



Analiza și Prognoză Seriilor de Timp

## Capitolul 3: Modele ARIMA

Serii de Timp Nestaționare

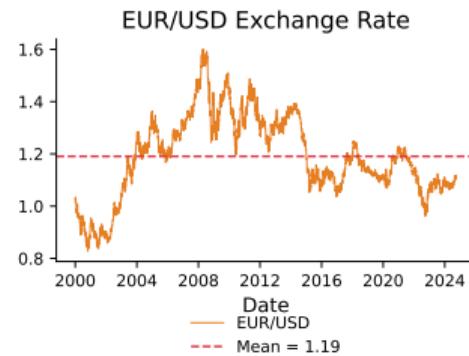
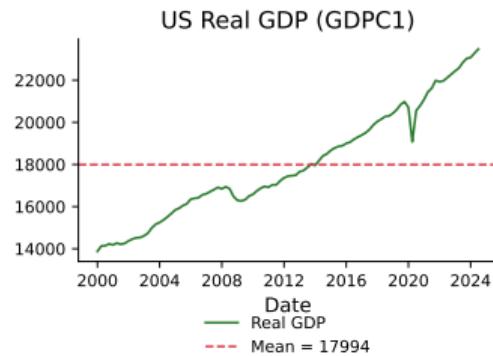


# Structura Cursului

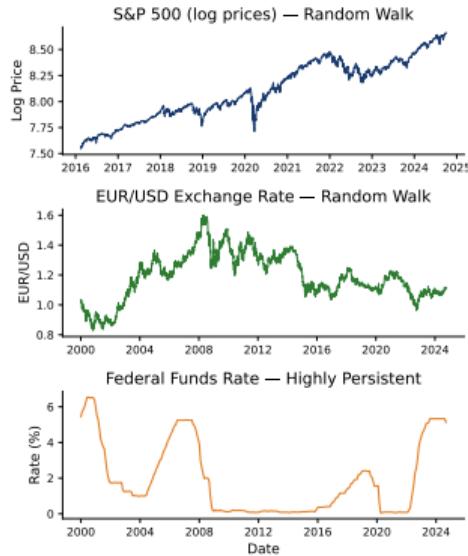
- 1 Nestaționaritatea în Seriile de Timp
- 2 Diferențierea și Operatorul Diferență
- 3 Modele ARIMA( $p,d,q$ )
- 4 Teste de Rădăcină Unitate
- 5 Identificarea Modelului ARIMA
- 6 Estimarea ARIMA
- 7 Verificare Diagnostic
- 8 Prognoză cu ARIMA
- 9 Studiu de Caz: Prognoză PIB SUA
- 10 Studiu de Caz: Date Reale
- 11 Rezumat

# Exemplu Motivațional: Datele Nestaționare Sunt Pretutindeni

Non-stationary data: sample mean is meaningless



- Prețurile acțiunilor, PIB, cursurile de schimb prezintă **tenduri sau comportament rătăcitor**
- Media din eșantion (linia roșie) este lipsită de sens pentru un mers aleatoriu
- Modelele ARMA standard **nu pot** gestiona aceste serii direct

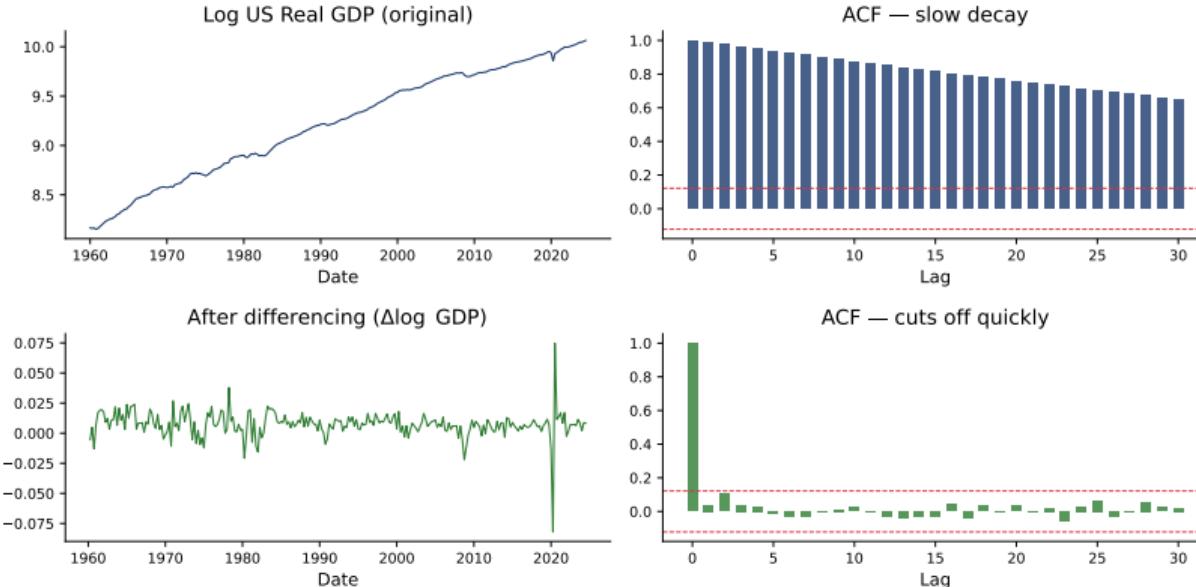


## Provocarea

Datele financiare și economice sunt de obicei **integrate** ( $I(1)$  sau aproape de rădăcină unitară):

- Prețuri de acțiuni: mers aleatoriu în logaritmi
- Cursuri de schimb: mers aleatoriu
- Rate ale dobânzii: foarte persistente (aproape de rădăcină unitară)

## Soluția: Diferențierea



### Observație Cheie

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

## Concepțe Fundamentale

- ① **Nestaționaritatea:** De ce contează și cum o detectăm
- ② **Teste de Rădăcină Unitate:** Testele ADF, PP, KPSS
- ③ **Diferențierea:** Transformarea cheie
- ④ **Modele ARIMA:** Combinarea diferențierii cu ARMA
- ⑤ **Metodologia Box-Jenkins:** Identificare → Estimare → Diagnosticare

## La Sfârșitul Acestui Curs

Veți fi capabili să modelați și să prognozați serii de timp nestaționare precum prețurile acțiunilor, PIB și cursurile de schimb folosind modele ARIMA.

# De Ce Contează Nestaționaritatea

## Problema

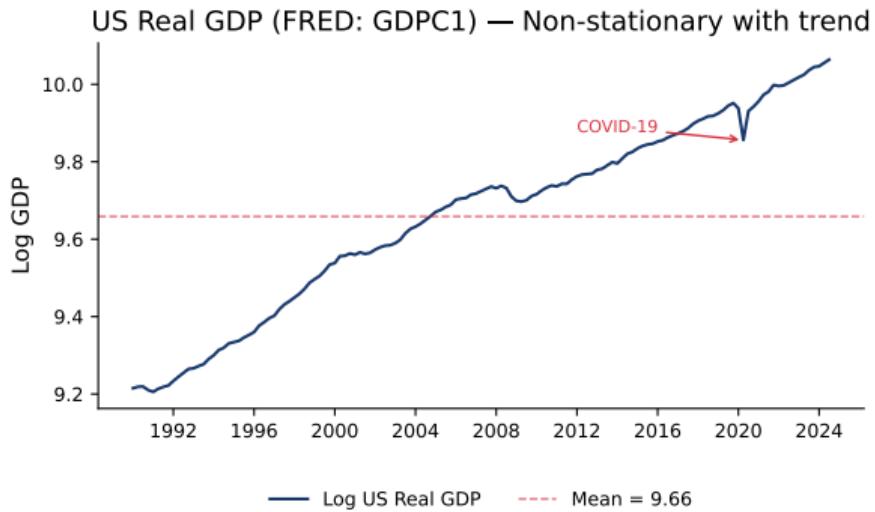
Multe serii de timp economice și financiare sunt **nestaționare**:

- PIB, prețuri de acțiuni, cursuri de schimb, indici de inflație
- Prezintă tendințe, 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 consistenti
- Inferența statistică devine invalidă

## Exemplu: PIB Real SUA



- **Trend** ascendent clar – media nu este constantă
- Acesta este un exemplu clasic de serie de timp **nestaționară**
- Nu putem aplica modele ARMA direct pe aceste date

# Tipuri de Nestaționaritate

## Trend Determinist

$$Y_t = \alpha + \beta t + \varepsilon_t$$

- Trendul este o funcție deterministă de timp
- Poate fi eliminat prin **regresie**
- Șocurile au efecte temporare

## Trend Stochastic (Rădăcină Unitate)

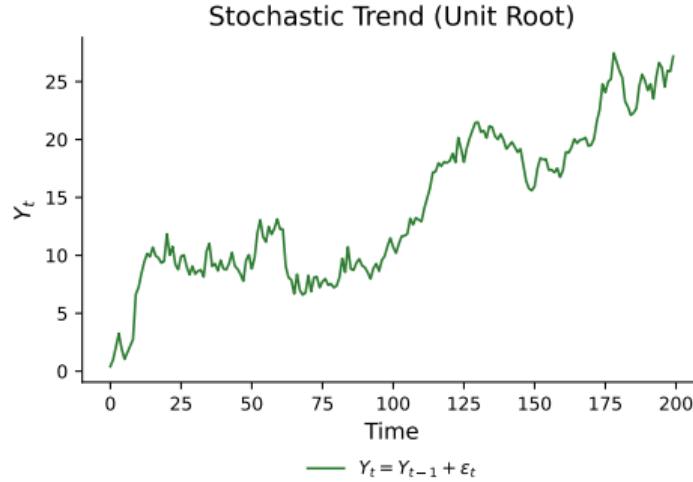
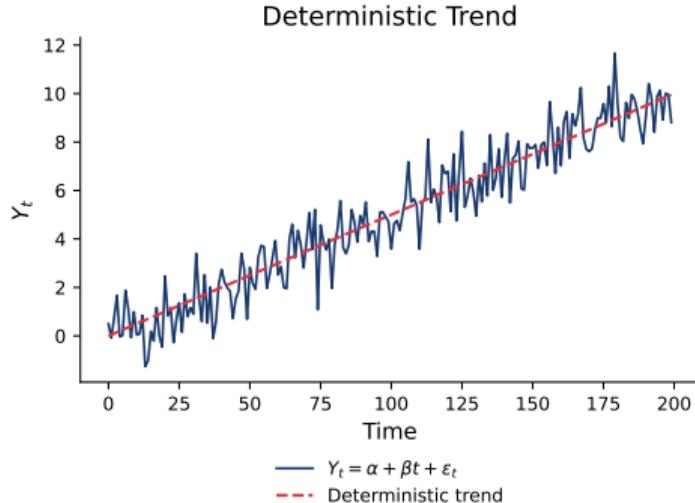
$$Y_t = Y_{t-1} + \varepsilon_t$$

- Proces de mers aleatoriu
- Trebuie eliminat prin **diferențiere**
- Șocurile au efecte permanente

## Distinctie 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



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

## Definiție 1 (Mers Aleatoriu)

Un **mers aleatoriu** este definit ca:

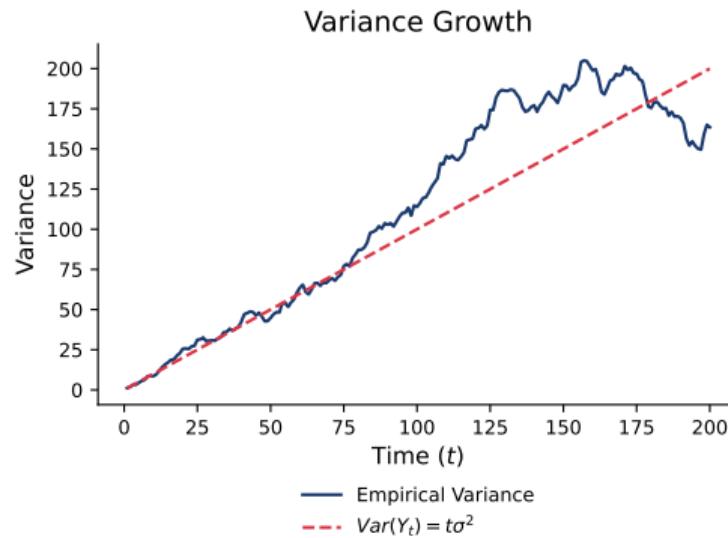
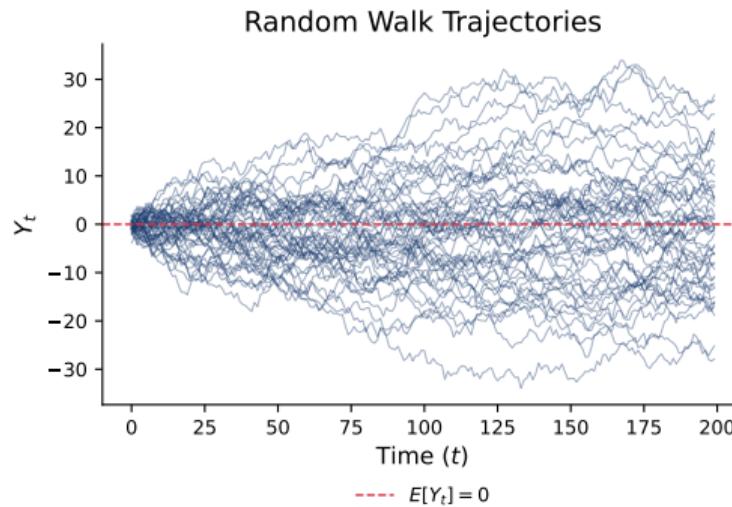
$$Y_t = Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

Cu condiția inițială  $Y_0 = 0$ , avem:  $Y_t = \sum_{i=1}^t \varepsilon_i$

## 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: Traекторii multiple rătăcesc imprevizibil fără revenire la medie. Dreapta: Varianța  $\text{Var}(Y_t) = t\sigma^2$  crește liniar — caracteristica definitorie a nestaționarității.

