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 putea:

- Identifica tiparele sezoniere în datele de tip serie de timp
- Aplica diferențierea sezonieră pentru a elimina rădăcinile unitare sezoniere
- Construi și estima modele SARIMA cu componente sezoniere
- Produce prognoze precise pentru serii temporale sezoniere

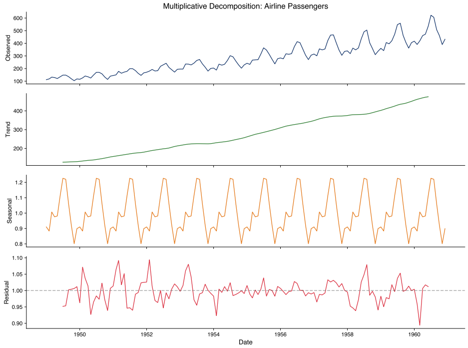
Structura cursului

De ce SARIMA? 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

TSA_ch4_motivation_seasonal

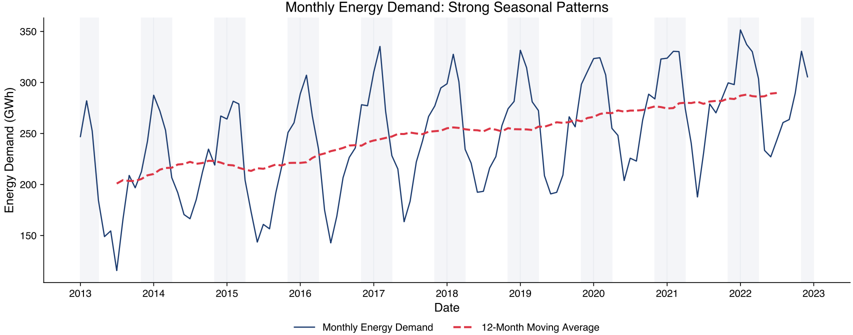
Înțelegerea componentelor sezoniere



- 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

TSA_ch4_motivation_decomposition

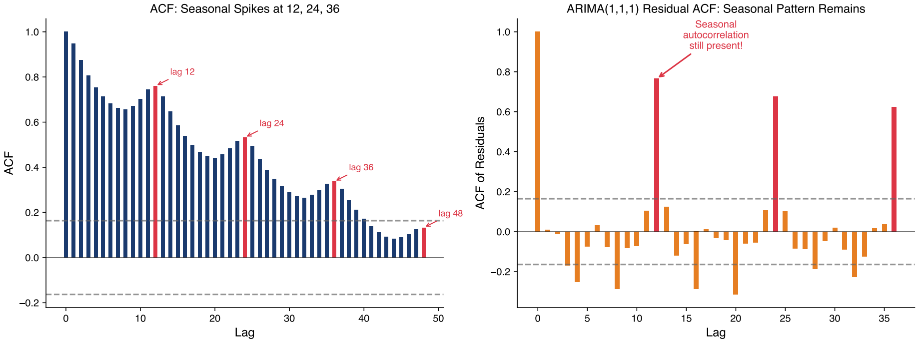
Aplicație reală: Tipare lunare



- 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

TSA_ch4_motivation_monthly

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

TSA_ch4_motivation_why_sarima

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

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

Ideea cheie

- SARIMA = ARIMA aplicată la **două frecvențe**: non-sezonieră (termen scurt) și sezonieră (termen lung)

Ce este sezonalitatea?

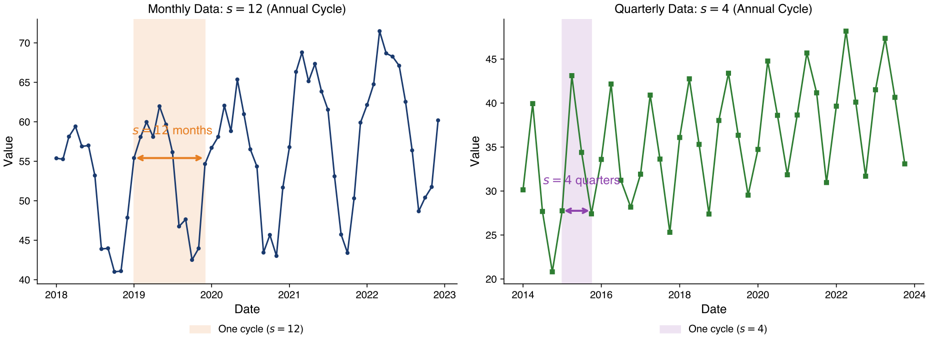
Definiție 1 (Sezonalitate)

