Analiza și Prognoză Seriilor de Timp

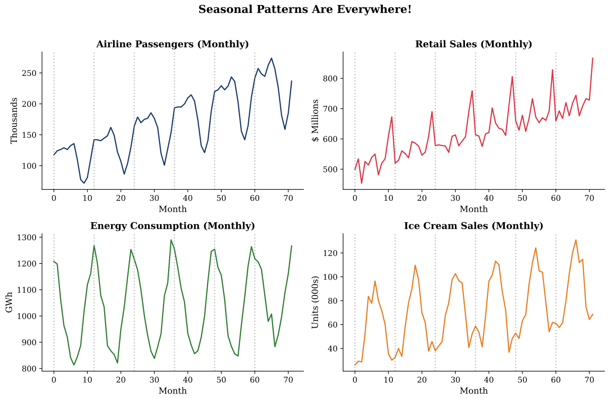
Capitolul 4: Modele SARIMA

Serii de Timp Sezoniere

Cuprins

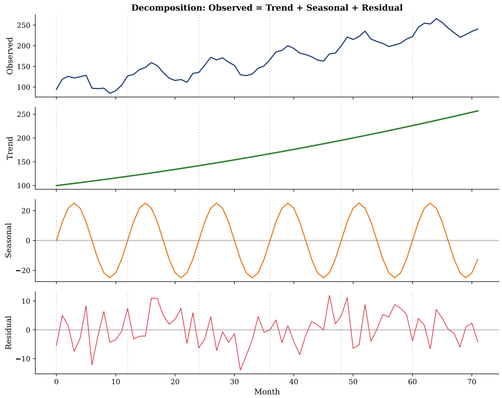
1. Sezonalitatea în seriile de timp
2. Diferențierea sezonieră
3. Modelul SARIMA
4. Tipare ACF și PACF sezoniere
5. Estimare și diagnosticare
6. Prognoză cu SARIMA
7. Aplicație pe date reale: Pasageri companiilor aeriene
8. Studiu de caz: Pasageri aerieni
9. Sumar

Exemplu motivațional: Sezonalitatea este peste tot



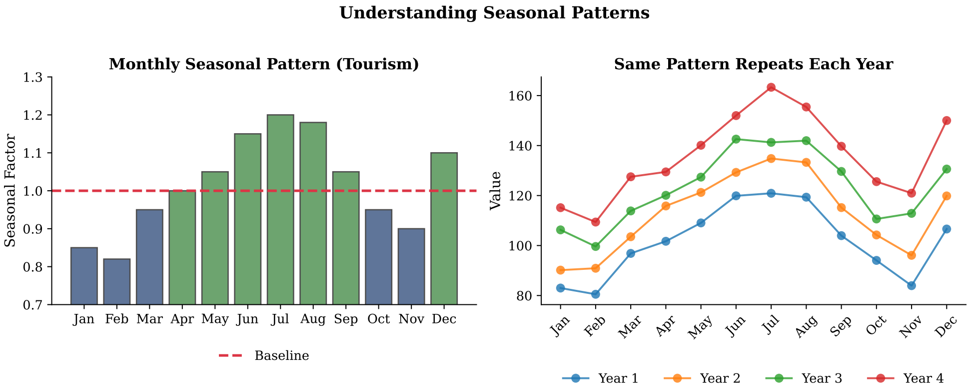
- Vânzările cu amănuntul prezintă **tipare anuale** clare: vârfuri în decembrie, minime în ianuarie
- Modelele ARIMA standard nu pot captura aceste **cicluri sezoniere repetitive**
- Ignorarea sezonalității duce la erori sistematice de prognoză

Înțelegerea componentelor sezoniere



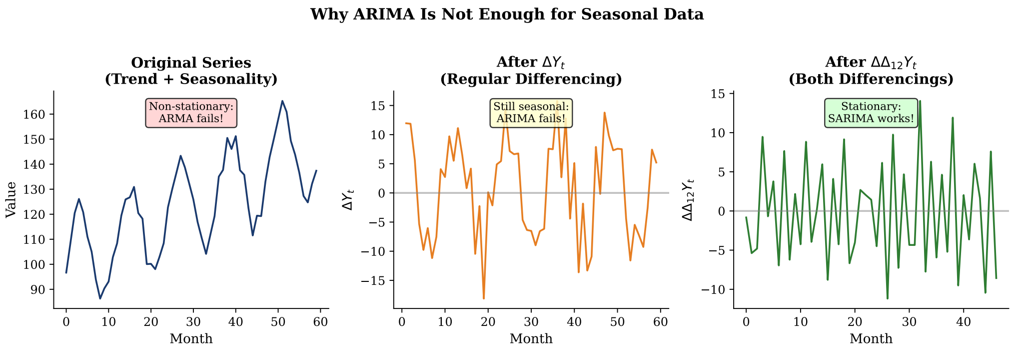
- Serie de timp sezonieră = **Trend** + **Tipar sezonier** + **Reziduuri**
- Descompunerea ajută la vizualizarea separată a fiecărei componente
- Modelele SARIMA captează atât dinamică trendului, cât și comportamentul sezonier

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 capacitatii

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 — modelul este incomplet
- SARIMA adaugă **termeni AR și MA sezonieri** pentru a captura aceste tipare

Ce vom învăța astăzi

Concepte

- Identificarea tiparelor sezoniere
- Operatorul de diferențiere sezonieră
- Notația SARIMA$(p, d, q)(P, D, Q)_s$
- Celebrul "Model Airline"
- Selecția modelului pentru date sezoniere

Abilități

- Diagnosticarea sezonalității din ACF/PACF
- Determinarea perioadei sezoniere s
- Alegerea ordinelor sezoniere (P, D, Q)
- Implementarea SARIMA în Python/R
- Prognoză seriilor de timp sezoniere

Ideea cheie

SARIMA = ARIMA aplicată la **două frecvențe**: nivelul obișnuit (pe termen scurt) și cel sezonier (pe termen lung)

Ce este sezonalitatea?

