



# Analiza și Prognoza Seriilor de Timp

## Capitolul 4: Modele SARIMA pentru Serii de Timp Sezoniere



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. **Identificați** tiparele sezoniere în datele de tip serie de timp
2. **Aplicați** diferențierea sezonieră pentru a elimina rădăcinile unitare sezoniere
3. **Construiți** și estimați modele SARIMA (Seasonal Autoregressive Integrated Moving Average) cu componente sezoniere
4. **Interpretați** tiparele ACF (Autocorrelation Function) / PACF (Partial ACF) sezoniere pentru identificarea modelului
5. **Evaluați** prognozele prin metoda ferestrei mobile (rolling) pentru date sezoniere
6. **Aplicați** metodologia completă pe date reale (pasageri aerieni)

## Cuprins

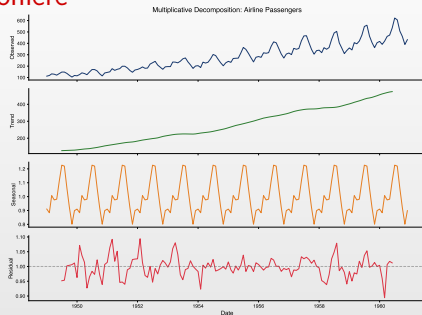
- Motivație
- Sezonalitatea în seriile de timp
- Diferențierea sezonieră
- Modelul SARIMA
- Tipare ACF și PACF sezoniere
- Estimare și validare
- Prognoza cu SARIMA
- Studiu de caz: Număr de pasageri
- Aspecte practice
- Utilizare IA
- Rezumat
- Quiz
- Bibliografie

## De ce SARIMA? Sezonalitatea este peste tot



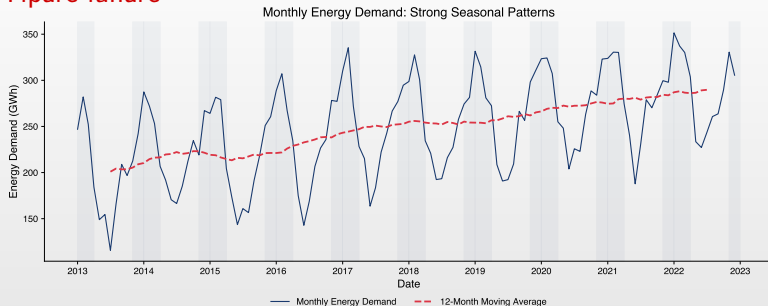
- Vânzările cu amănuntul prezintă **tipare anuale** clare: vârfuri în decembrie, minime în ianuarie
- Modelele ARIMA (Autoregressive Integrated Moving Average) standard nu pot captura aceste **cicluri sezoniere repetitive**
- Ignorarea sezonaliității duce la erori sistematice de prognoză

## Înțelegerea componentelor sezoniere



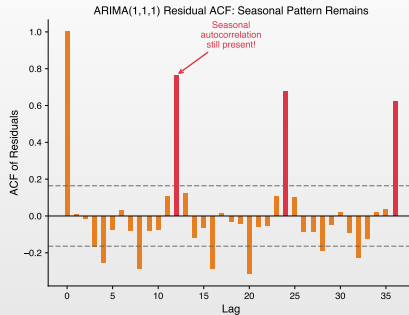
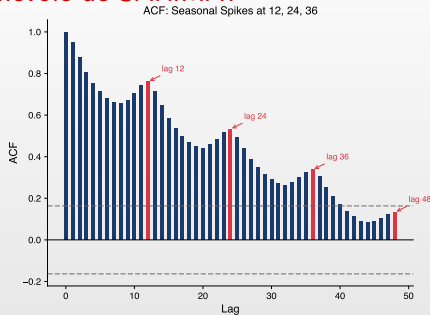
- ▣ Serie de timp sezonieră = **Trend + Tipar sezonier + Reziduuri**
- ▣ Descompunerea ajută la vizualizarea separată a fiecărei componente
- ▣ Modelele SARIMA captează atât dinamica trendului, cât și comportamentul sezonier

## Aplicație reală: Tipare lunare



- Cererea de energie prezintă o **sezonalitate lunară** puternică (cicluri de încălzire/răcire)
- Tiparul se repetă previzibil în fiecare an cu mici variații
- Companiile de utilități folosesc prognozele SARIMA pentru planificarea capacității

## De ce avem nevoie de SARIMA?



- **Stânga:** ACF sezonieră prezintă vârfuri la lag-urile 12, 24, 36... (tipar anual)
- **Dreapta:** Reziduurile ARIMA încă prezintă autocorelație sezonieră  $\Rightarrow$  modelul este incomplet
- SARIMA adaugă **termeni AR (autoregresivi) și MA (medie mobilă) sezonieri** pentru a captura aceste tipare

## Ce este sezonalitatea?

### Definiție 1 (Sezonalitate)

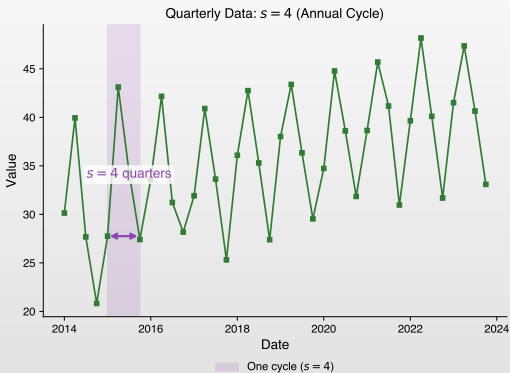
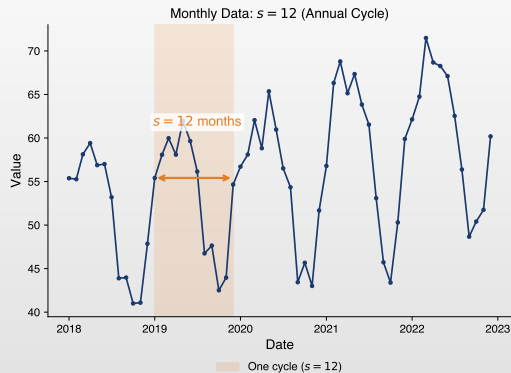
- O serie de timp prezintă **sezonalitate** când arată fluctuații regulate, periodice care se repetă pe o perioadă fixă  $s$  (perioadă sezonieră)

### Perioade sezoniere comune

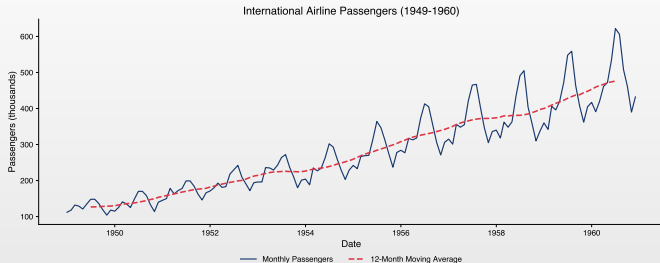
- Date lunare:  $s = 12$  (ciclu anual)
- Date trimestriale:  $s = 4$  (ciclu anual)
- Date săptămânale:  $s = 52$  (anual) sau  $s = 7$  (tipar săptămânal)
- Date zilnice:  $s = 7$  (tipar săptămânal)



## Ce este sezonalitatea? (vizualizare)



## Exemplu: Datele privind pasagerii companiilor aeriene



- ▣ Pasageri internaționali lunari (1949–1960)
- ▣ **Trend ascendent** clar și **amplitudine sezonieră crescătoare**
- ▣ Vârfurile din vară reflectă tiparele călătoriilor de vacanță

## Exemple de date sezoniere

### Serii economice

- ▣ Vânzări cu amănuntul (vârfuri de sărbători)
- ▣ Turism (vara/iarna)
- ▣ Producție agricolă
- ▣ Consum de energie
- ▣ Ocuparea forței de muncă (cicluri de angajare)

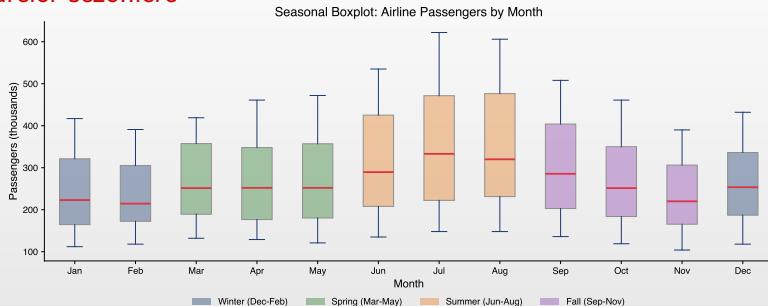
### Alte domenii

- ▣ Vreme/temperatură
- ▣ Trafic pe site-uri web
- ▣ Admisii la spital
- ▣ Utilizarea transportului
- ▣ Cererea de electricitate

### De ce contează

- ▣ Ignorarea sezonității duce la prognoze distorsionate și inferență invalidă.

