Analiza și Prognoza Seriilor de Timp

# Capitolul 1: Introducere în Serii de Timp

Fundamente și Concepte

# Obiective de Învățare

**La sfârșitul acestui capitol, veți fi capabili să:**

1. **Înțelegeți** procesele stochastice și realizările lor
2. **Definiți** staționaritatea strictă și slabă (de covarianță)
3. **Distingeți** între zgomot alb și mers aleatoriu
4. **Calculați** și interpretați funcția de autocorelație (ACF)
5. **Aplicați** funcția de autocorelație parțială (PACF)
6. **Utilizați** operatorul lag și diferențierea
7. **Efectuați** testul ADF pentru rădăcină unitate
8. **Aplicați** testul KPSS și interpretați rezultatele combinate

# Structura Capitolului

# Proces Stochastic: Definiție

## Definiție 1 (Proces Stochastic)

Un **proces stochastic** este o colecție de variabile aleatoare indexate după timp:

$$\{X_t(\omega) : t \in \mathcal{T}, \omega \in \Omega\}$$

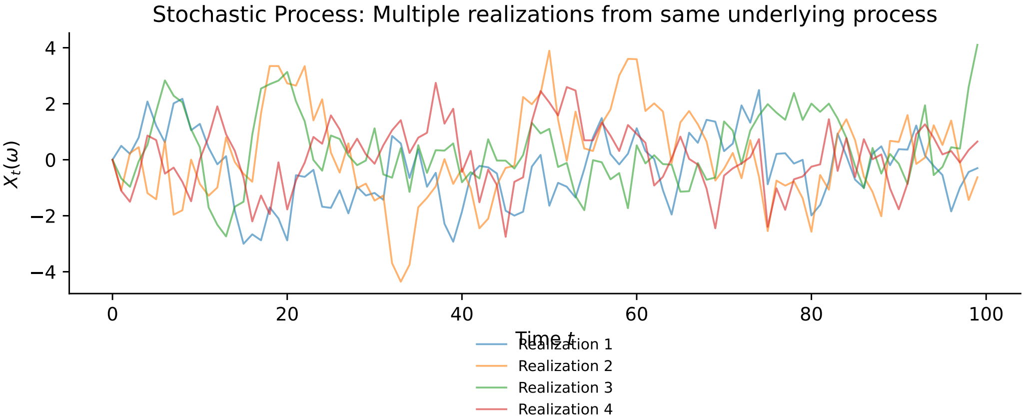
unde $\Omega$ este spațiul de selecție al rezultatelor posibile.

**Două perspective:**

- $\omega$ **fix**: O *realizare* sau *traiectorie de selecție* $\{X_t(\omega)\}_{t \in \mathcal{T}}$
- $t$ **fix**: O *variabilă aleatoare* $X_t$ cu distribuția $F_t(x)$

**Observație cheie:** O serie de timp pe care o observăm este **o realizare** a procesului stochastic subiacent. Folosim această singură realizare pentru a deduce proprietățile procesului.

# Proces Stochastic: Ilustrație Vizuală




Fiecare linie este o realizare diferită din același proces stochastic subiacent. Observăm doar o realizare dar vrem să înțelegem procesul.

# Momentele unui Proces Stochastic

**Primele două momente caracterizează proprietățile slabe:**

**Funcția de Medie:** $\mu_t = \mathbb{E}[X_t]$

**Funcția de Autocovarianță (ACVF):**

$$\gamma(t, s) = \mathrm{Cov}(X_t, X_s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)]$$

**Funcția de Autocorelație (ACF):**

$$\rho(t, s) = \frac{\gamma(t, s)}{\sqrt{\mathrm{Var}(X_t) \cdot \mathrm{Var}(X_s)}}$$

**Proprietăți:** $\rho(t, s) \in [-1, 1]$ și $\rho(t, t) = 1$

# De Ce Contează Staționaritatea

**Staționaritatea** este o ipoteză fundamentală pentru analiza seriilor de timp:

**Fără Staționaritate:**

- Media, varianța se schimbă în timp
- Trecutul poate să nu prezică viitorul
- Metodele standard eșuează
- Corelații false

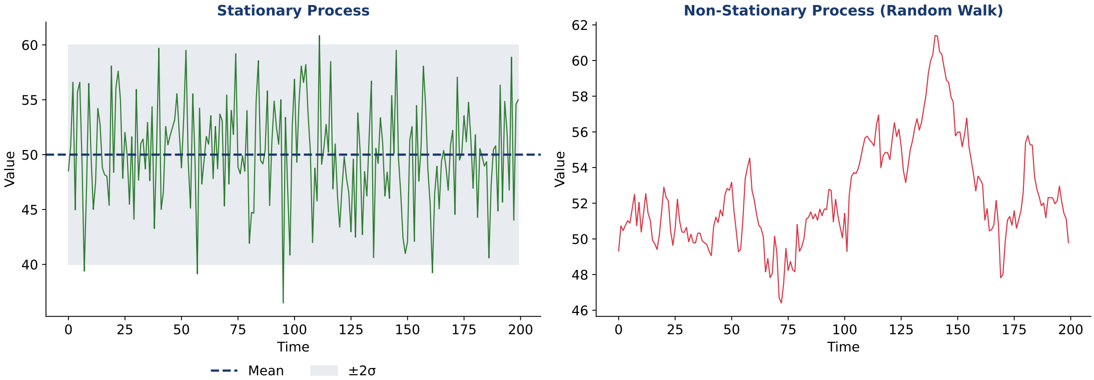
**Cu Staționaritate:**

- Proprietăți statistice constante
- Putem estima din o singură realizare
- Inferență validă posibilă
- Modelele sunt semnificative

**Principiu Cheie**

Majoritatea modelelor de serii de timp (ARMA, ARIMA, etc.) necesită staționaritate. Seriile nestaționare trebuie transformate (de ex., diferențiere) înainte de modelare.

# Staționar vs Nestaționar: Comparație Vizuală




- **Staționar**: Medie și varianță constante – fluctuează în jurul unui nivel fix
- **Nestaționar**: Media și/sau varianța se schimbă în timp
- Inspecția vizuală este primul pas; testele formale (ADF, KPSS) confirmă

# Staționaritate Strictă

**Definiție 2 (Staționaritate Strictă (Puternică))**

Un proces $\{X_t\}$ este **strict staționar** dacă pentru toți $k$, toți $t_1, \ldots, t_k$ și toți $h$:

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h})$$

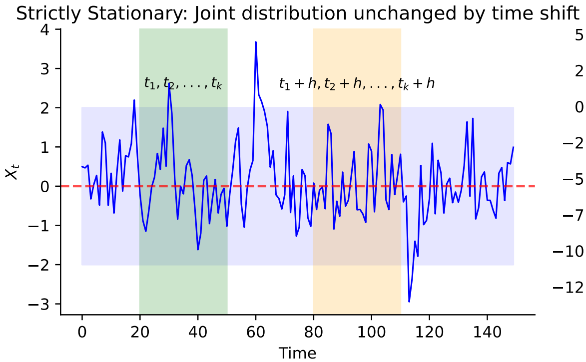
**Interpretare:** Distribuția comună a oricărei colecții de observații este **invariantă la deplasări temporale**.

