








Serii de timp

Capitolul 1: Procese stochastice și staționaritate



Daniel Traian PELE

Academia de Studii Economice din București
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Academia Română, Institutul de Prognoză Economică
MSCA Digital Finance

Acronime folosite în acest material

- **ACF**: Autocorrelation Function (funcția de autocorelație)
- **AI**: Artificial Intelligence (inteligență artificială)
- **AR**: AutoRegressive (model) — model autoregresiv
- **ARIMA**: AutoRegressive Integrated Moving Average (model) — model autoregresiv integrat cu medie mobilă
- **ARMA**: AutoRegressive Moving Average (model) — model autoregresiv cu medie mobilă
- **BET**: Bucharest Exchange Trading (index) — indicele de referință al Bursei de Valori București
- **BNR**: Banca Națională a României
- **BVB**: Bursa de Valori București
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity (heteroscedasticitate condiționată autoregresivă generalizată)
- **HICP**: Harmonised Index of Consumer Prices (indicele armonizat al prețurilor de consum)
- **IAPC**: Indicele armonizat al prețurilor de consum
- **MA**: Moving Average (medie mobilă)
- **PACF**: Partial AutoCorrelation Function (funcția de autocorelație parțială)
- **PIB**: Produsul intern brut
- **SARIMA**: Seasonal ARIMA (model) — model ARIMA sezonier
- **UE**: Uniunea Europeană
- **VAR**: Vector AutoRegression (model vector autoregresiv)

Întrebarea de azi și traseul

- **Întrebarea**: observăm o singură traiectorie a unei serii, o singură istorie; ce putem afla din ea despre mecanismul care a generat-o?
  - răspunsul se sprijină pe două ipoteze: **staționaritatea** și **ergodicitatea**
  - instrumentul principal: **funcția de autocorelație** (ACF)
- **Traseul** capitolului
  - procese stochastice și momentele lor; staționaritate strictă și slabă
  - zgomotul alb și mersul aleator; operatorul lag, diferențierea, descompunerea Wold, ergodicitatea
  - ACF și PACF de selecție cu benzi de încredere; testele Box–Pierce și Ljung–Box
  - transformări (logaritm, diferențiere, Box–Cox) și o primă analiză a unor serii reale
- Seminarul 1 are loc înaintea acestui curs: secțiunea „Noțiuni necesare azi” dă definițiile; aici le deducem și le explicăm

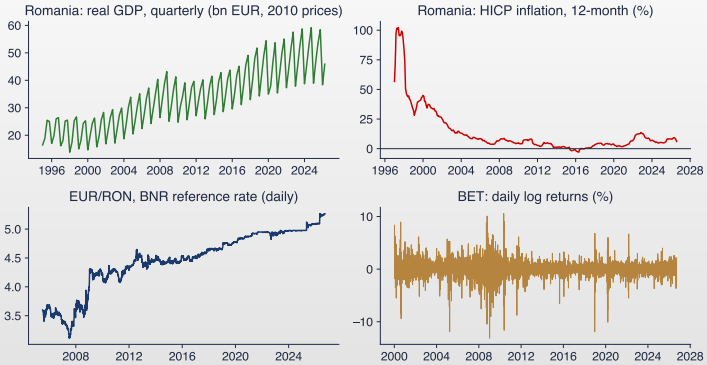
Rezultatele învățării

- Definiți un proces stochastic și funcțiile lui de medie, autocovarianță și autocorelație
- Deosebiți staționaritatea strictă de cea slabă și verificați staționaritatea slabă pentru procese simple
- Deduceți momentele zgomotului alb, ale unei medii mobile și ale unui mers aleator
- Calculați și interpretați ACF și PACF de selecție cu benzile $\pm 1{,}96/\sqrt{T}$
- Testați autocorelația cu statisticile Box–Pierce și Ljung–Box
- Alegeți o transformare (logaritm, diferență, Box–Cox) care apropie o serie reală de staționaritate

Bibliografie și instrumente

- Manual: Huang și Petukhina (2022), cap. 1–2 (concepte, analiză exploratorie)
  - manual însoțitor, gratuit online: Hyndman și Athanasopoulos (2021), cap. 2 (grafice pentru serii de timp) și cap. 3 (transformări)
- Teorie: Brockwell și Davis (2016), cap. 1; Hamilton (1994), cap. 3
- Quantlet-urile Python ale capitolului: Quantlets/Ch_01
  - ACF, PACF și teste portmanteau cu `statsmodels`
- Notebook-ul cursului: deschideți în Google Colab
- Curs video: Applied Time Series Analysis with Python

Patru serii ale acestui curs



- PIB-ul real al României, T1 1995–T2 2026 (Eurostat); inflația IAPC (Eurostat); cursul EUR/RON, cursul de referință BNR; randamentele logaritmice zilnice ale BET din 2000
- Patru frecvențe (trimestrială, lunară, zilnică) și patru tipuri de comportament

TSA_ch1_real_series

Interpretarea celor patru serii

- **PIB**: de la 16,5 la 45,9 mld. EUR pe trimestru; un trend și o sezonalitate a cărei amplitudine crește odată cu nivelul
  - media seriei se schimbă în timp
- **Inflația**: 102% în iunie 1997, −3,0% în mai 2016, 6,1% în august 2026
  - o mișcare lentă, persistentă: luna aceasta seamănă cu luna trecută
- **EUR/RON**: de la 3,60 la 5,26 lei; evoluează fără să revină la un nivel fix
  - un candidat pentru un mers aleator (secțiunea 3)
- **Randamentele BET**: oscilează în jurul unui nivel constant, apropiat de 0; extremele: −13,1% pe 7 ianuarie 2009, +10,6% pe 10 mai 2010
  - dar amplitudinea se schimbă: abaterea standard este 3,62% în septembrie 2008–martie 2009 și 0,66% în 2017

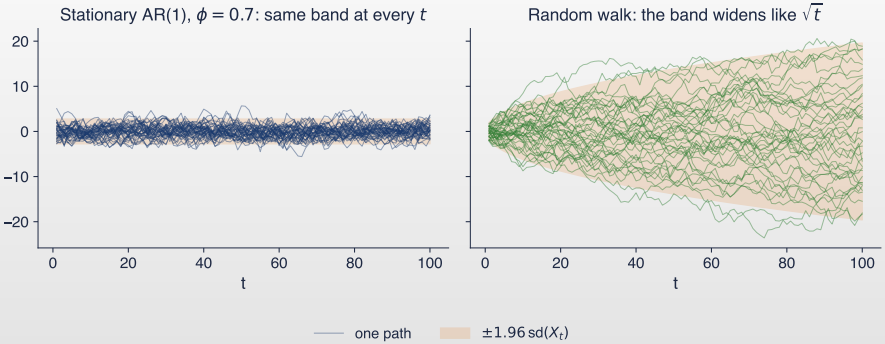
Seria de timp ca o realizare

- **Serie de timp**: observații $x_1, x_2, \ldots, x_T$ ordonate în timp, la intervale egale
  - T = volumul eșantionului; t = indicele de timp (zi, lună, trimestru)
- Ideea de bază: fiecare x_t este valoarea observată a unei **variabile aleatoare** X_t
  - randamentul BET de mîine este necunoscut azi: are o distribuție de valori posibile
  - istoria s-a desfășurat o singură dată: vedem cîte o singură valoare din fiecare distribuție
- Consecința: avem nevoie de ipoteze care leagă distribuțiile de la date diferite
  - altfel, T observații ar estima T medii diferite

Procesul stochastic

- **Definiție**: un **proces stochastic** este o familie de variabile aleatoare $\{X_t : t \in \mathcal{T}\}$ definite pe același spațiu de probabilitate $(\Omega, \mathcal{F}, P)$
  - $\mathcal{T} = \mathbb{Z}$ sau $\{1, \ldots, T\}$: **timp discret**, cazul acestui curs
  - Ω: mulțimea „istoriilor" posibile ω; P: probabilitățile lor
- Două moduri de a privi $X_t(\omega)$
  - fixăm t: X_t este o variabilă aleatoare (toate valorile posibile la data t)
  - fixăm ω: $t \mapsto X_t(\omega)$ este o **traiectorie** (realizare): ceea ce observăm
- Procesul este descris de **distribuțiile finit-dimensionale**: distribuția comună a lui $(X_{t_1}, \ldots, X_{t_k})$ pentru orice date $t_1, \ldots, t_k$

Un ansamblu de traiectorii



- Stînga: 40 de traiectorii ale lui $X_t = 0{,}7X_{t-\mathbf{1}} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$ i.i.d. (independente și identic distribuite); dreapta: 40 de traiectorii ale lui $X_t = X_{t-\mathbf{1}} + \varepsilon_t$, $X_{\mathbf{0}} = 0$
- Banda colorată: $\pm 1{,}96$ abateri standard ale lui X_t la fiecare t

TSA_ch1_processes

Interpretarea ansamblului de traiectorii

- O secțiune verticală la data t arată **distribuția lui** X_t; o traiectorie arată **o singură istorie**
  - momentele procesului (media, varianța) sînt medii pe traiectorii, la t fixat
- Stînga: aceeași bandă la orice t: $\mathrm{Var}(X_t) = 1/(1-0{,}7^2) = 1{,}96$, banda $\pm 2{,}74$
  - distribuția lui X_t nu depinde de t: o primă imagine a **staționarității**
- Dreapta: banda se lărgește; pe 5000 de traiectorii simulate, $\mathrm{Var}(X_{100}) = 100{,}8$ (teoretic: 100)
  - distribuția se schimbă cu t: mersul aleator **nu este staționar**

Funcțiile de medie, autocovarianță și autocorelație

- Pentru un proces cu $E[X_t^2] < \infty$ pentru orice t:
  - **funcția de medie**: $\mu_t = E[X_t]$
  - **funcția de autocovarianță**: $\gamma(t,s) = \mathrm{Cov}(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)]$
  - **funcția de autocorelație**: $\rho(t,s) = \gamma(t,s)/\sqrt{\gamma(t,t)\gamma(s,s)}$
- „Auto": covarianța seriei cu ea însăși la două date
  - $\gamma(t,t) = \mathrm{Var}(X_t)$; $\rho(t,t) = 1$; $|\rho(t,s)| \leq 1$ (inegalitatea Cauchy–Schwarz)
- **Exemplu**: $X_t = a + bt + \varepsilon_t$, ε_t i.i.d. cu media 0 și varianța σ^2
  - $\mu_t = a + bt$ depinde de t; $\gamma(t,t) = \sigma^2$, $\gamma(t,s) = 0$ pentru $t \neq s$