## Vizualizarea tiparelor sezoniere



- Diagrama box plot relevă un tipar sezonier consistent
- Iulie–August: cele mai mari numere de pasageri (călătorii de vară)
- Noiembrie–Februarie: cele mai mici numere (lunile de iarnă)

## Sezonalitate deterministă vs stochastică

## Sezonalitate deterministă

- ▣ **Tipar fix:**  $Y_t = \sum_{j=1}^s \gamma_j D_{jt} + \varepsilon_t$ 
  - $D_{jt}$  sunt variabile dummy sezoniere
- ▣ **Notă:** se omite o variabilă dummy și se include o constantă, sau se impune  $\sum \gamma_j = 0$ , pentru a evita colinearitatea perfectă
- ▣ Tiparul constant în timp
- ▣ Aceeași amplitudine în fiecare an
- ▣ Eliminare prin regresie pe dummy-uri
- ▣ ACF: scădere bruscă la lag-uri sezoniere
- ▣ **Exemplu:** Înscrierile universitare cresc în fiecare septembrie cu aceeași valoare

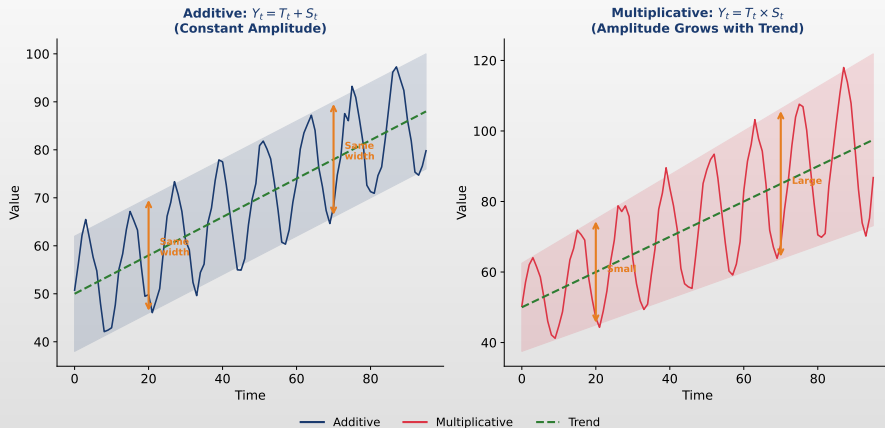
## Sezonalitate stochastică

- ▣ **Tipar în evoluție:**  $\Delta_s Y_t = Y_t - Y_{t-s}$ 
  - Prezintă structură de dependență
- ▣ Tiparul evoluează în timp
- ▣ Amplitudinea poate crește sau scădea
- ▣ Necesită diferențiere sezonieră
- ▣ ACF: scădere lentă la lag-uri sezoniere
- ▣ **Exemplu:** Vânzările cu amănuntul cresc tot mai mult în fiecare decembrie

## Cum decidem?

- ▣ Scădere lentă a ACF la lag-urile  $s, 2s, 3s, \dots \Rightarrow$  stochastică (folosiți  $\Delta_s$ )
- ▣ Scădere bruscă  $\Rightarrow$  deterministă (folosiți dummy-uri)
- ▣ Confirmați cu testele HEGY (Hylleberg-Engle-Granger-Yoo) sau Canova-Hansen

## Sezonalitate aditivă vs multiplicativă (vizualizare)



## Sezonalitate aditivă vs multiplicativă

Aditivă:  $Y_t = T_t + S_t + \varepsilon_t$

- ▣ Amplitudinea sezonieră **constantă**
- ▣ Nu necesită transformare
- ▣ Ex: temperaturi, înscrieri universitare

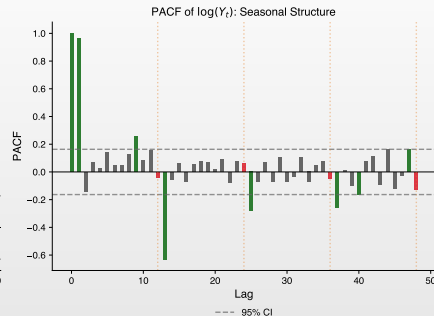
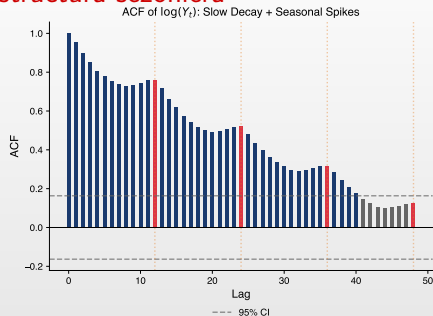
Multiplicativă:  $Y_t = T_t \cdot S_t \cdot \varepsilon_t$

- ▣ Amplitudinea **crește cu nivelul**
- ▣ Necesită transformare log (Box-Cox)
- ▣ Ex: Airline, vânzări cu amănuntul, PIB

### Prima decizie practică

- ▣ Amplitudinea crește cu trendul?  $\Rightarrow$  multiplicativă  $\Rightarrow$  aplicați log/Box-Cox *înainte* de diferențiere

## ACF relevă structura sezonieră



- Descreștere lentă la toate lag-urile indică nestăționaritate (trend)
- Vârfuri la lag-urile 12, 24, 36 confirmă tiparul sezonier ( $s = 12$ )
- ACF la lag-urile sezoniere: descreștere lentă  $\Rightarrow$  necesită diferențiere sezonieră



## Detectarea sezonaliții

### Metode vizuale

- ▣ Graficul seriei de timp – căutați tipare repetitive
- ▣ Graficul sub-seriilor sezoniere – comparați aceleași sezoane de-a lungul anilor
- ▣ Graficul ACF – vârfuri la lag-uri sezoniere ( $s, 2s, 3s, \dots$ )

### Teste statistice

- ▣ Teste de rădăcină unitară sezonieră (HEGY, Canova-Hansen, OCSB<sup>a</sup>)
- ▣ Testul F pentru variabile dummy sezoniere
- ▣ Testul Kruskal-Wallis (neparametric)

### Semnatura ACF

- ▣ Sezonalitate puternică: ACF prezintă vârfuri semnificative la lag-urile  $s, 2s, 3s, \dots$

---

<sup>a</sup>Osborn-Chui-Smith-Birchenhall — testul implicit din `auto_arima`

## Testul F pentru variabile dummy sezoniere: intuiție

### Ce face acest test?

- ▣ **Scop:** testează dacă valorile medii diferă semnificativ între sezoane
- ▣ **Logică:** dacă media din ianuarie  $\neq$  media din februarie  $\neq \dots \neq$  media din decembrie
- ▣ Diferențe semnificative între sezoane  $\Rightarrow$  sezonalitate prezentă
- ▣ **Metodă:** compară un model CU variabile dummy sezoniere vs. un model FĂRĂ

### Modelele comparate

- ▣ **Restricționat:**  $Y_t = \alpha + \varepsilon_t$     **Nerestricționat:**  $Y_t = \alpha + \sum_{j=1}^{s-1} \gamma_j D_{jt} + \varepsilon_t$
- ▣ unde  $D_{jt} = 1$  dacă observația  $t$  este în sezonul  $j$ , 0 altfel

### De reținut

- ▣ Dacă adăugarea dummy sezoniere **reduce semnificativ** erorile de predicție  $\Rightarrow$  sezonalitate prezentă

## Testul F pentru variabile dummy sezoniere: formula și exemplu

### Formula statisticii F

- **Formula:**  $F = \frac{(SSR_R - SSR_U)/(s-1)}{SSR_U/(n-s)} \sim F_{s-1, n-s}$ 
  - ▶  $SSR_R$ : suma pătratelor reziduurilor din modelul restricționat (fără dummy)
  - ▶  $SSR_U$ : suma pătratelor reziduurilor din modelul nerestricționat (cu dummy)
  - ▶  $s - 1$ : numărul de restricții (lunar: 11, trimestrial: 3)

### Exemplu numeric (date lunare, n=120)

- $SSR_R = 15000$ ,  $SSR_U = 8500$ ,  $s = 12$
- $F = \frac{(15000 - 8500)/11}{8500/108} = \frac{590.9}{78.7} = 7.51$
- Valoare critică  $F_{0.05, 11, 108} \approx 1.87$ . Cum  $7.51 > 1.87$ : **Respingem  $H_0 \Rightarrow$  Sezonalitate.**

## Testul Kruskal-Wallis: intuiție

### Ce face acest test?

- ▣ **Test neparametric:** verifică dacă observațiile din diferite sezoane provin din aceeași distribuție (nu necesită ipoteze despre forma distribuției)
- ▣ **Mecanism:** ordonează toate observațiile și înlocuiește valorile cu ranguri (1, 2, ...,  $N$ )
- ▣ **Verificare:** dacă rangurile sunt distribuite uniform între sezoane; compară suma rangurilor per sezon cu distribuția așteptată
- ▣ **Concluzie:** dacă un sezon are constant ranguri mai mari/mici  $\Rightarrow$  sezonalitate

### De ce să-l folosim în locul testului F?

- ▣ **Fără ipoteza de normalitate** – funcționează cu orice distribuție; ideal pentru date financiare cu cozi groase
- ▣ **Robust la valori extreme** – rangurile limitează influența oricărei observații individuale

### Limitare

- ▣ Mai puțin puternic decât testul F când datele *sunt* distribuite normal

## Testul Kruskal-Wallis: formula și exemplu

### Statistică de test

$$\square H = \frac{12}{N(N+1)} \sum_{j=1}^s \frac{R_j^2}{n_j} - 3(N+1) \quad \text{unde } N = \text{total obs.}, n_j = \text{obs. în sezonul } j, R_j = \text{suma rangurilor}$$

### Exemplu: Vânzări trimestriale (n=20, s=4)

- Date ordonate 1-20. Sumele rangurilor: T1:  $R_1 = 15$ , T2:  $R_2 = 35$ , T3:  $R_3 = 70$ , T4:  $R_4 = 90$
- $H = \frac{12}{20 \times 21} \left( \frac{15^2}{5} + \frac{35^2}{5} + \frac{70^2}{5} + \frac{90^2}{5} \right) - 3(21) = 19.6$
- Valoarea critică  $\chi_{0.05,3}^2 = 7.81$ . Deoarece  $19.6 > 7.81$ : **Respingem  $H_0 \Rightarrow$  Sezonalitate.**

### În Python

- **Implementare:** `scipy.stats.kruskal(q1, q2, q3, q4)`

## Testul HEGY: ce problemă rezolvă?

### Întrebarea fundamentală

- ▣ **Problemă:** având o serie sezonieră, trebuie să știm tipul de diferențiere (alegerea greșită duce la supra-/sub-diferențiere)
- ▣ **Diferențiere obișnuită**  $(1 - L)? \Rightarrow$  setăm  $d = 1$ ; **Diferențiere sezonieră**  $(1 - L^s)? \Rightarrow$  setăm  $D = 1$
- ▣ **HEGY:** testează pentru ambele tipuri de rădăcini unitare simultan! (Hylleberg, Engle, Granger, Yoo, 1990)

### De ce să nu folosim doar ADF?

- ▣ **ADF:** testează doar pentru o rădăcină unitară obișnuită la frecvența zero; ignoră componentele sezoniere
- ▣ **Limitare:** datele sezoniere pot avea rădăcini unitare la frecvențe sezoniere pe care ADF le omite.