- 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)

Sezonalitatea: Ilustrare vizuală



- **Stânga**: Date lunare cu $s = 12$ (ciclu anual); **Dreapta**: Date trimestriale cu $s = 4$
- Tiparul se repetă la fiecare s perioade $\succ$ această regularitate este exploatată de SARIMA

TSA_ch4_def_seasonality

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)

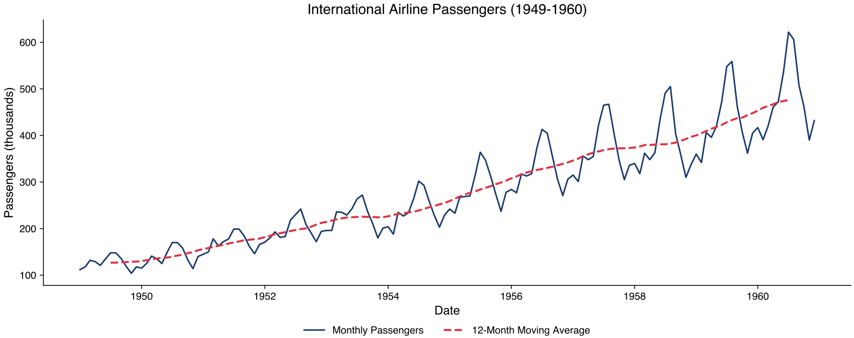
Alte domenii

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

De ce contează

- Ignorarea sezonalității duce la prognoze distorsionate și inferența invalidă!

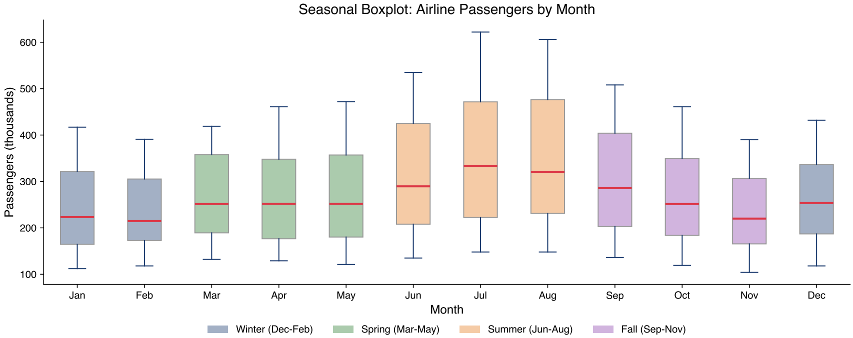
Exemplu: Datele privind pasagerii companiilor aeriene



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

TSA_ch4_airline_data

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)

TSA_ch4_seasonal_boxplot

Sezonalitate deterministă vs stochastică

Sezonalitate deterministă

- **Tipar sezonier fix**: $Y_t = \sum_{j=1}^{s} \gamma_j D_{jt} + \varepsilon_t$
  - D_{jt} sunt variabile dummy sezoniere
- **Tiparul** constant în timp
- **Eliminare**: prin regresie

Sezonalitate stochastică

- **Tipar sezonier în evoluție**: $\Delta_s Y_t = Y_t - Y_{t-s}$
  - Prezintă structura de dependență
- **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 ($s, 2s, 3s, \ldots$)

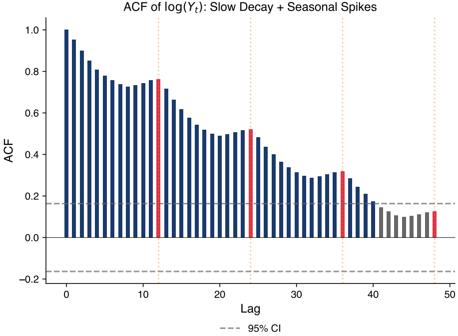
Teste statistice

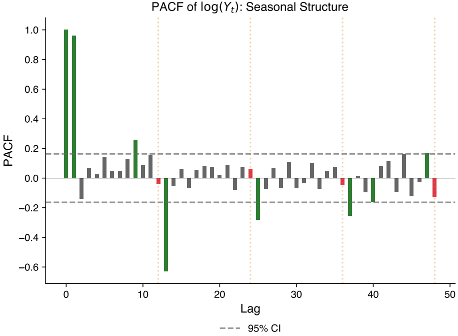
- Teste de rădăcină unitară sezonieră (HEGY, CH, OCSB)
- Testul F pentru variabile dummy sezoniere
- Testul Kruskal-Wallis (neparametric)

