (2013); CC BY 2.0; Wikimedia Commons**

"[W]hen the tests allow for variation in β that is unrelated to size, the relation between market β and average return is flat, even when β is the only explanatory variable."
— *Fama & French (1992)*

- Problema capitolului: ce determină diferențele dintre randamentele medii ale activelor?
  - 25 de portofolii americane sortate după capitalizare și raportul B/M (valoare contabilă / valoare de piață), iul. 1963 – iul. 2026
  - randamente medii în exces, anualizate: 3,9% pentru firmele mici de creștere (*small growth*), 13,3% pentru firmele mici de valoare (*small value*)
  - explică beta această diferență de aproximativ 9,5 puncte procentuale?
- Răspunsul CAPM (Capital Asset Pricing Model): prima de risc a unui activ este proporțională cu expunerea la riscul de piață, beta
- Structura argumentului
  - în CAPM, componenta randamentului în exces neexplicată de beta, alfa, este nulă
  - estimăm beta și alfa, apoi testăm ipoteza $\alpha = 0$ pe datele empirice
  - respingerea ipotezei conduce la modelele multifactoriale și la evaluarea robustetii factorilor propuși

▸ **Graficul central: dreapta pieței titlurilor (SML) în date**

Ce conține capitolul

Parcursul capitolului

1. Ce explică randamentele așteptate?
2. De la Markowitz la CAPM
3. Estimarea lui beta
4. Testarea CAPM
5. Modele multifactoriale
6. Regresii Fama–MacBeth
7. Testarea multiplă și replicarea factorilor
8. Investiții factoriale
9. Concluzii

Rezultatele învățării

- calculați un beta cu erori robuste la heteroscedasticitate și autocorelație și îl contractați spre medie (*shrinkage*)
- calculați testul simultan al valorilor alfa ale CAPM și prime de risc cu erori robuste
- interpretați un factor nou în raport cu pragurile testării multiple și cu dovezile de replicare
- interpretați un ETF (*Exchange-Traded Fund*, fond tranzacționat la bursă) factorial ca expuneri la factori și alfa

Subiecte suplimentare: Aprofundări; răspunsurile la verificările rapide: Anexă

Ce este un model factorial?

- Un **model factorial** explică randamentul activului i prin cîteva variabile comune $f_{1,t}, \ldots, f_{K,t}$, numite **factori**

$$R^e_{i,t} = R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,1} f_{1,t} + \cdots + \beta_{i,K} f_{K,t} + \varepsilon_{i,t}$$

  - $R^e_{i,t}$: randamentul în exces al activului i în perioada t; $R_{f,t}$: rata fără risc
  - K: numărul de factori
  - $\beta_{i,k}$: **încărcarea factorială** (*factor loading*): expunerea activului i la factorul k
  - α_i: abaterea **medie** a randamentului în exces de la cea explicată de factori
  - $\varepsilon_{i,t}$: abaterea din **perioada** t, specifică activului (idiosincratică), cu medie zero
- Dacă factorii sînt randamente în exces ale unor portofolii tranzacționate, un model care evaluează corect activul implică $\alpha_i = 0$

La ce folosește un model factorial?

- Ideea capitolului: randamentul mediu este **plata pentru risc**
  - investitorii cer o primă pentru riscurile comune: factorii (piața, mărimea, valoarea ...)
  - la expuneri egale la ceilalți factori, o expunere β mai mare la un factor ridică randamentul mediu dacă prima acelui factor este pozitivă
  - restricția de evaluare este $\mathbb{E}[R_i^e] = \boldsymbol{\beta}_i^\top \boldsymbol{\lambda}$, cu $\boldsymbol{\lambda}$ primele factorilor; ce rămîne, în medie, neexplicat de expuneri este α
- Două utilizări în acest capitol
  - Explicarea unui activ: pentru XLK (*Technology Select Sector SPDR Fund*), ETF-ul sectorului de tehnologie, unei creșteri de un punct procentual a randamentului în exces al pieței americane îi corespunde o creștere estimată de 1,28 puncte (Secțiunea 3)
  - Evaluarea unui fond: rămîne α_i pozitiv după ce ținem cont de expunerile lui (Secțiunile 4 și 8)?
- Datele și notațiile pentru randamente sînt în Aprofundări

Ce portofoliu deține un investitor medie–varianță?

- Întrebarea secțiunii: dacă investitorii aleg doar după medie și varianță, ce portofoliu dețin?
- A doua întrebare: ce randament mediu rezultă atunci pentru fiecare activ?
- Pasul 1: frontiera lui Markowitz (1952) dă portofoliul cu cel mai mare raport Sharpe
- Pasul 2: CAPM (Sharpe (1964); Lintner (1965)) spune că acest portofoliu este chiar portofoliul pieței
- Aplicăm pe cele nouă ETF-uri sectoriale din SUA (Statele Unite ale Americii) cu istoric lung (Select Sector SPDR), disponibile din decembrie 1998
  - sînt investibile, iar seriile lor sînt continue din 1999 pînă în 2026
  - reclasificările GICS (Global Industry Classification Standard) din 2016 (imobiliare) și 2018 (servicii de comunicații) au mutat companii în afara acestor nouă fonduri
  - de aceea, cele nouă nu mai acoperă întregul indice S&P 500 (*Standard & Poor's 500*), al celor 500 de mari companii americane

Frontiera medie–varianță

- Markowitz (1952): portofoliile de pe frontieră se obțin rezolvînd

$$\min_{\mathrm{w}} \ \mathrm{w}^\top \boldsymbol{\Sigma}\, \mathrm{w} \quad \text{cu condițiile} \quad \mathrm{w}^\top \boldsymbol{\mu} = \mu^*,\ \mathrm{w}^\top 1 = 1$$

  - w: ponderile
  - $\boldsymbol{\mu}$: randamentele în exces așteptate
  - $\boldsymbol{\Sigma}$: matricea de covarianță
- Ponderile negative sînt **poziții short**
- Soluția: $\mathrm{w}^* = \boldsymbol{\Sigma}^{-1}(\eta_1 \boldsymbol{\mu} + \eta_2 1)$
  - η_1, η_2: multiplicatorii Lagrange, fixați de cele două restricții
- Ponderile w^* sînt liniare în μ^*
  - orice portofoliu de pe frontieră este o combinație a două portofolii de pe frontieră (teorema celor două fonduri)
- În planul (σ_p, μ_p) frontiera este o hiperbolă
  - ramura de deasupra portofoliului de varianță minimă este **frontiera eficientă**

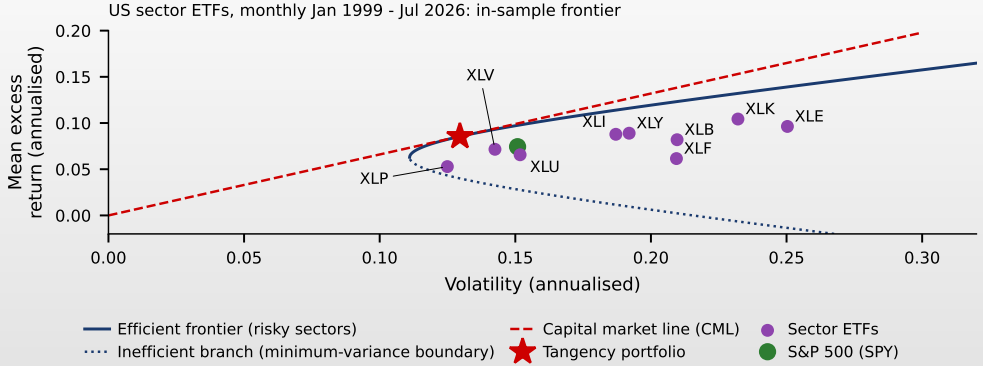
Raportul Sharpe și portofoliul tangent

- **Raportul Sharpe** $SR_p = \mathbb{E}[R_p - R_f]/\sigma_p$: randamentul în exces pe unitatea de volatilitate
- **Portofoliul tangent**: portofoliul cu cel mai mare raport Sharpe

$$w_{\text{tan}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \,/\, 1^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$$

  - formula este valabilă dacă $\boldsymbol{\Sigma}$ este pozitiv definită și numitorul este pozitiv
  - R_p, σ_p: randamentul și volatilitatea portofoliului p
  - numitorul rescalează $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ astfel încît ponderile să aibă suma 1
- **CML** (Capital Market Line, dreapta pieței de capital): $\mu_p = SR_{\text{tan}}\,\sigma_p$
  - toate combinațiile dintre activul fără risc și portofoliul tangent
- Dacă piața este portofoliul tangent, randamentul mediu al oricărui activ este liniar în beta
  - Seminarul 3, problema A2

Frontiera eficientă a sectoarelor din SUA (1/2)



Legenda: Efficient frontier (risky sectors): frontiera eficientă a sectoarelor; Inefficient branch: ramura ineficientă a frontierei de varianță minimă; Tangency portfolio: portofoliul tangent; in-sample frontier: frontiera estimată în eșantion (*in-sample*), pe tot eșantionul 1999–2026

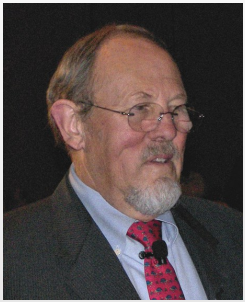
MFM_ch3_frontier ▶ Interactiv

Frontiera eficientă a sectoarelor din SUA (2/2)

- Axele graficului: volatilitatea și randamentul mediu în exces, anualizate
  - ca fracții, cu punct zecimal, ca pe axe ($0.10 = 10\%$)
  - 9 ETF-uri sectoriale, lunar, ianuarie 1999 – iulie 2026 ($T = 331$ luni)
- Portofoliul tangent (steaua roșie): medie 8,55%, volatilitate 12,96%, raport Sharpe 0,66
- SPY (cercul verde): medie 7,44%, volatilitate 15,08%, raport Sharpe 0,49
  - ETF-ul care replică S&P 500, folosit ca aproximare a pieței
- CAPM (slide-urile următoare) cere ca piața să fie chiar portofoliul tangent
  - în acest eșantion, SPY se află sub frontieră
- Raport Sharpe 0,66 față de 0,49 în acest eșantion: ar putea diferența să provină din eroarea de estimare?
  - portofoliul tangent este optimizat pe aceleași date pe care este evaluat
  - testul formal: Secțiunea 4
- Portofoliul tangent are poziții short mari
  - $-0{,}37$ în financiar (XLF) și $-0{,}21$ în materiale (XLB)
  - ce conține fiecare sector și ponderile complete: Aprofundări
- Ponderile extreme apar pentru că mediile sînt estimate cu zgomot (Capitolul 4)

MFM_ch3_frontier

CAPM: Sharpe, Lintner, Black



William F. Sharpe, Premiul Nobel pentru Economie 1990, cu Markowitz și Miller

Foto: Larry D. Moore (2007); CC BY 4.0; Wikimedia Commons

- **CAPM** a fost formulat de Sharpe (1964) și Lintner (1965)
- Toți investitorii dețin același portofoliu tangent
  - dacă toți aleg portofolii eficiente medie–varianță și au aceleași așteptări
- Cum fiecare activ trebuie deținut de cineva, portofoliul tangent este chiar **portofoliul pieței**
- Randamentul în exces așteptat este proporțional cu **beta**

$$\mathbb{E}[R_i] - R_f = \beta_i \left(\mathbb{E}[R_m] - R_f\right), \qquad \beta_i = \frac{\mathrm{Cov}(R_i, R_m)}{\mathrm{Var}(R_m)}$$

  - R_i, R_m: randamentele activului i și ale portofoliului pieței
  - $\mathbb{E}[R_m] - R_f$: **prima de risc a pieței**
  - β_i: sensibilitatea activului i la piață, $\beta_i > 1$ amplifică mișcările pieței
- Black (1972): CAPM cu beta zero (Aprofundări: o SML plată)
  - o versiune fără împrumut nelimitat la rata fără risc

Implicațiile CAPM

- Ipotezele CAPM
  - investitori medie–variață cu aceleași așteptări, o singură perioadă
  - împrumut și plasament nelimitate la R_f
  - fără taxe sau costuri de tranzacționare
- Doar beta aduce o **primă de risc**: un randament mediu mai mare pentru riscul purtat
  - la același beta, modelul prezice același randament așteptat, oricare ar fi celelalte caracteristici ale activului
- Riscul idiosincratic nu aduce nicio primă
  - se elimină prin diversificare, deci nimeni nu plătește pentru el
- **SML** (Security Market Line, dreapta pieței titlurilor): orice activ se află pe dreapta (beta, medie)
  - panta dreptei este egală cu prima pieței
  - SML este afirmația testabilă a CAPM (Secțiunea 4)
- Spre deosebire de SML, CML conține doar portofoliile eficiente
- Care portofoliu este „piața”?
  - răspunsul determină ce poate stabili un test al CAPM (critica lui Roll, Secțiunea 4)

De la CAPM la regresie: ce este alfa

- CAPM este o afirmație despre medii; exprimată în randamente în exces, relația nu are termen liber

$$\mathbb{E}[R_i^e] = \beta_i \, \mathbb{E}[R_m^e], \qquad R^e = R - R_f$$

- În date avem randamente lunare, deci estimăm o regresie, **modelul de piață**

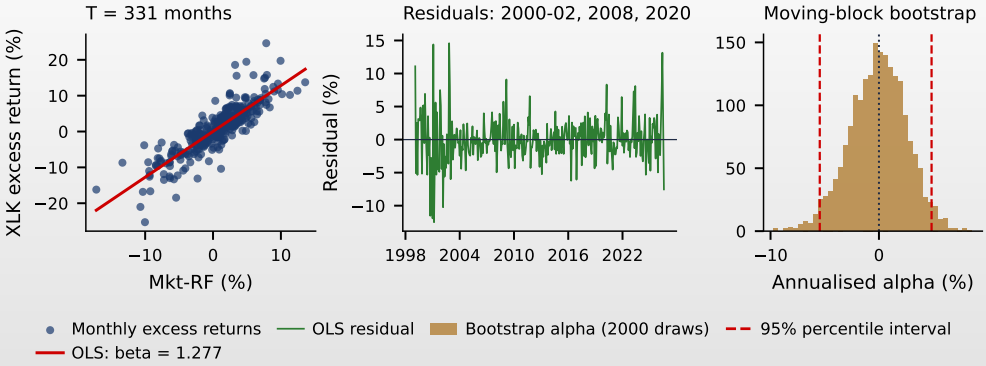
$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i \left(R_{m,t} - R_{f,t}\right) + \varepsilon_{i,t}$$

  - **alfa**, α_i: randamentul mediu în exces pe care beta nu îl explică
  - în medie, $\alpha_i = \mathbb{E}[R_i^e] - \beta_i \, \mathbb{E}[R_m^e]$: distanța verticală a activului față de SML
- Toată lumea vrea alfa
  - ce performanță rămîne după ajustarea pentru riscul de piață? Două fonduri cu același beta: cel cu randament mediu mai mare are alfa mai mare
  - expunerea la piață se poate obține cu costuri reduse printr-un fond indexat; administratorii activi sînt plătiți pentru alfa
- CAPM spune că, în echilibru, $\alpha_i = 0$ pentru orice activ
  - Secțiunea 3: estimăm β_i și α_i
  - Secțiunea 4: testăm dacă toate valorile alfa sînt zero

Cît de precis putem măsura beta?

- Întrebarea secțiunii: cu cît se mișcă un activ cînd piața se mișcă cu 1%?
  - cît de sigură este această cifră?
- Beta intră în costul capitalului și în acoperirea riscului
- Exemplul de lucru: XLK, ETF-ul de tehnologie, sectorul cu cel mai mare beta din eșantion
- Pașii secțiunii
  - erori standard robuste
  - de ce depinde un beta estimat
  - beta pe ferestre mobile (*rolling windows*)
  - contracția: estimațiile sînt trase spre media activelor
  - aplicarea la BVB (Bursa de Valori București)

Beta ETF-ului de tehnologie (XLK) în date



XLK pe Mkt-RF (randamentul în exces al pieței americane), randamente lunare în exces, ian. 1999 – iul. 2026. Stînga: dreapta OLS (*Ordinary Least Squares*, metoda celor mai mici pătrate), $\hat{\beta} = 1{,}277$; mijloc: reziduurile OLS, cu vîrfuri în 2000–2002, 2008 și 2020; dreapta: distribuția bootstrap (prin reeșantionare) a lui alfa (detalii în Aprofundări)

MFM_ch3_seminar ▶ Interactiv

Modelul de piață și erorile HAC

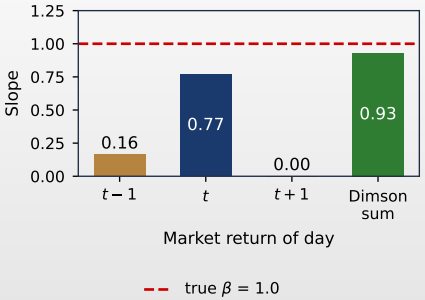
- Modelul de piață $R^e_{i,t} = \alpha_i + \beta_i R^e_{m,t} + \varepsilon_{i,t}$
- **OLS**: $\hat{\beta}_i = \widehat{\text{Cov}}(R^e_i, R^e_m)/\widehat{\text{Var}}(R^e_m)$
  - Ordinary Least Squares, metoda celor mai mici pătrate
- Descompunerea $\sigma^2_i = \beta^2_i \sigma^2_m + \sigma^2_{\varepsilon,i}$
  - risc sistematic (cu primă de risc în CAPM) plus risc idiosincratic (diversificabil)
  - σ^2_i, σ^2_m: variațele activului și ale pieței, $\sigma^2_{\varepsilon,i}$: varianța reziduurilor
  - $R^2 = \beta^2_i \sigma^2_m / \sigma^2_i$ este ponderea sistematică
- De ce erori **HAC** (Newey & West (1987))?
  - HAC: Heteroskedasticity and Autocorrelation Consistent, robuste la heteroscedasticitate și autocorelație
  - gruparea volatilității (*volatility clustering*, Capitolul 1) face reziduurile heteroscedastice
  - produsul $R^e_{m,t}\varepsilon_{i,t}$ poate fi autocorelat
- HAC schimbă matricea de covarianță estimată a coeficienților, nu estimația OLS
- XLK pe Mkt-RF, lunar 1999–2026: $\hat{\beta} = 1{,}277$ în toate cele trei cazuri
  - eroarea standard: 0,041 (clasică), 0,054 (White, doar heteroscedasticitate), 0,082 (Newey–West, 5 laguri)
  - ponderile, numărul de laguri și ipotezele: Aprofundări

De ce depinde un beta estimat (1/2)

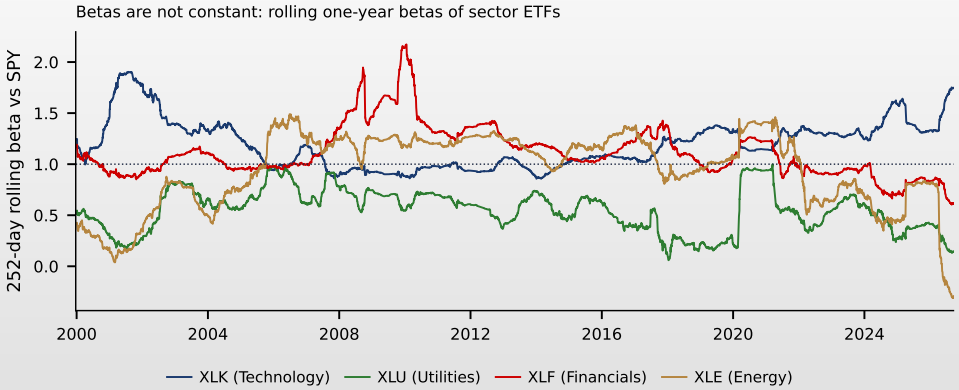
- Frecvența și fereastra: de obicei 3–5 ani de date lunare sau 1–2 ani de date zilnice
  - o fereastră mai lungă reduce eroarea standard
  - dar amestecă regimuri diferite ale lui beta (Beta nu este constant)
- Aproximarea pieței: SPY (S&P 500) sau Mkt-RF pentru active din SUA, BET (*Bucharest Exchange Trading*, indicele de referință al Bursei de Valori București) pentru acțiuni românești
- Curățarea datelor: un eveniment de capital neajustat creează un randament fals
  - Exemplu: 7 septembrie 2023 este data ex a dividendului special de 1,7225 lei pe acțiune al Fondului Proprietatea (FP)
  - În acea zi prețul de închidere scade cu 81% (randament fals, din prețuri), dar randamentul total al zilei, cu dividendul, este +6,0%

De ce depinde un beta estimat (2/2)

- Tranzacționarea asincronă: prețurile de închidere ale activelor nu sînt înregistrate în același moment
- O acțiune nelichidă reacționează cu întîrziere la știrile pieței, deci beta OLS zilnic este deplasat spre 0
- Beta Dimson (Dimson (1979)) regresează randamentul acțiunii pe trei randamente ale pieței
  - din ziua anterioară, din aceeași zi și din ziua următoare
  - $R_{i,t} = a + b_{-\mathbf{1}} R_{m,t-\mathbf{1}} + b_{\mathbf{0}} R_{m,t} + b_{+\mathbf{1}} R_{m,t+\mathbf{1}} + \varepsilon_{i,t}$
  - adună pantele: $\beta^D = b_{-\mathbf{1}} + b_{\mathbf{0}} + b_{+\mathbf{1}}$
- Graficul arată o simulare
  - o acțiune cu $\beta = 1$ se tranzacționează doar în 75% din zile
- Panta pe randamentul pieței din aceeași zi este 0,77, deci subestimează beta
- Panta pe randamentul pieței din ziua anterioară este 0,16
  - o parte din reacție vine cu o zi întîrziere
- Beta Dimson este 0,93, mult mai aproape de 1



