



# Analiza și Prognoza Seriilor de Timp

## Capitolul 3: Modele ARIMA pentru Date Nestaționare



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

## Obiective de Învățare

La finalul acestui capitol, veți fi capabili să:

1. **Înțelegeți** conceptul de nestăționaritate și implicațiile sale
2. **Aplicați** diferențierea pentru a obține staționaritate
3. **Folosiți** testul Augmented Dickey-Fuller (ADF) pentru detectarea rădăcinii unitate
4. **Construiți**, estimați și prognozați cu modele  $ARIMA(p, d, q)$
5. **Evaluați** prognozele prin metoda ferestrei rolling
6. **Aplicați** metodologia Box-Jenkins pe date reale (PIB SUA)

## Surse de date și instrumente software

### Surse de date

- ▣ **FRED** (Federal Reserve)
  - ▶ PIB real SUA (GDPC1), rate dobânzi
- ▣ **Yahoo Finance**
  - ▶ Prețuri acțiuni, cursuri de schimb
- ▣ **Eurostat / BNR**
  - ▶ Date macroeconomice europene
- ▣ **Statsmodels datasets**
  - ▶ Sunspots, Nile, Macrodata

### Python

- ▣ statsmodels — modele ARIMA
- ▣ pmdarima — selecție automată
- ▣ pandas-datareader — descărcare FRED
- ▣ matplotlib — vizualizare
- ▣ scipy — teste statistice

### Resurse

- ▣ [github.com/QuantLet/TSA/TSA\\_ch3](https://github.com/QuantLet/TSA/TSA_ch3)
- ▣ [quantlet.com](https://quantlet.com)

## Structura capitolului

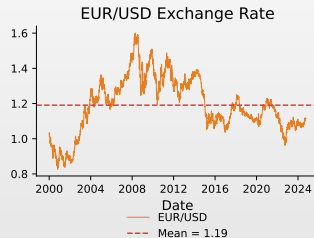
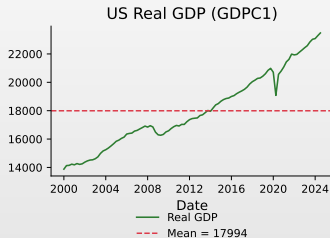
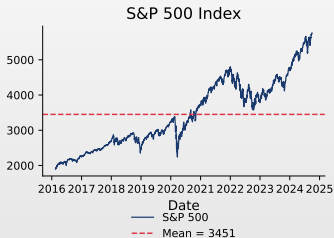
- ▣ Motivație
- ▣ Nestaționaritatea în seriile de timp
- ▣ Diferențierea și Operatorul diferență
- ▣ Modele ARIMA(p,d,q)
- ▣ Teste de rădăcină unitate
- ▣ Identificarea modelului ARIMA
- ▣ Estimarea ARIMA
- ▣ Diagnosticul modelului
- ▣ Prognoza cu ARIMA
- ▣ Studiu de Caz: Prognoza PIB SUA
- ▣ Utilizare IA
- ▣ Rezumat
- ▣ Quiz

## De ce ARIMA? Datele nestaționare sunt pretutindeni

- ▣ Prețurile acțiunilor, PIB, cursurile de schimb prezintă **trenduri** sau **mers aleatoriu**
- ▣ Media din eșantion (linia roșie) este lipsită de sens pentru un mers aleator ( $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ )
- ▣ Modelele ARMA standard **nu pot** gestiona aceste serii direct

## De ce ARIMA? Datele nestaționare sunt pretutindeni

Non-stationary data: sample mean is meaningless



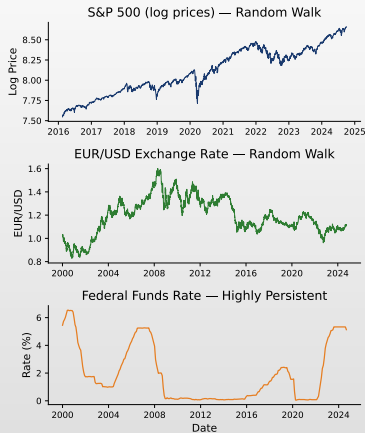
 TSA\_ch3\_motivation\_nonstationary

## Aplicații practice

### Provocarea

- ▣ Prețuri de acțiuni: mers aleator (preț logaritmice)
- ▣ Cursuri de schimb: mers aleator
- ▣ Rate ale dobânzii: foarte persistente, aproape de rădăcină unitate

## Aplicații practice



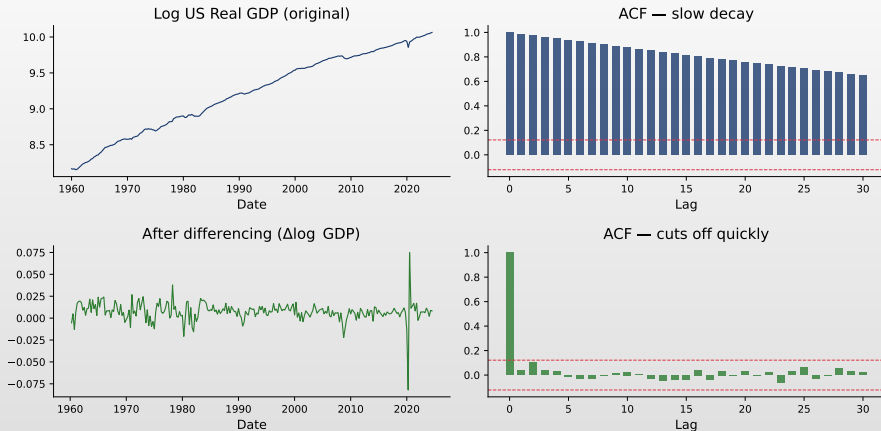


## Soluția: diferențierea

### Observație Cheie

- **Diferențierea** transformă o serie nestaționară într-una staționară:  $\Delta Y_t = Y_t - Y_{t-1}$
- ACF se schimbă de la descreștere lentă la descreștere rapidă!

## Soluția: diferențierea



## Ce vom învăța astăzi

### Concepte Fundamentale

1. **Nestaționaritatea:** de ce contează și cum o detectăm
2. **Teste de rădăcină unitate:** ADF, PP, KPSS
3. **Diferențierea:** transformarea cheie
4. **Modele ARIMA:** combinarea diferențierii cu ARMA
5. **Metodologia Box-Jenkins:** Identificare → Estimare → Diagnoză

### La Sfârșitul Acestui Curs

Veți putea modela și prognoza serii de timp nestaționare precum prețurile acțiunilor, PIB-ul și cursurile de schimb folosind modele ARIMA.

## De ce contează nestaționaritatea

### Problema

- ▣ Multe serii de timp economice și financiare sunt **nestaționare**:
  - ▶ PIB, prețuri de acțiuni, cursuri de schimb, indici de inflație
  - ▶ Prezintă trenduri, medii în schimbare sau varianță în creștere

### Consecințele Nestaționarității

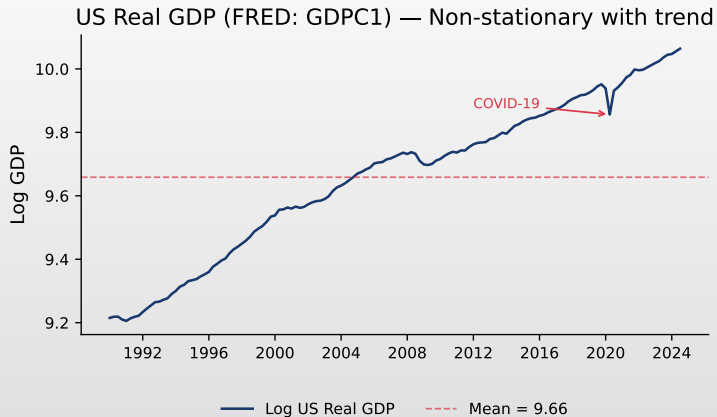
- ▣ Modelele ARMA standard presupun staționaritate
- ▣ Regresia OLS cu date nestaționare duce la **regresie falsă**
- ▣ Momentele din eșantion (medie, varianță, ACF) nu sunt estimatori consistenți
- ▣ Inferența statistică devine invalidă

## Exemplu: PIB real SUA

### Observații

- ▣ **Trend** ascendent clar  $\nearrow$  media nu este constantă
- ▣ Exemplu clasic de serie **nestaționară**
- ▣ Nu putem aplica modele ARMA direct

## Exemplu: PIB real SUA



## Tipuri de nestaționaritate

### Trend determinist

- ▣ **Model:**  $Y_t = \alpha + \beta t + \varepsilon_t$
- ▣ **Trend:** funcție deterministă de timp
  - Poate fi eliminat prin regresie
- ▣ **Șocuri:** au efecte temporare

### Trend stochastic (Rădăcină Unitate)

- ▣ **Model:**  $Y_t = Y_{t-1} + \varepsilon_t$
- ▣ **Tip:** proces de mers aleator
  - Trebuie eliminat prin diferențiere
- ▣ **Șocuri:** au efecte permanente

### Distincție Cheie

- ▣ Identificarea corectă este crucială: eliminarea trendului prin regresie pentru un proces cu rădăcină unitară sau diferențierea unui proces staționar în trend duc ambele la specificare greșită!