## 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:

$$Y_t = Y_{t-1} + \varepsilon_t = Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \dots = \sum_{i=1}^t \varepsilon_i$$

Luând varianța:

$$\text{Var}(Y_t) = \text{Var}\left(\sum_{i=1}^t \varepsilon_i\right) = \sum_{i=1}^t \text{Var}(\varepsilon_i) + 2 \sum_{i < j} \text{Cov}(\varepsilon_i, \varepsilon_j)$$

Deoarece  $\varepsilon_t$  sunt independente (zgomot alb), toate covariantele sunt zero:

$$\text{Var}(Y_t) = \sum_{i=1}^t \sigma^2 = \boxed{t\sigma^2}$$

## Nestaționaritate

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

## Demonstrație: Autocovarianța Mersului Aleatoriu

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

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

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

Doar termenii cu  $i = j$  supraviețuiesc (când  $i \leq t - k$ ).

ACF:

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

## Definiție 2 (Mers Aleatoriu cu Drift)

Un mers aleatoriu cu drift include un termen constant:

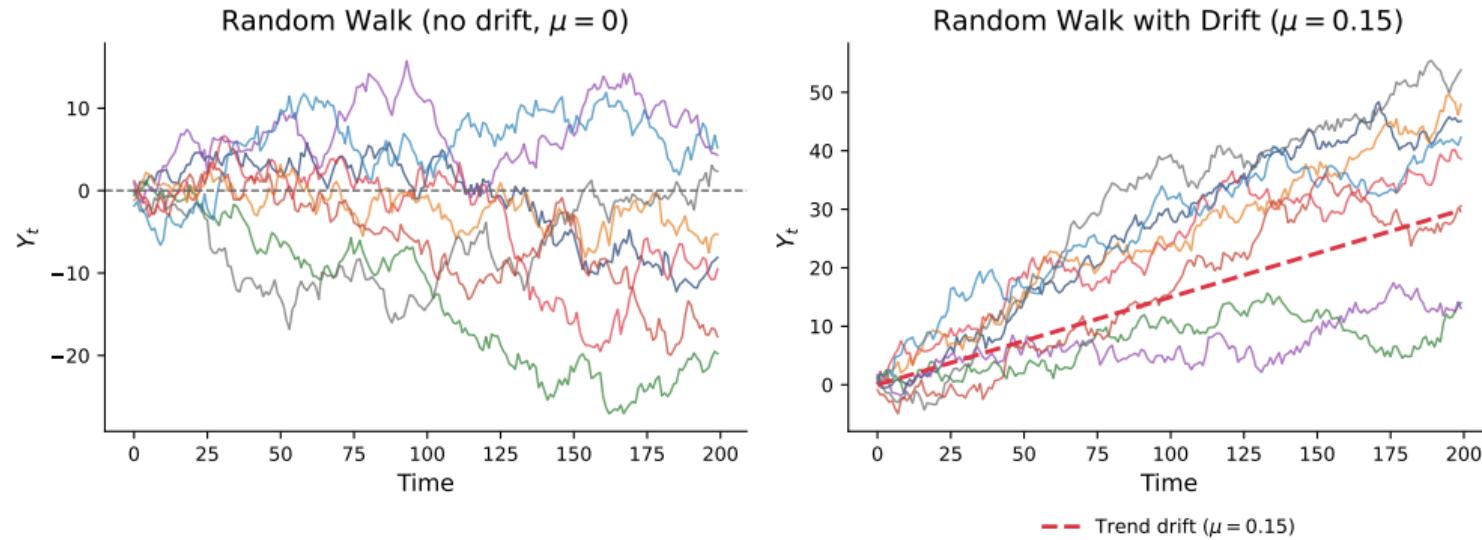
$$Y_t = \mu + Y_{t-1} + \varepsilon_t$$

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"

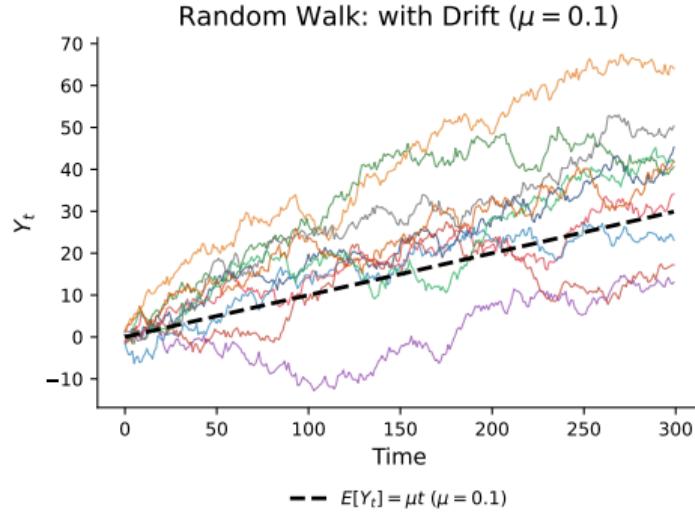
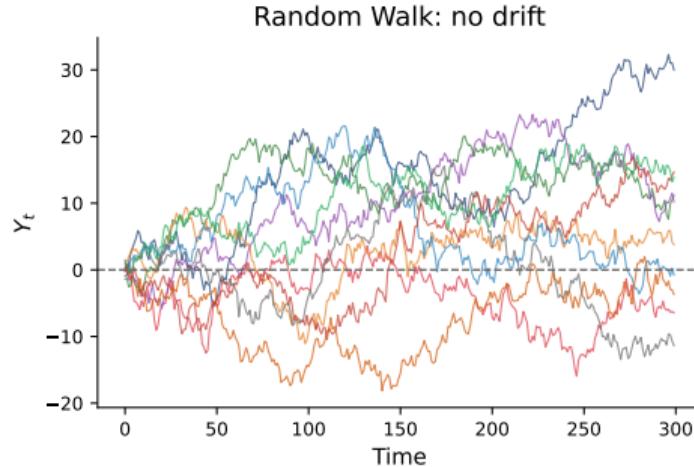
## Mers Aleatoriu cu Drift: Ilustrație Vizuală



### Comparație

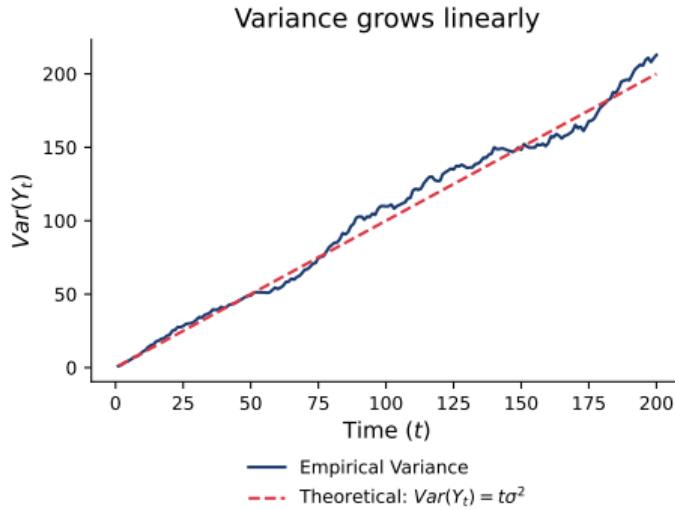
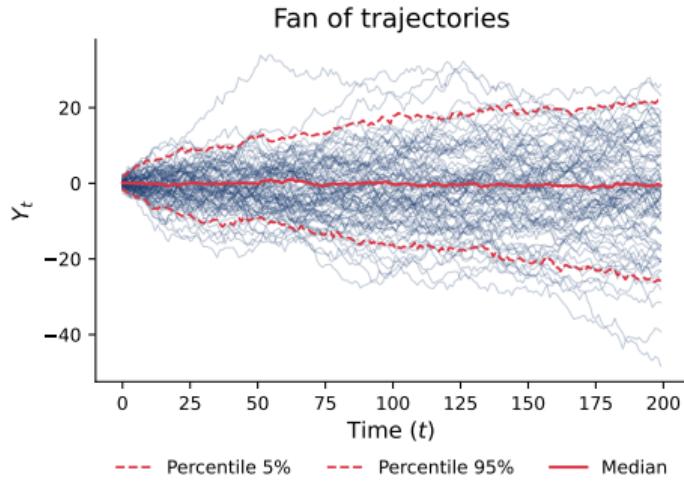
Fără drift (albastru): rătăcește în jurul lui zero fără direcție. Cu drift  $\mu > 0$  (roșu): trend ascendent sistematic. Ambele sunt nestaționare — drift-ul adaugă trend determinist la rătăcirea stochastică.

# Simularea Mersurilor Aleatorii



- **Stânga:** Mersuri aleatorii pure – fără drift, rătăcesc imprevizibil
- **Dreapta:** Mersuri aleatorii cu drift – trend ascendent în medie
- Fiecare traекторie este unică; incertitudinea crește în timp

## Creșterea Varianței: De Ce Mersurile Aleatorii Sunt Nestaționare



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

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

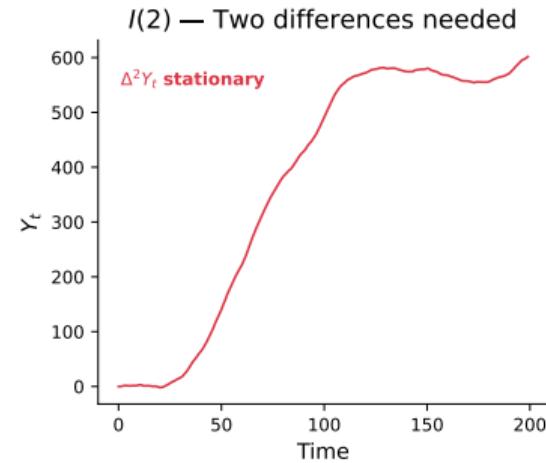
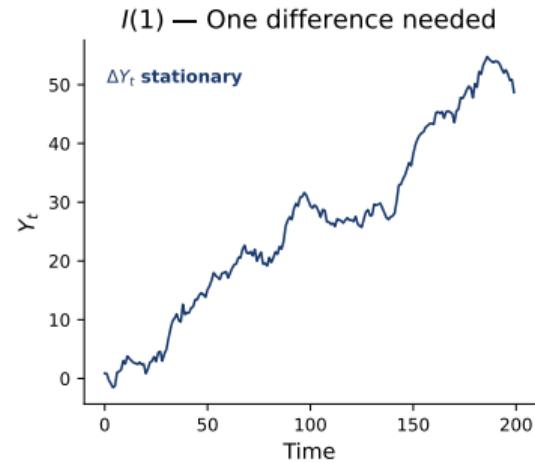
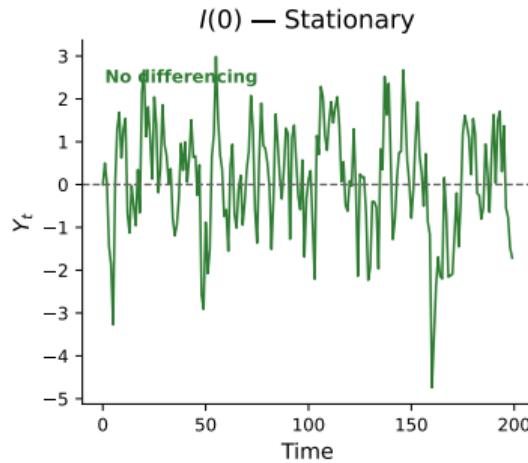
O serie de timp  $\{Y_t\}$  este **integrată de ordin  $d$** , scrisă  $Y_t \sim I(d)$ , dacă:

- $Y_t$  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 două diferență este staționară (mai rar)

# Proces Integrat: Ilustrație Vizuală



## Ordinul de Integrare

$I(0)$ : Staționar — nicio diferențiere necesară.  $I(1)$ : O diferență necesară (mers aleatoriu).  $I(2)$ : Două diferențe necesare. Majoritatea seriilor economice sunt  $I(0)$  sau  $I(1)$ .