Beta nu este constant (1/2)



MFM_ch3_rolling_beta ▶ Interactiv

Beta nu este constant (2/2)

- Graficul: beta OLS față de SPY pe ferestre mobile de 252 de randamente logaritmice zilnice (circa un an)
- Beta sectorului financiar (XLF) a variat între 0,60 și 2,17
- Beta XLF a atins maximul imediat după criza bancară din 2008–2009
- Beta sectorului tehnologic (XLK) a variat între 0,86 și 1,90
- Beta XLK a fost ridicat în anii bulei dot-com (2000–2001)
  - în ultima fereastră de 252 de zile este 1,75
- Beta sectorului energetic (XLE) a variat între $-0{,}31$ și 1,49
- Beta XLE este negativ în ultima fereastră
  - covarianța randamentelor XLE cu SPY a devenit negativă
- Beta utilităților (XLU), sectorul defensiv clasic, a variat între 0,06 și 1,00
- Un singur beta pe tot eșantionul ascunde aceste regimuri
- Modelele cu beta variabil în timp sînt tratate în Capitolele 5–6

MFM_ch3_rolling_beta

Contracția lui beta: Blume și Vasicek (1/2)

- **Regresia spre medie**: un beta estimat foarte mare sau foarte mic conține și zgomot
  - deci în perioada următoare tinde să se apropie de media secțiunii transversale, apropiată de 1
  - **contracția**: estimația este trasă dinainte spre această medie
- Blume (1971): regresia valorilor beta dintr-o perioadă pe cele din perioada anterioară
- De aici regula folosită de practicieni: $\beta^{\text{Blume}} = 0{,}33 + 0{,}67\,\hat{\beta}$
- Vasicek (1973): o medie ponderată între estimația activului și media $\bar{\beta}$

$$\beta_i^V = w_i \hat{\beta}_i + (1 - w_i)\bar{\beta}, \qquad w_i = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \text{se}(\hat{\beta}_i)^2}$$

  - $\bar{\beta}$: media valorilor beta ale tuturor activelor (media distribuției a priori)
  - w_i: ponderea estimației, mai mare cînd $\hat{\beta}_i$ este precis (SE mică)
  - $\text{se}(\hat{\beta}_i)$: eroarea lui standard
  - σ_β^2: varianța între active a parametrilor β_i, nu a estimărilor $\hat{\beta}_i$

Contracția lui beta: Blume și Vasicek (2/2)

- Varianța estimațiilor între active adună varianța adevărată și zgomotul
  - $\text{Var}_{cs}(\hat{\beta}) = \sigma_\beta^2 + \overline{\text{se}^2}$
  - $\overline{\text{se}^2}$: media pătratelor erorilor standard
- Deci $\sigma_\beta^2 = \text{Var}_{cs}(\hat{\beta}) - \overline{\text{se}^2}$ (estimația Bayes empirică)
  - derivarea: Seminarul 3, problema A3
- Exemplu: XLK față de SPY, lunar 1999–2012: $\hat{\beta} = 1{,}415$, SE Newey–West 0,139
  - alt reper (SPY, nu Mkt-RF) și altă perioadă (1999–2012) decît beta 1,277 din slide-urile anterioare
- Distribuția a priori, din cele 9 sectoare: $\bar{\beta} = 0{,}949$ și $\sigma_\beta^2 = 0{,}1159 - 0{,}0077 = 0{,}1082$
- Ponderea estimației proprii: $w = 0{,}1082/(0{,}1082 + 0{,}139^2) = 0{,}849$
  - beta Vasicek: $\beta^V = 0{,}849 \times 1{,}415 + 0{,}151 \times 0{,}949 = 1{,}344$
- Beta Blume: $0{,}33 + 0{,}67 \times 1{,}415 = 1{,}278$
- Beta estimat pe perioada următoare, 2013–2026: 1,172
  - ambele ajustări sînt mai aproape de el decît beta neajustat
  - pe toate cele 9 sectoare, rădăcina erorii pătratice medii a prognozei (RMSE, root mean squared error): OLS 0,135, Blume 0,109, Vasicek 0,114

Verificare rapidă: un beta, eroarea și contracția lui

- O acțiune, 120 de randamente lunare: $\hat{\beta} = 0{,}90$, SE clasică 0,05, SE Newey–West 0,09
  1. Calculați ambele intervale de 95%, cu multiplicatorul 1,96. Are acțiunea un risc sistematic mai mic decît piața (beta sub 1) după fiecare dintre ele?
  2. Ce interval raportați și de ce nu se schimbă estimația punctuală?
- O secțiune transversală de valori beta
  - $\mathrm{Var}_{cs}(\hat{\beta}) = 0{,}10$, $\overline{\mathrm{se}^2} = 0{,}02$, media $\bar{\beta} = 1{,}0$
  3. Calculați $\sigma^2_{\hat{\beta}}$ și beta Vasicek pentru un activ cu $\hat{\beta} = 1{,}6$, o dată cu SE 0,2 și o dată cu SE 0,1
  4. Care dintre cele două este contractată mai puternic și de ce? Comparați ambele cu beta Blume
- O acțiune nelichidă de la BVB, randamente zilnice, pantele pe BET
  - din ziua anterioară: 0,25; din aceeași zi: 0,55; din ziua următoare: 0,00
  5. Ce beta Dimson raportați?
  6. De ce este panta din aceeași zi mai mică decît beta Dimson?
- Răspunsuri: Anexă

Este CAPM respins de date?

- Întrebarea secțiunii: se află randamentele medii pe dreapta $\mathbb{E}[R_i^e] = \beta_i \, \mathbb{E}[R_m^e]$?
  - adică: au toate activele alfa zero?
- Active de test: cele 25 de portofolii mărime $\times$ B/M (Fama & French (1993))
- De ce aceste portofolii: contrastează sistematic mărimea și B/M, caracteristici asociate în literatură cu randamente medii diferite
  - randamente medii în exces, anualizate, între 3,9% și 13,3%
- Pașii secțiunii
  - SML în date
  - testul GRS (testul Gibbons–Ross–Shanken), care verifică simultan toate valorile alfa
  - limitele testului

Activele de test: 25 de portofolii mărime × B/M

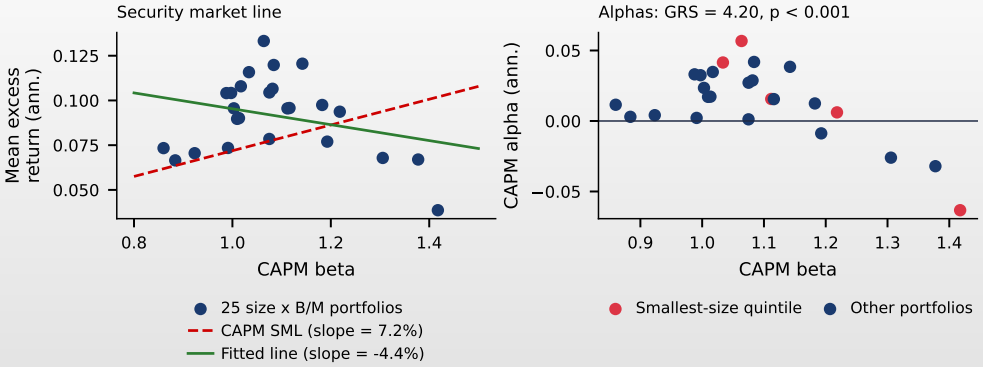
	BM1 (creștere)	BM2	BM3	BM4	BM5 (valoare)
ME1 (mici)	3,9	9,4	9,6	11,6	13,3
ME2	6,7	9,7	10,7	10,8	12,1
ME3	6,8	9,6	9,0	10,4	12,0
ME4	7,7	7,8	9,0	10,4	10,4
ME5 (mari)	7,3	7,1	7,3	6,7	9,6

randament mediu anual în exces, %; iul. 1963 – iul. 2026
ME: capitalizarea (*market equity*); BM: chintila de B/M

- În fiecare iunie t, acțiunile NYSE (*New York Stock Exchange*), AMEX (*American Stock Exchange*) și NASDAQ (*National Association of Securities Dealers Automated Quotations*) sînt sortate de două ori, independent (Fama & French (1993))
  - după mărime: capitalizarea de la sfîrșitul lui iunie t
  - după B/M: valoarea contabilă din anul fiscal $t-1$ împărțită la capitalizarea din decembrie $t-1$
- Pragurile sînt chintilele NYSE (percentilele 20, 40, 60 și 80)
  - altfel, microcapitalizările de la NASDAQ ar umple grupele
- Fiecare dintre cele 25 de intersecții este un portofoliu ponderat după capitalizare, deținut din iulie t pînă în iunie $t+1$
- Pe fiecare rînd, coloana de valoare (BM5) are un randament mediu mai mare decît cea de creștere (BM1); extremele: firmele mici de creștere (3,9% pe an) și de valoare (13,3%)
- Factorii SMB (*Small Minus Big*, mici minus mari) și HML (Secțiunea 5) folosesc aceleași caracteristici, deci explică aceste portofolii aproape prin construcție: un R^2 mare nu dovedește mult (Lewellen, Nagel & Shanken (2010))

▶ Interactiv

Dreapta pieței titlurilor (SML) în date (1/2)



Legenda: CAPM SML: dreapta CAPM, cu panta egală cu prima pieței; Fitted line: dreapta estimată pe date; Alphas: valorile alfa ale celor 25 de portofolii; Smallest-size quintile: cvintila firmelor cele mai mici

MFM_ch3_sml ▶ Interactiv

Dreapta pieței titlurilor (SML) în date (2/2)

- Cele 25 de portofolii, resortate în fiecare iunie și ponderate după valoare
  - iulie 1963 – iulie 2026, $T = 757$
- Dreapta CAPM are panta 7,2% pe an
  - randamentele și valorile alfa: fracții anuale, 0,05 = 5%; beta nu are unitate de măsură
- Dreapta estimată este regresia transversală a randamentelor medii pe beta
  - panta ei este $-4{,}4\%$
  - beta mai mare, randament mediu **mai mic**
- Valorile beta acoperă doar 0,86–1,42
- Alfa este $+5{,}7\%$ pe an pentru firmele mici cu B/M mare (de valoare)
  - și $-6{,}3\%$ pe an pentru firmele mici cu B/M mic (de creștere)
- 10 din 25 de valori alfa au $|t| > 1{,}96$ (erori HAC)
- În acest eșantion, portofoliile cu beta mai mare nu au avut, în medie, randamente mai mari
- Întrebarea următoare: pot fi toate aceste abateri compatibile cu zero?

MFM_ch3_sml

De la o singură valoare alfa la testul GRS (1/2)

- **Alfa lui Jensen** (Jensen (1968)): termenul liber al modelului de piață, α_i
  - CAPM implică $\alpha_i = 0$ pentru orice activ
- Un test t separat pentru fiecare activ: alfa XLK are $t = 0{,}13$
  - Cu multe active, unele valori alfa par semnificative doar din întîmplare (Secțiunea 7)
- **Testul GRS** (Gibbons, Ross & Shanken (1989)): un singur test pentru toate activele
  - $H_0 : \alpha_1 = \cdots = \alpha_N = 0$

$$W = \frac{T - N - 1}{N} \, \frac{\hat{\boldsymbol{\alpha}}^\top \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\alpha}}}{1 + \hat{\mu}_m^2 / \hat{\sigma}_m^2} \sim F(N,\, T - N - 1)$$

- Notațiile
  - N: numărul activelor de test
  - T: numărul de luni
  - $\hat{\boldsymbol{\alpha}}$: cele N valori alfa estimate
  - $\hat{\boldsymbol{\Sigma}}$: covariança reziduurilor estimată prin verosimilitate maximă (împărțire la T)
  - $\hat{\mu}_m / \hat{\sigma}_m$: raportul Sharpe al pieței

De la o singură valoare alfa la testul GRS (2/2)

- $\hat{\boldsymbol{\alpha}}^\top \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\alpha}}$: o sumă ponderată a pătratelor valorilor alfa
  - un alfa cîntărește mai mult cînd reziduurile activului sînt puțin volatile
- Decizia: H_0 se respinge cînd W depășește cuantila de 95% a distribuției $F(N, T-N-1)$
- Distribuția este exactă, condiționat de factori
  - erori i.i.d. cu distribuția Normală multivariată, independente de factori
  - medie zero și covariânță constantă nesingulară
- Forma cu rapoarte Sharpe (lunare, momente împărțite la T)

$$\hat{\boldsymbol{\alpha}}^\top \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\alpha}} = \widehat{SR}^2(f, \text{active}) - \widehat{SR}^2(f)$$

  - W este acest cîștig de raport Sharpe maxim la pătrat, normalizat: un W mare arată că piața este departe de frontiera construită din piață și activele de test
- Pentru $N = 1$, $W = t^2$, pătratul statisticii t clasice (OLS, nu HAC) a lui alfa
  - $F(1, T-2)$ este distribuția pătratului unei variabile Student-t cu $T-2$ grade de libertate
  - GRS generalizează testul t

Testul GRS pe cele 25 de portofolii

- Cele 25 de portofolii mărime $\times$ B/M, iulie 1963 – iulie 2026 ($T = 757$, $N = 25$)
  - $W = 4{,}20$, cu $p = 8 \times 10^{-11}$
- CAPM este respins: abaterile valorilor alfa sînt mari în raport cu incertitudinea lor
- Respingerea privește modelul împreună cu indicele folosit ca piață
  - critica lui Roll, slide-ul următor

Critica lui Roll

- Roll (1977): CAPM se referă la portofoliul **adevărat** al pieței, care cuprinde întreaga avuție
- Indicii bursieri folosiți ca piață omit
  - obligațiunile, activele imobiliare, firmele private și capitalul uman
- Beta calculat față de **orice** portofoliu eficient explică exact randamentele medii ale tuturor activelor
  - rezultat matematic: Seminarul 3, problema A2
- Deci un test CAPM verifică de fapt dacă indicele folosit ca piață este eficient
- Respingerea poate veni din model sau din alegerea indicelui

Alegerea activelor de test schimbă concluzia

- Pe cele 9 ETF-uri sectoriale din Secțiunea 2, cu Mkt-RF ca piață, CAPM nu este respins
  - ianuarie 1999 – iulie 2026, $T = 331$
  - GRS $= 0{,}53$, cu $p = 0{,}85$
- Adăugarea sectoarelor crește pătratul raportului Sharpe lunar cu 0,015
  - mai puțin decît 0,029, cît ar produce zgomotul singur (Anexă)
- Pe aceleași 331 de luni, cele 25 de portofolii mărime $\times$ B/M îl resping
  - GRS $= 2{,}91$, $p = 9 \times 10^{-6}$
- Sectoarele nu sînt construite pentru a contrasta mărimea și B/M
  - formarea după sector nu impune un contrast de capitalizare sau B/M comparabil cu sortarea 5×5
- Puterea testului depinde de abaterile de la model în raport cu covariația reziduurilor, de T și de N
- O nerespingere nu confirmă CAPM și nu dovedește singură lipsa de putere
- Activele de test se aleg pentru contraste economice stabilite înainte de test, nu după dispersia mediilor observate

Verificare rapidă: citirea unei statistici GRS

- Un factor, $N = 4$ active de test, $T = 65$ de luni, rapoarte Sharpe lunare
  - factorul singur: 0,15
  - portofoliul tangent al factorului și celor 4 active: 0,25
  1. Calculați W din forma cu rapoarte Sharpe. Cuantila de 95% a lui $F(4, 60)$ este circa 2,53: respingeți?
  2. Calculați creșterea așteptată sub H_0 a pătratului raportului Sharpe, $N(1 + \widehat{SR}^2(f))/(T - N - 3)$, și comparați-o cu cea observată
  3. Aceleași rapoarte Sharpe, $T = 650$ de luni: recalculați W (cuantila de 95% a lui $F(4, 645)$: circa 2,39). Ce s-a schimbat, distanța sau dovezile?
- Răspunsuri: Anexă

Teoria evaluării prin arbitraj

- Întrebarea secțiunii: beta de piață nu explică randamentele medii (Secțiunea 4)
  - de cîți factori și de care avem nevoie?
- **APT**: Arbitrage Pricing Theory (teoria evaluării prin arbitraj), Ross (1976)
- Ipoteza: randamentele urmează un model cu K factori

$$R_{i,t} = \mathbb{E}[R_i] + \boldsymbol{\beta}_i^\top \mathsf{f}_t + \varepsilon_{i,t}, \qquad \mathbb{E}[\mathsf{f}_t] = 0$$

  - riscurile specifice se pot elimina prin diversificare
  - $\boldsymbol{\beta}_i$: cele K încărcări ale activului i
  - f_t: componentele neanticipate ale factorilor
- **Absența arbitrajului** implică $\mathbb{E}[R_i] - R_f = \boldsymbol{\beta}_i^\top \boldsymbol{\lambda}$
  - absența arbitrajului: nu există profit sigur fără capital investit
- Relația este aproximativă: ține pentru aproape toate activele, nu exact pentru fiecare
- $\boldsymbol{\lambda}$: **primele de risc**, randamentul așteptat suplimentar pe unitatea fiecărei expuneri
- Teoria nu spune **care** sînt factorii: ei trebuie găsiți în date

Construcția factorilor: portofolii long–short

mărime: mediana NYSE; B/M: 30% / 40% / 30%

	Mic (creștere)	Neutru	Mare (valoare)
Mici	S_L	S_N	S_H
Mari	B_L	B_N	B_H

$$\text{SMB} = \tfrac{1}{3}(S_L + S_N + S_H) - \tfrac{1}{3}(B_L + B_N + B_H)$$

$$\text{HML} = \tfrac{1}{2}(S_H + B_H) - \tfrac{1}{2}(S_L + B_L)$$

- Un **factor construit prin sortare** este diferența dintre randamentele a două portofolii, $f_t = R_{\textbf{long},t} - R_{\textbf{short},t}$
  - ce se vinde finanțează ce se cumpără
- Sensul se fixează pentru fiecare factor
  - HML: B/M mare minus B/M mic
  - SMB: capitalizare mică minus capitalizare mare
- Fama–French: în fiecare iunie, acțiunile sînt împărțite în 2×3 grupe
  - cu datele contabile din anul anterior
  - fiecare grupă este un portofoliu ponderat după capitalizare
- Pragul de mărime este mediana capitalizării acțiunilor de la NYSE (New York Stock Exchange)
- Pragurile NYSE și ponderarea după capitalizare împiedică microcapitalizările să domine portofoliile
  - **SMB** (Small Minus Big): acțiuni mici minus acțiuni mari
  - **HML** (High Minus Low): B/M mare (valoare) minus B/M mic (creștere), în medie pe firme mici și mari, deci la aceeași mărime
- Aceeași construcție pe profitabilitate și investiții dă RMW și CMA (slide-ul următor; sensul fiecăruia este precizat acolo)

Factorii adăugați la modelul cu trei factori

- **RMW** (Robust Minus Weak): profitabilitate operațională mare minus mică
- **CMA** (Conservative Minus Aggressive): creștere mică minus mare a activelor
- **ROE** (Return On Equity, rentabilitatea capitalului propriu): factorul de profitabilitate al modelului q
- **MOM** (momentum, notat și UMD, Up Minus Down)
  - acțiunile cu randament mare minus cele cu randament mic în lunile $t-12$ pînă la $t-2$
  - ultima lună se omite
- Cîștigătorii trecuți continuă să cîștige 3–12 luni (Jegadeesh & Titman (1993))
- SMB din modelul cu trei factori (FF3) vine dintr-o singură sortare (mărime și B/M)
  - SMB din modelul cu cinci factori (FF5) este media a trei sortări
  - deci FF3 nu este un caz particular al FF5

Modelele factoriale standard

Model	Factori	Referință	Motivație
CAPM	Mkt-RF	Sharpe (1964)	echilibru
FF3	Mkt-RF, SMB, HML	Fama & French (1993)	efectele de mărime și B/M (Fama & French (1992))
Carhart	FF3 + MOM	Carhart (1997)	evaluarea fondurilor
modelul q	piață, mărime, investiții, ROE	Hou, Xue & Zhang (2015)	deciziile de investiții ale firmelor
FF5	Mkt-RF, SMB, HML, RMW, CMA	Fama & French (2015)	modelul de actualizare a dividendelor

- FF3 și FF5: modelele Fama–French cu trei, respectiv cinci factori
  - regresia FF3: $R^e_{i,t} = \alpha_i + b_i\,\mathrm{MKT}_t + s_i\,\mathrm{SMB}_t + h_i\,\mathrm{HML}_t + \varepsilon_{i,t}$
  - MKT = Mkt-RF; b_i, s_i, h_i: încărcările pe cei trei factori
- Un model bun lasă valori alfa apropiate de zero în regresiile pe serii de timp
  - GRS testează dacă abaterile sînt mari în raport cu incertitudinea lor
- Motivația FF5, modelul de actualizare a dividendelor: Aprofundări