Semnatura ACF

- Sezonalitate puternică: ACF prezintă vârfuri semnificative 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: descreștere lentă $\succ$ necesită diferențiere sezonieră

TSA_ch4_acf_seasonality

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 $\neq \ldots \neq$ media din decembrie $\succ$ sezonalitate
- **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

Ideea cheie

- Dacă adăugarea variabilelor dummy sezoniere **reduce semnificativ** erorile de predicție, atunci sezonalitatea este 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)

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

Testul Kruskal-Wallis: intuiție

Ce face acest test?

- **Test neparametric**: verifică dacă observațiile din diferite sezoane provin din aceeași distribuție
- **Mecanism**: ordonează toate observațiile de la cea mai mică la cea mai mare
- **Verificare**: dacă rangurile sunt distribuite uniform între sezoane
- **Concluzie**: dacă un sezon are constant ranguri mai mari/mici $\succ$ 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 puțin puternic decât 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: $R_4 = 90$
- $H = \frac{12}{20 \times 21} \left(\frac{15^2}{5} + \frac{35^2}{5} + \frac{70^2}{5} + \frac{90^2}{5} \right) - 3(21) = 19.6$
- Valoarea critică $\chi^2_{0.05,3} = 7.81$. Deoarece $19.6 > 7.81$: **Respingem** $H_0 \succ$ Sezonalitate!

În Python

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

Testul HEGY: ce problemă rezolvă?

Întrebarea cheie

- **Problemă**: având o serie sezonieră, trebuie să știm tipul de diferențiere
- **Diferențiere obișnuită** $(1 - L)$? $\succ$ setăm $d = 1$; **Diferențiere sezonieră** $(1 - L^s)$? $\succ$ setăm $D = 1$
- **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
- **Limitare**: 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 auxiliară HEGY

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

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

Ipoteze

- $H_0 : \pi_1 = 0$: rădăcină unitară la frecvența 0
- $H_0 : \pi_2 = 0$: rădăcină unitară la frecvența π
- $H_0 : \pi_3 = \pi_4 = 0$: rădăcină unitară la frecvența $\pm\pi/2$

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

- **Rezultate HEGY**: $t_1 = -1.52$, $t_2 = -4.21$, $F_{34} = 2.15$
- $t_1 = -1.52 > -2.88$: Nu putem respinge $\succ$ **necesită** $d = 1$
- $t_2 = -4.21 < -2.88$: Respingem $\succ$ fără rădăcină unitară la π
- $F_{34} = 2.15 < 6.57$: Nu putem respinge $\succ$ **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 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 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 $\hat{\omega}_i$ = estimator consistent al densității spectrale la frecvența λ_i

Decizie

- **Regula**: respingem H_0 (staționaritate) dacă $LM >$ valoare critică $\succ$ 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 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

Ideea cheie

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

Operatorul de diferență sezonieră

Definiție 2 (Diferență sezonieră)

- **Operatorul de diferență sezonieră** Δ_s este definit ca:

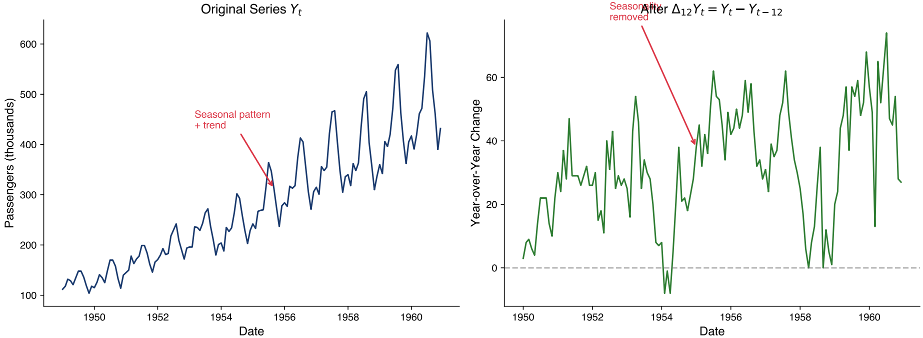
$$\Delta_s Y_t = (1 - L^s) Y_t = Y_t - Y_{t-s}$$