### HEGY testează frecvențe multiple

- ▣ **Trimestrial:** testează la  $0, \pi, \pm\pi/2$  (3 frecvențe  $\Rightarrow$  o rădăcină obișnuită + două sezoniere)
- ▣ **Lunar:** testează la  $0, \pi, \pm\pi/6, \pm\pi/3, \pm\pi/2, \pm2\pi/3, \pm5\pi/6$  (6 frecvențe)

## Testul HEGY: Formula de regresie (trimestrial)

### Regresia auxiliară HEGY

- ▣ **Date trimestriale** ( $s = 4$ ):  $\Delta_4 y_t = \pi_1 z_{1,t-1} + \pi_2 z_{2,t-1} + \pi_3 z_{3,t-2} + \pi_4 z_{4,t-2} + \sum_{j=1}^k \phi_j \Delta_4 y_{t-j} + \varepsilon_t$
- ▣ Lag-urile  $k$ : alese prin AIC / BIC

### Variabile transformate

- ▣  $z_{1t}$ :  $(1 + L + L^2 + L^3)y_t = y_t + y_{t-1} + y_{t-2} + y_{t-3}$  - frecvență zero (trend)
- ▣  $z_{2t}$ :  $-(1 - L + L^2 - L^3)y_t = -y_t + y_{t-1} - y_{t-2} + y_{t-3}$  - frecvență  $\pi$  (alternanță semestrială)
- ▣  $z_{3t}, z_{4t}$ : componentele de frecvență  $\pm\pi/2$  (ciclu anual)

### Ipoteze

- ▣  $H_0 : \pi_1 = 0$ : rădăcină unitară la frecvența 0  $\Rightarrow$  necesită  $d = 1$
- ▣  $H_0 : \pi_2 = 0$ : la frecvența  $\pi \Rightarrow$  necesită  $D = 1$
- ▣  $H_0 : \pi_3 = \pi_4 = 0$ : rădăcină unitară la frecvența  $\pm\pi/2 \Rightarrow$  necesită  $D = 1$

## Testul HEGY: Reguli de decizie cu exemple

Valori critice HEGY (5%,  $n=100$ , cu constanta)

| Test                             | Statistică | Valoare critică | Dacă NU este respins... |
|----------------------------------|------------|-----------------|-------------------------|
| $t_1$ ( $\pi_1 = 0$ )            | t-stat     | -2.88           | Necesită $d = 1$        |
| $t_2$ ( $\pi_2 = 0$ )            | t-stat     | -2.88           | Necesită $D = 1$        |
| $F_{34}$ ( $\pi_3 = \pi_4 = 0$ ) | F-stat     | 6.57            | Necesită $D = 1$        |

### Exemplu: PIB trimestrial

- ▣ **Rezultate HEGY:**  $t_1 = -1.52$ ,  $t_2 = -4.21$ ,  $F_{34} = 2.15$
- ▣  $t_1 = -1.52 > -2.88$ : Nu putem respinge  $\Rightarrow$  **necesită**  $d = 1$
- ▣  $t_2 = -4.21 < -2.88$ : Respingem  $\Rightarrow$  fără rădăcină unitară la  $\pi$
- ▣  $F_{34} = 2.15 < 6.57$ : Nu putem respinge  $\Rightarrow$  **necesită**  $D = 1$
- ▣ **Concluzie:** Folosim SARIMA cu  $d = 1$ ,  $D = 1$



## Testul Canova-Hansen: opusul testului HEGY

### HEGY vs Canova-Hansen: Ipoteze nule diferite

|                 | HEGY                              | Canova-Hansen                          |
|-----------------|-----------------------------------|--|
| $H_0$           | Rădăcină unitară sezonieră        | <b>Fără</b> rădăcină unitară sezonieră |
| $H_1$           | Fără rădăcină unitară sezonieră   | Rădăcină unitară sezonieră             |
| Respingem $H_0$ | Folosim variabile dummy sezoniere | Folosim diferențiere $(1 - L^s)$       |
| Nu respingem    | Folosim diferențiere $(1 - L^s)$  | Folosim variabile dummy sezoniere      |

### De ce contează?

- HEGY: "Demonstrați că *nu* există rădăcină unitară" (conservator față de diferențiere)
- CH: "Demonstrați că *există* rădăcină unitară" (conservator față de variabile dummy)
- Folosiți **ambele** teste pentru concluzii robuste.

## Testul Canova-Hansen: formula

### Procedura de testare

- ▣ **Pas 1:** Regresam  $y_t$  pe variabile dummy sezoniere:  $y_t = \sum_{j=1}^s \gamma_j D_{jt} + u_t$
- ▣ **Pas 2:** Calculam sumele parțiale la frecvența sezonieră  $\lambda_i$ :
  - ▶  $S_{it}^{(c)} = \sum_{j=1}^t \hat{u}_j \cos(\lambda_i j), \quad S_{it}^{(s)} = \sum_{j=1}^t \hat{u}_j \sin(\lambda_i j)$

### Statistică de test LM (Lagrange Multiplier)

- ▣  $LM_i = \frac{1}{T^2 \hat{\omega}_i} \left[ \sum_{t=1}^T (S_{it}^{(c)})^2 + \sum_{t=1}^T (S_{it}^{(s)})^2 \right]$
- ▣ unde  $\hat{\omega}_i$  = estimator consistent al densității spectrale la frecvența  $\lambda_i$

### Decizie

- ▣ **Regula:** respingem  $H_0$  dacă  $LM > \text{valoare critică} \Rightarrow \text{necesară diferențierea sezonieră}$

## Sumar: Alegerea testului de sezonalitate potrivit

### Comparație teste de sezonalitate

| Test           | $H_0$                        | Dacă respingem      | Cel mai bun pentru          |
|----------------|------------------------------|---------------------|-----------------------------|
| Test F         | Fără sezonalitate            | Sezonalitate există | Date normale                |
| Kruskal-Wallis | Fără diferență între sezoane | Sezonalitate există | Non-normale, valori extreme |
| HEGY           | Rădăcină unitară există      | Folosim dummy       | Determinarea $d, D$         |
| Canova-Hansen  | Fără rădăcină unitară        | Folosim $(1 - L^s)$ | Confirmarea stabilității    |

### Concluzie

- **Test F/Kruskal-Wallis:** "*Există sezonalitate?*"
- **HEGY/CH:** "*Ce tip?*" (deterministă vs stochastică)
- **În practică:** pentru serii cu sezonalitate evidentă, boxplot + ACF sunt suficiente
- Testele formale sunt esențiale când sezonalitatea este ambiguă

## Transformarea Box-Cox: Stabilizarea varianței

### Familia de transformări Box-Cox

- ▣ **Formula:**  $Y_t^{(\lambda)} = (Y_t^\lambda - 1)/\lambda$  dacă  $\lambda \neq 0$ ;  $Y_t^{(\lambda)} = \ln(Y_t)$  dacă  $\lambda = 0$
- ▣ **Cazuri speciale:**  $\lambda = 1$  (fără transformare),  $\lambda = 0$  (logaritm),  $\lambda = 0.5$  (rădăcină pătrată);  $\lambda < 1$  comprimă valorile mari

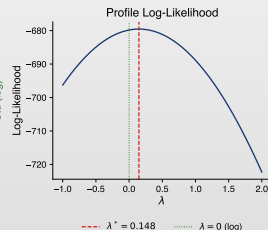
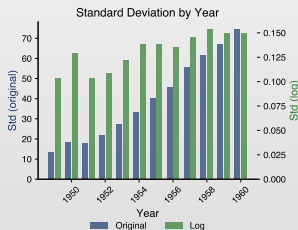
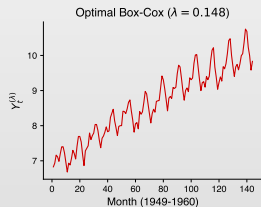
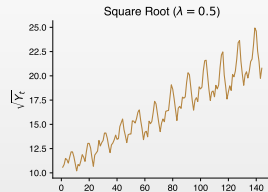
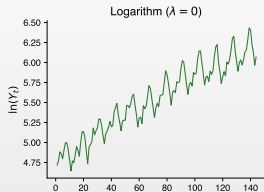
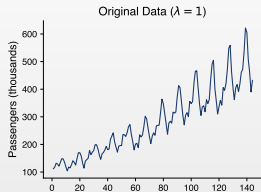
### Selectarea automată a lui $\lambda$

- ▣ **Verosimilitate profilată:** maximizează log-verosimilitatea în funcție de  $\lambda$  (cea mai utilizată metodă)
- ▣ **Metoda Guerrero (1993):** minimizează coeficientul de variație al sub-seriilor sezoniere; preferată pentru date sezoniere
- ▣ **Python:** `boxcox(y)` din `scipy.stats`; **R:** `BoxCox.lambda_(y)` din pachetul `forecast`

### De ce nu doar logaritm?

- ▣ Log ( $\lambda = 0$ ) presupune varianță proporțională cu nivelul – nu este întotdeauna cazul ( $\lambda = 0.3$  poate fi mai bun)
- ▣ Box-Cox alege transformarea optimală bazat pe date, nu pe ipoteze

## Box-Cox pe datele Airline: Exemplu complet



## Box-Cox pe datele Airline: Exemplu complet

### Rezultat pentru Airline passengers

- ▣  $\hat{\lambda} = 0.148 \approx 0 \Rightarrow \log$  e aproape optim
- ▣ Abaterea standard pe an: de la crescătoare (original) la stabilă (log)

### Corecția de bias la back-transformare

- ▣ Pe scara log:  $\hat{y}_{T+h}$  este **mediana**, nu media
- ▣ Corecție:  $\hat{Y}_{T+h} = \exp\left(\hat{y}_{T+h} + \frac{\sigma_h^2}{2}\right)$
- ▣ Fără corecție: prognoze sistematic sub-estimate.