Construcția factorului momentum

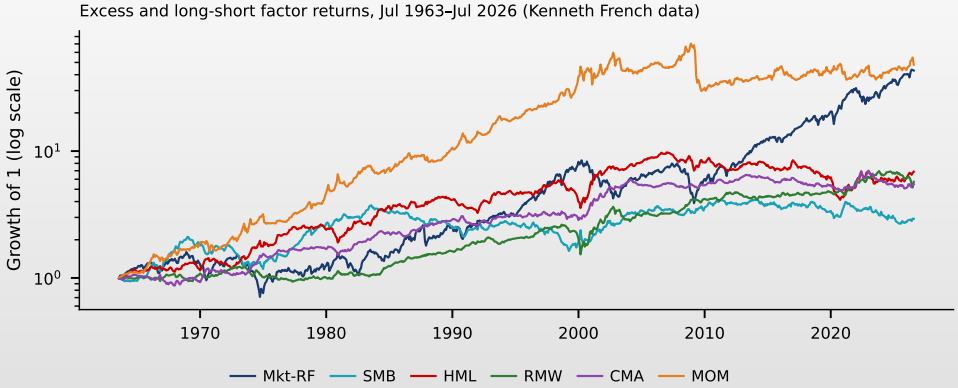
mărime: mediana NYSE
randament în $t-12 \ldots t-2$: 30% / 40% / 30%

	Perdanți	Neutru	Cîștigători
Mici	S_L	S_N	S_W
Mari	B_L	B_N	B_W

$\text{UMD} = \frac{1}{2}(S_W + B_W) - \frac{1}{2}(S_L + B_L)$

- Aceeași sortare 2×3 ca la SMB și HML, cu randamentul trecut în locul B/M
- Pragurile de randament sînt percentilele 30 și 70 ale acțiunilor NYSE
- Grupele se refac în fiecare lună, nu o dată pe an
  - randamentul trecut se schimbă lunar, datele contabile doar anual
- **Perioada de formare**: randamentul cumulat în lunile $t-12$ pînă la $t-2$, adică 11 luni
- Luna $t-1$ se omite din cauza **inversării pe termen scurt** (*short-term reversal*): pe o lună, randamentele se inversează (Jegadeesh (1990))
  - cauze: oscilația bid–ask (*bid-ask bounce*: prețul tranzacției alternează între prețul de cumpărare și cel de vînzare) și presiunea de lichiditate
  - inclusă, luna $t-1$ ar amesteca două efecte de semn opus
- UMD este momentum **transversal**: cîștigători minus perdanți la aceeași dată
- **Momentum-ul pe serii de timp** (*time-series momentum*) compară fiecare activ cu propriul trecut (Moskowitz, Ooi & Pedersen (2012))
  - long dacă randamentul pe 12 luni este pozitiv, short altfel

Randamentele cumulate ale celor șase factori, 1963–2026 (1/2)



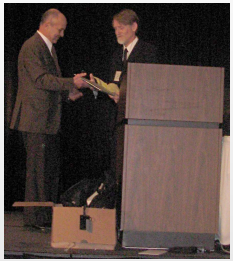
MFM_ch3_factor_cum ▶ Interactiv

Randamentele cumulate ale celor șase factori, 1963–2026 (2/2)

- Randamentele lunare ale factorilor, cumulate pornind de la 1, pe scară logaritmică
- Factorii sînt randamente în exces sau long–short
  - curbele nu includ dobînda pe capital sau pe garanția pozițiilor short
  - deci nu arată averea unui investitor real
- Valoarea finală a unei unități investite, iulie 1963 – iulie 2026
  - Mkt-RF 43,2, MOM 48,0, HML 6,9
  - RMW 5,7, CMA 5,5, SMB 2,9
- Momentum a atins maximul înainte de 2009
  - a pierdut mult cînd piața și-a revenit brusc în 2009: un **crah de momentum**
- Factorul valoare a stagnat sau a pierdut, aproximativ din 2007 pînă în 2020
- Randamente anuale medii
  - Mkt-RF 7,2%, MOM 7,3%, HML 3,6%
  - RMW 3,1%, CMA 3,0%, SMB 2,2%

MFM_ch3_factor_cum

Cele două etape



Eugene Fama, premiul Morgan Stanley–AFA (American Finance Association), 2008

Foto: TonyTheTiger (2008); CC BY-SA 3.0; Wikimedia Commons

- Întrebarea: aduce expunerea la un factor o primă de risc?
  - adică: explică valorile beta diferențele dintre randamentele medii, $\mathbb{E}[R_i^e] = \lambda_{\mathbf{0}} + \boldsymbol{\beta}_i^\top \boldsymbol{\lambda}$?
- De ce Fama & MacBeth (1973)?
  - factor tranzacționat: un model corect impune $\lambda = \mathbb{E}[f]$; estimația liberă poate diferi
  - factor netranzacționat (de exemplu, inflația): doar regresia transversală dă λ
- **Etapa 1** (serii de timp, cîte o regresie pe activ)
  - $R_{i,t}^e = a_i + \boldsymbol{\beta}_i^\top \mathrm{f}_t + \varepsilon_{i,t}$ dă matricea $\hat{\mathrm{B}}$ $(N \times K)$
- **Etapa 2** (transversal, cîte o regresie în fiecare lună pe $\mathrm{X} = [1, \hat{\mathrm{B}}]$)

$$R_{i,t}^e = \lambda_{\mathbf{0},t} + \hat{\boldsymbol{\beta}}_i^\top \boldsymbol{\lambda}_t + u_{i,t}, \qquad \hat{\boldsymbol{\lambda}} = \tfrac{\mathbf{1}}{T} \textstyle\sum_t \hat{\boldsymbol{\lambda}}_t$$

  - $\lambda_{\mathbf{0}}$: zero dacă modelul este corect; în versiunea Black (1972), randamentul portofoliului cu beta zero minus R_f
- Intuiția etapei 2: într-o lună, au cîștigat mai mult portofoliile mai expuse la HML?
  - panta acelei luni este prima realizată; media celor T pante este prima estimată
  - exemplul folosește beta din întregul eșantion: o descriere, nu o prognoză
  - pe cele 25 de portofolii: prima HML 4,0% pe an, media factorului 3,6%; cît de precisă este?

Erori standard pentru primele de risc (1/2)

- Patru surse de incertitudine, patru corecții
  - Fama–MacBeth: variația lunară a pantelor
  - Shanken: valorile beta sînt estimate
  - HAC: dependența în timp
  - Kan–Robotti–Shanken (KRS): posibile erori de evaluare (model greșit)
- **Fama–MacBeth**: estimațiile lunare $\hat{\lambda}_{k,t}$ formează un eșantion de volum T

$$\text{se}_{FM}(\hat{\lambda}_k) = s(\hat{\lambda}_{k,t})/\sqrt{T}$$

  - $s(\cdot)$: abaterea standard empirică pe cele T luni
  - robuste la corelația reziduurilor între active în aceeași lună
  - dacă $\hat{\lambda}_{k,t}$ este autocorelat, se folosește Newey–West pe această serie
- **Erori în variabile**: etapa 2 folosește estimările $\hat{\beta}_i$ în locul parametrilor β_i
  - deci se_{FM} subestimează incertitudinea

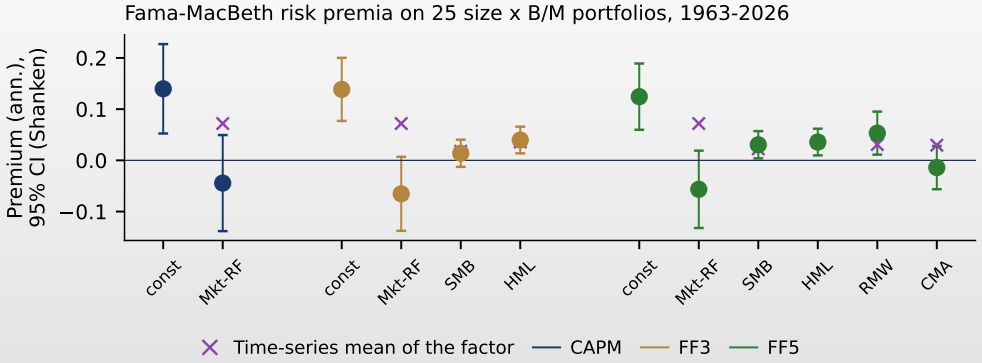
Erori standard pentru primele de risc (2/2)

- Corecția Shanken (1992) mărește cu factorul $(1 + c)$ partea variației care nu provine din variabilitatea factorilor

$$\text{Var}_{\text{Sh}}(\hat{\boldsymbol{\lambda}}) = (1 + c)\Big[\text{Var}_{FM}(\hat{\boldsymbol{\lambda}}) - \tfrac{1}{T}\boldsymbol{\Sigma}_f\Big] + \tfrac{1}{T}\boldsymbol{\Sigma}_f, \qquad c = \hat{\boldsymbol{\lambda}}^\top \boldsymbol{\Sigma}_f^{-1}\hat{\boldsymbol{\lambda}}$$

  - Var_{FM}: varianța Fama–MacBeth, $s^2(\hat{\boldsymbol{\lambda}}_t)/T$
  - $\boldsymbol{\Sigma}_f$: covarianța factorilor
  - c: pătratul raportului Sharpe al primelor, pe lună (0,03 pentru FF3)
- Un c mic (prime mici față de volatilitatea factorilor, ca la FF3 lunar): corecția este mică
- Un c mare (de exemplu, un factor macro cu volatilitate mică față de prima estimată): statisticile t Fama–MacBeth sînt supraestimate
- Erorile Fama–MacBeth, Newey–West și Shanken presupun că modelul este corect; corecția KRS pentru un model greșit: Aprofundări

Primele de risc estimate (1/2)



Legenda: Fama-MacBeth risk premia: primele de risc Fama–MacBeth, cu intervale de 95%; Time-series mean of the factor: media factorului în timp

MFM_ch3_fama_macbeth ▶ Interactiv

Primele de risc estimate (2/2)

- Barele: intervale de încredere de 95% cu erori Shanken
  - în fracții anuale ($0.1 = 10\%$)
- Crucile: media în timp a fiecărui factor tranzacționat
  - egală cu λ dacă modelul este corect
  - comparația cu crucea este vizuală: și media este estimată, din aceleași luni
- Prime estimate (25 de portofolii, 1963–2026), cu SE în paranteză
  - CAPM: termenul liber 14,0% (4,4%), piața −4,4% (4,8%)
  - FF3: SMB 1,4% (1,4%), HML 4,0% (1,3%)
  - FF5: RMW 5,3% (2,1%), CMA −1,4% (2,2%)
- HML: prima estimată 4,0%, media factorului 3,6%
  - valori apropiate
- Piața (FF3): prima estimată −6,5%, media factorului 7,2%
  - valori îndepărtate
- Valorile beta de piață diferă puțin între portofolii, deci panta pe ele este imprecisă
  - 0,86–1,42 în CAPM
  - 0,89–1,14 în FF3
- Prima pieței este deci **slab identificată**
  - termenul liber preia nivelul mediu al randamentelor (Lewellen, Nagel & Shanken (2010))
- Rezultatul arată o asociere între expuneri și randamentele medii, nu sursa economică a riscului

MFM_ch3_fama_macbeth

Trei verificări ale unui model factorial *(aprofundare)*

- Trei verificări (Lewellen, Nagel & Shanken (2010))
  - Lărgiți mulțimea activelor de test (Aprofundări)
  - Raportați și R^2 **GLS** (*Generalised Least Squares*, metoda celor mai mici pătrate generalizate): ca GRS, ține cont de covariațe, dar folosește inversa covarianței randamentelor activelor, nu a reziduurilor
  - Impuneți ca prima unui factor tranzacționat să fie media lui și eliminați termenul liber: FF3 pe cele 55 de active (25 de portofolii și 30 de industrii) dă $R^2 = -0{,}27$ (nu este nici R^2 GLS, nici R^2 OLS liber)
- R^2 transversal neponderat: $R^2 = 1 - \sum_i (\bar{R}_i^e - \hat{\mu}_i)^2 / \sum_i (\bar{R}_i^e - \bar{R}^e)^2$
  - $\hat{\mu}_i$: media prezisă
  - $\bar{R}^e$: media celor N active
- $R^2 < 0$ cînd modelul greșește mai mult decît simpla medie $\bar{R}^e$
  - OLS cu termen liber nu poate face mai rău decît ea
- Acțiuni individuale în locul portofoliilor (Ang, Liu & Schwarz (2020))
  - valori beta mai dispersate
  - dar estimate mai zgomotos

Verificare rapidă: prime, Shanken și un R^2 negativ

- Un factor netranzacționat (de exemplu, creșterea producției industriale)
  - volatilitatea $\sigma_f = 1\%$ pe lună, $T = 400$ de luni
  - etapa 2 dă $\hat{\lambda} = 0{,}5\%$ pe lună, cu $s(\hat{\lambda}_t) = 3\%$
  1. Calculați se_{FM} și statistica t Fama–MacBeth
  2. Calculați c, varianța Shanken și statistica t Shanken. Trece factorul pragul $t > 3$ (ipoteza exercițiului; motivarea: Secțiunea 7)?
  3. De ce ar fi c foarte mic pentru un factor tranzacționat cu $\sigma_f = 4{,}5\%$ și același $\hat{\lambda}$?
- Trei portofolii au randamentele medii în exces 2%, 4%, 6% pe an
  - un model cu prime impuse și fără termen liber estimează 6%, 4%, 2%
  4. Calculați SST (*Total Sum of Squares*, suma totală a pătratelor abaterilor de la medie), SSE (*Sum of Squared Errors*, suma pătratelor erorilor) și R^2
  5. De ce nu poate fi negativ R^2-ul unei regresii OLS cu termen liber?
- Răspunsuri: ▸ Anexă

Grădina zoologică a factorilor

- Întrebarea: cîți dintre factorii publicați sînt reali?
  - fiecare a fost testat ca și cum ar fi fost singurul
- Cochrane (2011) a numit lista tot mai lungă de factori un *factor zoo*
  - Harvey, Liu & Zhu (2016) catalogează 316 factori publicați
- Un experiment controlat: testăm 300 de ipoteze nule adevărate la 5%
  - ne așteptăm, în medie, la 15 respingeri false
- Inventarul publicațiilor (316) nu este familia testată, a cărei mărime este necunoscută
- **Căutarea oportunistă în date** (*data mining*, *p-hacking*): se încearcă multe variante pînă cînd una pare semnificativă
- Corecția Bonferroni împarte 5% la numărul de teste
  - pentru 316 teste fixate ipotetic în prealabil: $p \leq 0{,}00016$, adică $|t| > 3{,}78$
- Harvey, Liu & Zhu (2016) propun $|t| > 3{,}0$ pentru un factor nou
  - pragul care rezultă din corecțiile pentru testarea multiplă (slide-ul următor) aplicate tuturor factorilor testați
- Pragul crește cu fiecare factor publicat: familia testată se mărește
  - un prag prea mic lasă să treacă factori falși, care apoi nu se replică (studiul de caz despre replicare)

Controlul descoperirilor false (1/2)

- **FDR**: ponderea așteptată a respingerilor false în totalul respingerilor
  - FDR: False Discovery Rate, rata descoperirilor false
- **FWER**: probabilitatea de a avea cel puțin o respingere falsă (Capitolul 2)
  - FWER: Family-Wise Error Rate, rata erorii pe familia de teste
- Notațiile tabelului (slide-ul următor)
  - M: numărul de teste
  - α: nivelul de semnificație al testelor (aici α nu este alfa lui Jensen)
  - $p_{(1)} \leq \cdots \leq p_{(M)}$: valorile p (*p-values*) ordonate
  - $H_{(j)}$: ipoteza cu a j-a cea mai mică valoare p
  - **BH**: procedura Benjamini–Hochberg
  - **BY**: procedura Benjamini–Yekutieli

Controlul descoperirilor false (2/2)

Procedura	Controlează	Respinge $H_{(j)}$ dacă	Validă pentru
Bonferroni	FWER	$p_{(j)} \leq \alpha/M$	orice dependență
Holm (Holm (1979))	FWER	$p_{(k)} \leq \alpha/(M-k+1)$, $\forall k \leq j$	orice dependență
BH (Benjamini & Hochberg (1995))	FDR	$j \leq$ cel mai mare k cu $p_{(k)} \leq k\alpha/M$	dependență pozitivă
BY (Benjamini & Yekutieli (2001))	FDR	ca BH, pragurile împărțite la $\sum_{k=1}^{M} 1/k$	orice dependență

- Dependență pozitivă: statisticile de test tind să crească împreună
  - independența este un caz particular
  - condiția tehnică se numește PRDS, Positive Regression Dependence on a Subset (Benjamini & Yekutieli (2001))
- FWER: potrivit cînd nicio descoperire falsă nu este acceptabilă
- FDR: potrivit cînd căutăm printre mulți factori candidați
  - și acceptăm o mică proporție de descoperiri false

Verificare rapidă: numărarea descoperirilor

- Cinci factori candidați
  - valori p ordonate: 0,004; 0,012; 0,020; 0,030; 0,200
  - nivel de semnificație $\alpha = 5\%$
  1. Cîte descoperiri dau regula naivă, Bonferroni, Holm și BH? Scrieți fiecare prag cu care comparați
  2. La BY, pragurile BH se împart la $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}$: cîte descoperiri rămîn?
  3. Dacă toate cele cinci ipoteze nule ar fi adevărate și testele independente, care este probabilitatea ca cel puțin o valoare p să fie sub 0,05?
  4. Ce procedură ați folosi într-o căutare printre mulți factori candidați?
  5. Cum limitați la 5% probabilitatea de a respinge cel puțin o ipoteză nulă adevărată?
- Răspunsuri: ▸ Anexă

Se replică factorii publicați?

- **Anomalie**: un randament mediu pe care CAPM nu îl explică
- **Două critici** la adresa cercetării despre factori (Jensen, Kelly & Pedersen (2023))
  - Rezultatele nu se pot reproduce: Hou, Xue & Zhang (2020) reușesc să replice doar 35% dintre anomaliile studiate
  - Rezultatele sînt produsul testării multor factori fără corecție: Harvey, Liu & Zhu (2016) conchid că majoritatea descoperirilor raportate sînt probabil false
- **Întrebările**
  - au factorii publicați valori alfa CAPM semnificative în datele originale?
  - sînt aceste valori alfa semnificative și în afara eșantionului inițial (*out-of-sample*) și în alte țări?
  - rămîn ele semnificative după ce se ține cont de testarea multiplă?

De ce contează replicarea pentru acest capitol

- Dacă majoritatea factorilor ar fi descoperiri false, s-ar sprijini pe zgomot
  - modelele multifactoriale (Secțiunea 5)
  - ETF-urile factoriale (Secțiunea 8)
- Feng, Giglio & Xiu (2020) arată că majoritatea factorilor noi nu aduc nimic în plus față de cei existenți (▸ Anexă)
- Jensen, Kelly & Pedersen (2023) văd în schimb grădina zoologică a factorilor ca multe măsurători corelate ale acelorași cîteva teme economice
- Metoda lor, Bayes empiric: alfa fiecărui factor este micșorat spre zero și tras spre media factorilor din aceeași temă
- Mărimea testată în studiul de caz este alfa lui Jensen (Secțiunea 4)
- Reperul clasic (non-bayesian) este corecția Benjamini–Yekutieli (Secțiunea 7)

Majoritatea factorilor se replică, dar alfa scade după eșantionul inițial (1/3)

Rate de replicare (Figura 1)

Hou, Xue & Zhang (2020), randamente brute	35,0%
Jensen, Kelly & Pedersen (2023), randamente brute	55,6%
Randamente brute, 119 semnificativi inițial	61,3%
Alfa CAPM, OLS	82,4%
Alfa CAPM, Benjamini–Yekutieli	75,6%
Alfa CAPM, Bayes empiric	82,4%
Global, Bayes empiric	82,4%

- De la 35% (Hou, Xue & Zhang (2020)) la 82,4%
  - Cu un eșantion mai lung, 15 factori în plus și o construcție diferită, rata urcă la 55,6%
  - Dacă se păstrează doar factorii semnificativi în studiile inițiale, rata devine 61,3%
  - Dacă se testează alfa CAPM în locul randamentului brut, rata devine 82,4%
  - acesta este cel mai mare pas

Majoritatea factorilor se replică, dar alfa scade după eșantionul inițial (2/3)

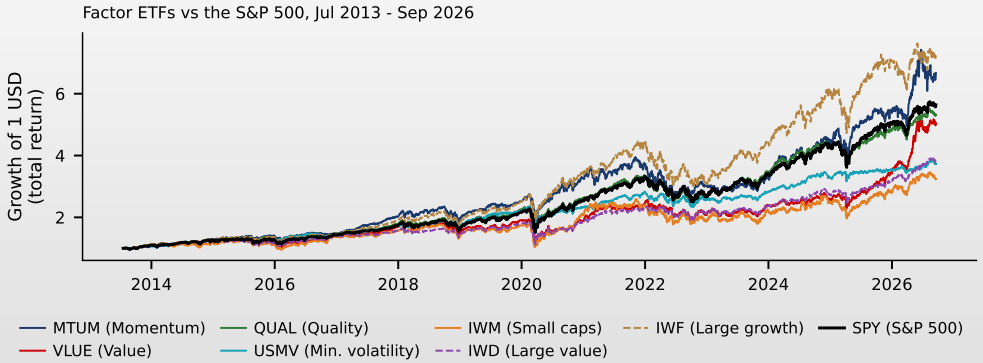
- Bayes empiric dă tot 82,4% (rezultatele complete: Anexă)
  - micșorarea spre zero pierde unele descoperiri
  - informația împrumutată de la temă adaugă altele
  - efectele se compensează
- Proporția estimată de descoperiri false (FDR bayesian) printre factorii cu $z \geq 1{,}96$ este doar 0,1%
  - deși pragul rămîne 1,96
- Rata de replicare se referă la cei 119 factori semnificativi în studiile inițiale, nu la toți factorii propuși

Majoritatea factorilor se replică, dar alfa scade după eșantionul inițial (3/3)

- **Scăderea** după perioada studiului inițial
  - În medie, alfa unui factor scade de la 0,49% pe lună în eșantion (al studiului inițial) la 0,26% pe lună în afara eșantionului, adică cu 47%
  - Panta regresiei alfa în afara eșantionului pe alfa în eșantion este 0,26 (Figura 8), un coeficient distinct de media de 0,26% pe lună
  - un alfa inițial mai mare cu 1 punct procentual pe lună rămîne ulterior mai mare doar cu 0,26 puncte procentuale
  - Contracția statistică singură ar explica o pantă de circa 0,9, deci scăderea reală este mai mare decît zgomotul
- Cei mai puternici factori scad cel mai mult
  - semn de arbitraj sau de căutare oportunistă în date (Aprofundări)
- Două rezultate distincte (Jensen, Kelly & Pedersen (2023))
  - reproducerea: pentru 98 dintre cei 119 factori inițial semnificativi, alfa CAPM (OLS) reproduce rezultatul (82,4%)
  - mărimea efectului: alfa mediu scade de la 0,49% la 0,26% pe lună (-47%); cauza scăderii rămîne deschisă

ETF-urile factoriale față de piață (1/2)

Întrebarea: cîștigă un investitor care cumpără factorii prin ETF-uri mai mult decît cu piața?



