



# Analiza și Prognoza Seriilor de Timp

## Seminar 4: Modele SARIMA



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

## Cuprins Seminar

Test de Recapitulare

Probleme Practice

Exemple Rezolvate

Analiză pe Date Reale

Subiecte de Discuție

Exercițiu cu asistență AI

Exerciții pentru Studiu Individual

## Test 1: Diferențierea Sezonieră

### Întrebare

Pentru date lunare cu sezonalitate anuală, ce face operatorul  $(1 - L^{12})$ ?

- A) Ia 12 diferențe consecutive
- B) Calculează  $Y_t - Y_{t-12}$
- C) Face media pe 12 luni
- D) Elimină primele 12 observații

## Test 1: Diferențierea Sezonieră

### Întrebare

Pentru date lunare cu sezonalitate anuală, ce face operatorul  $(1 - L^{12})$ ?

- A) Ia 12 diferențe consecutive
- B) Calculează  $Y_t - Y_{t-12}$
- C) Face media pe 12 luni
- D) Elimină primele 12 observații

Răspuns: B – Calculează  $Y_t - Y_{t-12}$

Operatorul de diferență sezonieră:

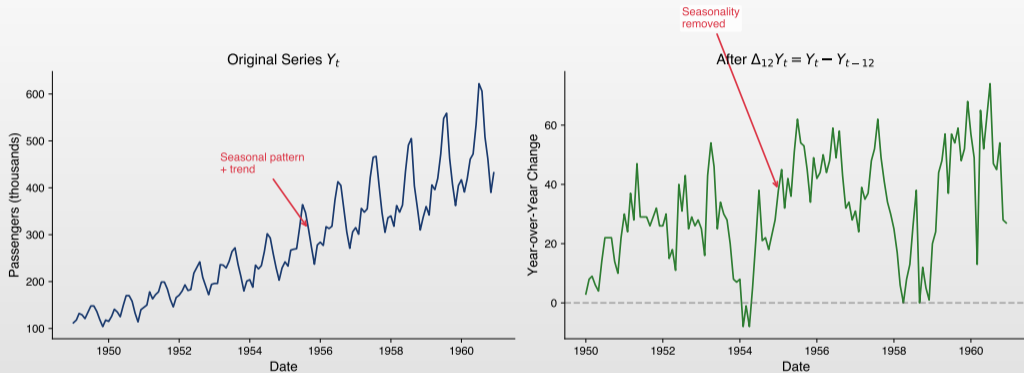
$$(1 - L^{12})Y_t = Y_t - L^{12}Y_t = Y_t - Y_{t-12}$$

**Exemplu** (vânzări ianuarie):  $Y_{\text{Ian2025}} - Y_{\text{Ian2024}}$

**Efect:** Elimină tiparul sezonier anual stabil

**Notă:**  $(1 - L^s)$  pentru orice perioadă sezonieră  $s$  (trimestrial:  $s = 4$ , săptămânal:  $s = 52$ )

## Vizual: Diferența Sezonieră



Diferențierea sezonieră elimină modelele anuale comparând aceleași perioade între ani. [TSA\\_ch4\\_seasonal\\_differencing](#)

## Test 2: Notăția SARIMA

### Întrebare

Ce reprezintă  $SARIMA(1, 1, 1) \times (1, 1, 1)_{12}$ ?

- A) 12 modele ARIMA diferite
- B) ARIMA cu 12 termeni AR și 12 termeni MA
- C) ARIMA(1,1,1) cu ARIMA(1,1,1) sezonier la perioada 12
- D) Un model care necesită 12 ani de date

## Test 2: Răspuns

Răspuns: C – ARIMA(1,1,1) cu ARIMA(1,1,1) sezonier la perioada 12

**SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_s$  Notation**

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

**Regular (Non-Seasonal)**

$p$	= AR order	(Number of AR lags)
$d$	= Differencing	(Regular differences)
$q$	= MA order	(Number of MA lags)

**Seasonal**

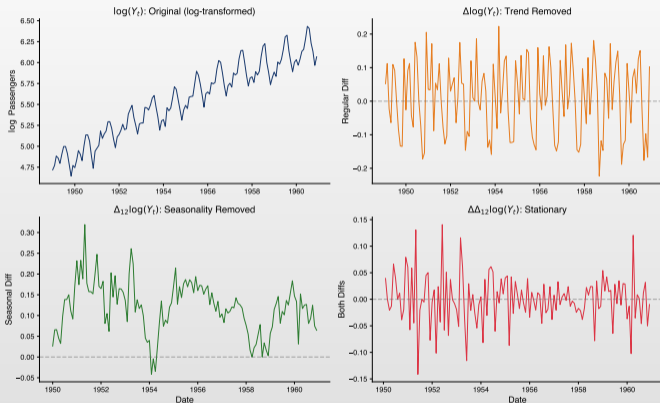
$P$	= Seasonal AR	(SAR lags at $s, 2s, \dots$ )
$D$	= Seasonal Diff	$(1-L^s)^D$
$Q$	= Seasonal MA	(SMA lags at $s, 2s, \dots$ )
$s$	= Period	(Seasonal period)

**Example: SARIMA(1, 1, 1)  $\times$  (0, 1, 1) $_{12}$**

Monthly data with: AR(1), MA(1), one regular diff,  
one seasonal diff at lag 12, seasonal MA(1)

$$(1 - \phi_1 L)(1 - \Phi_1 L^{12})(1 - L)(1 - L^{12})Y_t = (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t$$

## Vizual: Structura Modelului SARIMA



SARIMA combină componentele ARIMA obișnuite cu componentele sezoniere la lag-ul  $s$ .

 TSA\_ch4\_sarima\_estimation

## Test 3: Modelul Airline

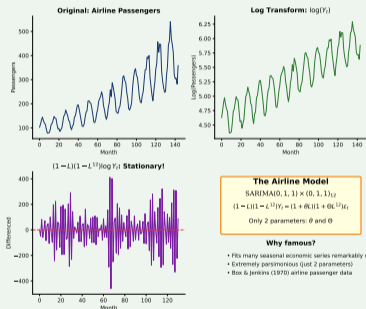
### Întrebare

“Modelul airline” se referă la  $SARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ . Câți parametri are (excluzând varianța)?

- A) 2 parametri
- B) 4 parametri
- C) 6 parametri
- D) 12 parametri

## Test 3: Răspuns

Răspuns: A – 2 parametri ( $\theta_1$  și  $\Theta_1$ )



**Modelul airline:**  $(1 - L)(1 - L^{12})Y_t = (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t$

Se potrivește remarcabil de bine pe multe serii economice sezoniere (Box & Jenkins, 1970)

## Test 4: ACF-ul Datelor Sezoniere

### Întrebare

Pentru date lunare cu sezonalitate puternică, unde vă așteptați să vedeți vârfuri ACF semnificative?