## Descompunerea STL: Alternative moderne

### STL: Seasonal-Trend decomposition using Loess (Cleveland et al., 1990)

- ▣ **Avantaje:** sezonalitate variabilă în timp, robustă la valori aberante, orice perioadă s
- ▣ **Algoritm:** regresie locală ponderată (loess) iterativă

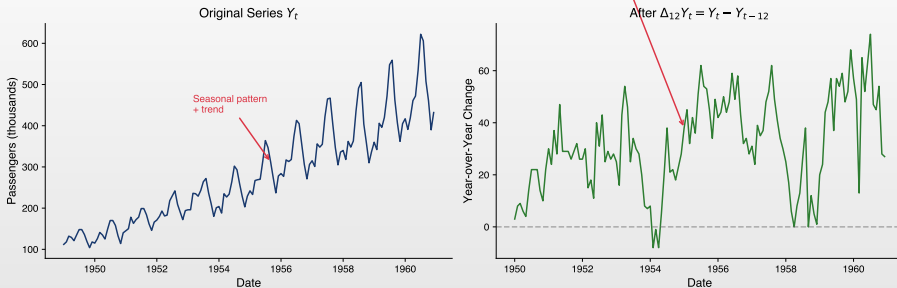
### Parametri

- ▣ **Fereastra sezonieră** (`seasonal`): controlează cât de rapid se schimbă sezonalitatea
- ▣ **Fereastra de trend** (`trend`): netezirea componentei de trend
- ▣ **Robustitate** (`robust=True`): reduce influența valorilor aberante

### Utilizare practică

- ▣ STL pentru explorare și pre-procesare; SARIMA pentru modelare și prognoză
- ▣ Python: `STL(y, period=12).fit()` din `statsmodels`

## Diferența sezonieră: Ilustrare vizuală



- **Stânga:** Seria originală cu tipar sezonier clar
- **Dreapta:** După  $\Delta_{12} = (1 - L^{12})$ , tiparul sezonier este eliminat
  - Comparația an-la-an elimină efectele sezoniere



## Operatorul de diferență sezonieră

### Definiție 2 (Diferență sezonieră)

- **Operatorul de diferență sezonieră**  $\Delta_s$  este definit ca:

$$\Delta_s Y_t = (1 - L^s) Y_t = Y_t - Y_{t-s}$$

- unde  $L^s Y_t = Y_{t-s}$  este operatorul de lag sezonier

### Exemple

- **Date lunare** ( $s = 12$ ):  $\Delta_{12} Y_t = Y_t - Y_{t-12}$ 
  - ▶ Compară fiecare lună cu aceeași lună din anul trecut
- **Date trimestriale** ( $s = 4$ ):  $\Delta_4 Y_t = Y_t - Y_{t-4}$ 
  - ▶ Compară fiecare trimestru cu același trimestru din anul trecut

## Demonstrație: diferențierea sezonieră elimină sezonalitatea deterministă

### Afirmație

- Dacă  $Y_t = \mu_t + \varepsilon_t$  unde  $\mu_t = \mu_{t-s}$  (medie periodică, se repetă la fiecare  $s$  perioade), atunci  $\Delta_s Y_t$  elimină media sezonieră

### Demonstrație

- Fie  $Y_t = \mu_t + \varepsilon_t$  cu  $\mu_t$  de perioadă  $s$ . Aplicăm  $\Delta_s$ :  
$$\Delta_s Y_t = (\mu_t + \varepsilon_t) - (\mu_{t-s} + \varepsilon_{t-s}) = \varepsilon_t - \varepsilon_{t-s} \quad (\text{deoarece } \mu_t = \mu_{t-s})$$

### Proprietățile lui $\Delta_s Y_t = \varepsilon_t - \varepsilon_{t-s}$

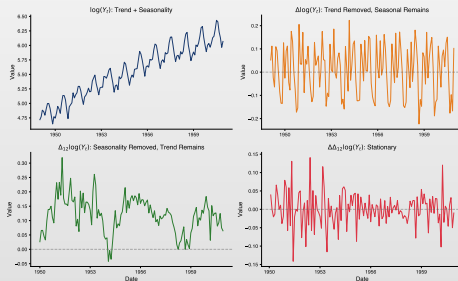
- $\mathbb{E}[\Delta_s Y_t] = 0$  (medie constantă);  $\text{Var}(\Delta_s Y_t) = 2\sigma^2$  (varianță constantă) – staționaritate de ordin 2
- **Autocovarianța:**  $\gamma(s) = -\sigma^2$ ,  $\gamma(k) = 0$  pentru  $k \neq 0, s$  (ACF: un singur vârf la lag-ul  $s$ , apoi zero)

### Rezultat

- **Concluzie:** diferențierea sezonieră transformă tiparul sezonier periodic în MA(1) la lag-ul sezonier

## Efectul operațiilor de diferențiere

- Diferențierea obișnuită elimină trendul dar tiparul sezonier rămâne
- Diferențierea sezonieră elimină sezonabilitatea dar tiparul de trend rămâne
- **Ambele diferențe** sunt necesare pentru a atinge staționaritatea



## Combinarea diferențierii obișnuite și sezoniere

### Diferențiere completă

- ▣ **Serii cu trend și sezonalitate:**  $\Delta\Delta_s Y_t = (1 - L)(1 - L^s) Y_t$

### Dezvoltare

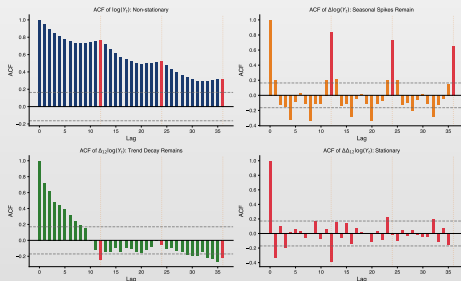
- ▣ **General:**  $(1 - L)(1 - L^s) Y_t = Y_t - Y_{t-1} - Y_{t-s} + Y_{t-s-1}$
- ▣ **Date lunare ( $s = 12$ ):**  $\Delta\Delta_{12} Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$

### Ordinea diferențierii

- ▣  $d$ : numărul de diferențe obișnuite (eliminarea trendului)
- ▣  $D$ : numărul de diferențe sezoniere (eliminarea trendului sezonier)

## ACF înainte și după diferențiere

- ACF originală: descreștere lentă indică nestaționaritate
- După  $\Delta$ : vârfuri sezoniere rămân la lag-urile 12, 24, 36
- După  $\Delta_{12}$ : descreșterea de trend rămâne la lag-urile inițiale
- După  $\Delta\Delta_{12}$ : ACF se oprește brusc  $\Rightarrow$  staționară



## Integrare sezonieră

### Definiție 3 (Proces integrat sezonier)

- O serie  $Y_t$  este **integrată sezonier** de ordinul  $(d, D)_s$ , scrisă  $Y_t \sim I(d, D)_s$ , dacă:

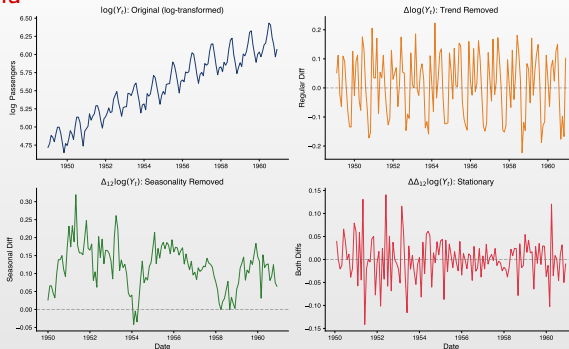
$$(1 - L)^d (1 - L^s)^D Y_t$$

- este staționară

### Cazuri comune

- $I(1, 0)_{12}$ : Doar rădăcină unitară obișnuită (lunară)
- $I(0, 1)_{12}$ : Doar rădăcină unitară sezonieră
- $I(1, 1)_{12}$ : Atât rădăcină unitară obișnuită cât și sezonieră

## SARIMA: Ilustrare vizuală



- Originală  $\Rightarrow$  diferență obișnuită (elimină trendul)  $\Rightarrow$  diferență sezonieră (elimină sezonalitatea)
- Aplicați diferențierea minimă necesară pentru staționaritate