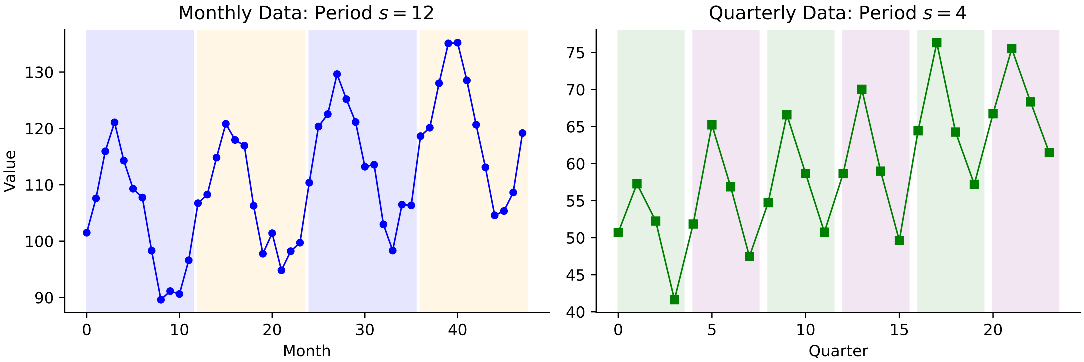
Definitie 1 (Sezonalitate)

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)

Sezonalitatea: Ilustrare vizuala



Perioade Sezoniere

Stânga: Date lunare cu $s = 12$ (ciclu anual). Dreapta: Date trimestriale cu $s = 4$. Tiparul se repetă la fiecare s perioade — această regularitate este exploatată de modelele SARIMA.

Exemple de date sezoniere

Serii economice

- Vânzări cu amănuntul (vârfuri de sarbatori)
- Turism (vara/iarna)
- Productie agricola
- Consum de energie
- Ocuparea fortei de munca (cicluri de angajare)

Alte domenii

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

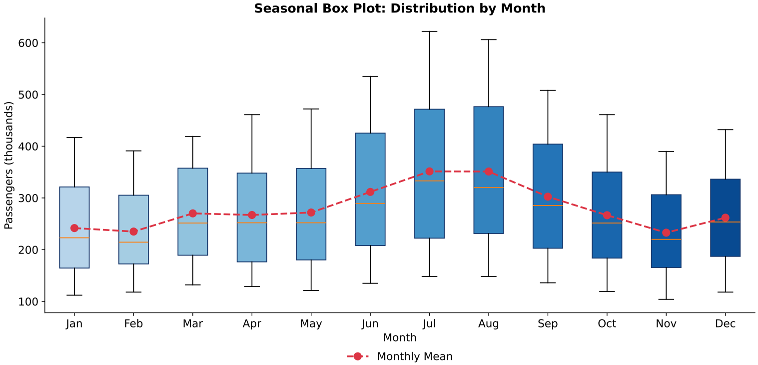
De ce contează

Ignorarea sezonalității duce la prognoze distorsionate și inferența invalidă!

Exemplu: Datele privind pasagerii companiilor aeriene

- Pasageri internaționali lunari ai companiilor aeriene (1949–1960)
- **Trend ascendent** clar și **amplitudine sezonieră crescătoare**
- Vârfurile din vara reflecta tiparele călătoriilor de vacanta

Vizualizarea tiparelor sezoniere



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

Sezonalitate deterministă vs stochastică

Sezonalitate deterministă

Tipar sezonier fix: $Y_t = \sum_{j=1}^{s} \gamma_j D_{jt} + \varepsilon_t$ unde D_{jt} sunt variabile dummy sezoniere.

Proprietăți:

- Tiparul constant în timp
- Eliminăt prin regresie

Sezonalitate stochastică

Tipar sezonier în evolutie: $\Delta_s Y_t = Y_t - Y_{t-s}$ prezintă structura de dependență.

Proprietăți:

- Tiparul evoluează în timp
- Necesită diferențiere sezonieră

Detectarea sezonalității

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 $(s, 2s, 3s, \ldots)$

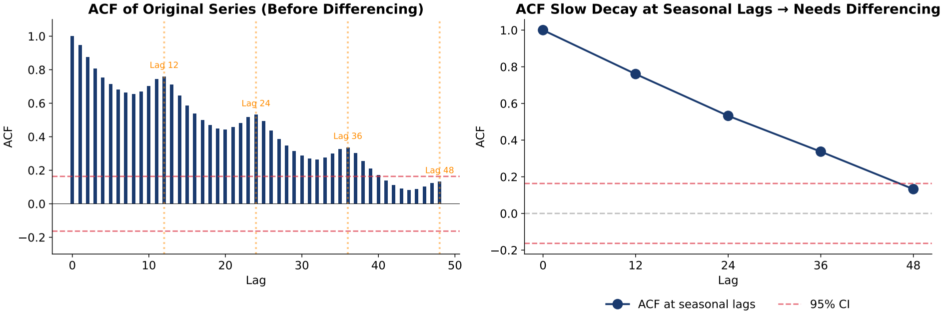
Teste statistice

- Teste de rădăcină unitară sezonieră (HEGY, CH, OCSB)
- Testul F pentru variabile dummy sezoniere
- Testul Kruskal-Wallis (neparametric)

Semnatura ACF

Sezonalitate puternică: ACF prezintă vârfuri semnificâtive la lag-urile $s, 2s, 3s, \ldots$

ACF relevă structura sezonieră



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

Testul F pentru variabile dummy sezoniere: Intuiție

Ce face acest test?

Testează dacă **valorile medii diferă semnificâțiv între sezoane**.

- Dacă media din ianuarie $\neq$ media din februarie $\neq$ ... $\neq$ media din decembrie $\Rightarrow$ sezonalitate
- Compară un model CU variabile dummy sezoniere vs. un model FARA

Modelele comparate

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

Ideea cheie

Dacă adăugarea variabilelor dummy sezoniere **reduce semnificâțiv** erorile de predicție, atunci sezonalitatea este prezentă.

Testul F pentru variabile dummy sezoniere: Formula și exemplu

Formula statisticii F

$$F = \frac{(SSR_R - SSR_U)/(s-1)}{SSR_U/(n-s)} \sim F_{s-1,n-s}$$

- 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)

$SSR_R = 15000$, $SSR_U = 8500$, $s = 12$

$$F = \frac{(15000 - 8500)/11}{8500/108} = \frac{590.9}{78.7} = 7.51$$

Valoarea critică $F_{0.05,11,108} \approx 1.87$. Deoarece $7.51 > 1.87$: **Respingem** $H_0 \Rightarrow$ Sezonalitate prezentă!

Testul Kruskal-Wallis: Intuiție

Ce face acest test?

Un test **neparametric** care verifică dacă observațiile din diferite sezoane provin din aceeași distribuție.

- Funcționează prin **ordonarea** tuturor observațiilor de la cea mai mică la cea mai mare
- Verifică dacă rangurile sunt distribuite uniform între sezoane
- Dacă un sezon are în mod constant ranguri mai mari/mici $\Rightarrow$ sezonalitate

De ce să-l folosim în locul testului F?

- **Fără ipoteza de normalitate** – funcționează cu orice distribuție
- **Robust la valori extreme** – valorile extreme nu distorsionează rezultatele

Limitare

Mai putin puternic decat testul F când datele SUNT distribuite normal.