Recapitulare: serii de timp și procese stochastice

- O serie de timp este o traiectorie a unui proces stochastic $\{X_t\}$
- Momentele sînt medii pe traiectorii, la date fixate: μ_t, $\gamma(t,s)$, $\rho(t,s)$
- Cu o singură traiectorie, trebuie să presupunem că aceste funcții nu se schimbă în timp

Staționaritatea strictă

- **Definiție**: $\{X_t\}$ este **strict staționar** dacă, pentru orice $k \geq 1$, orice date $t_1, \ldots, t_k$ și orice deplasare h:
  - $(X_{t_1}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_k+h})$, unde $\stackrel{d}{=}$ înseamnă distribuții comune egale
- Consecințe
  - $k = 1$: toate variabilele X_t au aceeași distribuție
  - $k = 2$: distribuția comună a lui (X_t, X_{t+h}) depinde doar de lagul h
- Exemplu: un șir i.i.d. este strict staționar
  - definiția privește distribuții întregi; momentele nici nu trebuie să existe (un șir i.i.d. Cauchy)
- Greu de verificat din date: ar trebui cunoscute toate distribuțiile comune

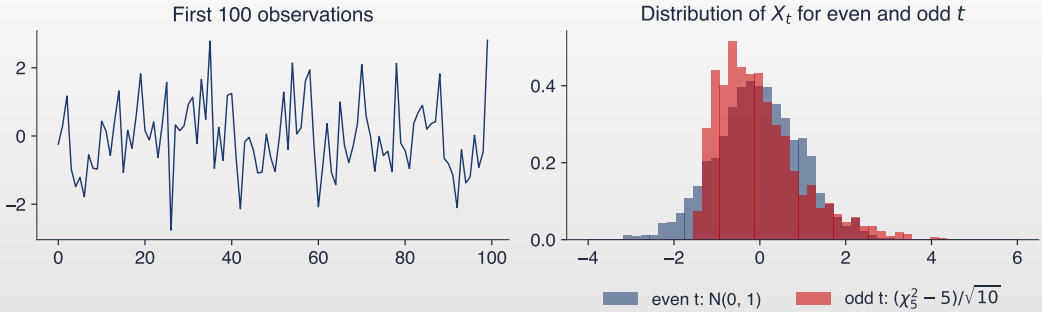
Staționaritatea slabă

- **Definiție** Khintchine (1934): $\{X_t\}$ este **slab staționar** (staționar în covarianță, staționar de ordinul doi) dacă
  - (i) $E[X_t^2] < \infty$ pentru orice t
  - (ii) $E[X_t] = \mu$, aceeași pentru orice t
  - (iii) $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ depinde doar de lagul h, nu de t
- Atunci funcțiile au un singur argument, **lagul** h:
  - $\gamma(h) = \text{Cov}(X_t, X_{t+h})$, $\gamma(0) = \text{Var}(X_t)$
  - $\rho(h) = \gamma(h)/\gamma(0)$: **ACF** (funcția de autocorelație)
- În acest curs, „staționar" înseamnă slab staționar, dacă nu se precizează altfel

Staționaritatea strictă și cea slabă

- **Strictă + varianță finită $\Rightarrow$ slabă**
  - distribuții egale dau medii egale; distribuții comune egale ale lui (X_t, X_{t+h}) dau covariațe egale
- **Slabă $\not\Rightarrow$ strictă**
  - staționaritatea slabă fixează doar primele două momente; forma distribuției se poate schimba (slide-ul următor)
- **Strictă $\not\Rightarrow$ slabă** dacă varianța este infinită
  - șir i.i.d. Student-t cu 2 grade de libertate: strict staționar, fără varianță finită
- **Procese gaussiene**: toate distribuțiile finit-dimensionale sînt Normale multivariate
  - o distribuție Normală este determinată de medii și covariațe, deci pentru ele slabă $\Leftrightarrow$ strictă

Slab, dar nu strict staționar



- X_t independente: $N(0,1)$ pentru t par, $(\chi^2_5 - 5)/\sqrt{10}$ pentru t impar; ambele au media 0 și varianța 1
- Eșantion de 4000: varianța 0,99 (par) și 1,00 (impar); asimetria $-0{,}02$ și 1,30 (teoretic: 0 și 1,26)

TSA_ch1_processes

Interpretarea contraexemplului

- Slab staționar: $E[X_t] = 0$, $\gamma(0) = 1$, $\gamma(h) = 0$ pentru $h \neq 0$, la orice t
  - ACF nu poate deosebi acest proces de zgomotul alb gaussian
- Nu este strict staționar: distribuția lui X_t alternează între una simetrică și una asimetrică la dreapta
  - $P(X_t < -1{,}6)$: circa 5% pentru t par, 0 pentru t impar ($X_t \geq -5/\sqrt{10} = -1{,}58$)
- Lecția: momentele de ordinul doi rezumă dependența, dar nu întreaga distribuție; riscul din cozi cere mai mult (Capitolul 5)

Proprietățile funcției de autocovarianță

- Pentru un proces slab staționar:
  - **simetrie**: $\gamma(-h) = \gamma(h)$, deoarece $\text{Cov}(X_t, X_{t-h}) = \text{Cov}(X_{t-h}, X_t)$
  - **mărginire**: $|\gamma(h)| \leq \gamma(0)$, deci $|\rho(h)| \leq 1$
  - **pozitiv semidefinire**: $\sum_{i,j} a_i a_j \gamma(i-j) \geq 0$ pentru orice numere reale $a_1, \ldots, a_n$
- Demonstrația celei de-a treia proprietăți
  - $0 \leq \text{Var}\left(\sum_i a_i X_i\right) = \sum_{i,j} a_i a_j \text{Cov}(X_i, X_j) = \sum_{i,j} a_i a_j \gamma(i-j)$
- Consecință: nu orice șir este o ACF
  - $\rho(1) = 0{,}9$, $\rho(2) = -0{,}9$ este imposibil: cu $a = (1, -1, 1)$,
    $\text{Var}(X_1 - X_2 + X_3) = \gamma(0)(3 - 4\rho(1) + 2\rho(2)) = -2{,}4\,\gamma(0) < 0$

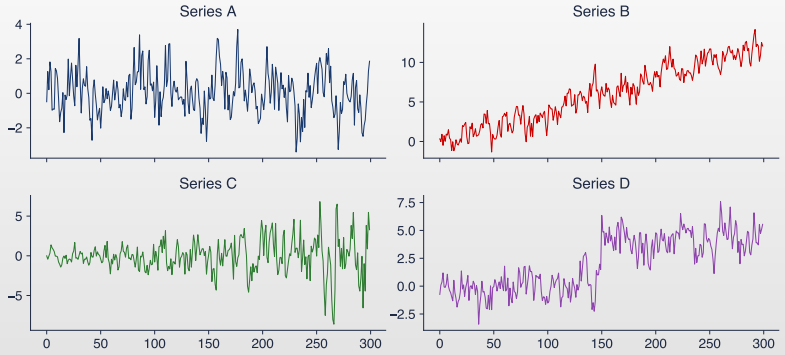
Exemplu rezolvat: o medie mobilă de ordinul 1

- $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, $\{\varepsilon_t\}$ necorelate, media 0, varianța σ^2: procesul **MA(1)** (medie mobilă)
  - θ: ponderea șocului din perioada anterioară; ε_t: șocul din perioada t
  - media: $E[X_t] = 0$
  - $\gamma(0) = E[(\varepsilon_t + \theta\varepsilon_{t-1})^2] = \sigma^2(1+\theta^2)$ (termenul încrucișat are media 0)
  - $\gamma(1) = E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t+1} + \theta\varepsilon_t)] = \theta\sigma^2$
  - $\gamma(h) = 0$ pentru $|h| \geq 2$: nu există niciun șoc comun
- Niciuna nu depinde de t: MA(1) este slab staționar pentru orice θ
  - $\rho(1) = \theta/(1+\theta^2)$; cu $\theta = 0{,}6$, $\sigma^2 = 1$: $\gamma(0) = 1{,}36$, $\rho(1) = 0{,}441$
  - $|\rho(1)| \leq 0{,}5$ pentru orice θ (maximul la $\theta = 1$)

Exemplu rezolvat: o autoregresie de ordinul 1

- $X_t = \phi X_{t-1} + \varepsilon_t$, $|\phi| < 1$: procesul **AR(1)** (autoregresiv); presupunem că este staționar și aflăm momentele
  - ϕ: ponderea valorii din perioada anterioară; ε_t: șocuri necorelate, cu media 0 și varianța σ^2
  - media: $\mu = \phi\mu + 0$, deci $\mu = 0$
  - varianța: $\gamma(0) = \phi^2\gamma(0) + \sigma^2$, deci $\gamma(0) = \sigma^2/(1-\phi^2)$
  - înmulțim cu X_{t-h} și aplicăm media: $\gamma(h) = \phi\gamma(h-1)$, deci $\rho(h) = \phi^{|h|}$
- Valori: $\phi = 0{,}8$, $\sigma^2 = 1$
  - $\gamma(0) = 2{,}78$; $\rho(1) = 0{,}8$, $\rho(5) = 0{,}328$, $\rho(10) = 0{,}107$: descreștere geometrică
- Cu $|\phi| \geq 1$ formula varianței nu mai funcționează: nu există o soluție staționară de această formă (Capitolele 2–3)

Întrebare pentru sală: care serii sînt staționare?



- Patru serii simulate, cu cîte 300 de observații; la baza tuturor stă același tip de zgomot AR(1) cu $\phi = 0{,}5$
- Care dintre condițiile (ii) și (iii) ale staționarității slabe nu este îndeplinită pentru fiecare serie?

