Analiza și Prognoza Seriilor de Timp

Capitolul 4: Modele SARIMA

Serii de Timp Sezoniere

Cuprins

Exemplu motivational: Sezonalitatea este peste tot

charts/ch4_motivation_seasonal.pdf

- Vanzarile cu amanuntul prezinta **tipare anuale** clare: varfuri in decembrie, minime in ianuarie
- Modelele ARIMA standard nu pot captura aceste **cicluri sezoniere repetitive**
- Ignorarea sezonalitatii duce la erori sistematice de prognoza

Intelegerea componentelor sezoniere

charts/ch4_motivation_decomposition.pdf

- Serie de timp sezoniera = **Trend** + **Tipar sezonier** + **Reziduuri**
- Descompunerea ajuta la vizualizarea separata a fiecarei componente
- Modelele SARIMA capteaza atat dinamica trendului, cat si comportamentul sezonier

Aplicatie reala: Tipare lunare

charts/ch4_motivation_monthly.pdf

- Cererea de energie prezinta o **sezonalitate lunara** puternica (cicluri de incalzire/racire)
- Tiparul se repeta previzibil in fiecare an cu mici variatii
- Companiile de utilitati folosesc prognozele SARIMA pentru planificarea capacitatii

De ce avem nevoie de SARIMA?

charts/ch4_motivation_why_sarima.pdf

- **Stanga**: ACF sezoniera prezinta varfuri la lag-urile 12, 24, 36... (tipar anual)
- **Dreapta**: Reziduurile ARIMA inca prezinta autocorelatie sezoniera — modelul este incomplet
- SARIMA adauga **termeni AR si MA sezonieri** pentru a captura aceste tipare

Ce vom invata astazi

Concepte

- Identificarea tiparelor sezoniere
- Operatorul de diferentiere sezoniera
- Notatia SARIMA$(p, d, q)(P, D, Q)_s$
- Celebrul "Model Airline"
- Selectia modelului pentru date sezoniere

Abilitati

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

Ideea cheie

SARIMA = ARIMA aplicat la **doua frecvente**: nivelul obisnuit (pe termen scurt) si cel sezonier (pe termen lung)

Ce este sezonalitatea?

Definitie 1 (Sezonalitate)

O serie de timp prezinta **sezonalitate** cand arata fluctuatii regulate, periodice care se repeta pe o perioada fixa s (perioada sezoniera).

Perioade sezoniere comune

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

Sezonalitatea: Ilustrare vizuala

charts/ch4_def_seasonality.pdf

Exemple de date sezoniere

Serii economice

- Vanzari cu amanuntul (varfuri 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 conteaza

Ignorarea sezonalitatii duce la prognoze distorsionate si inferenta invalida!

Exemplu: Datele privind pasagerii companiilor aeriene

charts/ch4_airline_data.pdf

- Pasageri internationali lunari ai companiilor aeriene (1949–1960)
- **Trend ascendent** clar si **amplitudine sezoniera crescatoare**
- Varfurile din vara reflecta tiparele calatoriilor de vacanta

Vizualizarea tiparelor sezoniere

charts/ch4_seasonal_boxplot.pdf

- Diagrama box plot releva un tipar sezonier consistent de-a lungul anilor
- Iulie–August prezinta cele mai mari numere de pasageri (calatorii de vara)
- Noiembrie–Februarie prezinta cele mai mici numere (lunile de iarna)

Sezonalitate determinista vs stochastica

Sezonalitate determinista

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

Proprietati:

- Tiparul constant in timp
- Eliminat prin regresie

Sezonalitate stochastica

Tipar sezonier in evolutie: $\Delta_s Y_t = Y_t - Y_{t-s}$ prezinta structura de dependenta.

Proprietati:

- Tiparul evolueaza in timp
- Necesita diferentiere sezoniera

Detectarea sezonalitatii

Metode vizuale

- Graficul seriei de timp – cautati tipare repetitive
- Graficul sub-seriilor sezoniere – comparati aceleasi sezoane de-a lungul anilor
- Graficul ACF – varfuri la lag-uri sezoniere $(s, 2s, 3s, \ldots)$

Teste statistice

- Teste de radacina unitara sezoniera (HEGY, CH, OCSB)
- Testul F pentru variabile dummy sezoniere
- Testul Kruskal-Wallis (neparametric)

Semnatura ACF

Sezonalitate puternica: ACF prezinta varfuri semnificative la lag-urile $s, 2s, 3s, \ldots$

ACF releva structura sezoniera

charts/ch4_acf_seasonality.pdf

- **Descrestere lenta** la toate lag-urile indica nestationaritate (trend)
- **Varfuri la lag-urile 12, 24, 36** confirma tiparul sezonier ($s = 12$)
- ACF la lag-urile sezoniere prezinta descrestere lenta $\Rightarrow$ necesita diferentiere sezoniera

Testul F pentru variabile dummy sezoniere: Intuitie

Ce face acest test?

