








Serii de timp

Capitolul 0: Introducere: componente și netezire exponențială



Daniel Traian PELE

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

Acronime folosite în acest material (1/3)

- **ACF**: Autocorrelation Function (funcția de autocorelație)
- **AI**: Artificial Intelligence (inteligență artificială)
- **API**: Application Programming Interface (interfață de programare a aplicațiilor)
- **AR**: AutoRegressive (model) — model autoregresiv
- **ARCH**: AutoRegressive Conditional Heteroskedasticity (heteroscedasticitate condiționată autoregresivă)
- **ARFIMA**: AutoRegressive Fractionally Integrated Moving Average (model) — model autoregresiv fracționar integrat cu medie mobilă
- **ARIMA**: AutoRegressive Integrated Moving Average (model) — model autoregresiv integrat cu medie mobilă
- **ARMA**: AutoRegressive Moving Average (model) — model autoregresiv cu medie mobilă
- **ASE**: Academia de Studii Economice din București
- **BET**: Bucharest Exchange Trading (index) — indicele de referință al Bursei de Valori București
- **BNR**: Banca Națională a României
- **BVB**: Bursa de Valori București
- **COVID-19**: COronaVIrus Disease 2019 (boala provocată de coronavirus, 2019)
- **CSIE**: Facultatea de Cibernetică, Statistică și Informatică Economică
- **ECTS**: European Credit Transfer and Accumulation System (sistemul european de credite transferabile)

Acronime folosite în acest material (2/3)

- **EOD**: End Of Day (daily closing data) — date de sfîrșit de zi, adică prețuri de închidere
- **EODHD**: EOD Historical Data (market-data provider) — furnizor de date de piață
- **ETS**: Error, Trend, Seasonality (exponential smoothing state space models) — eroare, trend, sezonalitate (modele de netezire exponențială)
- **FRED**: Federal Reserve Economic Data (baza de date economice a Rezervei Federale din St. Louis)
- **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
- **INS**: Institutul Național de Statistică
- **LLM**: Large Language Model (model lingvistic de mari dimensiuni)
- **LOESS**: LOcally Estimated Scatterplot Smoothing (netezire locală prin regresie ponderată)
- **LPPL**: Log-Periodic Power Law (legea de putere log-periodică)
- **MA**: Moving Average (medie mobilă)
- **MAE**: Mean Absolute Error (eroarea absolută medie)
- **MAPE**: Mean Absolute Percentage Error (eroarea procentuală absolută medie)

Acronime folosite în acest material (3/3)

- **MASE**: Mean Absolute Scaled Error (eroarea absolută medie scalată)
- **PIB**: Produsul intern brut
- **RMSE**: Root Mean Squared Error (rădăcina erorii pătratice medii)
- **SARIMA**: Seasonal ARIMA (model) — model ARIMA sezonier
- **SES**: Simple Exponential Smoothing (netezire exponențială simplă)
- **STL**: Seasonal and Trend decomposition using Loess (descompunerea în sezonalitate și trend cu Loess)
- **SUA**: Statele Unite ale Americii
- **TBATS**: Trigonometric seasonality, Box–Cox, ARMA errors, Trend, Seasonal components (model) — model cu sezonalitate trigonometrică, transformare Box–Cox, erori ARMA, trend și componente sezoniere
- **TRAMO-SEATS**: Time series Regression with ARIMA noise, Missing values and Outliers; Signal Extraction in ARIMA Time Series (metodă de ajustare sezonieră pe baza modelelor ARIMA)
- **TVA**: Taxa pe valoarea adăugată
- **UE**: Uniunea Europeană
- **VAR**: Vector AutoRegression (model vector autoregresiv)
- **VECM**: Vector Error Correction Model (model vectorial cu corecția erorii)

Traseul acestui capitol

- **Întrebări de pornire**
  - ce este o serie de timp și unde o întîlnim?
  - ce tipare se repetă în datele economice, financiare, energetice și climatice?
  - cum prognozăm o serie și de unde știm că prognoza este bună?
- **Partea I**
  - organizarea cursului
  - serii de timp pe date reale și o scurtă istorie
  - componentele: trend, sezonalitate, ciclu, zgomot
- **Partea a II-a**
  - notație, transformări și o primă privire asupra ACF
  - netezirea exponențială; prognoze și evaluarea lor
  - instrumente și contribuția posibilă a AI
- **Rezultatele învățării**: după acest capitol puteți
  - explica felul în care este organizat și evaluat cursul
  - recunoaște într-un grafic trendul, sezonalitatea, ciclul și zgomotul
  - calcula rate de creștere, o medie mobilă și o autocorelație de selecție
  - construi prognoze naive și prognoze prin netezire exponențială
  - evalua prognozele pe un set de test, cu MAE, RMSE și MASE
- Seminarul 0 are loc **înaintea** acestui curs și exersează primele instrumente pe date

Cursul pe scurt

- **Program**
  - Licență, programele Informatică economică și Cibernetică economică, anul III, semestrul 2
  - Facultatea de Cibernetică, Statistică și Informatică Economică (CSIE), ASE București
  - 4 credite ECTS; 2 ore de curs și 1 oră de seminar pe săptămînă; anul universitar 2026/2027
- **Curs**
  - Prof. Daniel Traian Pele, danpele@ase.ro
- **Cunoștințe necesare**
  - probabilități și statistică, econometrie (regresia liniară), Python
- **Site-ul cursului**
  - danpele.github.io/Time-Series-Analysis: slide-uri, seminarii, notebook-uri, Quantlets, quiz-uri



Academia de Studii Economice din București (ASE)

Foto: Joe Mabel (2014); CC BY 3.0; Wikimedia Commons

Evaluare

- **Examen scris: 70%**
  - citirea și interpretarea rezultatelor obținute în Python, derivări scurte, alegerea modelului potrivit pentru o serie dată
  - acoperă cursurile și seminariile Capitolelor 0–10 și 15
- **Proiect de echipă: 20%**
  - 2–4 studenți analizează și prognozează serii de timp reale cu metodele cursului
  - un repository GitHub, un raport scurt, o prezentare și o susținere orală
- **Prezență și quiz-uri: 10%**
  - cursuri și seminarii
  - cîte un quiz pentru fiecare capitol, pe site: 20 de întrebări extrase dintr-o bancă de 24, cu explicație pentru fiecare răspuns
  - rezolvate cu contul ASE; numele și scorul se înregistrează pentru notă

Manuale

- **Manualul de bază**
  - Huang & Petukhina (2022), *Applied Time Series Analysis and Forecasting with Python*, Springer
  - capitolele cursului urmează ordinea manualului: componente și netezire, staționaritate, ARMA, ARIMA, sezonalitate, volatilitate, modele multivariate
- **Manual însoțitor, gratuit**
  - Hyndman & Athanasopoulos (2021), *Forecasting: Principles and Practice* (ediția a 3-a, online)
  - referința pentru descompunere, netezire exponențială și evaluarea prognozei (Capitolul 0)
- **Teorie**
  - Brockwell & Davis (2016), *Introduction to Time Series and Forecasting*
  - Hamilton (1994), *Time Series Analysis*: referința standard a econometriei seriilor de timp

Harta cursului: Capitolele 0–15

Nr.	Capitol
0	Introducere: componente și netezire exponențială
1	Procese stochastice și staționaritate
2	Modele ARMA
3	Rădăcini unitare și modele ARIMA
4	Sezonalitate și prognoză: SARIMA, TBATS, Prophet
5	Volatilitate condiționată: ARCH și GARCH
6	Modele VAR și cauzalitate Granger
7	Cointegrare și VECM

Nr.	Capitol
8	Memorie lungă și ARFIMA
9	Învățare automată pentru serii de timp
10	Modele în spațiul stărilor, filtrul Kalman și modele Markov switching
11	Foundation models pentru serii de timp*
12	Analiză spectrală*
13	Bule speculative: modele LPPL*
14	Modele GARCH multivariate*
15	Recapitulare și pregătire pentru examen

- Modele univariate (0–5), mai multe serii (6, 7), extensii (8–10), recapitulare (15)
- Studiu individual, marcat cu *: Capitolele 11–14 (materiale mai scurte și un quiz, fără seminar obligatoriu)

Materiale și instrumente

- **Pentru fiecare capitol**
  - slide-urile de curs și de seminar, în română și în engleză
  - un notebook de curs și unul de seminar (Python), care se deschid în Google Colab cu un singur clic
  - un quiz de 20 de întrebări
- **Quantlets**: fiecare grafic are un folder public cu codul, un fișier de descriere (`Metainfo.txt`) și graficul
  - pictograma de sub fiecare grafic deschide Quantlet-ul corespunzător pe GitHub
- **Date**: surse publice, citite direct în cod
  - Eurostat, INS, BNR, FRED (Federal Reserve Bank of St. Louis)
  - date zilnice de piață de la EODHD (EOD Historical Data), salvate în repository-ul cursului pînă la 18 septembrie 2026

Seminarii

- **Trei părți**
  - A: calcule scurte pe hîrtie; B: date reale, fiecare cerință încheindu-se cu o întrebare de interpretare; C: o idee deschisă și critica unui răspuns generat de AI
- **Două tipuri de exerciții**
  - **[Rezolvat]**: rezolvarea completă în slide-uri și în notebook, un model de urmat
  - **[Propus]**: îl rezolvați dumneavoastră; rezolvarea se discută la seminar
- **Seminarul nu se notează**: are rol de exercițiu pentru examen și pentru proiect

Proiectul de echipă

- **Întrebarea**: o întrebare de prognoză sau de modelare pe serii de timp reale, aleasă de echipă
- **Metode**: cele din curs
  - descompunere și netezire, ARIMA și SARIMA, GARCH, VAR sau VECM, cauzalitate Granger, funcții de răspuns la impuls, evaluarea prognozei
  - sînt recomandate datele românești (INS, BNR, Eurostat, BVB)
- **Livrabile**
  - un repository GitHub al cărui cod reproduce fiecare rezultat numeric și fiecare grafic
  - un raport scurt, o prezentare și fișierul `AI_USE.md`
- **Criterii de evaluare**
  - o întrebare clară, metode și diagnosticări corecte, o evaluare onestă în afara eșantionului
  - **susținere orală**: fiecare membru explică codul și rezultatele