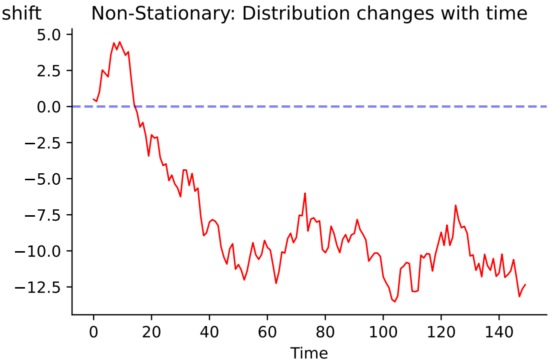
**Implicații:**

- Toate distribuțiile marginale $F_{X_t}(x)$ sunt identice
- $\mathbb{E}[X_t] = \mu$ (medie constantă)
- $\text{Var}(X_t) = \sigma^2$ (varianță constantă)
- Distribuțiile comune depind doar de *diferențele* temporale

**Notă:** Staționaritatea strictă este o condiție puternică, adesea impractică de verificat.

# Staționaritate Strictă: Ilustrație Vizuală




Staționar: oricare două ferestre au aceeași distribuție comună. Nestaționar: distribuția se schimbă în timp.

# Staționaritate Slabă (de Covarianță)

## Definiție 3 (Staționaritate Slabă)

Un proces $\{X_t\}$ este **slab staționar** (sau staționar de covarianță) dacă:

1. $\mathbb{E}[X_t] = \mu$ (medie constantă)
2. $\text{Var}(X_t) = \sigma^2 < \infty$ (varianță constantă, finită)
3. $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ (covarianța depinde doar de lag-ul $h$)

**Proprietate cheie:** Autocovarianța este o funcție doar de lag:

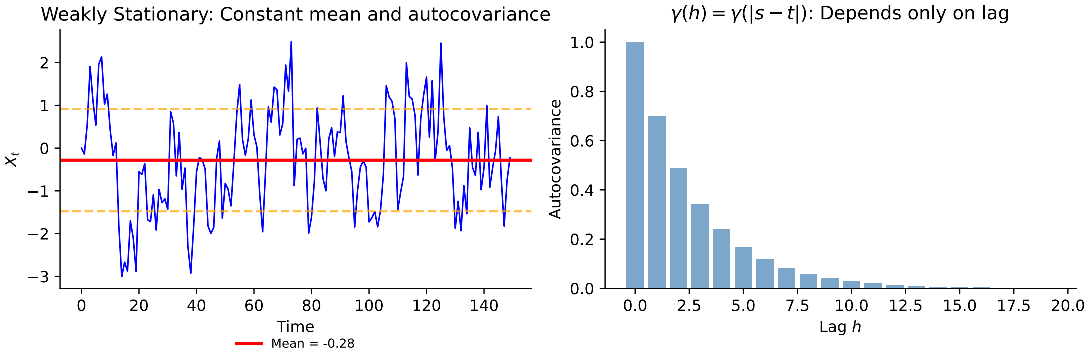
$$\gamma(h) = \text{Cov}(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)]$$

**Funcția de autocorelație:**

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\text{Cov}(X_t, X_{t+h})}{\text{Var}(X_t)}$$

Notă: $\rho(0) = 1$ și $\rho(h) = \rho(-h)$ (simetrie)

# Staționaritate Slabă: Ilustrație Vizuală




Stânga: medie și varianță constante. Dreapta: autocovarianța depinde doar de lag-ul $h$, nu de timpul $t$.

## Proprietățile Funcției de Autocovarianță

Pentru un proces slab staționar, ACVF $\gamma(h)$ satisface:

1. **Simetrie:** $\gamma(h) = \gamma(-h)$
2. **Maxim la zero:** $|\gamma(h)| \leq \gamma(0)$
3. **Definit nenegativ**

**Implicație:** Nu orice funcție poate fi o funcție de autocovarianță.

# Procesul de Zgomot Alb

## Definiție 4 (Zgomot Alb)

Un proces $\{\varepsilon_t\}$ este **zgomot alb**, notat $\varepsilon_t \sim WN(0, \sigma^2)$, dacă:

1. $\mathbb{E}[\varepsilon_t] = 0$ pentru toți $t$
2. $\text{Var}(\varepsilon_t) = \sigma^2$ pentru toți $t$
3. $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ pentru $t \neq s$

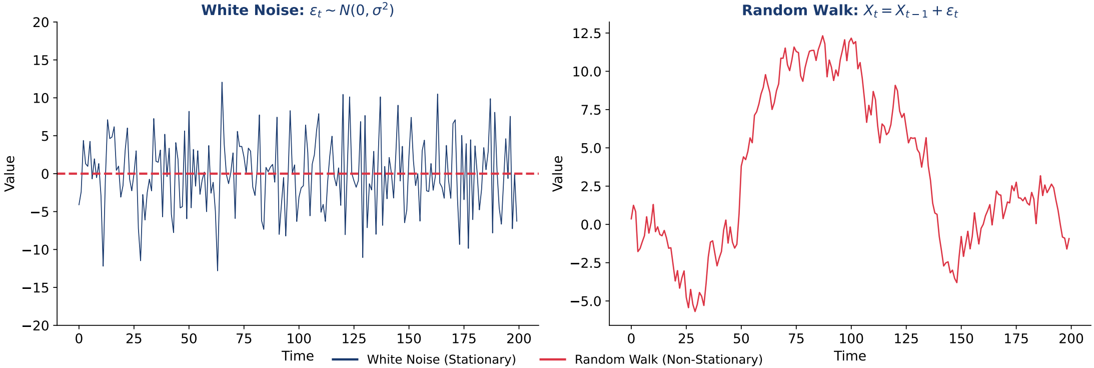
**ACF al Zgomotului Alb:**

$$\rho(h) = \begin{cases} 1 & \text{dacă } h = 0 \\ 0 & \text{dacă } h \neq 0 \end{cases}$$