Testeaza daca **valorile medii difera semnificativ intre sezoane**.

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

Modelele comparate

Restrictionat: $Y_t = \alpha + \varepsilon_t$ **Nerestictionat**: $Y_t = \alpha + \sum_{j=1}^{s-1} \gamma_j D_{jt} + \varepsilon_t$
unde $D_{jt} = 1$ daca observatia t este in sezonul j, 0 altfel.

Ideea cheie

Daca adaugarea variabilelor dummy sezoniere **reduce semnificativ** erorile de predictie, atunci sezonalitatea este prezenta.

Testul F pentru variabile dummy sezoniere: Formula si 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 patratelor reziduurilor din modelul restrictionat (fara dummy)
- SSR_U = Suma patratelor reziduurilor din modelul nerestrictionat (cu dummy)
- $s-1$ = numarul de restrictii (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 critica $F_{0.05,11,108} \approx 1.87$. Deoarece $7.51 > 1.87$: **Respingem** $H_0 \Rightarrow$ Sezonalitate prezenta!

Testul Kruskal-Wallis: Intuitie

Ce face acest test?

Un test **neparametric** care verifica daca observatiile din diferite sezoane provin din aceeasi distributie.

- Functioneaza prin **ordonarea** tuturor observatiilor de la cea mai mica la cea mai mare
- Verifica daca rangurile sunt distribuite uniform intre sezoane
- Daca un sezon are in mod constant ranguri mai mari/mici $\Rightarrow$ sezonalitate

De ce sa-l folosim in locul testului F?

- **Fara ipoteza de normalitate** – functioneaza cu orice distributie
- **Robust la valori extreme** – valorile extreme nu distorsioneaza rezultatele

Limitare

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

Testul Kruskal-Wallis: Formula si exemplu

Statistica 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. in sezonul j, $R_j =$ suma rangurilor.

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

In Python

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

Testul HEGY: Ce problema rezolva?

Intrebarea cheie

Avand o serie de timp sezoniera, trebuie sa stim:

1. Are nevoie de **diferentiere obisnuita** $(1 - L)$? $\Rightarrow$ setam $d = 1$
2. Are nevoie de **diferentiere sezoniera** $(1 - L^s)$? $\Rightarrow$ setam $D = 1$

HEGY testeaza pentru **ambele** tipuri de radacini unitare simultan!

De ce sa nu folosim doar ADF?

ADF testeaza doar pentru o radacina unitara **obisnuita** la frecventa zero. Datele sezoniere pot avea radacini unitare la **frecvente sezoniere** pe care ADF le omite!

HEGY testeaza frecvente multiple

Trimestrial: testeaza la 0, π, $\pm\pi/2$. Lunar: testeaza la 0, π, $\pm\pi/6$, $\pm\pi/3$, $\pm\pi/2$, $\pm 2\pi/3$, $\pm 5\pi/6$.

Testul HEGY: Formula de regresie (Trimestrial)

Regresia auxiliara HEGY

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

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

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	Statistica	Valoare critica	Daca NU este respins...
t_1 $(\pi_1 = 0)$	t-stat	-2.88	Necesita $d = 1$
t_2 $(\pi_2 = 0)$	t-stat	-2.88	Necesita $D = 1$
F_{34} $(\pi_3 = \pi_4 = 0)$	F-stat	6.57	Necesita $D = 1$

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 $\Rightarrow$ **necesita** $d = 1$
- $t_2 = -4.21 < -2.88$: Respingem $\Rightarrow$ fara radacina unitara la π
- $F_{34} = 2.15 < 6.57$: Nu putem respinge $\Rightarrow$ **necesita** $D = 1$

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	Radacina unitara sezoniera	**Fara** radacina unitara sezoniera
H_1	Fara radacina unitara sezoniera	Radacina unitara sezoniera
Respingem H_0	Folosim variabile dummy sezoniere	Folosim diferentiere $(1 - L^s)$
Nu respingem	Folosim diferentiere $(1 - L^s)$	Folosim variabile dummy sezoniere

De ce conteaza?