Politica privind AI

- **Instrumentele AI sînt permise și trebuie declarate**
  - AI (artificial intelligence, inteligență artificială): aici, asistenți construiți pe un LLM (model lingvistic de mari dimensiuni), care scriu text și cod
  - fiecare utilizare se consemnează în `AI_USE.md`: instrumentul, promptul, ce s-a păstrat, ce s-a corectat
- **Răspundeți pentru fiecare rezultat numeric și pentru fiecare referință**
  - greșeli tipice ale AI: referințe inventate, formule greșite, cod care rulează, dar calculează altceva
  - fiecare referință: deschideți DOI-ul (Digital Object Identifier) și verificați titlul
  - fiecare rezultat numeric: recalculați-l cu propriul cod
- **Susținerea orală**
  - o linie de cod pe care nu o puteți explica nu este considerată muncă proprie

Definiție

- **Serie de timp**: observații ale unei variabile, înregistrate în momente succesive și ordonate în timp
- **Notație**: $y_1, y_2, \ldots, y_T$
  - t: indicele de timp (trimestru, lună, zi); T: numărul de observații
  - **frecvența**: numărul de observații pe an (4 trimestriale, 12 lunare, aproximativ 252 pentru zilele de tranzacționare)
- **Exemplu**: PIB-ul real al României, trimestrial, miliarde EUR în prețurile anului 2010
  - T1 2024: 38,7; T2 2024: 45,6; T3 2024: 52,8; T4 2024: 59,2; T1 2025: 38,9
  - ordinea contează: dacă amestecăm cele cinci valori, pierdem informația
- PIB: produsul intern brut; T1–T4: cele patru trimestre ale unui an

Trei tipuri de date

Tip	Structură	Exemplu
Date transversale	multe unități, un singur moment	PIB-ul celor 27 de țări UE în 2025
Serie de timp	o unitate, multe momente	PIB-ul României, în fiecare trimestru din 1995
Date panel	multe unități, multe momente	PIB-ul celor 27 de țări UE, în fiecare trimestru

- **Particularitatea seriilor de timp**
  - observațiile **nu sînt independente**: trimestrul acesta seamănă cu cel anterior
  - această dependență este chiar informația pe care o folosim pentru prognoză
  - formulele clasice pentru eșantioane independente (erori standard, teste) trebuie adaptate
- UE: Uniunea Europeană

Importanța seriilor de timp

Domeniu	Serie	Întrebare tipică
Economie	PIB, inflație, șomaj	Unde va fi inflația peste un an?
Finanțe	cursuri de schimb, indici bursieri, dobînzi	Cît de riscant este BET mîine?
Energie	producția și consumul de electricitate, prețuri	De cîtă electricitate este nevoie în ianuarie viitor?
Climă	concentrația de CO_2, temperaturi	Cît de repede crește CO_2, dincolo de oscilația sezonieră?

- Deciziile depind de prognoze: banca centrală stabilește dobînda pe baza prognozei inflației; operatorul de rețea planifică producția pe baza prognozei cererii

Patru obiective ale analizei seriilor de timp

- **Descriere**: trend, sezonalitate, cicluri, observații neobișnuite
  - grafice, descompunere, autocorelație (acest capitol și Capitolul 1)
- **Modelare**: un model probabilist care ar fi putut genera datele
  - ARMA, ARIMA, GARCH (Capitolele 2–5)
- **Prognoză**: valori încă neobservate, împreună cu incertitudinea lor
  - obiectivul care străbate întregul curs
- **Explicare și control**: cum răspunde o serie la modificarea alteia
  - modele VAR, cauzalitate Granger, cointegrare (Capitolele 6 și 7)
- Box & Jenkins (1970) au formulat aceleași obiective: identificare, estimare, verificare, prognoză

Frecvență, fluxuri și stocuri, date ajustate

- **Flux**: măsurat pe o perioadă; **stoc**: măsurat la un moment
  - fluxuri: PIB-ul unui trimestru, electricitatea produsă într-o lună; se adună în timp
  - stocuri: cursul de schimb dintr-o zi, concentrația de CO_2, nivelul unui indice; le mediem, nu le adunăm
- **Schimbarea frecvenței**
  - din lunar în trimestrial: suma pentru fluxuri, media sau ultima valoare pentru stocuri
- **Date ajustate sezonier**
  - institutul de statistică elimină tiparul regulat din interiorul anului și efectele de calendar (numărul de zile lucrătoare)
  - seria neajustată este cea măsurată; seria ajustată este o estimare, revizuită cînd apar date noi

Datele folosite în acest capitol

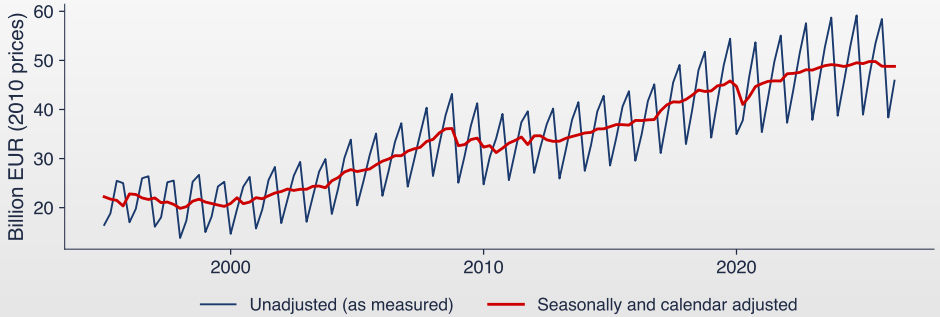
Serie	Sursă	Frecvență	Eșantion
PIB-ul real al României	Eurostat	trimestrial	T1 1995 – T2 2026
IAPC, prețurile de consum	Eurostat	lunar	1996 – august 2026
Producția de electricitate, România	Eurostat	lunar	ianuarie 2008 – iulie 2026
EUR/RON	cursul de referință BNR	zilnic	2005 – 2026
BET, S&P 500	EODHD	zilnic	2000 – 2026
CO2 la Mauna Loa	setul de date statsmodels	săptămînal, mediat lunar	1958 – 2001

- IAPC: indicele armonizat al prețurilor de consum (HICP, Harmonised Index of Consumer Prices), calculat cu aceeași metodă în toate țările UE
- BET: indicele de referință al BVB; S&P 500: 500 de companii mari din SUA
- statsmodels: biblioteca Python de modele statistice folosită pe tot parcursul cursului

Recapitulare: seria de timp

- O serie de timp este o variabilă observată în timp; ordinea observațiilor conține informație
- Observațiile consecutive depind unele de altele: pe acest fapt se sprijină prognoza
- Obiective: descriere, modelare, prognoză, explicare
- Fluxurile se adună, stocurile se mediază; seriile ajustate sînt estimări

PIB-ul real al României



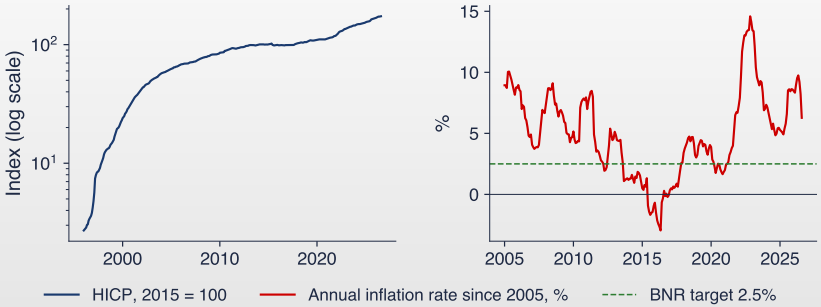
- Trimestrial, T1 1995 – T2 2026, volume înlănțuite în prețurile anului 2010 (miliarde EUR); albastru: seria măsurată; roșu: seria ajustată sezonier și cu efectele de calendar (Eurostat)

TSA_ch0_examples

Interpretarea: PIB

- **Trend**: economia a crescut de aproximativ 2,3 ori din 1995 pînă în 2025
  - o creștere reală medie de 2,8% pe an
- **Sezonalitate**: în fiecare an, trimestrul I este mult sub trimestrul IV
  - iarna: mai puține construcții și lucrări agricole; toamna: recolta și activitatea de la sfîrșit de an
  - oscilația crește odată cu nivelul PIB: un tipar multiplicativ
- **Ciclu**: recesiunile întrerup trendul
  - PIB-ul anual a scăzut cu 5,5% în 2009 (criza financiară globală) și cu 3,6% în 2020 (pandemia de COVID-19)
- Ultima valoare, T2 2026: 45,9 miliarde EUR neajustat, 48,8 ajustat

Prețurile de consum și inflația în România



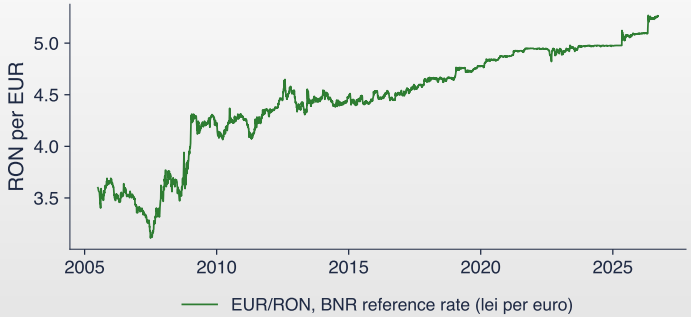
- Stînga: IAPC, 2015 = 100, scară logaritmică; dreapta: rata anuală a inflației $100\,(P_t/P_{t-\mathbf{12}} - 1)$ din 2005, față de ținta BNR de 2,5%; P_t: IAPC în luna t

TSA_ch0_examples

Interpretarea: inflația

- **Nivelul și rata oferă informații diferite**
  - indicele prețurilor aproape nu scade niciodată: trendul lui este inflația acumulată
  - rata inflației crește și scade: este seria pe care o țintește banca centrală
- **Episoade**
  - iunie 1997: 178%, după liberalizarea prețurilor din 1997
  - mai 2016: $-3{,}0\%$, după reducerile de TVA din 2015–2016
  - noiembrie 2022: 14,6%, șocul prețurilor la energie
  - august 2026: 6,3%
- O rată calculată pe 12 luni elimină tiparul sezonier, dar reacționează cu întîrziere: reflectă întregul an trecut
- TVA: taxa pe valoarea adăugată