## Definiție 4 (Prima Diferență)

**Operatorul primei diferențe**  $\Delta$  este definit ca:  $\Delta Y_t = Y_t - Y_{t-1} = (1 - L)Y_t$ , unde  $L$  este operatorul lag ( $LY_t = Y_{t-1}$ ).

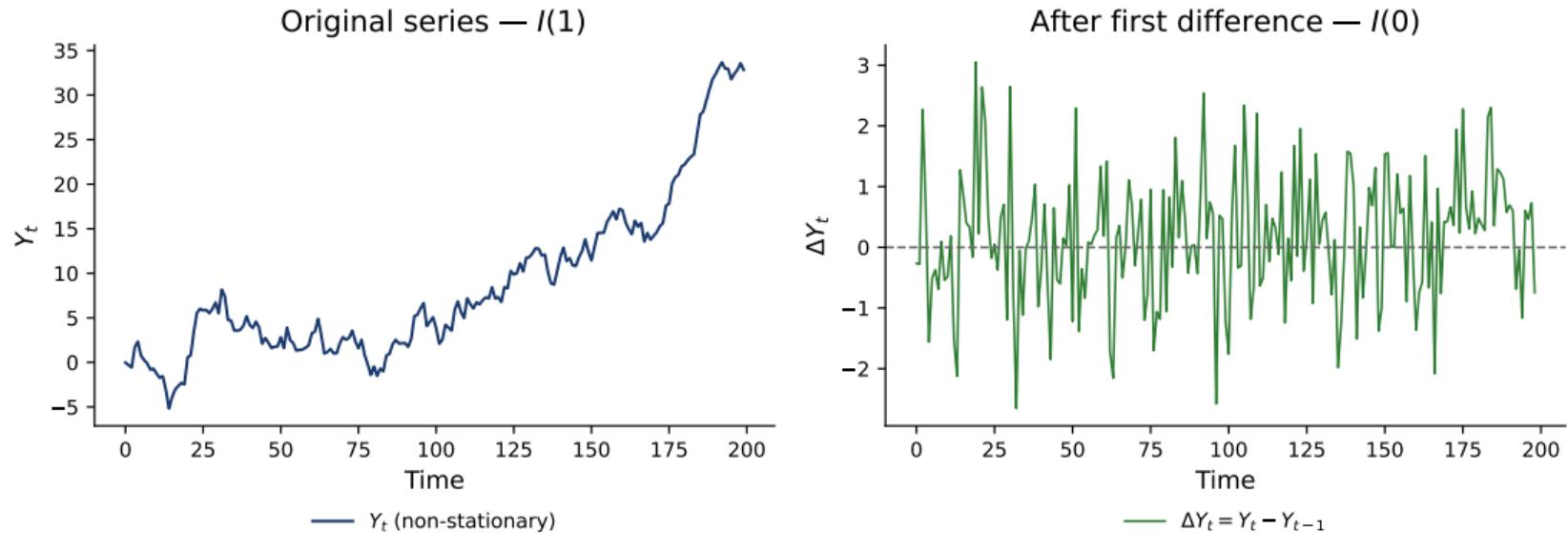
## Diferențe de Ordin Superior

- A două 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



Stânga: serie nestaționară. Dreapta: după prima diferență, seria devine staționară.

## Exemplu: Diferențierea unui Mers Aleatoriu

### Mers Aleatoriu la Zgomot Alb

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

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

Prima diferență este zgomot alb – un proces staționar!

### Interpretare

- Un mers aleatoriu este  $I(1)$
- O diferență îl transformă în  $I(0)$
- "Schimbările" într-un mers aleatoriu 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 pentru Mers Aleatoriu cu Drift:**  $Y_t = \mu + Y_{t-1} + \varepsilon_t$

Prima diferență este:

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

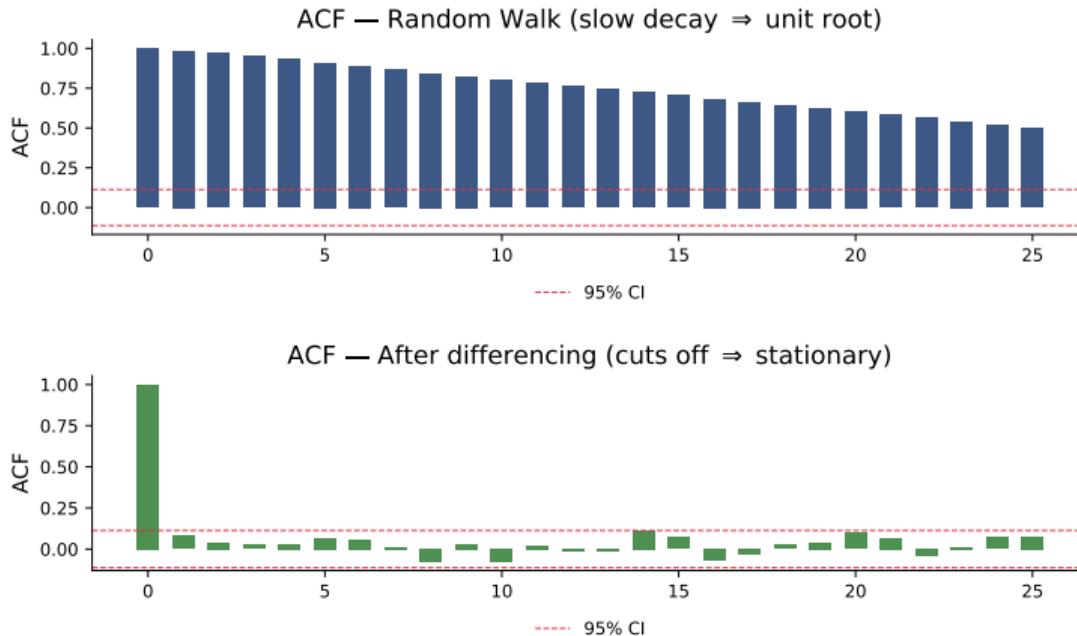
Verificăm condițiile de staționaritate:

- ① **Media:**  $\mathbb{E}[\Delta Y_t] = \mu$  (constantă, nu depinde de  $t$ ) ✓
- ② **Varianța:**  $\text{Var}(\Delta Y_t) = \text{Var}(\varepsilon_t) = \sigma^2$  (constantă) ✓
- ③ **Autocovarianța:**  $\text{Cov}(\Delta Y_t, \Delta Y_{t-k}) = \text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$  pentru  $k \neq 0$  ✓

### Principiu General

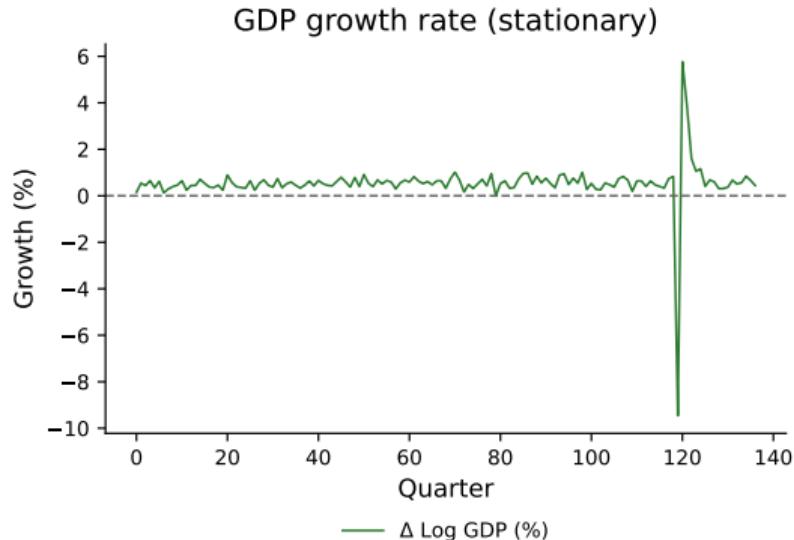
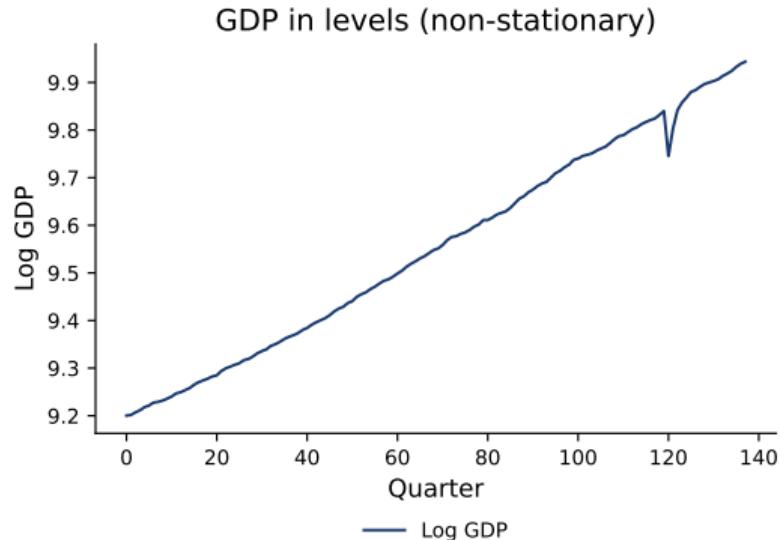
Diferențierea elimină “memoria” care face varianța să se acumuleze. Pentru procese  $I(d)$ , sunt necesare  $d$  diferențe.

## Diagnostic ACF: Detectarea Nestaționarității



- **Sus:** ACF mers aleatoriu scade foarte lent  $\Rightarrow$  rădăcină unitară
- **Jos:** După diferențiere, ACF se anulează  $\Rightarrow$  staționar

## Diferențierea în Practică: Exemplul PIB



- **Stânga:** PIB în niveluri – trend ascendent clar (nestaționar)
- **Dreapta:** Rata de creștere PIB (diferența logaritmică) – fluctuează în jurul mediei (staționar)
- Diferențierea elimină trendul și obține staționaritate

## 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  $\theta = 1$  (la granița non-invertibilității)!

## Definiție 5 (ARIMA(p,d,q))

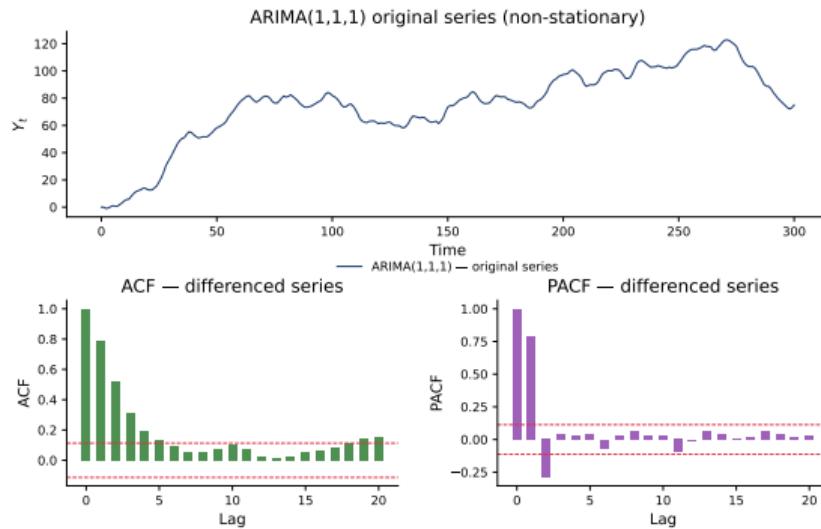
O serie de timp  $\{Y_t\}$  urmează un proces **ARIMA(p,d,q)** dacă:

$$\phi(L)(1 - L)^d Y_t = c + \theta(L)\varepsilon_t$$

unde:

- $\phi(L) = 1 - \phi_1L - \phi_2L^2 - \cdots - \phi_pL^p$  (polinomul AR)
- $\theta(L) = 1 + \theta_1L + \theta_2L^2 + \cdots + \theta_qL^q$  (polinomul MA)
- $d$  este ordinul de integrare (numărul de diferențe)
- $\varepsilon_t \sim WN(0, \sigma^2)$

# ARIMA: Ilustrație Vizuală



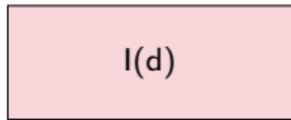
## Interpretare

Sus: seria ARIMA originală (nestaționară). Jos: după diferențiere de  $d$  ori, ACF/PACF dezvăluie ordinele AR și MA pentru componenta staționară.