**Tipuri:**

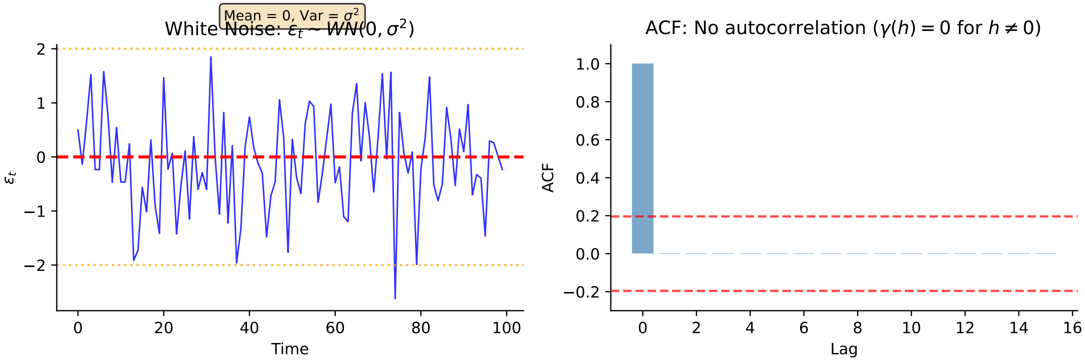
- **Zgomot alb slab**: Necorelat (condițiile de mai sus)
- **Zgomot alb puternic**: Independent și identic distribuit (i.i.d.)
- **Zgomot alb Gaussian**: $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$

# Zgomot Alb vs Mers Aleatoriu: Comparație




- **Zgomot alb**: Fluctuează în jurul lui zero – staționar, varianță constantă
- **Mers aleatoriu**: Suma cumulativă a zgomotului alb – rătăcește, nestaționar
- Mersul aleatoriu este cel mai simplu proces nestaționar (rădăcină unitate)

# Zgomot Alb: Ilustrație Vizuală




Stânga: zgomotul alb fluctuează în jurul lui zero cu varianță constantă. Dreapta: ACF arată nicio autocorelație (toate zero după lag 0).

# Procesul de Mers Aleatoriu

**Definiție:** $X_t = X_{t-1} + \varepsilon_t$ unde $\varepsilon_t \sim WN(0, \sigma^2)$, $X_0 = 0$

**Forma explicită:** $X_t = \sum_{i=1}^{t} \varepsilon_i$

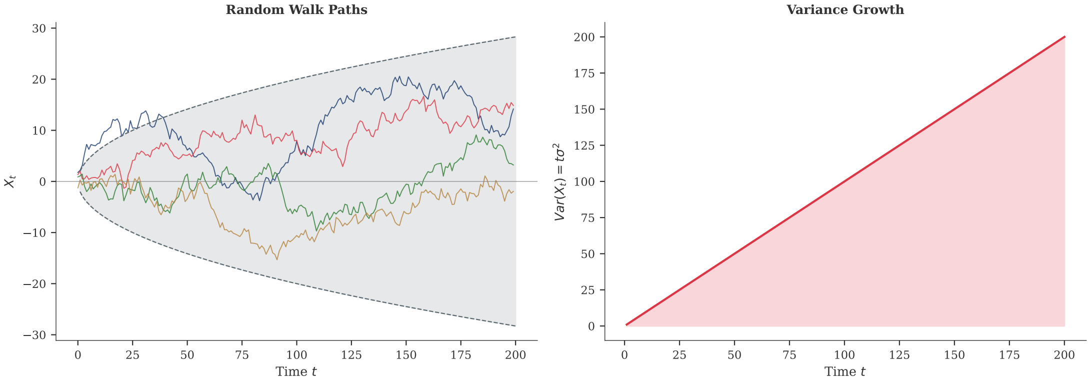
**Proprietăți:**

- $\mathbb{E}[X_t] = 0$ (medie constantă)
- $\text{Var}(X_t) = t\sigma^2$ (varianța crește în timp!)
- $\text{Cov}(X_t, X_s) = \min(t, s) \cdot \sigma^2$

**Nestaționar!**

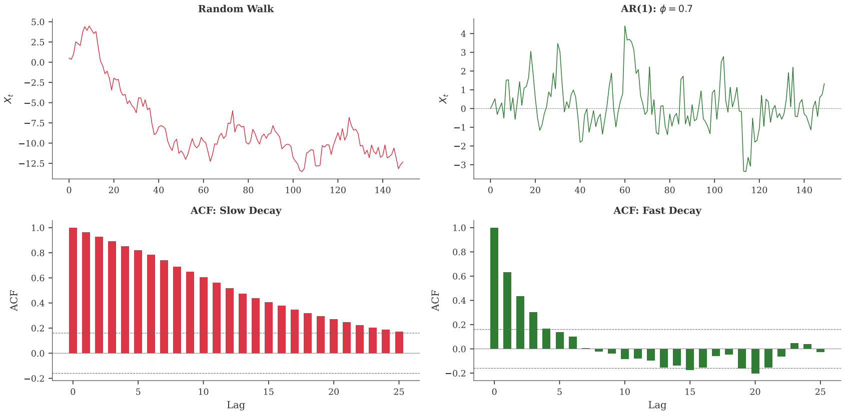
Mersul aleatoriu **nu este staționar** deoarece varianța depinde de $t$.

# Mers Aleatoriu: Vizualizare




**Stânga:** Traiectorii multiple divergă în timp. **Dreapta:** Varianța crește liniar: $\text{Var}(X_t) = t\sigma^2$.

# Staționar vs Nestaționar: Comparație




**Diagnostic cheie:** ACF al procesului staționar scade rapid; ACF al mersului aleatoriu scade foarte lent.

# Funcția de Autocorelație din Eșantion

**ACF din eșantion la lag-ul $h$:**

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

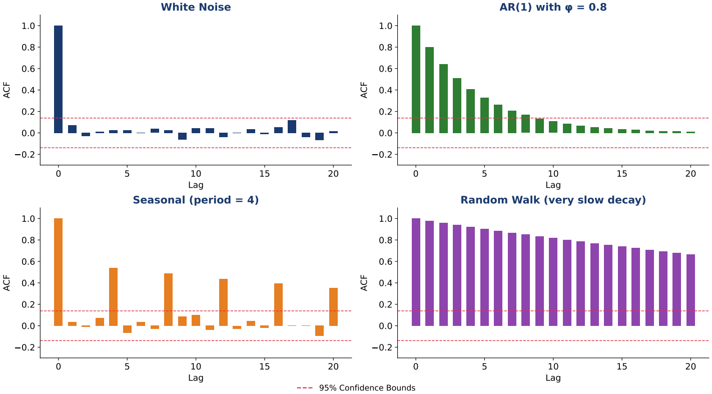
**Proprietăți:**

- $\hat{\rho}(0) = 1$ întotdeauna
- $|\hat{\rho}(h)| \leq 1$

**Test de semnificație:** Sub zgomot alb, $\hat{\rho}(h) \approx N(0, 1/T)$

**Limite 95%:** $\pm 1.96/\sqrt{T}$

# Tipare ACF pentru Diferite Procese




- **Zgomot alb**: ACF scade la zero imediat (nicio dependență)
- **AR(1)**: ACF scade exponențial – indică structură autoregresivă
- **Sezonier**: ACF arată vârfuri la lag-uri sezoniere (de ex., 12, 24 pentru lunar)
- **Mers aleatoriu**: ACF scade foarte lent – semn de nestaționaritate

# Funcția de Autocorelație Parțială (PACF)

**PACF** $\phi_{hh}$: Corelația dintre $X_t$ și $X_{t+h}$ după eliminarea efectului liniar al $X_{t+1}, \ldots, X_{t+h-1}$.