Testul Kruskal-Wallis: Formula și exemplu

Statistică de test

$H = \frac{12}{N(N+1)} \sum_{j=1}^{s} \frac{R_j^2}{n_j} - 3(N+1)$ unde N = total obs., n_j = obs. în sezonul j, R_j = 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: $R_4 = 90$

$$H = \frac{12}{20 \times 21} \left(\frac{15^2}{5} + \frac{35^2}{5} + \frac{70^2}{5} + \frac{90^2}{5} \right) - 3(21) = 12.6$$

Valoarea critică $\chi^2_{0.05,3} = 7.81$. Deoarece $12.6 > 7.81$: **Respingem** $H_0 \Rightarrow$ Sezonalitate!

In Python

```
scipy.stats.kruskal(q1, q2, q3, q4)
```

Testul HEGY: Ce problemă rezolvă?

Întrebarea cheie

Având o serie de timp sezonieră, trebuie să știm:

1. Are nevoie de **diferențiere obișnuită** $(1 - L)$? $\Rightarrow$ setam $d = 1$
2. Are nevoie de **diferențiere sezonieră** $(1 - L^s)$? $\Rightarrow$ setam $D = 1$

HEGY testează pentru **ambele** tipuri de rădăcini unitare simultan!

De ce să nu folosim doar ADF?

ADF testează doar pentru o rădăcină unitară **obișnuită** la frecvența zero. Datele sezoniere pot avea rădăcini unitare la **frecvențe sezoniere** pe care ADF le omite!

HEGY testează frecvențe multiple

Trimestrial: testează la 0, π, $\pm\pi/2$. Lunar: testează la 0, π, $\pm\pi/6$, $\pm\pi/3$, $\pm\pi/2$, $\pm2\pi/3$, $\pm5\pi/6$.

Testul HEGY: Formula de regresie (Trimestrial)

Regresia auxiliara HEGY

Pentru date trimestriale ($s = 4$), eștimam:

$$\Delta_4 y_t = \pi_1 z_{1,t-1} + \pi_2 z_{2,t-1} + \pi_3 z_{3,t-2} + \pi_4 z_{4,t-2} + \sum_{j=1}^{k} \phi_j \Delta_4 y_{t-j} + \varepsilon_t$$

Variabile transformate

$$z_{1t} = (1 + L + L^2 + L^3)y_t = y_t + y_{t-1} + y_{t-2} + y_{t-3}$$
$$z_{2t} = -(1 - L + L^2 - L^3)y_t = -y_t + y_{t-1} - y_{t-2} + y_{t-3}$$
$$z_{3t} = -(1 - L^2)y_t = -y_t + y_{t-2} \quad ; \quad z_{4t} = -(L - L^3)y_t = -y_{t-1} + y_{t-3}$$

Ipoteze

$H_0 : \pi_1 = 0$ (frecv. 0), $H_0 : \pi_2 = 0$ (frecv. π), $H_0 : \pi_3 = \pi_4 = 0$ (frecv. $\pm\pi/2$)

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

Sa presupunem ca HEGY da: $t_1 = -1.52$, $t_2 = -4.21$, $F_{34} = 2.15$

- $t_1 = -1.52 > -2.88$: Nu putem respinge $\Rightarrow$ **necesită** $d = 1$
- $t_2 = -4.21 < -2.88$: Respingem $\Rightarrow$ fără rădăcină unitară la π
- $F_{34} = 2.15 < 6.57$: 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	**Fără** 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 ca NU există rădăcină unitară" (conservator fata de diferențiere)
- CH: "Demonstrați ca EXISTA rădăcină unitară" (conservator fata de variabile dummy)
- Folosiți **ambele** teste pentru concluzii robuste!

Testul Canova-Hansen: Formula

Procedura de testare

1. Regresam y_t pe variabile dummy sezoniere: $y_t = \sum_{j=1}^{s} \gamma_j D_{jt} + u_t$
2. Calculam sumele partiale la frecvența sezonieră λ_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

$$LM_i = \frac{1}{T^2 \hat{\omega}_i} \left[\sum_{t=1}^{T} (S_{it}^{(c)})^2 + \sum_{t=1}^{T} (S_{it}^{(s)})^2 \right]$$

unde $\hat{\omega}_i$ = eștimator consistent al densității spectrale la frecvența λ_i.

Decizie

Respingem H_0 (staționaritate) dacă $LM >$ valoare critică $\Rightarrow$ este necesară diferențierea sezonieră.

Sumar: Alegerea testului de sezonalitate potrivit

Test	H_0	Dacă respingem	Cel mai bun pentru
Test F	Fără sezonalitate	Sezonalitate există	Date normale
Kruskal-Wallis	Fără dif. sezonieră	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

Ideea cheie

Test F/Kruskal-Wallis: "*Există sezonalitate?*"
HEGY/Canova-Hansen: "*Ce tip?*" (deterministă vs stochastică)

Operatorul de diferența sezonieră

Definitie 2 (Diferenta sezonieră)

Operatorul de diferența sezonieră Δ_s este definit ca:

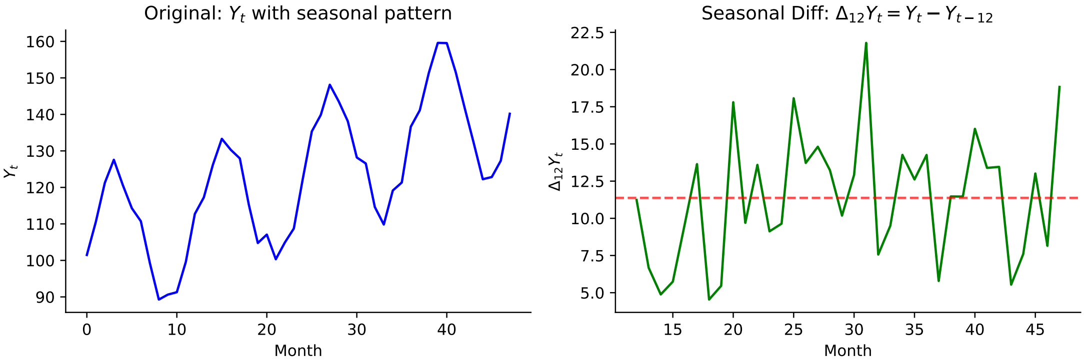
$$\Delta_s Y_t = (1 - L^s)Y_t = Y_t - Y_{t-s}$$

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

Diferenta sezonieră: Ilustrare vizuala



Efectul Diferențierii Sezoniere

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

Demonstrație: Diferențierea Sezonieră Elimină Sezonalitatea Deterministă

Afirmație: Dacă $Y_t = \mu_t + \varepsilon_t$ unde $\mu_t = \mu_{t-s}$ (medie periodică), atunci $\Delta_s Y_t$ elimină media sezonieră.