TSA_ch1_processes

Răspuns: doar seria A este staționară

- **A**: AR(1) staționar: media și varianța constante, ACF depinde doar de lag
- **B**: trend determinist, $\mu_t = 0{,}04\,t$: condiția (ii) nu este îndeplinită
  - o serie **staționară în jurul trendului**: staționară în jurul unei drepte, după eliminarea trendului
- **C**: abaterea standard crește de 7,5 ori de-a lungul eșantionului: condiția (iii) nu este îndeplinită la $h = 0$
- **D**: un salt de nivel de $+4$ la $t = 150$: condiția (ii) nu este îndeplinită
  - o ruptură structurală; în date reale: o schimbare de regim sau de politică (Nilul, secțiunea 8)
- În practică, graficul seriei este prima verificare; testele formale apar în Capitolul 3

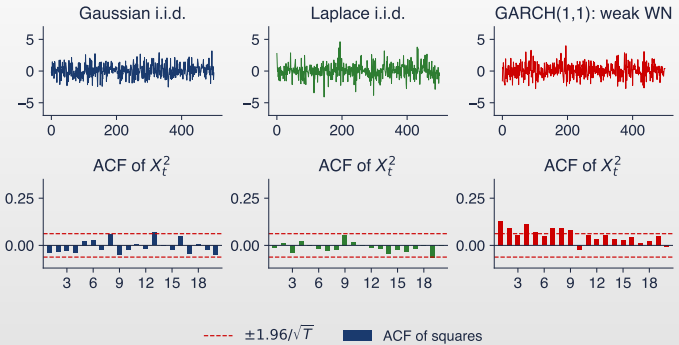
Recapitulare: staționaritate

- Strictă: toate distribuțiile comune sînt invariante la deplasare; slabă: media constantă, γ depinde doar de lag
- Strictă + varianță finită $\Rightarrow$ slabă; pentru procesele gaussiene coincid
- MA(1) este întotdeauna staționar; AR(1) este staționar pentru $|\phi| < 1$, cu $\rho(h) = \phi^{|h|}$
- Trendurile, varianța variabilă și rupturile încalcă staționaritatea

Zgomotul alb

- **Definiție**: $\{\varepsilon_t\}$ este **zgomot alb**, $\varepsilon_t \sim \text{WN}(0, \sigma^2)$, dacă
  - $E[\varepsilon_t] = 0$, $\text{Var}(\varepsilon_t) = \sigma^2$ și $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ pentru $t \neq s$
  - deci $\rho(0) = 1$ și $\rho(h) = 0$ pentru orice $h \neq 0$
- Trei variante, de la cea mai slabă la cea mai tare
  - zgomot alb **slab**: doar necorelat; dependența neliniară este permisă (de exemplu, în pătrate)
  - zgomot alb **i.i.d.** (tare): independent și identic distribuit
  - zgomot alb **gaussian**: i.i.d. $N(0, \sigma^2)$; pentru variabile comun Normale, necorelat înseamnă independent
- Zgomotul alb este piesa de bază a tuturor modelelor din acest curs: ARMA, ARIMA, GARCH, VAR

Trei tipuri de zgomot alb



- Sus: 500 din cele 1000 de valori simulate, varianța 1 în toate trei; jos: ACF a pătratelor X_t^2
- GARCH(1,1) (Capitolul 5): $\sigma_t^2 = 0{,}05 + 0{,}10\,\varepsilon_{t-1}^2 + 0{,}85\,\sigma_{t-1}^2$, $\varepsilon_t = \sigma_t z_t$; σ_t^2: varianța din ziua t, condiționată de trecut; z_t: i.i.d. $N(0,1)$

TSA_ch1_processes

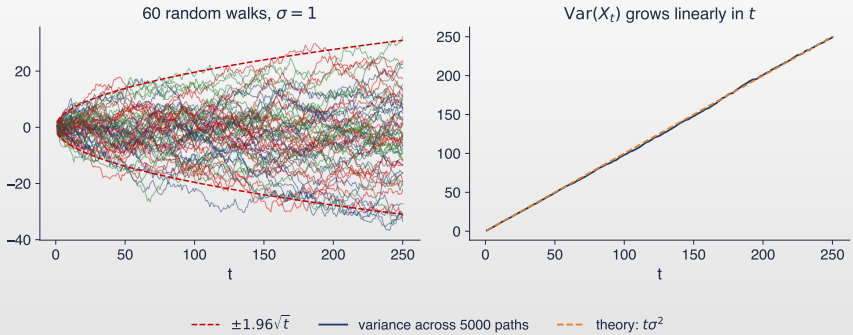
Interpretarea celor trei zgomote albe

- Toate trei sînt zgomot alb prin construcție; testul Ljung–Box $Q^*(10)$ (secțiunea 6) pentru X_t: p = 0,278 (gaussian), 0,986 (Laplace), 0,049 (GARCH)
  - pentru seria GARCH, ipoteza este respinsă la 5%, deși seria este necorelată: cînd varianța se schimbă, testul obișnuit respinge prea des
- Laplace: cozi mai groase (kurtosis-ul 5,2, față de 3,0), dar tot i.i.d.: pătratele sînt necorelate (p = 0,680)
  - cozile groase singure nu creează dependență
- GARCH: pătratele sînt puternic corelate, $Q^*(10) = 73{,}2$, $p < 0{,}001$
  - un zgomot alb slab: semnul și nivelul nu pot fi anticipate, mărimea poate fi (volatility clustering)
  - randamentele zilnice se comportă așa (BET, secțiunea 5)

Mersul aleator

- **Definiție**: $X_t = X_{t-1} + \varepsilon_t$, $t \geq 1$, $X_0 = 0$, $\varepsilon_t \sim \text{WN}(0, \sigma^2)$; prin substituție, $X_t = \sum_{i=1}^{t} \varepsilon_i$
  - fiecare șoc rămîne pentru totdeauna în nivel: $\partial X_{t+h}/\partial \varepsilon_t = 1$ pentru orice $h \geq 0$
- Momentele (sînt suficiente șocuri necorelate)
  - $E[X_t] = 0$; $\text{Var}(X_t) = \sum_{i=1}^{t} \text{Var}(\varepsilon_i) = t\sigma^2$
  - pentru $t \leq s$: $\text{Cov}(X_t, X_s) = \text{Cov}\big(\sum_{i \leq t} \varepsilon_i, \sum_{j \leq s} \varepsilon_j\big) = t\sigma^2$
  - $\rho(t, s) = t/\sqrt{ts} = \sqrt{t/s}$
- **Nestaționar**: varianța crește cu t, iar corelația depinde de t, nu doar de $s - t$
  - exemplu: $\rho(X_{100}, X_{110}) = \sqrt{100/110} = 0{,}953$, dar $\rho(X_{10}, X_{20}) = \sqrt{0{,}5} = 0{,}707$

Mersuri aleatoare și varianța lor



- Stînga: 60 de traiectorii cu $\sigma = 1$ și banda $\pm 1{,}96\sqrt{t}$; dreapta: varianța lui X_t pe 5000 de traiectorii și dreapta $t\sigma^2$
- La $t = 250$: varianța simulată 249,1 (teoretic: 250); proporția traiectoriilor în afara benzii: 5,0%

TSA_ch1_processes

Interpretarea mersurilor aleatoare

- Dispersia crește ca $\sqrt{t}$: incertitudinea despre nivel se acumulează
  - o prognoză a nivelului peste 100 de zile are varianța $100\sigma^2$
- Traiectoriile par să aibă trenduri și cicluri, deși sînt generate exclusiv de șocuri aleatoare
  - aceeași lecție ca la Slutzky (1937): sumele de șocuri aleatoare pot semăna cu ciclurile economice
- Valoarea simulată $\rho(X_{100}, X_{110}) = 0{,}953$, față de 0,953 teoretic
  - nivelurile depărtate în timp rămîn puternic corelate: o ACF care scade foarte încet

Mersul aleator cu derivă

- $X_t = c + X_{t-1} + \varepsilon_t$, $X_0 = 0$: $X_t = ct + \sum_{i=1}^{t} \varepsilon_i$; c este **deriva**
  - $E[X_t] = ct$: un trend liniar în medie; $\text{Var}(X_t) = t\sigma^2$, ca înainte
- Prima diferență este staționară: $\Delta X_t = X_t - X_{t-1} = c + \varepsilon_t$
  - o serie **staționară în diferențe**, cu **trend stochastic**
- Comparație cu seria staționară în jurul trendului $Y_t = a + bt + u_t$, u_t staționar; a: termenul liber, b: panta trendului
  - ambele au o medie liniară; în Y_t șocurile se sting, în X_t rămîn pentru totdeauna
  - deosebirea celor două este problema rădăcinii unitare din Capitolul 3
- Logaritmii prețurilor acțiunilor și ai indicilor sînt adesea apropiați de un mers aleator cu o derivă mică și pozitivă

Bursa de Valori București

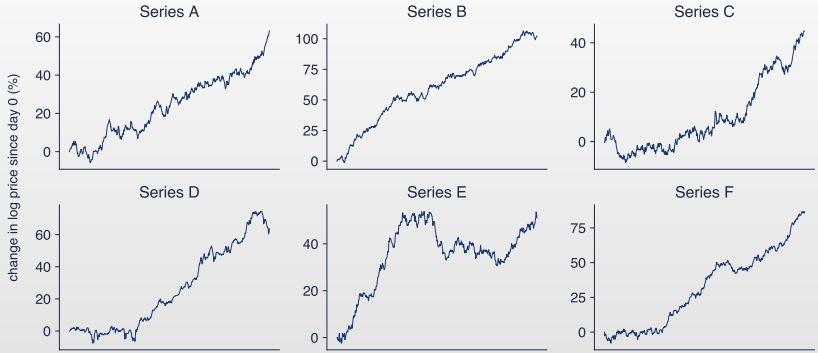


Bursa de Valori București (BVB), 2024

Foto: Corina Chitu (2024); CC BY-SA 4.0; Wikimedia Commons

- Indicele BET: indicele de referință al BVB (Bursa de Valori București), calculat din 19 septembrie 1997
- Închideri zilnice din 2000: 6 670 de randamente logaritmice $r_t = 100\,\Delta \ln P_t$ (P_t: nivelul de închidere din ziua t); abaterea standard 1,40% pe zi
- Este logaritmul prețului BET un mers aleator? Un prim test, vizual, pe slide-ul următor

Întrebare pentru sală: care serie este cea reală?