## Definiția modelului SARIMA

### Definiție 4 ( $\text{SARIMA}(p, d, q) \times (P, D, Q)_s$ )

- Modelul **Seasonal ARIMA** este:

$$\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^DY_t = c + \theta(L)\Theta(L^s)\varepsilon_t$$

### Componente

- $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ : AR non-sezonier
- $\Phi(L^s) = 1 - \Phi_1 L^s - \dots - \Phi_P L^{Ps}$ : AR sezonier
- $\theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q$ : MA non-sezonier
- $\Theta(L^s) = 1 + \Theta_1 L^s + \dots + \Theta_Q L^{Qs}$ : MA sezonier
- $(1-L)^d$ : Diferențiere obișnuită;  $(1-L^s)^D$ : Diferențiere sezonieră



## Demonstrație: structura multiplicativă sezonieră

### De ce multiplicativă?

- Considerăm  $\text{SARIMA}(1, 0, 0) \times (1, 0, 0)_s$ :  $(1 - \phi L)(1 - \Phi L^s)Y_t = \varepsilon_t$

### Dezvoltăm produsul

- $(1 - \phi L)(1 - \Phi L^s)Y_t = Y_t - \phi Y_{t-1} - \Phi Y_{t-s} + \phi\Phi Y_{t-s-1}$
- Rezultat:** modelul include un termen de interacțiune  $\phi\Phi Y_{t-s-1}$

### Interpretare (lunar, $s = 12$ )

- $Y_{t-1}$ : luna trecută;  $Y_{t-12}$ : aceeași lună anul trecut;  $Y_{t-13}$ : interacțiunea ambelor

### Parcimonie

- Multiplicativă:** 2 parametri ( $\phi, \Phi$ ); **Aditivă:** ar necesita 3+ parametri

## Notația SARIMA

### Specificație completă

- ▣ **SARIMA**( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_s$ : are 7 parametri de specificat

### Cei 7 parametri

| Parametru | Semnificație                              |
|-----------|---|
| $p, d, q$ | Ordine AR, diferențiere, MA non-sezoniere |
| $P, D, Q$ | Ordine AR, diferențiere, MA sezoniere     |
| $s$       | Perioada sezonieră                        |

### Exemplu

- ▣ **SARIMA**(1, 1, 1)  $\times$  (1, 1, 1) $_{12}$ : date lunare cu AR(1), MA(1), AR sezonier(1), MA sezonier(1)
- ▣ Include atât diferențiere obișnuită ( $d = 1$ ) cât și sezonieră ( $D = 1$ )

## Modele SARIMA comune

Modelul Airline:  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_s$

- ▣ **Ecuatia:**  $(1 - L)(1 - L^s)Y_t = (1 + \theta L)(1 + \Theta L^s)\varepsilon_t$
- ▣ **Origine:** model clasic (Box & Jenkins, 1970)

$\text{SARIMA}(1, 0, 0) \times (1, 0, 0)_s$

- ▣ **Ecuatia:**  $(1 - \phi L)(1 - \Phi L^s)Y_t = \varepsilon_t$
- ▣ **Descriere:** AR sezonier și non-sezonier pur

$\text{SARIMA}(0, 1, 1) \times (0, 1, 0)_s$

- ▣ **Ecuatia:**  $(1 - L)(1 - L^s)Y_t = (1 + \theta L)\varepsilon_t$
- ▣ **Descriere:** random walk + dif. sezonieră + MA(1)

## ACF/PACF pentru modele sezoniere

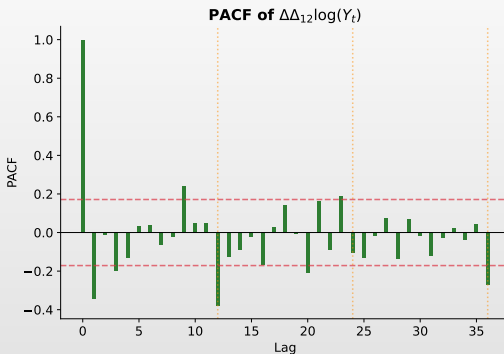
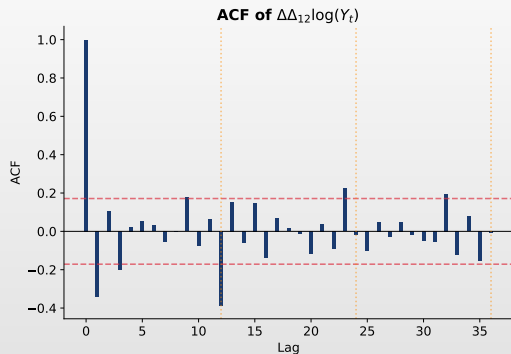
## Principiu

- ▣ **Modelele sezoniere:** prezintă tipare la ambele tipuri de lag-uri
- ▣ **Lag-uri non-sezoniere:**  $1, 2, 3, \dots$
- ▣ **Lag-uri sezoniere:**  $s, 2s, 3s, \dots$

## Tipare ACF/PACF sezoniere

| Model      | ACF                        | PACF                       |
|------------|----------------------------|----------------------------|
| SAR( $P$ ) | Scade la $s, 2s, \dots$    | Se oprește după $P_s$      |
| SMA( $Q$ ) | Se oprește după $Q_s$      | Scade la $s, 2s, \dots$    |
| SARMA      | Scade la lag-uri sezoniere | Scade la lag-uri sezoniere |

## Exemplu: ACF/PACF pentru modelul Airline



## Exemplu: ACF/PACF pentru modelul Airline

ACF:  $\Delta\Delta_{12} \log(Y_t)$

- ▣ Vârf la lag 1  $\leftarrow$  MA(1),  $\theta$
- ▣ Vârf la lag 12  $\leftarrow$  SMA(1),  $\Theta$
- ▣ Restul  $\approx$  zero

PACF: descreștere exponențială

- ▣ Scade la lag-urile 1, 2, 3, ...
- ▣ Scade la lag-urile 12, 24, 36
- ▣  $\Rightarrow$  **MA, nu AR**

- ▣ **Concluzie:** ACF se oprește brusc  $\Rightarrow$  componentă MA; PACF scade  $\Rightarrow$  nu AR
- ▣ **Model identificat:**  $(0, 1, 1) \times (0, 1, 1)_{12}$

## Ghid de identificare a modelului

### Proces pas cu pas

- ▣ **Pas 1:** examinați ACF pentru descreștere lentă la lag-uri sezoniere  $\Rightarrow$  diferențiere sezonieră
- ▣ **Pas 2:** după diferențiere, verificați tiparele ACF/PACF
- ▣ **Pas 3:** comportamentul non-sezonier la lag-urile  $1, 2, \dots, s - 1$
- ▣ **Pas 4:** comportamentul sezonier la lag-urile  $s, 2s, 3s, \dots$

### Sfaturi practice

- ▣ Începeți cu  $d \leq 1$  și  $D \leq 1$
- ▣ De obicei  $P, Q \leq 2$  este suficient
- ▣ Folosiți criterii informaționale (AIC, BIC) pentru selecția finală
- ▣ Algoritmii Auto-SARIMA pot ajuta

## Metode de estimare

### Estimare prin verosimilitate maximă

- ▣ **Abordare standard** pentru SARIMA:
  - ▶ MLE (Maximum Likelihood Estimation) condiționată (condiționată de valorile inițiale)
  - ▶ MLE exactă (prin filtrul Kalman)

### Considerații computaționale

- ▣ Mai mulți parametri decât ARIMA  $\Rightarrow$  mai multe date necesare
- ▣ Parametrii sezonieri estimați din lag-urile  $s, 2s, \dots$
- ▣ Necesită suficiente cicluri sezoniere (cel puțin 3–4 ani de date lunare)



## Verosimilitate exactă: descompunerea erorilor de predicție

### De ce filtrul Kalman?

- ▣ **SARIMA**: are structura unui model în spațiul stărilor
- ▣ **Filtrul Kalman**: calculează recursiv erorile de predicție  $v_t$  și varianțele lor  $f_t$ , fără a condiționa pe valori inițiale

### Log-verosimilitatea exactă (prediction error decomposition)

- ▣ **Formula**:  $\ell(\theta) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \left[ \ln(f_t) + \frac{v_t^2}{f_t} \right]$
- ▣  $v_t$ :  $Y_t - \hat{Y}_{t|t-1}$  (inovația);  $f_t$ :  $\text{Var}(v_t)$  (varianța inovației)

### Avantaje față de MLE condiționată

- ▣ Nu necesită alegerea valorilor inițiale
- ▣ Fiecare termen  $\ln(f_t)$  ponderează diferit observațiile (varianță variabilă la început)
- ▣ Esențial pentru serii scurte; implementat în `statsmodels.tsa.SARIMAX()`

## Staționaritate și invertibilitate

### Condiții de staționaritate

- ▣ **Cerință:** polinoamele AR trebuie să aibă rădăcini în afara cercului unitate
- ▣ **Non-sezonier:**  $\phi(z) = 0 \Rightarrow |z| > 1$
- ▣ **Sezonier:**  $\Phi(z^s) = 0 \Rightarrow |z| > 1$

### Condiții de invertibilitate

- ▣ **Cerință:** polinoamele MA trebuie să aibă rădăcini în afara cercului unitate
- ▣ **Non-sezonier:**  $\theta(z) = 0 \Rightarrow |z| > 1$
- ▣ **Sezonier:**  $\Theta(z^s) = 0 \Rightarrow |z| > 1$