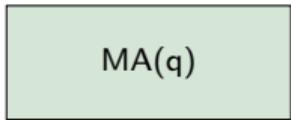
# Componentele ARIMA



AR(p)



I(d)



MA(q)

Autoregresiv  
Memorie

Integrare  
Diferențiere

Medie Mobilă  
Șocuri

## Cazuri Speciale

- $\text{ARIMA}(p,0,q) = \text{ARMA}(p,q)$  – staționar
- $\text{ARIMA}(0,1,0) = \text{Mers aleatoriu}$
- $\text{ARIMA}(0,1,1) = \text{IMA}(1,1)$  – netezire exponențială
- $\text{ARIMA}(1,1,0) = \text{ARI}(1,1)$  –  $\text{AR}(1)$  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 stochastic
- 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  $-1 < \theta_1 < 0$ ).

## 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  $d \geq 1$

## De Ce Testăm?

Înainte de a potrivi un model ARIMA, trebuie să determinăm:

- ① Este seria staționară? (Este  $d = 0$ ?)
- ② 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 – ipoteză nulă inversată)

## 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 unitară,  $\phi = 1$ , nestaționar)
- $H_1: \gamma < 0$  (staționar,  $|\phi| < 1$ )

## Problemă Cheie

Sub  $H_0$ , statistică  $t$  **nu** urmează o distribuție  $t$  standard! Trebuie folosite valorile critice 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

Dacă există dinamică AR dincolo de AR(1), reziduurile DF vor fi autocorelate.

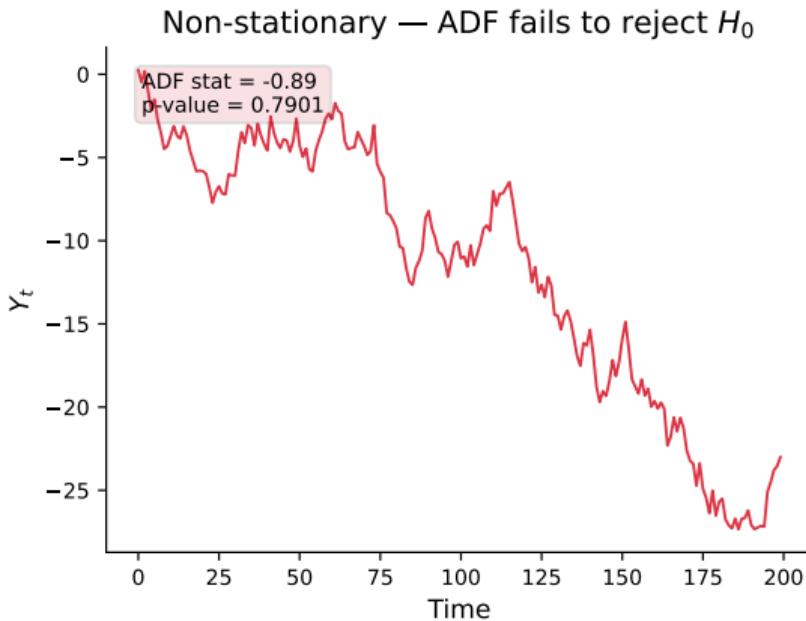
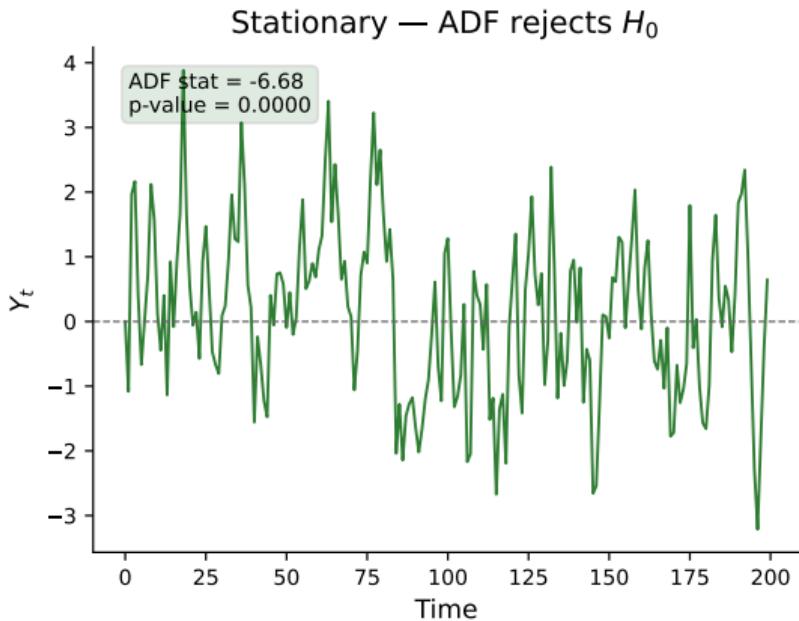
## Definiție 6 (Testul ADF)

Adăugați diferențe întârziate:  $\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{j=1}^k \delta_j \Delta Y_{t-j} + \varepsilon_t$   
Testați  $H_0 : \gamma = 0$  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



Stânga: serie staționară – ADF respinge rădăcina unitară. Dreapta: nestaționară – ADF nu respinge.

| 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  $\Rightarrow$  Respingem  $H_0$  (staționar)
- Statistică de test  $\geq$  valoare critică  $\Rightarrow$  Nu respingem (rădăcină unitară)

# Testul Phillips-Perron (PP)

## Motivație

Ca și ADF, testează  $H_0$ : Rădăcină unitate vs  $H_1$ : Staționar, dar folosește o **corecție non-parametrică** pentru corelația serială în loc de adăugarea diferențelor întârziate.

## Statistică de Test

Testul PP modifică statistică  $t$  DF:

$$Z_t = t_{\hat{\gamma}} \cdot \sqrt{\frac{\hat{\sigma}^2}{\hat{\lambda}^2}} - \frac{T(\hat{\lambda}^2 - \hat{\sigma}^2)(se(\hat{\gamma}))}{2\hat{\lambda}^2 \cdot s}$$

unde  $\hat{\lambda}^2$  este o estimare consistentă a varianței pe termen lung folosind Newey-West.

## Avantaje față de ADF

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

## Ipoteze Înversate

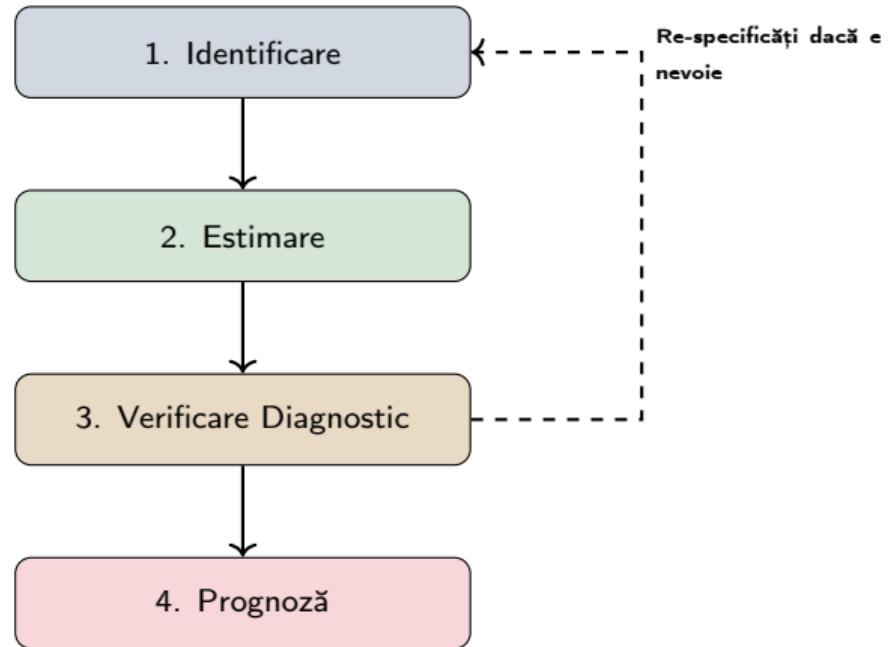
Spre deosebire de ADF:  $H_0$ : Staționar vs  $H_1$ : Rădăcină unitate

## Procedura KPSS

Descompunem:  $Y_t = \xi t + r_t + \varepsilon_t$  unde  $r_t = r_{t-1} + u_t$ . Testăm dacă  $\text{Var}(u_t) = 0$ .

## Utilizare Complementară cu ADF

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



## Pasul 1: Determinarea lui $d$

### Procedură

- ❶ Reprezentăți grafic seria de timp – căutați trenduri, varianță în schimbare
- ❷ Examinați ACF – descreștere lentă sugerează nestaționaritate
- ❸ Aplicați teste de rădăcină unitară (ADF, KPSS)
- ❹ Dacă nestaționară, diferențiați și repetăți

### 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

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 sunt neclare, comparați modelele folosind:

- AIC =  $-2 \ln(L) + 2k$ ; BIC =  $-2 \ln(L) + k \ln(n)$

Mai mic este mai bun. BIC penalizează complexitatea mai mult.

## Selectie 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 unitară 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

Selectia automată este utilă dar nu infailibilă. Verificați întotdeauna diagnosticele!

## Estimarea prin Maximum de Verosimilitate (MLE)

Abordarea standard pentru ARIMA:

- Presupune  $\varepsilon_t \sim N(0, \sigma^2)$
- Maximizează funcția de verosimilitate
- Oferă estimatori consistenti, 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

## Staționaritate și Invertibilitate

Modelul ARIMA estimat ar trebui să satisfacă:

- **Staționaritate AR:** Rădăcinile lui  $\phi(z) = 0$  în afără cercului unitate
- **Invertibilitate MA:** Rădăcinile lui  $\theta(z) = 0$  în afără 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

## Ce Trebuie Verificat

Dacă modelul este corect, reziduurile  $\hat{\epsilon}_t$  ar trebui să fie zgomot alb:

- ① Medie zero
- ② Varianță constantă
- ③ Fără autocorelație
- ④ (Optional) Normalitate

## Instrumente de Diagnostic

- **ACF/PACF rezidual:** Nu ar trebui să arate vârfuri semnificative
- **Testul Ljung-Box:** Testează autocorelația la lag-uri multiple
- **Graficul Q-Q:** Verifică ipoteza de normalitate
- **Rezidual vs potrivit:** Verifică heteroscedasticitatea

# Testul Ljung-Box

## Definiție 7 (Statistică Q Ljung-Box)

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

## Utilizare

- Alegeti  $m \approx \ln(n)$  sau  $m = 10$  pentru trimestrial,  $m = 20$  pentru lunar
- Grade de libertate ajustate pentru parametrii estimati
- Respingeti dacă  $Q(m)$  depășește valoarea critică

## Dacă Testul Eșuează

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

## Prognoză cu MSE Minim

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

## Prognoză ARIMA(1,1,1)

Model:  $(1 - \phi_1 L)(1 - L)Y_t = c + (1 + \theta_1 L)\varepsilon_t$

Prognoză un pas:  $\hat{Y}_{T+1|T} = c + Y_T + \phi_1(Y_T - Y_{T-1}) + \theta_1 \hat{\varepsilon}_T$

Pentru  $h > 1$ : înlocuiți  $\varepsilon_{T+j}$  necunoscut cu 0,  $Y_{T+j}$  necunoscut cu  $\hat{Y}_{T+j|T}$

# Intervale de Prognoză

## Incertitudinea Prognozei

Varianța erorii de prognoză la  $h$  pași:  $\text{Var}(e_{T+h}) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$ , unde  $\psi_j$  sunt coeficienții 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 proguozei crește nelimitat când  $h \rightarrow \infty$ . Intervalele se largesc în timp!

# Prognoze pe Termen Lung pentru ARIMA

## Comportament când $h \rightarrow \infty$

Pentru ARIMA(p,1,q) cu drift c:

- Prognoze punctuale: Trend liniar cu pantă = drift
- Intervale de prognoză: Lățimea crește cu  $\sqrt{h}$

Pentru ARIMA(p,1,q) fără drift:

- Prognoze punctuale: Converg la ultimul nivel
- Intervale de prognoză: Tot cresc nelimitat

## Implicație Practică

Prognozele ARIMA sunt cele mai fiabile pentru orizonturi scurte. Prognozele pe termen lung au benzi de incertitudine foarte largi.