- Un panou: logaritmul prețului BET în ultimele 500 de zile de tranzacționare; cinci panouri: mersuri aleatoare cu aceeași derivă și aceeași abatere standard zilnică
- Care panou este BET?

TSA_ch1_processes

Răspuns: seria D

- Seria D este BET, 11 septembrie 2024–18 septembrie 2026: variația zilnică medie 0,13%, abaterea standard 0,99%
  - cu ochiul liber, nu poate fi deosebită de mersurile aleatoare
- Datele nu sînt exact un mers aleator: autocorelația de ordinul 1 a variațiilor zilnice este 0,12; p-value-ul testului Ljung–Box $Q^*(10)$ este 0,018
  - abaterile mici, invizibile în nivel, apar în ACF a diferențelor
- Lecția: studiem variațiile, nu nivelul; instrumentele sînt ACF și testele portmanteau (secțiunile 5–6)

Recapitulare: zgomotul alb și mersul aleator

- Zgomotul alb: media 0, varianța constantă, fără autocorelație; slab, i.i.d. sau gaussian
- Necorelat nu înseamnă independent: pătratele unui zgomot alb slab pot fi corelate
- Mersul aleator: $\mathrm{Var}(X_t) = t\sigma^2$, $\rho(t,s) = \sqrt{t/s}$: nestaționar; diferența lui este zgomot alb
- Seriile staționare în jurul trendului și cele staționare în diferențe arată la fel, dar reacționează diferit la șocuri

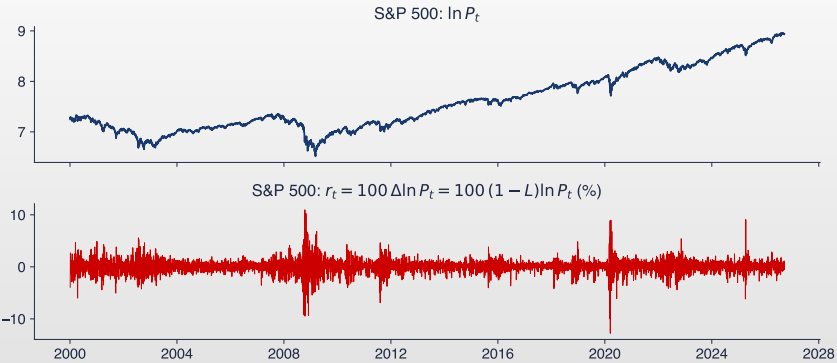
Operatorul lag

- **Definiție**: $LX_t = X_{t-1}$; puteri: $L^k X_t = X_{t-k}$, $L^0 = 1$ (notat și B, de la *backshift*)
  - o constantă nu se schimbă: $Lc = c$
- **Polinoame în lag**: $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$
  - p: gradul (lagul cel mai mare); $\phi_1, \ldots, \phi_p$: coeficienți reali; $\phi(L)X_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}$
  - AR(1): $(1 - \phi L)X_t = \varepsilon_t$; MA(1): $X_t = (1 + \theta L)\varepsilon_t$
  - se înmulțesc ca polinoamele obișnuite: $(1 - L)(1 + L) = 1 - L^2$
- **Inversare**: pentru $|\phi| < 1$, $(1 - \phi L)^{-1} = 1 + \phi L + \phi^2 L^2 + \ldots$ (o serie geometrică)
  - deci AR(1) se scrie $X_t = \sum_{j \geq 0} \phi^j \varepsilon_{t-j}$: o medie mobilă infinită
- Limbajul Capitolelor 2–4: modelele ARMA, ARIMA și SARIMA se scriu cu polinoame în lag

Diferențierea

- **Prima diferență**: $\Delta X_t = (1-L)X_t = X_t - X_{t-1}$; **a doua**: $\Delta^2 X_t = (1-L)^2 X_t = X_t - 2X_{t-1} + X_{t-2}$
  - **diferența sezonieră** cu perioada s: $\Delta_s X_t = (1-L^s)X_t = X_t - X_{t-s}$ ($s = 4$ pentru trimestre, $s = 12$ pentru luni)
- **Exemplu rezolvat**: $x = (10, 12, 15, 14, 18)$
  - $\Delta x = (2, 3, -1, 4)$; $\Delta^2 x = (1, -4, 5)$; fiecare diferențiere pierde o observație
- Efectele diferențierii
  - un trend liniar: $\Delta(a + bt) = b$; un trend pătratic necesită Δ^2
  - un mers aleator: $\Delta X_t = \varepsilon_t$; un tipar sezonier care se repetă exact: Δ_s
- **Proces integrat**: $X_t \sim I(d)$ dacă $\Delta^d X_t$ este staționar, iar $\Delta^{d-1} X_t$ nu este
  - zgomotul alb este $I(0)$, mersul aleator este $I(1)$; testarea lui d: Capitolul 3

S&P 500: nivelul și diferența lui



- Sus: logaritmul indicelui S&P 500 din 2000; jos: randamentele logaritmice zilnice, 6 714 zile pînă la 18 septembrie 2026

TSA_ch1_transformations

Interpretarea transformării S&P 500

- Logaritmul prețului: ACF de selecție 0,999 la lagul 1 și încă 0,973 la lagul 50
  - comportamentul unui mers aleator: nivelul rămîne corelat cu întregul său trecut
- Randamentele logaritmice: media 0,025% pe zi, abaterea standard 1,21%; ACF la lagul 1: $-0{,}099$
  - un nivel constant: o singură diferențiere a eliminat trendul stochastic
- Nu încă zgomot alb în sens strict: perioadele liniștite alternează cu cele agitate (2008, 2020)
  - randamentele sînt apropiate de un zgomot alb slab; varianța lor este subiectul Capitolului 5

Descompunerea Wold

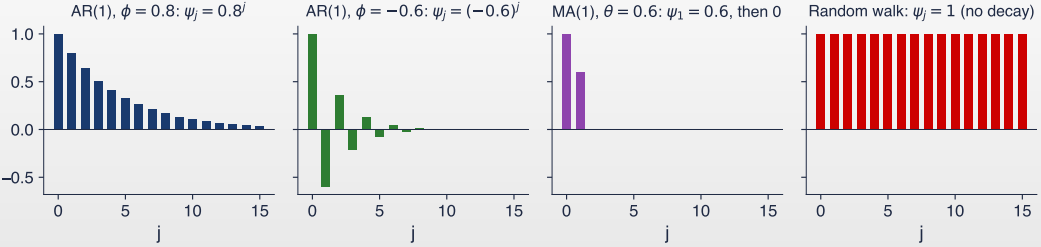
- **Teoremă** Wold (1938): orice proces slab staționar cu media 0 se poate scrie
  - $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \eta_t$, cu $\psi_0 = 1$, $\sum_j \psi_j^2 < \infty$
  - ε_t: zgomot alb, eroarea celei mai bune prognoze liniare a lui X_t din trecutul său (**inovația**)
  - η_t: o parte deterministă, perfect previzibilă din trecutul îndepărtat (de exemplu, un ciclu fix)
- Semnificația
  - orice serie staționară este o sumă ponderată de șocuri trecute: MA(∞)
  - ponderile ψ_j = efectul unui șoc după j perioade (răspunsul la impuls)
  - modelele ARMA (Capitolul 2) aproximează șirul infinit ψ_j cu puțini parametri



Herman Wold, Uppsala, 1969

Foto: Uppsala-Bild (1969); CC BY 4.0; Wikimedia Commons

Ponderile Wold ale a patru procese



- ψ_j, $j = 0, \ldots, 15$: AR(1) cu $\phi = 0{,}8$ și $\phi = -0{,}6$, MA(1) cu $\theta = 0{,}6$ și mersul aleator
- AR(1), $\phi = 0{,}8$: $\psi_5 = 0{,}328$, $\psi_{10} = 0{,}107$; $\mathrm{Var}(X_t) = \sigma^2 \sum_j \psi_j^2 = 2{,}78\,\sigma^2$

TSA_ch1_wold_ergodicity

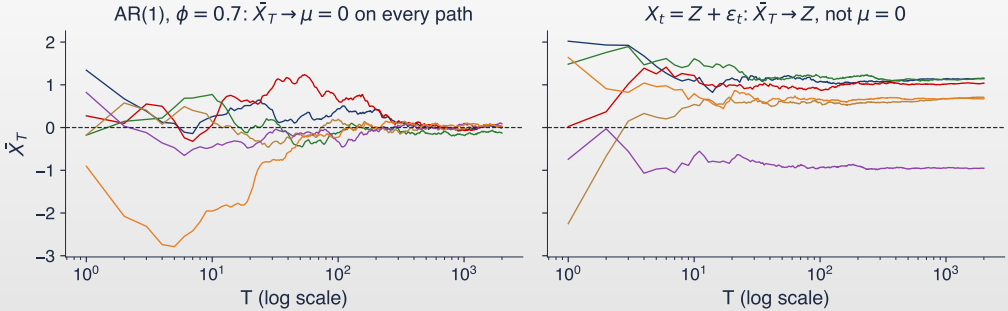
Interpretarea ponderilor Wold

- Procesele staționare: $\psi_j \to 0$, deci șocurile se sting, iar varianța $\sigma^2 \sum_j \psi_j^2$ este finită
  - $\phi < 0$: efectul își schimbă semnul alternativ; MA(1): efectul se oprește după o perioadă
- Mersul aleator: $\psi_j = 1$ pentru orice j, $\sum_j \psi_j^2 = \infty$: fără formă Wold, fără staționaritate
  - sensul economic: un șoc permanent (un șoc tehnologic asupra PIB-ului, o știre în preț)
- ACF este determinată de ponderi: $\gamma(h) = \sigma^2 \sum_j \psi_j \psi_{j+h}$