Demonstrație: Fie $Y_t = \mu_t + \varepsilon_t$ unde μ_t are perioadă s. Aplicăm diferența sezonieră:

$$\begin{aligned}\Delta_s Y_t &= Y_t - Y_{t-s} = (\mu_t + \varepsilon_t) - (\mu_{t-s} + \varepsilon_{t-s}) \\ &= \mu_t - \mu_{t-s} + \varepsilon_t - \varepsilon_{t-s} \\ &= 0 + \varepsilon_t - \varepsilon_{t-s} \quad (\text{deoarece } \mu_t = \mu_{t-s})\end{aligned}$$

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)
- Autocovarianța: $\gamma(s) = -\sigma^2$, $\gamma(k) = 0$ pentru $k \neq 0, s$

Rezultat

Diferențierea sezonieră transformă tiparul sezonier periodic în MA(1) la lag-ul sezonier.

Combinarea diferențierii obișnuite și sezoniere

Diferențiere completa

Pentru serii cu atât trend cât și sezonalitate:

$$\Delta\Delta_s Y_t = (1 - L)(1 - L^s)Y_t$$

Dezvoltare

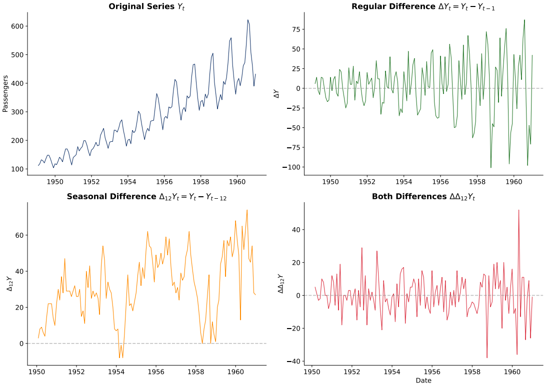
$(1 - L)(1 - L^s)Y_t = Y_t - Y_{t-1} - Y_{t-s} + Y_{t-s-1}$
Pentru date lunare ($s = 12$):

$$\Delta\Delta_{12} Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$$

Ordinea diferențierii

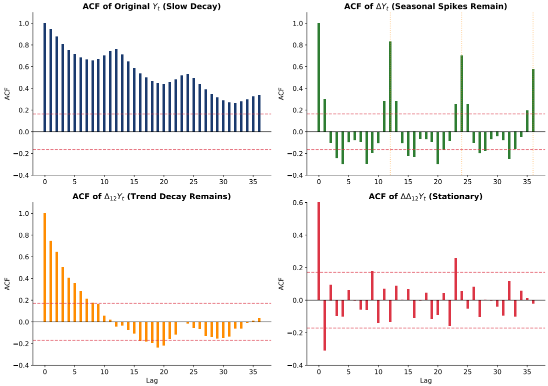
- d: numărul de diferențe obișnuite (eliminarea trendului)
- D: numărul de diferențe sezoniere (eliminarea trendului sezonier)

Efectul operatiilor de diferențiere



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

ACF înainte și după diferențiere



- ACF originală: descreștere lentă indică nestaționaritate
- După Δ: vârfuri sezoniere raman la lag-urile 12, 24, 36
- După Δ_{12}: descreșterea de trend ramane la lag-urile inițiale
- După $\Delta\Delta_{12}$: ACF se opreste brusc $\Rightarrow$ **staționară**

Integrare sezonieră

Definitie 3 (Proces integrat sezonier)

O serie Y_t este **integrata sezonier** de ordinul $(d, D)_s$, scrisa $Y_t \sim I(d, D)_s$, 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

Definitia modelului SARIMA

Definitie 4 (SARIMA$(p, d, q) \times (P, D, Q)_s$)

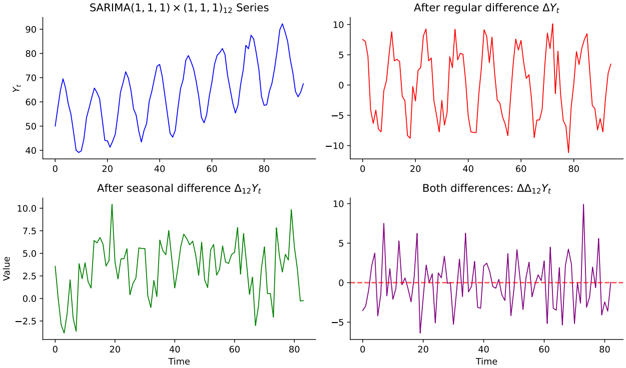
Modelul **Seasonal ARIMA** este:

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

Componente

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

SARIMA: Ilustrare vizuala



Strategia de Diferențiere

Transformare progresivă: Originală → diferență obișnuită (elimină trendul) → diferență sezonieră (elimină sezonalitatea) → ambele. Aplicați diferențierea minimă necesară pentru a obține staționaritate.

Demonstrație: Structura Multiplicâtivă Sezonieră

De ce multiplicâtivă? Considerăm SARIMA$(1,0,0)\times(1,0,0)_s$:

$$(1-\phi L)(1-\Phi L^s)Y_t=\varepsilon_t$$

Dezvoltăm produsul:

$$\begin{aligned}(1-\phi L)(1-\Phi L^s)Y_t &= Y_t-\phi LY_t-\Phi L^sY_t+\phi\Phi L^{s+1}Y_t\\ &= Y_t-\phi Y_{t-1}-\Phi Y_{t-s}+\phi\Phi Y_{t-s-1}\end{aligned}$$

Rezultat: Modelul include un **termen de interacțiune** $\phi\Phi Y_{t-s-1}$

Interpretare (Lunar, $s=12$)

Y_t depinde de:

- Y_{t-1}: Luna trecută (dinamică pe termen scurt)
- Y_{t-12}: Aceeași lună anul trecut (efectul sezonier)
- Y_{t-13}: Interacțiunea ambelor efecte

Parsimonie

Forma multiplicâtivă: 2 parametri (ϕ,Φ). Forma aditivă ar necesită 3+ parametri.

Notația SARIMA

Specificâție completa

SARIMA$(p, d, q) \times (P, D, Q)_s$ are 7 parametri de specificăt:

Parametru	**Semnificâție**
p	Ordinul AR non-sezonier
d	Ordinul diferențierii non-sezoniere
q	Ordinul MA non-sezonier
P	Ordinul AR sezonier
D	Ordinul diferențierii sezoniere
Q	Ordinul MA sezonier
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), și atât diferențiere obișnuită cât și sezonieră.