## Validarea modelului

### Analiza reziduurilor

- ▣ **Scop:** verificați că reziduurile sunt zgomot alb
- ▣ **Graficul reziduurilor:** în timp (fără tipare)
- ▣ **ACF:** a reziduurilor (fără vârfuri semnificative)
- ▣ **Ljung-Box:** la lag-uri multiple inclusiv sezoniere
- ▣ **Normalitate:** grafic Q-Q, testul Jarque-Bera

### Important

- ▣ **Verificați ACF** la ambele lag-uri non-sezoniere și sezoniere.
- ▣ **ACF semnificativă** la lag-ul 12 sugerează modelare sezonieră inadecvată

## Criterii de selecție a modelului

### Criterii informaționale

- ▣ **AIC:**  $-2 \ln(L) + 2k$
- ▣ **BIC:**  $-2 \ln(L) + k \ln(n)$
- ▣ **AICc:**  $AIC + \frac{2k(k+1)}{n-k-1}$  (corectat pentru eșantioane mici)
- ▣ **Parametri:**  $k = p + q + P + Q + 1$  (plus 1 pentru varianță)  
unde:  $L$  = maximul funcției de verosimilitate,  $k$  = nr. parametri,  $n$  = dimensiunea eșantionului

### Auto-SARIMA

- ▣ **Python:** `pmdarima.auto_arima()` cu `seasonal=True`
- ▣ **Funcție:** caută automat  $(p, d, q) \times (P, D, Q)_s$  optim

## Algoritmul Hyndman-Khandakar (auto\_arima)

Cum funcționează selecția automată? (Hyndman & Khandakar, 2008)

1.  $d$ : teste KPSS (Kwiatkowski-Phillips-Schmidt-Shin) succesive ( $d = 0, 1, 2$ );  $D$ : testul OCSB sau Canova-Hansen ( $D = 0, 1$ )
2. **Căutare stepwise**: pornește de la modelul inițial, explorează modele vecine
3. **Criteriu**: AICc (corect pentru eșantioane mici)

### Strategia de căutare

- ▣ **Model inițial**: SARIMA(2,  $d$ , 2)(1,  $D$ , 1)<sub>s</sub> sau SARIMA(0,  $d$ , 0)(0,  $D$ , 0)<sub>s</sub>
- ▣ **Variații testate**:  $\pm 1$  pentru fiecare ordin ( $p, q, P, Q$ )
- ▣ **Oprire**: când niciun model vecin nu îmbunătățește AICc
- ▣ **Complexitate**:  $O(20-30)$  modele evaluate (vs.  $O(k^4)$  pentru grid search)

Python: `pm.auto_arima(y, seasonal=True, m=12, stepwise=True, trace=True)`

- ▣ Setăți `stepwise=False` pentru căutare exhaustivă (mai lentă, uneori mai bună)

## Proгноze punctuale

### Calculul prognozei

- ▣ **Metoda:** prognozele SARIMA sunt calculate recursiv
- ▣ **Erori viitoare:** înlocuiți  $\varepsilon_{T+h}$  cu 0
- ▣ **Valori viitoare:** înlocuiți  $Y_{T+h}$  cu prognozele  $\hat{Y}_{T+h|T}$
- ▣ **Valori trecute:** folosiți  $Y_T, Y_{T-1}, \dots$  cunoscute

### Tiparul sezonier în prognoze

- ▣ **Proprietate:** prognozele SARIMA captează în mod natural sezonabilitatea
- ▣ **Pe termen scurt:** influențate de valorile recente
- ▣ **Pe termen lung:** revin la tiparul sezonier

## Intervale de prognoză

### Cuantificarea incertitudinii

- ▣ **Interval**  $(1 - \alpha)\%$ :  $\hat{Y}_{T+h|T} \pm z_{\alpha/2} \sqrt{\text{Var}(e_{T+h})}$
- ▣ **Varianța**: calculată din reprezentarea  $\text{MA}(\infty)$

### Proprietăți

- ▣ Intervalele se largesc cu orizontul de prognoză
- ▣ Pentru serii  $I(1, 1)_s$ : intervalele cresc nelimitat
- ▣ Tiparul sezonier vizibil în prognozele punctuale
- ▣ Incertitudinea captează atât variația de trend cât și cea sezonieră

## Proгноze pe orizont lung

### Comportamentul când $h \rightarrow \infty$

- ▣ Prognozele punctuale converg la tiparul sezonier determinist
- ▣ Dacă există derivă: trend linear + tipar sezonier
- ▣ Intervalele de prognoză continuă să se lărgescă

### Implicație practică

- ▣ Pe termen scurt: SARIMA captează atât dinamica pe termen scurt cât și sezonul
- ▣ Pe termen mediu: Prognoze sezoniere bune, incertitudine crescătoare
- ▣ Pe termen lung: Reflectă în principal tiparul sezonier, intervale largi



## Benchmark-ul Seasonal Naive

### Definiție: Proгноза Seasonal Naive

- ▣ **Formula:**  $\hat{Y}_{T+h} = Y_{T+h-s}$  (ultimul sezon observat)
- ▣ **Exemplu lunar:** Proгноза pentru Martie 2025 = valoarea din Martie 2024
- ▣ **Interpretare:** “Cel mai simplu model care respectă sezonalitatea”

### De ce este esențial?

- ▣ Orice model SARIMA **trebuie** să depășească benchmark-ul seasonal naive
- ▣ Dacă nu îl depășește  $\Rightarrow$  complexitatea modelului nu este justificată
- ▣ Surprinzător de eficient pentru multe serii cu sezonalitate stabilă

### Regulă de aur

- ▣ Raportați **întotdeauna** performanța SARIMA relativ la seasonal naive
- ▣ Acesta este **primul lucru** pe care un recenzor sau un manager îl verifică

## Metrica MASE: Evaluare corectă pentru serii sezoniere

MASE — Mean absolute scaled error (Hyndman & Koehler, 2006)

- ▣ **Formula:** 
$$\text{MASE} = \frac{\frac{1}{H} \sum_{h=1}^H |e_{T+h}|}{\frac{1}{T-s} \sum_{t=s+1}^T |Y_t - Y_{t-s}|}$$
- ▣ **Numărător:** eroarea medie absolută a modelului
- ▣ **Numitor:** eroarea medie absolută a seasonal naive (pe datele de antrenament)

### Interpretare

- ▣  $\text{MASE} < 1$ : Modelul este **mai bun** decât seasonal naive
- ▣  $\text{MASE} = 1$ : Modelul este **echivalent** cu seasonal naive
- ▣  $\text{MASE} > 1$ : Modelul este **mai rău** — renunțați la el.

### De ce MASE și nu MAPE?

- ▣ MAPE (Mean Absolute Percentage Error): nedefinit pentru  $Y_t = 0$ ; asimetric; dependent de scară
- ▣ MASE: funcționează cu orice date; simetric; comparabil între serii diferite

## Evaluarea prognozei: Rolling Forecast Origin

### Cross-validare pentru serii de timp sezoniere

- ▣ **Principiu:** re-estimare model  $\rightarrow$  prognoză  $h$  pași  $\rightarrow$  avansare 1 pas  $\rightarrow$  repetare
- ▣ **Fereastră fixă:** antrenament pe ultimii  $w$  observații (dimensiune constantă)
- ▣ **Fereastră expandabilă:** antrenament de la început până la  $T + i$  (crește)

### Procedura pas cu pas

1. Antrenează SARIMA pe  $Y_1, \dots, Y_T$ ; prognozează  $\hat{Y}_{T+1}, \dots, \hat{Y}_{T+h}$
2. Antrenează SARIMA pe  $Y_1, \dots, Y_{T+1}$ ; prognozează  $\hat{Y}_{T+2}, \dots, \hat{Y}_{T+h+1}$
3. ... repetă de  $N$  ori; calculează RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), MASE pe toate cele  $N$  prognoze

### Important

- ▣ Minim  $N \geq 2s$  origini (2 cicluri sezoniere complete) pentru rezultate fiabile
- ▣ Niciodată nu “privim în viitor” — datele de test sunt strict ulterioare antrenamentului

## SARIMA vs Holt-Winters/ETS: Când folosim ce?

### Comparație — ETS (Error-Trend-Seasonality)

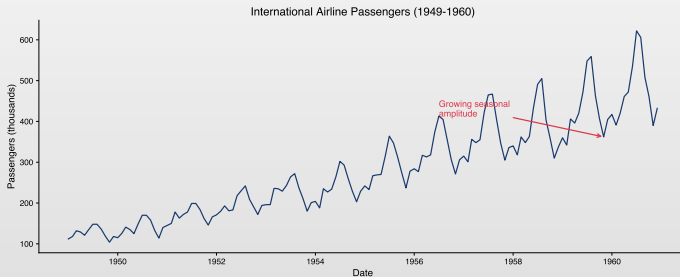
| Criteriu      | SARIMA                             | ETS / Holt-Winters                          |
|---------------|------------------------------------|---|
| Abordare      | Box-Jenkins (ACF/PACF, cf. Cap. 3) | Netezire exponențială                       |
| Sezonalitate  | Stochastică (diferențiere)         | Aditivă sau multiplicativă                  |
| Interpretare  | Coef. AR/MA                        | Ponderi de netezire $\alpha, \beta, \gamma$ |
| Flexibilitate | Foarte flexibil (7 param.)         | Mai puțin flexibil                          |
| Automatizare  | auto_arima                         | ets() / ExponentialSmoothing                |

### Ghid practic de selecție

- ▣ **SARIMA preferat:** serii cu autocorelație complexă, sezonalitate stochastică
- ▣ **ETS preferat:** serii scurte, sezonalitate stabilă, prognoze rapide
- ▣ **Cel mai bun:** comparați ambele pe date out-of-sample și alegeți câștigătorul