Ergodicitatea

- Staționaritatea spune că momentele sînt constante; mai trebuie să le estimăm dintr-o **singură** traiectorie
  - media pe ansamblu: $\mu = E[X_t]$ (pe traiectorii); media în timp: $\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t$ (de-a lungul unei traiectorii)
- **Definiție**: un proces staționar este **ergodic pentru medie** dacă $\bar{X}_T \to \mu$ (în medie pătratică) cînd $T \to \infty$
  - o condiție suficientă: $\gamma(h) \to 0$ cînd $h \to \infty$ (corelația dispare)
  - dacă $\sum_h |\gamma(h)| < \infty$, atunci $T\,\mathrm{Var}(\bar{X}_T) \to \sum_{h=-\infty}^{\infty} \gamma(h)$
- **Contraexemplu**: $X_t = Z + \varepsilon_t$, $Z \sim N(0,1)$ extras o singură dată, ε_t zgomot alb
  - staționar, cu $\gamma(h) = 1$ pentru orice $h \neq 0$; $\bar{X}_T \to Z$, nu $\mu = 0$

Mediile în timp: proces ergodic și neergodic



- Media cumulată $\bar{X}_T$ pe 6 traiectorii, T pînă la 2 000: AR(1) cu $\phi = 0{,}7$ (stînga) și $X_t = Z + \varepsilon_t$ (dreapta)
- Valorile finale: între $-0{,}13$ și $0{,}09$ în stînga; între $-0{,}95$ și $1{,}16$ în dreapta, aproape de Z extras

TSA_ch1_wold_ergodicity

Interpretarea ergodicității

- Stînga: fiecare traiectorie „învață" media adevărată 0; o istorie lungă este la fel de bună ca multe istorii
  - convergența este lentă cînd corelația este puternică: $\mathrm{Var}(\bar{X}_T) \approx \gamma(0)(1+\phi)/((1-\phi)T)$
- Dreapta: fiecare traiectorie converge la propriul Z; oricîte date dintr-o singură traiectorie nu dezvăluie $\mu = 0$
  - o economie cu un „tip" permanent și neobservat se comportă astfel
- Ergodicitatea nu poate fi testată dintr-o singură traiectorie; o presupunem ori de cîte ori estimăm momente dintr-o singură serie

Recapitulare: operatorul lag, diferențierea și descompunerea Wold

- $L^k X_t = X_{t-k}$; $\Delta = 1 - L$; $\Delta_s = 1 - L^s$; $I(d)$: staționar după d diferențieri
- Wold: un proces staționar este un MA(∞) al inovațiilor sale, cu $\sum \psi_j^2 < \infty$
- Ergodicitatea permite ca mediile în timp să estimeze momentele pe ansamblu; cere corelații care se sting

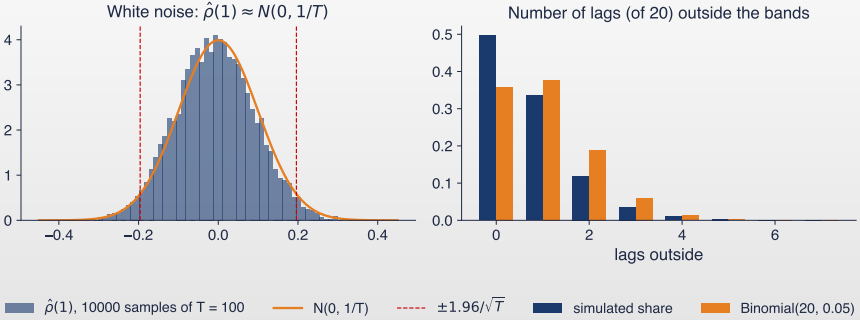
Autocovarianța și autocorelația de selecție

- Din $x_1, \ldots, x_T$, cu media de selecție $\bar{x} = \frac{1}{T}\sum_t x_t$:
  - $\hat{\gamma}(h) = \frac{1}{T}\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x})$, $0 \le h < T$
  - $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$: **ACF de selecție**; graficul lui $\hat{\rho}(h)$ în funcție de h este **corelograma**
- Împărțitorul T în locul lui $T - h$
  - șirul $\hat{\gamma}(h)$ este atunci pozitiv semidefinit, ca o autocovarianță adevărată Brockwell și Davis (2016)
  - prețul: o mică deplasare spre 0 la laguri mari; folosim $h \le T/4$
- **Exemplu rezolvat**: $x = (3, 5, 4, 6, 7)$, $\bar{x} = 5$, abaterile $(-2, 0, -1, 1, 2)$
  - $\hat{\gamma}(0) = 10/5 = 2{,}0$; $\hat{\gamma}(1) = (0 + 0 - 1 + 2)/5 = 0{,}2$; $\hat{\gamma}(2) = (2 + 0 - 2)/5 = 0{,}0$
  - $\hat{\rho}(1) = 0{,}10$, $\hat{\rho}(2) = 0{,}00$; cu $T = 5$ banda este $\pm 0{,}88$: mult prea puține date

Benzi de încredere pentru ACF de selecție

- **Rezultatul asimptotic** din Bartlett (1946): dacă X_t este i.i.d. cu varianță finită, pentru T mare
  - $\hat{\rho}(1), \ldots, \hat{\rho}(m)$ sînt aproximativ independente, $N(0, 1/T)$
  - banda de 95%: $\pm 1{,}96/\sqrt{T}$; o bară în afara ei respinge $\rho(h) = 0$ la 5%, **doar pentru acel lag**
- Pentru un MA(q), la laguri $h > q$: $\mathrm{Var}(\hat{\rho}(h)) \approx \frac{1}{T}\left(1 + 2\sum_{k=1}^{q} \rho(k)^2\right)$
  - benzile mai largi din `plot_acf` din `statsmodels` (Capitolul 2)
- Limitele benzii
  - cu 20 de laguri și nicio corelație, ne așteptăm ca aproximativ o bară să iasă din bandă din întîmplare
  - banda presupune date i.i.d.; pentru un zgomot alb slab (randamente GARCH) banda adevărată este mai largă

Distribuția de selecție a ACF



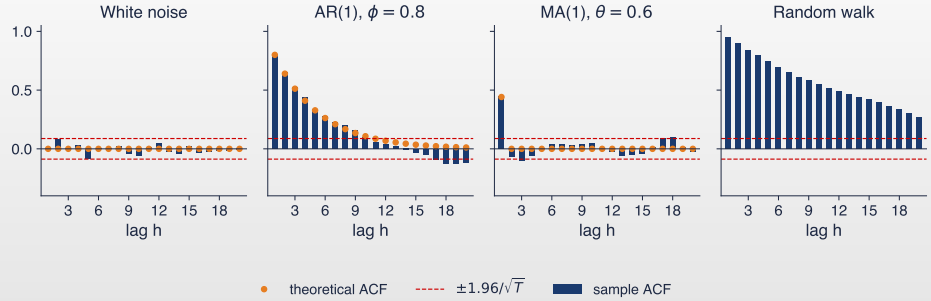
- 10000 de eșantioane de zgomot alb gaussian cu $T = 100$: $\hat{\rho}(1)$ (stînga) și numărul de laguri, din 20, în afara lui $\pm 0{,}196$ (dreapta)

TSA_ch1_acf_pacf

Interpretarea simulării

- $\hat{\rho}(1)$: media $-0{,}008$ (teoretic $-1/T = -0{,}01$), abaterea standard 0,098 (teoretic $1/\sqrt{T} = 0{,}1$)
  - aproximarea lui Bartlett este deja bună la $T = 100$
- Cel puțin una din 20 de bare iese din bandă în 50% dintre eșantioane, deși nu există nicio corelație
  - testele independente ar da $1 - 0{,}95^{20} = 64\%$; la laguri mari $\mathrm{Var}(\hat{\rho}(h)) < 1/T$, deci ies ceva mai puține bare
- Nu supra-interpretați bare izolate: judecați tiparul sau folosiți un test comun (secțiunea 6)

ACF teoretică și de selecție pentru patru procese



- $T = 500$ de valori simulate; bare: ACF de selecție; puncte: ACF teoretică; banda $\pm 0{,}088$

TSA_ch1_acf_pacf

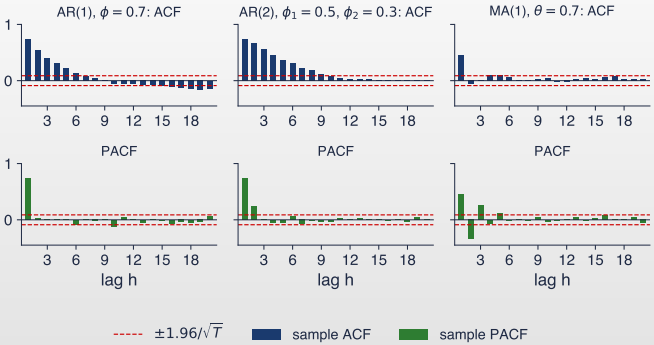
Interpretarea celor patru corelograme

- Zgomotul alb: toate barele sînt mici; cea mai mare, $\hat{\rho}(2) = 0{,}08$, este zgomot
- AR(1), $\phi = 0{,}8$: descreștere geometrică, $\hat{\rho}(1) = 0{,}81$, $\hat{\rho}(2) = 0{,}65$ (teoretic 0,80 și 0,64)
- MA(1), $\theta = 0{,}6$: un singur vîrf, $\hat{\rho}(1) = 0{,}44$ (teoretic 0,44), apoi nimic: ACF **se anulează** după lagul 1
- Mersul aleator: $\hat{\rho}(1) = 0{,}95$, încă 0,55 la lagul 10: descreșterea lentă, aproape liniară, a unei serii nestaționare
  - ACF de selecție a unui mers aleator nu are corespondent teoretic: $\rho(h)$ nu este definit

Autocorelația parțială

- **Definiție**: **PACF** (funcția de autocorelație parțială) la lagul h, ϕ_{hh}, este ultimul coeficient al celei mai bune prognoze liniare a lui X_t din $X_{t-1}, \ldots, X_{t-h}$:
  - $X_t = \phi_{h1} X_{t-1} + \cdots + \phi_{hh} X_{t-h} + e_t$; $\phi_{h1}, \ldots, \phi_{hh}$: coeficienții regresiei pe h laguri; e_t: eroarea de prognoză
  - corelația dintre X_t și X_{t-h} după eliminarea efectului liniar al lagurilor intermediare
- Primele două valori
  - $\phi_{11} = \rho(1)$; $\phi_{22} = \dfrac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}$
  - cazul general: recursia Durbin–Levinson; PACF de selecție folosește $\hat{\rho}(h)$ în locul lui $\rho(h)$
- **Exemplu rezolvat**
  - AR(1): $\rho(2) = \rho(1)^2$, deci $\phi_{22} = 0$: odată cunoscut X_{t-1}, X_{t-2} nu mai aduce nimic
  - AR(2) cu $\phi_1 = 0{,}5$, $\phi_2 = 0{,}3$: $\rho(1) = 0{,}714$, $\rho(2) = 0{,}657$, $\phi_{22} = (0{,}657 - 0{,}510)/(1 - 0{,}510) = 0{,}300 = \phi_2$
- Aceeași bandă ca pentru ACF, în cazul zgomotului alb: $\pm 1{,}96/\sqrt{T}$