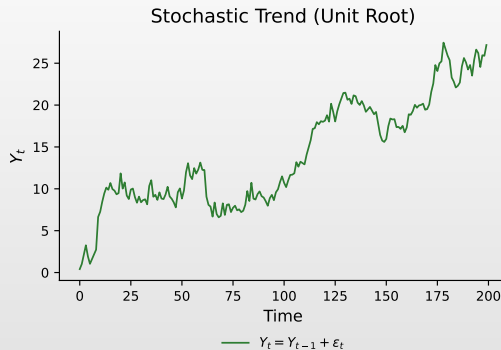
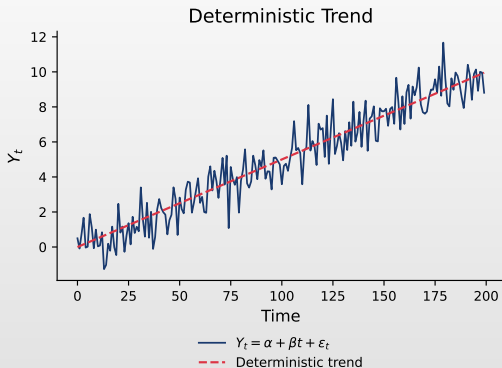
## Vizualizarea diferenței

### Observații

- ▣ **Stânga:** Trend determinist — abaterile de la trend sunt temporare
- ▣ **Dreapta:** Trend stochastic — șocurile se acumulează permanent
- ▣ Ambele arată similar, dar necesită tratamente **diferite!**



## Vizualizarea diferenței



## Procesul de mers aleator

### Definiție 1 (Mers aleatoriu)

- ▣ **Definiție:**  $Y_t = Y_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \sim WN(0, \sigma^2)$
- ▣ **Condiție inițială:**  $Y_0 = 0 \succ Y_t = \sum_{i=1}^t \varepsilon_i$

### Proprietățile Mersului Aleatoriu

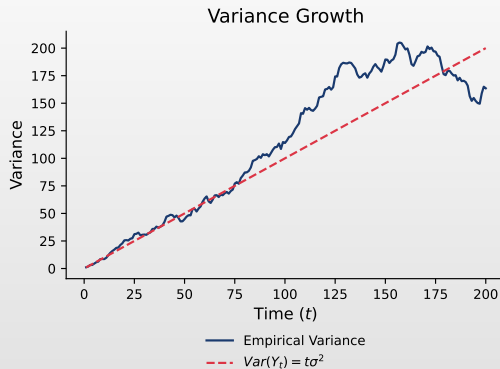
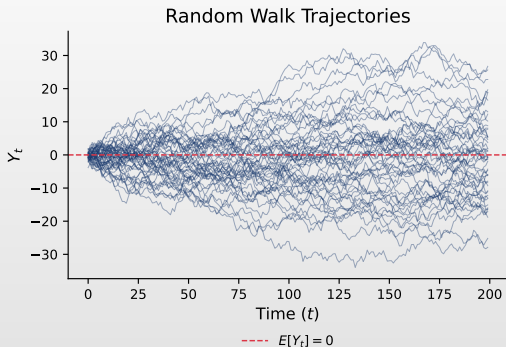
- ▣  $\mathbb{E}[Y_t] = 0$  (medie constantă)
- ▣  $\text{Var}(Y_t) = t\sigma^2$  (varianța crește în timp!)
- ▣  $\text{Cov}(Y_t, Y_{t-k}) = (t-k)\sigma^2$  pentru  $k \leq t$
- ▣ ACF:  $\rho_k = \sqrt{\frac{t-k}{t}} \rightarrow 1$  când  $t \rightarrow \infty$

## Mers aleatoriu: ilustrație vizuală

### Proprietăți cheie

- ▣ **Stânga:** Traiectorii rătăcesc imprevizibil, fără revenire la medie
- ▣ **Dreapta:**  $\text{Var}(Y_t) = t\sigma^2$  crește liniar — caracteristica definitorie a nestaționarității

## Mers aleatoriu: ilustrație vizuală



## Demonstrație: varianța mersului aleatoriu

### Afirmație

- Pentru  $Y_t = Y_{t-1} + \varepsilon_t$  cu  $Y_0 = 0$ :  $\text{Var}(Y_t) = t\sigma^2$

### Demonstrație

- Prin substituție recursivă:  $Y_t = Y_{t-1} + \varepsilon_t = \dots = \sum_{i=1}^t \varepsilon_i$
- Luând varianța:  $\text{Var}(Y_t) = \sum_{i=1}^t \text{Var}(\varepsilon_i) + 2 \sum_{i < j} \text{Cov}(\varepsilon_i, \varepsilon_j)$
- Deoarece  $\varepsilon_t$  sunt independente, toate covarianțele sunt zero
- $\succ \text{Var}(Y_t) = \sum_{i=1}^t \sigma^2 = \boxed{t\sigma^2}$

### Nestaționaritate

- Varianța depinde de  $t \succ$  încalcă cerința staționarității ( $\text{Var}(Y_t) = \gamma(0)$  constant)

## Demonstrație: autocovarianța mersului aleatoriu

### Afirmație

$$\square \text{Cov}(Y_t, Y_{t-k}) = (t-k)\sigma^2 \text{ pentru } k \leq t$$

### Demonstrație

$$\square \text{Folosind } Y_t = \sum_{i=1}^t \varepsilon_i \text{ și } Y_{t-k} = \sum_{i=1}^{t-k} \varepsilon_i:$$

$$\text{Cov}(Y_t, Y_{t-k}) = \sum_{i=1}^t \sum_{j=1}^{t-k} \text{Cov}(\varepsilon_i, \varepsilon_j) = \sum_{i=1}^{t-k} \text{Var}(\varepsilon_i) = \boxed{(t-k)\sigma^2}$$

$$\square \text{Doar termenii cu } i = j \text{ supraviețuiesc (când } i \leq t-k)$$

### ACF

$$\square \rho(k) = \frac{\text{Cov}(Y_t, Y_{t-k})}{\sqrt{\text{Var}(Y_t)\text{Var}(Y_{t-k})}} = \frac{(t-k)\sigma^2}{\sqrt{t\sigma^2 \cdot (t-k)\sigma^2}} = \sqrt{\frac{t-k}{t}}$$

## Mers aleatoriu cu drift

### Definiție 2 (Mers aleatoriu cu drift)

- ▣ **Model:**  $Y_t = \mu + Y_{t-1} + \varepsilon_t$
- ▣ **Echivalent:**  $Y_t = Y_0 + \mu t + \sum_{i=1}^t \varepsilon_i$

### Proprietăți

- ▣  $\mathbb{E}[Y_t] = Y_0 + \mu t$  (media crește liniar)
- ▣  $\text{Var}(Y_t) = t\sigma^2$  (varianța tot crește)
- ▣ Drift-ul  $\mu$  creează un trend ascendent sau descendent
- ▣ Tot nestaționar în ciuda faptului că are un “trend”

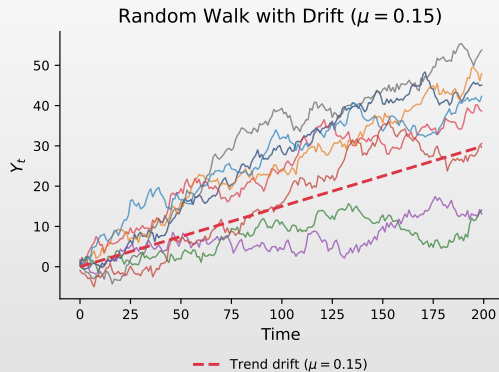
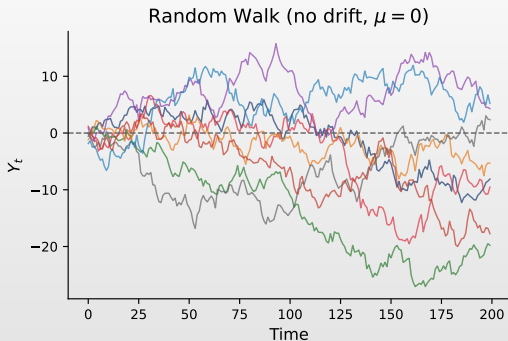
## Mers aleatoriu cu drift: ilustrație vizuală

### Comparație

- ▣ **Fără drift** (albastru): rătăcește în jurul lui zero
- ▣ **Cu drift**  $\mu > 0$  (roșu): trend ascendent sistematic
- ▣ Ambele sunt nestaționare — drift-ul adaugă trend determinist la rătăcirea stochastică



## Mers aleatoriu cu drift: ilustrație vizuală

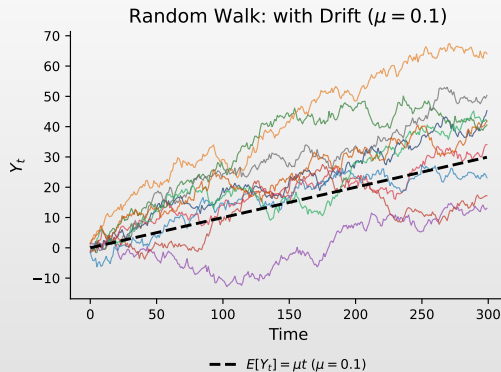
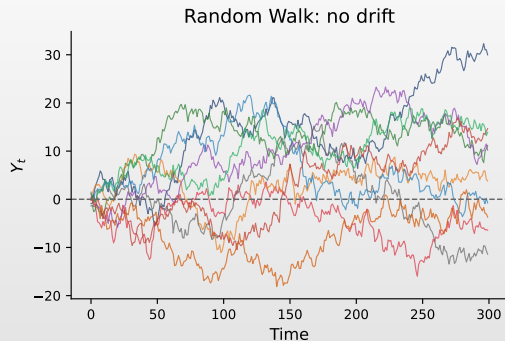


## Simularea măsurilor aleatorii

### Observații

- ▣ **Stânga:** Mersuri aleatorii pure  $\succ$  fără drift, rătăcesc imprevizibil
- ▣ **Dreapta:** Cu drift  $\succ$  trend ascendent în medie
- ▣ Fiecare traiectorie este unică — incertitudinea crește în timp

## Simularea măsurilor aleatorii

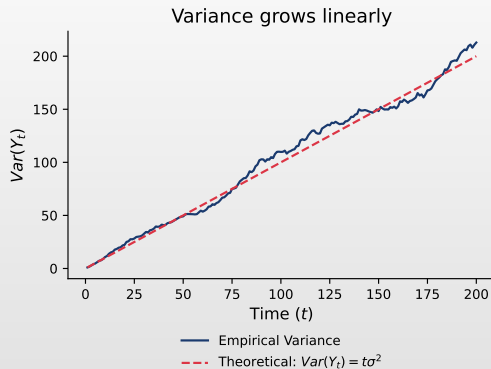
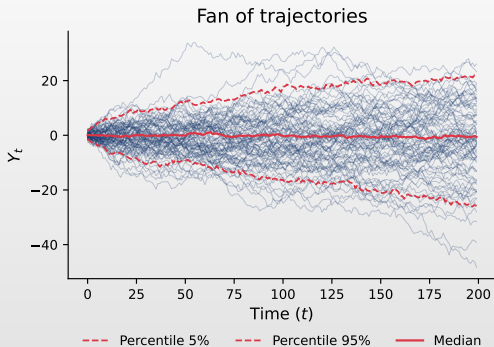