- A) Doar la lag 1
- B) Doar la lag 12
- C) La lag-urile 12, 24, 36, ...
- D) Distribuite aleatoriu

## Test 4: ACF-ul Datelor Sezoniere

### Întrebare

Pentru date lunare cu sezonalitate puternică, unde vă așteptați să vedeți vârfuri ACF semnificative?

- A) Doar la lag 1
- B) Doar la lag 12
- C) La lag-urile 12, 24, 36, ...
- D) Distribuite aleatoriu

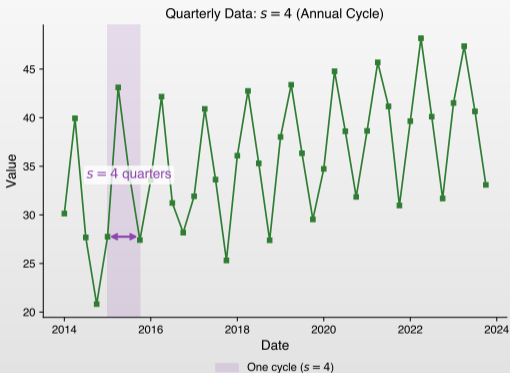
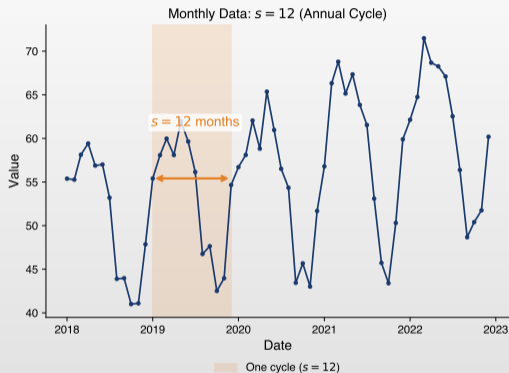
Răspuns: C – La lag-urile 12, 24, 36, ...

**Intuiție:** Ianuarie 2024 este similar cu ianuarie 2023, 2022, etc.

**Model ACF:** Vârfuri la lag-urile  $s, 2s, 3s, \dots$  ( $\rho_{12}, \rho_{24}, \rho_{36} \neq 0$ )

**Diagnostic:** Descreștere lentă la lag-urile sezoniere  $\Rightarrow D = 1$ ; Întrerupere după lag  $s \Rightarrow Q = 1$

## Vizual: Modele de Sezonalitate



Modelele sezoniere se repetă la intervale regulate (lunar, trimestrial, etc.) și pot fi aditive sau multiplicative.

[TSA\\_ch4\\_seasonal\\_patterns](#)



## Test 5: Structura Multiplicativă

### Întrebare

În SARIMA, ce înseamnă “structură multiplicativă”?

- A) Amplitudinea sezonieră crește proporțional
- B) Polinoamele obișnuite și sezoniere sunt înmulțite
- C) Înmulțim datele cu factori sezonieri
- D) Modelul este estimat folosind înmulțirea

## Test 5: Structura Multiplicativă

### Întrebare

În SARIMA, ce înseamnă “structură multiplicativă”?

- A) Amplitudinea sezonieră crește proporțional
- B) Polinoamele obișnuite și sezoniere sunt înmulțite
- C) Înmulțim datele cu factori sezonieri
- D) Modelul este estimat folosind înmulțirea

Răspuns: B – Polinoamele obișnuite și sezoniere sunt înmulțite

**SARIMA multiplicativ:**  $\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D Y_t = \theta(L)\Theta(L^s)\varepsilon_t$

**Exemplu:**  $(1 - \phi_1 L)(1 - \Phi_1 L^{12}) = 1 - \phi_1 L - \Phi_1 L^{12} + \phi_1 \Phi_1 L^{13}$

**Termenul încrucișat**  $\phi_1 \Phi_1 L^{13}$ : Captează interacțiunea între dinamica pe termen scurt și lung

## Test 6: Diferențierea Sezonieră vs Obișnuită

### Întrebare

Când ați aplica atât diferențierea obișnuită ( $d = 1$ ) cât și sezonieră ( $D = 1$ )?

- A) Când datele au doar un trend
- B) Când datele au doar sezonabilitate
- C) Când datele au atât trend cât și nestaționaritate sezonieră
- D) Niciodată – se anulează reciproc

## Test 6: Diferențierea Sezonieră vs Obișnuită

### Întrebare

Când ați aplica atât diferențierea obișnuită ( $d = 1$ ) cât și sezonieră ( $D = 1$ )?

- A) Când datele au doar un trend
- B) Când datele au doar sezonabilitate
- C) Când datele au atât trend cât și nestăționaritate sezonieră
- D) Niciodată – se anulează reciproc

Răspuns: C – Atât trend cât și nestăționaritate sezonieră

**Combinat:**  $W_t = (1 - L)(1 - L^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$

**Când este necesar:** ACF cu descreștere lentă la lag-urile 1,2,3...  $\Rightarrow d = 1$ ; la lag-urile 12,24,36...  $\Rightarrow D = 1$

**Exemple:** Pasageri aerieni, vânzări retail, cerere de energie

## Test 7: Detectarea Sezonalității din ACF

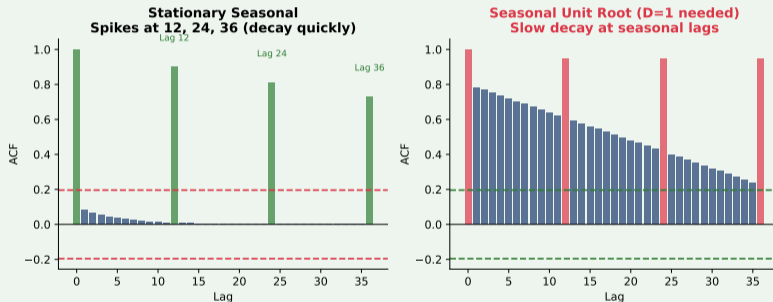
### Întrebare

ACF-ul unei serii de timp lunare arată descreștere lentă la lag-urile 12, 24 și 36. Ce sugerează aceasta?