ACF și PACF pentru trei procese



- $T = 500$ de valori simulate din AR(1) ($\phi = 0{,}7$), AR(2) ($\phi_1 = 0{,}5$, $\phi_2 = 0{,}3$) și MA(1) ($\theta = 0{,}7$); sus: ACF de selecție; jos: PACF de selecție

TSA_ch1_acf_pacf

Interpretarea ACF și PACF

- AR(1): ACF descrește; PACF are un singur vîrf, $\hat{\phi}_{11} = 0{,}73$, apoi $\hat{\phi}_{22} = 0{,}03$
  - AR(2): două vîrfuri în PACF, $\hat{\phi}_{22} = 0{,}24$ (teoretic 0,3), apoi $\hat{\phi}_{33} = 0{,}01$
- MA(1): imaginea în oglindă; un singur vîrf în ACF ($\hat{\rho}(1) = 0{,}46$), o PACF care alternează și descrește ($-0{,}34$, $0{,}26$, ...)
- Regula de identificare din Capitolul 2 Huang și Petukhina (2022):
  - PACF se anulează după p: AR(p); ACF se anulează după q: MA(q); ambele descresc: ARMA

Date reale: trei piețe

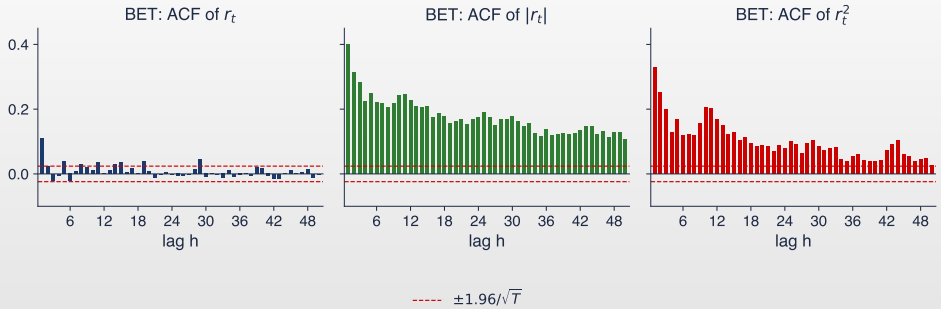


Banca Națională a României (BNR), care publică cursul de referință EUR/RON

Foto: Ștefan Jurcă (2015); CC BY 2.0; Wikimedia Commons

- Urmează: ACF a randamentelor logaritmice zilnice ale BET din 2000, a valorilor lor absolute și a pătratelor lor
- Aceeași analiză pentru EUR/RON (cursul de referință BNR) și S&P 500: Seminarul 1, cerința B2
- Întrebarea: sînt randamentele BET zgomot alb? De ce tip?

BET: ACF a randamentelor, a valorilor absolute și a pătratelor



- Randamente logaritmice zilnice, 6 670 de zile, lagurile 1–50; banda $\pm 0{,}024$

TSA_ch1_acf_pacf

Interpretarea corelogramelor BET

- Randamentele: $\hat{\rho}(1) = 0{,}109$, clar în afara benzii; $Q^*(10) = 108$
  - semnificativă statistic, dar mică: randamentul de ieri explică aproximativ $\hat{\rho}(1)^2 = 1{,}2\%$ din varianța celui de azi
  - un semn tipic al unei piețe mai puțin lichide (tranzacționare nesincronă)
- Valorile absolute: $\hat{\rho}(1) = 0{,}40$ și încă 0,11 la lagul 50; pătratele: $\hat{\rho}(1) = 0{,}33$
  - mișcările mari urmează după mișcări mari: volatility clustering, o memorie lungă a mărimii randamentelor
- Concluzie: randamentele BET sînt apropiate de un zgomot alb **slab**; media este greu de anticipat, varianța nu (Capitolul 5)

Recapitulare: ACF și PACF de selecție

- $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$, cu împărțitorul T; folosim laguri pînă la aproximativ $T/4$
- Pentru date i.i.d.: $\hat{\rho}(h) \approx N(0, 1/T)$, banda $\pm 1{,}96/\sqrt{T}$; o bară din 20 iese din întîmplare
- PACF: efectul direct al lagului h; AR(p) se anulează în PACF, MA(q) în ACF
- Randamentele: dependență liniară slabă, dependență neliniară puternică

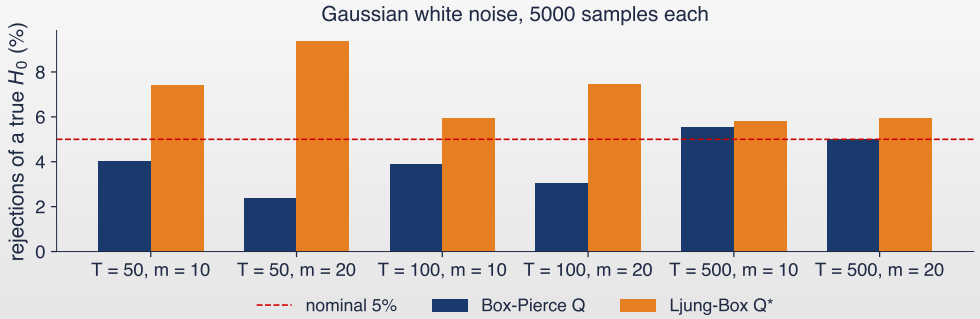
Testarea simultană a mai multor autocorelații

- H_0: $\rho(1) = \cdots = \rho(m) = 0$ (seria este zgomot alb pînă la lagul m); H_1: cel puțin un $\rho(h) \neq 0$
  - un test **portmanteau**: o singură statistică „poartă" m autocorelații
- **Statistica Box–Pierce**, din Box și Pierce (1970): $Q(m) = T \sum_{h=1}^{m} \hat{\rho}(h)^2$
  - în ipoteza H_0, fiecare $\sqrt{T}\hat{\rho}(h) \approx N(0,1)$, independente: suma a m pătrate este $\approx \chi^2(m)$
- **Statistica Ljung–Box**, din Ljung și Box (1978): $Q^*(m) = T(T+2) \sum_{h=1}^{m} \frac{\hat{\rho}(h)^2}{T-h}$
  - ponderea $(T+2)/(T-h)$ corectează varianța mică a lui $\hat{\rho}(h)$ la laguri mari; aceeași limită $\chi^2(m)$
- Respingem H_0 la 5% dacă $Q^* > \chi^2_{0,95}(m)$; de exemplu, $\chi^2_{0,95}(10) = 18{,}31$
  - $\chi^2(m)$: distribuția hi-pătrat cu m grade de libertate; $\chi^2_{0,95}(m)$: cuantila ei de 95%
  - p-value: probabilitatea, în ipoteza H_0, de a obține o statistică cel puțin la fel de mare ca cea observată; respingem dacă este sub 0,05
  - pentru reziduurile unui ARMA(p,q): $\chi^2(m-p-q)$ (Capitolul 2)

Exemplu rezolvat: statistica Ljung–Box

- $T = 100$; $\hat{\rho}(1) = 0{,}25$, $\hat{\rho}(2) = 0{,}12$, $\hat{\rho}(3) = -0{,}08$; banda $\pm 0{,}196$
  - doar $\hat{\rho}(1)$ este în afara benzii
- Box–Pierce: $Q(3) = 100\,(0{,}0625 + 0{,}0144 + 0{,}0064) = 8{,}33$
  - Ljung–Box: $Q^*(3) = 100 \cdot 102\left(\frac{0{,}0625}{99} + \frac{0{,}0144}{98} + \frac{0{,}0064}{97}\right) = 8{,}61$
- Valoarea critică $\chi^2_{0{,}95}(3) = 7{,}81$: respingem H_0; p-value-ul este 0,035
  - seria nu este zgomot alb; dependența este concentrată la lagul 1
- Alegerea lui m: prea mic ratează lagurile îndepărtate; prea mare diluează dovezile
  - alegeri uzuale: $m = 10$ pentru date nesezoniere, $m = 2s$ pentru date sezoniere cu perioada s Hyndman și Athanasopoulos (2021)

Box–Pierce și Ljung–Box în eșantioane mici



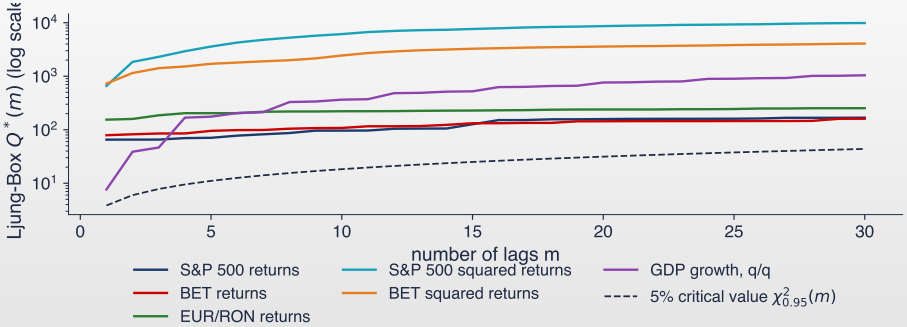
- Proporția din 5000 de eșantioane de zgomot alb gaussian în care fiecare test respinge o ipoteză H_0 adevărată, la nivelul de 5%

TSA_ch1_portmanteau

Interpretarea simulării mărimii testelor

- $T = 50$, $m = 20$: Box–Pierce respinge în 2,4% dintre eșantioane, Ljung–Box în 9,4%
  - Box–Pierce respinge prea rar, deci poate rata dependența; Ljung–Box respinge prea des cînd m este mare față de T
- $T = 500$, $m = 20$: 5,0% și 5,9%: ambele aproape de 5%
- În practică: folosim Ljung–Box (implicit în programe) și păstrăm m mult sub T (de exemplu, $m \leq T/5$)