## Creșterea varianței: de ce măsurile aleatorii sunt nestaționare

### Observații

- ▣ **Stânga:** Evantaiul de traiectorii arată incertitudinea crescând
- ▣ **Dreapta:** Varianța crește liniar:  $\text{Var}(Y_t) = t\sigma^2$
- ▣ Aceasta violează staționaritatea — varianța ar trebui să fie constantă

## Creșterea varianței: de ce măsurile aleatorii sunt nestaționare



## Procese integrate

### Definiție 3 (Proces Integrat de Ordin $d$ )

- **Notăție:**  $Y_t \sim I(d)$  dacă:
  - ▶  $Y_t$  este nestaționară
  - ▶  $(1 - L)^d Y_t = \Delta^d Y_t$  este staționară
  - ▶  $(1 - L)^{d-1} Y_t$  este încă nestaționară

### Cazuri Comune

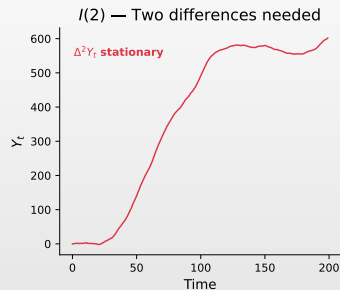
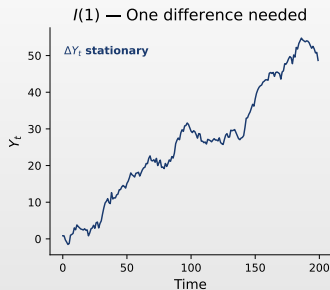
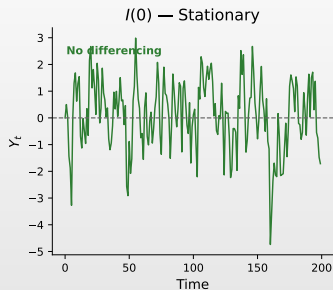
- $I(0)$ : Proces staționar (de ex., ARMA)
- $I(1)$ : Prima diferență este staționară (cel mai frecvent pentru date economice)
- $I(2)$ : A doua diferență este staționară (mai rar)

## Proces integrat: ilustrație vizuală

### Ordinul de integrare

- $I(0)$ : Staționar  $\succ$  nicio diferențiere necesară
- $I(1)$ : O diferență necesară (mers aleator)
- $I(2)$ : Două diferențe necesare
- Majoritatea seriilor economice sunt  $I(0)$  sau  $I(1)$

## Proces integrat: ilustrație vizuală



TSA\_ch3\_def\_integrated



## Operatorul diferență

### Definiție 4 (Prima Diferență)

- ▣ **Operator:**  $\Delta Y_t = Y_t - Y_{t-1} = (1 - L)Y_t$
- ▣ **Notăție:**  $L$  este operatorul lag ( $LY_t = Y_{t-1}$ )

### Diferențe de Ordin Superior

- ▣ A doua diferență:  $\Delta^2 Y_t = \Delta(\Delta Y_t) = (1 - L)^2 Y_t$
- ▣  $\Delta^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$
- ▣ Diferența de ordin  $d$ :  $\Delta^d Y_t = (1 - L)^d Y_t$

### Rezultat Cheie

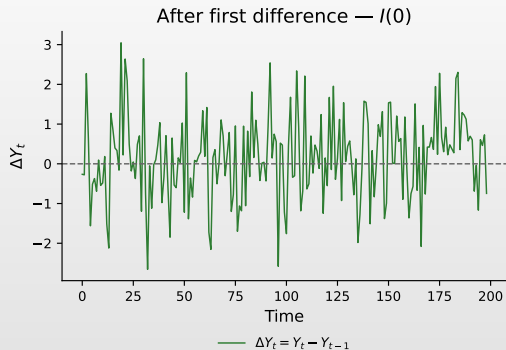
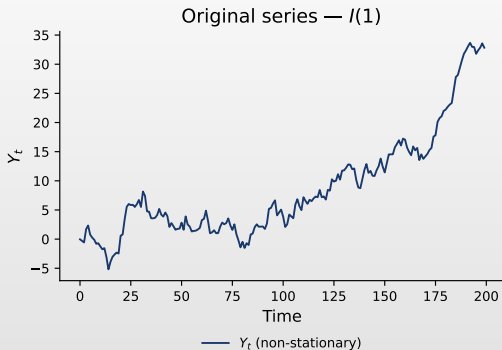
- ▣ Dacă  $Y_t \sim I(d)$ , atunci  $\Delta^d Y_t \sim I(0)$  (staționar)

## Prima diferență: ilustrație vizuală

### Observație

- ▣ **Stânga:** serie nestaționară
- ▣ **Dreapta:** după prima diferență, seria devine staționară

## Prima diferență: ilustrație vizuală



## Exemplu: diferențierea unui mers aleator

### Mers aleatoriu la Zgomot Alb

- Fie  $Y_t = Y_{t-1} + \varepsilon_t$  (mers aleator). Luând prima diferență:

$$\Delta Y_t = Y_t - Y_{t-1} = \varepsilon_t$$

- Prima diferență este zgomot alb  $\succ$  un proces staționar!

### Interpretare

- Un mers aleator este  $I(1)$
- O diferență îl transformă în  $I(0)$
- “Schimbările” într-un mers aleator sunt staționare

## Demonstrație: diferențierea induce staționaritatea

### Afirmație

- Dacă  $Y_t \sim I(1)$ , atunci  $\Delta Y_t = Y_t - Y_{t-1}$  este staționar

### Demonstrație: Mers aleatoriu cu drift $Y_t = \mu + Y_{t-1} + \varepsilon_t$

- Prima diferență:  $\Delta Y_t = Y_t - Y_{t-1} = \mu + \varepsilon_t$
- **Media:**  $\mathbb{E}[\Delta Y_t] = \mu$  (constantă) ✓
- **Varianța:**  $\text{Var}(\Delta Y_t) = \sigma^2$  (constantă) ✓
- **Autocovarianța:**  $\text{Cov}(\Delta Y_t, \Delta Y_{t-k}) = 0$  pentru  $k \neq 0$  ✓

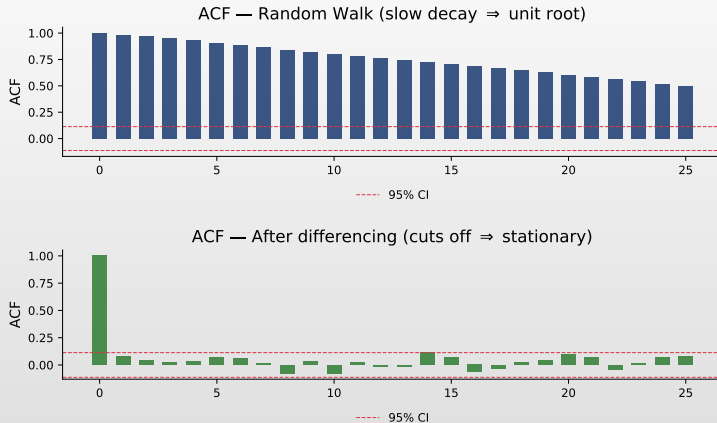
### Principiu General

- Diferențierea elimină “memoria” care face varianța să se acumuleze
- Pentru procese  $I(d)$ , sunt necesare  $d$  diferențe

## ACF: detectarea nestaționarității

- ▣ **Sus:** ACF mers aleator scade foarte lent  $\succ$  rădăcină unitate
- ▣ **Jos:** După diferențiere, ACF se anulează  $\succ$  staționar

## ACF: detectarea nestaționarității

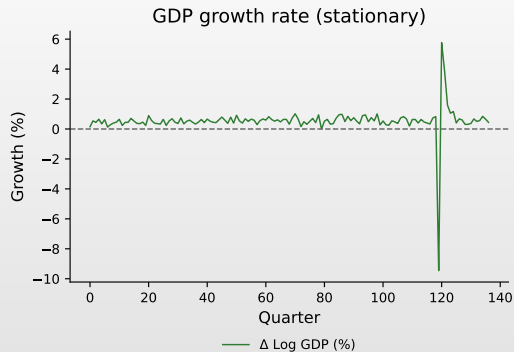
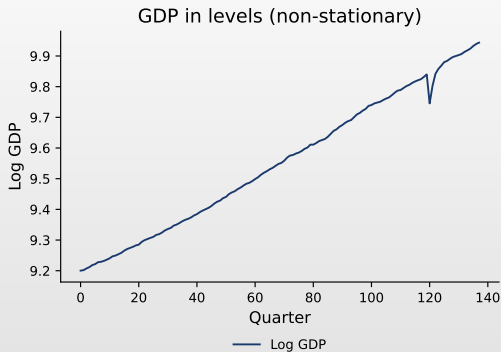


## Diferențierea în practică: exemplul PIB

- ▣ **Stânga:** PIB în valori absolute  $\succ$  trend ascendent clar (nestaționar)
- ▣ **Dreapta:** Rata de creștere PIB (diferența logaritmică)  $\succ$  fluctuează în jurul mediei (staționar)
- ▣ Diferențierea elimină trendul și obține staționaritate



## Diferențierea în practică: exemplul PIB



## Supra-diferențierea

### Avertisment: Supra-diferențierea

- Diferențierea mai mult decât este necesar introduce probleme:
  - ▶ Creează autocorelație negativă artificială
  - ▶ Inflează varianța
  - ▶ Pierde informație

### Exemplu

- Dacă  $Y_t \sim I(1)$ , atunci  $\Delta Y_t \sim I(0)$ . Dar dacă diferențiem din nou:

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = \varepsilon_t - \varepsilon_{t-1}$$

- Acesta este un MA(1) cu  $\theta_1 = -1$  (la granița non-invertibilității)!