- 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

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

TSA_ch4_def_seasonal_diff

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:
$$\Delta_s Y_t = (\mu_t + \varepsilon_t) - (\mu_{t-s} + \varepsilon_{t-s}) = 0 + \varepsilon_t - \varepsilon_{t-s} \quad \text{(deoarece } \mu_t = \mu_{t-s}\text{)}$$

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

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

Combinarea diferențierii obișnuite și sezoniere

Diferențiere completă

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

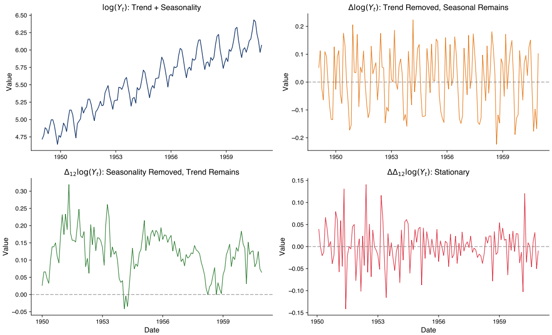
Dezvoltare

- **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)

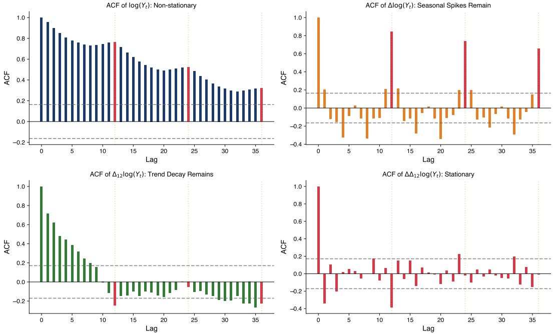
Efectul operațiilor de diferențiere



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

TSA_ch4_differencing_effect

ACF înainte și după diferențiere



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

TSA_ch4_acf_differencing

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

Definiția modelului SARIMA

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

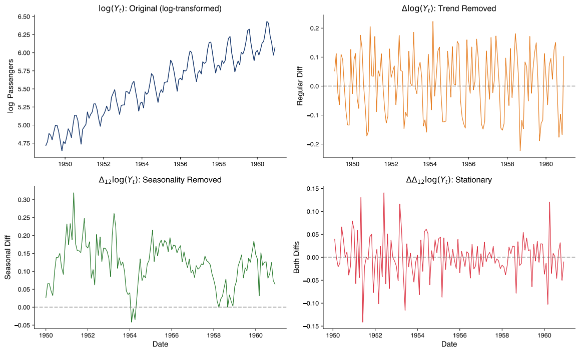
- Modelul **Seasonal ARIMA** este:

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

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 vizuală



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

TSA_ch4_def_sarima

Demonstrație: structura multiplicativă sezonieră

De ce multiplicativă?

- Considerăm 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

Parsimonie

- **Multiplicativă**: 2 parametri (ϕ,Φ); **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), și atât diferențiere obișnuită cât și sezonieră

Modele SARIMA comune

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

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

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

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

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

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

Structura multiplicativă

De ce multiplicativă?

- **Principiu**: părțile sezonieră și non-sezonieră se înmulțesc
- **AR**: $\phi(L)\Phi(L^s)$ **MA**: $\theta(L)\Theta(L^s)$

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

- **Model**: $(1-\phi L)(1-\Phi L^{12})Y_t = \varepsilon_t$
- **Dezvoltare**: $Y_t - \phi Y_{t-1} - \Phi Y_{t-12} + \phi\Phi Y_{t-13} = \varepsilon_t$
- **Termenul încrucișat** $\phi\Phi Y_{t-13}$ captează interacțiunea!