MFM_ch3_factor_etfs ▶ Interactiv

ETF-urile factoriale față de piață (2/2)

- Graficul: valoarea a 1 dolar american (USD) investit la 18 iulie 2013
  - cu dividendele reinvestite (randament total)
  - SPY este linia neagră
- **Smart beta**: ETF-uri doar long care cumpără acțiuni după reguli fixe
  - deci păstrează expunerea la piață și nu reproduc factorul academic long–short
- IWD (*iShares Russell 1000 Value ETF*) și IWF (*iShares Russell 1000 Growth ETF*): acțiunile de valoare, respectiv de creștere din Russell 1000 (cele mai mari 1000 de acțiuni americane); IWM (*iShares Russell 2000 ETF*): Russell 2000 (următoarele 2000)
- Ce explică o rată anuală compusă de creștere (CAGR) mai mare? Exemplul MTUM (*iShares MSCI USA Momentum Factor ETF*)
  - regresie zilnică pe FF5 + MOM, pînă în iulie 2026: expunere MOM 0,36
  - alfa anualizat $-0{,}2\%$, $t = -0{,}11$: date compatibile cu alfa zero
- **CAGR** (*Compound Annual Growth Rate*): $(V_T/V_0)^{1/n} - 1$ pe n ani
- Peste SPY (14,0%)
  - firmele mari de creștere (IWF): 16,3%
  - momentum (MTUM): 15,5%
- Sub SPY
  - QUAL (*iShares MSCI USA Quality Factor ETF*, ETF pe factorul calitate) 13,5%, VLUE (*iShares MSCI USA Value Factor ETF*, ETF pe factorul valoare) 13,1%, IWD 10,7%
  - USMV (*iShares MSCI USA Min Vol Factor ETF*, ETF de volatilitate minimă) 10,5%, IWM 9,3%
- Treisprezece ani reprezintă o singură traiectorie istorică
  - prea scurtă pentru a respinge primele factorilor estimate pe șase decenii

MFM_ch3_factor_etfs

▸ Aprofundări: alfa ETF-uri, timing

Concluzii: pașii unei analize factoriale

1. **Estimați expunerile** cu erori HAC
   - raportați eșantionul și aproximarea pieței
   - contractați spre medie valorile beta extreme (Blume (1971), Vasicek (1973))
2. **Testați toate valorile alfa deodată**
   - GRS măsoară, normalizat, cît crește pătratul raportului Sharpe maxim cînd adăugăm activele de test
   - CAPM eșuează pe cele 25 de portofolii mărime $\times$ B/M (SML plată, GRS respinge)
3. **Estimați primele** în două etape
   - cu erori care țin cont că beta este estimat și că modelul poate fi greșit
   - un R^2 mare pe cele 25 de portofolii nu dovedește mare lucru
4. **Protejați-vă de grădina zoologică a factorilor**: folosiți praguri corectate pentru testarea multiplă
5. **Verificați replicarea**: 82,4% dintre cei 119 factori inițial semnificativi se replică
   - dar alfa mediu scade cu 47% după eșantionul inițial
6. **Descompuneți ETF-urile factoriale**
   - ele vînd în mare parte expuneri la factori cunoscuți, nu alfa (MTUM: Secțiunea 8; restul: Aprofundări)

Verifică-ți înțelegerea

Întrebări de verificare

- De ce diferă eroarea standard HAC a unui beta de cea clasică, deși beta nu se schimbă? (Aprofundări)
- De ce poate GRS să nu respingă CAPM pe ETF-urile sectoriale, dar să-l respingă pe cele 25 de portofolii?
  - în aceleași luni (Secțiunea 4)
- De ce pare să aibă o primă un factor inutil în etapa 2? (Aprofundări)
- De ce poate rata de replicare Bayes empiric să depășească rata Benjamini–Yekutieli? (Secțiunea 7)

Legături cu alte capitole

- Capitolul 2, deja parcurs: un alfa nenul este fie o anomalie, fie semnul unui factor de risc omis
  - problema ipotezei comune
- Capitolul 4: modele factoriale de covariață și estimatori de contracție pentru Markowitz
- Capitolele 5–6: volatilitate, corelații și beta variabile în timp
- Capitolul 13: supraajustarea (*overfitting*) testelor istorice
  - și învățarea automată (*machine learning*) pentru secțiunea transversală a randamentelor

Decizia practică

- **Ce aleg**
  - beta cu erori HAC și contracție (XLK față de SPY, 1999–2012: $\hat{\beta} = 1{,}415$, Blume 1,278, Vasicek 1,344; realizat 2013–2026: 1,172)
  - testul GRS pe active alese pentru contraste economice stabilite înainte: 25 de portofolii mărime $\times$ B/M, nu doar sectoarele (GRS 2,91 față de 0,53, pe aceleași 331 de luni)
  - prime Fama–MacBeth cu erori KRS (cu laguri Newey–West); erorile FM și Shanken doar drept comparație; un factor nou doar peste pragul corectat pentru testarea multiplă
  - robustețea la specificare greșită și la autocorelație nu rezolvă identificarea slabă
- **În ce condiții**
  - erorile HAC presupun moment de ordin patru finit, pus sub semnul întrebării în Capitolul 1
  - GRS presupune erori cu distribuția Normală multivariată, independente de factori; respingerea privește și indicele folosit ca piață (Roll)
  - prima pieței rămîne slab identificată cînd valorile beta diferă puțin (0,89–1,14 în FF3)
- **Ce rezultate ar schimba decizia**
  - contracția: o eroare de prognoză mai mare decît la OLS pe mai multe active și perioade (acum, pe 9 sectoare, RMSE: OLS 0,135, Blume 0,109, Vasicek 0,114)
  - un model: un R^2 transversal negativ cu primele impuse și fără termen liber (FF3 pe 55 de active: $-0{,}27$)
  - un factor publicat: un alfa în afara eșantionului mic din punct de vedere economic, citit împreună cu intervalul lui (în medie, alfa scade cu 47%)

Aprofundări: subiecte suplimentare

- Datele capitolului și randamentele folosite
- Factorii ca factor stocastic de actualizare
- Factori latenți: PCA (*Principal Component Analysis*, analiza componentelor principale) și IPCA (*Instrumented Principal Component Analysis*, analiza instrumentată a componentelor principale)
- Studiu de caz: momentum-ul factorilor
- AI (inteligența artificială) pentru descoperire științifică
- Apoi: exemple rezolvate și detalii pentru fiecare secțiune, în ordinea capitolului

Datele: ce serii folosim

Serie	Ce conține	Frecvență	Perioadă
Factorii Fama–French	randamentele a cinci factori: piața, mărimea, valoarea, profitabilitatea și investițiile	lunară	1963–2026
Factorul momentum	acțiunile cu cele mai mari randamente în lunile $t-12$ pînă la $t-2$ minus cele cu cele mai mici	lunară	1963–2026
25 de portofolii	acțiuni americane sortate independent în 5 grupe după mărime și 5 după B/M (valoarea contabilă / valoarea de piață); cele 25 de intersecții formează portofoliile	lunară	1963–2026
ETF-uri sectoriale	cîte un ETF pentru fiecare sector din S&P 500 (indicele a 500 de mari companii americane): 9 sectoare din 1999, 11 din 2018	zilnică	1999–2026
ETF-uri factoriale	momentum, valoare, calitate, volatilitate minimă (din 2013); mărime, valoare și creștere, indici Russell (din 2000)	zilnică	2000–2026
Bursa de Valori București (BVB)	11 acțiuni mari (blue chips) și indicele bursei, BET	zilnică	2015–2026

- Simbolurile ETF-urilor sectoriale
  - XLB materiale, XLE energie, XLF financiar, XLI (*Industrial Select Sector SPDR Fund*, ETF pe sectorul industrial) industrie, XLK tehnologie
  - XLP (*Consumer Staples Select Sector SPDR Fund*, ETF pe sectorul bunurilor de consum de bază) bunuri de larg consum, XLU utilități, XLV (*Health Care Select Sector SPDR Fund*, ETF pe sectorul sănătății) sănătate, XLY (*Consumer Discretionary Select Sector SPDR Fund*, ETF pe sectorul bunurilor de consum ciclice) consum discreționar
  - din 2018 se adaugă XLRE imobiliare și XLC servicii de comunicații

Ce randamente folosim în acest capitol

- Notații
  - $R_{i,t} = P_{i,t}/P_{i,t-1} - 1$: randament simplu
  - $r_{i,t} = \ln(1 + R_{i,t})$: randament logaritmic
  - $P_{i,t}$: prețul ajustat pentru dividende și split-uri
- Regresiile pe factorii French folosesc randamente simple lunare
  - Factorii sînt randamente de portofolii, iar randamentul unui portofoliu se calculează din randamente simple
  - Media randamentelor logaritmice ar fi mai mică cu aproximativ $\sigma^2/2$: la o volatilitate anuală de 20%, circa 2 puncte procentuale pe an
- Randamente logaritmice zilnice folosesc
  - beta pe ferestre mobile
  - beta la BVB
  - analiza componentelor principale (Aprofundări)
- La BVB, rata fără risc zilnică este tratată ca o constantă
  - deci beta este același cu randamente brute sau în exces

Ce conțin sectoarele și ce pondere primesc în portofoliul tangent

ETF	Sector	Ce conține (exemple de companii)	Pondere
XLB	materiale	chimicale, materiale de construcții, ambalaje, metale și minerit (Linde, Sherwin-Williams, Freeport-McMoRan, Newmont)	−0,21
XLE	energie	petrol și gaze, servicii petroliere (Exxon Mobil, Chevron, ConocoPhillips)	+0,19
XLF	financiar	bănci, asigurări, administratori de active, procesatori de plăți (Berkshire Hathaway, JPMorgan Chase, Visa)	−0,37
XLI	industrie	aerospațial și apărare, utilaje, transport (GE Aerospace, Caterpillar, Union Pacific)	+0,18
XLK	tehnologie	software, hardware, semiconductori (Apple, Microsoft, Nvidia)	+0,13
XLP	bunuri de larg consum	alimente, băuturi, produse de uz casnic (Procter & Gamble, Coca-Cola, Walmart)	+0,16
XLU	utilități	electricitate, gaze naturale, apă (NextEra Energy, Southern, Duke Energy)	+0,27
XLV	sănătate	farmaceutice, biotehnologie, echipamente medicale (Eli Lilly, UnitedHealth, Johnson & Johnson)	+0,37
XLY	consum discreționar	comerț, auto, hoteluri, restaurante (Amazon, Tesla, Home Depot)	+0,29

- Sectoarele urmează clasificarea GICS (Global Industry Classification Standard) a companiilor din S&P 500
  - compoziția s-a schimbat: imobiliarele au ieșit din XLF în 2016; în 2018, Alphabet și Facebook (azi Meta), din XLK, și Netflix și Comcast, din XLY, au trecut în servicii de comunicații
- Semnul fiecărei ponderi depinde de întreaga matrice de covarianță, nu de o singură pereche de sectoare
  - exemplu: XLB și XLI au corelația 0,85, iar XLI are medie mai mare (8,78% față de 8,19%) și volatilitate mai mică (18,71% față de 20,96%)
- XLF are cel mai mic raport Sharpe dintre cele 9 sectoare (0,29) și corelația 0,80 cu XLI

MFM_ch3_frontier

Factorii ca factor stocastic de actualizare *(aprofundare)*



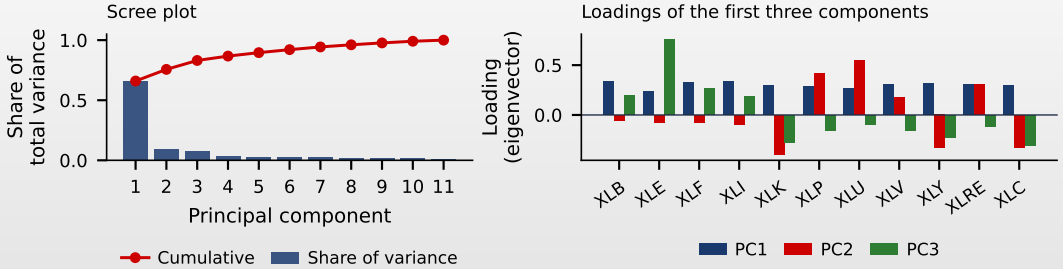
Lars Peter Hansen, Premiul Nobel pentru Economie 2013

Foto: Bengt Nyman (2013); CC BY 2.0; Wikimedia Commons

- Întrebarea: este un factor necesar pentru evaluarea activelor, odată ce avem ceilalți factori?
  - sinteză: Cochrane (2011)
- **SDF** m_t: evaluează toate randamentele în exces, $\mathbb{E}[m_t R^e_{i,t}] = 0$
  - SDF: Stochastic Discount Factor, factorul stocastic de actualizare
- m_t este mare în perioadele proaste
  - un activ care cîștigă tocmai atunci este o asigurare
  - deci are un randament mediu mic
- Un model factorial liniar este un SDF, $m_t = 1 - \mathrm{b}^\top(\mathrm{f}_t - \mathbb{E}\mathrm{f})$
  - b: coeficienții SDF
- Prin înlocuire: $\mathbb{E}[R^e_i] = \mathrm{Cov}(R^e_i, \mathrm{f}_t)^\top \mathrm{b} = \beta_i^\top \boldsymbol{\lambda}$
  - $\boldsymbol{\lambda} = \boldsymbol{\Sigma}_f \mathrm{b}$ (Secțiunea 6)
- Două teste diferite
  - $b_k = 0$: factorul k nu ajută la evaluare, ținînd cont de ceilalți factori
  - $\lambda_k = 0$: expunerea la factorul k nu aduce nicio primă
  - cu factori corelați, $\lambda_k \neq 0$ și $b_k = 0$ pot coexista
- b se estimează prin GMM (Hansen (1982); Aprofundări)
  - media empirică a lui $m_t(\mathrm{b})\,\mathrm{R}^e_t$ trebuie să fie zero pentru fiecare activ

PCA pe ETF-urile sectoriale din SUA (1/2)

Întrebarea: ce factori apar din randamente dacă nu îi alegem dinainte?



Stînga (Scree plot): ponderea fiecărei componente în varianța totală (bare, Share of variance) și ponderea cumulată (linie, Cumulative); dreapta: încărcările (eigenvector) ale primelor trei componente pe fiecare ETF

MFM_ch3_pca

▶ Interactiv

PCA pe ETF-urile sectoriale din SUA (2/2)

- **PCA** (Principal Component Analysis, analiza componentelor principale)
  - combinațiile liniare ale randamentelor care explică, pe rînd, cea mai mare parte a variației rămase
  - formula: slide-ul următor
- 11 ETF-uri sectoriale, randamente logaritmice zilnice
  - 20 iunie 2018 – 18 septembrie 2026 ($T = 2\,073$ zile comune)
- PC1 (*first Principal Component*, prima componentă principală) explică 66,0% din varianță
  - încărcări toate pozitive (0,24–0,34)
  - o medie aproape egală a sectoarelor
- PC1 este factorul pieței: corelația cu SPY este 0,954
- PC2 (9,7%): **sectoare defensive față de sectoare de creștere**
  - utilități $+0{,}55$, bunuri de larg consum $+0{,}42$, imobiliare $+0{,}31$
  - față de tehnologie $-0{,}40$, consum discreționar $-0{,}33$
- PC3 (7,4%) este dominată de energie
- Primele trei componente explică 83,1%
- Rezultatul descrie eșantionul; nu identifică o cauză

MFM_ch3_pca

Componentele principale

- **Factori latenți**: factori neobservați, estimați doar din randamente
  - fără a spune dinainte ce reprezintă
- PCA a matricei de corelație C a randamentelor standardizate

$$\mathsf{C} = \mathbf{\Gamma}\mathsf{D}\mathbf{\Gamma}^\top, \qquad \mathsf{y}_t = \mathbf{\Gamma}^\top \mathsf{z}_t$$

  - z_t: randamentele standardizate ale zilei t
  - y_t: componentele
  - $\mathbf{\Gamma}$: vectorii proprii pe coloane, adică **încărcările**
  - D: matricea diagonală a valorilor proprii ℓ_k
- Valorile proprii, ordonate: $\ell_1 \geq \cdots \geq \ell_N$
  - notate ℓ, ca să nu se confunde cu primele de risc λ
- Componenta k explică ℓ_k/N din varianța totală
- Exemplu, două active cu corelația ρ: $\ell = 1 \pm \rho$
  - vectorul $(1,1)/\sqrt{2}$ este „piața"
  - vectorul $(1,-1)/\sqrt{2}$ este diferența dintre ele
- În criză ρ crește și prima componentă domină

Limitele PCA în evaluarea activelor

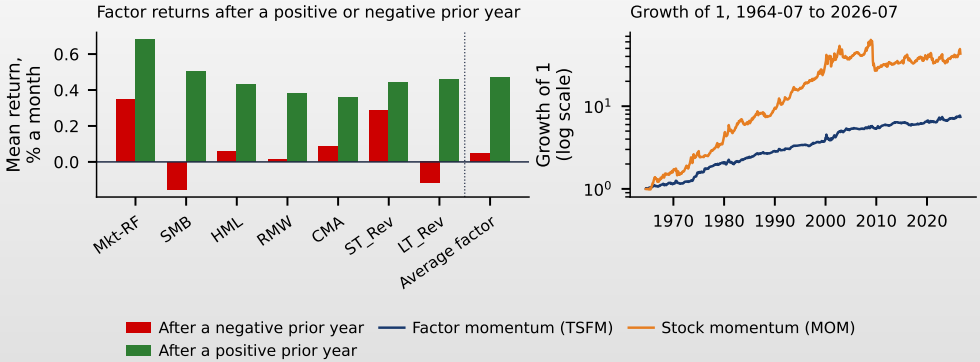
- Încărcările sînt **constante**
  - PCA nu poate urmări o acțiune a cărei mărime sau valoare se schimbă
- Componentele explică **varianța**, nu neapărat **randamentele medii**
- Acțiunile individuale intră și ies din eșantion (listări, delistări)
  - deci lipsesc multe observații
  - iar PCA clasic cere un panel complet
- Semnul fiecărei componente este arbitrar
  - schimbarea semnului unui vector propriu nu schimbă nimic
- Factorii latenți sînt determinați doar pînă la o rotație
  - orice combinație rotită a lor explică datele la fel de bine
  - deci interpretarea lor este o alegere
- Două extensii ale PCA țin cont și de randamentele medii (Anexă)
  - PCA cu primă de risc
  - estimatorul în trei etape

Momentum-ul factorilor: cel al acțiunilor doar sincronizează alți factori?

- **Întrebarea**: sînt randamentele factorilor autocorelate și este factorul momentum MOM (Secțiunea 5) doar o sincronizare a celorlalți factori?
- **De ce contează**
  - modelul Carhart (FF3 + MOM) și FF5 + MOM adaugă MOM ca sursă separată de risc
  - dacă MOM doar sincronizează alți factori, modelul are nevoie de alocare dinamică între factori (*factor timing*: mutarea capitalului după semnale), nu de un factor momentum separat
- **Afirmațiile principale** din Ehsani & Linnainmaa (2022)
  - factorul mediu cîștigă 0,06% pe lună după un an cu pierderi
  - și 0,51% pe lună după un an pozitiv
  - momentum-ul factorilor se concentrează în factorii care explică mai mult din secțiunea transversală
  - momentum-ul factorilor din componentele principale cu valori proprii mari înglobează majoritatea formelor de momentum al acțiunilor
  - concluzia autorilor: momentum nu este un factor de risc distinct, ci sincronizează ceilalți factori

Momentum-ul factorilor French, 1964–2026

Întrebarea: un factor care a crescut în ultimul an continuă să crească în luna următoare?