## Definiția ARIMA

### Definiție 5 (ARIMA(p,d,q))

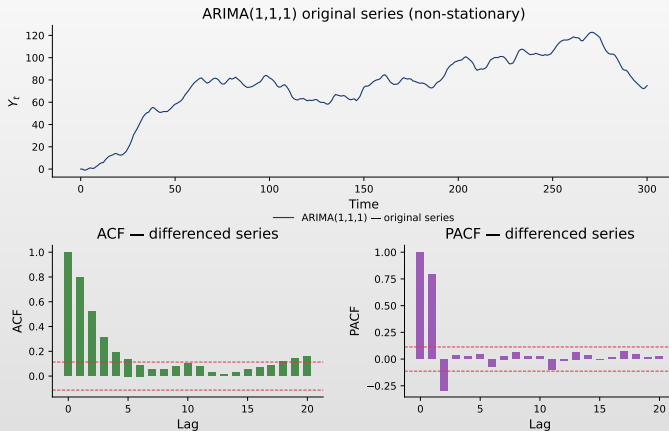
- ▣ **Model:**  $\phi(L)(1-L)^d Y_t = c + \theta(L)\varepsilon_t$
- ▣ **Polinom AR:**  $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$
- ▣ **Polinom MA:**  $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$
- ▣ **Integrare:**  $d$  este ordinul de integrare (numărul de diferențe)
- ▣ **Inovații:**  $\varepsilon_t \sim WN(0, \sigma^2)$

## ARIMA: ilustrație vizuală

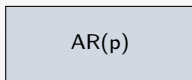
### Interpretare

- ▣ **Sus:** seria ARIMA originală (nestaționară)
- ▣ **Jos:** după diferențiere de  $d$  ori — ACF/PACF dezvăluie ordinele AR și MA

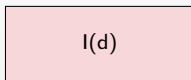
## ARIMA: ilustrație vizuală



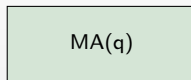
## Componentele ARIMA



Autoregresiv  
Memorie



Integrare  
Diferențiere



Medie Mobilă  
Șocuri

### Cazuri Speciale

- ▣  $ARIMA(p,0,q) = ARMA(p,q) \succ$  staționar
- ▣  $ARIMA(0,1,0) =$  Mers aleatoriu
- ▣  $ARIMA(0,1,1) = IMA(1,1) \succ$  netezire exponențială
- ▣  $ARIMA(1,1,0) = ARI(1,1) \succ AR(1)$  diferențiat

## Exemplu ARIMA(1,1,0)

### Model ARI(1,1)

- ▣  $\Delta Y_t = c + \phi_1 \Delta Y_{t-1} + \varepsilon_t$
- ▣ Echivalent:  $(1 - \phi_1 L)(1 - L)Y_t = c + \varepsilon_t$

### Interpretare

- ▣ **Schimbările** în  $Y_t$  urmează un proces AR(1)
- ▣ Dacă  $|\phi_1| < 1$ , schimbările sunt staționare
- ▣  $Y_t$  în sine are un trend stocastic
- ▣ Model comun pentru multe serii de timp economice

## Exemplu ARIMA(0,1,1)

### Model IMA(1,1)

- ▣  $\Delta Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$
- ▣ Echivalent:  $(1 - L)Y_t = c + (1 + \theta_1 L)\varepsilon_t$

### Conexiunea cu Netezirea exponențială

- ▣ Modelul IMA(1,1) este echivalent cu **netezirea exponențială simplă**:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t$$

- ▣ unde  $\alpha = 1 + \theta_1$  (pentru  $-1 < \theta_1 < 0$ )



## Rolul constantei în ARIMA

### Termenul Constant în ARIMA(p,d,q)

- ▣ Când  $d > 0$ , constanta  $c$  are o interpretare diferită:
- ▣  $\phi(L)(1-L)^d Y_t = c + \theta(L)\varepsilon_t$

### Implicații Importante

- ▣ Pentru  $d = 1$ :  $c$  reprezintă **drift-ul** (schimbarea medie):  $\mathbb{E}[\Delta Y_t] = \frac{c}{1 - \phi_1 - \dots - \phi_p}$
- ▣ Pentru  $d = 2$ :  $c$  afectează **curbura** trendului
- ▣ Adesea se presupune  $c = 0$  când  $d \geq 1$

## Testarea pentru rădăcini unitate

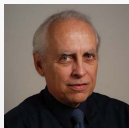
### De ce Testăm?

- ▣ **Scop:** înainte de a potrivi un model ARIMA, trebuie să determinăm:
  - ▶ Este seria staționară? (Este  $d = 0$ ?)
  - ▶ Dacă nu, câte diferențe sunt necesare? (Care este  $d$ ?)

### Teste comune de rădăcină unitate

- ▣ **Dickey-Fuller (DF)** și **Augmented Dickey-Fuller (ADF)**
- ▣ **Phillips-Perron (PP)**
- ▣ **KPSS** (test de staționaritate  $\succ$  ipoteză nulă inversată)

## Portret de cercetător: Dickey & Fuller



David Dickey (\*1945)

 [Wikipedia \(en\)](#)



Wayne Fuller (1931–2022)

 [Wikipedia \(en\)](#)

### Biografie

- ▣ **David Dickey:** statistician american la NC State University. Doctorand al lui Wayne Fuller la Iowa State
- ▣ **Wayne Fuller:** statistician american, profesor la Iowa State University
- ▣ Împreună au dezvoltat testul fundamental pentru rădăcini unitate în serii de timp

### Contribuții principale

- ▣ **Testul Dickey-Fuller** (1979) — testul fundamental pentru rădăcini unitate
- ▣ **Testul ADF** (Augmented Dickey-Fuller) — extensie cu diferențe întârziate
- ▣ **Tabele de valori critice** — distribuții non-standard sub ipoteza nulă
- ▣ Testarea riguroasă a ordinului de integrare pentru modelarea ARIMA

## Testul Dickey-Fuller

### Configurare

- ▣ Considerăm modelul AR(1):  $Y_t = \phi Y_{t-1} + \varepsilon_t$ . Scădem  $Y_{t-1}$ :
- ▣  $\Delta Y_t = (\phi - 1)Y_{t-1} + \varepsilon_t = \gamma Y_{t-1} + \varepsilon_t$ , unde  $\gamma = \phi - 1$

### Ipoteze

- ▣  $H_0: \gamma = 0$  (rădăcină unitate,  $\phi = 1$ , nestăţionar)
- ▣  $H_1: \gamma < 0$  (staţionar,  $|\phi| < 1$ )

### Problemă Cheie

- ▣ Sub  $H_0$ , statistica  $t$  **nu** urmează o distribuţie  $t$  standard!
- ▣ Trebuie folosite valorile critice Dickey-Fuller

## Variante ale testului Dickey-Fuller

### Trei Specificări

1. **Fără constantă, fără trend:**  $\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t$
2. **Cu constantă (drift):**  $\Delta Y_t = \alpha + \gamma Y_{t-1} + \varepsilon_t$
3. **Cu constantă și trend:**  $\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \varepsilon_t$

### Alegerea Specificării Corecte

- ☐ Examinați datele: au un trend vizibil?
- ☐ Includerea termenilor inutili reduce puterea
- ☐ Excluderea termenilor necesari duce la inferență incorectă

## Testul Augmented Dickey-Fuller (ADF)

### Problema cu DF Simplu

- ▣ **Limitare:** dacă există dinamică AR dincolo de AR(1), reziduurile DF vor fi autocorelate

### Definiție 6 (Testul ADF)

- ▣ **Ecuație:**  $\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{j=1}^k \delta_j \Delta Y_{t-j} + \varepsilon_t$
- ▣ **Test:**  $H_0 : \gamma = 0$  folosind valorile critice ADF

### Alegerea Lungimii Lag-ului $k$

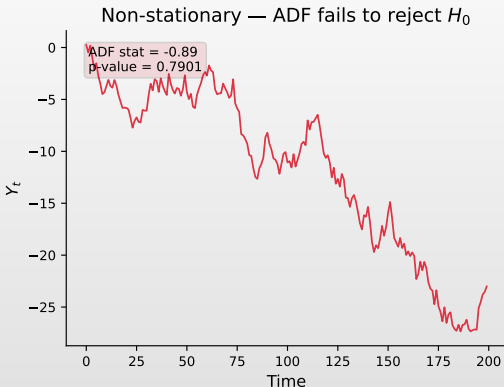
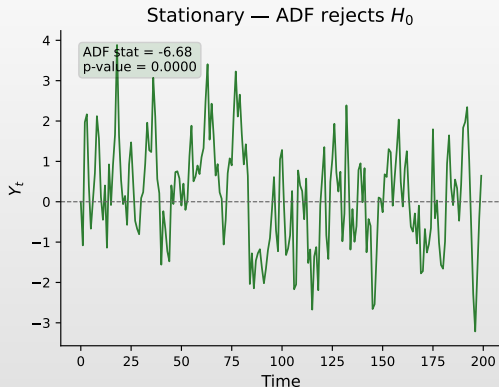
- ▣ Folosiți criterii informaționale (AIC, BIC)
- ▣ Începeți cu  $k_{max}$ , reduceți până ultimul lag este semnificativ

## Testul ADF: ilustrație vizuală

### Observație

- ▣ **Stânga:** serie staționară  $\succ$  ADF respinge rădăcina unitate
- ▣ **Dreapta:** nestaționară  $\succ$  ADF nu respinge

## Testul ADF: ilustrație vizuală





## Valori critice ADF

Model	1%	5%	10%
Fără constantă, fără trend	-2.58	-1.95	-1.62
Cu constantă	-3.43	-2.86	-2.57
Cu constantă și trend	-3.96	-3.41	-3.13

## Regula de Decizie

- Statistică de test  $<$  valoare critică  $\succ$  Respingem  $H_0$  (staționar)
- Statistică de test  $\geq$  valoare critică  $\succ$  Nu respingem (rădăcină unitate)

## Testul Phillips-Perron (PP)

### Motivație

- ▣ **Ipoteze:**  $H_0$ : Rădăcină unitate vs  $H_1$ : Staționar (ca ADF)
- ▣ **Diferență:** folosește o corecție non-parametrică pentru corelația serială
  - ▶ Nu adaugă diferențe întârziate ca ADF