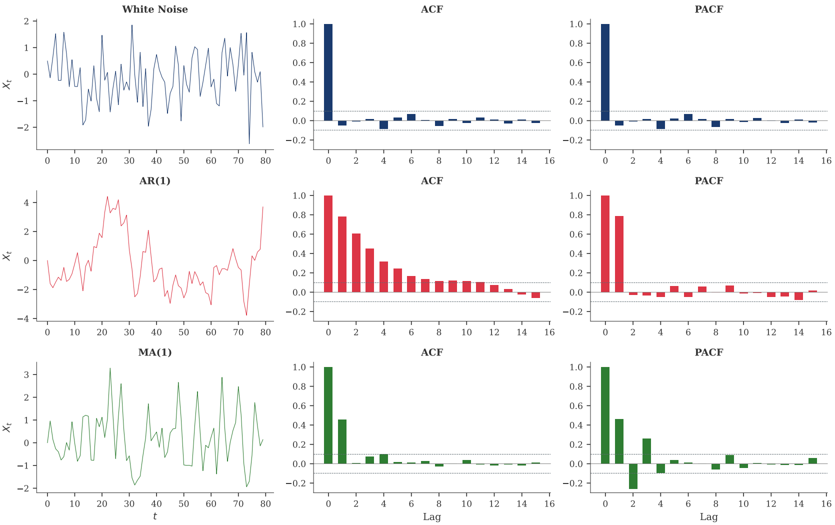
**Interpretare:**

- $\phi_{11} = \rho(1)$ (același ca ACF la lag 1)
- $\phi_{22} =$ corelația lui $X_t, X_{t+2}$ controlând pentru $X_{t+1}$
- Măsoară dependența *directă* la lag-ul $h$

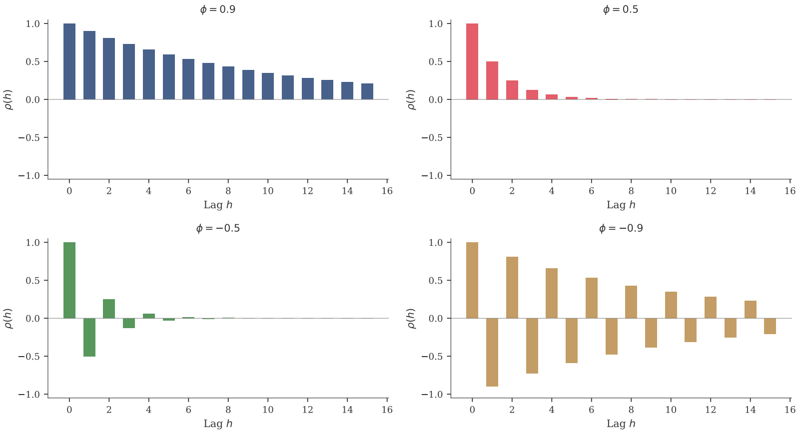
**Aplicație cheie:** Identificarea ordinului AR

- Pentru AR($p$): PACF **se întrerupe** după lag-ul $p$
- Pentru MA($q$): ACF **se întrerupe** după lag-ul $q$

# Tipare ACF și PACF



# ACF Teoretic pentru AR(1)




Pentru AR(1): $X_t = \phi X_{t-1} + \varepsilon_t$, ACF teoretic este $\rho(h) = \phi^h$.

# Operatorul Lag

## Definiție 5 (Operatorul Lag)

**Operatorul lag** (sau operatorul de întârziere) $L$ este definit prin:

$$LX_t = X_{t-1}$$

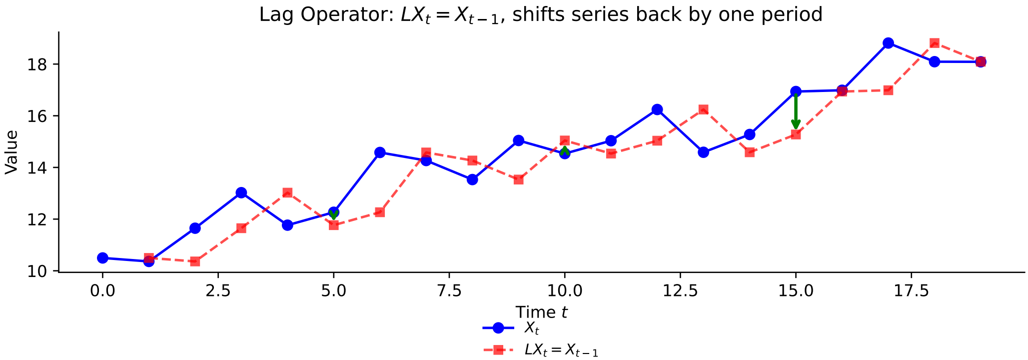
**Proprietăți:**

- $L^k X_t = X_{t-k}$ (întârziere cu $k$ perioade)
- $L^0 = I$ (identitate)
- $(1 - \phi L)X_t = X_t - \phi X_{t-1}$

**Exemple:**

- AR(1): $(1 - \phi L)X_t = \varepsilon_t$
- MA(1): $X_t = (1 + \theta L)\varepsilon_t$
- AR($p$): $(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)X_t = \varepsilon_t$

# Operatorul Lag: Ilustrație Vizuală




Operatorul lag $L$ deplasează fiecare observație înapoi cu o perioadă de timp: $LX_t = X_{t-1}$.