MFM_ch3_factor_momentum ▶ Interactiv

Momentum-ul factorilor: replicarea din curs

- **Datele**: șapte factori lunari din Kenneth French Data Library, iulie 1963 – iulie 2026
  - Mkt-RF, SMB, HML, RMW, CMA (FF5)
  - ST_Rev (short-term reversal): long perdanții lunii $t-1$, short cîștigătorii ei
  - LT_Rev (long-term reversal): long perdanții din lunile de la $t-60$ la $t-13$, short cîștigătorii lor
  - MOM rămîne în afara setului: este seria de explicat
- **Semnalul** $s_{i,t}$: randamentul compus al factorului i în lunile de la $t-12$ la $t-1$, ca la time-series momentum (Moskowitz, Ooi & Pedersen (2012))
- **TSFM** (time-series factor momentum): $\mathrm{TSFM}_t = \frac{1}{7}\sum_{i=1}^{7} \mathrm{sign}(s_{i,t})\, r_{i,t}$
  - long factorii cu anul anterior pozitiv, short ceilalți; prima lună iulie 1964 ($T = 745$)
- **Testele**
  - diferența mediilor după ani pozitivi și negativi, pe toate perechile factor–lună, cu erori grupate pe luni
  - regresii de spanning ale lui MOM pe FF5 și TSFM, cu erori Newey–West
- **Diferența față de lucrare**: lucrarea folosește mult mai mulți factori, inclusiv factori din componente principale; șapte factori dau un test de spanning mai slab

Momentum-ul factorilor: rezultatul și limitele

- Factorul mediu, randamentul mediu din luna următoare
  - după un an anterior negativ: 0,05% pe lună
  - după un an anterior pozitiv: 0,47% pe lună
  - diferența: 0,42 puncte procentuale, $t = 3{,}62$
  - lucrarea raportează 0,06% și 0,51%
- Media este mai mare după un an pozitiv la toți cei șapte factori
- TSFM: media 0,28% pe lună, $t = 5{,}48$
  - raportul Sharpe anual 0,67; MOM: 0,49
  - drawdown maxim -15%; MOM: -58% (septembrie 2009)
- Alfa FF5 a lui MOM: 0,74% pe lună ($t = 4{,}18$)
  - cu TSFM adăugat: 0,32% ($t = 2{,}27$), cu 57% mai mic
  - panta pe TSFM: 1,76 ($t = 13{,}8$)
- Alfa lui TSFM pe FF5 și MOM: 0,08% ($t = 1{,}81$)
- Rezultatul descrie eșantionul
  - randamentele factorilor sînt autocorelate pe orizontul de douăsprezece luni
  - șapte factori explică aproximativ jumătate din alfa lui MOM, nu tot
- Limite
  - un singur eșantion, fără costuri de tranzacționare
  - cei doi factori de reversal sînt construiți tot din randamente trecute

MFM_ch3_factor_momentum

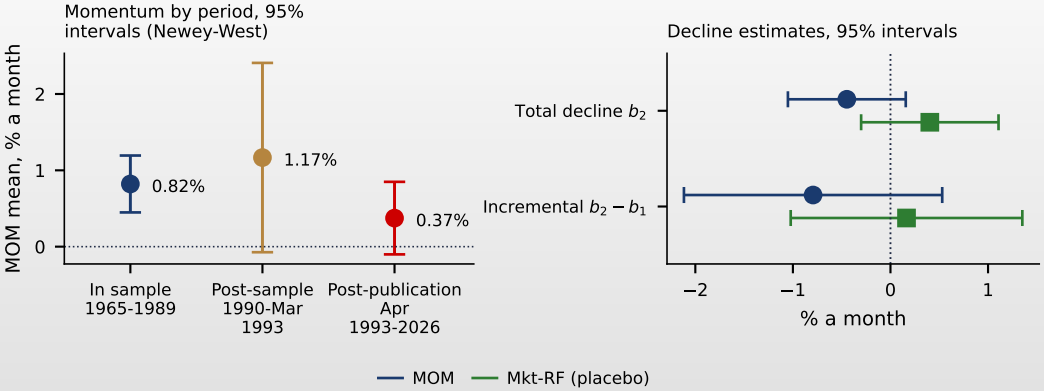
Întrebări deschise și bucla descoperirii

- **Întrebări deschise**
  - de ce scad primele factorilor după publicare: arbitraj, căutare oportunistă în date sau o schimbare a riscului?
  - diferă scăderea între factori și piețe?
- Stadiul cunoașterii
  - McLean & Pontiff (2016): pe 97 de predictori, randamentele sînt cu 26% mai mici în afara eșantionului și cu 58% mai mici după publicare
  - autorii atribuie diferența investitorilor care tranzacționează pe baza articolelor publicate
  - Chen & Zimmermann (2022): date și cod deschise pentru predictorii publicați
- **Bucla descoperirii**
  - Asistentul AI propune: un tabel al literaturii cu DOI (Digital Object Identifier, identificator digital de obiect), ipoteze concurente și ce prezice fiecare, codul regresiei, un plan de robustețe (prompturi: Anexă)
  - Cercetătorul fixează înainte de rulare: datele evenimentelor, lista factorilor, corecția pentru testarea multiplă și testul placebo (un factor la care nu ar trebui să apară efectul)
  - Înainte de interpretare: fiecare DOI se deschide în Crossref, statisticile t pe subperioade din Aprofundări se reproduc, iar versiunea datelor French este notată

Ministudiu de caz: arbitrajul și declinul momentum-ului (1/3)

- Specificația din McLean & Pontiff (2016): $\mathrm{MOM}_t = a + b_1 \mathrm{PostSample}_t + b_2 \mathrm{PostPub}_t + e_t$
  - lunar, ianuarie 1965 – august 2026 ($T = 740$)
  - erori Newey–West
- Variabile dummy
  - $\mathrm{PostSample}_t = 1$ în ianuarie 1990 – martie 1993 (după eșantion)
  - $\mathrm{PostPub}_t = 1$ din aprilie 1993 (după publicare)
- a: media în eșantion (1965–1989)
- b_2: scăderea totală față de media în eșantion
- $b_2 - b_1$: efectul incremental al publicării

Ministudiu de caz: arbitrajul și declinul momentum-ului (2/3)



MFM_ch3_ai_discovery ▶ Interactiv

Ministudiu de caz: arbitrajul și declinul momentum-ului (3/3)

- Media momentum în eșantion: $a = 0{,}82\%$ pe lună ($t = 4{,}32$)
- Scăderea după publicare: $b_2 = -0{,}45\%$ pe lună ($t = -1{,}45$)
  - interval $[-1{,}05\%; 0{,}16\%]$
  - adică o medie cu 54% mai mică
- Efectul incremental al publicării: $b_2 - b_1 = -0{,}79\%$ ($t = -1{,}17$)
- Placebo pe Mkt-RF: $b_2 = +0{,}40\%$ ($t = 1{,}12$)
- Scăderea este apropiată de 58% din McLean & Pontiff (2016)
  - dar cu un singur factor nu putem respinge $H_0 : b_2 = 0$ la 5%
- Rezultatul descrie eșantionul
  - o comparație înainte/după publicare; arbitrajul este o interpretare posibilă, nu o cauză demonstrată

MFM_ch3_ai_discovery

Proiect: scad primele factorilor după publicare?

- **Întrebări**
  - scad primele factorilor Fama–French și momentum după publicare?
  - diferă scăderea între factori?
- **Metoda**
  - o regresie comună pentru SMB, HML, RMW, CMA și MOM, cu variabile dummy proprii fiecărui factor pentru perioadele după eșantion și după publicare
  - Mkt-RF ca placebo
  - Erori robuste la corelația dintre factori și la autocorelație
  - dacă scăderea este mai mare acolo unde arbitrajul este mai ușor se poate testa doar cu date la nivel de acțiune (mărime, lichiditate, costul vînzării în lipsă)
- **Replicați întîi**
  - statisticile t pe subperioade
  - estimațiile din ministudiul de caz

Datele: surse și notații

Surse

- Factorii și cele 25 de portofolii: Kenneth French Data Library, randamente lunare în procente
- Ultima lună disponibilă: iulie 2026 (pentru momentum: august 2026)
- ETF-uri și acțiuni: prețuri zilnice de închidere ajustate (dividende reinvestite)
- Ultima zi disponibilă: 18 septembrie 2026

Notații

- **BET**: indicele de referință al BVB, un indice de preț (fără dividende)
- **Randament în exces**: $R^e_{i,t} = R_{i,t} - R_{f,t}$
- $R_{f,t}$: rata titlurilor de stat americane la o lună (RF în datele French)
- **Mkt-RF**: randamentul în exces al pieței americane de acțiuni, cu ponderi după capitalizare

- De ce aceste date
  - randamentele medii foarte diferite ale celor 25 de portofolii fac testul exigent
  - ETF-urile sînt investibile
  - BVB aplică metodele pe piața locală

Costul capitalului: care incertitudine contează?

- Costul capitalului prin CAPM: $R_f + \hat{\beta}_i \, (\widehat{\mathbb{E}[R_m]} - R_f)$
- Eroarea standard (SE) a primei piețe: $15{,}4\%/\sqrt{63} = 1{,}9\%$ pe an
  - 15,4%: volatilitatea anuală a pieței
  - 63 de ani de date
- Înmulțită cu beta XLK, această incertitudine dă 2,5% pe an în costul capitalului
  - $\hat{\beta} = 1{,}28$ (Aprofundări)
- Incertitudinea lui beta adaugă doar $0{,}082 \times 7{,}2\% = 0{,}6\%$ pe an
  - SE a lui beta înmulțită cu prima
- Concluzie: incertitudinea primei piețe cîntărește mult mai mult decît cea a lui beta

Erorile HAC: ponderi, laguri și ipoteze

- Estimatorul Newey–West (Newey & West (1987)) adună autocovariațele momentelor $R^e_{m,t}\varepsilon_{i,t}$ pînă la lagul L
  - cu ponderi Bartlett $1 - \ell/(L+1)$, care fac matricea estimată pozitiv semidefinită
- Regula pentru numărul de laguri: $L = \lfloor 4(T/100)^{2/9} \rfloor$ (Newey & West (1994))
  - pentru $T = 331$ se obține $L = 5$, pentru $T = 757$ se obține $L = 6$
- Estimațiile punctuale rămîn cele OLS
- Erorile HAC nu sînt obligatoriu mai mari decît cele clasice în orice eșantion
- Erorile HAC tratează dependența, dar presupun moment de ordin patru finit, pus sub semnul întrebării în Capitolul 1

Exemplu rezolvat: beta ETF-ului de tehnologie (1/2)

		Eroare standard		
	Estimație	**Clasică**	**White**	**Newey–West**
α (anual)	0,33%	2,27%	2,32%	2,45%
β	1,277	0,041	0,054	0,082

- XLK pe Mkt-RF, lunar ianuarie 1999 – iulie 2026, $T = 331$, $R^2 = 0{,}742$
- Coloanele erorilor standard
  - **Clasică**: erori presupuse homoscedastice și neautocorelate
  - **White**: robustă la heteroscedasticitate (White (1980))
  - **Newey–West**: robustă și la autocorelație, $L = 5$ laguri
- Eroarea standard robustă a lui β este dublul celei clasice
- Heteroscedasticitatea explică o parte din creștere (coloana White)
  - autocorelația produsului $R^e_{m,t}\varepsilon_t$ explică restul
- Erorile White și Newey–West presupun moment de ordin patru finit, pus sub semnul întrebării în Capitolul 1

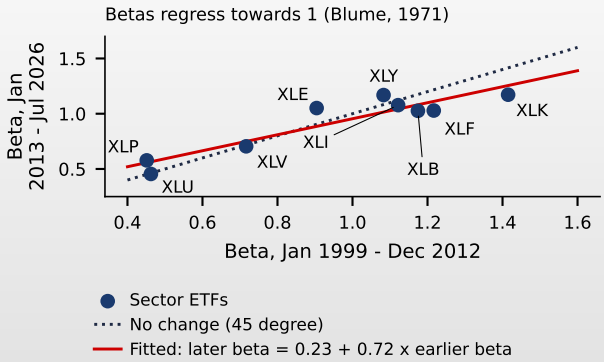
MFM_ch3_seminar

Exemplu rezolvat: beta ETF-ului de tehnologie (2/2)

- Statistica t a lui alfa: $t(\alpha) = 0{,}33/2{,}45 = 0{,}13$
- Intervalul bootstrap de 95% pentru alfa: $[-5{,}5\%; 4{,}9\%]$
  - SE bootstrap 2,61%, puțin peste 2,45%
  - bootstrap pe blocuri suprapuse de 12 luni din perechile $(R^e_{\text{XLK},t}, R^e_{m,t})$, care păstrează dependența în timp
  - 2 000 de reeșantionări (Seminarul 3, problema B1)

MFM_ch3_seminar

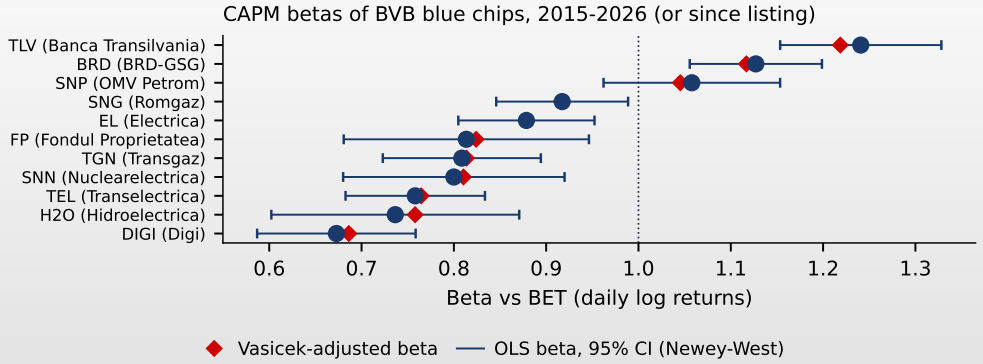
Revenirea valorilor beta spre unu



- Valori beta lunare ale celor 9 ETF-uri sectoriale față de SPY
  - 1999–2012 pe orizontală
  - 2013–2026 pe verticală
- Dreapta estimată: $\hat{\beta}_{\text{ulterior}} = 0{,}23 + 0{,}72\,\hat{\beta}_{\text{anterior}}$
  - foarte aproape de regula Blume $0{,}33 + 0{,}67\,\hat{\beta}$
- Eroarea prognozei: **RMSE** (Root Mean Squared Error, rădăcina erorii pătratice medii)
  - beta neajustat: RMSE 0,135
  - beta Blume: RMSE 0,109
  - beta Vasicek: RMSE 0,114
- Contracția reduce eroarea de prognoză cu 15–20%
- Cu doar 9 sectoare, concluzia este calitativă

MFM_ch3_beta_shrink

Beta CAPM la Bursa de Valori București (1/3)



Legenda: OLS beta, 95% CI: beta OLS cu intervalul de încredere (CI, confidence interval) de 95%, erori Newey–West; Vasicek-adjusted beta: beta ajustat Vasicek

MFM_ch3_bvb_betas

Beta CAPM la Bursa de Valori București (2/3)

- Randamente logaritmice zilnice pe BET, de la 6 ianuarie 2015 sau de la listare
  - DIGI: mai 2017, H2O (simbolul bursier al Hidroelectrica): iulie 2023
  - prețurile FP sînt ajustate pentru dividendul special
- Băncile sînt cele mai sensibile la piață
  - Banca Transilvania (TLV): 1,24, cu $R^2 = 0{,}62$
  - BRD: 1,13
- Cele mai puțin sensibile
  - Digi: 0,67, cu $R^2 = 0{,}21$
  - Hidroelectrica: 0,74, Transelectrica: 0,76
- R^2 este mare la TLV și pentru că TLV are pondere mare în BET
  - acțiunea face parte din propriul reper

MFM_ch3_bvb_betas

Beta CAPM la Bursa de Valori București (3/3)

- Contracția Vasicek schimbă puțin estimațiile
  - valorile beta zilnice sînt precise (SE între 0,04 și 0,07)
- Prețurile neactualizate (*stale prices*) deplasează beta zilnic spre zero
  - beta Dimson urcă de la 0,80 la 0,94 la Nuclearelectrica
  - și de la 0,81 la 0,905 la Transgaz (Seminarul 3, problema B4)
- BET este un indice de preț, iar prețurile acțiunilor includ dividendele
  - alfa este deplasat în sus cu aproximativ beta înmulțit cu randamentul dividendelor indicelui
  - beta se schimbă puțin cu BET-TR (*BET Total Return*, indicele BET cu dividendele reinvestite)

MFM_ch3_bvb_betas

O SML plată: Black și Betting Against Beta



Fischer Black (1938–1995)

Foto: Dalmatine at English Wikipedia (încărcată 2009); domeniu public; Wikimedia Commons

- Black (1972): fără împrumut nelimitat la R_f, randamentul rămîne liniar în beta
  - dar dreapta pornește de la randamentul unui portofoliu **cu beta zero**
  - acest randament este mai mare decît R_f
- Predicție: SML este **mai plată** decît dreapta CAPM
- Frazzini & Pedersen (2014): investitorii care vor mai mult risc, dar nu se pot împrumuta, cumpără active cu beta mare
  - acestea devin scumpe și au randamente viitoare mai mici
- Strategia lor, **BAB** (Betting Against Beta, pariul împotriva lui beta):

$$r_t^{\mathbf{BAB}} = \frac{R_{L,t} - R_{f,t}}{\beta_L} - \frac{R_{H,t} - R_{f,t}}{\beta_H}$$

  - $R_{L,t}$, $R_{H,t}$: randamentele portofoliilor cu beta mic (L) și cu beta mare (H)
  - β_L, β_H: valorile lor beta
- Strategia este long $1/\beta_L > 1$ unități în L și short $1/\beta_H < 1$ unități în H
  - poziția long este finanțată la R_f, iar încasările din short sînt plasate la R_f
  - strategia are beta zero și un randament mediu pozitiv în datele lor
- Simpla diferență beta mic minus beta mare, fără scalare, are beta negativ
  - media ei amestecă prima BAB cu o expunere negativă la piață (Seminarul 3, problema C2)

GRS ca test al raportului Sharpe

- O identitate algebrică (Gibbons, Ross & Shanken (1989)), cu toate momentele estimate prin verosimilitate maximă

$$\hat{\alpha}^\top \hat{\Sigma}^{-1} \hat{\alpha} = \widehat{SR}^2(f, \mathrm{R}) - \widehat{SR}^2(f)$$

  - $\widehat{SR}^2(f, \mathrm{R})$: pătratul raportului Sharpe al portofoliului tangent format din factor și cele N active de test
  - $\widehat{SR}^2(f)$: cel al factorului singur
- Deci GRS măsoară creșterea raportului Sharpe obținută prin adăugarea activelor de test

$$W = \frac{T - N - 1}{N} \cdot \frac{\widehat{SR}^2(f, \mathrm{R}) - \widehat{SR}^2(f)}{1 + \widehat{SR}^2(f)}$$

- Raportul Sharpe lunar al pieței (Mkt-RF) singure este 0,134 (0,47 pe an)
- Raportul Sharpe lunar al portofoliului tangent format din piață și cele 25 de portofolii este 0,405 (1,40 pe an)
- $W = 29{,}24 \times (0{,}405^2 - 0{,}134^2)/(1 + 0{,}134^2) = 4{,}20$
  - $29{,}24 = (757 - 25 - 1)/25$
- Sub CAPM, zgomotul ar crește $\widehat{SR}^2$ lunar în medie cu $N(1 + \widehat{SR}^2(f))/(T - N - 3) = 0{,}035$
  - observat: 0,146 (Anexă)

Metoda generalizată a momentelor (1/2)

- Un model economic implică $\mathbb{E}[\mathrm{g}_t(\boldsymbol{\theta})] = 0$ pentru un vector de condiții de momente g_t
- **GMM** (Hansen (1982)) alege $\hat{\boldsymbol{\theta}}$ astfel încît media empirică a momentelor să fie zero

$$\bar{\mathrm{g}}(\hat{\boldsymbol{\theta}}) = \tfrac{1}{T} \textstyle\sum_{t=1}^{T} \mathrm{g}_t(\hat{\boldsymbol{\theta}}) = 0$$

  - GMM: Generalized Method of Moments, metoda generalizată a momentelor
  - cazul exact identificat: tot atîtea momente cîți parametri
  - cu mai multe momente decît parametri, nu toate pot fi zero, iar $\bar{\mathrm{g}}$ se face cît mai mic, cu o matrice de ponderare
- Varianța sandwich în cazul exact identificat

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}) \approx \tfrac{1}{T}\, \mathrm{D}^{-1}\mathrm{S}\,\mathrm{D}^{-1\top}, \qquad \mathrm{D} = \mathbb{E}\big[\partial \mathrm{g}_t / \partial \boldsymbol{\theta}^{\top}\big], \qquad \mathrm{S} = \textstyle\sum_{j=-\infty}^{\infty} \mathrm{Cov}(\mathrm{g}_t, \mathrm{g}_{t-j})$$

  - D: jacobianul momentelor (derivatele lor după parametri)
  - S: varianța pe termen lung a momentelor, inclusiv autocovariațele, estimată Newey–West (Secțiunea 3)
- Cazul cel mai simplu, o medie: $g_t = R_t - \mu$ dă $D = -1$ și $\mathrm{Var}(\hat{\mu}) \approx S/T$
  - așa se obține statistica t HAC a randamentului mediu al unui factor (Aprofundări)

Metoda generalizată a momentelor (2/2)