## Ce este Prognosă Rulantă?

O tehnică pentru evaluarea acurateții progonzei în afără eșantionului:

- 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 eșantionului

## De ce Prognoze Rulante?

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

## Prognosă Rulantă: 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  $T$

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

Prognosă Punctuală la 1 Pas

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

## Prognosă la 2 Pași

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

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

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

## Numeric: Prognosă la 3 Pași

$$\hat{Y}_{T+3|T} = 108 + 5 \times \frac{0.6(1 - 0.6^3)}{1 - 0.6} = 108 + 5 \times 1.092 = \boxed{113.46}$$

### Varianța Erorii de Prognoză

Pentru ARIMA(1,1,0), varianța erorii de prognoză la  $h$  pași:

$$\text{Var}(e_{T+h|T}) = \sigma^2 \left( 1 + \sum_{j=1}^{h-1} \psi_j^2 \right)$$

unde  $\psi_j = \phi_1^{j-1} (1 + \phi_1 + \cdots + \phi_1^{j-1}) = \phi_1^{j-1} \cdot \frac{1 - \phi_1^j}{1 - \phi_1}$

### 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

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

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

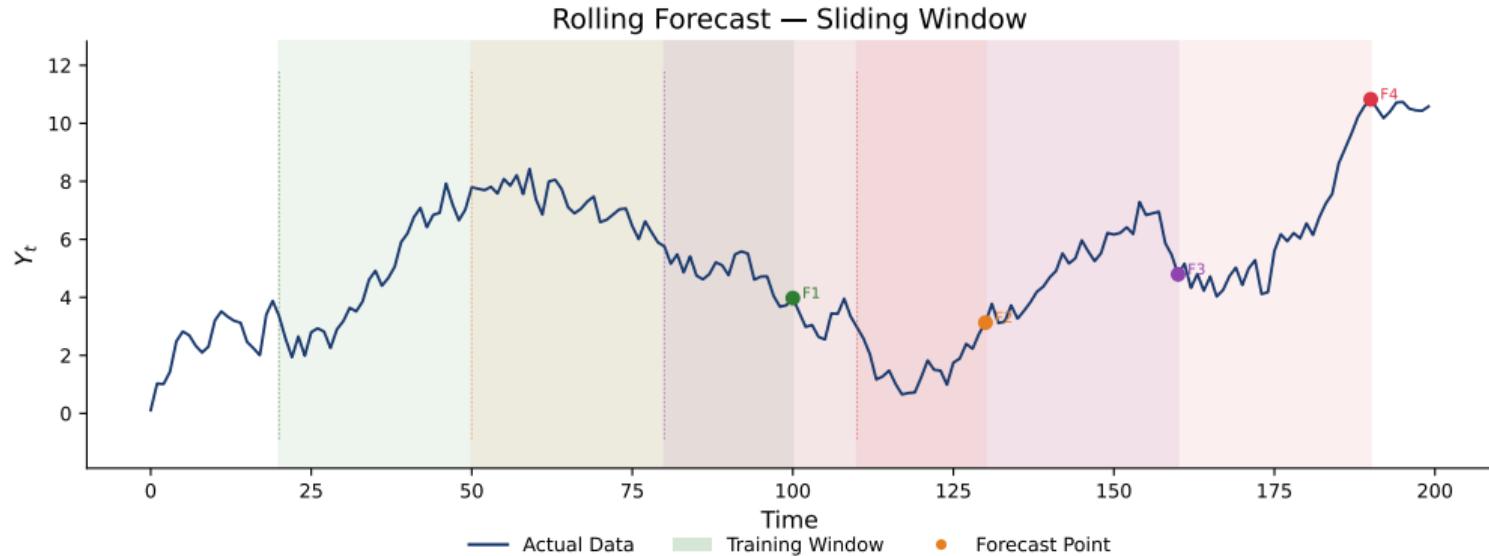
### IC la 2 Pași (pentru $\hat{Y}_{T+2|T} = 112.8$ )

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

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

Notă: IC se lărgește pe măsură ce orizontul crește!