### Statistică de Test

- ▣ **Formula:**  $Z_t = t_{\hat{\gamma}} \cdot \sqrt{\frac{\hat{\sigma}^2}{\hat{\lambda}^2}} - \frac{T(\hat{\lambda}^2 - \hat{\sigma}^2)(se(\hat{\gamma}))}{2\hat{\lambda}^2 \cdot s}$
- ▣ **Notăție:**  $\hat{\lambda}^2$  este estimarea consistentă a varianței pe termen lung (Newey-West)

### Avantaje față de ADF

- ▣ Robust la heteroscedasticitate și corelație serială
- ▣ Nu necesită selectarea lungimii lag-ului (folosește lățime de bandă)

## Testul KPSS

### Ipoteze Inversate

- ▣ **Spre deosebire de ADF:**  $H_0$ : Staționar vs  $H_1$ : Rădăcină unitate
  - ▶ Ipoteza nulă este inversată față de ADF/PP

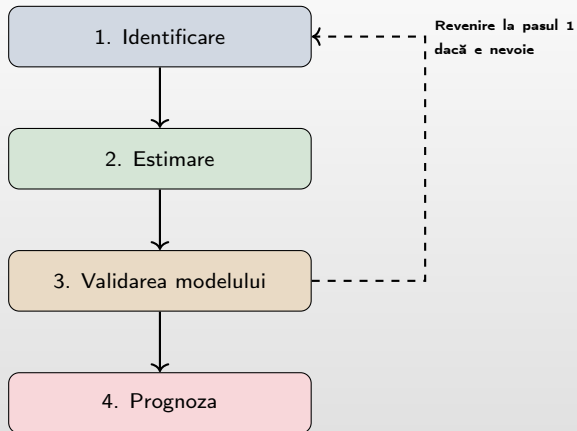
### Procedura KPSS

- ▣ **Descompunere:**  $Y_t = \xi t + r_t + \varepsilon_t$  unde  $r_t = r_{t-1} + u_t$
- ▣ **Test:** verificăm dacă  $\text{Var}(u_t) = 0$ 
  - ▶ Dacă da, componenta aleatoare  $r_t$  este constantă

### Utilizare Complementară cu ADF

- ▣ ADF respinge, KPSS nu respinge  $\succ$  Staționar
- ▣ ADF nu respinge, KPSS respinge  $\succ$  Rădăcină unitate
- ▣ Ambele resping sau niciunul  $\succ$  Neconcludent

## Metodologia Box-Jenkins



## Pasul 1: Determinarea lui $d$

### Procedură

1. Reprezentați grafic seria de timp  $\succ$  căutați trenduri, varianță în schimbare
2. Examinați ACF  $\succ$  descreștere lentă sugerează nestaționaritate
3. Aplicați teste de rădăcină unitate (ADF, KPSS)
4. Dacă nestaționară, diferențiați și repetați

### Ghiduri Practice

- ▣ Majoritatea seriilor economice:  $d = 1$  este suficient
- ▣ Rar avem nevoie de  $d > 2$
- ▣ Dacă ACF al  $\Delta Y_t$  tot scade lent, încercați  $d = 2$
- ▣ Atenție la supra-diferențiere (ACF cu  $\rho_1 \approx -0.5$ )

## Pasul 2: Determinarea lui $p$ și $q$

### După Diferențiere

▣ **Principiu:** odată ce  $W_t = \Delta^d Y_t$  este staționar, folosiți ACF/PACF pentru a identifica ARMA( $p, q$ )

Model	ACF	PACF
AR( $p$ )	Scade exponențial	Se anulează după lag $p$
MA( $q$ )	Se anulează după lag $q$	Scade exponențial
ARMA( $p, q$ )	Scade	Scade

### Criterii informaționale

- ▣ **Când:** tiparele ACF/PACF sunt neclare
- ▣ **AIC:**  $-2 \ln(L) + 2k$ ; **BIC:**  $-2 \ln(L) + k \ln(n)$  ( $L$  = verosimilitate,  $k$  = parametri,  $n$  = eșantion)
- ▣ Mai mic este mai bun; BIC penalizează complexitatea mai mult

## Algoritmi Auto-ARIMA

### Selecție automată a modelului

- ▣ Software-ul modern poate selecta automat  $(p, d, q)$ :
  - ▶ Python: `pmdarima.auto_arima()`    R: `forecast::auto.arima()`

### Cum funcționează Auto-ARIMA

1. Folosește teste de rădăcină unitate pentru a determina  $d$
2. Potrivește modele pentru diverse combinații  $(p, q)$
3. Selectează modelul cu cel mai mic AIC/BIC
4. Opțional folosește căutare pas cu pas pentru eficiență

**Atenție:** Selecția automată este utilă dar nu garantată — verificați întotdeauna validitatea modelului!

## Metode de estimare

### Estimarea prin metoda verosimilității maxime (MLE)

- ▣ Abordarea standard pentru ARIMA:
  - ▶ Presupune  $\varepsilon_t \sim N(0, \sigma^2)$
  - ▶ Maximizează funcția de verosimilitate
  - ▶ Oferă estimatori consistenți, eficienți
  - ▶ Furnizează erori standard pentru inferență

### MLE Condiționată vs Exactă

- ▣ **MLE Condiționată:** Condiționează pe valorile inițiale
- ▣ **MLE Exactă:** Tratează valorile inițiale ca necunoscute
- ▣ Diferența diminuează pe măsură ce dimensiunea eșantionului crește



## Log-verosimilitatea condiționată

### Funcția de log-verosimilitate gaussiană

- ▣  $\ell(\boldsymbol{\theta}, \sigma^2) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T e_t^2(\boldsymbol{\theta})$
- ▣  $e_t(\boldsymbol{\theta}) = X_t - \hat{X}_{t|t-1}$  sunt **erorile de predicție la un pas**
- ▣  $\boldsymbol{\theta} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, c)$

### Exemplu: ARIMA(1,1,1)

- ▣ Erorile de predicție:  $e_t = \Delta X_t - \phi_1 \Delta X_{t-1} - \theta_1 e_{t-1} - c$
- ▣ MLE condiționată: fixează  $e_0 = 0$ , calculează  $e_1, \dots, e_T$ , maximizează  $\ell$

### Estimarea lui $\sigma^2$

- ▣ La parametrii optimi  $\hat{\boldsymbol{\theta}}$ :  $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T e_t^2(\hat{\boldsymbol{\theta}})$

## Restricții asupra parametrilor

### Staționaritate și invertibilitate

- ▣ Modelul ARIMA estimat ar trebui să satisfacă:
  - ▶ **Staționaritate AR:** Rădăcinile lui  $\phi(z) = 0$  în afara cercului unitate
  - ▶ **Invertibilitate MA:** Rădăcinile lui  $\theta(z) = 0$  în afara cercului unitate

### Verificare în Practică

- ▣ Majoritatea software-ului raportează:
  - ▶ Coeficienți estimați cu erori standard
  - ▶ Rădăcinile polinoamelor AR și MA
  - ▶ Avertisment dacă este detectată aproape-rădăcină-unitate

## Diagnosticul modelului ARIMA

### Verificări esențiale (aceleași ca pentru ARMA, cf. Cap. 2)

- Dacă modelul este corect, reziduurile  $\hat{\varepsilon}_t$  trebuie să fie zgomot alb:
  1. **ACF/PACF rezidual**: fără vârfuri semnificative
  2. **Testul Ljung-Box**:  $p\text{-value} > 0.05$   $\succ$  fără autocorelare
  3. **Graficul Q-Q**: verificarea normalității
  4. **Heteroscedasticitate**: varianță constantă a reziduurilor

### Aspecte specifice ARIMA

- Testul Ljung-Box: alegeți  $m \approx \ln(n)$  sau  $m = 10$  (trimestrial),  $m = 20$  (lunar)
- Grade de libertate:  $\chi^2(m - p - q)$ , ajustate pentru  $p$  și  $q$  estimați
- Dacă testul eșuează: adăugați termeni AR/MA sau verificați rupturi structurale

## Testul Ljung-Box

### Definiție 7 (Statistica Q Ljung-Box)

$Q(m) = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k}$ . Sub  $H_0$  (fără autocorelare):  $Q(m) \sim \chi^2(m-p-q)$

### Utilizare

- ▣ Alegeți  $m \approx \ln(n)$  sau  $m = 10$  trimestrial,  $m = 20$  lunar
- ▣ Gradele de libertate se ajustează pentru parametrii estimați
- ▣ Respingeți dacă  $Q(m)$  depășește valoarea critică

### Dacă testul eșuează

Luați în considerare adăugarea de termeni AR sau MA, sau verificați rupturile structurale.

## Proгноze punctuale

### Proгноза cu MSE Minim

- ▣ Prognoză optimă la  $h$  pași:  $\hat{Y}_{T+h|T} = \mathbb{E}[Y_{T+h} | Y_T, Y_{T-1}, \dots]$

### Proгноза ARIMA(1,1,1)

- ▣ **Model:**  $(1 - \phi_1 L)(1 - L)Y_t = c + (1 + \theta_1 L)\varepsilon_t$
- ▣ **Un pas:**  $\hat{Y}_{T+1|T} = c + Y_T + \phi_1(Y_T - Y_{T-1}) + \theta_1 \hat{\varepsilon}_T$
- ▣ **Multi-pas:** înlocuiți  $\varepsilon_{T+j}$  cu 0,  $Y_{T+j}$  cu  $\hat{Y}_{T+j|T}$

## Intervale de prognoză

### Incertitudinea prognozei

- ▣ Varianța erorii la  $h$  pași:  $\text{Var}(e_{T+h}) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$ , unde  $\psi_j$  sunt coeficienții  $\text{MA}(\infty)$

### Intervale de încredere

- ▣ Sub normalitate, interval  $(1 - \alpha)\%$ :  $\hat{Y}_{T+h|T} \pm z_{\alpha/2} \sqrt{\text{Var}(e_{T+h})}$

### Proprietate Cheie pentru Serii I(1)

- ▣ Pentru procese integrate, varianța prognozei crește nelimitat când  $h \rightarrow \infty$
- ▣ Intervalele se largesc în timp!