- HEGY: "Demonstrati ca NU exista radacina unitara" (conservator fata de diferentiere)
- CH: "Demonstrati ca EXISTA radacina unitara" (conservator fata de variabile dummy)
- Folositi **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 frecventa sezoniera λ_i: $S_{it}^{(c)} = \sum_{j=1}^{t} \hat{u}_j \cos(\lambda_i j)$, $\quad S_{it}^{(s)} = \sum_{j=1}^{t} \hat{u}_j \sin(\lambda_i j)$

Statistica 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 densitatii spectrale la frecventa λ_i.

Decizie

Respingem H_0 (stationaritate) daca $LM >$ valoare critica $\Rightarrow$ este necesara diferentierea sezoniera.

Sumar: Alegerea testului de sezonalitate potrivit

Test	H_0	Daca respingem	Cel mai bun pentru
Test F	Fara sezonalitate	Sezonalitate exista	Date normale
Kruskal-Wallis	Fara dif. sezoniera	Sezonalitate exista	Non-normale, valori extreme
HEGY	Radacina unitara exista	Folosim dummy	Determinarea d, D
Canova-Hansen	Fara radacina unitara	Folosim $(1 - L^s)$	Confirmarea stabilitatii

Ideea cheie

Test F/Kruskal-Wallis: "*Exista sezonalitate?*"
HEGY/Canova-Hansen: "*Ce tip?*" (determinista vs stochastica)

Operatorul de diferenta sezoniera

Definitie 2 (Diferenta sezoniera)

Operatorul de diferenta sezoniera 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 ($s = 12$): $\Delta_{12} Y_t = Y_t - Y_{t-12}$
  Compara fiecare luna cu aceeasi luna din anul trecut
- Date trimestriale ($s = 4$): $\Delta_4 Y_t = Y_t - Y_{t-4}$
  Compara fiecare trimestru cu acelasi trimestru din anul trecut

Diferenta sezoniera: Ilustrare vizuala

charts/ch4_def_seasonal_diff.pdf

Combinarea diferentierii obisnuite si sezoniere

Diferentiere completa

Pentru serii cu atat trend cat si 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 diferentierii

- d: numarul de diferente obisnuite (eliminarea trendului)
- D: numarul de diferente sezoniere (eliminarea trendului sezonier)

Efectul operatiilor de diferentiere

charts/ch4_differencing_effect.pdf

- Diferentierea obisnuita elimina trendul dar tiparul sezonier ramane
- Diferentierea sezoniera elimina sezonalitatea dar tiparul de trend ramane
- **Ambele diferente** sunt necesare pentru a atinge stationaritatea

ACF inainte si dupa diferentiere

charts/ch4_acf_differencing.pdf

- ACF originala: descrestere lenta indica nestationaritate
- Dupa Δ: varfuri sezoniere raman la lag-urile 12, 24, 36
- Dupa Δ_{12}: descresterea de trend ramane la lag-urile initiale
- Dupa $\Delta\Delta_{12}$: ACF se opreste brusc $\Rightarrow$ **stationara**

Integrare sezoniera

Definitie 3 (Proces integrat sezonier)

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

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

este stationara.

Cazuri comune

- $I(1, 0)_{12}$: Doar radacina unitara obisnuita (lunara)
- $I(0, 1)_{12}$: Doar radacina unitara sezoniera
- $I(1, 1)_{12}$: Atat radacina unitara obisnuita cat si sezoniera

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
- $(1-L)^d$: Diferentiere obisnuita; $(1-L^s)^D$: Diferentiere sezoniera

SARIMA: Ilustrare vizuala

charts/ch4_def_sarima.pdf

Notatia SARIMA

Specificatie completa

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

Parametru	**Semnificatie**
p	Ordinul AR non-sezonier
d	Ordinul diferentierii non-sezoniere
q	Ordinul MA non-sezonier
P	Ordinul AR sezonier
D	Ordinul diferentierii sezoniere
Q	Ordinul MA sezonier
s	Perioada sezoniera

Exemplu

SARIMA$(1, 1, 1) \times (1, 1, 1)_{12}$: Date lunare cu AR(1), MA(1), AR sezonier(1), MA sezonier(1), si atat diferentiere obisnuita cat si sezoniera.

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 si 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. sezoniera + MA(1)

Structura multiplicativa

De ce multiplicativa?

Partile sezoniera si non-sezoniera se **inmultesc**:

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

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 $\phi\Phi Y_{t-13}$ capteaza interactiunea!