Interpretare

- **Avantaj**: structura multiplicativă permite modelarea parsimonioasă a tiparelor sezoniere complexe cu puțini parametri

ACF/PACF pentru modele sezoniere

Ideea cheie

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

Tipare ACF/PACF sezoniere

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

Exemplu: ACF/PACF pentru modelul Airline

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

- **Diferențiere**: $W_t = (1 - L)(1 - L^{12})Y_t$
- **Model**: $W_t = (1 + \theta L)(1 + \Theta L^{12})\varepsilon_t$

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 interacțiunea $\theta \cdot \Theta$)
- 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

- **Pas 1**: examinați ACF pentru descreștere lentă la lag-uri sezoniere $\succ$ diferențiere sezonieră
- **Pas 2**: după diferențiere, verificați tiparele ACF/PACF
- **Pas 3**: comportamentul non-sezonier la lag-urile $1, 2, \ldots, s-1$
- **Pas 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 informaționale (AIC, BIC) pentru selecția finală
- Algoritmii Auto-SARIMA pot ajuta

Metode de estimare

Estimare prin verosimilitate maximă

- **Abordare standard** pentru SARIMA:
  - MLE condiționată (condiționată de valorile inițiale)
  - MLE exactă (prin filtrul Kalman)

Considerații computationale

- Mai mulți parametri decât ARIMA $\succ$ mai multe date necesare
- Parametrii sezonieri estimați din lag-urile $s, 2s, \ldots$
- 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 state-space
- **Filtrul Kalman**: calculează recursiv erorile de predicție v_t și variantele lor f_t, fără a condiționa pe valori inițiale

Log-verosimilitatea exactă (prediction error decomposition)

- **Formula**: $\ell(\boldsymbol{\theta}) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left[\ln(f_t) + \frac{v_t^2}{f_t}\right]$
- 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 unde valorile inițiale contează
- Implementat implicit în `statsmodels.tsa.SARIMAX()` cu `method='mle'`

Staționaritate și invertibilitate

Condiții de staționaritate

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

Condiții de invertibilitate

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

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

Prognoze punctuale

Calculul prognozei

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

Tiparul sezonier în prognoze

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

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 variația de trend cât și cea sezonieră

Prognoze pe orizont lung

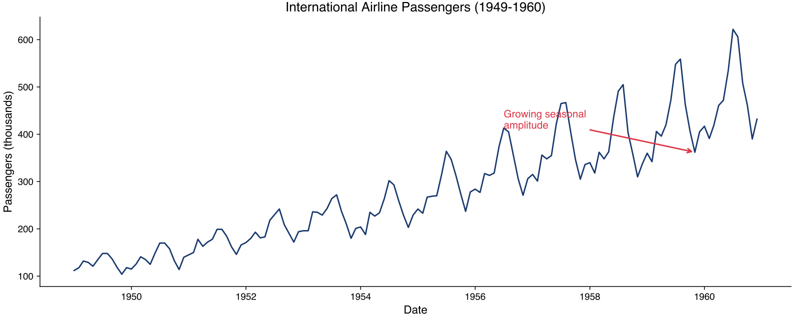
Comportamentul când $h \to \infty$

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

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

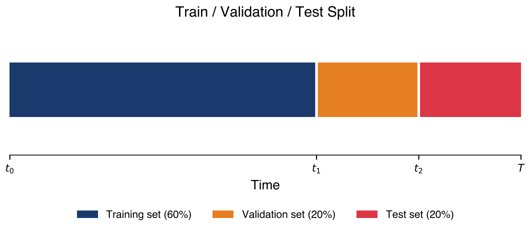
Studiu de caz: Date despre pasagerii aerieni



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

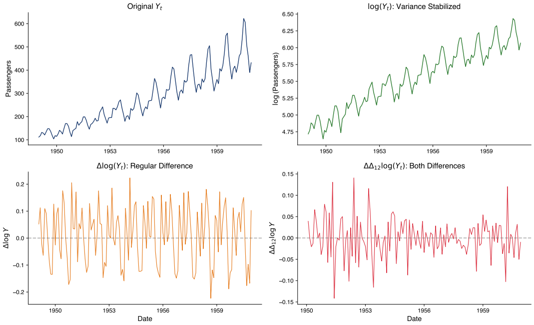
TSA_ch4_case_raw_data

Strategia de împărțire a datelor



- **Set de antrenare (70%)** — Estimarea parametrilor modelului
  - Estimare coeficienți SARIMA (ϕ, θ, Φ, Θ)
  - Porțiunea cea mai mare asigură estimări fiabile ale parametrilor
- **Set de validare (15%)** — Selectarea celui mai bun model
  - Comparare modele candidate (ordine diferite)
  - Alegere model cu cea mai mică eroare de validare
- **Set de test (15%)** — Evaluare finală
  - Performanță out-of-sample imparțială; niciodată folosit în dezvoltare