Modele SARIMA comune

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

$(1-L)(1-L^s)Y_t = (1+\theta L)(1+\Theta L^s)\varepsilon_t$ – Model clasic (Box & Jenkins, 1970)

SARIMA$(1,0,0) \times (1,0,0)_s$

$(1-\phi L)(1-\Phi L^s)Y_t = \varepsilon_t$ – AR sezonier și non-sezonier pur

SARIMA$(0,1,1) \times (0,1,0)_s$

$(1-L)(1-L^s)Y_t = (1+\theta L)\varepsilon_t$ – Random walk + dif. sezonieră + MA(1)

Structura multiplicâțivă

De ce multiplicâțivă?

Părțile sezonieră și non-sezonieră se **înmulțesc**:

$$\phi(L)\Phi(L^s) \quad \text{si} \quad \theta(L)\Theta(L^s)$$

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

$(1-\phi L)(1-\Phi L^{12})Y_t = \varepsilon_t$
Dezvoltand: $Y_t - \phi Y_{t-1} - \Phi Y_{t-12} + \phi\Phi Y_{t-13} = \varepsilon_t$
Termenul incrucist $\phi\Phi Y_{t-13}$ captează interactiunea!

Interpretare

Structura multiplicâțivă permite modelarea parsimonioasă a tiparelor sezoniere complexe cu puțini parametri.

ACF/PACF pentru modele sezoniere

Ideea cheie

Modelele sezoniere prezintă tipare la ambele:

- Lag-uri non-sezoniere: $1, 2, 3, \ldots$
- Lag-uri sezoniere: $s, 2s, 3s, \ldots$

Model	**ACF**	**PACF**
SAR(P)	Descreste la $s, 2s, \ldots$	Se oprește după Ps
SMA(Q)	Se oprește după Qs	Descreste la $s, 2s, \ldots$
SARMA	Descreste la lag-uri sezoniere	Descreste la lag-uri sezoniere

Exemplu: ACF/PACF pentru modelul Airline

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

După diferențiere $W_t = (1-L)(1-L^{12})Y_t$:

$$W_t = (1+\theta L)(1+\Theta L^{12})\varepsilon_t$$

Tiparul ACF așteptat

- Vârf la lag-ul 1 (de la θ)
- Vârf la lag-ul 12 (de la Θ)
- Vârf la lag-ul 13 (de la interactiunea $\theta \cdot \Theta$)
- Toate celelalte lag-uri aproape de zero

Tiparul PACF așteptat

- Descreștere exponențială la lag-urile $1, 2, 3, \ldots$
- Descreștere exponențială la lag-urile $12, 24, 36, \ldots$

Ghid de identificare a modelului

Proces pas cu pas

1. Examinați ACF pentru descreștere lentă la lag-uri sezoniere $\Rightarrow$ diferențiere sezonieră
2. După diferențiere, verificați tiparele ACF/PACF
3. Comportamentul non-sezonier la lag-urile $1, 2, \ldots, s-1$
4. Comportamentul sezonier la lag-urile $s, 2s, 3s, \ldots$

Sfaturi practice

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

Metode de estimare

Estimare prin verosimilitate maxima

Abordare standard pentru SARIMA:

- MLE condiționată (condiționată de valorile inițiale)
- MLE exacta (prin filtrul Kalman)

Considerații computationale

- Mai multi parametri decat ARIMA $\Rightarrow$ mai multe date necesare
- Parametrii sezonieri eștimati din lag-urile $s, 2s, \ldots$
- Necesită suficiente cicluri sezoniere (cel putin 3-4 ani de date lunare)

Staționaritate și invertibilitate

Condiții de staționaritate

Atât polinoamele AR non-sezoniere cât și sezoniere trebuie să aibă rădăcini în afara cercului unitate:

- $\phi(z) = 0 \Rightarrow |z| > 1$
- $\Phi(z^s) = 0 \Rightarrow |z| > 1$

Condiții de invertibilitate

Atât polinoamele MA non-sezoniere cât și sezoniere trebuie să aibă rădăcini în afara cercului unitate:

- $\theta(z) = 0 \Rightarrow |z| > 1$
- $\Theta(z^s) = 0 \Rightarrow |z| > 1$

Verificarea diagnosticării

Analiza reziduurilor

După ajustarea SARIMA, verificați ca reziduurile sunt zgomot alb:

1. Graficul reziduurilor în timp (fără tipare)
2. ACF a reziduurilor (fără vârfuri semnificâtive)
3. Testul Ljung-Box la lag-uri multiple inclusiv sezoniere
4. Teste de normalitate (grafic Q-Q, Jarque-Bera)

Important

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

Criterii de selectie a modelului

Criterii informationale

Comparați modelele SARIMA concurente folosind:

- $\text{AIC} = -2\ln(L) + 2k$
- $\text{BIC} = -2\ln(L) + k\ln(n)$
- $\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}$ (corectat pentru esantioane mici)

unde $k = p + q + P + Q + 1$ (plus 1 pentru varianța).

Auto-SARIMA

`pmdarima.auto_arima()` din Python cu `seasonal=True` cauta automat $(p, d, q) \times (P, D, Q)_s$ optim.

Prognoze punctuale

Calculul prognozei

Prognozele SARIMA sunt calculate recursiv:

- Înlocuiți ε_{T+h} viitor cu 0
- Înlocuiți Y_{T+h} viitor cu prognozele $\hat{Y}_{T+h|T}$
- Folosiți valorile trecute cunoscute $Y_T, Y_{T-1}, \ldots$

Tiparul sezonier în prognoze

Prognozele SARIMA captează în mod natural sezonalitatea:

- Pe termen scurt: influențate de valorile recente
- Pe termen lung: revin la tiparul sezonier

Intervale de prognoză

Cuantificarea incertitudinii

Interval de predicție $(1-\alpha)\%$:

$$\hat{Y}_{T+h|T} \pm z_{\alpha/2}\sqrt{\text{Var}(e_{T+h})}$$

Varianta calculata din reprezentărea $\text{MA}(\infty)$.