- A) Seria este staționară
- B) Seria necesită doar diferențierea obișnuită
- C) Seria are o rădăcină unitară sezonieră necesitând  $D = 1$
- D) Seria este zgomot alb

## Test 7: Răspuns

Răspuns: C – Rădăcină unitară sezonieră necesitând  $D = 1$



**Stânga:** Sezonieră staționară (descreștere rapidă la lag-urile sezoniere)

**Dreapta:** Rădăcină unitară sezonieră (descreștere lentă  $\Rightarrow$  necesită  $D = 1$ )

## Test 8: Sezonalitate Multiplicativă vs Aditivă

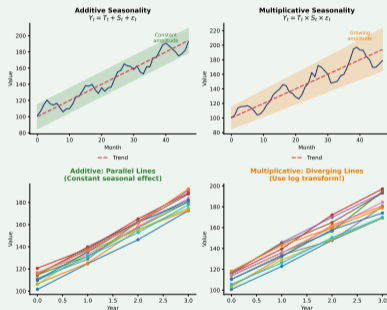
### Întrebare

Dacă amplitudinea sezonieră a unei serii de timp crește proporțional cu nivelul, aceasta indică:

- A) Sezonalitate aditivă – folosiți  $(1 - L^s)$
- B) Sezonalitate multiplicativă – folosiți transformarea log
- C) Fără sezonality prezentă
- D) Nevoie doar de diferențiere obișnuită

## Test 8: Răspuns

Răspuns: B – Sezonalitate multiplicativă, folosiți transformarea log



**Multiplicativă:** Amplitudinea sezonieră crește cu nivelul (linii divergente)

**Soluție:** Aplicați transformarea log înainte de a ajusta SARIMA

## Test 9: Graficul Subseriilor Sezoniere

### Întrebare

Într-un grafic de subserii sezoniere, ce indică sezonalitatea multiplicativă?

- A) Liniile pentru fiecare lună sunt paralele
- B) Liniile pentru fiecare lună divergează (dispersia crește în timp)
- C) Toate lunile au aceeași medie
- D) Liniile sunt orizontale

## Test 9: Graficul Subseriilor Sezoniere

### Întrebare

Într-un grafic de subserii sezoniere, ce indică sezonalitatea multiplicativă?

- A) Liniile pentru fiecare lună sunt paralele
- B) Liniile pentru fiecare lună divergează (dispersia crește în timp)
- C) Toate lunile au aceeași medie
- D) Liniile sunt orizontale

Răspuns: B – Liniile divergează (dispersia crește în timp)

**Graficul subseriilor:** Grupează datele pe luni, reprezintă valorile fiecărei luni de-a lungul anilor

**Paralele**  $\Rightarrow$  Aditivă; **Divergente**  $\Rightarrow$  Multiplicativă; **Orizontale**  $\Rightarrow$  Fără trend

**Acțiune:** Dacă multiplicativă, aplicați log înainte de a ajusta SARIMA

## Test 10: Invertibilitatea în SARIMA

### Întrebare

Pentru ca  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$  să fie invertibil, care condiție trebuie îndeplinită?

- A)  $|\theta_1| < 1$  doar
- B)  $|\Theta_1| < 1$  doar
- C) Atât  $|\theta_1| < 1$  cât și  $|\Theta_1| < 1$
- D) Nu există condiție de invertibilitate pentru modelele MA

## Test 10: Invertibilitatea în SARIMA

## Întrebare

Pentru ca  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$  să fie invertibil, care condiție trebuie îndeplinită?

- A)  $|\theta_1| < 1$  doar
- B)  $|\Theta_1| < 1$  doar
- C) Atât  $|\theta_1| < 1$  cât și  $|\Theta_1| < 1$
- D) Nu există condiție de invertibilitate pentru modelele MA

Răspuns: C – Atât  $|\theta_1| < 1$  cât și  $|\Theta_1| < 1$

**Invertibilitate:** Toate rădăcinile MA în afara cercului unitate

**MA multiplicativ:**  $(1 + \theta_1 L)(1 + \Theta_1 L^{12})$

**Rădăcini:** Obişnuită  $|z| = |-1/\theta_1| > 1 \Leftrightarrow |\theta_1| < 1$ ; Sezonieră  $|\Theta_1| < 1$

**Ambele** condiții necesare pentru invertibilitate generală.

## Test 11: Testul HEGY

### Întrebare

Testul HEGY este folosit pentru:

- A) Estimarea parametrilor SARIMA
- B) Testarea rădăcinilor unitare la diferite frecvențe (trend și sezoniere)
- C) Verificarea normalității reziduurilor
- D) Compararea modelelor SARIMA folosind criterii informaționale

## Test 11: Testul HEGY

### Întrebare

Testul HEGY este folosit pentru:

- A) Estimarea parametrilor SARIMA
- B) Testarea rădăcinilor unitare la diferite frecvențe (trend și sezoniere)
- C) Verificarea normalității reziduurilor
- D) Compararea modelelor SARIMA folosind criterii informaționale

Răspuns: B – Testarea rădăcinilor unitare la diferite frecvențe

**Testul HEGY** (Hylleberg-Engle-Granger-Yoo, 1990):

Testează la: Frecvența zero ( $\omega = 0$ )  $\Rightarrow d = 1$ ; Nyquist ( $\omega = \pi$ ); Sezonieră  $\Rightarrow D = 1$

**Decizie:** Respingeți toate  $\Rightarrow$  variabile dummy sezoniere; Nu respingeți sezoniera  $\Rightarrow$  diferențiere sezonieră

## Test 12: Identificarea MA Sezonier

### Întrebare

După aplicarea  $(1 - L)(1 - L^{12})$ , ACF arată un singur vârf semnificativ doar la lag 12 (fără vârf la lag 1). PACF descrește la lag-urile sezoniere. Aceasta sugerează:

- A) SARIMA(0, 1, 0)  $\times$  (0, 1, 1)<sub>12</sub>
- B) SARIMA(0, 1, 1)  $\times$  (0, 1, 0)<sub>12</sub>
- C) SARIMA(1, 1, 0)  $\times$  (1, 1, 0)<sub>12</sub>
- D) SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub>

## Test 12: Identificarea MA Sezonier

### Întrebare

După aplicarea  $(1 - L)(1 - L^{12})$ , ACF arată un singur vârf semnificativ doar la lag 12 (fără vârf la lag 1). PACF descrește la lag-urile sezoniere. Aceasta sugerează:

- A) SARIMA(0, 1, 0)  $\times$  (0, 1, 1)<sub>12</sub>
- B) SARIMA(0, 1, 1)  $\times$  (0, 1, 0)<sub>12</sub>
- C) SARIMA(1, 1, 0)  $\times$  (1, 1, 0)<sub>12</sub>
- D) SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub>