## Diferențierea

**Prima diferență:** $\Delta X_t = X_t - X_{t-1} = (1 - L)X_t$

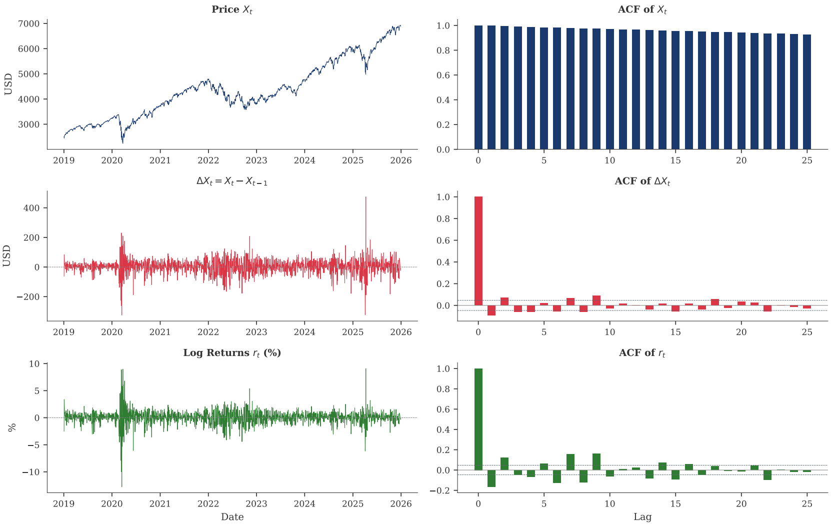
**De ce diferențiem?**

- Elimină trendul și rădăcina unitate
- Mers aleatoriu: $\Delta X_t = \varepsilon_t$ (zgomot alb)

**Proces integrat:** $X_t \sim I(d)$ dacă $\Delta^d X_t$ este staționar

- $I(0)$: Staționar (nu necesită diferențiere)
- $I(1)$: Necesită o diferențiere
- $I(2)$: Necesită două diferențieri

# Efectul Diferențierii: S&P 500



# Testul Augmented Dickey-Fuller (ADF)

**Model:** $\Delta X_t = \alpha + \gamma X_{t-1} + \sum_{i=1}^{p} \delta_i \Delta X_{t-i} + \varepsilon_t$

**Ipoteze:**

- $H_0$: $\gamma = 0$ (rădăcină unitate)
- $H_1$: $\gamma < 0$ (staționar)

**Decizie:**

- $\tau <$ valoare critică $\Rightarrow$ Respingem $H_0$ $\Rightarrow$ Staționar
- $\tau \geq$ valoare critică $\Rightarrow$ Nestaționar

Valori critice: distribuția Dickey-Fuller (nu normală)

**Statistica de test:**

$$\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

# Testul KPSS

**Model:** $X_t = \xi t + r_t + \varepsilon_t$ unde $r_t = r_{t-1} + u_t$

**Ipoteze (opuse față de ADF):**

- $H_0$: $\sigma_u^2 = 0$ (staționar)
- $H_1$: $\sigma_u^2 > 0$ (rădăcină unitate)

**Statistica de test:**

$$LM = \frac{\sum_{t=1}^{T} S_t^2}{T^2 \hat{\sigma}^2}$$

unde $S_t = \sum_{i=1}^{t} \hat{e}_i$

**Decizie:**

- $LM >$ valoare critică $\Rightarrow$ Respingem $H_0$ $\Rightarrow$ Nestaționar
- $LM \leq$ valoare critică $\Rightarrow$ Staționar

**Notă:** KPSS complementează ADF—folosiți ambele pentru concluzii robuste.

# Folosirea ADF și KPSS Împreună

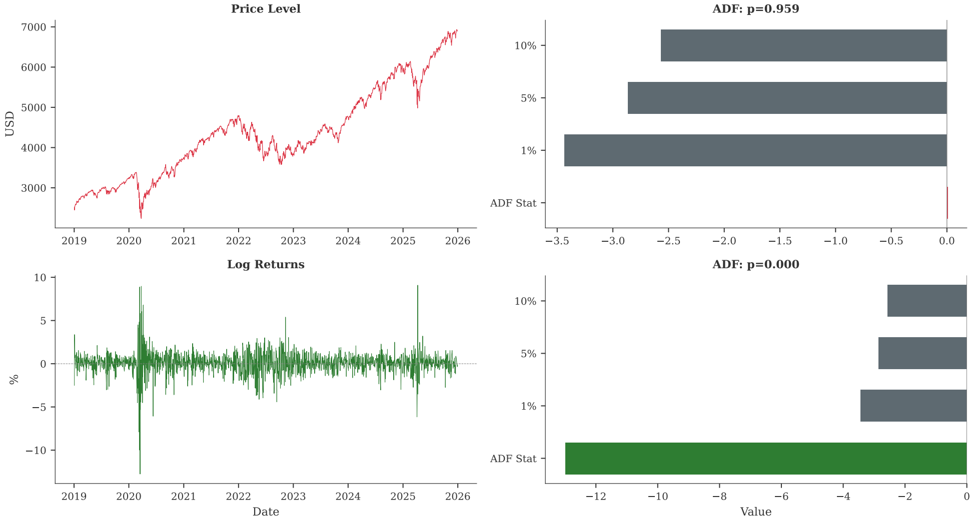
**Testare confirmatorie** pentru concluzii robuste:

| ADF | KPSS | Concluzie |
|---|---|---|
| Respingem $H_0$ | Nu respingem $H_0$ | Staționar |
| Nu respingem $H_0$ | Respingem $H_0$ | Rădăcină Unitate |
| Respingem $H_0$ | Respingem $H_0$ | Neconcludent |
| Nu respingem $H_0$ | Nu respingem $H_0$ | Neconcludent |

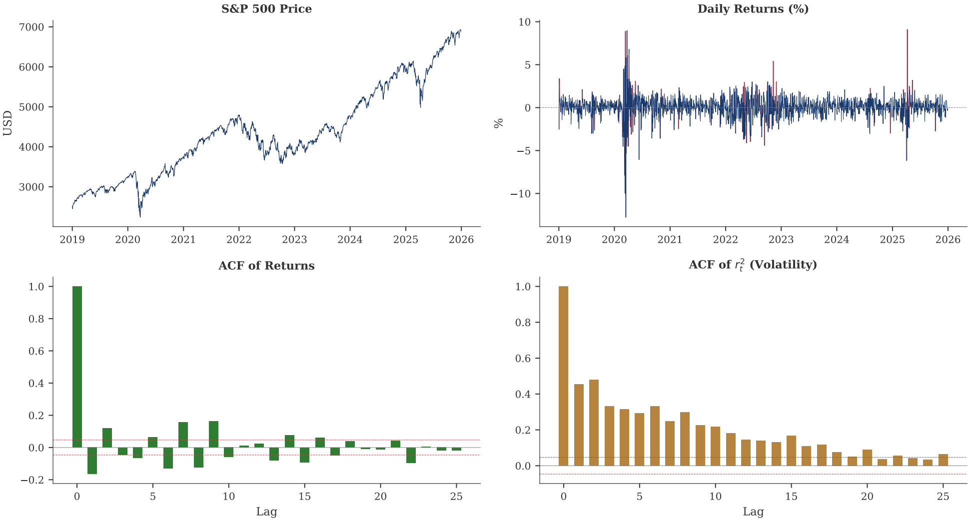
**Flux de lucru recomandat:**

1. Rulați testul ADF (nulă = rădăcină unitate)
2. Rulați testul KPSS (nulă = staționar)
3. Dacă rezultatele coincid, procedați cu încredere
4. Dacă neconcludent, considerați teste alternative (PP, DF-GLS)