## Ilustrație Fereastră Rulantă



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

## 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 rulantă
    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))
```

## Obiectiv

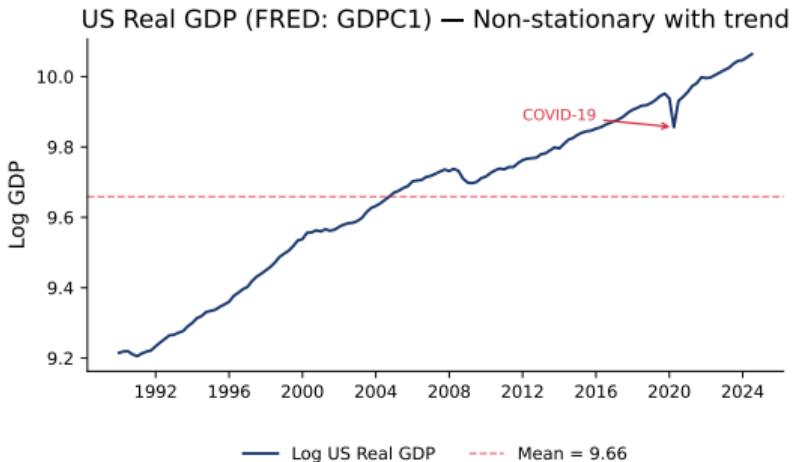
Prognosă PIB Real al SUA folosind metodologia Box-Jenkins

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

## Date

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

## Pasul 1: Analiza Inițială a Datelor



### Observații

- Trend ascendent clar  $\Rightarrow$  medie neconstantă
- Varianța pare relativ stabilă (după transformare log)
- Scădere notabilă în 2020 (pandemia COVID-19)
- **Concluzie:** Seria este nestaționară, necesită diferențiere

## Pasul 2: Testarea Rădăcinii Unitate

### Test ADF pe Log PIB în Niveluri

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

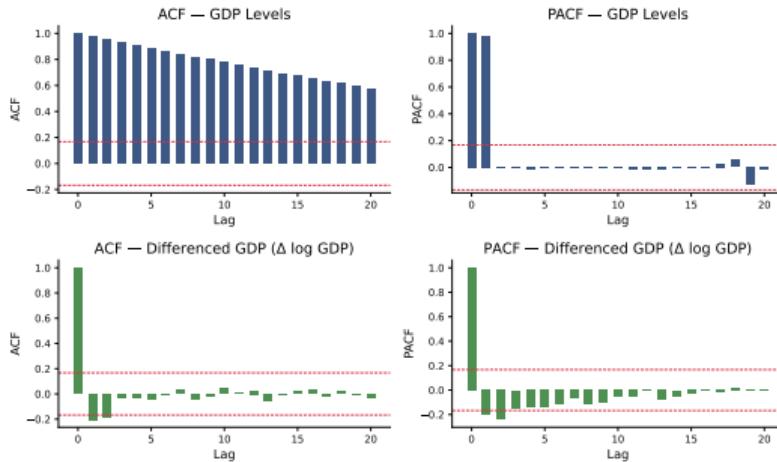
### Test ADF pe Prima Diferență (Rata de Creștere)

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

### Concluzie

PIB este  $I(1) \Rightarrow$  Folosim  $d = 1$  în modelul ARIMA

## Pasul 3: Identificarea Modelului prin ACF/PACF



### Analiza Seriei Diferențiate

- ACF: Spike semnificativ la lag 1, apoi se anulează  $\Rightarrow$  sugerează MA(1)
- PACF: Spike semnificativ la lag 1, scade  $\Rightarrow$  sugerează AR(1)
- Modele candidate: ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1)

## Pasul 4: Estimarea Modelului

### Compararea Modelelor folosind Criterii Informaționale

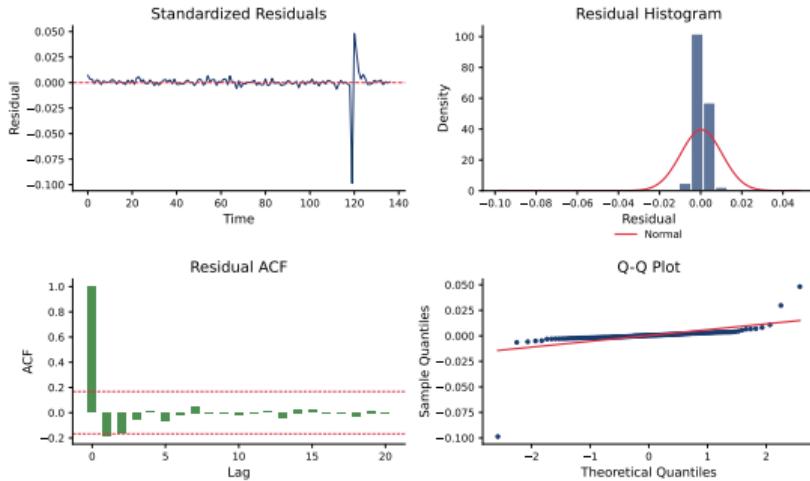
| Model               | AIC           | BIC           | Log-Lik      |
|---------------------|---------------|---------------|--------------|
| ARIMA(1,1,0)        | -725.2        | -719.5        | 364.6        |
| ARIMA(0,1,1)        | -724.8        | -719.2        | 364.4        |
| <b>ARIMA(1,1,1)</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: Verificare Diagnostică



### Analiza Reziduurilor

- Test Ljung-Box:  $Q(10) = 5.8$ , p-value = 0.83  $\Rightarrow$  Fără autocorelare
- Test Jarque-Bera:  $JB = 156.4$ , p-value < 0.001  $\Rightarrow$  Non-normal (outlier COVID)
- Concluzie: Modelul trece verificările de autocorelare; outlierii sunt aşteptați

## Pasul 6: Prognoză cu Intervale de Încredere

### Ultimele Valori Observe (Log PIB)

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

### Prognoză la 1 Pas (2024T3)

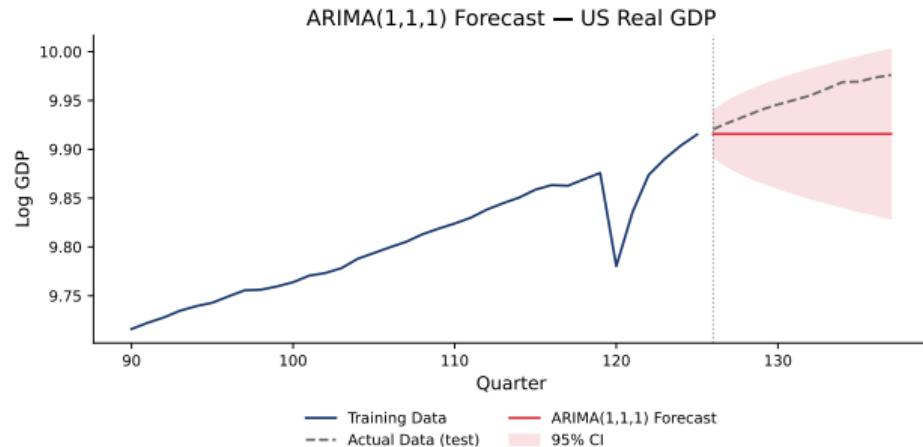
$$\hat{\Delta}Y_{T+1} = \hat{\phi}_1 \Delta Y_T + \hat{\theta}_1 \hat{\varepsilon}_T = 0.35(0.017) + 0.58(0.004) = 0.0083 \\ \hat{Y}_{T+1} = 9.973 + 0.0083 = \boxed{9.981}$$

### Interval de Încredere 95%

$$IC = 9.981 \pm 1.96 \times \sqrt{0.000156} = [9.957, 10.006]$$

În niveluri: Prognoză PIB = \$21,652 mld, IC = [\$21,142 mld, \$22,175 mld]

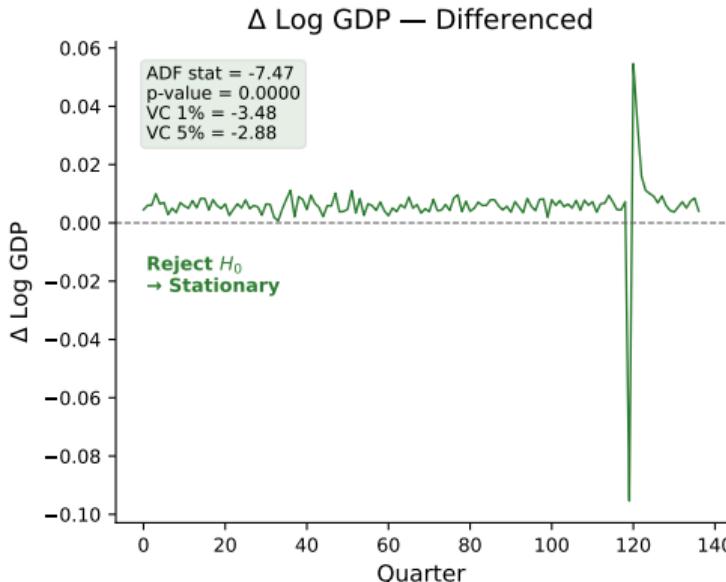
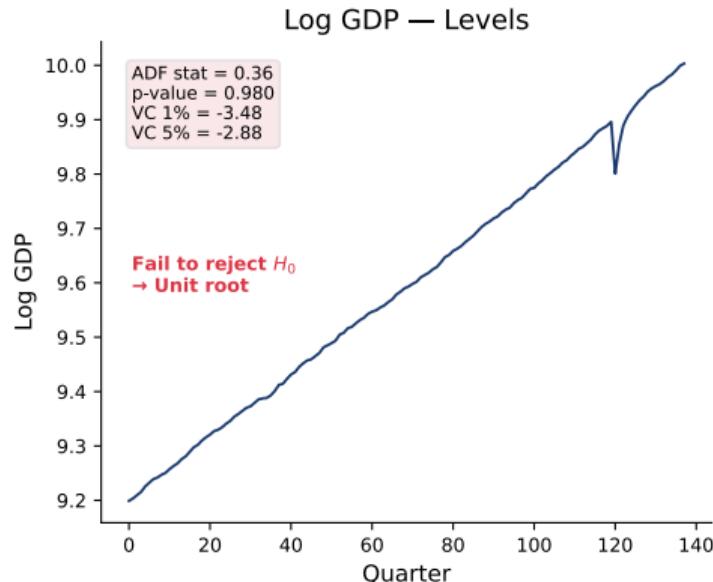
## Pasul 7: Evaluarea Prognozei



### Performanță Out-of-Sample (Ultimele 12 Trimestre)

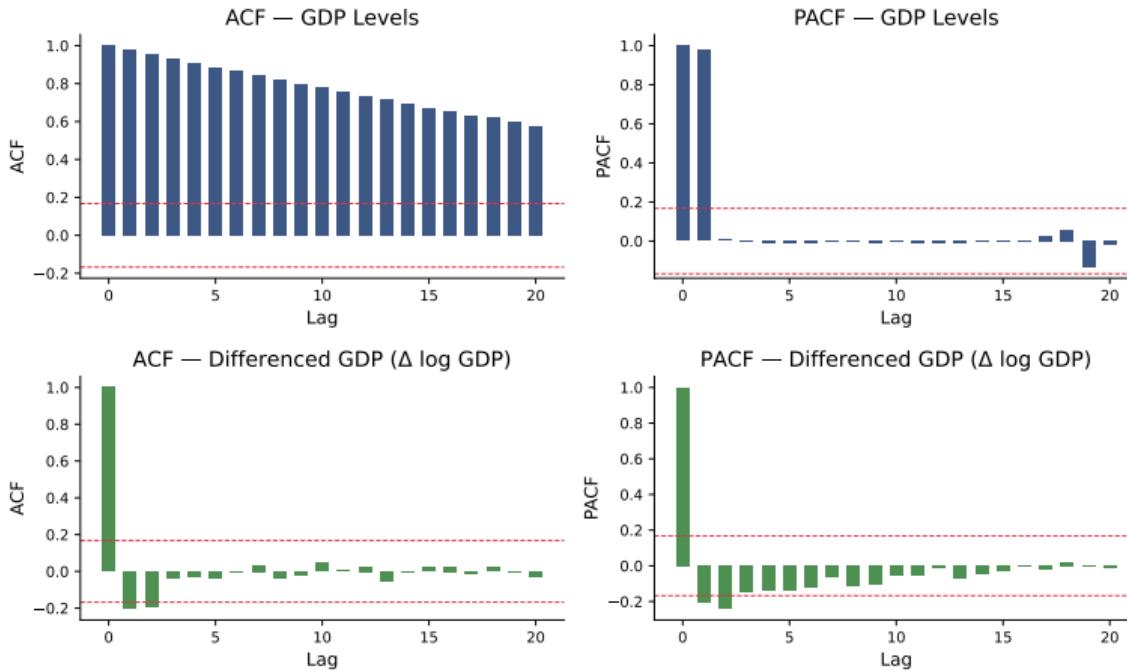
- RMSE = 0.0486 (scală log)  $\approx 4.86\%$  eroare
- MAE = 0.0430 (scală log)  $\approx 4.30\%$  eroare
- Acuratețe direcție = 91% (a prezis corect creștere/scădere)

## Rezultatele Testului de Rădăcină Unitate



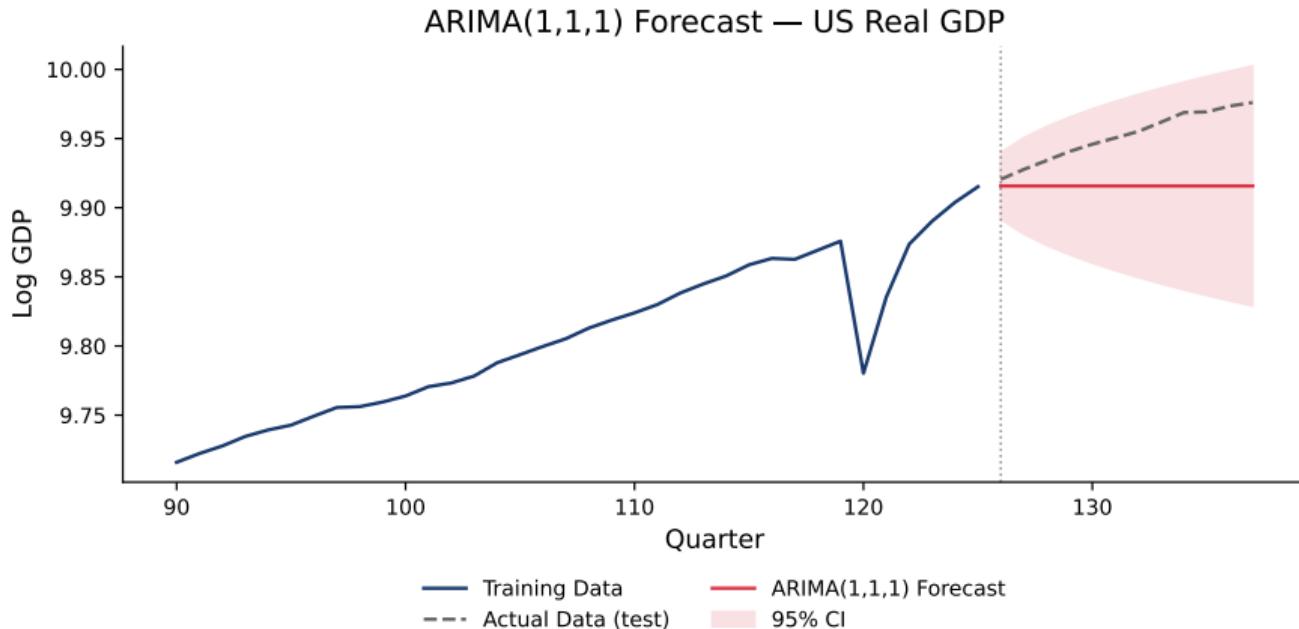
- PIB în niveluri: Nu putem respinge rădăcina unitară (nestaționar)
- Creștere PIB: Respingem rădăcina unitară la nivel de 1% (staționar)

# ACF/PACF: Niveluri vs Diferențiat



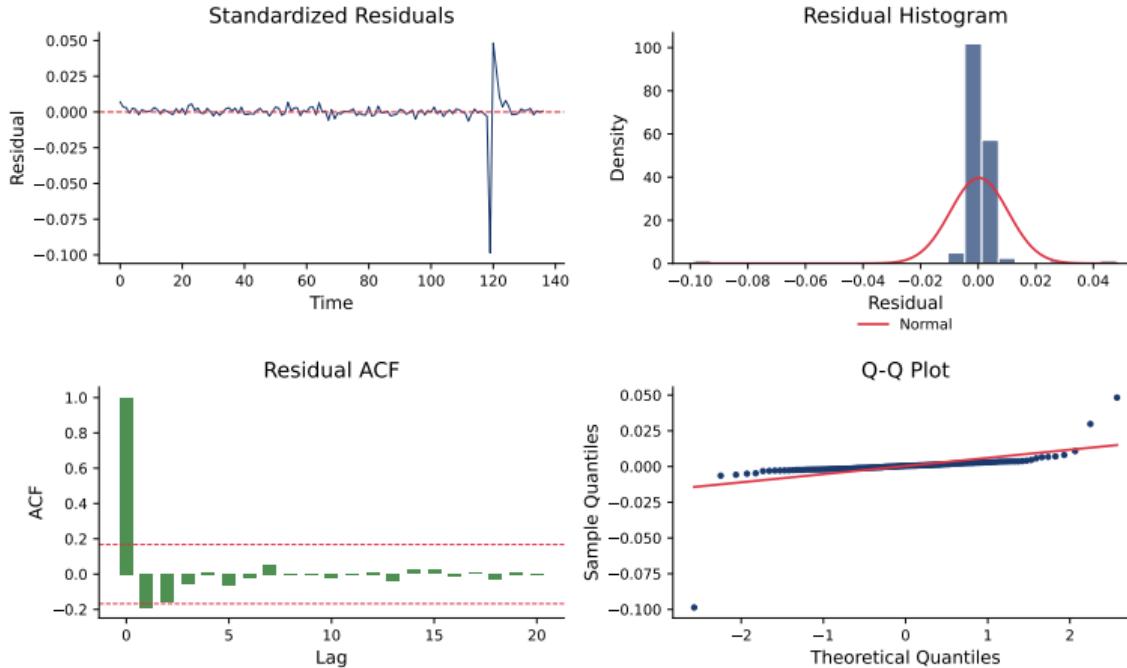
- **Sus:** Descreștere lentă ACF în niveluri sugerează nestaționaritate
- **Jos:** După diferențiere, ACF/PACF ajută la identificarea lui  $p$  și  $q$

## Prognosă ARIMA: Real vs Prezis



- ARIMA(1,1,1) captează dinamică trendului
- Intervalele de încredere se largesc cu orizontul de prognoză

# Diagnosticarea Modelului

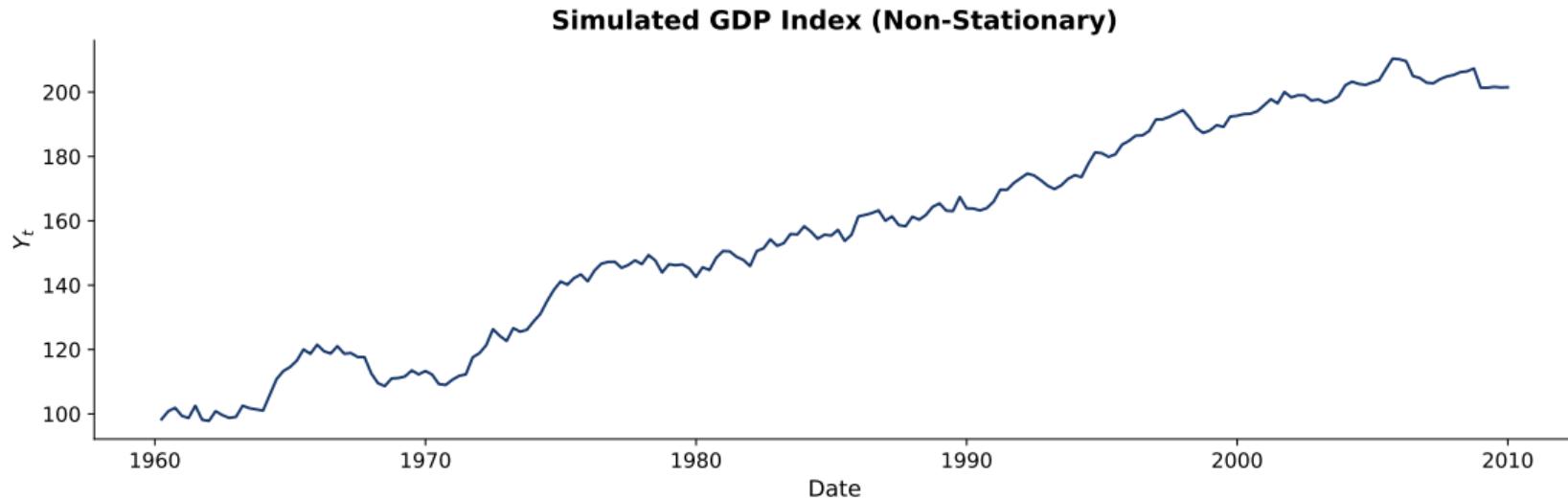


- Reziduurile par aleatorii; ACF în limitele benzilor
- Graficul Q-Q arată normalitate aproximativă

## Exemplu Auto-ARIMA