Răspuns: A – SARIMA(0, 1, 0)  $\times$  (0, 1, 1)<sub>12</sub>

**Model:** Lag-uri obișnuite – fără vârfuri în ACF/PACF; Lag-uri sezoniere – ACF se anulează la  $s$ , PACF descrește

**Interpretare:** Fără MA obișnuit ( $q = 0$ ); MA(1) sezonier indicat ( $Q = 1$ )

**Model:**  $(1 - L)(1 - L^{12})Y_t = (1 + \Theta_1 L^{12})\varepsilon_t$

## Test 13: Supradiferențierea

### Întrebare

După diferențiere, ACF arată un vârf negativ mare la lag 1 sau lag  $s$ . Aceasta indică de obicei:

- A) Modelul necesită mai mulți termeni AR
- B) Seria a fost supradiferențiată
- C) Seria este perfect staționară
- D) Prezența heteroscedasticității

## Test 13: Supradiferențierea

### Întrebare

După diferențiere, ACF arată un vârf negativ mare la lag 1 sau lag  $s$ . Aceasta indică de obicei:

- A) Modelul necesită mai mulți termeni AR
- B) Seria a fost supradiferențiată
- C) Seria este perfect staționară
- D) Prezența heteroscedasticității

Răspuns: B – Seria a fost supradiferențiată

**Semnătură:** ACF la lag 1  $\approx -0.5 \Rightarrow$  supradiferențiere la  $d$ ; ACF la lag  $s \approx -0.5 \Rightarrow$  supradiferențiere la  $D$   
**De ce?**  $\Delta^2 Y_t = \varepsilon_t - \varepsilon_{t-1}$  este MA(1) cu  $\theta = -1$ , dând  $\rho_1 = -0.5$

**Corecție:** Reduceți  $d$  sau  $D$  cu unu și re-examinați ACF/PACF

## Test 14: Orizontul de Prognoză

### Întrebare

Pentru un model SARIMA cu  $D = 1$ , ce se întâmplă cu intervalele de încredere ale prognozei când orizontul  $h \rightarrow \infty$ ?

- A) Converg la o lățime fixă
- B) Cresc fără limită
- C) Se micșorează la zero
- D) Oscilează sezonier

## Test 14: Orizontul de Prognoză

### Întrebare

Pentru un model SARIMA cu  $D = 1$ , ce se întâmplă cu intervalele de încredere ale prognozei când orizontul  $h \rightarrow \infty$ ?

- A) Converge la o lăţime fixă
- B) Cresc fără limită
- C) Se micşorează la zero
- D) Oscilează sezonier

Răspuns: B – Cresc fără limită

**Proprietatea rădăcinii unitare:** Orice rădăcină unitară cauzează varianţă de prognoză nemărginită

**Pentru SARIMA cu  $D = 1$ :**  $\text{Var}(\hat{Y}_{T+h} - Y_{T+h}) \rightarrow \infty$  când  $h \rightarrow \infty$

**Intuiţie:** Şocurile sezoniere se acumulează; prognozele pe termen lung au IC-uri largi

## Test 15: Selectarea Perioadei Sezoniere

### Întrebare

Aveți date zilnice care arată modele săptămânale clare. Ce perioadă sezonieră  $s$  ar trebui să folosiți într-un model SARIMA?

- A)  $s = 12$  (lunar)
- B)  $s = 7$  (săptămânal)
- C)  $s = 365$  (anual)
- D)  $s = 24$  (orar)

## Test 15: Selectarea Perioadei Sezoniere

### Întrebare

Aveți date zilnice care arată modele săptămânale clare. Ce perioadă sezonieră  $s$  ar trebui să folosiți într-un model SARIMA?

- A)  $s = 12$  (lunar)
- B)  $s = 7$  (săptămânal)
- C)  $s = 365$  (anual)
- D)  $s = 24$  (orar)

Răspuns: B –  $s = 7$  (săptămânal)

Date	Model	Perioada $s$
Zilnice	Săptămânal	7
Lunare	Anual	12
Trimestriale	Anual	4

**Regulă:**  $s$  = observații per ciclu al modelului dominant

## Test 16: Componenta AR Sezonieră

### Întrebare

În componenta sezonieră  $\Phi(L^s) = 1 - \Phi_1 L^s$ , ce ne spune coeficientul  $\Phi_1 = 0.8$ ?

- A) 80% din valoarea perioadei curente vine din perioada anterioară
- B) Există 80% corelație între observațiile consecutive
- C) Persistență sezonieră puternică: valoarea așteptată condiționată depinde de 80% din valoarea aceleiași perioade de anul trecut
- D) Modelul sezonier explică 80% din varianță

## Test 16: Componenta AR Sezonieră

### Întrebare

În componenta sezonieră  $\Phi(L^s) = 1 - \Phi_1 L^s$ , ce ne spune coeficientul  $\Phi_1 = 0.8$ ?

- A) 80% din valoarea perioadei curente vine din perioada anterioară
- B) Există 80% corelație între observațiile consecutive
- C) Persistență sezonieră puternică: valoarea așteptată condiționată depinde de 80% din valoarea aceleiași perioade de anul trecut
- D) Modelul sezonier explică 80% din varianță

Răspuns: C – Persistență sezonieră puternică ( $\Phi_1 = 0.8$ )

**SAR(1):**  $Y_t = \Phi_1 Y_{t-12} + \varepsilon_t$

**Cu  $\Phi_1 = 0.8$ :**  $Y_{Ian2024} = 0.8 \cdot Y_{Ian2023} + \varepsilon_t$

**Interpretare:** Dependență liniară puternică ( $\Phi_1 = 0.8$ ) de aceeași perioadă din anul trecut. Notă:  $\Phi_1$  este un coeficient de regresie, nu  $R^2$ ; autocorelația la lag 12 este  $\rho_{12} = \Phi_1 = 0,8$

**Staționaritate:** Necesită  $|\Phi_1| < 1$  (satisfăcută aici)

## Test 17: Staționaritatea Sezonieră

### Întrebare

Un proces sezonier cu  $\Phi_1 = 1$  în  $\text{SARIMA}(0, 0, 0) \times (1, 0, 0)_{12}$  este:

- A) Staționar
- B) Are o rădăcină unitară sezonieră (integrat sezonier)
- C) Exploziv
- D) Nedefinit

## Test 17: Staționaritatea Sezonieră

### Întrebare

Un proces sezonier cu  $\Phi_1 = 1$  în  $\text{SARIMA}(0, 0, 0) \times (1, 0, 0)_{12}$  este:

- A) Staționar
- B) Are o rădăcină unitară sezonieră (integrat sezonier)
- C) Exploziv
- D) Nedefinit

Răspuns: B – Are o rădăcină unitară sezonieră

**Model:**  $Y_t = Y_{t-12} + \varepsilon_t$  (mers aleatoriu sezonier)

**Proprietăți:** Varianța crește cu timpul; fiecare lună urmează propriul său mers aleatoriu; necesită  $D = 1$

**Analogie:** Ca mersul aleatoriu obișnuit dar la frecvența sezonieră

## Test 18: Compararea Modelelor

### Întrebare

Modelul A:  $\text{SARIMA}(1, 1, 1) \times (1, 1, 1)_{12}$  are  $\text{AIC} = 520$ . Modelul B:  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$  are  $\text{AIC} = 525$ . Care afirmație este cea mai corectă?