Proprietăți cheie

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

Prognoze pe orizont lung

Comportamentul când $h \to \infty$

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

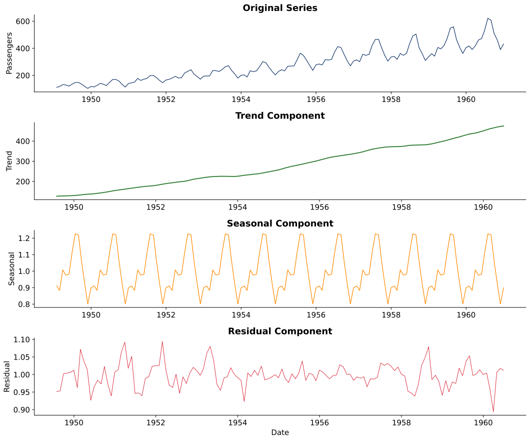
Implicație practică

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

Datele privind pasagerii companiilor aeriene

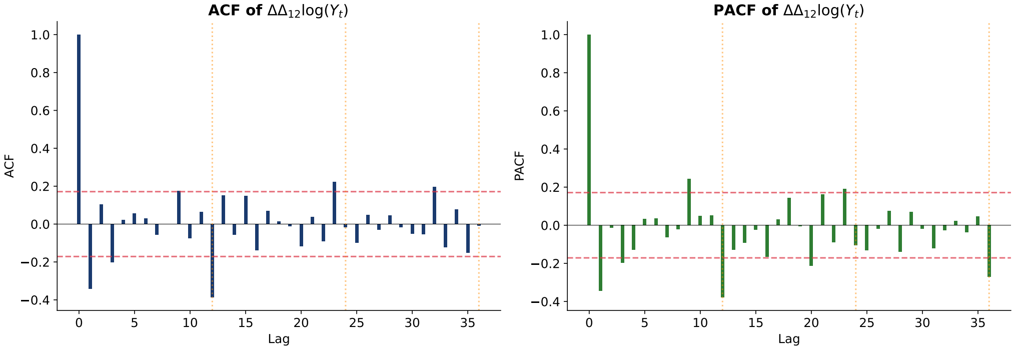
- Set de date clasic: Pasageri internaționali lunari ai companiilor aeriene (1949-1960)
- Trend ascendent clar și amplitudine sezonieră crescătoare

Descompunerea sezonieră



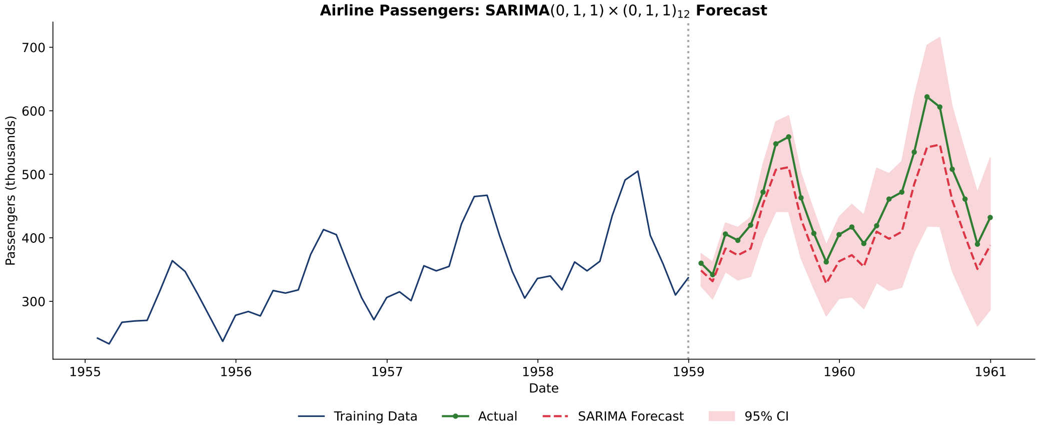
- Trend: Crestere puternică ascendenta
- Sezonalitate: Vârfuri de vară (călătorii de vacanta)
- Rezidual: Variație aleatoare după eliminarea trendului și sezonului

Analiza ACF/PACF



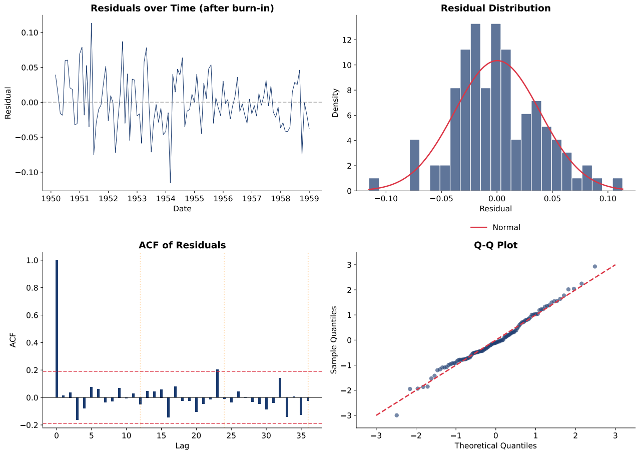
- După diferențierea $\Delta\Delta_{12}$: vârfuri la lag-urile 1 și 12
- Sugerează SARIMA$(0,1,1)\times(0,1,1)_{12}$ (Modelul Airline)

Rezultatele prognozei SARIMA



- SARIMA captează atât trendul cât și tiparul sezonier
- Prognozele prezintă vârfuri și minime sezoniere corespunzătoare

Diagnosticarea modelului



- Reziduurile par aleatorii; ACF în limite la toate lag-urile
- Modelul captează adecvat structura sezonieră