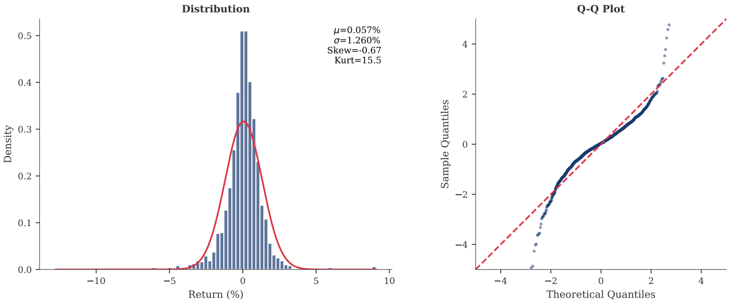
# Testul ADF: Vizualizare cu S&P 500



# Analiza S&P 500: Prezentare Generală



# Fapte Stilizate ale Randamentelor Financiare



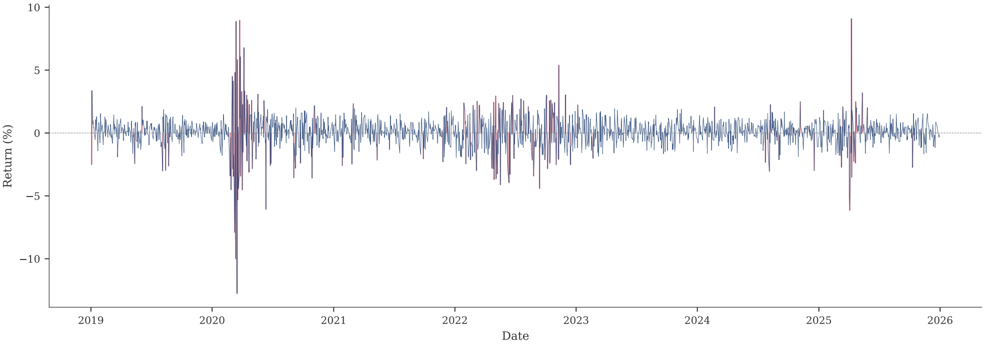

**Proprietăți observate:**

- Asimetrie negativă (coadă stângă)
- Kurtoză excesivă ($\gg 3$)
- Cozi groase (heavy tails)

**Implicații:**

- Distribuția normală inadecvată
- Evenimente extreme mai probabile
- Necesită distribuție Student-t sau similară

# Gruparea Volatilității




## Fapt Stilizat

Randamentele mari (pozitive sau negative) tind să fie urmate de randamente mari. Această **grupare a volatilității** motivează modelele ARCH/GARCH (capitolele viitoare).

# Concluzii Cheie

1. **Proces stochastic** = colecție de variabile aleatoare indexate după timp
2. **Realizare** = o traiectorie observată din procesul stochastic subiacent
3. **Staționaritate slabă**: Medie constantă, varianță constantă, autocovarianță depinde doar de lag
4. **Zgomot alb**: Proces necorelat cu medie zero și varianță constantă
5. **Mers aleatoriu**: Suma cumulativă a zgomotului alb — nestaționar
6. **ACF/PACF**: Instrumente esențiale pentru identificarea structurii de dependență
7. **Operatorul lag**: $LX_t = X_{t-1}$; diferențiere: $\Delta X_t = X_t - X_{t-1}$
8. **Testul ADF**: $H_0$: rădăcină unitate (nestaționar)
9. **Testul KPSS**: $H_0$: staționar — folosiți împreună cu ADF pentru concluzii robuste

# Formule Importante I

## Staționaritate Slabă

$\mathbb{E}[X_t] = \mu$ (constantă), $\quad \mathrm{Var}(X_t) = \sigma^2$ (constantă), $\quad \gamma(t,s) = \gamma(|t-s|)$

## Zgomot Alb

$\mathbb{E}[\varepsilon_t] = 0, \quad \mathrm{Var}(\varepsilon_t) = \sigma^2, \quad \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0$ pentru $t \neq s$

## Mers Aleatoriu

$X_t = X_{t-1} + \varepsilon_t \quad \Rightarrow \quad \mathrm{Var}(X_t) = t\sigma^2$ (nestaționar)

## Autocovarianță și Autocorelație

$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}) \qquad \rho(h) = \frac{\gamma(h)}{\gamma(0)}$

# Formule Importante II

## Operatorul Lag

$LX_t = X_{t-1}, \quad L^k X_t = X_{t-k}$

## Diferențiere

$\Delta X_t = (1 - L)X_t = X_t - X_{t-1}$

## Testul ADF

$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \sum_{i=1}^{p} \delta_i \Delta X_{t-i} + \varepsilon_t$

$H_0 : \gamma = 0$ (rădăcină unitate) vs $H_1 : \gamma < 0$ (staționar)

## Testul KPSS

$H_0$: Seria este staționară vs $H_1$: Seria are rădăcină unitate

# Previzualizare Capitolul Următor

**Capitolul 2: Modele ARMA**

- Modele Autoregresive (AR)
- Modele de Medie Mobilă (MA)
- Modele ARMA combinate
- Identificarea modelului folosind ACF/PACF
- Estimarea parametrilor
- Diagnosticarea modelului
- Prognoza

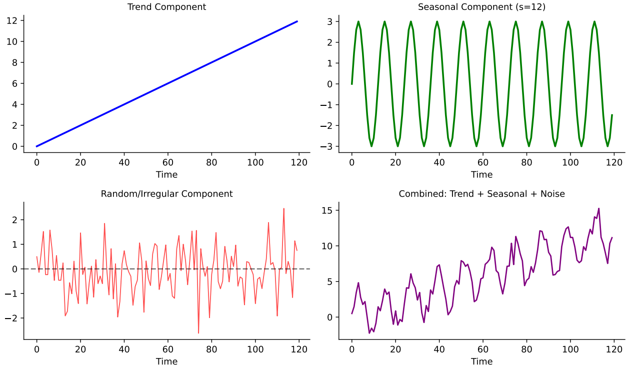
# Întrebarea Quiz 1

## Întrebare

O serie de timp $Y_t$ arată mișcare ascendentă de-a lungul anilor plus tipare repetitive în fiecare trimestru. Ce componente sunt prezente?