- A) Modelul A este întotdeauna mai bun deoarece are AIC mai mic
- B) Modelul B ar trebui preferat datorită parcimoniei în ciuda AIC mai mare
- C) Diferența AIC de 5 sugerează că Modelul A este substanțial mai bun
- D) Nu putem compara modele cu ordine diferite

## Test 18: Compararea Modelelor

### Întrebare

Modelul A:  $\text{SARIMA}(1, 1, 1) \times (1, 1, 1)_{12}$  are  $\text{AIC} = 520$ . Modelul B:  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$  are  $\text{AIC} = 525$ . Care afirmație este cea mai corectă?

- A) Modelul A este întotdeauna mai bun deoarece are AIC mai mic
- B) Modelul B ar trebui preferat datorită parcimoniei în ciuda AIC mai mare
- C) Diferența AIC de 5 sugerează că Modelul A este substanțial mai bun
- D) Nu putem compara modele cu ordine diferite

Răspuns: C – Diferența AIC de 5 sugerează că Modelul A este substanțial mai bun

**Regulă empirică:**  $\Delta\text{AIC} < 2$ : echivalente;  $2-10$ : anumite dovezi;  $> 10$ : dovezi puternice

**Aici:**  $\Delta\text{AIC} = 5$  sugerează Modelul A semnificativ mai bun

**Întotdeauna:** Verificați și diagnosticele reziduurilor și performanța prognozei.

## Test 19: Modele Sezoniere în Reziduuri

### Întrebare

După ajustarea unui model SARIMA, observați vârfuri ACF semnificative la lag-urile 12 și 24 în reziduuri. Ce indică aceasta?

- A) Modelul este corect specificat
- B) Componenta sezonieră este inadecvată
- C) Datele nu sunt sezoniere
- D) A apărut supraajustarea

## Test 19: Modele Sezoniere în Reziduuri

### Întrebare

După ajustarea unui model SARIMA, observați vârfuri ACF semnificative la lag-urile 12 și 24 în reziduuri. Ce indică aceasta?

- A) Modelul este corect specificat
- B) Componenta sezonieră este inadecvată
- C) Datele nu sunt sezoniere
- D) A apărut supraajustarea

### Răspuns: B – Componenta sezonieră este inadecvată

**Diagnostic:** Reziduurile bune ar trebui să fie zgomot alb (fără ACF semnificativ)

**ACF sezonier în reziduuri:** Modelul nu a capturat structura sezonieră; încercați să creșteți  $P$  sau  $Q$ ; verificați că  $D$  este corect

**Acțiune:** Încercați SARIMA cu ordin sezonier mai mare, verificați Ljung-Box la lag-urile sezoniere

## Test 20: Prognoză Practică

### Întrebare

Prognozați vânzări retail lunare cu  $SARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ . Pentru prognoza la 13 luni înainte, care observații istorice sunt cele mai influente?

- A) Doar cea mai recentă observație
- B) Observația de aceeași lună din anul trecut
- C) Toate observațiile în mod egal
- D) Doar observațiile din aceeași lună din toți anii anteriori

## Test 20: Prognoză Practică

### Întrebare

Prognozați vânzări retail lunare cu  $SARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ . Pentru prognoza la 13 luni înainte, care observații istorice sunt cele mai influente?

- A) Doar cea mai recentă observație
- B) Observația de aceeași lună din anul trecut
- C) Toate observațiile în mod egal
- D) Doar observațiile din aceeași lună din toți anii anteriori

Răspuns: B – Observația de aceeași lună din anul trecut

**Pentru 13 luni înainte:** Cea mai influentă este  $Y_{T-11}$  (aceeași lună anul trecut), de asemenea  $Y_T$  și  $Y_{T-12}$

**Intuiție:** “Ianuarie viitor arată ca ianuarie trecut, ajustat pentru trendul recent”

## Întrebări Adevărat/Fals (1-6)

### Întrebare

Determinați dacă fiecare afirmație este Adevărată sau Falsă:

1. Perioada sezonieră  $s$  pentru date trimestriale cu modele anuale este  $s = 4$ .
2. Modelele SARIMA pot gestiona doar o singură frecvență sezonieră.
3. Dacă AIC selectează  $\text{SARIMA}(1, 1, 1) \times (1, 1, 1)_{12}$  și BIC selectează modelul airline, BIC greșește întotdeauna.
4. Testul Kruskal-Wallis poate detecta sezonalitatea fără a presupune normalitate.
5. După ajustarea unui model SARIMA, reziduurile nu ar trebui să arate ACF semnificativ la lag-urile sezoniere.
6. Transformarea logaritmică convertește sezonalitatea multiplicativă în aditivă.

*Răspunsul pe slide-ul următor...*

## Soluții Adevărat/Fals (1-6)

### Răspunsuri

1. **ADEVĂRAT**: Datele trimestriale cu ciclu anual au  $s = 4$  trimestre pe an.
2. **ADEVĂRAT**: SARIMA standard gestionează un  $s$ ; sezonaliități multiple necesită TBATS sau termeni Fourier.
3. **FALS**: BIC penalizează complexitatea mai mult; modelul mai simplu poate fi mai bun pentru interpretare/proгноză.
4. **ADEVĂRAT**: Kruskal-Wallis este neparametric, comparând distribuțiile între sezoane.
5. **ADEVĂRAT**: ACF-ul reziduurilor ar trebui să fie în limitele de încredere la TOATE lag-urile inclusiv cele sezoniere.
6. **ADEVĂRAT**:  $\log(T \times S \times \varepsilon) = \log T + \log S + \log \varepsilon$  (formă aditivă).

## Problema 1: Expandarea Diferenței Sezoniere

### Exercițiu

Expandați complet  $(1 - L)(1 - L^{12})Y_t$ . Ce observații sunt implicate?

## Problema 1: Expandarea Diferenței Sezoniere

### Exercițiu

Expandăți complet  $(1 - L)(1 - L^{12})Y_t$ . Ce observații sunt implicate?