EUR/RON, cursul de referință BNR



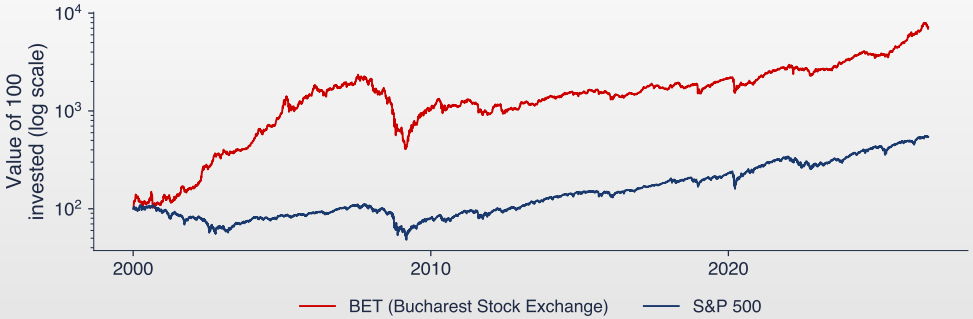
- Lei pentru un euro, în fiecare zi lucrătoare de la 4 iulie 2005 la 18 septembrie 2026 (5 353 de observații)

TSA_ch0_examples

Interpretarea: EUR/RON

- **Un trend fără o formă fixă**
  - de la 3,5986 la 5,2644 lei; valoarea minimă 3,1112 pe 2 iulie 2007
  - perioade lungi de calm, întrerupte de salturi (toamna lui 2008, 2012, 2025)
- **Fără sezonalitate**
  - dacă euro ar fi, previzibil, mai ieftin în fiecare primăvară, participanții la piață l-ar cumpăra primăvara și tiparul ar dispărea
- **Variațiile zilnice**
  - abaterea standard a variației logaritmice zilnice: 0,28%; cea mai mare variație: 2,93% pe 3 octombrie 2008
- Un astfel de trend, format din șocuri acumulate, se numește **trend stochastic** (Capitolele 1 și 3)

BET și S&P 500 din 2000



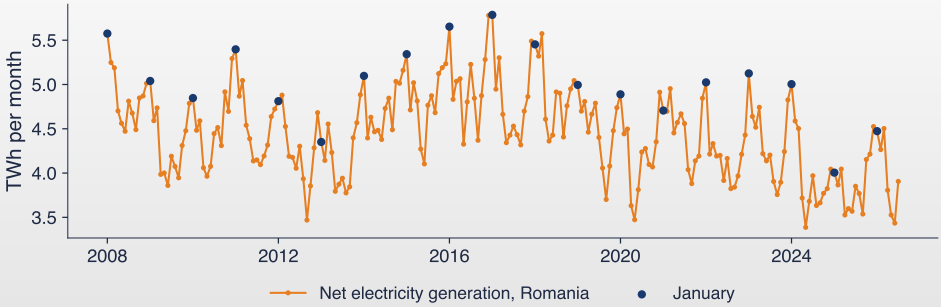
- Valoarea a 100 de unități investite pe 5 ianuarie 2000, prețuri de închidere, scară logaritmică: distanțele verticale egale sînt modificări procentuale egale

TSA_ch0_examples

Interpretarea: indicii bursieri

- **Creștere pe termen lung, cu scăderi adînci**
  - 100 de unități investite s-au înmulțit de aproximativ 71 de ori în BET și de 5,5 ori în S&P 500 (indici de preț, fără dividende)
  - 2008–2009: BET a pierdut mai mult de trei sferturi din valoare; revenirea a durat ani
- **Rolul scării logaritmice**
  - pe o scară liniară, primii ani par plați, iar variațiile recente par uriașe
  - logaritmul transformă o creștere procentuală constantă într-o dreaptă
- Prețurile activelor tranzacționate se prognozează greu; **riscul** lor se prognozează mai ușor (Capitolul 5)

Producția de electricitate în România



Producția netă de electricitate, TWh pe lună, ianuarie 2008 – iulie 2026 (Eurostat); puncte: luna ianuarie a fiecărui an

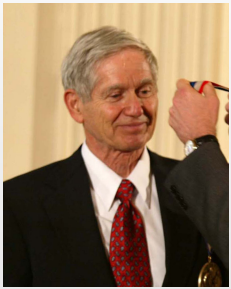
TSA_ch0_examples

Interpretarea: electricitatea

- **Sezonalitate puternică**
  - luna cu media cea mai mare: ianuarie (5,03 TWh); cea mai mică: septembrie (4,06 TWh)
  - cererea de iarnă pentru încălzire și iluminat; raportul celor două medii este 1,24
- **Un nivel în scădere**
  - 58,5 TWh în 2008, 46,7 TWh în 2025: mai multe importuri, o cerere industrială mai mică
- **Ani atipici**
  - anii secetoși reduc producția hidro; iernile blînde reduc cererea
- TWh: terawatt-oră, 10^{12} wați-oră; GWh: gigawatt-oră, 10^{9} wați-oră

Curba Keeling

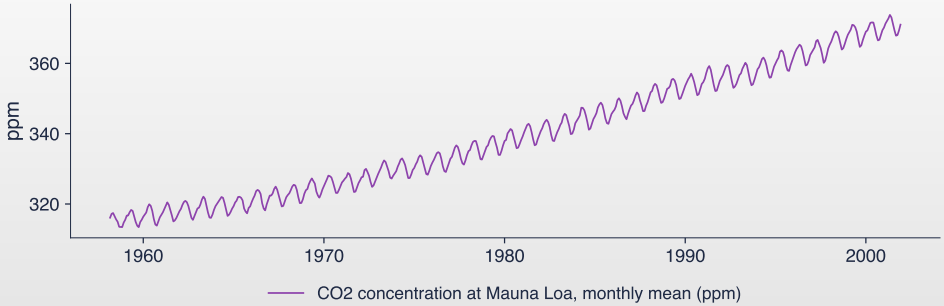
- **1958**: Charles David Keeling începe măsurarea CO_2 la Mauna Loa, Hawaii, departe de surse locale (Keeling (1960))
  - cea mai lungă serie continuă de măsurători ale CO_2 din atmosferă
- **Două tipare într-o singură serie**
  - un trend crescător: emisiile din combustibili fosili se acumulează
  - o undă anuală: plantele din emisfera nordică absorb CO_2 vara și îl eliberează iarna
- ppm (parts per million): molecule de CO_2 la un milion de molecule de aer uscat
- CO_2: dioxid de carbon



Charles David Keeling (1928–2005), 2001

Foto: National Science Foundation (2001); domeniu public; Wikimedia Commons

CO2 atmosferic la Mauna Loa



- Mediile lunare ale măsurătorilor săptămînale, martie 1958 – decembrie 2001 (526 de luni; setul de date clasic din statsmodels)

TSA_ch0_examples

Interpretarea: CO2

- **Trend**: de la 316,1 la 371,0 ppm
  - în medie 1,26 ppm pe an; panta însăși crește în timp
- **Sezonalitate**: o undă anuală regulată, de cîțiva ppm
  - unda își păstrează aproximativ mărimea în timp ce nivelul crește: un tipar aditiv
- **Aproape fără zgomot**
  - o măsurătoare fizică atentă, spre deosebire de PIB sau de cursul de schimb
- O serie netedă, cu trend și sezonalitate clare: cazul cel mai simplu pentru descompunere și netezire

Tipare în șase serii

Serie	Trend	Sezonalitate	Ciclu	Zgomot
PIB (neajustat)	crescător	puternică, multiplicativă	recesiuni	mic
Rata inflației	descrescător după 1997	eliminată de rata pe 12 luni	șocuri	mediu
EUR/RON	stochastic	–	–	mediu
BET	crescător, cu prăbușiri	–	avînturi și prăbușiri	mare
Electricitate	descrescător	puternică, vîrf iarna	–	mediu
CO2	crescător, accelerat	regulată, aditivă	–	foarte mic

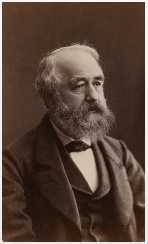
- **Întrebare pentru sală**: care dintre aceste serii ați prognoza-o cel mai precis cu un an înainte?

Tipare în șase serii

Serie	Trend	Sezonalitate	Ciclu	Zgomot
PIB (neajustat)	crescător	puternică, multiplicativă	recesiuni	mic
Rata inflației	descrescător după 1997	eliminată de rata pe 12 luni	șocuri	mediu
EUR/RON	stochastic	–	–	mediu
BET	crescător, cu prăbușiri	–	avînturi și prăbușiri	mare
Electricitate	descrescător	puternică, vîrf iarna	–	mediu
CO2	crescător, accelerat	regulată, aditivă	–	foarte mic

- **Întrebare pentru sală**: care dintre aceste serii ați prognoza-o cel mai precis cu un an înainte?
- **Răspuns**: CO2: un trend neted și o sezonalitate stabilă, puțin zgomot; BET este cea mai grea: fără sezonalitate și cu zgomot mare

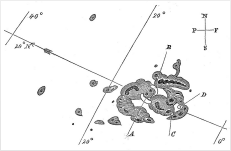
Petele solare: prima serie de timp celebră



Rudolf Wolf (1816–1893)

Foto: Emil Gassler (1888–1893); domeniu public; ETH-Bibliothek, Wikimedia Commons

- **1848**: astronomul elvețian Rudolf Wolf definește un indice zilnic al activității solare, pe baza numărului de pete solare
- îl reconstituie pînă în 1749: numerele de pete solare Wolf (ulterior Wolfer)
- seria crește și scade în valuri de aproximativ 11 ani, dar nici lungimea, nici înălțimea valurilor nu sînt constante
- întrebarea: un semnal periodic ascuns, peste care se suprapune zgomot, sau altceva?