```
# Selectie automata a modelului
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())
```

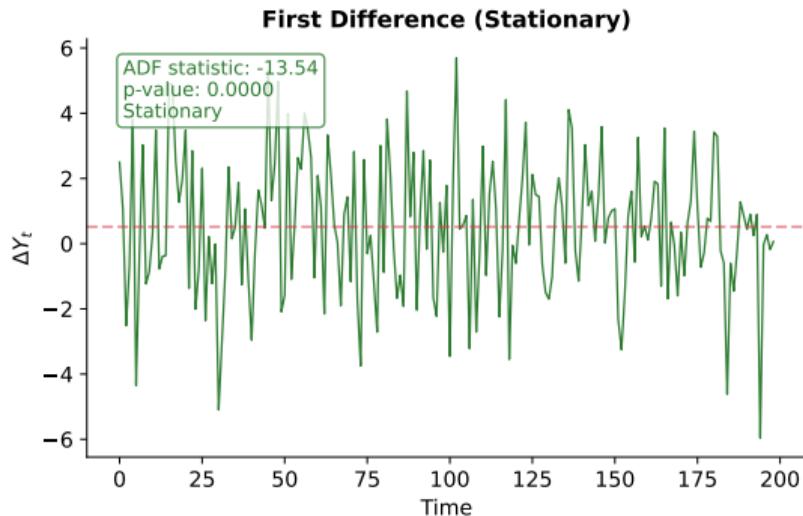
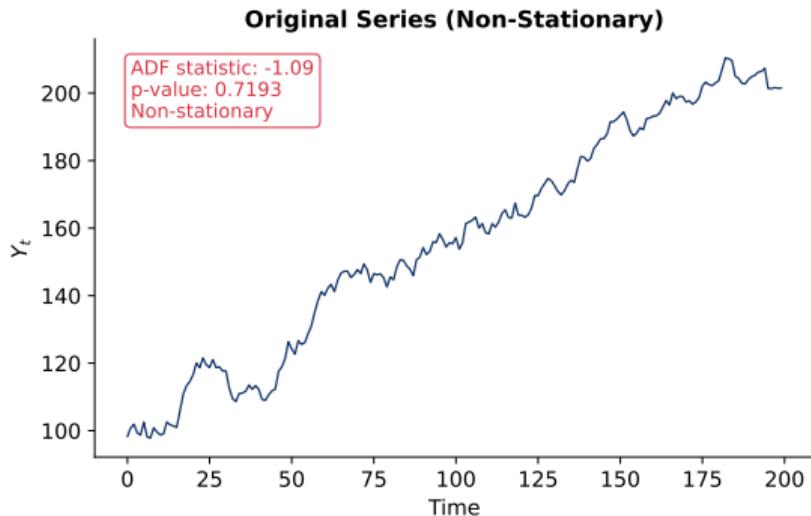
## Studiu de Caz: Indice PIB (Simulat)



### Descrierea Datelor

**Indice PIB trimestrial simulat (1960–2010):** Serie nestaționară cu trend ascendent. Demonstrează necesitatea diferențierii înainte de modelarea ARIMA.

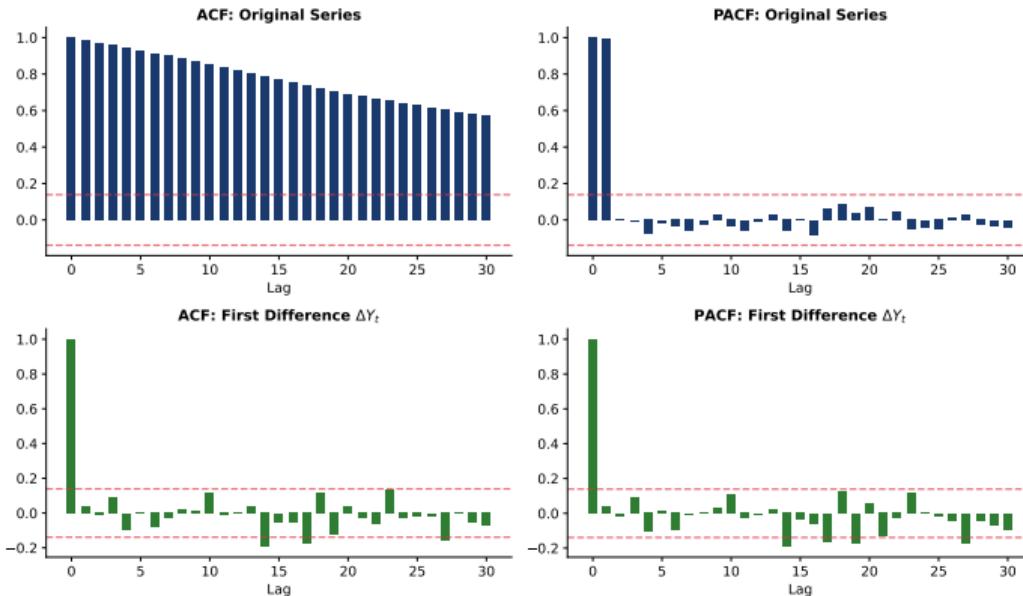
## Pasul 1: Testul ADF pentru Staționaritate



### Rezultate Test ADF

Seria originală: p-value mare  $\Rightarrow$  nu respingem  $H_0$  (rădăcină unitară prezentă). Prima diferență: p-value  $< 0.01$   $\Rightarrow$  respingem  $H_0 \Rightarrow d = 1$  este suficient.

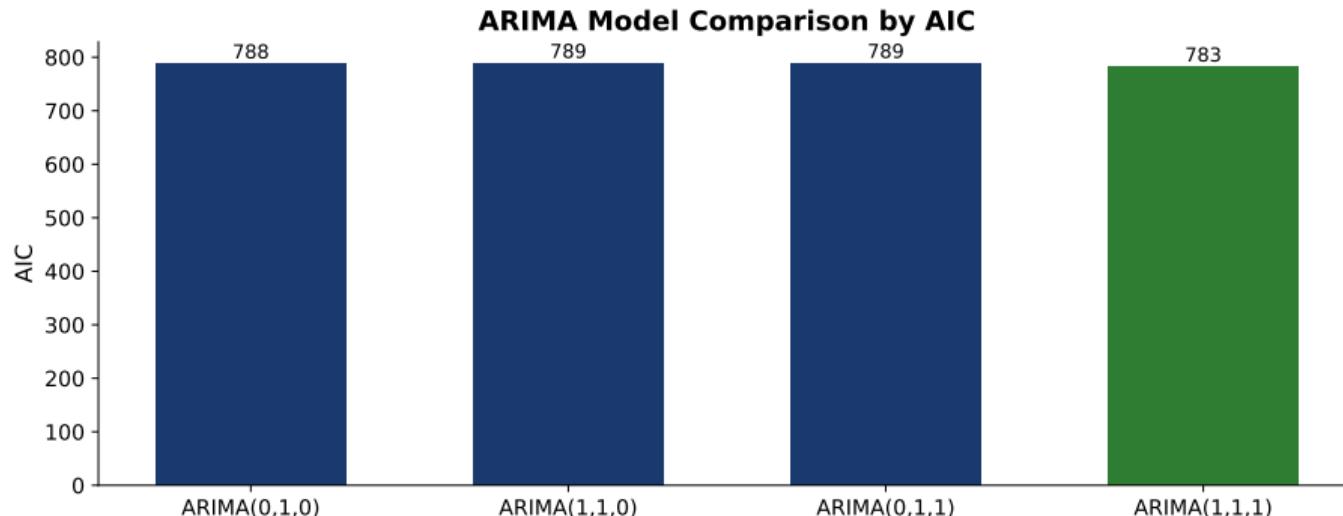
## Pasul 2: ACF/PACF Înainte și După Diferențiere



### Identificare

**Sus:** ACF cu descreștere lentă  $\Rightarrow$  nestacionaritate. **Jos:** După diferențiere, ACF și PACF sugerează ARMA de ordin mic.

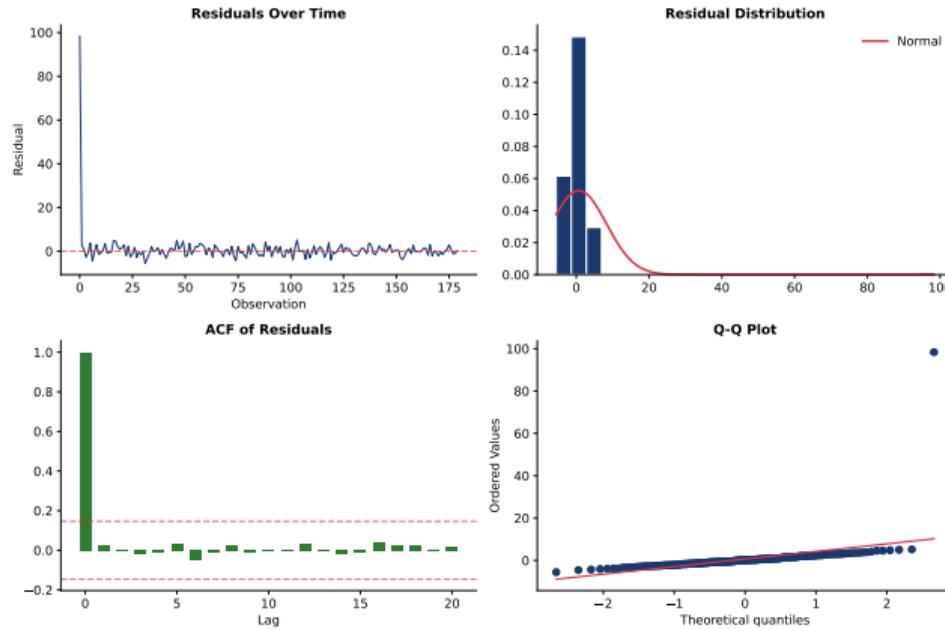
## Pasul 3: Compararea Modelelor ARIMA



### Selectia Modelului

Comparăm ARIMA(0,1,0), ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1). Modelul cu cel mai mic AIC este selectat.

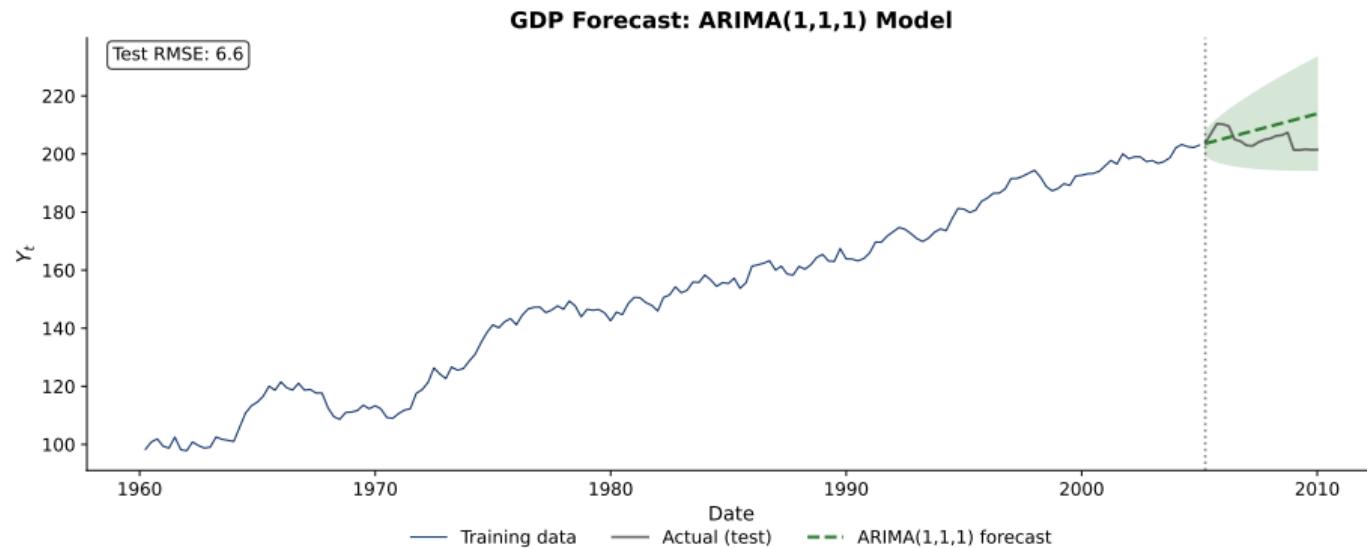
## Pasul 4: Verificarea Diagnostică



### Diagnostice ARIMA(1,1,1)

Reziduurile sunt aproximativ zgomot alb: nicio autocorelație semnificativă, distribuție aproape normală.

## Pasul 5: Prognoză



### Rezultate

- Prognozele ARIMA(1,1,1) urmează trendul datelor
- Intervalele de încredere se largesc cu orizontul (caracteristic pentru I(1))
- Incertitudinea crește deoarece erorile se acumulează prin diferențiere

## Puncte Principale

- ❶ **Nestăționaritatea** este frecventă în datele economice – trebuie abordată
- ❷ **Diferențierea** transformă  $I(d)$  în  $I(0)$
- ❸ **ARIMA(p,d,q)** combină diferențierea cu modelarea ARMA
- ❹ **Testele de rădăcină unitară** (ADF, KPSS) ajută la determinarea lui  $d$
- ❺ **Metodologia Box-Jenkins:** Identificare → Estimare → Diagnosticare
- ❻ **Prognozele** pentru serii  $I(1)$  au incertitudine în creștere

## Pașii Următori

Capitolul 4 va extinde ARIMA pentru a gestiona sezonialitatea: modele SARIMA.

## Întrebarea Quiz 1

### Întrebare

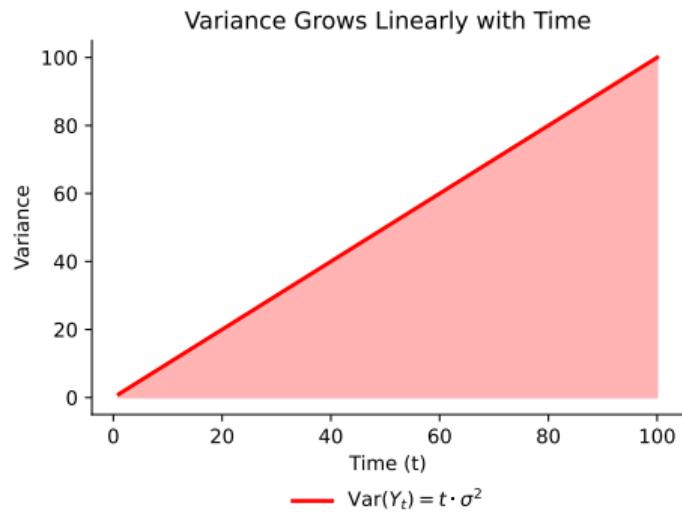
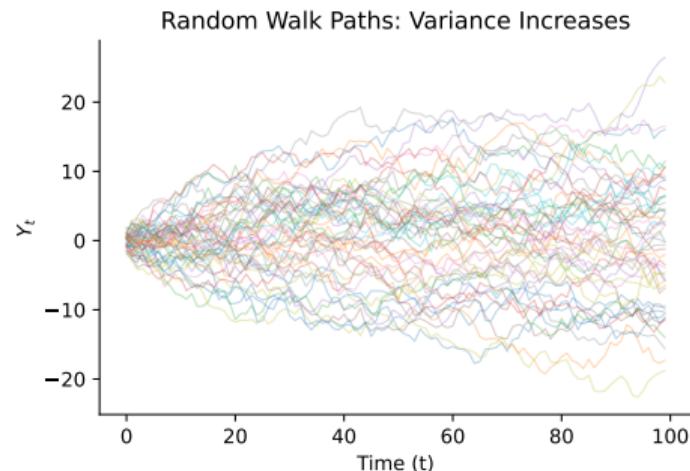
O serie de timp  $Y_t$  urmează un mers aleatoriu:  $Y_t = Y_{t-1} + \varepsilon_t$ . Care este  $\text{Var}(Y_t)$ ?

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