### Soluție

$$(1 - L)(1 - L^{12}) = 1 - L - L^{12} + L^{13}$$

$$\text{Prin urmare: } (1 - L)(1 - L^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$$

**Interpretare:** Aceasta este diferența diferențelor:

- ▣ Mai întâi diferența sezonieră:  $Y_t - Y_{t-12}$  (anul acesta vs anul trecut)
- ▣ Apoi diferența obișnuită:  $(Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13})$

## Problema 2: Expandarea Modelului Airline

### Exercițiu

Scrieți ecuația completă pentru modelul airline SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub>:  
$$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t$$

## Problema 2: Expandarea Modelului Airline

### Exercițiu

Scrieți ecuația completă pentru modelul airline SARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub>:  
 $(1 - L)(1 - L^{12})Y_t = (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t$

### Soluție

Expandați partea MA:  $(1 + \theta_1 L)(1 + \Theta_1 L^{12}) = 1 + \theta_1 L + \Theta_1 L^{12} + \theta_1 \Theta_1 L^{13}$

Modelul complet:  $Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \Theta_1 \varepsilon_{t-12} + \theta_1 \Theta_1 \varepsilon_{t-13}$

**Notă:** Termenul încrucișat  $\theta_1 \Theta_1 L^{13}$  este interacțiunea multiplicativă între componentele MA obișnuite și sezoniere.

## Problema 3: Numărarea Parametrilor

### Exercițiu

Câți parametri (excluzând  $\sigma^2$ ) sunt în  $\text{SARIMA}(2, 1, 1) \times (1, 0, 1)_4$ ?

### Problema 3: Numărarea Parametrilor

#### Exercițiu

Câți parametri (excluzând  $\sigma^2$ ) sunt în  $\text{SARIMA}(2, 1, 1) \times (1, 0, 1)_4$ ?

#### Soluție

- ▣ AR obișnuit( $p = 2$ ):  $\phi_1, \phi_2 \Rightarrow 2$  parametri
- ▣ MA obișnuit( $q = 1$ ):  $\theta_1 \Rightarrow 1$  parametru
- ▣ AR sezonier( $P = 1$ ):  $\Phi_1 \Rightarrow 1$  parametru
- ▣ MA sezonier( $Q = 1$ ):  $\Theta_1 \Rightarrow 1$  parametru

**Total: 5 parametri**

Notă: Ordinele de diferențiere ( $d = 1, D = 0$ ) nu adaugă parametri – sunt transformări aplicate datelor.

## Problema 4: Prognoza SARIMA

### Exercițiu

Dat modelul airline cu  $\theta_1 = -0.4$  și  $\Theta_1 = -0.6$ , și:

▣  $Y_T = 500, Y_{T-1} = 495, Y_{T-11} = 480, Y_{T-12} = 470$

▣  $\varepsilon_T = 5, \varepsilon_{T-11} = -3, \varepsilon_{T-12} = 2$

Prognozați  $Y_{T+1}$ .

## Problema 4: Prognoza SARIMA

### Exercițiu

Dat modelul airline cu  $\theta_1 = -0.4$  și  $\Theta_1 = -0.6$ , și:

$$\square Y_T = 500, Y_{T-1} = 495, Y_{T-11} = 480, Y_{T-12} = 470$$

$$\square \varepsilon_T = 5, \varepsilon_{T-11} = -3, \varepsilon_{T-12} = 2$$

Prognozați  $Y_{T+1}$ .

### Soluție

Din model:  $Y_{T+1} = Y_T + Y_{T-11} - Y_{T-12} + \varepsilon_{T+1} + \theta_1 \varepsilon_T + \Theta_1 \varepsilon_{T-11} + \theta_1 \Theta_1 \varepsilon_{T-12}$

Setând  $\mathbb{E}[\varepsilon_{T+1}] = 0$ :

$$\begin{aligned}\hat{Y}_{T+1} &= 500 + 480 - 470 + 0 + (-0.4)(5) + (-0.6)(-3) + (-0.4)(-0.6)(2) \\ &= 510 - 2 + 1.8 + 0.48 = 510.28\end{aligned}$$

## Problema 5: Identificarea Perioadei Sezoniere

### Exercițiu

Potrivești fiecare tip de date cu perioada sezonieră tipică s:

1. Date trimestriale de PIB
2. Vânzări retail lunare
3. Rezervări săptămânale la restaurante
4. Cerere zilnică de electricitate

## Problema 5: Identificarea Perioadei Sezoniere

### Exercițiu

Potrivii fiecare tip de date cu perioada sezonieră tipică  $s$ :

1. Date trimestriale de PIB
2. Vânzări retail lunare
3. Rezervări săptămânale la restaurante
4. Cerere zilnică de electricitate

### Soluție

1. PIB trimestrial:  $s = 4$  (ciclu anual pe 4 trimestre)
2. Vânzări retail lunare:  $s = 12$  (ciclu anual pe 12 luni)
3. Rezervări săptămânale la restaurante:  $s = 7$  (ciclu săptămânal) sau  $s = 52$  (anual)
4. Cerere zilnică de electricitate:  $s = 7$  (model săptămânal) sau  $s = 365$  (anual)

**Notă:** Unele serii au modele sezoniere multiple (de ex., datele zilnice pot avea cicluri săptămânale și anuale)

## Exemplu: Analiza Vânzărilor Retail Lunare

### Scenariu

Aveți 5 ani de date de vânzări retail lunare cu vârfuri clare în decembrie și scăderi în ianuarie. Construiți un model SARIMA potrivit.

### Abordare Pas cu Pas

1. **Inspecție vizuală:** Graficul arată trend ascendent + vârfuri puternice în decembrie
2. **Perioada sezonieră:** Date lunare cu model anual  $\Rightarrow s = 12$
3. **Transformare:** Considerați  $\log(Y_t)$  dacă amplitudinea sezonieră crește cu nivelul
4. **Diferențiere:** Încercați  $(1 - L)(1 - L^{12})Y_t$  – verificați ACF/PACF
5. **Selectarea modelului:** Începeți cu modelul airline, comparați prin AIC

## Exemplu: Interpretarea ACF/PACF pentru Date Sezoniere

### Modele Observate (după diferențiere)

- ▣ ACF: Semnificativ la lag-urile 1, 12; se anulează după lag 1 și lag 12
- ▣ PACF: Semnificativ la lag-urile 1, 12, 13; descrește la multiplii de 12

### Interpretare

**Componenta obișnuită:** ACF se anulează la 1  $\Rightarrow$  MA(1)

**Componenta sezonieră:** ACF semnificativ doar la lag 12  $\Rightarrow$  MA(1) sezonier

**Model sugerat:** SARIMA(0,  $d$ , 1)  $\times$  (0,  $D$ , 1)<sub>12</sub> – modelul airline.

**Verificare alternativă:** Dacă PACF ar fi arătat anulare la lag-urile sezoniere în loc de ACF, considerați termeni AR sezonieri.