Pete solare desenate de Richard Carrington, 1 septembrie 1859

Desen: Richard Carrington (1859); domeniu public; Wikimedia Commons

Yule (1926, 1927): corelațiile fără sens și modelul autoregresiv

- **1926**: *corelațiile fără sens* dintre seriile de timp (Yule (1926))
  - două serii fără nicio legătură, dar cu trend, pot avea o corelație apropiată de 1
  - exemplul lui: ponderea căsătoriilor oficiate de Biserica Angliei și rata mortalității în Anglia și Țara Galilor
  - astăzi: regresia falsă (Capitolul 3) și cointegrarea (Capitolul 7)
- **1927**: numerele lui Wolfer explicate prin propriul trecut (Yule (1927))
  - în locul unor sinusoide ascunse: $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$, cu perturbații aleatoare ε_t
  - y_t: numărul de pete solare din anul t; ϕ_1, ϕ_2: coeficienți estimați din date, care fixează lungimea și amortizarea valurilor
  - primul model **autoregresiv**: AR(2) (Capitolul 2)
  - perturbațiile deplasează valurile, deci perioada variază, ca în date

Slutsky (1937): cicluri din șocuri aleatoare

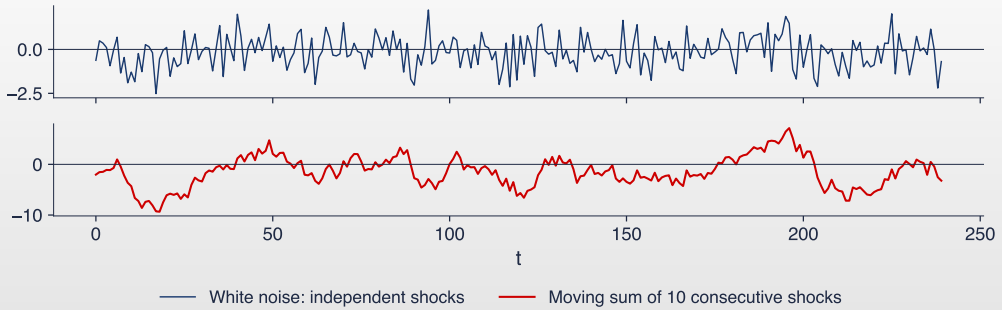


Eugen Slutsky (1880–1948)

Foto: autor necunoscut; domeniu public; Wikimedia Commons

- **1927** (Moscova), **1937** în engleză: *Însumarea cauzelor aleatoare ca sursă a proceselor ciclice* (Slutzky (1937))
- **Experimentul**
  - se iau numere aleatoare independente $\varepsilon_{\mathbf{1}}, \varepsilon_{\mathbf{2}}, \ldots$ (cifrele unei extrageri la loterie)
  - fiecare număr se înlocuiește cu suma ultimelor 10: $y_t = \varepsilon_t + \varepsilon_{t-\mathbf{1}} + \cdots + \varepsilon_{t-\mathbf{9}}$
  - rezultatul seamănă cu ciclurile economice
- Astăzi, acesta este un proces de **medie mobilă**: MA(9) (Capitolul 2)

Experimentul lui Slutsky, simulat



- Sus: 240 de șocuri independente, cu distribuția Normală standard; jos: suma mobilă a cîte 10 șocuri consecutive

TSA_ch0_slutsky

Interpretarea: experimentul lui Slutsky

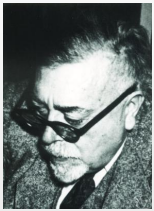
- **Nu am introdus niciun ciclu, totuși apar valuri**
  - suma mobilă traversează zero în sus de 16 ori în 240 de perioade: valuri de lungime neregulată
  - valorile vecine au în comun 9 din cele 10 șocuri, deci sînt puternic corelate
- **Implicații**
  - un ciclu într-un grafic nu dovedește existența unei cauze periodice
  - netezirea (mediile mobile) poate *crea* cicluri aparente: atenție la datele netezite
- Yule (AR) și Slutsky (MA) au oferit cele două componente de bază ale modelelor ARMA

Kolmogorov și Wiener: cea mai bună prognoză liniară



Andrei Kolmogorov (1903–1987)

Foto: Konrad Jacobs; CC BY-SA 2.0 de; Oberwolfach, Wikimedia Commons



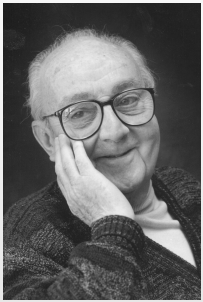
Norbert Wiener (1894–1964)

Foto: Konrad Jacobs; CC BY-SA 2.0 de; Oberwolfach, Wikimedia Commons

- **1941**: teoria interpolării și extrapolării șirurilor staționare (Kolmogorov (1941))
- **1942** (raport secret), **1949**: filtrul Wiener, dezvoltat pentru conducerea tirului antiaerian (Wiener (1949))
- Amîndoi răspund la întrebarea: ce sumă ponderată a valorilor trecute prognozează viitorul cu cea mai mică eroare pătratică medie? Strămoșul filtrului Kalman (Capitolul 10)

Box și Jenkins (1970): o metodă pentru practică

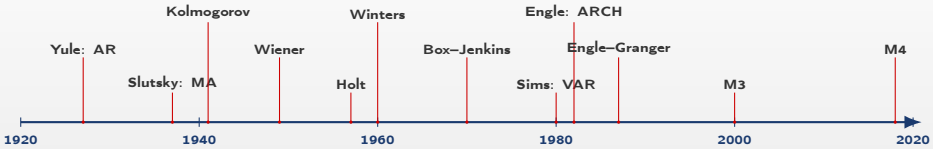
- **1970**: George Box și Gwilym Jenkins publică *Time Series Analysis: Forecasting and Control* (Box & Jenkins (1970))
- **Ciclul Box–Jenkins**
  - **identificare**: alegerea modelului din grafice și din ACF
  - **estimare**: estimarea parametrilor
  - **verificare**: sînt reziduurile un zgomot imprevizibil?
  - **prognoză**: doar după ce verificările au trecut
- Modelele ARIMA și versiunea lor sezonieră, SARIMA, provin din această carte (Capitolele 2–4)
- Box: *toate modelele sînt greșite, dar unele sînt utile*



George E.P. Box (1919–2013)

Foto: DavidMCEddy (2011); CC BY-SA 3.0; Wikimedia Commons

Cronologia analizei seriilor de timp



- 1927 Yule, 1937 Slutsky, 1941 Kolmogorov, 1949 Wiener: fundamentele (Capitolele 1 și 2)
- 1957 Holt (1957), 1960 Winters (1960): netezirea exponențială, apărută în industrie, pentru vînzări și stocuri (acest capitol)
- 1970 Box–Jenkins (Capitolele 2–4); 1980 Sims (1980) (Capitolul 6); 1982 Engle (1982) (Capitolul 5); 1987 Engle & Granger (1987) (Capitolul 7)
- 2000 și 2018: competițiile de prognoză M3 și M4 compară metodele pe mii de serii reale (Makridakis & Hibon (2000); Makridakis, Spiliotis & Assimakopoulos (2020))

Patru componente

- **Trendul** T_t: evoluția pe termen lung a nivelului
  - creșterea PIB, creșterea CO2; își poate schimba lent direcția
- **Sezonalitatea** S_t: un tipar care se repetă cu o **perioadă fixă și cunoscută** m
  - $m = 4$ pentru date trimestriale, $m = 12$ pentru date lunare, $m = 7$ pentru date zilnice cu tipar săptămînal
- **Ciclul** C_t: creșteri și scăderi **fără perioadă fixă**, de obicei mai lungi de un an
  - ciclurile economice: expansiuni și recesiuni de 2–10 ani; de obicei sînt unite cu trendul într-o componentă **trend-ciclu**
- **Componenta neregulată** R_t (reziduală, zgomot): ce rămîne
  - ar trebui să pară imprevizibilă; dacă nu, descompunerea a omis ceva

Ciclu sau sezonalitate?

- **Întrebare pentru sală**: producția de electricitate este mare în fiecare ianuarie
  - este un ciclu sau sezonalitate?

Ciclu sau sezonalitate?

- **Întrebare pentru sală**: producția de electricitate este mare în fiecare ianuarie
  - este un ciclu sau sezonalitate?
- **Răspuns**: sezonalitate: perioada este fixă (12 luni) și cunoscută dinainte
  - o recesiune apare într-un moment necunoscut și durează un timp necunoscut: acesta este un ciclu
- **Importanța diferenței**
  - sezonalitatea poate fi prognozată cu ani înainte: și luna ianuarie viitoare va avea valori mari
  - ciclurile se prognozează greu: momentul lor se schimbă (Slutsky)
  - petele solare: un *ciclu* de 11 ani, nu o sezonalitate, pentru că durata lui variază

Modelul aditiv și modelul multiplicativ

- **Aditiv**: $y_t = T_t + S_t + R_t$
  - oscilația sezonieră are o mărime constantă, în unitățile lui y (CO2: cîțiva ppm în fiecare an)
- **Multiplicativ**: $y_t = T_t \times S_t \times R_t$
  - oscilația sezonieră este un **procent** constant din nivel (PIB: trimestrul I cu aproximativ un sfert sub trend)
  - S_t și R_t sînt factori în jurul lui 1: $S_t = 0{,}8$ înseamnă 20% sub trend
- **Logaritmul le leagă**: $\ln y_t = \ln T_t + \ln S_t + \ln R_t$
  - o serie multiplicativă devine aditivă după logaritmare
- **Alegerea**: priviți graficul
  - oscilația sezonieră crește odată cu nivelul? Multiplicativ (sau logaritmi); altfel, aditiv

Medii mobile centrate

- **Media mobilă de ordin** k (impar): media a k valori vecine
  - $\hat{T}_t = \frac{1}{k}\sum_{j=-q}^{q} y_{t+j}$, cu $k = 2q+1$; netezește zgomotul
  - q: numărul de valori de fiecare parte a lui t; j: poziția față de t; notația $\hat{}$ indică o estimare
- **Proprietatea esențială**: o medie mobilă pe un sezon complet elimină tiparul sezonier
  - fiecare trimestru intră o singură dată, deci efectele sezoniere se compensează
- **Perioadă pară** $m = 4$: o medie mobilă 2×4, pentru a rămîne centrată pe un trimestru
  - $\hat{T}_t = \frac{1}{8}y_{t-2} + \frac{1}{4}y_{t-1} + \frac{1}{4}y_t + \frac{1}{4}y_{t+1} + \frac{1}{8}y_{t+2}$
  - cinci trimestre, cele două capete cu pondere înjumătățită; pentru date lunare, 2×12
- Primele și ultimele $m/2$ valori nu au medie centrată: estimarea trendului lipsește la ambele capete