Implementare în Python

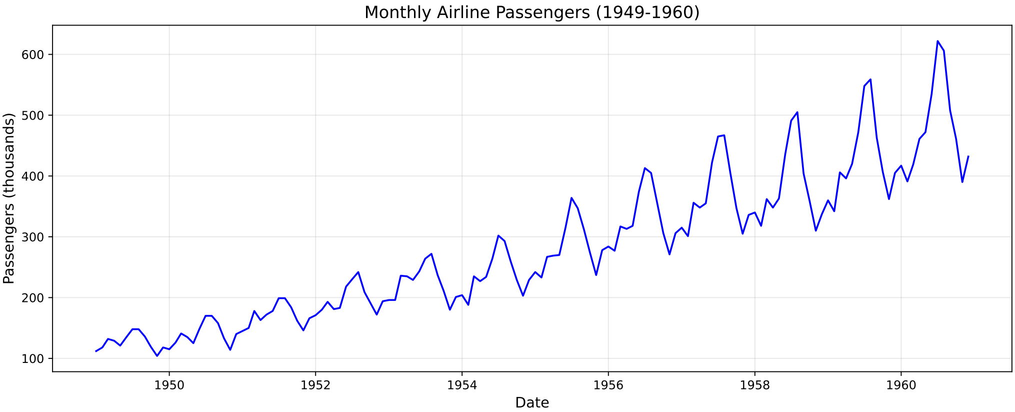
Ajustarea SARIMA în Python

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
model = SARIMAX(y, order=(0,1,1), seasonal_order=(0,1,1,12))
results = model.fit()
forecast = results.get_forecast(steps=24)
```

Nota

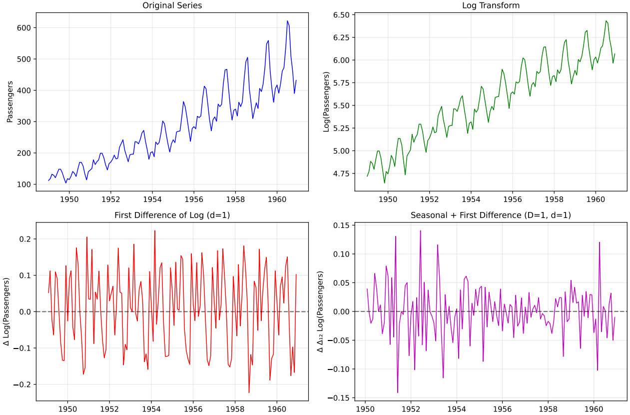
Exemple complete în Python cu comentarii sunt furnizate în caietele Jupyter.

Studiu de caz: Date despre pasagerii aerieni



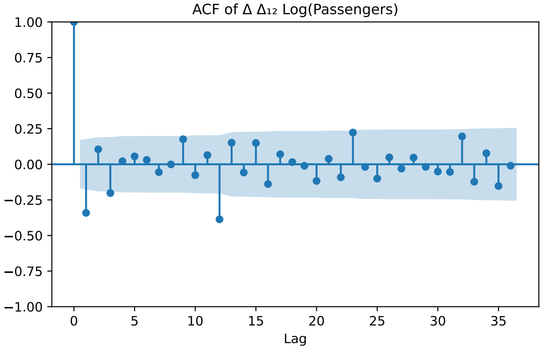
- Setul de date clasic Box-Jenkins: pasageri aerieni lunari (1949-1960)
- Trend ascendent clar și amplitudine sezonieră crescătoare
- Sezonalitatea multiplicâțivă sugerează transformarea logaritmică

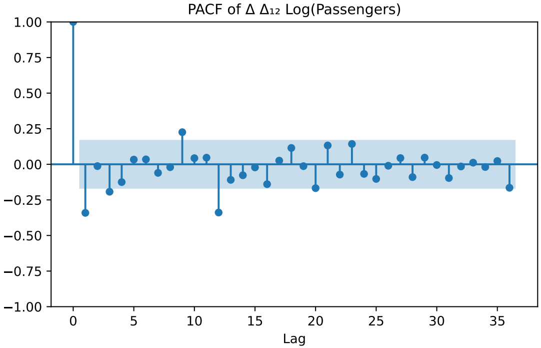
Pasul 1: Transformări



- Transformarea log stabilizează varianța (multiplicâțiv → aditiv)
- Prima diferență elimină trendul; diferența sezonieră elimină sezonalitatea
- Seria dublu diferențiată pare staționară

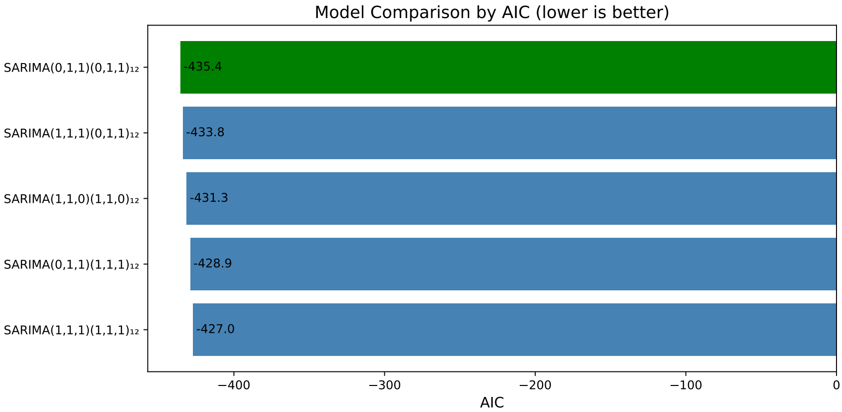
Pasul 2: Analiza ACF/PACF





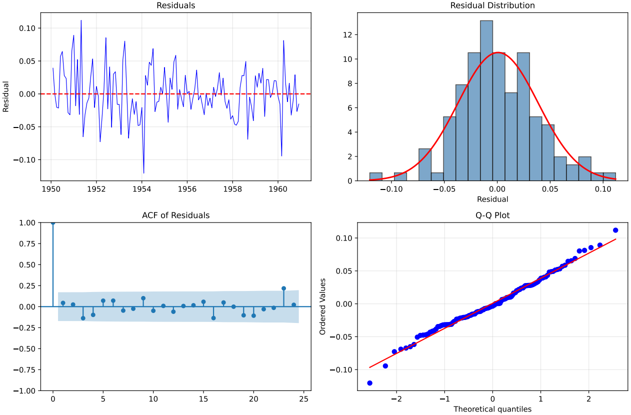
- ACF: Vârf semnificâțiv la lag 1 și lag 12 $\Rightarrow$ MA(1), SMA(1)
- PACF: Tipar de descreștere exponențială confirmă structura MA
- Sugerează SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ (modelul airline)

Pasul 3: Compararea modelelor



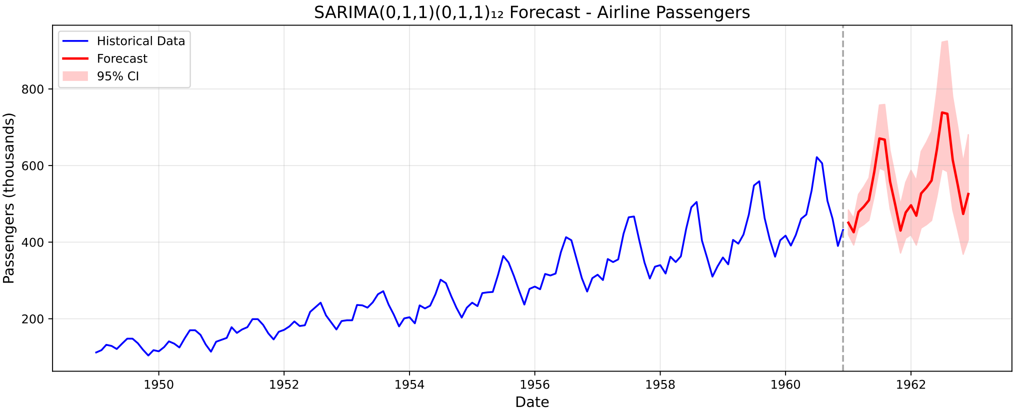
- Comparăm modelele SARIMA cândidate folosind criteriul AIC
- SARIMA$(0,1,1) \times (0,1,1)_{12}$ oferă cea mai bună ajustare (AIC minim)
- Acesta este faimosul "model airline" identificat de Box & Jenkins

Pasul 4: Diagnosticarea reziduurilor



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

Pasul 5: Prognoză



- 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

Concluzii cheie

Puncte principale

1. **Sezonalitatea** este comuna în datele economice și de afaceri
2. **Diferențierea sezonieră** $(1 - L^s)$ elimină sezonalitatea stochastică
3. **SARIMA**$(p, d, q) \times (P, D, Q)_s$ extinde ARIMA pentru date sezoniere
4. **Structura multiplicâțivă** captează interactiunile sezon-trend
5. **ACF/PACF** prezintă tipare la ambele lag-uri obișnuite și sezoniere
6. **Selecția modelului**: Folosiți AIC/BIC sau algoritmi auto-SARIMA

Pașii următorii

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

Întrebarea 1

Întrebare

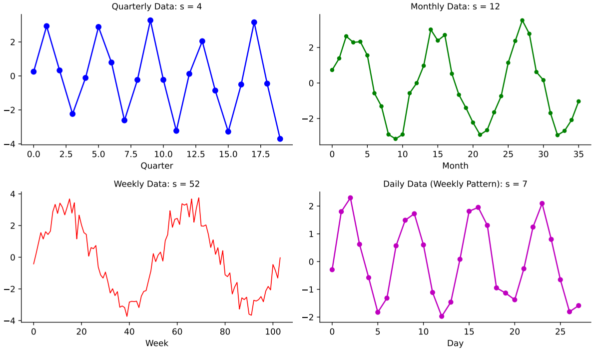
Pentru date lunare cu sezonalitate anuala, care este perioadă sezonieră s?

- **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

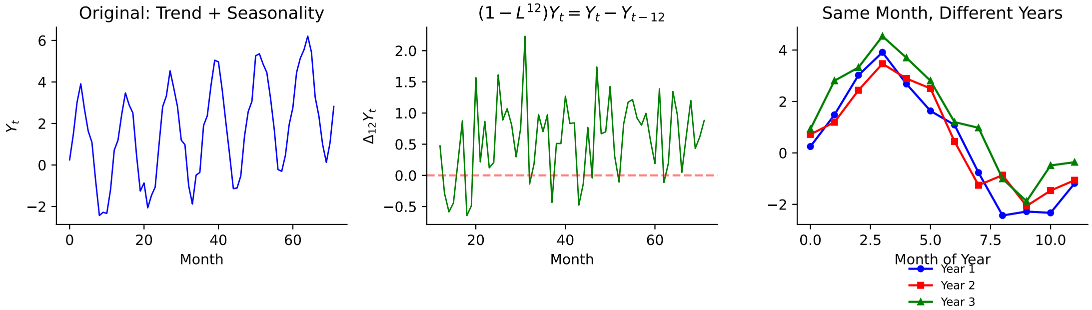
Ce face operatorul de diferența sezonieră $(1 - L^{12})$ unei serii lunare?

- 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