## Exemplu: Implementare Python

### Ajustarea SARIMA în Python

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
import pmdarima as pm

# Ajustare manuală
model = SARIMAX(y, order=(0,1,1), seasonal_order=(0,1,1,12))
results = model.fit()
print(results.summary())

# Selecție automată
auto_model = pm.auto_arima(y, seasonal=True, m=12,
                           start_p=0, max_p=2,
                           start_q=0, max_q=2,
                           d=1, D=1,
                           trace=True)
```

## Exemplu: Interpretarea Rezultatelor SARIMA

### Rezultate Exemplu statsmodels

```

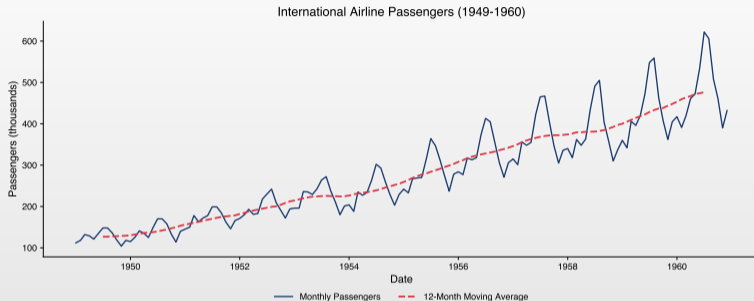
SARIMAX Results
=====
Model:          SARIMAX(0,1,1)x(0,1,1,12)    AIC:    1348.52
                                           BIC:    1358.21
=====
              coef    std err          z      P>|z|
-----
ma.L1        -0.4018     0.072    -5.58    0.000
ma.S.L12     -0.5521     0.081    -6.82    0.000
sigma2       1254.3201   142.856     8.78    0.000

```

### Interpretare

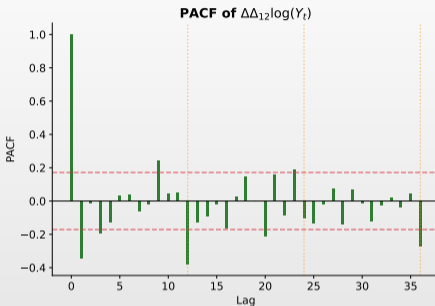
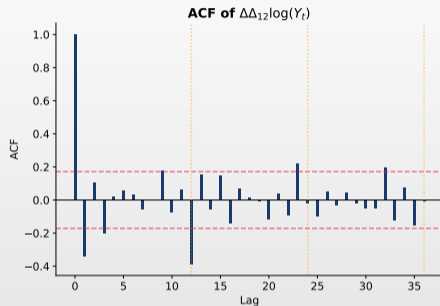
- ▣  $\hat{\theta}_1 = -0.40$ : MA negativ înseamnă că șocurile pozitive reduc valoarea perioadei următoare
- ▣  $\hat{\Theta}_1 = -0.55$ : Corelația pentru aceeași sezon este captată
- ▣ Ambii coeficienți semnificativi ( $p < 0.001$ );  $|\theta|, |\Theta| < 1$  – invertibil

## Studiu de Caz: Pasageri Aerieni (1949–1960)



- ▣ Set de date clasic Box-Jenkins: 144 observații lunare
- ▣ **Trend ascendent** clar și **tipar sezonier** (vârfuri vara)
- ▣ Amplitudinea sezonieră **crește cu nivelul**  $\Rightarrow$  sezonalitate multiplicativă
- ▣ Sugerează: transformare logaritmică + modelare SARIMA

## Analiza ACF/PACF După Diferențiere



- După  $(1 - L)(1 - L^{12})\log(Y_t)$ : seria pare staționară
- Vârf semnificativ la lag 1 în ACF  $\Rightarrow$  componentă MA(1)
- Vârf semnificativ la lag 12 în ACF  $\Rightarrow$  componentă MA(1) sezonieră
- Modelul sugerează: **SARIMA**(0, 1, 1)(0, 1, 1)<sub>12</sub> (modelul airline)

## Rezultate Estimare SARIMA: Date Pasageri Aerieni

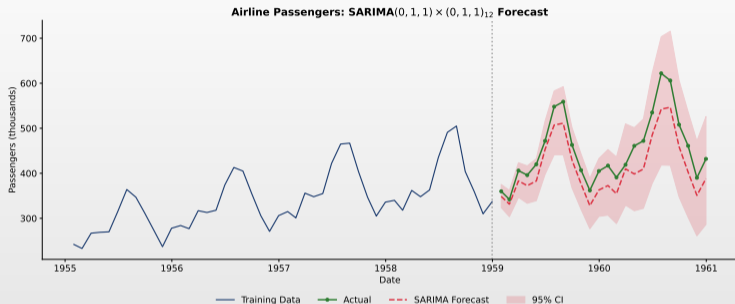
Model: SARIMA(0, 1, 1)(0, 1, 1)<sub>12</sub> pe log(Pasageri)

Parametru	Estimat	Eroare Std.	z-stat	valoare-p
$\theta_1$ (MA.L1)	-0.4018	0.0896	-4.48	< 0.001
$\Theta_1$ (MA.S.L12)	-0.5569	0.0731	-7.62	< 0.001
$\sigma^2$	0.00135	—	—	—

## Statistici de Ajustare a Modelului

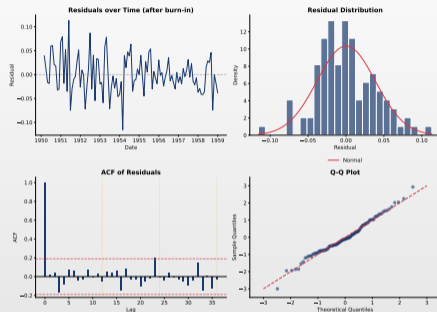
- Log-Verosimilitate: 244.70
- AIC: -483.40, BIC: -474.53
- Ambii coeficienți MA semnificativi și în limitele de invertibilitate

## Prognoză: 24 Luni Înainte



- Prognozele captează atât trendul cât și tiparul sezonier
- Intervalele de încredere de 95% se lărgesc pe orizontul de prognoză
- Vârfurile sezoniere (iulie-august) și scăderile (februarie) clar vizibile
- Modelul extrapolează cu succes tiparul sezonier multiplicativ

## Diagnostic Model



- Reziduurile par aleatoare fără modele sistematice
- Distribuție aproximativ normală (graficul Q-Q aproape de diagonală)
- ACF-ul reziduurilor în limitele de încredere – fără autocorelație semnificativă
- Testul Ljung-Box:  $p > 0.05$  la toate lag-urile testate  $\Rightarrow$  model adecvat

## Discuție: Sezonalitate Deterministă vs Stochastică

### Întrebare Cheie

Când ar trebui să folosiți variabile dummy sezoniere vs SARIMA pentru date sezoniere?

### Considerații

**Variabile dummy sezoniere** (determinate):

- Model fix, care se repetă în fiecare an
- Același efect decembrie în fiecare an
- Potrivite când sezonalitatea este stabilă

**SARIMA** (stochastic):

- Model sezonier în evoluție
- Decembrie anul acesta depinde de decembrie anul trecut
- Mai bun când amplitudinea sezonieră variază

## Discuție: Transformarea Logaritmică

### Întrebare Cheie

Când ar trebui să luați logaritmi înainte de a ajusta SARIMA?

### Îndrumări

**Folosiți transformarea log când:**

- ▣ Fluctuațiile sezoniere cresc cu nivelul (sezonalitate multiplicativă)
- ▣ Varianța crește în timp
- ▣ Datele sunt strict pozitive (prețuri, vânzări, numărători)

**Evitați log când:**

- ▣ Tiparul sezonier este aditiv (amplitudine constantă)
- ▣ Datele conțin zerouri sau negative
- ▣ Deja pe o scală de rate/proporții

**Sfat:** Comparați AIC-ul modelelor cu și fără transformare log.

## Discuție: Sezonalități Multiple

### Provocare

Datele zilnice de vânzări pot avea atât modele săptămânale (7 zile) cât și anuale (365 zile). Cum gestionați aceasta?