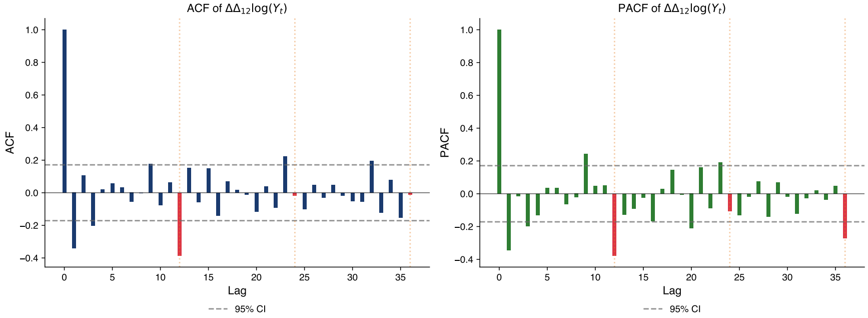
Pasul 1: Transformări



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

TSA_ch4_case_transformations

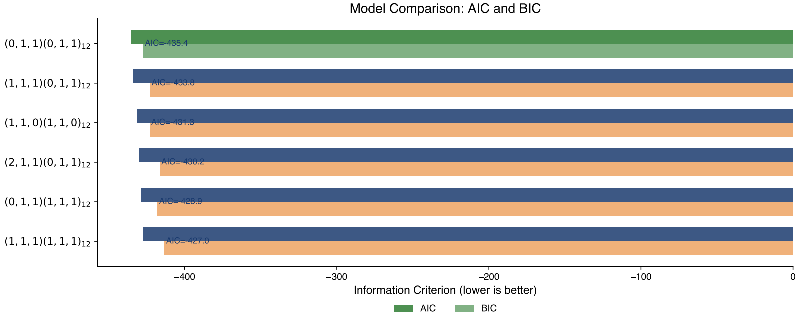
Pasul 2: Analiza ACF/PACF



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

TSA_ch4_case_acf_pacf

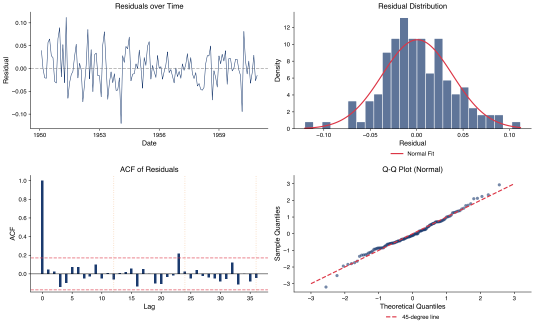
Pasul 3: Compararea modelelor



- Comparăm modelele SARIMA candidate 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

TSA_ch4_case_model_comparison

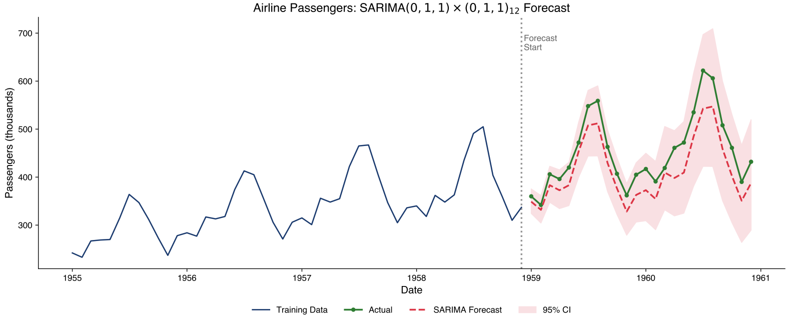
Pasul 4: Diagnosticul reziduurilor



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

TSA_ch4_case_diagnostics

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

TSA_ch4_case_forecast

Concluzii cheie

Puncte principale

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

Pașii următorii

- **Capitolul 5**: va acoperi modelarea volatilității $\succ$ modele ARCH, GARCH și extensii asimetrice