- OLS: $\mathrm{g}_t = \mathrm{x}_t(y_t - \mathrm{x}_t^\top \boldsymbol{\theta})$ dă $\mathrm{D} = -\mathbb{E}[\mathrm{x}_t \mathrm{x}_t^\top]$
  - erorile White cînd S nu are laguri
  - Newey–West în rest (Aprofundări)
- **GRS ca GMM**: parametrii (α_i, β_i), $i = 1, \ldots, N$
  - $2N$ momente $\varepsilon_t \otimes (1, f_t)$: fiecare reziduu înmulțit cu 1 și cu factorul ($\otimes$: produsul Kronecker)
  - sistemul este exact identificat, deci estimațiile sînt chiar cele OLS, activ cu activ; GMM schimbă doar varianța lor
  - parametrii ordonați activ cu activ, $(\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots)$, dau $\mathrm{D} = -\mathrm{I}_N \otimes (\mathrm{Z}^\top \mathrm{Z}/T)$, unde Z este matricea $T \times 2$ cu rîndurile $(1, f_t)$
- $\hat{\mathrm{V}}_\alpha$: blocul matricei sandwich care corespunde valorilor alfa, calculat cu un $\hat{\mathrm{S}}$ HAC
- Statistica Wald $W_{\mathrm{GMM}} = \hat{\boldsymbol{\alpha}}^\top \hat{\mathrm{V}}_\alpha^{-1} \hat{\boldsymbol{\alpha}} \to \chi^2_N$ sub H_0
- Același aparat dă erorile primelor robuste la specificarea greșită (Anexă)
  - și estimează modelul cu factor de actualizare din Aprofundări

Încălcarea ipotezelor GRS

- **Cozi groase și gruparea volatilității** (*volatility clustering*) (Capitolul 1): distribuția F nu mai este exactă
  - Bootstrap pe reziduuri sub H_0: se reeșantionează rînduri întregi de reziduuri și se reconstruiesc randamentele cu $\alpha = 0$ (Seminarul 3, problema B7)
  - Bootstrap-ul este robust la lipsa normalității doar dacă erorile sînt i.i.d. și independente de factori
  - Testul Wald GMM (Aprofundări), cu $\hat{V}_\alpha$ HAC: robust la heteroscedasticitate și autocorelație
- **Prea multe active față de numărul de luni**: $\hat{\boldsymbol{\Sigma}}$ are rangul cel mult $T - 2$
  - cînd $N \geq T - 1$, $\hat{\boldsymbol{\Sigma}}$ nu se poate inversa și GRS nu se poate calcula
  - puterea testului scade mult înainte de această limită
- Pentru multe active există teste care nu inversează $\hat{\boldsymbol{\Sigma}}$ (Pesaran & Yamagata (2024); Anexă)

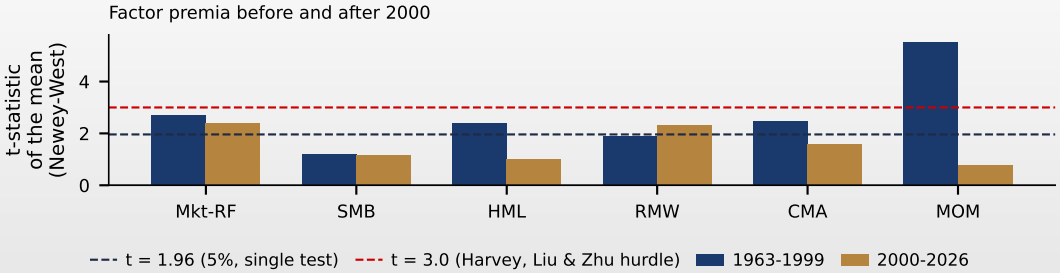
Motivația FF5: modelul de actualizare a dividendelor

- Fama & French (2015) pornesc de la evaluarea prin actualizarea dividendelor

$$\frac{M_t}{B_t} = \frac{\sum_{\tau \geq 1} \mathbb{E}[Y_{t+\tau} - dB_{t+\tau}]/(1+r)^\tau}{B_t}$$

  - M: capitalizarea, B: valoarea contabilă, Y: profitul, dB: investițiile
  - r: randamentul așteptat, rata de actualizare
- La M/B dat, un profit așteptat mare (RMW) sau investiții așteptate mici (CMA) implică un randament așteptat r mare
- Relația motivează factorii, dar nu arată că ei sînt riscuri cu primă

Stabilitatea primelor în timp (1/2)



MFM_ch3_factor_tstats

Stabilitatea primelor în timp (2/2)

- Barele: statistica t Newey–West a randamentului lunar mediu
  - 1963–1999 față de 2000–2026
  - tăietură arbitrară, nu data publicării
  - scăderea după publicare: Aprofundări
- Liniile: $t = 1{,}96$ (5%, un singur test) și $t = 3$
  - $t = 3$: pragul pentru testarea multiplă al Harvey, Liu & Zhu (2016) (Secțiunea 7)
- A doua perioadă are 319 luni față de 438
  - deci t scade cu factorul $\sqrt{319/438} \approx 0{,}85$ chiar la o medie neschimbată
  - de aceea comparăm și mediile
- Momentum: t scade de la 5,53 la 0,75
  - media anuală scade de la 10,7% la 2,5%
- Valoare (HML): t scade de la 2,39 la 1,02
  - media anuală scade de la 4,2% la 2,8%
- Profitabilitatea (RMW) este excepția
  - t crește de la 1,90 la 2,31
- Prima pieței rămîne semnificativă în ambele perioade ($t = 2{,}70$ și 2,39)

MFM_ch3_factor_tstats

Erori standard cînd modelul este greșit specificat

- Erorile Fama–MacBeth, Newey–West și Shanken presupun că modelul este corect
  - adică erorile de evaluare sînt zero: $\mathrm{e} = \boldsymbol{\mu} - \mathrm{X}\boldsymbol{\gamma} = 0$
  - $\boldsymbol{\mu}$: randamentele medii în exces, $\boldsymbol{\gamma} = (\lambda_0, \boldsymbol{\lambda}^\top)^\top$
- Dacă modelul este greșit ($\mathrm{e} \neq 0$), zgomotul din valorile beta estimate adaugă variață primelor
- Erorile standard robuste la această problemă sînt cele Kan–Robotti–Shanken (Kan, Robotti & Shanken (2013); formula: Anexă)
- Robust la specificare greșită nu înseamnă HAC: dependența temporală cere și autocovariațele momentelor (Newey–West)

Exemplu rezolvat: patru erori standard pentru prima de valoare

FF3, HML	$\hat{\lambda}$	SE	t	Semnificativ la 5%?
Fama–MacBeth	3,97%	1,33%	2,99	da
Newey–West (6 laguri)	3,97%	1,62%	2,45	da
Shanken	3,97%	1,33%	2,99	da
Kan–Robotti–Shanken (fără laguri)	3,97%	1,32%	3,00	da
Kan–Robotti–Shanken (6 laguri)	3,97%	1,61%	2,47	da

- 25 de portofolii mărime $\times$ B/M, 1963–2026
- Corecția Shanken nu schimbă eroarea, pentru că $c = 0{,}03$ pe lună
- Cel mai mult contează autocorelația pantelor lunare $\hat{\lambda}_t$
  - Newey–West cu 6 laguri mărește SE de la 1,33% la 1,62%
- Pentru RMW din FF5, Kan–Robotti–Shanken crește eroarea de la 2,1% la 2,6%, iar cu 6 laguri la 3,0%
  - t scade de la 2,5 la 2,1, apoi la 1,8, sub 1,96
- Pragul $t > 3$ (Secțiunea 7) nu este depășit clar cu nicio eroare standard

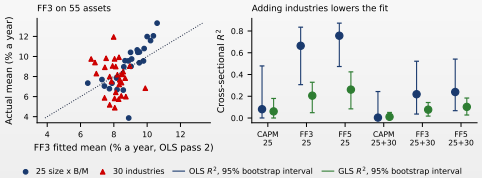
Factori inutili și factori slabi (1/2)

- Un **factor inutil** g_t este independent de toate randamentele
  - toți parametrii β_i sînt zero
  - estimările $\hat{\beta}_i$ sînt doar zgomot, de ordinul $1/\sqrt{T}$
- Etapa 2 regresează randamentele pe acest zgomot
  - $\hat{\lambda}_g$ nu tinde spre zero
  - testul t declară prea des că factorul are o primă cînd modelul este greșit (Kan & Zhang (1999))
- Experimentul: pe cele 25 de portofolii, unde GRS respinge CAPM (Secțiunea 4)
  - estimăm CAPM plus un factor inventat, extras independent din distribuția Normală
  - Repetăm experimentul de 2 000 de ori și testăm la nivelul 5% dacă prima factorului inventat este zero
  - Un test corect ar respinge în aproximativ 5% din repetări
  - Cu erori Fama–MacBeth, prima falsă pare semnificativă în 60% din repetări
  - Cu erori Shanken, prima falsă pare semnificativă în 27% din repetări

Factori inutili și factori slabi (2/2)

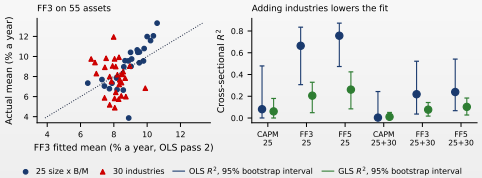
- Remediul: testați întîi, în etapa 1, $H_0 : \beta_g = 0$ pe toate cele N active
  - statistica: $W = \hat{\beta}_g^\top \widehat{\text{Var}}(\hat{\beta}_g)^{-1} \hat{\beta}_g \sim \chi^2_N$
  - dacă nu respingeți, prima factorului nu se poate estima
  - Testul Wald respinge doar în 6,7% din repetări, aproape de 5%
- **Identificare slabă**
  - valorile beta sînt aproape de zero (factor slab)
  - sau aproape egale între active (piața, în exemplul nostru)
  - Există teste robuste la identificarea slabă (Kleibergen (2009))
  - Cu multe active, consistența etapei 2 depinde de cît de puternici sînt factorii (Anatolyev & Mikusheva (2022))

Un test mai sever: adăugarea a 30 de portofolii pe industrii (1/2)



- Cele 25 de portofolii au o structură factorială puternică (Lewellen, Nagel & Shanken (2010))
  - trei factori explică în medie 91% din variația lor în timp (între 79% și 95%)
  - deci mai multe modele care surprind variația comună de mărime și valoare pot obține un R^2 transversal ridicat
- Stînga: media efectivă (vertical) față de media prezisă de FF3 (orizontal), % pe an
  - un model exact ar pune toate punctele pe diagonală
  - industriile (triunghiuri roșii) se împrăștie vertical

Un test mai sever: adăugarea a 30 de portofolii pe industrii (2/2)



- Dreapta: R^2 transversal cu intervale bootstrap de 95%
  - cu 30 de industrii adăugate, R^2 OLS al FF3 scade de la 0,66 la 0,22
  - R^2 OLS al FF5 scade de la 0,76 la 0,24
- Intervalele R^2 FF3: [0,31; 0,84] pe 25 de portofolii, [0,04; 0,52] pe 55 de active
  - intervalele marginale nu stabilesc semnificația diferenței: un test cere aceleași luni extrase pentru ambele mulțimi de active și un interval direct pentru ΔR^2
- R^2 GLS ponderează erorile cu inversa covarianței randamentelor
  - GLS: Generalised Least Squares, metoda celor mai mici pătrate generalizate
  - R^2 GLS nu depășește niciodată 0,26

Compararea modelelor: rapoarte Sharpe maxime (1/2)

Raport Sharpe maxim (anual)	CAPM	FF3	Carhart	FF5	FF5 + MOM
1963–2026 ($T = 757$)	0,47	0,66	0,97	0,97	1,16
2000–2026 ($T = 319$)	0,48	0,54	0,67	0,97	1,02

- Barillas & Shanken (2018): modelul mai bun este cel ai cărui factori permit cel mai mare raport Sharpe
  - $SR^2(\mathrm{f}) = \boldsymbol{\mu}_f^\top \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\mu}_f$
  - $\boldsymbol{\mu}_f$, $\boldsymbol{\Sigma}_f$: mediile și covarianța factorilor
- Motivul vine din GRS
  - valorile alfa ale unui model sînt mici exact cînd factorii lui se apropie de raportul Sharpe maxim
  - comparînd direct $SR^2(\mathrm{f})$, activele de test nu mai contează
  - spre deosebire de distanța Hansen–Jagannathan, calculată pe activele de test (Anexă)

Compararea modelelor: rapoarte Sharpe maxime (2/2)

- Cînd un model îl include pe celălalt, SR^2 în eșantion nu poate decît să crească
- Atunci se face o **regresie de spanning**: factorul adăugat se regresează pe ceilalți
  - un alfa nenul arată că el aduce ceva nou
  - HML pe ceilalți cinci factori: $\alpha = 0{,}9\%$ pe an ($t = 0{,}75$)
  - din 2000: $\alpha = -1{,}7\%$ ($t = -0{,}81$), deci nicio contribuție incrementală semnificativă a HML, ca în Fama & French (2015)
- FF5 față de Carhart (niciunul nu îl include pe celălalt): diferența dintre SR^2 anuale este 0,00
  - intervalul de 95%: $[-0{,}57; 0{,}66]$
- Din 2000: diferența este 0,50, cu intervalul de 95% $[-0{,}24; 1{,}48]$
  - deci nici aici nu este semnificativă
  - bootstrap staționar, Politis & Romano (1994); Seminarul 3, problema B8

Exemplu rezolvat: zece factori candidați (1/2)

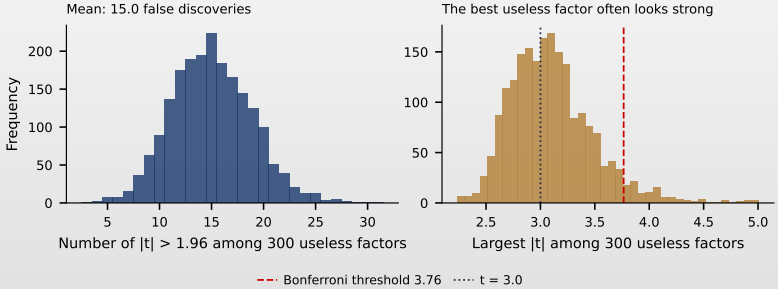
- Zece factori candidați, nivel $\alpha = 5\%$
  - valorile p ordonate: 0,0004; 0,002; 0,006; 0,011; 0,018; 0,027; 0,041; 0,070; 0,15; 0,62
- Regula naivă respinge orice ipoteză cu $p \leq 0{,}05$
  - găsește **7** factori semnificativi
- Bonferroni împarte 5% la cele 10 teste, deci pragul este 0,005
  - trec doar **2** factori

Exemplu rezolvat: zece factori candidați (2/2)

- Holm compară a j-a valoare p cu $0{,}05/(11-j)$
  - se oprește la primul care nu trece
- Primele trei valori p trec
  - $0{,}0004 \leq 0{,}005$
  - $0{,}002 \leq 0{,}0056$
  - $0{,}006 \leq 0{,}00625$
- Al patrulea, 0,011, depășește pragul 0,0071
  - deci Holm se oprește la **3** factori
- BH compară a j-a valoare p cu $0{,}005\,j$
  - cel mai mare j care trece este 6 ($0{,}027 \leq 0{,}030$)
  - deci BH declară semnificativi primii **6** factori

Simulare: 300 de factori inutili (1/2)

- Experimentul: 300 de factori fără nicio primă reală
  - statistica t a fiecăruia este o extragere independentă din distribuția Normală standard
- Repetăm experimentul de 2 000 de ori și numărăm, de fiecare dată, cîți factori trec pragul



Legenda: false discoveries: descoperiri false; The best useless factor often looks strong: cel mai bun factor inutil pare adesea puternic; Bonferroni threshold: pragul Bonferroni

MFM_ch3_false_discoveries

Simulare: 300 de factori inutili (2/2)

- La pragul clasic $|t| > 1{,}96$ trec, în medie, 15,0 factori inutili, adică 5% din 300
- La pragul $|t| > 3$ trec, în medie, 0,83 factori inutili
- În 57% din repetări, cel puțin un factor inutil trece pragul $|t| > 3$
- În jumătate din repetări, cel mai bun factor inutil are $|t|$ peste 3,06
- Pragul Bonferroni pentru 300 de teste este $|t| > 3{,}76$
- Pragul corect depinde de cîte teste s-au încercat efectiv, inclusiv cele nepublicate

MFM_ch3_false_discoveries

Numărul de factori (1/2)

- Model factorial aproximativ $x_{it} = \mathbf{\Lambda}_i^\top \mathsf{F}_t + e_{it}$
  - numărul de active N și cel de perioade T sînt ambele mari
  - x_{it}: randamentul standardizat, F_t: factorii, $\mathbf{\Lambda}_i$: încărcările, e_{it}: partea idiosincratică
  - $V(k)$: media pătratelor reziduurilor după primele k componente
- Bai & Ng (2002): alege k care minimizează $\ln V(k) + k \frac{N+T}{NT} \ln \min(N, T)$
  - criteriul de informație IC_{p2}: potrivire plus o penalizare pe factor
- Ahn & Horenstein (2013): alege k unde raportul ℓ_k / ℓ_{k+1} este maxim
  - adică acolo unde valorile proprii scad cel mai brusc
- Regula lui Kaiser (1960) ($\ell_k > 1$) nu are o justificare teoretică
  - chiar și zgomotul pur produce valori proprii împrăștiate în jurul lui 1

Numărul de factori (2/2)

Panel	N	T	Kaiser	Bai–Ng	Ahn–Horenstein
11 ETF-uri sectoriale, zilnic 2018–2026	11	2 073	2	5	1
100 portofolii mărime $\times$ B/M, lunar 1963–2026	95	757	3	4	1

- Randamente standardizate
  - cel mult $k = 5$ factori pentru sectoare și $k = 8$ pentru portofolii
  - 5 portofolii cu luni lipsă sînt eliminate
- Pe cele 95 de portofolii, Ahn–Horenstein alege doar factorul pieței (raport 14,6)
  - al doilea vîrf este la $k = 3$ (raport 3,5), ca în FF3
- Bai–Ng pe 11 sectoare alege numărul maxim permis, $k = 5$
  - penalizarea este gîndită pentru multe active
  - deci cu doar 11 răspunsul nu este de încredere (Anexă)

PCA instrumentată (1/2)

- **IPCA**: Instrumented Principal Component Analysis (PCA instrumentată), Kelly, Pruitt & Su (2019)
- Încărcările sînt funcții de L caracteristici observate $z_{i,t}$ (mărime, B/M, momentum, ...)

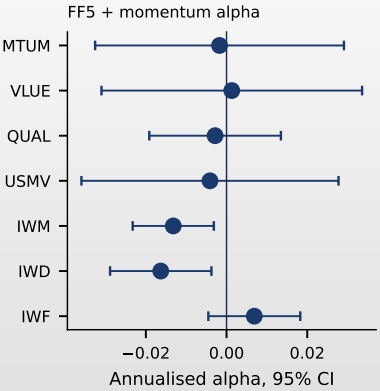
$$R^e_{i,t+1} = \alpha_{i,t} + \boldsymbol{\beta}^\top_{i,t} f_{t+1} + \varepsilon_{i,t+1}, \qquad \boldsymbol{\beta}^\top_{i,t} = z^\top_{i,t} \boldsymbol{\Gamma}_\beta$$

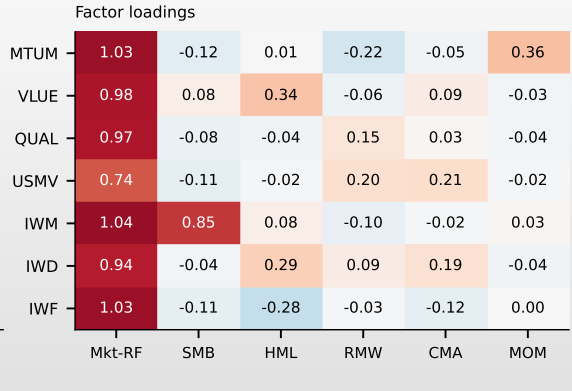
- $\boldsymbol{\Gamma}_\beta$ ($L \times K$) și factorii f_{t+1} minimizează $\sum_{i,t} (R^e_{i,t+1} - z^\top_{i,t} \boldsymbol{\Gamma}_\beta f_{t+1})^2$
  - prin pași alternativi
  - cu $\boldsymbol{\Gamma}_\beta$ fixat, f_{t+1} vine dintr-o regresie transversală
  - cu factorii fixați, $\boldsymbol{\Gamma}_\beta$ vine dintr-o regresie pe panel

PCA instrumentată (2/2)

- Întrebarea testată: prezic caracteristicile randamentele pentru că măsoară expunerea la risc (prin beta)?
  - sau direct, ca alfa ($\alpha_{i,t} = \mathbf{z}_{i,t}^\top \mathbf{\Gamma}_\alpha$)?
- Rezultatul, pe acțiuni americane: cîțiva factori IPCA explică randamentele medii mai bine
  - decît modelele cu tot atîția factori observați
- $\mathbf{\Gamma}_\alpha$ nu este semnificativ
  - de unde titlul articolului: „caracteristicile sînt covariațe"
- Motivația: IPCA rezolvă două limite ale PCA de pe slide-ul „Limitele PCA"
  - încărcări constante
  - cerința unui panel complet
  - extinderi: Anexă

Valorile alfa ale ETF-urilor factoriale (1/2)





MFM_ch3_etf_alphas

Valorile alfa ale ETF-urilor factoriale (2/2)