### Abordări

1. **SARIMA imbricat:** Modelați la frecvența mai scurtă, includeți mai lungă ca exogenă
2. **Modele TBATS/BATS:** Gestionează explicit sezonality multiple
3. **Termeni Fourier:** Adăugați termeni sin/cos pentru fiecare frecvență sezonieră
4. **Prophet/similare:** Instrumente moderne proiectate pentru sezonality multiple

**Notă:** SARIMA standard gestionează doar o perioadă sezonieră. Pentru sezonality complexă, considerați metode specializate.

## Discuție: Prognozarea Datelor Sezoniere

### Întrebare Cheie

Care sunt provocările unice ale prognozării seriilor de timp sezoniere?

### Provocări și Soluții

- ▣ **Orizontul contează:** Prognoza pe 12 luni înseamnă prezicerea unui ciclu complet
- ▣ **Incertitudinea crește:** Prognozele sezoniere compun incertitudinea obișnuită
- ▣ **Puncte de cotitură:** Captarea când sezoanele ating vârf/minim
- ▣ **Rupturi structurale:** COVID-19 a perturbat multe modele sezoniere

#### Bune practici:

- ▣ Folosiți validare încrucișată cu origine mobilă
- ▣ Comparați cu benchmark-ul naiv sezonier
- ▣ Raportați intervale de prognoză, mai ales la orizonturi sezoniere

## Exercițiu AI: Gândire critică

### Prompt de testat în ChatGPT / Claude / Copilot

“Descarcă de pe FRED producția lunară de electricitate din SUA (seria IPG2211A2N) din 2010-01 până în 2024-12 (180 observații). Identifică perioada de sezonalitate cu ACF și periodograma. Aplică transformarea Box-Cox dacă e nevoie. Estimează un model  $SARIMA(p, d, q) \times (P, D, Q)_{12}$  cu `auto_arima`. Împarte datele în antrenare (2010–2023) și test (2024), evaluează cu RMSE și MASE relativ la seasonal naive. Vreau cod Python complet cu grafice.”

### Exercițiu:

1. Rulați prompt-ul într-un LLM la alegere și analizați critic răspunsul.
2. Identifică corect perioada sezonieră  $s = 12$ ? Folosește ACF sau periodograma?
3. Aplică Box-Cox *înainte* de diferențiere? Justifică alegerea lui  $\lambda$ ?
4. Compară cu un baseline seasonal naive, sau doar raportează RMSE fără context?
5. Verifică reziduurile la lag-urile sezoniere (12, 24, 36)?

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

## Exerciții pentru Acasă

1. **Teoretic:** Arătați că  $(1 - L)(1 - L^4)$  poate fi scris ca  $(1 - L - L^4 + L^5)$  și explicați ce face această transformare datelor trimestriale cu sezonality anuală.
2. **Calcul:** Pentru  $SARIMA(1, 0, 0) \times (1, 0, 0)_4$  cu  $\phi_1 = 0.5$  și  $\Phi_1 = 0.8$ , scrieți polinomul AR complet și identificați toți coeficienții nenuli.
3. **Aplicat:** Descărcați datele lunare despre pasagerii aerieni și:
  - ▶ Reprezentați grafic seria și identificați trend/sezonality
  - ▶ Aplicați transformările potrivite
  - ▶ Ajustați modelul airline și interpretați coeficienții
  - ▶ Generați prognoze pe 24 de luni cu intervale de încredere
4. **Comparație:** Ajustați atât  $SARIMA(0, 1, 1) \times (0, 1, 1)_{12}$  cât și  $SARIMA(1, 1, 0) \times (1, 1, 0)_{12}$  pe datele despre pasagerii aerieni. Comparați folosind AIC, BIC și diagnosticele reziduurilor. Care este preferat?

## Indicii pentru Soluții

### Indicii

1. Expandați  $(1 - L)(1 - L^4) = 1 \cdot 1 - 1 \cdot L^4 - L \cdot 1 + L \cdot L^4 = 1 - L - L^4 + L^5$
2. Polinomul AR:  $(1 - \phi_1 L)(1 - \Phi_1 L^4) = 1 - 0.5L - 0.8L^4 + 0.4L^5$
3. Pentru datele pasagerilor aerieni:
  - ▶ Folosiți transformarea log (sezonalitate multiplicativă)
  - ▶ Atât  $d = 1$  cât și  $D = 1$  sunt necesare
  - ▶ Estimări tipice:  $\theta_1 \approx -0.4$ ,  $\Theta_1 \approx -0.6$
4. Modelul airline bazat pe MA se potrivește de obicei mai bine decât modelul AR sezonier pur pentru aceste date (AIC mai mic).

## Concluzii Cheie din Acest Seminar

### Puncte Principale

1. Diferențierea sezonieră  $(1 - L^s)$  elimină sezonalitatea stochastică
2. Notăția SARIMA:  $(p, d, q) \times (P, D, Q)_s$  separă obișnuitul de sezonier
3. Modelul airline este surprinzător de eficient pentru multe seturi de date
4. Structura multiplicativă creează termeni de interacțiune
5. ACF/PACF arată modele atât la lag-urile obișnuite cât și la cele sezoniere
6. Transformarea log adesea necesară pentru sezonalitatea multiplicativă

### Pașii Următori

Capitolul 5 va acoperi seriile de timp multivariate: modele VAR, cauzalitatea Granger și cointegrarea.