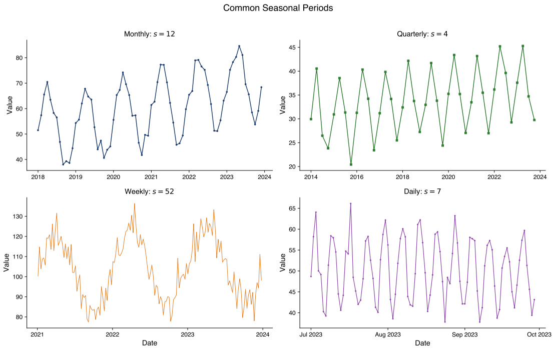
Întrebarea 1

Întrebare

- Pentru date lunare cu sezonalitate anuală, care este perioada 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

TSA_ch4_quiz1_seasonal_periods

Întrebarea 2

Întrebare

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

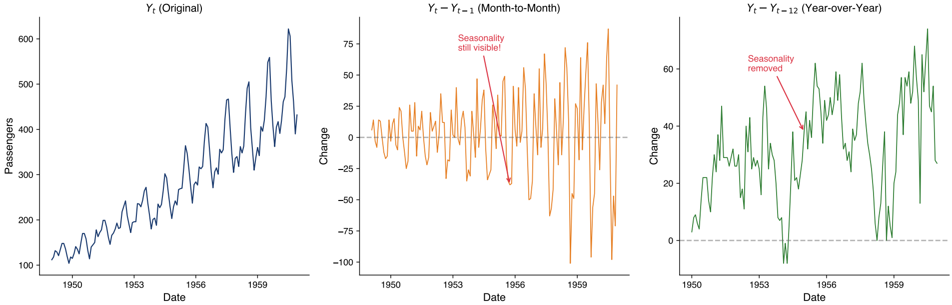
(A) Calculează $Y_t - Y_{t-1}$ (schimbarea luna-la-luna)

(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

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

TSA_ch4_quiz2_seasonal_diff

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) 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 $\succ$ AR(p), d diferențe, MA(q)
- $(P, D, Q)_s$: Sezonier $\succ$ SAR(P), D dif. sezoniere, SMA(Q)
- **Non-sezonier** $(1, 1, 1)$: AR(1), o diferență obișnuită, MA(1)
- **Sezonier** $(1, 1, 1)_{12}$: 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 estimați (excluzând varianța)?

(A) 1

(B) 2

(C) 4

(D) 12

Întrebarea 4: Răspuns

Răspuns corect: (B)

- **Răspuns**: 2 parametri

Structura modelului

- **Model**: $(1-L)(1-L^{12})Y_t = (1+\theta_1 L)(1+\Theta_1 L^{12})\varepsilon_t$
- θ_1: coeficient MA non-sezonier
- Θ_1: coeficient MA sezonier
- **Total**: 2 parametri (plus σ^2)

De ce "Modelul Airline"?

- **Origine**: Box & Jenkins (1970) au folosit acest model pentru a prognoza pasagerii companiilor aeriene internaționale
- **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?

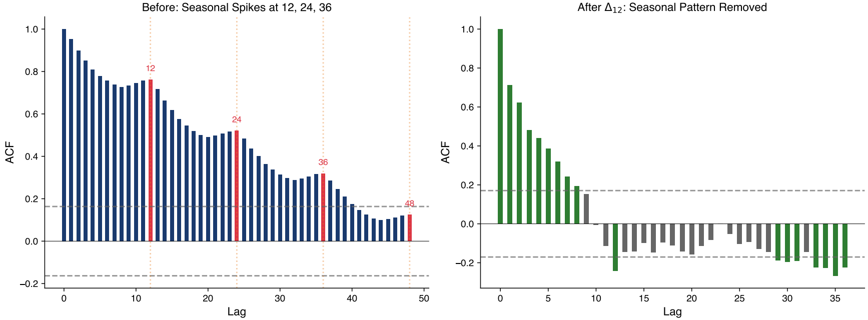
(A) Seria are o rădăcină unitară

(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 = sezonalitate stochastică
- **Soluție**: aplicați $(1 - L^{12})$ pentru a o elimina

TSA_ch4_quiz5_seasonal_acf

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?

(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**: 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 statistică
- **Quantinar**: `https://quantinar.com` ≻ Platformă de învățare metode cantitative
- **GitHub TSA**: `https://github.com/QuantLet/TSA/tree/main/TSA_ch4` ≻ Cod Python pentru acest curs

Vă Mulțumim!

Întrebări?

Graficele au fost generate folosind Python (statsmodels, matplotlib)

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

Quantlet Quantinar