Statisticile Ljung–Box pentru serii reale



- $Q^*(m)$ pentru $m = 1, \ldots, 30$ (scară logaritmică), față de valoarea critică de 5% $\chi^2_{0,95}(m)$ (linie întreruptă)

TSA_ch1_portmanteau

Interpretarea statisticilor Ljung–Box

Seria	T	$\hat{\rho}(1)$	$\hat{\rho}(4)$	$Q^*(10)$	p
S&P 500, randamente zilnice	6 714	−0,10	−0,03	96,3	$< 0{,}001$
S&P 500, randamente la pătrat	6 714	0,31	0,31	6132,8	$< 0{,}001$
BET, randamente zilnice	6 670	0,11	−0,01	108,0	$< 0{,}001$
BET, randamente la pătrat	6 670	0,33	0,13	2440,0	$< 0{,}001$
EUR/RON, randamente zilnice	5 352	0,17	−0,06	221,7	$< 0{,}001$
PIB România, creștere t/t	125	−0,24	0,96	364,8	$< 0{,}001$
PIB România, creștere an/an	122	0,63	0,03	75,0	$< 0{,}001$
IAPC România, inflația lunară	260	0,40	0,12	126,7	$< 0{,}001$
Nil, debitul anual	100	0,50	0,24	88,1	$< 0{,}001$
Pete solare, anual	309	0,82	−0,28	627,4	$< 0{,}001$

- Toate seriile resping ipoteza de zgomot alb; pentru pătratele randamentelor statistica este de zeci de ori mai mare decît pentru randamente
- Respingerea spune doar că există **o anumită** corelație; ACF arată unde (PIB: lagul 4, sezonul)

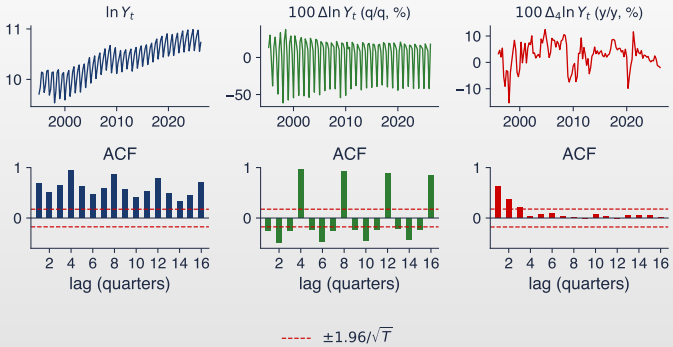
Recapitulare: teste portmanteau

- Box–Pierce $Q(m)$ și Ljung–Box $Q^*(m)$ însumează pătratele autocorelațiilor de selecție; ambele $\approx \chi^2(m)$ pentru zgomot alb
- Ljung–Box are o mărime mai bună în eșantioane mici; păstrăm m mult sub T
- O respingere este un semnal pentru modelare, nu un model; în prezența efectelor GARCH, testul randamentelor respinge prea des

Motivele transformării unei serii

- Scopuri
  - stabilizarea variației (oscilații sezoniere care cresc odată cu nivelul)
  - transformarea creșterii exponențiale într-un trend liniar
  - eliminarea unui trend stochastic sau a unui tipar sezonier (diferențiere)
- **Logaritmul**: $\ln Y_t$; diferența lui este rata de creștere în termeni continui
  - $100\,\Delta \ln Y_t \approx$ variația procentuală, pentru variații mici: $\ln 1{,}03 = 0{,}0296$, circa 3%
  - nu și pentru cele mari: $\ln 1{,}30 = 0{,}262$, $\ln 0{,}70 = -0{,}357$; variațiile logaritmice se adună în timp, procentele nu
- Ordinea operațiilor pentru o serie pozitivă, cu trend și sezonalitate
  - întîi logaritmul (varianța), apoi diferențele (trendul, sezonul); niciodată invers

PIB-ul real al României: logaritm, creștere trimestrială și anuală



- Date trimestriale, neajustate sezonier, prețuri din 2010 (Eurostat), T1 1995–T2 2026; sus: seriile; jos: ACF de selecție

TSA_ch1_transformations

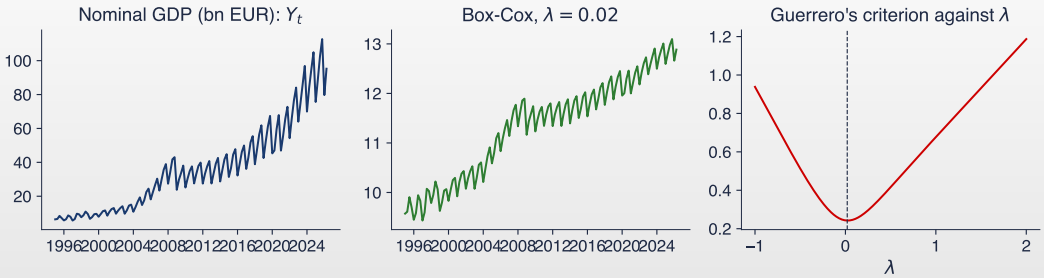
Interpretarea transformărilor PIB

- $\ln Y_t$: ACF 0,68 la lagul 1 și 0,87 la lagul 8, cu vîrfuri la fiecare 4 trimestre: trend și sezon
- $\Delta \ln Y_t$: trendul a dispărut, sezonul domină: $\hat{\rho}(4) = 0{,}96$, $\hat{\rho}(2) = -0{,}49$; abaterea standard 27,5%
  - rata trimestru față de trimestru a datelor neajustate măsoară mai ales sezonul
- $\Delta_4 \ln Y_t$ (creșterea anuală): media 2,7%, ACF se stinge după 2–3 trimestre ($\hat{\rho}(1) = 0{,}63$, $\hat{\rho}(4) = 0{,}03$)
  - apropiată de staționaritate, cu persistența ciclului economic; T2 2020: $-9{,}9\%$; ultimul trimestru: $-2{,}0\%$

Transformarea Box–Cox

- **Definiție** Box și Cox (1964), pentru $y > 0$: $w = \dfrac{y^\lambda - 1}{\lambda}$ pentru $\lambda \neq 0$, $w = \ln y$ pentru $\lambda = 0$
  - $\lambda = 1$: forma nu se schimbă; $\lambda = 0{,}5$: radical; $\lambda = 0$: logaritm; $\lambda = -1$: inversă
  - continuă în λ: $(y^\lambda - 1)/\lambda \to \ln y$ cînd $\lambda \to 0$
- **Exemplu rezolvat**: $y = 100$ și $y = 400$
  - $\lambda = 0{,}5$: $w = 18$ și 38; $\lambda = 0$: $w = 4{,}605$ și 5,991
  - cu cît λ este mai mic, cu atît valorile mari sînt comprimate mai mult
- Alegerea lui λ: Guerrero (1993), ca în Hyndman și Athanasopoulos (2021)
  - împărțim seria pe ani; alegem λ astfel încît $\text{sd}_i/\text{medie}_i^{1-\lambda}$ să fie cît mai constant de la un an la altul; sd_i, medie_i: abaterea standard și media din anul i
- Prognozele se fac pentru w și se transformă înapoi: $y = (\lambda w + 1)^{1/\lambda}$; valoarea obținută este mediana, nu media

Box–Cox pentru PIB-ul nominal al României



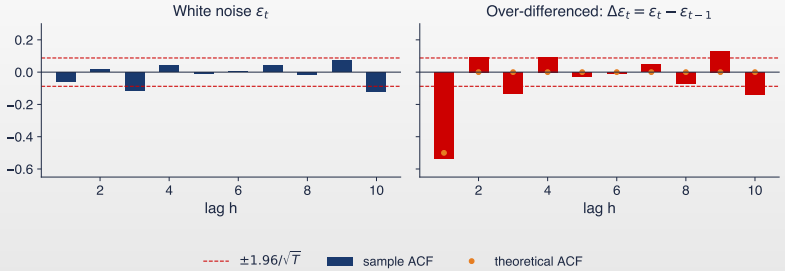
- PIB-ul nominal trimestrial, neajustat sezonier, de la 6,4 la 95,2 mld. EUR; criteriul lui Guerrero pe blocuri anuale de 4 trimestre

TSA_ch1_transformations

Interpretarea alegerii Box–Cox

- $\hat{\lambda}$ al lui Guerrero $= 0{,}02$: practic logaritmul (criteriul 0,243 la $\hat{\lambda}$, 0,243 la $\lambda = 0$, 0,68 la $\lambda = 1$)
  - alegem valoarea rotundă și ușor de interpretat: $\lambda = 0$
- Logaritmul se potrivește deoarece oscilația sezonieră este o proporție stabilă din nivel, circa 39% în primii cinci ani și 41% în ultimii cinci
  - sezonalitatea multiplicativă devine aditivă după logaritmare
- Box–Cox corectează doar varianța; trendul și sezonul tot trebuie eliminate prin diferențiere

Supradiferențierea



- ACF de selecție pentru 500 de valori de zgomot alb (stînga) și pentru prima lui diferență (dreapta)

TSA_ch1_transformations

Interpretarea supradiferențierii

- $\Delta\varepsilon_t = \varepsilon_t - \varepsilon_{t-1}$ este un MA(1) cu $\theta = -1$: $\rho(1) = -1/2$; aici $\hat{\rho}(1) = -0{,}54$
  - varianța se dublează: 0,95 devine 2,01
- Un $\hat{\rho}(1)$ puternic negativ, apropiat de $-0{,}5$, după diferențiere: un semn că seria nu avea nevoie de ea
  - la fel se întîmplă cu o serie staționară în jurul trendului: $\Delta(a + bt + \varepsilon_t) = b + \varepsilon_t - \varepsilon_{t-1}$
- Diferențiem cît este nevoie, nu mai mult; Capitolul 3 dă testele formale pentru numărul de diferențieri