Exemplu rezolvat: o medie mobilă 2×4

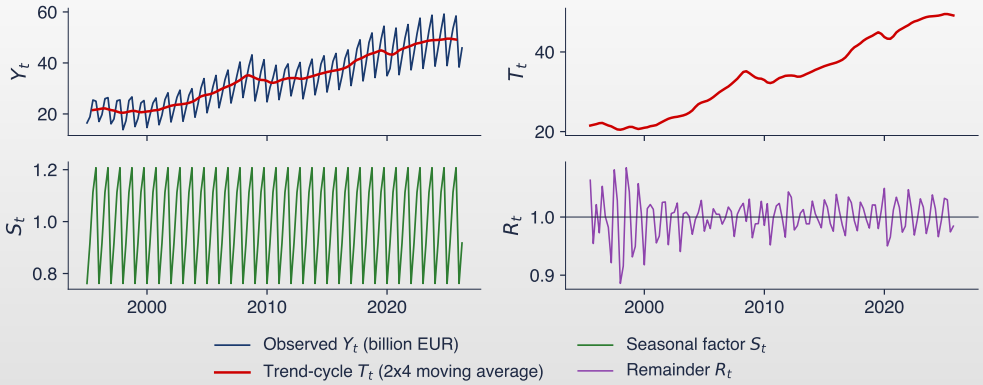
Trimestru	y_t
T1 2024	38,7
T2 2024	45,6
T3 2024	52,8
T4 2024	59,2
T1 2025	38,9

- PIB-ul real al României, miliarde EUR (Eurostat, neajustat)
- **Trendul în T3 2024**
  - $\hat{T}_t = \frac{1}{4}\left(\frac{38{,}7}{2} + 45{,}6 + 52{,}8 + 59{,}2 + \frac{38{,}9}{2}\right) = 49{,}10$
- **Raportul sezonier**
  - $y_t / \hat{T}_t = 52{,}8/49{,}10$: trimestrul III este peste trend
- Media acestor rapoarte pe toți anii oferă factorul sezonier al fiecărui trimestru

Descompunerea clasică, pas cu pas

- **Pasul 1**: trend-ciclul $\hat{T}_t$, printr-o medie mobilă centrată $2 \times m$
- **Pasul 2**: eliminarea trendului
  - multiplicativ: $y_t / \hat{T}_t$; aditiv: $y_t - \hat{T}_t$
- **Pasul 3**: factorii sezonieri $\hat{S}_1, \ldots, \hat{S}_m$
  - media valorilor fără trend din fiecare sezon, pe toți anii
  - normalizare: media lor este 1 (multiplicativ) sau 0 (aditiv)
- **Pasul 4**: componenta neregulată $\hat{R}_t = y_t / (\hat{T}_t \hat{S}_t)$ sau $y_t - \hat{T}_t - \hat{S}_t$
- **Seria ajustată sezonier**: $y_t / \hat{S}_t$ sau $y_t - \hat{S}_t$
  - limite: același tipar sezonier în fiecare an, fără trend la capete, sensibilitate la valori extreme

Descompunerea PIB-ului României



- Descompunere clasică multiplicativă, $m = 4$: seria observată și trend-ciclul, trend-ciclul, factorii sezonieri, componenta neregulată

TSA_ch0_decomposition

Interpretarea: factorii sezonieri ai PIB

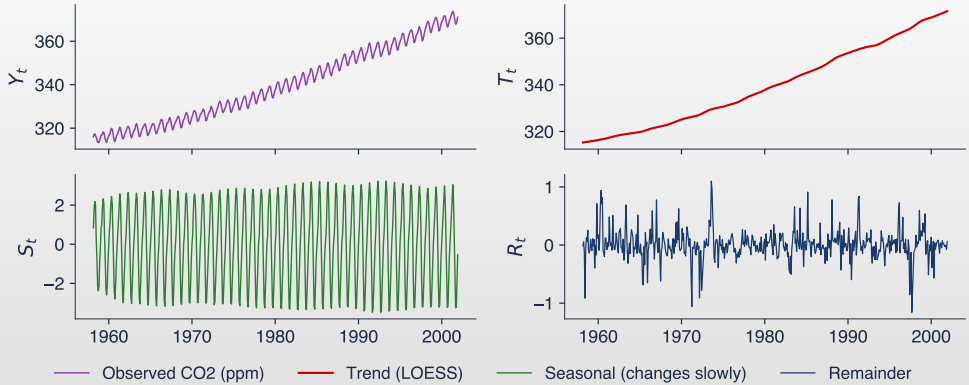
Trimestru	$\hat{S}_q$	față de trend
T1	0,762	$-23{,}8\%$
T2	0,918	$-8{,}2\%$
T3	1,113	$+11{,}3\%$
T4	1,207	$+20{,}7\%$

- **Citirea factorilor**
  - $\hat{S}_q$: factorul sezonier estimat al trimestrului q
  - un trimestru I tipic se află cu 23,8% sub trend, un trimestru IV tipic cu 20,7% peste trend
  - trimestrul IV este cu aproximativ 58% mai mare decît trimestrul I, din motive pur sezoniere
- **Componenta neregulată**
  - abaterea standard este de aproximativ 3,3%; mai mare în anii 1990 și în 2020
- Consecință: comparați un trimestru cu același trimestru din anul anterior, niciodată cu trimestrul precedent al seriei neajustate

STL: o descompunere flexibilă

- **STL**: Seasonal and Trend decomposition using LOESS, descompunere în sezonalitate și trend cu LOESS (Cleveland et al. (1990))
  - LOESS (locally estimated scatterplot smoothing): o regresie ponderată, estimată în jurul fiecărui punct
- **Avantaje față de metoda clasică**
  - tiparul sezonier se poate schimba lent de la un an la altul
  - o estimare a trendului pînă la ultima observație
  - o variantă robustă: valorile extreme ajung în componenta neregulată și nu deformează sezonalitatea
- **Limite**
  - doar aditivă: pentru serii multiplicative se aplică lui $\ln y_t$
  - nu tratează efectele de calendar (zile lucrătoare, Paște)
- Institutele de statistică folosesc X-13ARIMA-SEATS sau TRAMO-SEATS, metode construite pe modele ARIMA (Capitolul 4)

Descompunerea STL a seriei CO2



- STL robustă cu perioada 12, pe seria lunară CO2: seria observată, trendul, sezonalitatea, componenta neregulată (ppm)

TSA_ch0_decomposition

Interpretarea: STL pentru CO2

- **Trendul**: neted, ușor convex: creșterea se accelerează
- **Sezonalitatea**: maximul în mai, minimul în octombrie
  - iarna și primăvara: descompunerea frunzelor căzute eliberează CO2; vara: fotosinteza îl absoarbe
  - amplitudinea anuală crește de la 4,9 ppm (1959) la 6,3 ppm (2001): STL permite modificarea sezonalității
- **Componenta neregulată**: abaterea standard 0,25 ppm
  - mică în raport cu sezonalitatea: trendul și sezonalitatea explică aproape totul
  - verificați vîrfurile mai mari în istoricul stației înainte de a le considera zgomot

Recapitulare: componente și descompunere

- Trend, sezonalitate (perioadă fixă), ciclu (fără perioadă fixă), componenta neregulată
- Aditiv dacă oscilația sezonieră este constantă, multiplicativ dacă ea crește odată cu nivelul; logaritmul transformă un model în celălalt
- O medie mobilă centrată $2 \times m$ elimină sezonalitatea și estimează trend-ciclul
- PIB-ul României: trimestrul I cu aproximativ 23,8% sub trend, trimestrul IV cu aproximativ 20,7% peste trend
- STL permite modificarea tiparului sezonier și este robustă la valori extreme

Notația folosită în curs

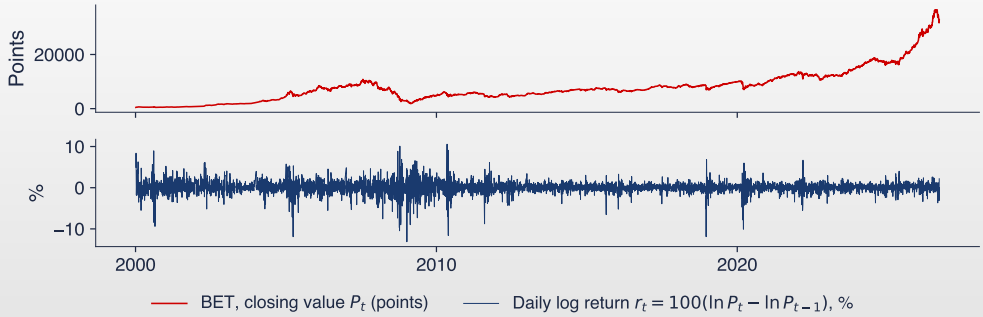
Simbol	Semnificație
y_t, $t = 1, \ldots, T$	observația din momentul t; T observații în eșantion
$\bar{y} = \frac{1}{T}\sum_t y_t$	media de selecție
$Ly_t = y_{t-\mathbf{1}}$	**operatorul lag**: deplasează seria cu o perioadă înapoi; $L^k y_t = y_{t-k}$
$\Delta y_t = y_t - y_{t-\mathbf{1}}$	**diferența de ordinul întîi**, $\Delta = 1 - L$
$\Delta_m y_t = y_t - y_{t-m}$	**diferența sezonieră**, $\Delta_m = 1 - L^m$
$\hat{y}_{T+h\mid T}$	prognoza lui y_{T+h} făcută cu datele pînă la T; h este **orizontul**
$e_t = y_t - \hat{y}_t$	eroarea de prognoză

- Notația $\hat{}$ indică estimările și prognozele; literele grecești (ϕ, θ, α) marchează parametrii