Interpretare

Structura multiplicativa permite modelarea parsimoniasa a tiparelor sezoniere complexe cu putini parametri.

ACF/PACF pentru modele sezoniere

Ideea cheie

Modelele sezoniere prezinta 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 opreste dupa Ps
SMA(Q)	Se opreste dupa Qs	Descreste la $s, 2s, \ldots$
SARMA	Descreste la lag-uri sezoniere	Descreste la lag-uri sezoniere

Exemplu: ACF/PACF pentru modelul Airline

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

Dupa diferentiere $W_t = (1-L)(1-L^{12})Y_t$:

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

Tiparul ACF asteptat

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

Tiparul PACF asteptat

- Descrestere exponentiala la lag-urile $1, 2, 3, \ldots$
- Descrestere exponentiala la lag-urile $12, 24, 36, \ldots$

Ghid de identificare a modelului

Proces pas cu pas

1. Examinati ACF pentru descrestere lenta la lag-uri sezoniere $\Rightarrow$ diferentiere sezoniera
2. Dupa diferentiere, verificati 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

- Incepeti cu $d \leq 1$ si $D \leq 1$
- De obicei $P, Q \leq 2$ este suficient
- Folositi criterii informationale (AIC, BIC) pentru selectia finala
- Algoritmii Auto-SARIMA pot ajuta

Metode de estimare

Estimare prin verosimilitate maxima

Abordare standard pentru SARIMA:

- MLE conditionata (conditionata de valorile initiale)
- MLE exacta (prin filtrul Kalman)

Consideratii computationale

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

Stationaritate si invertibilitate

Conditii de stationaritate

Atat polinoamele AR non-sezoniere cat si sezoniere trebuie sa aiba radacini in afara cercului unitate:

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

Conditii de invertibilitate

Atat polinoamele MA non-sezoniere cat si sezoniere trebuie sa aiba radacini in afara cercului unitate:

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

Verificarea diagnosticarii

Analiza reziduurilor

Dupa ajustarea SARIMA, verificati ca reziduurile sunt zgomot alb:

1. Graficul reziduurilor in timp (fara tipare)
2. ACF a reziduurilor (fara varfuri semnificative)
3. Testul Ljung-Box la lag-uri multiple inclusiv sezoniere
4. Teste de normalitate (grafic Q-Q, Jarque-Bera)

Important

Verificati ACF la **ambele** lag-uri non-sezoniere si sezoniere!
ACF semnificativa la lag-ul 12 sugereaza modelare sezoniera inadecvata.

Criterii de selectie a modelului

Criterii informationale

Comparati 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 varianta).

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:

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

Tiparul sezonier in prognoze

Prognozele SARIMA capteaza in mod natural sezonalitatea:

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

Intervale de prognoza

Cuantificarea incertitudinii

Interval de predictie $(1-\alpha)\%$:

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

Varianta calculata din reprezentarea MA(∞).

Proprietati cheie

- Intervalele se largesc cu orizontul de prognoza
- Pentru serii $I(1,1)_s$: intervalele cresc nelimitat
- Tiparul sezonier vizibil in prognozele punctuale
- Incertitudinea capteaza atat variatia de trend cat si cea sezoniera

Prognoze pe orizont lung

Comportamentul cand $h \to \infty$

- Prognozele punctuale converg la tiparul sezonier determinist
- Daca exista deriva: trend linear + tipar sezonier
- Intervalele de prognoza continua sa se largeasca

Implicatie practica

- Pe termen scurt: SARIMA capteaza atat nivelul cat si sezonul
- Pe termen mediu: Prognoze sezoniere bune, incertitudine crescatoare
- Pe termen lung: Reflecta in principal tiparul sezonier, intervale largi

Datele privind pasagerii companiilor aeriene

charts/ch4_airline_data.pdf

- Set de date clasic: Pasageri internationali lunari ai companiilor aeriene (1949-1960)
- Trend ascendent clar si amplitudine sezoniera crescatoare

Descompunerea sezoniera

charts/ch4_decomposition.pdf

- Trend: Crestere puternica ascendenta
- Sezonalitate: Varfuri de vara (calatorii de vacanta)
- Rezidual: Variatie aleatoare dupa eliminarea trendului si sezonului

Analiza ACF/PACF

charts/ch4_acf_pacf.pdf

- Dupa diferentierea $\Delta\Delta_{12}$: varfuri la lag-urile 1 si 12
- Sugereaza SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ (Modelul Airline)

Rezultatele prognozei SARIMA

charts/ch4_sarima_forecast.pdf

- SARIMA capteaza atat trendul cat si tiparul sezonier
- Prognozele prezinta varfuri si minime sezoniere corespunzatoare

Diagnosticarea modelului

charts/ch4_diagnostics.pdf

- Reziduurile par aleatorii; ACF in limite la toate lag-urile
- Modelul capteaza adecvat structura sezoniera

Implementare in Python

Ajustarea SARIMA in 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 in Python cu comentarii sunt furnizate in caietele Jupyter.