Recapitulare: transformări

- Întîi logaritmul, pentru seriile pozitive ale căror oscilații cresc cu nivelul; apoi diferențele
- $100\,\Delta \ln Y_t$: creșterea în %; Δ_4, Δ_{12}: creșterea anuală pentru date trimestriale și lunare
- Box–Cox cu λ ales prin metoda Guerrero; pentru PIB-ul României $\hat{\lambda} \approx 0$, logaritmul
- Supradiferențierea creează $\rho(1) \approx -0{,}5$ și mărește varianța

Studiu de caz: Yule (1927) și Slutzky (1937)

- **Întrebarea** anilor 1920: de unde provin ciclurile regulate ale petelor solare și ale activității economice?
  - viziunea dominantă: componente periodice ascunse (sume de sinusoide), estimate cu periodograma
- Yule (1927): numărul de pete solare depinde de propriile două valori anterioare plus o perturbație aleatoare
  - $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$: prima autoregresie; un pendul amortizat lovit de șocuri aleatoare
  - un ciclu fără nicio sinusoidă: șocurile îl întrețin, dinamica îi dă perioada
- Slutzky (1937): sumele mobile ale unor numere pur aleatoare (o serie de loterie) produc unde netede, asemănătoare ciclurilor
  - procesul de medie mobilă (MA) și un avertisment: ciclurile din date pot fi create prin mediere
- Moștenirea: procese staționare construite din zgomot alb; Wold (1938) a unificat cele două viziuni (secțiunea 4); Huang și Petukhina (2022), cap. 3

Nilul și petele solare

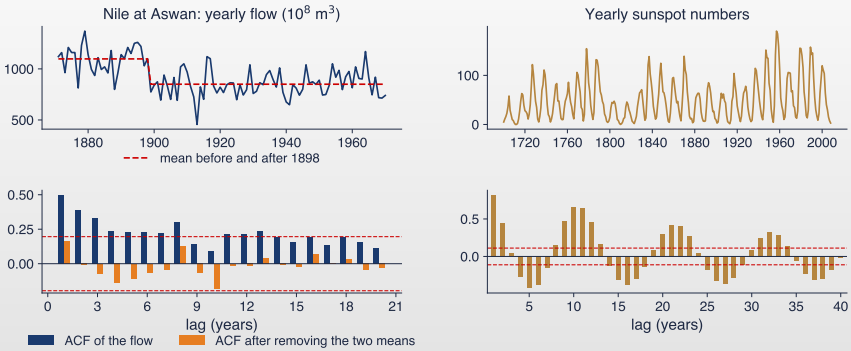


Nilometrul de pe insula Roda, Cairo: nivelurile Nilului se înregistrează aici de peste o mie de ani

Foto: Berthold Werner (2010); CC BY-SA 3.0; Wikimedia Commons

- **Nilul**: debitul anual la Aswan, 1871–1970, 100 de observații; o scădere în jurul anului 1898 Cobb (1978)
- **Petele solare**: valori anuale, 1700–2008, 309 observații; seria care a inspirat autoregresia Yule (1927)
- Ambele sînt incluse în `statsmodels`; ambele oferă o lecție despre interpretarea ACF

Nilul și petele solare: seriile și ACF



⊡ Stînga: debitul Nilului cu media înainte și după 1898, și două ACF; dreapta: numărul de pete solare și ACF pînă la lagul 40

TSA_ch1_textbook_series

Interpretarea celor două serii clasice

- Nilul: $\hat{\rho}(1) = 0{,}50$ și o descreștere lentă; $Q^*(10) = 88{,}1$ ($p < 0{,}001$)
  - după eliminarea celor două medii (1098 și 850): $\hat{\rho}(1) = 0{,}16$, $Q^*(10) = 13{,}0$ ($p = 0{,}225$)
  - o ACF care scade lent poate proveni dintr-o **rupturã**, nu din persistență sau dintr-o rădăcină unitară
- Petele solare: o undă amortizată, minim $-0{,}43$ la lagul 5, vîrf 0,66 la lagul 10
  - ciclul solar de circa 11 ani; o serie staționară cu un ciclu, modelată de Yule (1927) printr-un AR(2)
- Interpretăm ACF împreună cu graficul seriei și cu ce știm despre modul în care au fost produse datele

Contribuția posibilă a AI

- **Cod**: o primă versiune a unui script care descarcă o serie, o reprezintă grafic împreună cu ACF și PACF și aplică testul Ljung–Box pentru mai multe valori ale lui m
- **Explicații**: o a doua explicație a unei definiții sau a motivului pentru care o corelogramă arată într-un anumit fel
- **Explorare**: aceleași diagnostice pe multe serii deodată (toate acțiunile BVB, toate ratele inflației din UE)
- Exemplu de prompt
  - `Write Python code that loads the Romanian monthly HICP from Eurostat, computes 100 times the monthly log difference since 2005, plots the series with its ACF and PACF up to lag 36, and reports the Ljung-Box statistic for m = 12 and m = 24.`

Verificări necesare

- Împărțitorul lui $\hat{\gamma}(h)$ (T, nu $T-h$) și dacă lagul 0 este inclus în grafic
- Tipul benzilor desenate: $\pm 1{,}96/\sqrt{T}$ sau benzile Bartlett pentru MA; ele răspund la întrebări diferite
- Că testul Ljung–Box se aplică transformării staționare, nu nivelului; și cu $m-p-q$ grade de libertate pentru reziduuri
- Că „semnificativ" nu este citit ca „mare" sau „profitabil"
- Datele, frecvențele și unitățile seriilor (ajustate sezonier sau nu, baza indicelui, procente sau zecimale)
- Fiecare referință citată: trebuie să existe; verificați DOI-ul

Idei de reținut

- O serie de timp este o traiectorie a unui proces stochastic; inferența dintr-o singură traiectorie cere staționaritate și ergodicitate
- Staționaritatea slabă: media constantă și o autocovarianță care depinde doar de lag
- Zgomotul alb (slab, i.i.d., gaussian) este piesa de bază; mersul aleator este suma lui cumulată și nu este staționar
- ACF și PACF de selecție cu benzile $\pm 1{,}96/\sqrt{T}$; testul Ljung–Box testează simultan mai multe laguri
- Seriile reale au nevoie de transformări: logaritm sau Box–Cox pentru variață, diferențe pentru trend și sezon
- Randamentele: aproape necorelate, dar pătratele lor sînt puternic corelate

Formule de reținut

Mărimea	Formula
Autocovarianța, ACF	$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}), \quad \rho(h) = \gamma(h)/\gamma(0)$
MA(1)	$\gamma(0) = \sigma^2(1+\theta^2), \quad \rho(1) = \theta/(1+\theta^2), \quad \rho(h) = 0,\ \lvert h\rvert \geq 2$
AR(1)	$\gamma(0) = \sigma^2/(1-\phi^2), \quad \rho(h) = \phi^{\lvert h\rvert}$
Mersul aleator	$\mathrm{Var}(X_t) = t\sigma^2, \quad \mathrm{Cov}(X_t, X_s) = \min(t,s)\,\sigma^2$
Wold	$X_t = \sum_{j\geq 0}\psi_j\varepsilon_{t-j} + \eta_t, \quad \sum_j \psi_j^2 < \infty$
ACF de selecție	$\hat\rho(h) = \sum_{t=1}^{T-h}(x_t-\bar x)(x_{t+h}-\bar x)/\sum_{t=1}^{T}(x_t-\bar x)^2, \quad \pm 1{,}96/\sqrt{T}$
PACF	$\phi_{11} = \rho(1), \quad \phi_{22} = (\rho(2)-\rho(1)^2)/(1-\rho(1)^2)$
Ljung–Box	$Q^*(m) = T(T+2)\sum_{h=1}^{m}\hat\rho(h)^2/(T-h) \sim \chi^2(m)$
Box–Cox	$w = (y^\lambda - 1)/\lambda, \quad w = \ln y\ (\lambda = 0)$

Autoevaluare

- **Întrebare**: $X_t = \varepsilon_t - 0{,}5\varepsilon_{t-1}$, $\sigma^2 = 4$. Cît sînt $\gamma(0)$ și $\rho(1)$?
  - **Răspuns**: $\gamma(0) = 4 \times 1{,}25 = 5$; $\rho(1) = -0{,}5/1{,}25 = -0{,}4$
- **Întrebare**: o corelogramă cu 40 de laguri pentru o serie cu $T = 400$ arată două bare ușor în afara lui $\pm 0{,}098$. Este seria corelată?
  - **Răspuns**: nu neapărat: aproximativ 2 din 40 de bare ies din întîmplare; aplicăm testul Ljung–Box
- **Întrebare**: după diferențiere, $\hat{\rho}(1) = -0{,}48$ și nimic altceva nu este semnificativ. Ce sugerează acest lucru?
  - **Răspuns**: supradiferențiere: seria inițială era probabil deja staționară
- Urmează: Capitolul 2, modele ARMA: identificarea cu ACF și PACF, estimarea, diagnosticarea și prognoza

Bibliografie (1/2)

- Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society*, 8(1), 27–41. doi:10.2307/2983611
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, 26(2), 211–243. doi:10.1111/j.2517-6161.1964.tb00553.x
- Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332), 1509–1526. doi:10.1080/01621459.1970.10481180
- Brockwell, P. J., & Davis, R. A. (1991). *Time Series: Theory and Methods* (2nd ed.). Springer. doi:10.1007/978-1-4419-0320-4
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (3rd ed.). Springer. doi:10.1007/978-3-319-29854-2
- Cobb, G. W. (1978). The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika*, 65(2), 243–251. doi:10.1093/biomet/65.2.243
- Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, 12(1), 37–48. doi:10.1002/for.3980120104
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press. doi:10.2307/j.ctv14jx6sm

Bibliografie (2/2)

- Huang, C., & Petukhina, A. (2022). *Applied Time Series Analysis and Forecasting with Python*. Springer. doi:10.1007/978-3-031-13584-2
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. otexts.com/fpp3
- Khintchine, A. (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, 109, 604–615. doi:10.1007/BF01449156
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303. doi:10.1093/biomet/65.2.297
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications* (4th ed.). Springer. doi:10.1007/978-3-319-52452-8
- Slutzky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica*, 5(2), 105–146. doi:10.2307/1907241
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksell. archive.org
- Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society A*, 226, 267–298. doi:10.1098/rsta.1927.0007