Transformări: logaritmi, diferențe, rate de creștere

- **Logaritmul** $\ln y_t$: stabilizează o variantă care crește odată cu nivelul
- **Rata de creștere pe o perioadă**: $g_t = 100\,(y_t / y_{t-\mathbf{1}} - 1)$
  - aproximarea logaritmică: $100\,\Delta \ln y_t \approx g_t$ pentru variații mici
- **Rata anuală de creștere** a unei serii trimestriale: $100\,(y_t / y_{t-\mathbf{4}} - 1)$
  - compară aceleași trimestre, deci sezonalitatea se anulează
- **Exemplu rezolvat**: PIB-ul României, T1 2024 = 38,7, T4 2024 = 59,2, T1 2025 = 38,9
  - pe un trimestru: $100\,(38{,}9/59{,}2 - 1) \approx -34\%$: nu o prăbușire, ci doar iarna
  - pe un an: $100\,(38{,}9/38{,}7 - 1) \approx +0{,}5\%$: economia aproape nu a crescut
- **Randamentul logaritmic** al unui preț: $r_t = 100\,(\ln P_t - \ln P_{t-\mathbf{1}})$, în %; P_t: prețul din ziua t

De la prețuri la randamente: BET



- Sus: valoarea de închidere a indicelui BET din 2000; jos: randamentele logaritmice zilnice, în % (6 670 de zile de tranzacționare)

TSA_ch0_acf

Interpretarea: prețuri și randamente

- **Prețul are trend; randamentele nu au**
  - randamentele fluctuează în jurul unei medii de 0,064% pe zi, cu abaterea standard de 1,40% (aproximativ 22% pe an)
- **Zile extreme**
  - cea mai slabă: $-13{,}1\%$ pe 7 ianuarie 2009; cea mai bună: $+10{,}6\%$ pe 10 mai 2010
- **Perioadele calme alternează cu cele agitate**
  - variațiile mari apar grupat în 2008–2010 și în martie 2020: *volatility clustering* (Capitolul 5)
- Diferențierea logaritmului prețului elimină trendul: majoritatea modelelor din curs se aplică unor astfel de serii transformate

Funcția de autocorelație de selecție

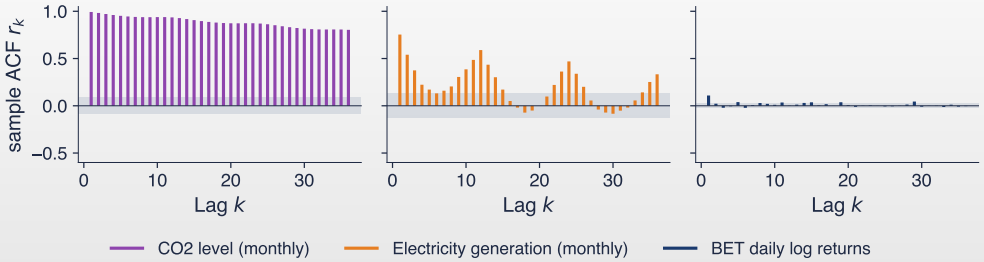
- **Întrebarea**: cît de puternic este legat y_t de propria valoare de acum k perioade?
- **Autocovarianța de selecție** la lagul k
  - $c_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$
  - k: lagul, în perioade; $\bar{y}$: media de selecție; produsul este pozitiv cînd y_t și y_{t-k} se află de aceeași parte a mediei
- **Autocorelația de selecție**: $r_k = c_k/c_0$, între -1 și 1
  - c_0: varianța de selecție (autocovarianța la lagul 0)
  - **ACF** (autocorrelation function, funcția de autocorelație) este șirul $r_1, r_2, \ldots$ reprezentat în funcție de k (**corelograma**)
- **Banda de referință** $\pm 1{,}96/\sqrt{T}$
  - dacă seria ar fi un zgomot independent, aproximativ 95% dintre valorile r_k ar cădea în interiorul ei; 1,96: cuantila de 97,5% a distribuției Normale standard
  - teoria (staționaritate, zgomot alb, testul Ljung–Box) urmează în Capitolul 1

Exemplu rezolvat: r_1 calculat de mînă

t	y_t	$y_t - \bar{y}$	$(y_t - \bar{y})(y_{t-1} - \bar{y})$
1	2	-2	–
2	4	0	0
3	3	-1	0
4	5	1	-1
5	4	0	0
6	6	2	0
sumă	24	0	-1

- $T = 6$, $\bar{y} = 24/6 = 4$
- suma pătratelor abaterilor: $4 + 0 + 1 + 1 + 0 + 4 = 10$
- $r_1 = \dfrac{-1/6}{10/6} = -0{,}1$
- la lagul 2: produsele $(-1)(-2) + (1)(0) + (0)(-1) + (2)(1) = 4$, deci $r_2 = 0{,}4$
- **Interpretare**
  - banda $\pm 1{,}96/\sqrt{6} = \pm 0{,}80$: cu 6 observații, nicio valoare nu se deosebește de 0

Trei corelograme



- ACF de selecție pentru lagurile 1–36; zona colorată: banda $\pm 1{,}96/\sqrt{T}$. Stînga: nivelul CO2; centru: producția lunară de electricitate; dreapta: randamentele logaritmice zilnice ale BET

TSA_ch0_acf

Interpretarea: trei corelograme

- **Trend**: CO2, $r_1 = 0{,}993$, încă 0,80 la lagul 36
  - o descreștere lentă, aproape liniară, este semnătura unui trend: nivelul rămîne corelat cu valorile de acum cîțiva ani
- **Sezonalitate**: electricitatea, $r_6 = 0{,}13$, $r_{12} = 0{,}59$
  - vîrfuri la 12, 24 și 36 de luni: luna ianuarie de anul acesta seamănă cu ianuarie de anul trecut
- **Aproape fără memorie**: randamentele BET, $r_1 = 0{,}109$, $r_2 = 0{,}022$, banda $\pm 0{,}024$
  - autocorelații mici, dar semnificative: greu de exploatat după costurile de tranzacționare
  - randamentele în valoare *absolută* au $r_1 = 0{,}40$: mărimea variațiilor este previzibilă, semnul lor mult mai puțin (Capitolul 5)
- ACF a unei serii cu trend este puțin informativă: întîi transformați seria, apoi examinați ACF

Recapitulare: notație, transformări și ACF

- $Ly_t = y_{t-1}$, $\Delta = 1 - L$, $\Delta_m = 1 - L^m$; $\hat{y}_{T+h|T}$ este prognoza pentru orizontul h
- Logaritmii stabilizează varianța; diferențele elimină trendul; ratele anuale elimină sezonalitatea
- $r_k = c_k/c_0$, comparat cu banda $\pm 1{,}96/\sqrt{T}$
- Descreștere lentă: trend; vîrfuri la $m, 2m, \ldots$: sezonalitate; toate aproape de 0: memorie liniară redusă

Netezirea exponențială simplă (SES)

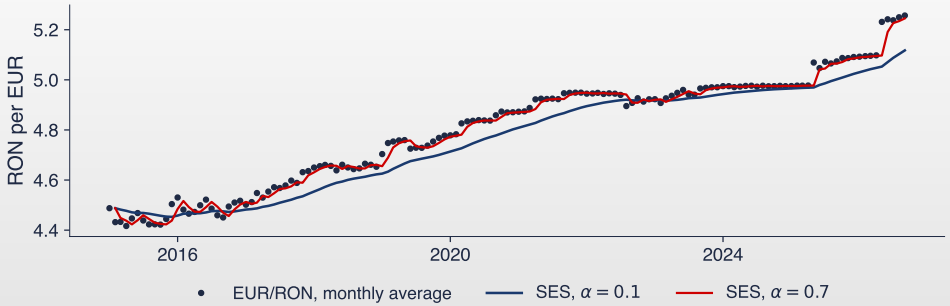
- **Ideea**: prognoza este o medie ponderată a trecutului, în care valorile recente au ponderi mai mari
- **Ecuația nivelului**: $\ell_t = \alpha y_t + (1-\alpha)\,\ell_{t-1}$, cu $0 < \alpha \leq 1$
  - ℓ_t: nivelul netezit în momentul t; α: constanta de netezire; ℓ_t se deplasează de la ℓ_{t-1} spre y_t cu fracțiunea α
  - prognoza pentru orice orizont: $\hat{y}_{T+h|T} = \ell_T$, o dreaptă orizontală
- **Originea denumirii „exponențială"**: substituiri repetate
  - $\ell_T = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots$
  - ponderile scad geometric; cu $\alpha = 0{,}5$: 0,5; 0,25; 0,125; ...
- **Constanta de netezire** α
  - α mic: memorie lungă, prognoze netede, reacție lentă; $\alpha = 1$: prognoza naivă $\hat{y}_{T+1} = y_T$
  - se estimează prin minimizarea sumei pătratelor erorilor de prognoză cu un pas înainte

Exemplu rezolvat: SES cu $\alpha = 0{,}5$

t	y_t	$\hat{y}_{t\mid t-\mathbf{1}} = \ell_{t-\mathbf{1}}$	$\ell_t = 0{,}5\, y_t + 0{,}5\, \ell_{t-\mathbf{1}}$
0	–	–	$\ell_{\mathbf{0}} = 10$
1	10	10	10
2	12	10	11
3	11	11	11
4	13	11	12
5	?	**12**	–

- nivelul inițial $\ell_{\mathbf{0}} = 10$ (prima observație)
- $t = 2$: $0{,}5 \times 12 + 0{,}5 \times 10 = 11$
- $t = 4$: $0{,}5 \times 13 + 0{,}5 \times 11 = 12$
- **Interpretare**
  - prognoza pentru $t = 5$ și pentru toate perioadele următoare este 12: SES nu extrapolează trendul
  - erorile cu un pas: 0, 2, 0, 2: prognozele rămîn în urma datelor crescătoare

SES pe cursul EUR/RON lunar



- Puncte: mediile lunare ale cursului de referință BNR, ianuarie 2015 – septembrie 2026; linii: prognozele SES cu un pas, $\hat{y}_{t|t-1}$, cu $\alpha = 0{,}1$ și $\alpha = 0{,}7$