## Studiu de caz: Date despre pasagerii aerieni

- ▣ Setul de date clasic Box-Jenkins: număr de pasageri lunari (1949-1960)
- ▣ Trend ascendent cu amplitudine sezonieră crescătoare (sezonalitate multiplicativă)
- ▣ Prima decizie: aplicăm transformarea logaritmică pentru a stabiliza varianța



## Strategia de împărțire a datelor

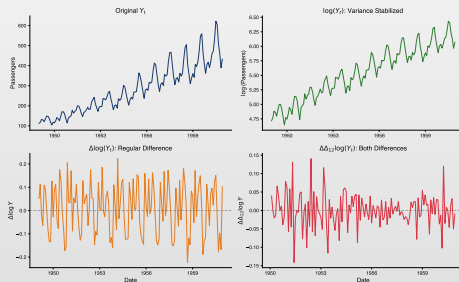
### Time Series Train/Validation/Test Split



- ▣ **Set de antrenare (70%):** estimare coeficienți SARIMA ( $\phi, \theta, \Phi, \Theta$ )
- ▣ **Set de validare (15%):** comparare modele candidate, alegere cea mai mică eroare
- ▣ **Set de test (15%):** evaluare finală out-of-sample, niciodată folosit în dezvoltare

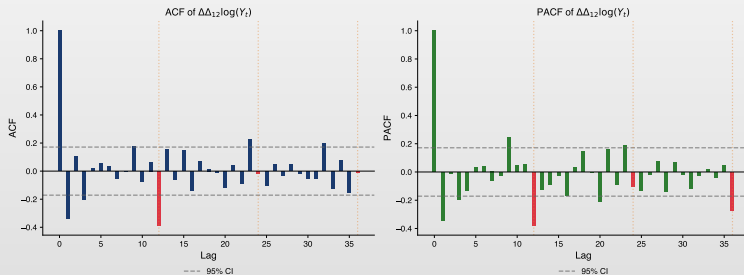
## Pasul 1: Transformări

- Transformarea log stabilizează varianța (multiplicativ  $\Rightarrow$  aditiv)
- Prima diferență elimină trendul; diferența sezonieră elimină sezonalitatea
- Seria dublu diferențiată pare staționară



## Pasul 2: Analiza ACF/PACF

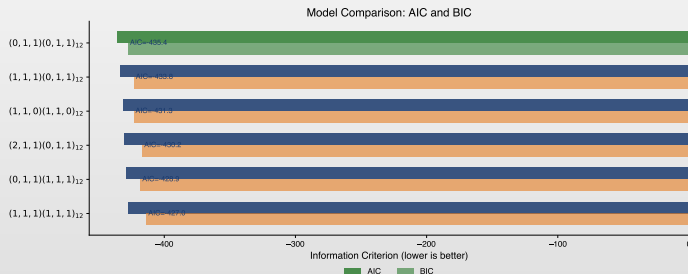
- ACF: Vârf semnificativ la lag 1 și lag 12  $\Rightarrow$  MA(1), SMA(1)
- PACF: Tipar de descreștere exponențială confirmă structura MA
- Sugerează SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> (modelul airline)





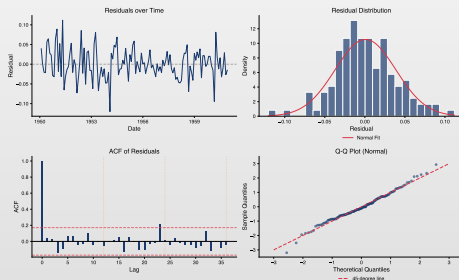
## Pasul 3: Compararea modelelor

- Comparăm modelele SARIMA candidate folosind criteriul AIC
- SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> oferă cea mai bună ajustare (AIC minim)
- Acesta este faimosul “model airline” identificat de Box & Jenkins



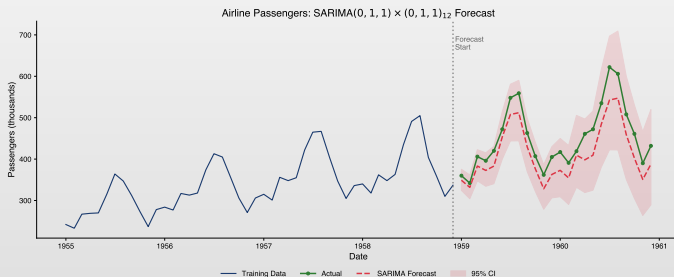
## Pasul 4: Diagnosticul reziduurilor

- Reziduurile par aleatorii fără autocorelație remanentă
- Graficul Q-Q arată normalitate aproximativă
- Modelul captează adecvat atât trendul cât și structura sezonieră



## Pasul 5: Prognoza

- Prognoză pe 24 de luni cu interval de încredere de 95%
- Modelul captează tiparul sezonier și trendul ascendent
- Intervalele de predicție se lărgesc corespunzător cu orizontul prognozei



## Capcane practice în modelarea SARIMA

### 1. Supra-diferențiere (Overdifferencing)

- ▣ **Simptom:** ACF la lag 1  $\approx -0.5$  (regulară) sau la lag  $s \approx -0.5$  (sezonieră)
- ▣ **Cauză:** aplicarea  $(1 - L)$  sau  $(1 - L^s)$  de prea multe ori
- ▣ **Soluție:** reduceți  $d$  sau  $D$  cu 1 și re-examinați ACF/PACF

### 2. Număr insuficient de date

- ▣ **Minimum:** 3–4 cicluri sezoniere complete (36–48 obs. lunare)
- ▣ **Recomandat:** 5+ cicluri sezoniere pentru estimare robustă
- ▣ Parametrii sezonieri  $\Phi, \Theta$  se estimează din lag-urile  $s, 2s, 3s, \dots$

### 3. Alte capcane frecvente

- ▣ **Anularea rădăcinilor:**  $\phi \approx \theta$  sugerează supraparametrizare
- ▣ **Parametri la limita invertibilității:**  $|\theta| \approx 1$  sau  $|\Theta| \approx 1$  indică probleme
- ▣ **Uitarea transformării inverse:** prognozele pe scara log trebuie retransformate.

## X-13ARIMA-SEATS: Ajustarea sezonieră oficială

### Ce este ajustarea sezonieră?

- ▣ **Scop:** elimină componenta sezonieră pentru a dezvălui tendința reală
- ▣ **Utilizatori:** Eurostat, US Census Bureau, BNR (Banca Națională a României)
- ▣ **INS** (Institutul Național de Statistică), BCE (Banca Centrală Europeană)
- ▣ **Exemplu:** "PIB-ul a crescut cu 0.3% față de trimestrul anterior" (date ajustate sezonier)

### X-13ARIMA-SEATS (US Census Bureau)

- ▣ **Etapa 1:** Identifică și estimează un model regARIMA (SARIMA + efecte calendaristice)
- ▣ **Etapa 2:** Extrage componenta sezonieră prin filtre SEATS sau X-11
- ▣ **Etapa 3:**  $Y_t^{ajustat} = Y_t - \hat{S}_t$  (aditiv) sau  $Y_t^{ajustat} = Y_t / \hat{S}_t$  (multiplicativ)

### De ce contează pentru economiști?

- ▣ Datele macroeconomice publicate sunt aproape întotdeauna ajustate sezonier
- ▣ Interpretarea greșită a datelor neajustate poate duce la concluzii eronate

## Exercițiu AI: Gândire critică

Prompt de testat în ChatGPT / Claude / Copilot

"Am seria AirPassengers din statsmodels (date lunare, pasageri aerieni internaționali, 1949–1960, 144 obs.). Identifică sezonaltatea, aplică transformarea Box-Cox dacă e nevoie, estimează un model SARIMA și prognozează 12 luni. Vreau cod Python complet cu grafice."

### Exercițiu:

1. Rulați prompt-ul într-un LLM la alegere și analizați critic răspunsul.
2. Verifică sezonaltatea cu ACF la lag-urile  $s, 2s, 3s$ ? Folosește descompunere STL?
3. Aplică Box-Cox *înainte* de diferențiere? Justifică alegerea lui  $\lambda$ ?
4. Cum alege ordinele  $(p, d, q) \times (P, D, Q)_s$ ? Doar `auto_arima` sau și ACF/PACF?
5. Evaluează cu MASE relativ la seasonal naive? Folosește rolling forecast?

**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

- Sezonalitatea în seriile de timp
  - ▶ Tipare repetitive la intervale regulate; aditivă vs multiplicativă
- Diferențierea sezonieră și transformarea Box-Cox
  - ▶  $(1 - L^s)$  elimină sezonalitatea stochastică; Box-Cox stabilizează varianța
- Modele SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ )<sub>s</sub>
  - ▶ Extind ARIMA cu componente sezoniere; selecție automată via `auto_arima`
- Prognoze și evaluare
  - ▶ Benchmark: MASE relativ la seasonal naive; rolling forecast out-of-sample

### Observație

- **Principiul parcimoniei:** Modelul Airline  $(0, 1, 1) \times (0, 1, 1)_{12}$  are doar 2 parametri
- Este remarcabil de eficient pentru multe serii economice sezoniere

## Ce urmează?

### Capitolul 5: Modelarea volatilității — GARCH

- ▣ **Volatilitate:** variația condiționată a randamentelor financiare
- ▣ **ARCH/GARCH:** modele pentru varianța condiționată
- ▣ **Extensii asimetrice:** GJR-GARCH, EGARCH (efectul de levier)
- ▣ **VaR:** Value-at-Risk bazat pe modele GARCH
- ▣ **Studiu de caz:** Volatilitatea randamentelor S&P 500

Întrebări?