- A Doar trend
- B Doar sezonalitate
- C Trend și Sezonalitate
- D Doar zgomot aleatoriu

# Întrebarea Quiz 1: Răspuns




**Răspuns Corect: (C) Trend și Sezonalitate**

Mișcare ascendentă = Trend; Tipare trimestriale = Sezonalitate (s=4)

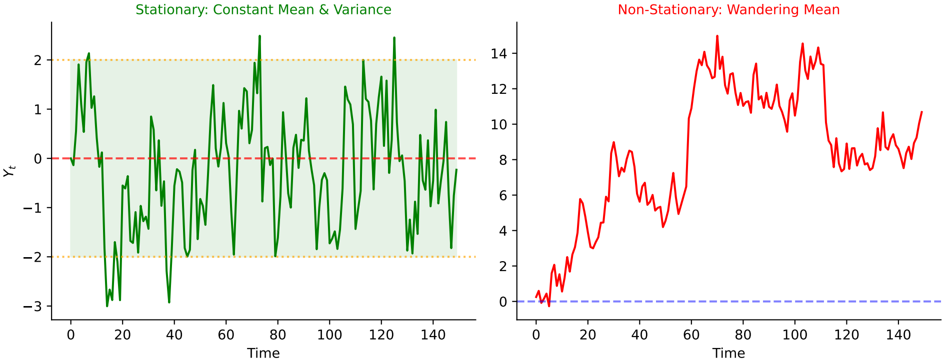
# Întrebarea Quiz 2

**Întrebare**

Care dintre următoarele este o caracteristică a unei serii de timp staționare?

- A. Media se schimbă în timp
- B. Varianța crește în timp
- C. Medie și varianță constante în timp
- D. Conține o componentă de trend

# Întrebarea Quiz 2: Răspuns




**Răspuns Corect: (C) Medie și varianță constante în timp**

Staționaritatea necesită: medie constantă, varianță constantă și autocovarianța depinde doar de lag.

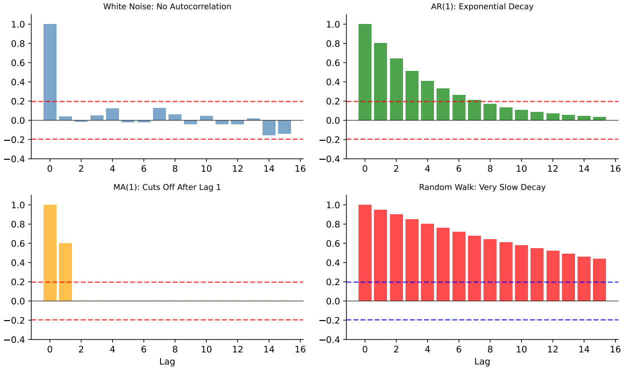
# Întrebarea Quiz 3

## Întrebare

Pentru un proces de zgomot alb, cum arată ACF la lag-uri $k > 0$?

- A. Descreștere exponențială
- B. Toate valorile semnificative și pozitive
- C. Toate valorile aproximativ zero (în interiorul benzilor de încredere)
- D. Alternare pozitiv și negativ

# Întrebarea Quiz 3: Răspuns




**Răspuns Corect: (C) Aproximativ zero în interiorul benzilor de încredere**

Zgomotul alb nu are autocorelație: $\rho_k = 0$ pentru toți $k \neq 0$.

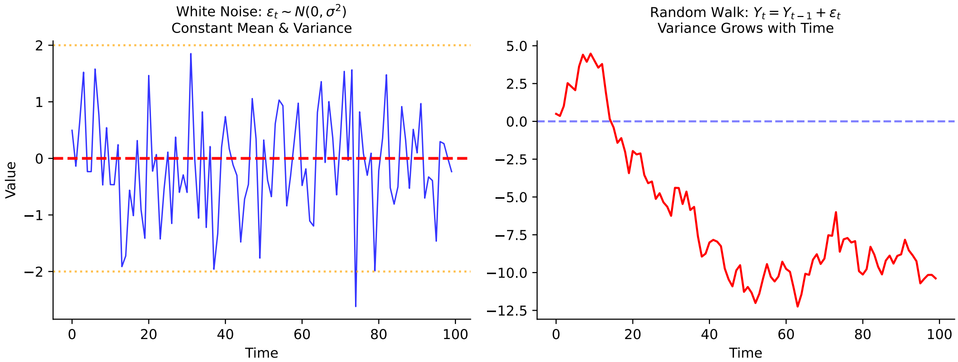
# Întrebarea Quiz 4

**Întrebare**

Care este diferența cheie între zgomotul alb și mersul aleatoriu?

- A. Zgomotul alb are trend, mersul aleatoriu nu
- B. Mersul aleatoriu este suma cumulativă a zgomotului alb
- C. Ambele sunt procese staționare
- D. Zgomotul alb are varianță mai mare

# Întrebarea Quiz 4: Răspuns




**Răspuns Corect: (B) Mers aleatoriu = suma cumulativă a zgomotului alb**

$Y_t = Y_{t-1} + \varepsilon_t = \sum_{i=1}^{t} \varepsilon_i$ unde $\varepsilon_t$ este zgomot alb.

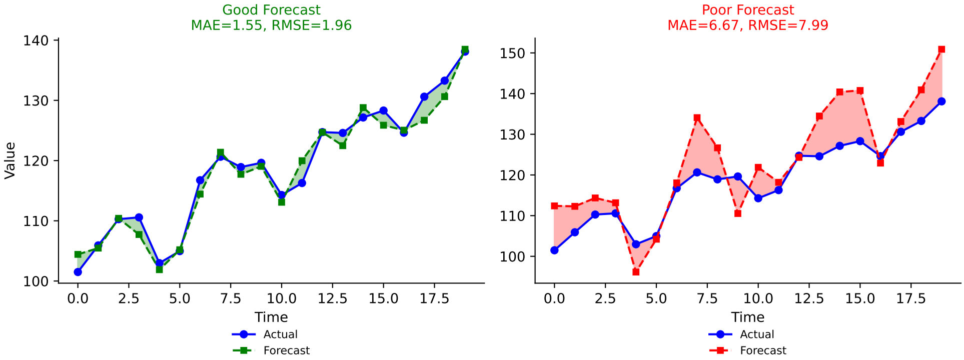
# Întrebarea Quiz 5

**Întrebare**

Care metrică de eroare a prognozei este cea mai sensibilă la erori mari (valori aberante)?