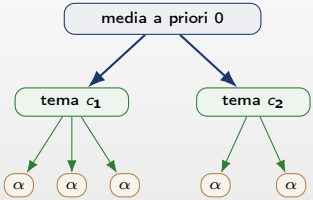
- Întrebarea: au ETF-urile un randament peste cel explicat de expunerile lor?
  - $R_t^e = \alpha + \boldsymbol{\beta}^\top \mathbf{f}_t + \varepsilon_t$, cu cei șase factori (FF5 + MOM)
- Regresii zilnice, erori Newey–West
  - fiecare ETF pe propriile zile de tranzacționare din 2000–2026
  - graficul: alfa anualizat cu interval de 95% și încărcările; pe axă, 0.02 înseamnă 2%
- Încărcările corespund denumirilor
  - MTUM pe MOM 0,36, VLUE pe HML 0,34
  - IWM pe SMB 0,85, IWF pe HML $-0{,}28$
- USMV (volatilitate minimă) are beta de piață 0,74
- Valori alfa aproape de zero
  - MTUM: $-0{,}2\%$ ($t = -0{,}11$)
  - QUAL: $-0{,}3\%$ ($t = -0{,}34$)
- IWM și IWD au alfa negativ, semnificativ individual la 5%; IWF are alfa pozitiv, nesemnificativ
  - IWM: $-1{,}3\%$ ($t = -2{,}57$), IWD: $-1{,}6\%$ ($t = -2{,}54$)
  - comisioanele (sub 0,2% pe an) explică doar o mică parte
  - reconstituirea periodică a indicilor Russell și construcția diferită de Fama–French: explicații posibile, necuantificate separat
- R^2 între 0,81 și 0,97
  - ETF-urile factoriale vînd, în mare parte, expuneri la factori cunoscuți (Seminarul 3, problema B11)

MFM_ch3_etf_alphas

Alocarea dinamică între factori, aglomerarea și costurile

- **Alocarea dinamică între factori**: mutarea capitalului între factori după semnale
  - de exemplu, cît de ieftin este un factor sau cum a evoluat recent
  - Primele sînt incerte și seriile scurte, deci este greu de demonstrat că alocarea dinamică funcționează în afara eșantionului (Capitolul 13)
- **Aglomerarea pozițiilor** (*crowding*): mulți investitori dețin aceleași poziții factoriale
  - Primele se pot micșora (McLean & Pontiff (2016)), iar aglomerarea poate amplifica pierderile la lichidări forțate
  - Crahul momentum-ului (*momentum crash*): după o scădere a pieței, portofoliul momentum are beta negativ, deci pierde mult cînd piața revine brusc, ca în 2009 (Daniel & Moskowitz (2016))
- **Implementarea**: randamentele factorilor academici nu scad două costuri
  - costurile de tranzacționare
  - costul împrumutului acțiunilor vîndute în lipsă

153 de factori grupați în 13 teme



alfa fiecărui factor preia informație de la ceilalți factori din tema sa

- **Date** (Secțiunea II a articolului): 153 de factori în 93 de țări
  - acțiunile sînt împărțite în treimi după fiecare caracteristică, cu ponderi după capitalizare, plafonate
  - se cumpără treimea care avea randamentul mai mare în studiul inițial și se vinde cealaltă treime extremă
  - fiecare factor este scalat la o volatilitate specifică (necorelată cu piața) de 10% pe an, ca valorile alfa să fie comparabile
- **Teme**: 13 grupe de factori americani
  - grupare ierarhică (*clustering*): factorii ale căror reziduuri CAPM sînt puternic corelate intră în aceeași temă
- **Modelul** (Secțiunea I): $f_{i,t} = \alpha_i + \beta_i r_t^m + \varepsilon_{i,t}$
  - $f_{i,t}$: randamentul factorului i; r_t^m: randamentul în exces al pieței; β_i: expunerea la piață; $\varepsilon_{i,t}$: zgomotul
  - $\alpha_i = c_j + s_n + w_i$: componenta temei j, a semnalului n (caracteristica, de exemplu B/M) și componenta specifică factorului i (semnalul n construit într-o anumită țară), toate cu media a priori 0

Contracția lui alfa și rata de replicare

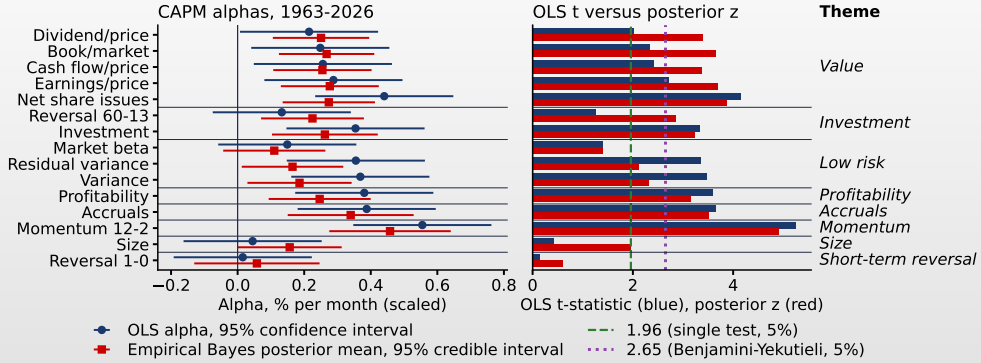
- Cazul unui singur factor: alfa estimat este micșorat spre zero

$$\mathbb{E}(\alpha \mid \hat{\alpha}) = \kappa\hat{\alpha}, \qquad \kappa = \frac{1}{1 + \sigma^2/(\tau^2 T)}$$

  - ecuația 5 din articol; ierarhia completă: Anexă
  - σ: volatilitatea reziduurilor
  - τ: dispersia valorilor alfa adevărate
  - T: numărul de luni
  - eșantioanele scurte și zgomotoase sînt contractate mai puternic, ca la Vasicek (Secțiunea 3): σ^2/T joacă rolul lui $\mathrm{se}(\hat{\beta}_i)^2$, iar τ^2 pe cel al lui σ_β^2

- **Rata de replicare**
  - dintre factorii semnificativi în studiile inițiale, ce proporție are alfa semnificativ și în datele noi
  - în varianta bayesiană, un factor este „semnificativ" dacă media a posteriori împărțită la abaterea standard a posteriori este cel puțin 1,96

Se replică 15 factori americani? (1/2)



Legenda: OLS alpha, 95% confidence interval: alfa OLS cu intervalul de încredere; Empirical Bayes posterior mean, 95% credible interval: media a posteriori Bayes empiric cu intervalul de credibilitate; OLS t versus posterior z: t OLS față de z a posteriori; Theme: tema

MFM_ch3_jkp_replication

Se replică 15 factori americani? Cum citim graficul

- Panoul stîng: alfa CAPM al fiecărui factor, în % pe lună
  - scalat la o volatilitate idiosincratică de 10% pe an
  - albastru: alfa OLS cu intervalul de încredere de 95%
  - roșu: media a posteriori Bayes empiric cu intervalul de credibilitate de 95%
- Panoul drept: t OLS (albastru) și z a posteriori (roșu)
  - z = media a posteriori împărțită la abaterea standard a posteriori
- Liniile punctate
  - 1,96: un singur test, 5%
  - 2,65: pragul Benjamini–Yekutieli, 5%
- 15 factori americani long–short din sortările ponderate după valoare ale Kenneth French Data Library
  - iulie 1963 – iulie 2026 ($T = 757$)
  - grupați pe temele articolului (etichetele autorilor)

Se replică 15 factori americani? (2/2)

- Rate de replicare
  - OLS: 73,3% (11 din 15 factori)
  - Benjamini–Yekutieli: 53,3% (8 factori; $|t|$ critic 2,65)
  - Bayes empiric: 86,7% (13 factori)
- Contracția trage alfa spre media temei, nu doar spre zero
- Exemplu: inversarea pe termen lung (tema investiții) crește de la 0,13% la 0,23% pe lună
  - t OLS este 1,25, dar z a posteriori este 2,85
- Nu se replică: beta de piață și inversarea pe termen scurt ($z \leq 1{,}41$)
- Mărimea trece la limită ($z = 1{,}96$)
  - și în articol mărimea este una dintre cele două teme cu rata sub 50%
- Rata se calculează aici pe toate cele 15 semnale (Anexă)
  - fără a verifica dacă fiecare a fost semnificativ în studiul inițial

MFM_ch3_jkp_replication

Ce nu poate arăta replicarea din curs

- **Limitele versiunii de curs**
  - 15 semnale americane în loc de 153 în 93 de țări
  - portofoliile din Kenneth French Data Library, nu treimile construite fără firmele foarte mici (microcapitalizări)
  - o singură țară, deci fiecare semnal apare o singură dată și componenta semnalului nu se poate estima (dispersia ei se fixează la zero)
  - rezultatul depinde de gruparea pe teme: dacă toți factorii ar fi într-o singură temă, rata Bayes empiric ar fi 100%
- **Limite precizate de autori** (Anexă)
  - CAPM este singurul reper
  - toate caracteristicile sînt construite după aceeași metodă
  - distribuția a priori ignoră factorii nepublicați
- **Întrebări deschise** (datele articolului se opresc în 2020)
  - s-au menținut valorile alfa și rata de replicare în 2021–2026?
  - pot factorii americani, folosiți ca distribuție a priori, să îmbunătățească estimările factorilor de la BVB, unde seriile sînt scurte?
- Datele și codul articolului: jkpfactors.com

Proiect: datele eșantionului și rezultatele cerute

- **Livrabile**
  - mai întîi, un tabel al evenimentelor: factorul, lucrarea inițială cu DOI, sfîrșitul eșantionului ei, data publicării, toate citite din lucrare (rîndul-model: MOM, Jegadeesh & Titman (1993))
  - o notă de preînregistrare (H_0, H_1, datele din tabel, lista factorilor, corecțiile), salvată în Git (commit) înainte de a vedea rezultatele
  - codul ca Quantlet: un director cu scriptul, sursa datelor și o scurtă descriere, ca în Quantlet-urile capitolului
  - un raport cu erori Newey–West, perioade placebo și puterea testului
  - puterea: probabilitatea de a respinge $b_2 = 0$ la 5% dacă scăderea reală ar fi, de exemplu, cei 58% din McLean & Pontiff (2016)
  - puterea se obține prin simulare sau aproximativ $\Phi(|\delta|/s - 1{,}96) + \Phi(-|\delta|/s - 1{,}96)$, cu $s = \mathrm{se}(\hat{b}_2)$
  - δ: scăderea presupusă, în unitățile lui b_2 (o scădere de 58%: $\delta = -0{,}58\times$ media de dinainte de publicare); Φ: funcția de repartiție a distribuției Normale standard
- **Jurnale**: prompturile în `AI_USE.md`, erorile depistate în `AI_ERRORS.md`

GRS sub ipoteza nulă

- Adăugarea a N active crește întotdeauna raportul Sharpe în eșantion
  - testul verifică dacă această creștere depășește ce ar produce zgomotul
- Forma Sharpe: $W = \frac{T-N-1}{N} \frac{\widehat{SR}^2(f,\mathrm{R}) - \widehat{SR}^2(f)}{1+\widehat{SR}^2(f)}$
- Sub H_0, cu reziduuri Normale i.i.d., $W \sim F(N, T-N-1)$
  - deci $\mathbb{E}[W] = \frac{T-N-1}{T-N-3}$
- Înmulțind cu $\frac{N(1+\widehat{SR}^2(f))}{T-N-1}$
  - $\mathbb{E}\left[\widehat{SR}^2(f,\mathrm{R}) - \widehat{SR}^2(f)\right] \approx \frac{N\,(1+\widehat{SR}^2(f))}{T-N-3}$
- 25 de portofolii, $T = 757$: $25 \times 1{,}018/729 = 0{,}035$ (lunar)
  - față de $0{,}405^2 - 0{,}134^2 = 0{,}146$ observat
- 9 ETF-uri sectoriale, $T = 331$: cîștig observat $0{,}191^2 - 0{,}146^2 = 0{,}015$
  - sub cel așteptat, $9 \times 1{,}021/319 = 0{,}029$

Înapoi

Teste pentru multe active

- $\hat{\mathbf{\Sigma}} = \hat{\mathbf{E}}^\top \hat{\mathbf{E}}/T$ are rangul cel mult $T - 2$
  - doi parametri în fiecare regresie
  - $\hat{\mathbf{E}}$: matricea $T \times N$ a reziduurilor
  - Exemplu: 100 de portofolii mărime $\times$ B/M pe 8 ani de date lunare ($T = 96$): $\hat{\mathbf{\Sigma}}$ este singulară, GRS nu există (Seminarul 3, problema B7)
- Teste pentru multe active care evită $\hat{\mathbf{\Sigma}}^{-1}$ (Pesaran & Yamagata (2024))
  - sume standardizate ale pătratelor statisticilor t
- Creșterea puterii prin selecția valorilor alfa mari: Fan, Liao & Yao (2015)

înapoi

Erori standard robuste la specificarea greșită

- Erorile Fama–MacBeth și Shanken presupun un model **corect**
  - toate erorile de evaluare $e = \mu - X\gamma$ sînt zero, cu $X = [1, B]$
- Aici GRS respinge restricțiile modelului cu factori tranzacționați ($\alpha = 0$)
  - relația transversală cu prime libere poate fi doar aproximativă: se estimează valoarea pseudo-adevărată, care minimizează erorile de evaluare
- Zgomotul din $\hat{B}$ se înmulțește cu e și adaugă un termen de variață (Kan, Robotti & Shanken (2013))
- Scriem ambele etape ca un singur sistem GMM exact identificat (Aprofundări)
  - momentele: $(R_t - a - Bf_t) \otimes (1, f_t)$ și $X^\top(R_t - X\gamma)$
  - varianța sandwich cu jacobianul complet este cea robustă
  - Cu $e = 0$ în jacobian se obține versiunea pentru un model corect specificat, robustă la heteroscedasticitate, a lui Jagannathan & Wang (1998)
  - matricea S a momentelor fără autocovariațe: robustă la specificare greșită, nu și la autocorelație; cu laguri Newey–West, și HAC
- FF5 pe cele 25 de portofolii (Secțiunea 6): erorile standard robuste, fără laguri (cu 6 laguri: RMW 3,0%, CMA 3,1%)
  - CMA: 2,2% → 2,7%
  - termenul liber: 3,3% → 3,8%; HML: schimbare sub 0,05 puncte procentuale
- Curs avansat: Generalized Method of Moments

înapoi

Distanța Hansen–Jagannathan și selecția bayesiană

- **Distanța Hansen–Jagannathan** (Hansen & Jagannathan (1997))

$$\delta^2 = \min_{\mathsf{b}} \; \mathsf{e}(\mathsf{b})^\top \, \mathbb{E}[\mathsf{R}^e \mathsf{R}^{e\top}]^{-1} \, \mathsf{e}(\mathsf{b}), \qquad \mathsf{e}(\mathsf{b}) = \mathbb{E}[m_t(\mathsf{b}) \, \mathsf{R}_t^e]$$

  - cît de departe este SDF-ul modelului de cel mai apropiat SDF care evaluează corect toate activele
  - Randamente în exces R_t^e cu normalizarea $\mathbb{E}[m] = 1$
  - Pentru randamente brute R_t: $\mathsf{e} = \mathbb{E}[m\mathsf{R}] - 1$, iar SDF-ul are nevoie de o scală liberă ($\mathbb{E}[m] = 1/(1 + R_f)$)
  - Aceeași matrice de ponderare pentru orice model, deci distanțele sînt comparabile
  - Statistica de test are o limită χ^2 ponderată
- Selecție printre foarte multe modele (Bryzgalova, Huang & Julliard (2023))
  - media bayesiană a circa 2×10^{15} de modele factoriale
  - robustă la factori slabi
- Kozak, Nagel & Santosh (2020) aplică o contracție coeficienților SDF
  - cîteva componente principale ale anomaliilor sînt suficiente
- Lecția comună: comparați toate modelele cu aceeași măsură și raportați incertitudinea ei

Înapoi

Grădina zoologică redusă prin LASSO cu dublă selecție

- Feng, Giglio & Xiu (2020): aduce un factor nou informație suplimentară față de factorii deja cunoscuți?
- **LASSO cu dublă selecție** (Least Absolute Shrinkage and Selection Operator)
  - LASSO: regresie cu o penalizare care pune exact la zero coeficienții variabilelor neimportante, deci selectează variabilele
  - Pasul 1: dintre sute de factori existenți, LASSO îi alege pe cei care explică randamentele medii ale activelor de test
  - Pasul 2: un al doilea LASSO îi alege pe cei corelați cu factorul nou, ca omiterea lor să nu distorsioneze rezultatul
  - Pasul 3: se testează coeficientul factorului nou în SDF, controlînd pentru toți factorii aleși în pașii 1 și 2
- Rezultat: majoritatea factorilor noi sînt redundanți
- Doar cîțiva factori noi explică randamente pe care cei vechi nu le explică
- Studiul de caz ajunge la o concluzie compatibilă
  - mulți factori sînt redundanți
  - informația se concentrează în cîteva teme

Înapoi

De ce nu concordă estimatorii numărului de factori

- Onatski (2010): un factor este real dacă valoarea lui proprie iese clar dintr-un interval
  - intervalul în care cad valorile proprii ale zgomotului
- Pentru zgomot i.i.d. standardizat, suportul spectral asimptotic (Marchenko–Pastur) este aproximativ $[(1-\sqrt{N/T})^2, (1+\sqrt{N/T})^2]$: un reper ilustrativ, nu testul Onatski
  - $= [0{,}86;\ 1{,}15]$ pentru $N = 11,\ T = 2\,073$
  - a doua valoare proprie, 1,06, este în interval
  - reperul presupune că toate seriile sînt zgomot; aici prima valoare proprie este 7,26, deci reperul nu clasifică direct a doua componentă
  - regula Kaiser ($\ell_k > 1$) reține două componente, dar nu stabilește dacă a doua este semnificativă
- Bai & Ng (2002): cu $N = 11,\ T = 2\,073$, penalizarea din IC_{p2} este 0,219 pe factor
- Pe sectoare, criteriul $IC_{p2}(k)$ pentru $k = 0, \ldots, 5$
  - $0;\ -0{,}86;\ -0{,}97;\ -1{,}12;\ -1{,}14;\ -1{,}17$
- Criteriul scade pînă la maximul permis, $k_{\max} = 5$
  - penalizarea este prea mică pentru a opri adăugarea factorilor slabi
- Raportați intervalul estimațiilor și sensibilitatea la $k_{\max}$, nu un singur număr
  - Seminarul 3, problema B3

înapoi

PCA cu primă de risc și estimarea primelor în trei etape

- **RP-PCA** (*Risk-Premium Principal Component Analysis*, analiza componentelor principale cu primă de risc; Lettau & Pelger (2020))
  - componentele principale ale matricei $\frac{1}{T}\mathbf{X}^\top\mathbf{X} + \gamma\bar{\mathbf{x}}\bar{\mathbf{x}}^\top$
  - $\mathbf{X}$: matricea $T \times N$ a randamentelor în exces, $\bar{\mathbf{x}}$: mediile lor, γ: ponderea randamentelor medii
  - $\gamma = -1$ dă PCA pe covariață
  - $\gamma > -1$ favorizează și factorii care explică randamentele medii
  - În datele autorilor, metoda găsește factori cu rapoarte Sharpe mai mari și erori de evaluare mai mici decît PCA
- **Estimatorul în trei etape** (Giglio & Xiu (2021))
  - prima unui factor observat g_t cînd modelul omite factori
  - (1) PCA pe randamentele activelor de test
  - (2) regresia transversală a mediilor pe valorile beta latente
  - (3) regresia lui g_t pe factorii latenți
  - prima lui g_t se obține din primele factorilor latenți și din coeficienții acestei regresii
- Ambele păstrează baza statistică a PCA
  - dar urmăresc explicarea randamentelor medii, nu doar a variației
- IPCA adaugă în plus caracteristicile activelor

Înapoi

IPCA și factorii criptoactivelor

- IPCA leagă PCA (statistica) de sortările după caracteristici (Fama–French)
  - beta variabil în timp, puțini factori
- Criptoactive: factorii piață, mărime și momentum (Liu, Tsyvinski & Wu (2022))
  - evaluează secțiunea transversală a criptomonedelor
- Reducerea dimensionalității: Pele et al. (2023), Jheng et al. (2026)
  - pe caracteristicile randamentelor criptoactivelor și activelor tradiționale
- Curs avansat: Multivariate Statistical Analysis

Înapoi

Construcția celor 153 de factori

- **Date** (Secțiunea II a articolului)
  - 138 din lista alcătuită de Hou, Xue & Zhang (2020) și 15 din alte studii
  - randamente americane din 1926, majoritatea piețelor dezvoltate din 1986
  - toate în dolari americani, în exces față de rata titlurilor de stat americane pe termen scurt
- **Construcția**, în fiecare țară și lună
  - treimi după caracteristică, cu praguri calculate fără firmele foarte mici (microcapitalizări)
  - ponderi după capitalizarea plafonată la percentila 80 a NYSE
  - perioada de deținere: o lună
  - cel puțin 5 acțiuni pe fiecare parte și 60 de observații lunare
- **Teme**: 13 grupe de factori americani (Secțiunea II.B)
  - grupare ierarhică, cu distanța $1-$ corelația reziduurilor CAPM și metoda Ward (unește grupele care cresc cel mai puțin varianța din interiorul grupelor)
  - Ward cere distanțe euclidiene: pentru serii standardizate, $d_{ij} = \sqrt{2(1-\rho_{ij})}$; autorii: R `ward.D` cu $1-\rho$; în SciPy, `ward` cu $\sqrt{2(1-\rho)}$ (echivalentul R: `ward.D2`)