## Întrebarea 1

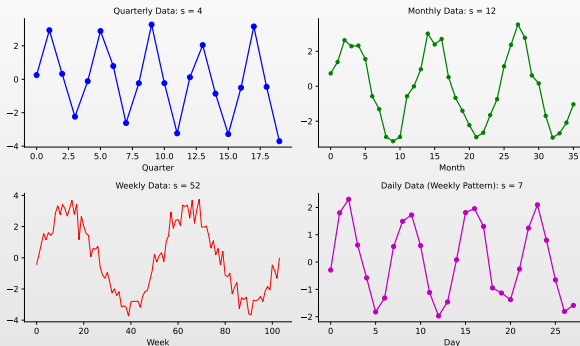
### Întrebare

□ Pentru date lunare cu sezonalitate anuală, care este perioada sezonieră  $s$ ?

### Variante de răspuns

(A)  $s = 4$       (B)  $s = 7$       (C)  $s = 12$       (D)  $s = 52$

## Întrebarea 1: Răspuns



Răspuns corect: (C)  $s = 12$  (12 luni pe an)

▣ **Perioade comune:** Trimestrial=4, Lunar=12, Săptămânal=52, Zilnic=7, Orar=24

## Întrebarea 2

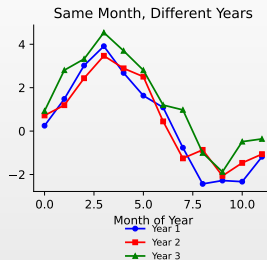
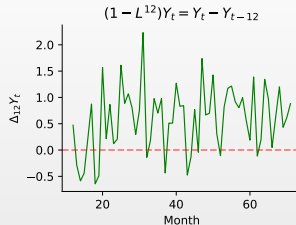
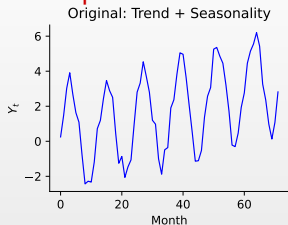
### Întrebare

□ Ce face operatorul de diferență sezonieră  $(1 - L^{12})$  unei serii lunare?

### Variante de răspuns

- (A) Calculează  $Y_t - Y_{t-1}$  (schimbarea luna-la-luna)
- (B) Calculează  $Y_t - Y_{t-12}$  (schimbarea an-la-an)
- (C) Calculează media mobilă pe 12 luni
- (D) Elimină doar componenta de trend

## Întrebarea 2: Răspuns



Răspuns corect: (B) Schimbarea an-la-an

- **Formula:**  $(1 - L^{12})Y_t = Y_t - Y_{t-12}$
- **Efect:** elimină tiparul sezonier prin compararea acelorași luni

## Întrebarea 3

### Întrebare

□ În notația  $SARIMA(1, 1, 1) \times (1, 1, 1)_{12}$ , ce reprezintă partea  $(1, 1, 1)_{12}$ ?

### Variante de răspuns

- (A) AR(1), o diferențiere, MA(1) pentru componenta non-sezonieră
- (B) AR sezonier(1), o diferențiere sezonieră, MA sezonier(1)
- (C) 12 termeni AR, 12 diferențe, 12 termeni MA
- (D) Modelul are 12 parametri în total

### Întrebarea 3: Răspuns

Răspuns corect: (B)

- ▣ **Răspuns:** AR sezonier(1), o diferențiere sezonieră, MA sezonier(1)

#### Descompunerea notației SARIMA

- ▣  $(p, d, q)$ : Non-sezonier  $\Rightarrow$  AR( $p$ ),  $d$  diferențe, MA( $q$ )
- ▣  $(P, D, Q)_s$ : Sezonier  $\Rightarrow$  SAR( $P$ ),  $D$  dif. sezoniere, SMA( $Q$ )
- ▣ **Non-sezonier** (1, 1, 1): AR(1), o diferență obișnuită, MA(1)
- ▣ **Sezonier** (1, 1, 1)<sub>12</sub>: SAR(1) la lag-ul 12, un  $\Delta_{12}$ , SMA(1) la lag-ul 12

## Întrebarea 4

### Întrebare

- ▣ “Modelul Airline” este  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$ . Câți parametri trebuie estimați (excluzând varianța)?

### Variante de răspuns

- (A) 1      (B) 2      (C) 4      (D) 12

## Întrebarea 4: Răspuns

Răspuns corect: (B)

- **Răspuns:** 2 parametri

### Structura modelului

- **Model:**  $(1 - L)(1 - L^{12})Y_t = (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t$
- $\theta_1$ : coeficient MA non-sezonier;  $\Theta_1$ : coeficient MA sezonier
- **Total:** 2 parametri (plus  $\sigma^2$ )

### De ce “modelul Airline”?

- **Origine:** Box & Jenkins (1970) l-au folosit pentru pasagerii aerieni internaționali
- **Impact:** remarcabil de eficient pentru multe serii economice sezoniere.



## Întrebarea 5

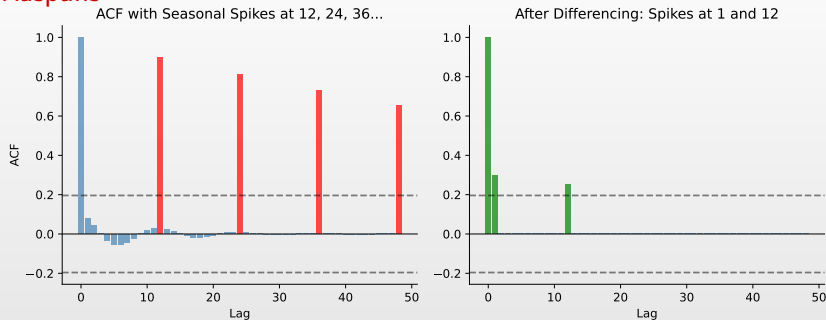
### Întrebare

- ☐ Observați vârfuri ACF semnificative la lag-urile 12, 24 și 36 într-o serie lunară. Ce sugerează aceasta?

### Variante de răspuns

- (A) Seria are o rădăcină unitară
- (B) Seria are sezonalitate anuală care necesită diferențiere sezonieră
- (C) Seria urmează un proces AR(36)
- (D) Seria este deja staționară

## Întrebarea 5: Răspuns



Răspuns corect: (B) Necesită diferențiere sezonieră

- **Diagnostic:** vârfuri ACF la 12, 24, 36 = sezonabilitate stohastică
- **Soluție:** aplicați  $(1 - L^{12})$  pentru a o elimina

## Întrebarea 6

### Întrebare

- După aplicarea  $(1 - L)(1 - L^{12})$  unei serii lunare, ACF prezintă un vârf semnificativ doar la lag-ul 1 și lag-ul 12. Ce model SARIMA este sugerat?

### Variante de răspuns

- (A)  $\text{SARIMA}(1, 1, 0) \times (1, 1, 0)_{12}$
- (B)  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$
- (C)  $\text{SARIMA}(1, 1, 1) \times (1, 1, 1)_{12}$
- (D)  $\text{SARIMA}(0, 1, 0) \times (0, 1, 0)_{12}$

## Întrebarea 6: Răspuns

Răspuns corect: (B)

- ▣ **Model:**  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$  (Modelul Airline)

### Reguli de identificare ACF/PACF

- ▣ **Regulă:** pentru procese MA, ACF se oprește brusc după lag-ul  $q$
- ▣ **Vârf ACF la lag-ul 1:** MA(1) pentru partea non-sezonieră
- ▣ **Vârf ACF la lag-ul 12:** SMA(1) pentru partea sezonieră
- ▣ **Combinat:**  $\text{MA}(1) \times \text{SMA}(1) = (0, d, 1) \times (0, D, 1)_{12}$
- ▣ **Cu  $d = 1, D = 1$ :**  $(0, 1, 1) \times (0, 1, 1)_{12}$

## Bibliografie I

### Modele sezoniere – lucrări fundamentale

- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*, 5th ed., Wiley.
- Hylleberg, S., Engle, R.F., Granger, C.W.J., & Yoo, B.S. (1990). Seasonal Integration and Cointegration, *Journal of Econometrics*, 44(1-2), 215–238.
- Canova, F., & Hansen, B.E. (1995). Are Seasonal Patterns Constant Over Time?, *Journal of Business & Economic Statistics*, 13(3), 237–252.

### Descompunere sezonieră și diagnoză

- Cleveland, R.B., Cleveland, W.S., McRae, J.E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess, *Journal of Official Statistics*, 6(1), 3–33.
- Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*, 3rd ed., OTexts.

## Bibliografie II

### Manuale și referințe suplimentare

- ▣ Shumway, R.H., & Stoffer, D.S. (2017). *Time Series Analysis and Its Applications*, 4th ed., Springer.
- ▣ Brockwell, P.J., & Davis, R.A. (2016). *Introduction to Time Series and Forecasting*, 3rd ed., Springer.
- ▣ Hyndman, R.J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R, *Journal of Statistical Software*, 27(3), 1–22.

### Resurse online și cod

- ▣ **Quantlet:** <https://quantlet.com> – Platformă de cod pentru metode cantitative
- ▣ **Quantinar:** <https://quantinar.com> – Platformă de învățare pentru metode cantitative
- ▣ **GitHub TSA:** [https://github.com/QuantLet/TSA/tree/main/TSA\\_ch4](https://github.com/QuantLet/TSA/tree/main/TSA_ch4) – Cod Python pentru acest capitol

# Vă Mulțumim!

Întrebări?

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



Quantlet



Quantinar