## Întrebarea Quiz 1: Răspuns

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

Varianța mersului aleatoriu crește liniar în timp — de aceea mersurile aleatorii sunt nestaționare.



## Întrebarea Quiz 2

### Întrebare

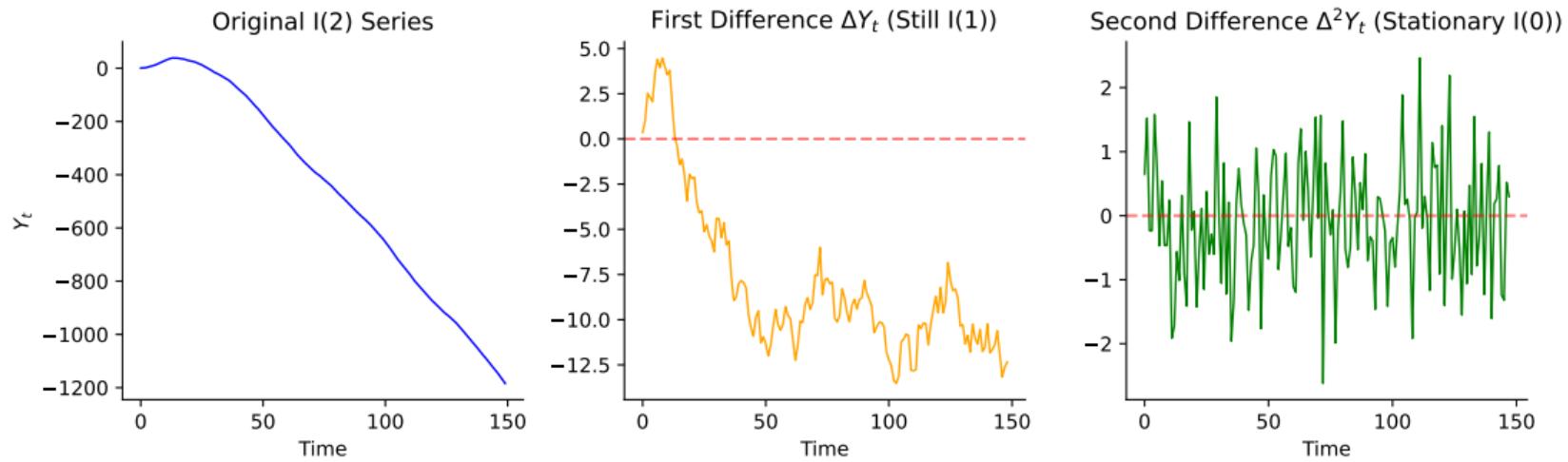
Dacă o serie  $Y_t$  este I(2), de câte ori trebuie diferențiată pentru a atinge staționaritatea?

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

## Întrebarea Quiz 2: Răspuns

Răspuns Corect: (C) 2 ori

I( $d$ ) înseamnă “integrată de ordin  $d$ ” — necesită  $d$  diferențe pentru staționaritate.



## Întrebarea Quiz 3

### Întrebare

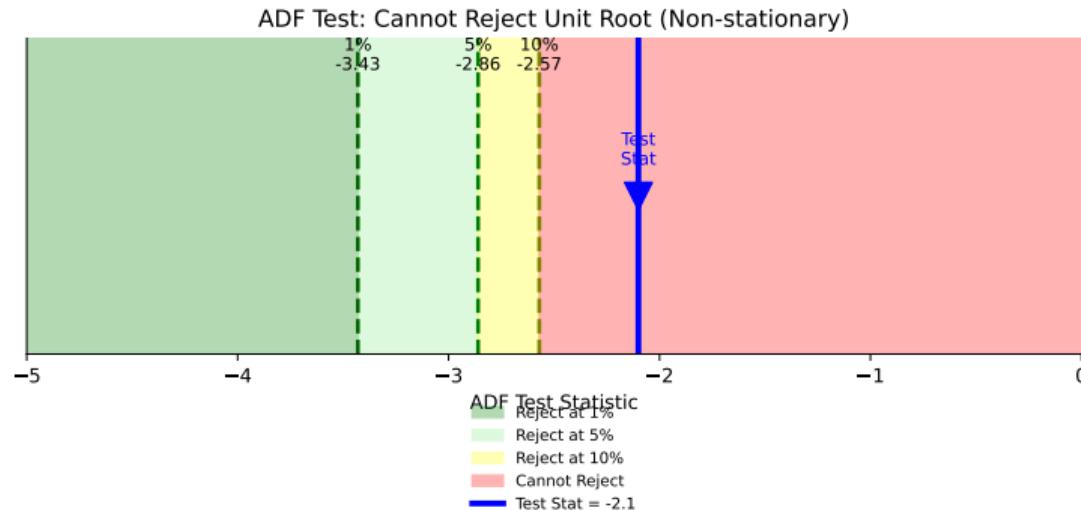
Rulați un test ADF și obțineți o statistică de test de  $-2.1$  cu valori critice:  $-3.43$  (1%),  $-2.86$  (5%),  $-2.57$  (10%). Ce concluzie trageți?

- A Respingem  $H_0$ : seria este staționară la toate nivelurile
- B Respingem  $H_0$ : seria este staționară doar la nivel de 10%
- C Nu respingem  $H_0$ : seria probabil are rădăcină unitară
- D Testul este neconcludent

## Întrebarea Quiz 3: Răspuns

Răspuns Corect: (C) Nu respingem  $H_0$ : seria are rădăcină unitară

Statistică de test  $-2.1 > -2.57$  (VC 10%)  $\Rightarrow$  Nu putem respinge la niciun nivel. Luați în considerare diferențierea.



## Întrebarea Quiz 4

### Întrebare

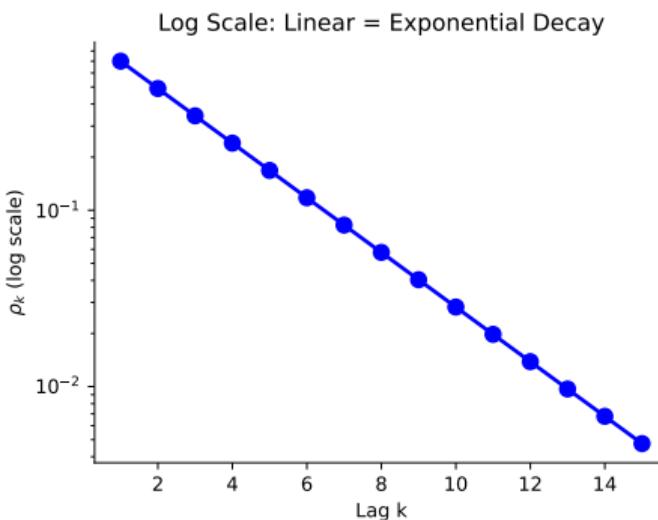
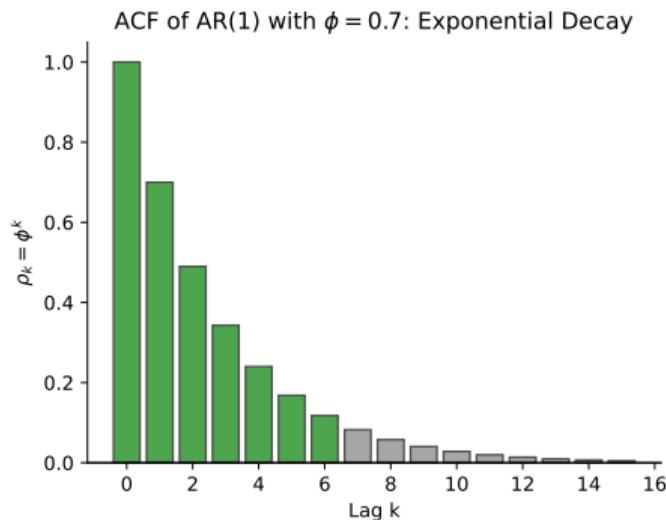
Pentru un model ARIMA(1,1,0), care este tiparul ACF al seriei diferențiate  $\Delta Y_t$ ?

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

## Întrebarea Quiz 4: Răspuns

Răspuns Corect: (B) Scade exponențial

$\text{ARIMA}(1,1,0) \Rightarrow \Delta Y_t$  urmează AR(1) cu ACF  $\rho_k = \phi_1^k$  (descreștere geometrică).



## Întrebarea Quiz 5

### Întrebare

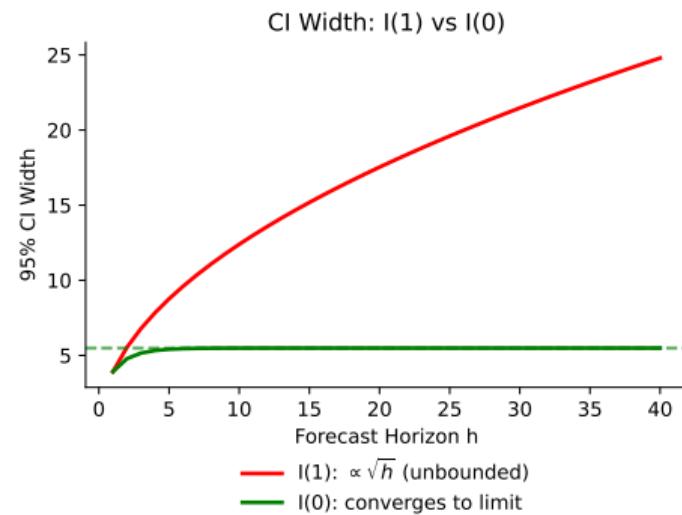
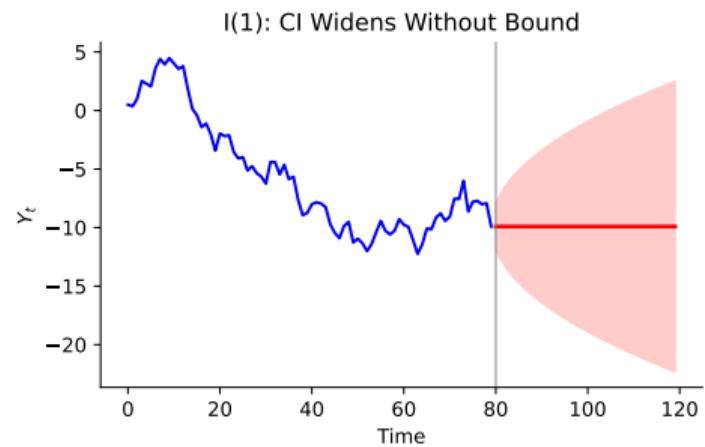
Ce se întâmplă cu intervalele de încredere ale prognozei ARIMA pe măsură ce orizontul  $h$  crește pentru o serie I(1)?

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

## Întrebarea Quiz 5: Răspuns

Răspuns Corect: (C) Se largesc nelimitat

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



## Referințe

-  Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*. 5th ed. Wiley.
-  Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
-  Enders, W. (2014). *Applied Econometric Time Series*. 4th ed. Wiley.
-  Hyndman, R.J. & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3rd ed. OTexts.