## Proгноze pe termen lung pentru ARIMA

### Comportament când $h \rightarrow \infty$

- ▣ **Cu drift**  $c$ : Proгноze punctuale  $\succ$  trend liniar; IC  $\succ$  lăţimea creşte cu  $\sqrt{h}$
- ▣ **Fără drift**: Proгноze punctuale  $\succ$  converg la ultima valoare observată; IC  $\succ$  tot cresc nelimitat

### Implicaţie Practică

- ▣ Proгноzele ARIMA sunt cele mai fiabile pentru orizonturi scurte
- ▣ Proгноzele pe termen lung au benzi de incertitudine foarte largi

## Proгноза rolling: concept

### Ce este Proгноза Rolling?

- ▣ Tehnică pentru evaluarea acurateții prognozei în afara eşantionului:
  1. Fixăm o **fereastră de antrenament** de dimensiune  $w$
  2. Estimăm modelul pe observațiile  $t = 1, \dots, w$
  3. Prognozăm  $h$  pași înainte:  $\hat{Y}_{w+h|w}$
  4. **Deplasăm** fereastra înainte cu o perioadă
  5. Repetăm până la sfârșitul eşantionului

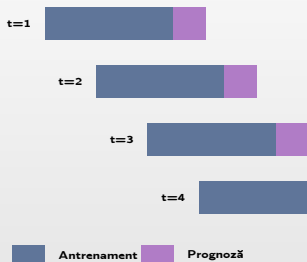
### De ce Proгноze Rolling?

- ▣ Mimează scenariul de prognoză în timp real
- ▣ Oferă multiple erori de prognoză pentru evaluare
- ▣ Evită supraajustarea pe întregul eşantion

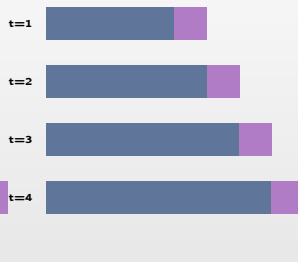


## Fereastră fixă vs expandabilă

### Fereastră Fixă (Rolling)



### Fereastră Expandabilă



Time → Time →

### Comparație

- **Fixă:** Fereastra alunecă înainte, dimensiune constantă — se adaptează la schimbări de regim
- **Expandabilă:** Fereastra crește în timp — folosește toate datele istorice

## Proгноză 1-pas vs multi-pas

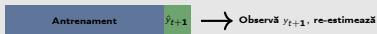
### 1-Pas Înainte (Recursiv)

- Prognozează doar perioada următoare
  - ▶ Re-estimează modelul după fiecare pas
  - ▶ Folosește valoarea reală odată observată
- Cea mai precisă pentru orizonturi scurte

### Multi-Pas (Direct)

- Prognozează mai multe perioade înainte
  - ▶ Fără re-estimare între pași
  - ▶ Folosește valori prognozate ca input
- Incertitudinea se acumulează cu orizontul

#### 1-Pas Înainte



#### Multi-Pas (h=3)



## Proгноза rolling: exemplu pas cu pas

Configurare: ARIMA(1,1,0) cu  $\phi_1 = 0.6$

▣ Model:  $\Delta Y_t = \phi_1 \Delta Y_{t-1} + \varepsilon_t$  unde  $\Delta Y_t = Y_t - Y_{t-1}$

Date la Momentul  $T$

▣  $Y_{T-2} = 100, Y_{T-1} = 103, Y_T = 108 \succ \Delta Y_{T-1} = 3, \Delta Y_T = 5$

Proгноза Punctuală la 1 Pas

$$\begin{aligned}\hat{\Delta Y}_{T+1|T} &= \phi_1 \cdot \Delta Y_T = 0.6 \times 5 = 3 \\ \hat{Y}_{T+1|T} &= Y_T + \hat{\Delta Y}_{T+1|T} = 108 + 3 = \boxed{111}\end{aligned}$$

## Proгноze punctuale multi-pas

### Proгноza la 2 Pași

$$\begin{aligned}\Delta \hat{Y}_{T+2|T} &= \phi_1 \cdot \Delta \hat{Y}_{T+1|T} = 0.6 \times 3 = 1.8 \\ \hat{Y}_{T+2|T} &= \hat{Y}_{T+1|T} + \Delta \hat{Y}_{T+2|T} = 111 + 1.8 = \boxed{112.8}\end{aligned}$$

### Formula Generală pentru Proгноza la $h$ Pași (ARIMA(1,1,0))

$$\begin{aligned}\Delta \hat{Y}_{T+h|T} &= \phi_1^h \cdot \Delta Y_T \\ \hat{Y}_{T+h|T} &= Y_T + \Delta Y_T \cdot \frac{\phi_1(1 - \phi_1^h)}{1 - \phi_1}\end{aligned}$$

### Numeric: Proгноza la 3 Pași

$$\hat{Y}_{T+3|T} = 108 + 5 \times \frac{0.6(1-0.6^3)}{1-0.6} = 108 + 5 \times 1.176 = \boxed{113.88}$$

## Intervale de încredere: formule

### Varianța erorii de prognoză

- ▣ Pentru ARIMA(1,1,0) la  $h$  pași:  $\text{Var}(e_{T+h|T}) = \sigma^2 \left( 1 + \sum_{j=1}^{h-1} \psi_j^2 \right)$
- ▣  $\psi_j = 1 + \phi_1 + \dots + \phi_1^j = \frac{1 - \phi_1^{j+1}}{1 - \phi_1}$  pentru  $j \geq 0$

### Interval de Încredere $(1 - \alpha)\%$

- ▣  $\hat{Y}_{T+h|T} \pm z_{\alpha/2} \cdot \sqrt{\text{Var}(e_{T+h|T})}$
- ▣ Pentru IC 95%:  $z_{0.025} = 1.96$

## Interval de încredere: exemplu numeric

Date:  $\sigma^2 = 4$ ,  $\phi_1 = 0.6$ ,  $\hat{Y}_{T+1|T} = 111$

## IC la 1 Pas

$$\text{Var}(e_{T+1|T}) = \sigma^2 = 4$$

$$\text{IC } 95\% = 111 \pm 1.96 \times \sqrt{4} = 111 \pm 3.92 = [107.08, 114.92]$$

IC la 2 Pași (pentru  $\hat{Y}_{T+2|T} = 112.8$ )

$$\psi_1 = 1 + \phi_1 = 1.6, \quad \text{Var}(e_{T+2|T}) = 4(1 + 1.6^2) = 14.24$$

$$\text{IC } 95\% = 112.8 \pm 1.96 \times \sqrt{14.24} = 112.8 \pm 7.40 = [105.40, 120.20]$$

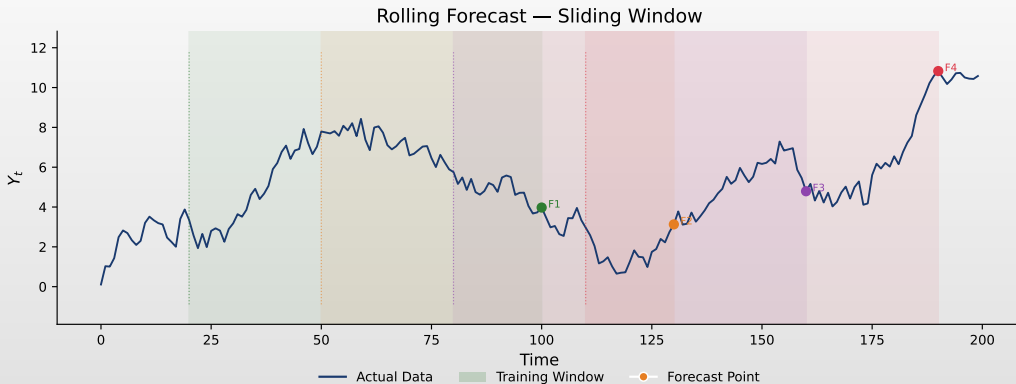
## Notă

- IC se lărgeste pe măsură ce orizontul de predicție crește!

## Ilustrație fereastră rolling

- Fiecare fereastră produce o prognoză la 1 pas
- Comparăm prognozele cu valorile reale pentru a calcula RMSE, MAE
- Fereastra rolling menține estimarea modelului actualizată

## Ilustrație fereastră rolling





## Proгноза Rolling: Cod Python

### Implementare

```
from statsmodels.tsa.arima.model import ARIMA

window_size = 100
forecasts, actuals = [], []

for t in range(window_size, len(y) - 1):
    train = y[:t] # Fereastra rolling
    model = ARIMA(train, order=(1,1,0)).fit()
    forecast = model.forecast(steps=1)[0]
    forecasts.append(forecast)
    actuals.append(y[t])

rmse = np.sqrt(np.mean((np.array(forecasts) - np.array(actuals))**2))
```

## Studiu de caz: analiză ARIMA completă

### Obiectiv

- Prognoză PIB Real al SUA folosind metodologia Box-Jenkins

### Etape

1. **Pasul 1:** vizualizarea datelor și verificarea staționarității
2. **Pasul 2:** Aplicarea testelor de rădăcină unitate (ADF, KPSS)
3. **Pasul 3:** Diferențiere dacă e necesar, identificare  $p$  și  $q$
4. **Pasul 4:** Estimarea modelului ARIMA
5. **Pasul 5:** Diagnosticul modelului
6. **Pasul 6:** Generarea prognozelor cu intervale de încredere
7. **Pasul 7:** Evaluarea acurateții prognozei

### Date

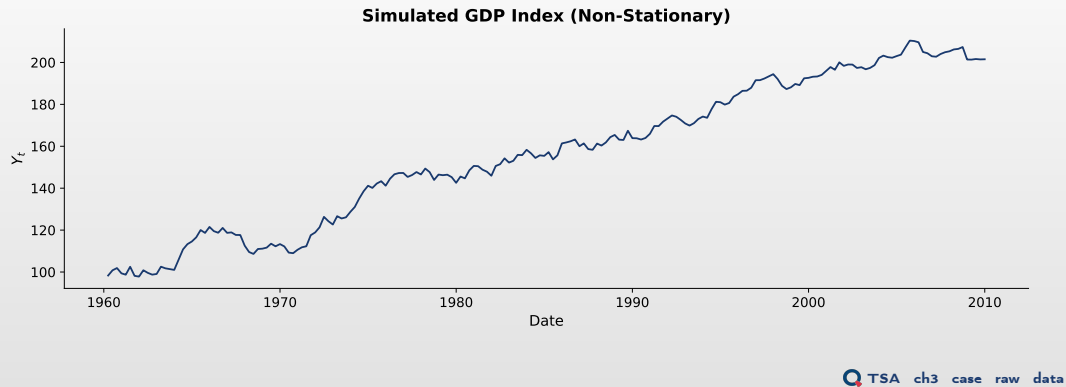
- PIB Real SUA (FRED: GDPC1), Trimestrial, 1990T1–2024T2,  $n = 138$