TSA_ch0_smoothing

Interpretarea: alegerea lui α

- $\alpha = 0{,}1$: netedă, dar mereu sub cursul crescător
  - ponderează observații din mai mulți ani: după saltul din 2025 are nevoie de multe luni pentru a ajunge la noul nivel
- $\alpha = 0{,}7$: urmărește îndeaproape datele, cu o lună întîrziere
- **Valoarea estimată** $\hat{\alpha} = 1{,}00$
  - cea mai bună prognoză SES pentru luna viitoare este media lunii curente: SES se reduce la prognoza naivă
  - tipic pentru cursurile de schimb și prețuri: ultima valoare conține aproape toată informația (un mers aleator, Capitolul 1)
- O valoare estimată a lui α apropiată de 1 este o informație despre serie, nu un eșec al metodei

Metodele Holt și Holt–Winters (1/2)

- **Holt (1957)**: nivel și **trend** b_t (Holt (1957))
  - $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$, $b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$
  - nivelul se actualizează spre y_t, pornind de la nivelul anterior plus panta anterioară
  - b_t: panta estimată, adică variația nivelului pe perioadă; se actualizează spre ultima variație $\ell_t - \ell_{t-1}$
  - $\beta \in (0, 1]$: constanta de netezire a pantei; un β mic dă o pantă care se schimbă lent
- **Prognoza**: $\hat{y}_{T+h|T} = \ell_T + h\, b_T$, o dreaptă
  - ultimul nivel plus de h ori ultima pantă
  - un trend **amortizat** înmulțește panta cu ϕ^j, $0 < \phi < 1$, la pasul j: dreapta se aplatizează pentru orizonturi lungi

Metodele Holt și Holt–Winters (2/2)

- **Holt–Winters (1960)**: adaugă factori sezonieri s_t cu perioada m (Winters (1960))
  - forma multiplicativă: $\hat{y}_{T+h|T} = (\ell_T + h\, b_T)\, s_{T+h-m}$
  - s_{T+h-m}: ultimul factor sezonier al aceluiași sezon (pentru $h \leq m$); $s = 1{,}1$ înseamnă 10% peste linia trendului
  - fiecare componentă are propria constantă de netezire: α (nivel), β (pantă), γ (sezonalitate)
- **ETS**: Error, Trend, Seasonal, adică eroare, trend, sezonalitate (Hyndman, Koehler, Snyder & Grose (2002))
  - fiecare metodă este un model statistic: N (none, absent), A (aditiv), M (multiplicativ), A_d (amortizat)
  - de exemplu, ETS(M,A_d,M): erori multiplicative, trend amortizat, sezonalitate multiplicativă; oferă intervale de prognoză și un criteriu informațional (Capitolul 2)
- O sinteză a 25 de ani de netezire exponențială: Gardner (1985)

Recapitulare: netezirea exponențială

- SES: $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$; ponderile scad geometric; prognoze orizontale
- α mic: netezire puternică și reacție lentă; $\alpha = 1$: prognoza naivă
- EUR/RON: $\hat{\alpha} = 1{,}00$, ultima valoare este cea mai bună prognoză SES
- Holt adaugă un trend, Holt–Winters o sezonalitate; ETS le transformă în modele statistice

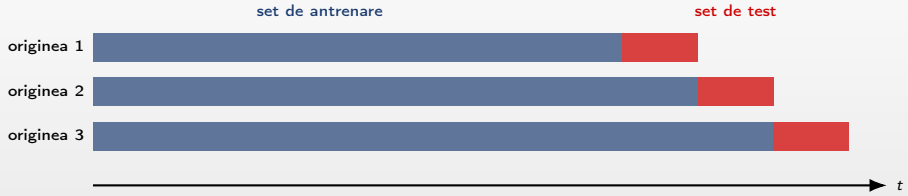
Prognoza: cadrul general

- **Prognoza** $\hat{y}_{T+h|T}$: o afirmație despre y_{T+h}, făcută în momentul T
  - poate folosi doar informația disponibilă în momentul T: **mulțimea de informație**
  - $h = 1, 2, \ldots$: **orizontul**; incertitudinea crește odată cu h
- **Prognoze punctuale și prognoze pe interval**
  - punctuală: un singur număr; pe interval: un interval care conține y_{T+h} cu o probabilitate stabilită (de exemplu, 95%)
- **Condițiile previzibilității**
  - un tipar stabil (CO2), nu unul determinat de știri (cursurile de schimb)
  - o prognoză care nu modifică rezultatul: o prognoză a prețului publicată pentru milioane de participanți la piață schimbă prețul

Patru metode de referință

- **Media**: $\hat{y}_{T+h|T} = \bar{y}$
  - viitorul seamănă cu trecutul mediu
- **Naivă**: $\hat{y}_{T+h|T} = y_T$
  - ultima valoare; optimă pentru un mers aleator (cursuri de schimb, prețuri)
- **Naivă sezonieră**: $\hat{y}_{T+h|T} = y_{T+h-m}$ (pentru $h \leq m$)
  - ianuarie viitor = ianuarie trecut; puternică pentru seriile cu sezonalitate pronunțată
- **Cu derivă**: $\hat{y}_{T+h|T} = y_T + h\,\dfrac{y_T - y_1}{T-1}$
  - ultima valoare plus variația medie pe perioadă din trecut
- Un model nou merită folosit doar dacă depășește aceste metode de referință pe date pe care nu le-a văzut

Setul de antrenare și setul de test



- **Împărțirea se face în timp, niciodată aleator**
  - setul de antrenare: primele observații, folosite pentru alegerea și estimarea metodei
  - setul de test: ultimele h observații, ascunse și folosite doar pentru măsurarea erorilor
  - o împărțire aleatoare îi permite modelului să vadă viitorul: erorile par prea mici
- **Origine mobilă** (validare încrucișată pentru serii de timp): se repetă, mutînd originea înainte
  - mediați erorile pe toate originile: o estimare mai stabilă a acurateței

Măsuri ale erorii de prognoză

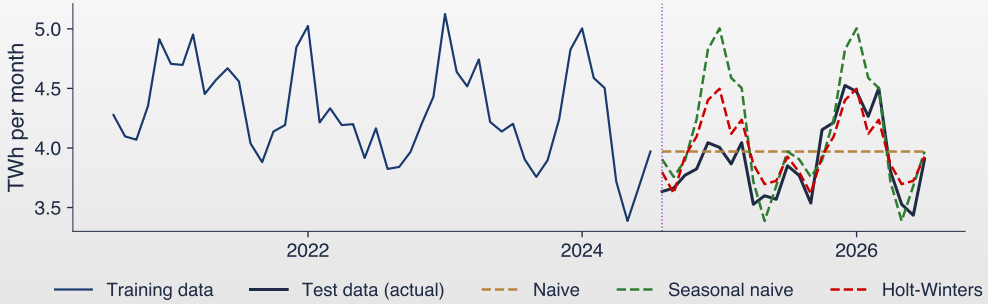
- Erorile pe setul de test: $e_{T+j} = y_{T+j} - \hat{y}_{T+j|T}$, $j = 1, \ldots, h$
- **MAE** (mean absolute error, eroarea absolută medie) $= \frac{1}{h} \sum_j |e_{T+j}|$
  - în unitățile seriei; ușor de explicat
- **RMSE** (root mean squared error, rădăcina erorii pătratice medii) $= \sqrt{\frac{1}{h} \sum_j e_{T+j}^2}$
  - penalizează mai mult erorile mari; întotdeauna $\geq$ MAE
- **MAPE** (mean absolute percentage error, eroarea procentuală absolută medie) $= \frac{100}{h} \sum_j |e_{T+j}/y_{T+j}|$
  - fără unități de măsură, dar inutilizabilă cînd y este aproape de 0 (inflație, randamente)
- **MASE** (mean absolute scaled error, eroarea absolută medie scalată) = MAE / MAE în eșantion a metodei naive (Hyndman & Koehler (2006)); pentru serii sezoniere, a metodei naive sezoniere (ca în M4)
  - sub 1: mai bună decît această metodă de referință pe datele de antrenare; comparabilă între serii

Exemplu rezolvat: erorile calculate de mînă

	T1	T2	T3	T4
anul 1	10	14	18	12
anul 2	11	15	20	13
anul 3 (test)	12	16	21	14
naivă	13	13	13	13
naivă sezonieră	11	15	20	13

- **Naivă**: erorile $-1, 3, 8, 1$
  - MAE $= (1 + 3 + 8 + 1)/4 = 3{,}25$
  - RMSE $= \sqrt{(1 + 9 + 64 + 1)/4} = \sqrt{18{,}75} = 4{,}33$
- **Naivă sezonieră**: erorile $1, 1, 1, 1$
  - MAE $=$ RMSE $= 1$
- **Interpretare**
  - prognoza naivă ignoră sezonalitatea: eroarea ei mare din trimestrul III domină RMSE
  - pentru datele sezoniere, metoda naivă sezonieră este reperul de depășit

Prognoza producției de electricitate



- Setul de test: august 2024 – iulie 2026 (24 de luni); datele de antrenare: toate lunile anterioare; linii întrerupte: prognoze făcute o singură dată, la sfîrșitul setului de antrenare

TSA_ch0_forecast

Interpretarea: compararea preciziei prognozelor

Metodă	MAE	RMSE	MAPE	MASE
Naivă	0,28	0,33	7,3%	0,82
Naivă sezonieră	0,28	0,38	7,2%	0,83
SES	0,28	0,32	7,1%	0,81
Holt–Winters	0,17	0,21	4,5%	0,51