$(1 - L^{12})Y_t = Y_t - Y_{t-12}$ elimină tiparul sezonier prin compararea acelorasi luni.



Întrebarea 3

Întrebare

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

- A. AR(1), o diferențiere, MA(1) la nivelul obișnuit
- 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)

AR sezonier(1), o diferențiere sezonieră, MA sezonier(1)

Descompunerea notației SARIMA

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

(p, d, q)	Non-sezonier: AR(p), d diferențe, MA(q)
$(P, D, Q)_s$	Sezonier: SAR(P), D dif. sezoniere, SMA(Q)

Pentru $(1, 1, 1) \times (1, 1, 1)_{12}$:

- Non-sezonier: AR(1), o diferența obișnuită, MA(1)
- Sezonier: SAR(1) la lag-ul 12, un Δ_{12}, SMA(1) la lag-ul 12

Întrebarea 4

Întrebare

"Modelul Airline" este SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$. Câți parametri trebuie eștimati (excluzand varianța)?

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

Întrebarea 4: Răspuns

Răspuns corect: (B)

2 parametri

Structura modelului

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

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

Parametri:

- θ_1: coeficient MA non-sezonier
- Θ_1: coeficient MA sezonier

Total: **2 parametri** (plus σ^2)

De ce "Modelul Airline"?

Box & Jenkins (1970) au folosit acest model pentru a prognoză pasagerii companiilor aeriene internationale. Este remarcabil de eficient pentru multe serii economice sezoniere!

Întrebarea 5

Întrebare

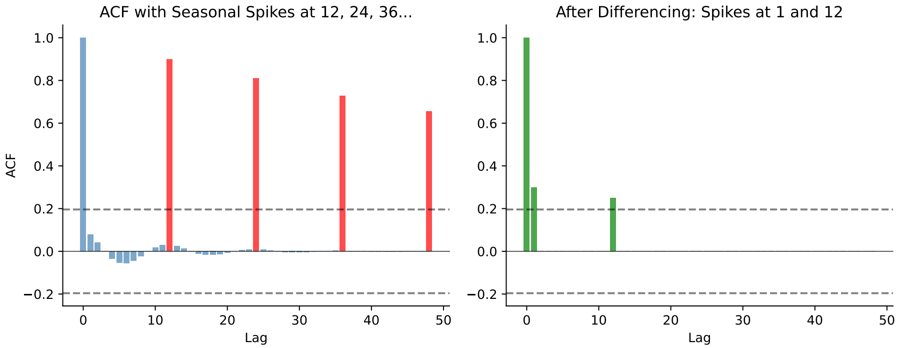
Observați vârfuri ACF semnificâtive la lag-urile 12, 24 și 36 intr-o serie lunară. Ce sugerează aceasta?

- A Seria are o rădăcină unitară
- B Seria are sezonalitate anuala 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

Vârfuri ACF la 12, 24, 36 = sezonalitate stochastică. Aplicați $(1 - L^{12})$ pentru a o elimină.



Întrebarea 6

Întrebare

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

- 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)

SARIMA$(0,1,1) \times (0,1,1)_{12}$ (Modelul Airline)

Reguli de identificare ACF/PACF

Pentru procese MA, ACF **se oprește brusc** după lag-ul q:

Tipar	**Sugerează**
Vârf ACF doar la lag-ul 1	MA(1) pentru partea non-sezonieră
Vârf ACF doar la lag-ul 12	SMA(1) pentru partea sezonieră

Combinat: MA(1) $\times$ SMA(1) $= (0, d, 1) \times (0, D, 1)_{12}$
Cu $d = 1$ și $D = 1$ (deja diferențiată): $(0, 1, 1) \times (0, 1, 1)_{12}$

Referinte

- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*. 5th ed. Wiley.
- Hyndman, R.J. & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3rd ed. OTexts.
- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
- Brockwell, P.J. & Davis, R.A. (2016). *Introduction to Time Series and Forecasting*. 3rd ed. Springer.