- A. MAE (Eroarea Medie Absolută)
- B. RMSE (Rădăcina Erorii Medii Pătratice)
- C. MAPE (Eroarea Medie Absolută Procentuală)
- D. Toate sunt la fel de sensibile

# Întrebarea Quiz 5: Răspuns




## Răspuns Corect: (B) RMSE

RMSE ridică la pătrat erorile, deci erorile mari au impact disproporționat: $\sqrt{\frac{1}{n}\sum e_t^2}$

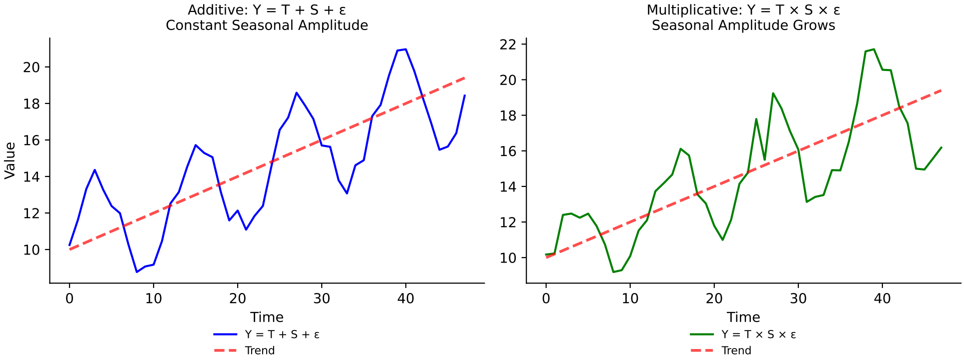
# Întrebarea Quiz 6

**Întrebare**

Când ar trebui să folosiți descompunerea multiplicativă în loc de cea aditivă?

- A. Când seria nu are trend
- B. Când amplitudinea sezonieră este constantă
- C. Când amplitudinea sezonieră crește odată cu nivelul seriei
- D. Când seria este staționară

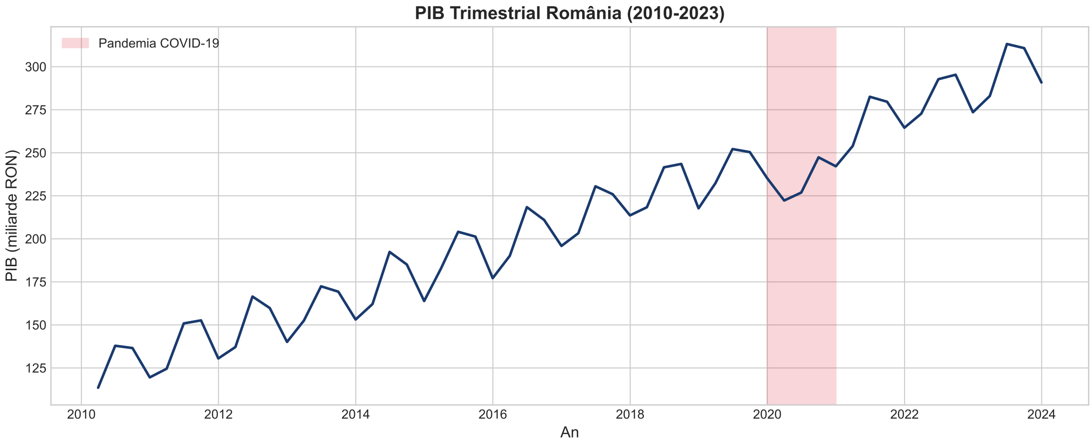
# Întrebarea Quiz 6: Răspuns




**Răspuns Corect: (C) Amplitudinea sezonieră crește odată cu nivelul**

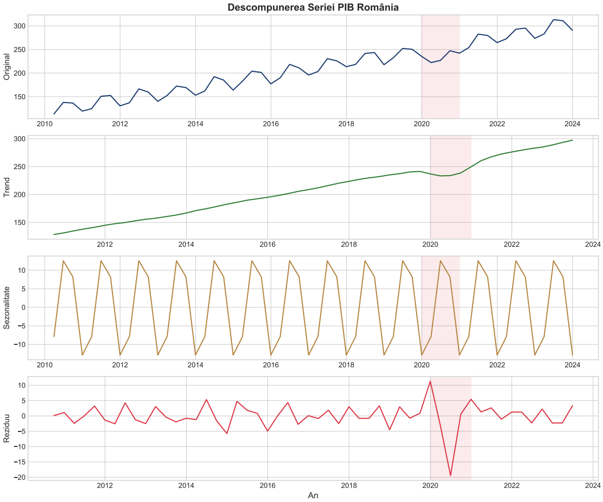
Multiplicativă: $Y_t = T_t \times S_t \times \varepsilon_t$ — oscilațiile sezoniere proporționale cu trendul.

# Studiu de Caz: PIB Trimestrial România




- **Date**: PIB trimestrial România, 2010–2023 (sursa: INS/Eurostat)
- **Observații**: Trend crescător, sezonalitate trimestrială, șoc COVID-19 în 2020

# Descompunerea Seriei PIB



# Interpretarea Componentelor

## Componente Identificate

- **Trend**: Creștere economică susținută
- **Sezonalitate**: Pattern trimestrial regulat (Q4 > Q1)
- **Reziduu**: Include șocul COVID-19 din 2020

## Lecții Învățate

- Descompunerea ajută la înțelegerea structurii datelor
- Șocurile externe (COVID) apar în componenta reziduală
- Sezonalitatea trebuie modelată explicit

## Următorii Pași

În capitolele următoare vom învăța să modelăm fiecare componentă: ARIMA pentru trend, SARIMA pentru sezonalitate.

# Referințe


- Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3rd ed., OTexts.
- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*. 5th ed., Wiley.
- Tsay, R.S. (2010). *Analysis of Financial Time Series*. 3rd ed., Wiley.
- Cleveland, R.B., Cleveland, W.S., McRae, J.E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition. *Journal of Official Statistics*, 6(1), 3-73.

# Surse de Date

## Date Reale Utilizate în Acest Capitol

- **Pasageri Aviație**: Set de date clasic Box-Jenkins, 1949–1960
- **S&P 500**: Yahoo Finance (SPY), date istorice
- **Pete Solare**: Set de date Statsmodels, observații lunare

## Software și Instrumente

- **Python**: statsmodels, pandas, matplotlib, yfinance
- **R**: pachetele forecast, tseries
- **Surse de Date**: Yahoo Finance, FRED Economic Data

# Vă Mulțumesc!

Întrebări?

*Grafice generate folosind Python (statsmodels, matplotlib)*

Materiale curs disponibile la: `https://github.com/danpele/Time-Series-Analysis`