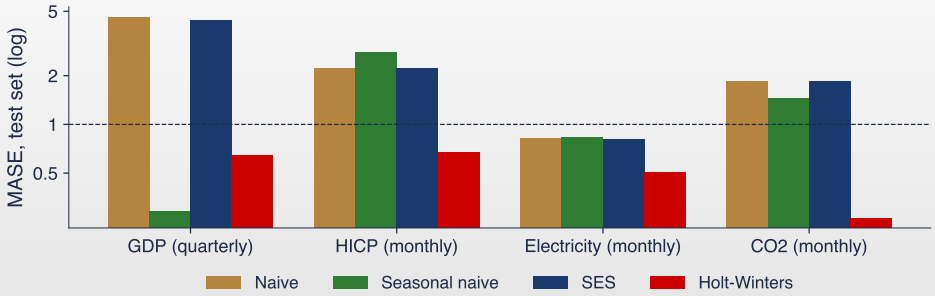
- MAE și RMSE în TWh pe lună

- **Holt–Winters** are cea mai mică eroare după fiecare măsură
  - combină nivelul în scădere cu vîrful de iarnă
- **Aici, metoda naivă sezonieră nu este mai bună decît cea naivă**
  - copiază iarna puternică din 2023–2024 peste perioada 2024–2026, cu un nivel mai scăzut
- O singură perioadă de test de 24 de luni este un singur eșantion: o origine mobilă oferă un răspuns mai solid

Studiu de caz: competiția M4

- **Întrebarea**: ce metode de prognoză funcționează cel mai bine pe date reale?
  - competițiile M, inițiate de Spyros Makridakis în 1982: prognozele sînt trimise fără acces la datele de test și sînt evaluate pe acestea
- **M4 (2018)**: 100 000 de serii (de la anuale la orare), 61 de metode (Makridakis, Spiliotis & Assimakopoulos (2020))
  - 12 dintre cele mai precise 17 metode au fost **combinații** de metode
  - cîștigătoarea a fost o metodă hibridă, care combină netezirea exponențială cu o rețea neuronală
  - metodele care folosesc exclusiv învățarea automată au avut rezultate slabe: niciuna nu a depășit o combinație simplă de metode de netezire exponențială
- **Relevanța pentru acest curs**
  - întîi metodele de referință; acuratețea se măsoară în afara eșantionului, pe multe serii, nu pe una singură
  - aceeași lecție apăruse și în M3 (Makridakis & Hibon (2000)): o metodă complexă nu este automat mai bună

O mică competiție M pe date românești



⊡ MASE pe setul de test (scară logaritmică) pentru patru metode și patru serii sezoniere; seturi de test: ultimele 8 trimestre (PIB) sau 24 de luni; linia întreruptă: MASE = 1

TSA_ch0_forecast

Interpretarea: mini-competiția

- **Holt–Winters câștigă pe 3 din 4 serii**
  - CO2: MASE 0,26, față de 1,45 pentru metoda naivă sezonieră; IAPC: 0,68, față de 2,24 pentru cea naivă
- **PIB: câștigă metoda naivă sezonieră**, MASE 0,29 față de 0,65
  - creșterea s-a oprit în 2025–2026: trendul estimat de Holt–Winters duce la prognoze prea mari
- **Metodele naivă și SES eșuează pe seriile sezoniere**
  - PIB: MASE 4,62 (naivă) și 4,39 (SES): ignoră sezonalitatea
- Ca în M4: nicio metodă nu câștigă peste tot, iar o metodă de referință simplă poate depăși un model aparent mai bun

Recapitulare: prognoza și evaluarea ei

- $\hat{y}_{T+h|T}$ folosește doar informația pînă la T; incertitudinea crește odată cu h
- Metode de referință: media, naivă, naivă sezonieră, cu derivă
- Împărțiți datele în timp; măsurați MAE, RMSE, MAPE și MASE doar pe setul de test
- M4 și mini-competiția noastră: metodele simple sînt puternice; combinațiile și Holt–Winters au rezultate bune

Python pentru serii de timp

- **Bibliotecile cursului**
  - **NumPy**: tablouri și algebră liniară (Harris et al. (2020))
  - **pandas**: tabele indexate după dată, schimbarea frecvenței, laguri, diferențe (McKinney (2010))
  - **matplotlib**: grafice
  - **statsmodels**: descompunere, netezire exponențială, ARIMA, VAR, teste de rădăcină unitară (Seabold & Perktold (2010))
  - **arch**: modele GARCH (Capitolul 5)
- **Google Colab**: notebook-uri în browser, fără instalare
  - deschideți un notebook al cursului cu un singur clic; *File* → *Save a copy in Drive* vă păstrează modificările
  - fiecare grafic din slide-uri are un Quantlet: același cod, care poate fi rulat în Colab

Primul script

```
import pandas as pd
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.holtwinters import ExponentialSmoothing

# Romanian real GDP, quarterly (Eurostat)
url = ('https://ec.europa.eu/eurostat/api/dissemination/'
       'sdmx/2.1/data/namq_10_gdp/Q.CLV10_MEUR.NSA.B1GQ.RO'
       '?format=SDMX-CSV')
t = pd.read_csv(url)
y = pd.Series(t['OBS_VALUE'].values / 1000,
              index=pd.PeriodIndex(t['TIME_PERIOD'], freq='Q'))
y = y.to_timestamp()

dec = seasonal_decompose(y, model='multiplicative', period=4)
print(dec.seasonal.iloc[:4])            # seasonal factors
yoy = 100 * (y / y.shift(4) - 1)         # annual growth, %
hw = ExponentialSmoothing(y, trend='add', seasonal='mul',
                          seasonal_periods=4).fit()
print(hw.forecast(4))                    # next four quarters
```

- Linie cu linie
- citirea seriei direct din API-ul Eurostat (fără cheie)
- `PeriodIndex`: trimestrele ca indice de timp
- descompunere clasică multiplicativă, cu $m = 4$
- `shift(4)`: valoarea din același trimestru al anului anterior
- Holt–Winters cu trend aditiv și sezonalitate multiplicativă
- API (application programming interface): interfață de programare a aplicațiilor

Contribuția posibilă a AI

- **Unde ajută un asistent AI**
  - o primă variantă a codului care citește și reprezintă grafic o serie Eurostat sau BNR
  - explicarea unui mesaj de eroare sau a unei funcții statsmodels
  - enumerarea explicațiilor posibile ale unui tipar (o scădere a PIB, un salt al cursului EUR/RON)
- **Greșeli frecvente în codul generat de AI pentru serii de timp**
  - împărțirea aleatoare în set de antrenare și set de test, care îi permite modelului să vadă viitorul
  - creșterea pe un trimestru numită „creștere anuală”; `shift(1)` în loc de `shift(4)`
  - RMSE fără rădăcina pătrată; MAPE pe o serie care trece prin zero
- Seminarul 0, exercițiul C2: găsiți trei astfel de greșeli într-un răspuns generat de AI

Verificări necesare

- **Datele**
  - reprezentați seria grafic înainte de orice calcul; verificați frecvența, unitățile de măsură, prima și ultima dată
  - ajustată sau neajustată? Un model pentru sezonalitate are nevoie de seria neajustată
- **Evaluarea**
  - setul de test urmează după setul de antrenare; se folosește o singură dată, la final
  - comparați cu prognozele naivă și naivă sezonieră înainte de a declara un succes
- **Afirmațiile**: verificările din politica privind AI (referințe, rezultate numerice, `AI_USE.md`)

Idei de reținut

- O serie de timp este ordonată în timp; dependența de trecut face posibilă prognoza
- Componente: trend, sezonalitate (perioadă fixă), ciclu (fără perioadă fixă), componenta neregulată; aditiv sau multiplicativ
- Yule și Slutsky: dependența și sumele de șocuri, originile modelelor AR și MA
- ACF dezvăluie trendul (descreștere lentă), sezonalitatea (vîrfuri la m) și memoria
- Netezirea exponențială acordă ponderi mai mari trecutului recent; evaluați orice prognoză față de metodele naive, pe un set de test

Autoevaluare, idee de proiect; urmează Capitolul 1

- **Autoevaluare**
  - De ce este înșelătoare creșterea față de trimestrul anterior a PIB-ului neajustat?
  - Ce înseamnă un factor sezonier de 0,76 pentru trimestrul I?
  - Ce spune despre o serie o valoare estimată a lui α apropiată de 1?
  - De ce trebuie ca setul de test să urmeze după setul de antrenare?
- **Idee de proiect**
  - prognozați producția lunară de electricitate a României cu un an înainte; comparați metoda naivă sezonieră, Holt–Winters și STL urmată de SES, cu origine mobilă
- **Capitolul 1: procese stochastice și staționaritate**
  - Ce este un proces stochastic?
  - Cînd are o serie medie și variață stabile?
  - Ce sînt zgomotul alb și mersul aleator?
  - Cum testăm că autocorelațiile sînt nule?

Bibliografie (1/3)

- Box, G.E.P., Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day; 4th ed. with G.C. Reinsel: Wiley (2008).
- Brockwell, P.J., Davis, R.A. (2016). *Introduction to Time Series and Forecasting* (3rd ed.). Springer.
- Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I. (1990). STL: a seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1), 3–73.
- Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.
- Engle, R.F., Granger, C.W.J. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(2), 251–276.
- Gardner, E.S. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting*, 4(1), 1–28.
- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
- Harris, C.R., Millman, K.J., van der Walt, S.J., et al. (2020). Array programming with NumPy. *Nature*, 585, 357–362.

Bibliografie (2/3)

- Holt, C.C. (1957; reprinted 2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1), 5–10.
- Huang, C., Petukhina, A. (2022). *Applied Time Series Analysis and Forecasting with Python*. Springer.
- Hyndman, R.J., Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.
- Hyndman, R.J., Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Hyndman, R.J., Koehler, A.B., Snyder, R.D., Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439–454.
- Keeling, C.D. (1960). The concentration and isotopic abundances of carbon dioxide in the atmosphere. *Tellus*, 12(2), 200–203.
- Kolmogorov, A.N. (1941). Interpolation and extrapolation of stationary random sequences. English translation in *Selected Works of A.N. Kolmogorov*, Vol. II, Springer (1992), 272–280.
- Makridakis, S., Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16(4), 451–476.
- Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74.

Bibliografie (3/3)

- McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 56–61.
- Seabold, S., Perktold, J. (2010). Statsmodels: econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference*, 92–96.
- Sims, C.A. (1980). Macroeconomics and reality. *Econometrica*, 48(1), 1–48.
- Slutzky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica*, 5(2), 105–146.
- Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press.
- Winters, P.R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3), 324–342.
- Yule, G.U. (1926). Why do we sometimes get nonsense-correlations between time-series? *Journal of the Royal Statistical Society*, 89(1), 1–63.
- 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.