Înapoi

Ierarhia bayesiană a valorilor alfa

- **Ierarhia** (Secțiunea I a articolului)
  - $\alpha_i = c_j + s_n + w_i$ (tema j, semnalul n, factorul i; ec. 21)
  - varianța a priori $\mathbf{\Omega} = \mathrm{M}\mathrm{M}^\top \tau_c^2 + \mathrm{Z}\mathrm{Z}^\top \tau_s^2 + \mathrm{I}\tau_w^2$ (ec. 23)
  - media a posteriori $(\mathbf{\Omega}^{-1} + T\mathbf{\Sigma}^{-1})^{-1} T\mathbf{\Sigma}^{-1}\hat{\boldsymbol{\alpha}}$ (Propoziția 4)
  - M, Z: matrice 0/1 care leagă fiecare factor de tema și semnalul lui
  - $\tau_c^2, \tau_s^2, \tau_w^2$: variațele celor trei componente ale lui alfa
  - $\mathbf{\Sigma}$: covarianța reziduurilor
  - estimare Bayes empirică: τ_c, τ_s, τ_w prin verosimilitate maximă pentru $\hat{\boldsymbol{\alpha}} \sim N(0, \mathbf{\Omega} + \hat{\mathbf{\Sigma}}/T)$ (Secțiunea I.C.2)

înapoi

Rezultatele complete

- Cu OLS se replică 98 din 119 factori americani
- Corecția Benjamini–Yekutieli cere $|t| > 2{,}7$ (Figura 4)
- FDR bayesian: 0,1%, cu intervalul de 95% $[0; 1\%]$
- Ponderea așteptată a factorilor adevărați: 94% (Secțiunea III.C)
- Rata de replicare pentru firmele cu capitalizare foarte mare și mare (Figura 5)
  - 77,3% și 79,8%
- 11 din 13 teme au o rată de cel puțin 50%
  - excepțiile sînt îndatorarea redusă și mărimea
- Rata mare și la firmele mari arată că rezultatul nu vine doar din firmele foarte mici
- Raportul Sharpe (lunar) al portofoliului tangent, SUA 1972–2020 (Figura 14)
  - piața: 0,13
  - toți factorii: 0,80

Înapoi

Replicarea din curs în detaliu

- Benjamini–Yekutieli dă cea mai mică rată, ca în rezultatele americane din articol (Figura 1)
- Emisiunile nete de acțiuni sînt trase spre media temei lor
  - de la 0,44% la 0,27% pe lună
- Abaterea standard a posteriori este 0,07%–0,10%, sub eroarea standard OLS de 0,11%
  - fiecare factor împrumută informație de la ceilalți factori ai temei
- **Dispersia a priori**: $\hat{\tau}_c = 0{,}28\%$, $\hat{\tau}_w = 0{,}04\%$ pe lună
  - față de 0,35% și 0,21% raportate în Tabelul I al articolului
  - cu 15 semnale rămîne puțină dispersie specifică factorilor
- **Teme**: cele 15 semnale se încadrează în 8 dintre cele 13 teme ale articolului
  - 5 în valoare, 3 în risc scăzut, 2 în investiții

Înapoi

Limite precizate de autori și extinderi

- **Limite precizate de autori**
  - CAPM este singurul model de referință: factorii de îndatorare redusă nu au alfa CAPM semnificativ, dar contribuie cînd sînt combinați cu celelalte teme (Secțiunea III.A)
  - Toate caracteristicile sînt construite după aceeași regulă, uneori altfel decît în articolul original (Secțiunea II.C)
  - Factorii încercați, dar nepublicați, ar schimba distribuția a priori: în scenariile lui Harvey, Liu & Zhu (2016), rata pentru SUA scade la 81,5%, respectiv 79,8% (Figura 10)
- **Completarea replicării de curs** (Seminarul 3, C3)
  - justificați semnul fiecărui factor pe baza studiului inițial
  - comparați dendrograma cu temele articolului
  - adăugați FDR bayesian și panta după eșantion (Figura 8)
- Datele și codul articolului
  - jkpfactors.com
  - github.com/bkelly-lab/ReplicationCrisis

Înapoi

Bucla descoperirii, prompturi reutilizabile (1/2)

- **Literatură**
  - `Rezumă dovezile din 2016 încoace despre scăderea după publicare și replicarea randamentelor anomaliilor. Pentru fiecare lucrare: DOI, eșantion, numărul de factori, estimația principală, identificarea scăderii. Marchează intrările nesigure.`
  - Verificați fiecare DOI în Crossref
  - Verificați fiecare cifră principală în rezumat
- **Ipoteză**
  - `Compară căutarea oportunistă în date, arbitrajul bazat pe informația publicată și riscul variabil în timp ca explicații ale scăderii. Pentru fiecare: tiparul prezis între perioadele din eșantion, după eșantion și după publicare, și între acțiunile cu limite mari și cele cu limite mici ale arbitrajului.`
- **Date și cod**
  - `Încarcă factorul lunar de momentum din Kenneth French Data Library; din 1965, regresează MOM pe o variabilă dummy după eșantion (1990-01 până la 1993-03) și una după publicare (din 1993-04), cu erori Newey-West.`
  - Replicați întîi o cifră cunoscută: momentum $t = 5{,}53$ înainte de 2000 și 0,75 după (graficul pe subperioade)

Înapoi

Bucla descoperirii, prompturi reutilizabile (2/2)

- **Robustețe**
  - `Scrie un plan de robustețe preînregistrat: cîte o dată de publicare pentru fiecare factor, corecția Holm între factori, excluderea crahului de momentum din 2009, data documentului de lucru în locul datei din revistă și perioada 1927-1962 ca placebo pentru Mkt-RF, SMB, HML și MOM (RMW și CMA încep în iulie 1963).`
- **Critică**
  - `Evaluează critic: un factor, o dată de publicare, t = -1,45. Care este puterea de a detecta o scădere de 58%? De ce o schimbă combinarea mai multor anomalii?`
- **Raport**
  - Media în eșantion, ambele variabile dummy cu erori Newey–West, placebo-ul, versiunea datelor (data descărcării)
  - Factori combinați: erori robuste la corelația dintre factori în aceeași lună și la autocorelație
  - Prompturile care au influențat o decizie intră în `AI_USE.md`
- Asistentul AI propune variantele de robustețe
- Cercetătorul alege dinainte care sînt raportate

Înapoi

Verificările cercetătorului

- **Data greșită a evenimentului**
  - data încheierii eșantionului, data difuzării documentului de lucru și data publicării în revistă sînt diferite
  - test: citiți-le din lucrarea însăși
- **Alegerea factorului după ce s-a văzut că a scăzut**: fixați dinainte lista factorilor
- Aplicați Holm sau BH pe toți factorii
- **Randamente autocorelate, cu crahuri**: erori Newey–West
- Reestimați fără anul 2009
- **Versiunile datelor**: biblioteca de factori este revizuită retroactiv
  - test: notați data descărcării
  - reproduceți statisticile t pe subperioade din capitol
- **Putere mică**: cu un singur factor, H_0 se respinge rar
- Raportați intervalul de încredere al scăderii, nu doar semnul ei

Înapoi

Răspunsurile la verificările rapide (1/2)

Beta, eroarea și contracția lui

- Clasic: $0{,}90 \pm 1{,}96 \times 0{,}05 = [0{,}802; 0{,}998]$: limita superioară este foarte aproape de 1, dar sub 1
- Newey–West: $0{,}90 \pm 1{,}96 \times 0{,}09 = [0{,}724; 1{,}076]$, conține 1: nicio dovadă că acțiunea are o sensibilitate la piață mai mică (riscul total include și componenta idiosincratică)
- Raportați Newey–West: reziduurile sînt heteroscedastice și autocorelate; HAC schimbă varianța estimată, nu estimația OLS
- $\sigma^2_\beta = 0{,}10 - 0{,}02 = 0{,}08$; SE 0,2: $w = 0{,}08/0{,}12 = 0{,}67$, $\beta^V = 0{,}67 \times 1{,}6 + 0{,}33 \times 1{,}0 = 1{,}40$; SE 0,1: $w = 0{,}08/0{,}09 = 0{,}89$, $\beta^V = 1{,}53$
- Estimația mai zgomotoasă este contractată mai puternic; Blume dă $0{,}33 + 0{,}67 \times 1{,}6 = 1{,}40$ indiferent de precizie
- Beta Dimson: $0{,}25 + 0{,}55 + 0{,}00 = 0{,}80$; prețul acțiunii încorporează randamentul pieței cu o zi întîrziere, deci panta din aceeași zi pierde partea de 0,25

Citirea unei statistici GRS

- $W = \frac{60}{4} \times (0{,}25^2 - 0{,}15^2)/(1 + 0{,}15^2) = 15 \times 0{,}040/1{,}0225 = 0{,}59 < 2{,}53$: nerespins
- Cîștigul așteptat $4 \times 1{,}0225/58 = 0{,}071$, peste cel observat, 0,040: distanța este în limitele zgomotului
- $T = 650$: $W = \frac{645}{4} \times 0{,}0391 = 6{,}31 > 2{,}39$: respins; distanța este aceeași, dovezile cresc cu T

Înapoi

Răspunsurile la verificările rapide (2/2)

Prime, Shanken și un R^2 negativ

- $\text{se}_{FM} = 3\%/\sqrt{400} = 0{,}15\%$, $t = 3{,}33$
- $c = 0{,}5^{\mathbf{2}}/1^{\mathbf{2}} = 0{,}25$; $\text{Var}_{\mathbf{Sh}} = 1{,}25 \times (0{,}0225 - 0{,}0025) + 0{,}0025 = 0{,}0275$, SE 0,166%, $t = 3{,}02$: doar puțin peste 3
- Factor tranzacționat: $c = 0{,}25/4{,}5^{\mathbf{2}} = 0{,}012$, deci corecția este neglijabilă
- Media 4%: $\text{SST} = 4 + 0 + 4 = 8$, $\text{SSE} = 16 + 0 + 16 = 32$, $R^{\mathbf{2}} = 1 - 32/8 = -3$; cu termen liber, pantele zero dau deja SSE = SST, iar OLS poate doar să micșoreze SSE

Numărarea descoperirilor

- Naiv 4; Bonferroni (0,01) 1; Holm: $0{,}004 \leq 0{,}01$, $0{,}012 \leq 0{,}0125$, $0{,}020 > 0{,}0167$ oprire: 2; pragurile BH $0{,}01j$: $0{,}030 \leq 0{,}040$ este ultima respingere: 4
- BY: suma este 2,283, pragurile $0{,}0044j$: trece doar 0,004: 1
- $1 - 0{,}95^{\mathbf{5}} = 0{,}23$
- O căutare printre candidați: BH (FDR)
- Holm sau Bonferroni (FWER) limitează la 5% probabilitatea a cel puțin unei respingeri false

Înapoi

Pentru aprofundare: cursuri și cod

Cursuri avansate pe quantinar.com

- Generalized Method of Moments
- Multivariate Statistical Analysis

Cod și date

- Toate graficele și tabelele: github.com/danpele/MFM/tree/main/Quantlets/Ch_03
- Datele factorilor: Kenneth French Data Library

Înapoi

Bibliografie (1/6)

- Ahn, S.C., Horenstein, A.R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3), 1203–1227.
- Anatolyev, S., Mikusheva, A. (2022). Factor models with many assets: strong factors, weak factors, and the two-pass procedure. *Journal of Econometrics*, 229(1), 103–126.
- Ang, A., Liu, J., Schwarz, K. (2020). Using stocks or portfolios in tests of factor models. *Journal of Financial and Quantitative Analysis*, 55(3), 709–750.
- Bai, J., Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.
- Barillas, F., Shanken, J. (2018). Comparing asset pricing models. *Journal of Finance*, 73(2), 715–754.
- Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57(1), 289–300.
- Benjamini, Y., Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4), 1165–1188.
- Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business*, 45(3), 444–455.
- Blume, M.E. (1971). On the assessment of risk. *Journal of Finance*, 26(1), 1–10.
- Bryzgalova, S., Huang, J., Julliard, C. (2023). Bayesian solutions for the factor zoo: we just ran two quadrillion models. *Journal of Finance*, 78(1), 487–557.

Bibliografie (2/6)

- Campbell, J.Y., Lo, A.W., MacKinlay, A.C. (1997). The Econometrics of Financial Markets. Princeton University Press.
- Carhart, M.M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1), 57–82.
- Chen, A.Y., Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review*, 11(2), 207–264.
- Cochrane, J.H. (2011). Presidential address: discount rates. *Journal of Finance*, 66(4), 1047–1108.
- Daniel, K., Moskowitz, T.J. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2), 221–247.
- Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, 7(2), 197–226.
- Ehsani, S., Linnainmaa, J.T. (2022). Factor momentum and the momentum factor. *Journal of Finance*, 77(3), 1877–1919.
- Fama, E.F., French, K.R. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2), 427–465.
- Fama, E.F., French, K.R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56.
- Fama, E.F., French, K.R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22.
- Fama, E.F., MacBeth, J.D. (1973). Risk, return, and equilibrium: empirical tests. *Journal of Political Economy*, 81(3), 607–636.
- Fan, J., Liao, Y., Yao, J. (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, 83(4), 1497–1541.

Bibliografie (3/6)

- Feng, G., Giglio, S., Xiu, D. (2020). Taming the factor zoo: a test of new factors. *Journal of Finance*, 75(3), 1327–1370.
- Frazzini, A., Pedersen, L.H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1), 1–25.
- Gibbons, M.R., Ross, S.A., Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57(5), 1121–1152.
- Giglio, S., Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy*, 129(7), 1947–1990.
- Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029–1054.
- Hansen, L.P., Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance*, 52(2), 557–590.
- Harvey, C.R., Liu, Y., Zhu, H. (2016). . . . and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5–68.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.

Bibliografie (4/6)

- Hou, K., Xue, C., Zhang, L. (2015). Digesting anomalies: an investment approach. *Review of Financial Studies*, 28(3), 650–705.
- Hou, K., Xue, C., Zhang, L. (2020). Replicating anomalies. *Review of Financial Studies*, 33(5), 2019–2133.
- Jagannathan, R., Wang, Z. (1998). An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *Journal of Finance*, 53(4), 1285–1309.
- Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance*, 45(3), 881–898.
- Jegadeesh, N., Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, 48(1), 65–91.
- Jensen, M.C. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance*, 23(2), 389–416.
- Jensen, T.I., Kelly, B., Pedersen, L.H. (2023). Is there a replication crisis in finance? *Journal of Finance*, 78(5), 2465–2518.
- Jheng, S.-L., Găman, Ș., Conda, A., Pele, D.T., Härdle, W.K., Teng, H.-W., Osterrieder, J. (2026). Crypto currency returns. Document de lucru SSRN.
- Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151.
- Kan, R., Robotti, C., Shanken, J. (2013). Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance*, 68(6), 2617–2649.
- Kan, R., Zhang, C. (1999). Two-pass tests of asset pricing models with useless factors. *Journal of Finance*, 54(1), 203–235.
- Kelly, B.T., Pruitt, S., Su, Y. (2019). Characteristics are covariances: a unified model of risk and return. *Journal of Financial Economics*, 134(3), 501–524.

Bibliografie (5/6)

- Kleibergen, F. (2009). Tests of risk premia in linear factor models. *Journal of Econometrics*, 149(2), 149–173.
- Kozak, S., Nagel, S., Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2), 271–292.
- Lettau, M., Pelger, M. (2020). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies*, 33(5), 2274–2325.
- Lewellen, J., Nagel, S., Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, 96(2), 175–194.
- Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1), 13–37.
- Liu, Y., Tsyvinski, A., Wu, X. (2022). Common risk factors in cryptocurrency. *Journal of Finance*, 77(2), 1133–1177.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77–91.
- McLean, R.D., Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, 71(1), 5–32.
- Moskowitz, T.J., Ooi, Y.H., Pedersen, L.H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2), 228–250.
- Newey, W.K., West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Newey, W.K., West, K.D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61(4), 631–653.
- Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics*, 92(4), 1004–1016.

Bibliografie (6/6)

- Pele, D.T., Wesselhöfft, N., Härdle, W.K., Kolossiatis, M., Yatracos, Y.G. (2023). Are cryptos becoming alternative assets? *The European Journal of Finance*, 29(10), 1064–1105.
- Pesaran, M.H., Yamagata, T. (2024). Testing for alpha in linear factor pricing models with a large number of securities. *Journal of Financial Econometrics*, 22(2), 407–460.
- Politis, D.N., Romano, J.P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428), 1303–1313.
- Roll, R. (1977). A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics*, 4(2), 129–176.
- Ross, S.A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), 341–360.
- Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies*, 5(1), 1–33.
- Sharpe, W.F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3), 425–442.
- Vasicek, O.A. (1973). A note on using cross-sectional information in Bayesian estimation of security betas. *Journal of Finance*, 28(5), 1233–1239.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838.
- Date: Kenneth R. French Data Library, mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Acronime folosite în acest material (1/4)

- **AI**: Artificial Intelligence — *inteligență artificială*
- **AMEX**: American Stock Exchange — *Bursa Americană de Valori*
- **APT**: Arbitrage Pricing Theory — *teoria evaluării prin arbitraj*
- **BAB**: Betting Against Beta — *strategia „pariu împotriva lui beta"*
- **BET**: Bucharest Exchange Trading (index) — *indicele de referință al Bursei de Valori București*
- **BET-TR**: BET Total Return (index) — *indicele BET cu dividendele reinvestite*
- **B/M**: Book-to-Market — *raportul dintre valoarea contabilă și valoarea de piață*
- **BH**: Benjamini–Hochberg (procedure) — *procedura Benjamini–Hochberg*
- **BM**: Book-to-Market (B/M quintile in the size–value chart) — *chintila raportului B/M din graficul mărime–valoare*
- **BVB**: Bursa de Valori București
- **BY**: Benjamini–Yekutieli (procedure) — *procedura Benjamini–Yekutieli*
- **CAGR**: Compound Annual Growth Rate — *rata anuală compusă de creștere*
- **CAPM**: Capital Asset Pricing Model — *modelul de evaluare a activelor financiare*
- **CI**: Confidence Interval — *interval de încredere*

Acronime folosite în acest material (2/4)

- **CMA**: Conservative Minus Aggressive (investment factor) — *factorul investiții: firme conservatoare minus agresive*
- **CML**: Capital Market Line — *dreapta pieței de capital*
- **DOI**: Digital Object Identifier — *identificator digital de obiect*
- **ETF**: Exchange-Traded Fund — *fond tranzacționat la bursă*
- **FDR**: False Discovery Rate — *rata descoperirilor false*
- **FF3**: Fama–French three-factor model — *modelul Fama–French cu trei factori*
- **FF5**: Fama–French five-factor model — *modelul Fama–French cu cinci factori*
- **FWER**: Family-Wise Error Rate — *probabilitatea de cel puțin o eroare de tip I în familia de teste*
- **GICS**: Global Industry Classification Standard — *standardul global de clasificare a industriilor*
- **GLS**: Generalised Least Squares — *metoda celor mai mici pătrate generalizate*
- **GMM**: Generalized Method of Moments — *metoda generalizată a momentelor*
- **GRS**: Gibbons–Ross–Shanken (test) — *testul Gibbons–Ross–Shanken*
- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors) — *erori standard robuste la heteroscedasticitate și autocorelație*
- **HML**: High Minus Low (value factor) — *factorul valoare: B/M mare minus mic*

Acronime folosite în acest material (3/4)

- **IPCA**: Instrumented Principal Component Analysis — *analiza instrumentată a componentelor principale*
- **LASSO**: Least Absolute Shrinkage and Selection Operator — *regresie penalizată L1 cu selecția variabilelor*
- **ME**: Market Equity (market capitalisation) — *capitalizarea bursieră*
- **MOM**: Momentum (factor) — *factorul momentum*
- **NASDAQ**: National Association of Securities Dealers Automated Quotations — *bursa americană NASDAQ (cotații automate ale Asociației Naționale a Dealerilor de Valori Mobiliare)*
- **NYSE**: New York Stock Exchange — *Bursa din New York*
- **OLS**: Ordinary Least Squares — *metoda celor mai mici pătrate*
- **OMV**: Österreichische Mineralölverwaltung — *Administrația Austriacă a Petrolului (grupul OMV)*
- **PC1**: first Principal Component — *prima componentă principală*
- **PC2**: second Principal Component — *a doua componentă principală*
- **PC3**: third Principal Component — *a treia componentă principală*
- **PCA**: Principal Component Analysis — *analiza componentelor principale*
- **PRDS**: Positive Regression Dependence on a Subset (of p-values) — *dependența pozitivă de regresie pe o submulțime (a valorilor p)*

Acronime folosite în acest material (4/4)

- **RF**: Risk-Free rate — *rata fără risc*
- **RMSE**: Root Mean Squared Error — *rădăcina erorii pătratice medii*
- **RMW**: Robust Minus Weak (profitability factor) — *factorul profitabilitate: firme robuste minus slabe*
- **ROE**: Return On Equity — *rentabilitatea capitalului propriu*
- **RP-PCA**: Risk-Premium Principal Component Analysis — *analiza componentelor principale cu primă de risc*
- **SDF**: Stochastic Discount Factor — *factorul stocastic de actualizare*
- **SE**: Standard Error — *eroarea standard*
- **SMB**: Small Minus Big (size factor) — *factorul mărime: firme mici minus mari*
- **SML**: Security Market Line — *dreapta pieței titlurilor*
- **SPDR**: Standard & Poor's Depositary Receipts — *certificate de depozit Standard & Poor's (familia de ETF-uri SPDR)*
- **SSE**: Sum of Squared Errors — *suma pătratelor erorilor*
- **SST**: Total Sum of Squares — *suma totală a pătratelor abaterilor de la medie*
- **TSFM**: Time-Series Factor Momentum — *momentum-ul factorilor în serie de timp*
- **SUA**: Statele Unite ale Americii
- **UMD**: Up Minus Down (momentum factor) — *factorul momentum: cîștigători minus perdanți*