## Pasul 1: analiza inițială a datelor

### Observații

- ▣ Trend ascendent clar  $\succ$  medie neconstantă
- ▣ Scădere notabilă în 2020 (COVID-19)
- ▣ **Concluzie:** nestaționar, necesită diferențiere

## Pasul 1: analiza inițială a datelor



## Pasul 2: testarea rădăcinii unitate

### Test ADF pe Log PIB (serie originală)

- ▣ Statistică test:  $-0.91$
- ▣ Valori critice:  $-3.48$  (1%),  $-2.88$  (5%),  $-2.58$  (10%)
- ▣ p-value:  $0.79$
- ▣ **Rezultat:** Nu putem respinge  $H_0$   $\succ$  **Rădăcină unitate prezentă**

### Test ADF pe Prima Diferență (Rata de Creștere)

- ▣ Statistică test:  $-13.24$
- ▣ p-value:  $< 0.001$
- ▣ **Rezultat:** Respingem  $H_0$  la 1%  $\succ$  **Staționar după diferențiere**

### Concluzie

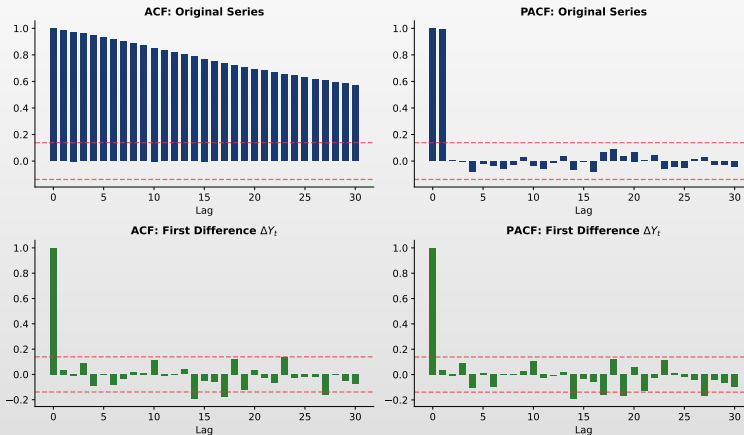
- ▣ PIB este  $I(1)$   $\succ$  Folosim  $d = 1$  în modelul ARIMA

## Pasul 3: Identificarea modelului prin ACF/PACF

### Analiza seriei diferențiate

- ▣ **ACF:** Vârf la lag 1, apoi se anulează — sugerează MA(1)
- ▣ **PACF:** Vârf la lag 1, scade — sugerează AR(1)
- ▣ **Candidate:** ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1)

## Pasul 3: Identificarea modelului prin ACF/PACF



## Pasul 4: Estimarea modelului

## Compararea modelelor folosind Criterii informaționale

Model	AIC	BIC	Log-Lik
ARIMA(1,1,0)	-725.2	-719.5	364.6
ARIMA(0,1,1)	-724.8	-719.2	364.4
<b>ARIMA(1,1,1)</b>	<b>-747.0</b>	<b>-738.5</b>	<b>376.5</b>

## Model Selectat: ARIMA(1,1,1)

$$(1 - 0.35L)(1 - L)Y_t = (1 + 0.58L)\varepsilon_t, \quad \hat{\sigma}^2 = 0.000156$$

- ▣  $\hat{\phi}_1 = 0.35$  (SE = 0.09), semnificativ la 1%
- ▣  $\hat{\theta}_1 = 0.58$  (SE = 0.08), semnificativ la 1%

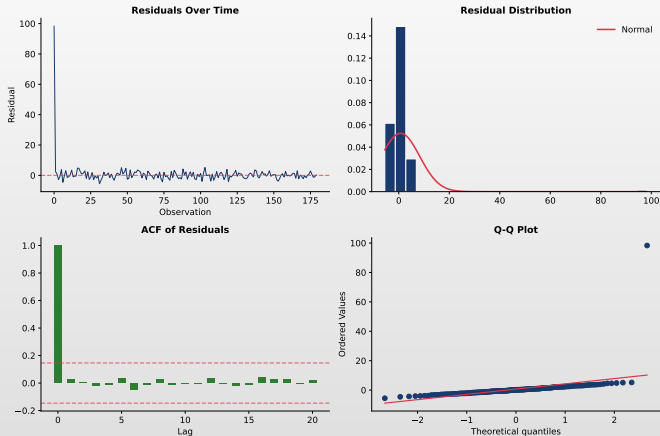


## Pasul 5: diagnosticul modelului

### Analiza reziduurilor

- Ljung-Box:  $Q(10) = 5.8$ ,  $p\text{-value} = 0.83$  — fără autocorelare
- JB: 156.4,  $p < 0.001$  — non-normal (outlier COVID)
- **Concluzie:** trece verificările de autocorelare

## Pasul 5: diagnosticul modelului



## Pasul 6: Prognoza cu intervale de încredere

### Ultimele Valori Observate (Log PIB)

- ▣  $Y_T = 9.973$  (2024T2),  $Y_{T-1} = 9.956$  (2024T1)
- ▣  $\Delta Y_T = 0.017$ ,  $\hat{\varepsilon}_T = 0.004$

### Prognoza la 1 Pas (2024T3)

$$\Delta \hat{Y}_{T+1} = \hat{\phi}_1 \Delta Y_T + \hat{\theta}_1 \hat{\varepsilon}_T = 0.35(0.017) + 0.58(0.004) = 0.0083$$

$$\hat{Y}_{T+1} = 9.973 + 0.0083 = \boxed{9.981}$$

### Interval de Încredere 95%

- ▣  $IC = 9.981 \pm 1.96 \times \sqrt{0.000156} = [9.957, 10.006]$
- ▣ În valori absolute: Prognoză PIB = \$21,652 mld, IC = [\$21,142 mld, \$22,175 mld]

## Pasul 7: Prognoză Rolling cu Train/Val/Test

Prognoză Rolling 1-Pas Înainte (Fereastră Expandabilă, IC 95%)

Train 70% → Val 15% → Test 15% | Fereastră expandabilă re-estimează modelul la fiecare pas

## Pasul 7: Prognoză Rolling cu Train/Val/Test

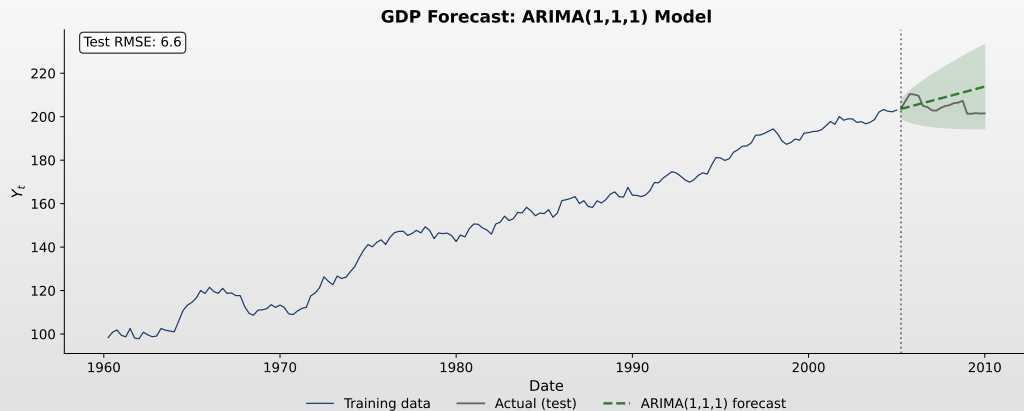


## Pasul 8: Evaluarea prognozei

### Performanță out-of-sample (ultimele 12 trimestre)

- $RMSE = 0.0486 \approx 4.86\%$  eroare
- $MAE = 0.0430 \approx 4.30\%$  eroare
- Acuratețe direcție = 91% — a prezis corect creștere/scădere

## Pasul 8: Evaluarea prognozei



## Implementare Python

### Biblioteci Cheie

```
import pandas as pd
import numpy as np
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.stattools import adfuller, kpss
import pmdarima as pm
```

### Exemplu Auto-ARIMA

```
model = pm.auto_arima(y, start_p=0, start_q=0,
                      max_p=3, max_q=3, d=None,
                      seasonal=False, trace=True)
print(model.summary())
```



## Exercițiu AI: Gândire critică

Prompt de testat în ChatGPT / Claude / Copilot

“Am o serie de timp cu PIB-ul trimestrial al unei țări (80 observații). Testează dacă este staționară, diferențiaz-o dacă e nevoie, estimează un model ARIMA și prognozează 8 trimestre. Vreau cod Python complet cu grafice.”

**Exercițiu:**

1. Rulați prompt-ul într-un LLM la alegere și analizați critic răspunsul.
2. Testează staționaritatea cu ADF *înainte* de a estima ARIMA? Folosește și KPSS?
3. Cum determină ordinul de diferențiere  $d$ ? Verifică supra-diferențierea?
4. Cum alege ordinele  $p$  și  $q$ ? Folosește ACF/PACF sau doar `auto_arima`?
5. Intervalele de încredere se lărgesc cu orizontul? (proprietate cheie  $I(1)$ )

**Atenție:** Codul generat de AI poate rula fără erori și arăta profesional. *Asta nu înseamnă că e corect.*

## Rezumat

### Ce am învățat în acest capitol

- Nestaționaritatea în seriile de timp
  - ▶ Trend determinist vs stochastic; consecințe asupra inferenței statistice
- Diferențierea și procesele integrate
  - ▶  $\Delta Y_t = Y_t - Y_{t-1}$ ; dacă  $Y_t \sim I(d)$ , atunci  $\Delta^d Y_t \sim I(0)$
- Modele ARIMA( $p, d, q$ ) și teste de rădăcină unitate
  - ▶ ADF, PP, KPSS; Box-Jenkins: identificare  $\rightarrow$  estimare  $\rightarrow$  validare
- Prognoze cu intervale de încredere
  - ▶ Pentru  $I(1)$ : IC se largesc nelimitat cu orizontul ( $\propto \sqrt{h}$ )

### Idee cheie

- **Diferențiați cu atenție:** O diferență este de obicei suficientă ( $d = 1$ ). Supra-diferențierea creează autocorelație artificială.

## Quiz rapid

### Verificați-vă cunoștințele

1. Ce tipuri de nestaționaritate cunoașteți și cum se tratează fiecare?
2. De ce varianța unui mers aleator crește în timp?
3. Care este diferența dintre testele ADF și KPSS?
4. Ce se întâmplă dacă supra-diferențiem o serie?
5. De ce intervalele de încredere ARIMA se largesc nelimitat pentru serii  $I(1)$ ?

## Răspunsuri quiz

### Răspunsuri

1. **Tipuri:** Trend determinist (regresie); trend stochastic/rădăcină unitate (diferențiere)
2. **Variantă:**  $Y_t = \sum_{i=1}^t \varepsilon_i \Rightarrow \text{Var}(Y_t) = t\sigma^2$  (șocurile se acumulează)
3. **ADF vs KPSS:** ADF:  $H_0$  = rădăcină unitate; KPSS:  $H_0$  = staționar. Se folosesc complementar.
4. **Supra-diferențiere:** Creează autocorelație negativă artificială ( $\rho_1 \approx -0.5$ ); MA(1) non-invertibil
5. **IC nelimitat:**  $\text{Var}(e_{T+h}) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \rightarrow \infty$  deoarece  $\psi_j$  nu se amortizează

## Ce urmează?

### Capitolul 4: Modele SARIMA pentru Date Sezoniere

- ▣ **Sezonalitatea:** tipare repetitive la intervale regulate
- ▣ **Diferențierea sezonieră:** operatorul  $(1 - L^s)$
- ▣ **SARIMA( $p, d, q$ )( $P, D, Q$ )<sub>s</sub>:** extensia sezonieră a ARIMA
- ▣ **Identificarea modelului:** ACF/PACF sezoniere
- ▣ **Studiu de caz:** Prognoza pasagerilor aerieni

Întrebări?

## Întrebarea 1

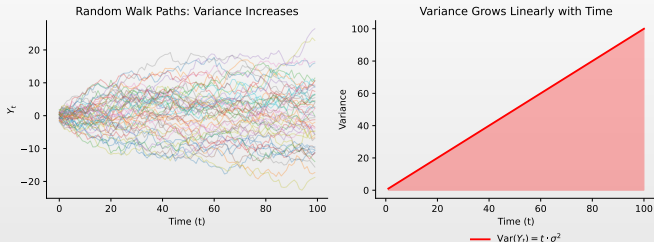
### Întrebare

□ O serie de timp  $Y_t$  urmează un mers aleator:  $Y_t = Y_{t-1} + \varepsilon_t$ . Care este  $\text{Var}(Y_t)$ ?

### Variante de răspuns

- (A)  $\sigma^2$  (constantă)
- (B)  $t \cdot \sigma^2$  (crește liniar în timp)
- (C)  $\sigma^2/t$  (scade în timp)
- (D)  $\sigma^{2t}$  (crește exponențial)

## Întrebarea 1: Răspuns



Răspuns Corect: (B)  $\text{Var}(Y_t) = t \cdot \sigma^2$

☒ Varianța mersului aleatoriu crește liniar în timp  $\succ$  de aceea măsurile aleatorii sunt nestaționare

## Întrebarea 2

### Întrebare

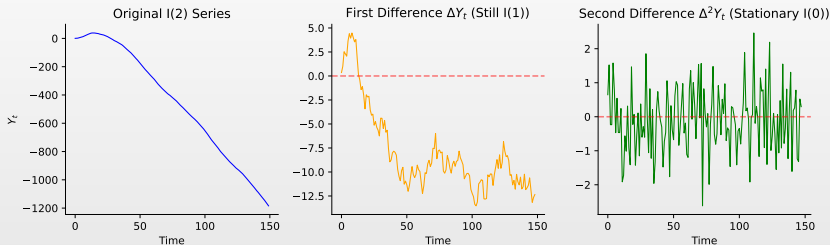
□ Dacă o serie  $Y_t$  este  $I(2)$ , de câte ori trebuie diferențiată pentru a atinge staționaritatea?

### Variante de răspuns

- (A) 0 ori (deja staționară)
- (B) 1 dată
- (C) 2 ori
- (D) Nu poate fi făcută staționară prin diferențiere



## Întrebarea 2: Răspuns



Răspuns Corect: (C) 2 ori

□  $I(d)$  înseamnă “integrată de ordin  $d$ ”  $\succ$  necesită  $d$  diferențe pentru staționaritate

## Întrebarea 3

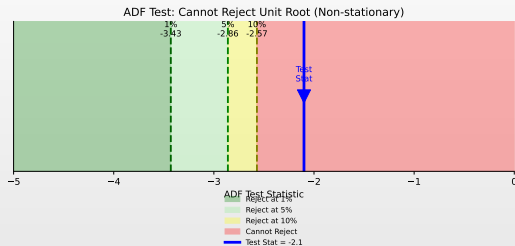
### Întrebare

- ☐ Rulați un test ADF și obțineți o statistică de test de  $-2.1$  cu valori critice:  $-3.43$  (1%),  $-2.86$  (5%),  $-2.57$  (10%). Ce concluzie trageți?

### Variante de răspuns

- (A) Respingem  $H_0$ : seria este staționară la toate pragurile de semnificație
- (B) Respingem  $H_0$ : seria este staționară doar la pragul de 10%
- (C) Nu respingem  $H_0$ : seria probabil are rădăcină unitate
- (D) Testul este neconcludent

### Întrebarea 3: Răspuns



Răspuns Corect: (C) Nu respingem  $H_0$

- Statistică de test  $-2.1 > -2.57$  (VC 10%)  $\succ$  Nu putem respinge la niciun prag de semnificație
- Luați în considerare diferențierea

## Întrebarea 4

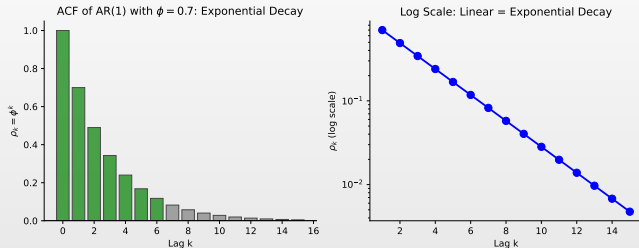
### Întrebare

□ Pentru un model ARIMA(1,1,0), care este tiparul ACF al seriei **diferențiate**  $\Delta Y_t$ ?

### Variante de răspuns

- (A) Se anulează după lag 1  
(B) Scade exponențial  
(C) Alternează în semn  
(D) Este zero la toate lag-urile

## Întrebarea 4: Răspuns



Răspuns Corect: (B) Scade exponențial

□ ARIMA(1,1,0)  $\succ$   $\Delta Y_t$  urmează AR(1) cu ACF  $\rho_k = \phi_1^k$  (descreștere geometrică)

TSA\_ch3\_quiz4\_acf\_decay

## Întrebarea 5

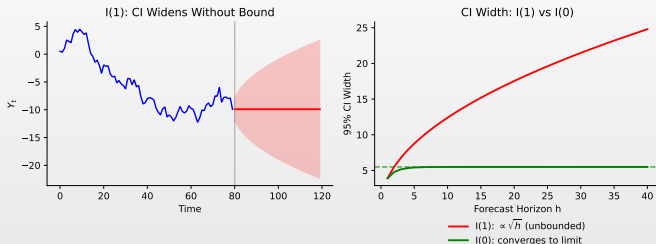
### Întrebare

- ☐ Ce se întâmplă cu intervalele de încredere ale prognozei ARIMA pe măsură ce orizontul  $h$  crește pentru o serie  $I(1)$ ?

### Variante de răspuns

- (A) Rămân constante
- (B) Se îngustează (mai multă precizie)
- (C) Se lărgesc nelimitat
- (D) Se lărgesc dar converg la o limită

## Întrebarea 5: Răspuns



Răspuns Corect: (C) Se largesc nelimitat

- Pentru  $I(1)$ : lăţimea IC  $\propto \sqrt{h}$  (nelimitată)
- Pentru  $I(0)$ : IC converg la o limită

## Bibliografie I

### Teste de rădăcină unitate

- Dickey, D.A., & Fuller, W.A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *JASA*, 74(366), 427–431.
- Phillips, P.C.B., & Perron, P. (1988). Testing for a Unit Root in Time Series Regression, *Biometrika*, 75(2), 335–346.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., & Shin, Y. (1992). Testing the Null Hypothesis of Stationarity, *Journal of Econometrics*, 54(1-3), 159–178.

### Modele ARIMA și selecție automată

- Box, G.E.P., & Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day.
- Hyndman, R.J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R, *Journal of Statistical Software*, 27(3), 1–22.



## Bibliografie II

### Manuale și referințe suplimentare

- ▣ Hamilton, J.D. (1994). *Time Series Analysis*, Princeton University Press.
- ▣ Shumway, R.H., & Stoffer, D.S. (2017). *Time Series Analysis and Its Applications*, 4th ed., Springer.
- ▣ Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*, 3rd ed., OTexts.

### Resurse online și cod

- ▣ **Quantlet**: <https://quantlet.com> > Platformă de cod pentru statistică
- ▣ **Quantinar**: <https://quantinar.com> > Platformă de învățare metode cantitative
- ▣ **GitHub TSA**: [https://github.com/QuantLet/TSA/tree/main/TSA\\_ch3](https://github.com/QuantLet/TSA/tree/main/TSA_ch3) > Cod Python pentru acest capitol

Întrebări?

Vă Mulțumim!

Întrebări?

Materialele cursului sunt disponibile la: <https://danpele.github.io/Time-Series-Analysis/>

 Quantlet

 Quantinar