Concluzii cheie

Puncte principale

1. **Sezonalitatea** este comuna in datele economice si de afaceri
2. **Diferentierea sezoniera** $(1 - L^s)$ elimina sezonalitatea stochastica
3. **SARIMA**$(p, d, q) \times (P, D, Q)_s$ extinde ARIMA pentru date sezoniere
4. **Structura multiplicativa** capteaza interactiunile sezon-trend
5. **ACF/PACF** prezinta tipare la ambele lag-uri obisnuite si sezoniere
6. **Selectia modelului**: Folositi AIC/BIC sau algoritmi auto-SARIMA

Pasii urmatori

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

Intrebarea 1

Intrebare

Pentru date lunare cu sezonalitate anuala, care este perioada sezoniera s?

- A. $s = 4$
- B. $s = 7$
- C. $s = 12$
- D. $s = 52$

Intrebarea 1: Raspuns

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

Perioade comune: Trimestrial=4, Lunar=12, Saptamanal=52, Zilnic=7, Orar=24

charts/ch4_quiz1_seasonal_periods.pdf

Intrebarea 2

Intrebare

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

- A. Calculeaza $Y_t - Y_{t-1}$ (schimbarea luna-la-luna)
- B. Calculeaza $Y_t - Y_{t-12}$ (schimbarea an-la-an)
- C. Calculeaza media mobila pe 12 luni
- D. Elimina doar componenta de trend

Intrebarea 2: Raspuns

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

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

charts/ch4_quiz2_seasonal_diff.pdf

Intrebarea 3

Intrebare

In notatia SARIMA$(1,1,1) \times (1,1,1)_{12}$, ce reprezinta partea $(1,1,1)_{12}$?

- A. AR(1), o diferentiere, MA(1) la nivelul obisnuit
- B. AR sezonier(1), o diferentiere sezoniera, MA sezonier(1)
- C. 12 termeni AR, 12 diferente, 12 termeni MA
- D. Modelul are 12 parametri in total

Intrebarea 3: Raspuns

Raspuns corect: (B)

AR sezonier(1), o diferentiere sezoniera, MA sezonier(1)

Descompunerea notatiei SARIMA

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

(p, d, q)	Non-sezonier: AR(p), d diferente, 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 diferenta obisnuita, MA(1)
- Sezonier: SAR(1) la lag-ul 12, un Δ_{12}, SMA(1) la lag-ul 12

Intrebarea 4

Intrebare

"Modelul Airline" este SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$. Cati parametri trebuie estimati (excluzand varianta)?

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

Intrebarea 4: Raspuns

Raspuns 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 prognoza pasagerii companiilor aeriene internationale. Este remarcabil de eficient pentru multe serii economice sezoniere!

Intrebarea 5

Intrebare

Observati varfuri ACF semnificative la lag-urile 12, 24 si 36 intr-o serie lunara. Ce sugereaza aceasta?

- A. Seria are o radacina unitara
- B. Seria are sezonalitate anuala care necesita diferentiere sezoniera
- C. Seria urmeaza un proces AR(36)
- D. Seria este deja stationara

Intrebarea 5: Raspuns

Raspuns corect: (B) Necesita diferentiere sezoniera

Varfuri ACF la 12, 24, 36 = sezonalitate stochastica. Aplicati $(1 - L^{12})$ pentru a o elimina.

charts/ch4_quiz5_seasonal_acf.pdf

Intrebarea 6

Intrebare

Dupa aplicarea $(1 - L)(1 - L^{12})$ unei serii lunare, ACF prezinta un varf semnificativ doar la lag-ul 1 si 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. $\text{SARIMA}(0, 1, 0) \times (0, 1, 0)_{12}$

Intrebarea 6: Raspuns

Raspuns corect: (B)

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

Reguli de identificare ACF/PACF

Pentru procese MA, ACF **se opreste brusc** dupa lag-ul q:

Tipar	**Sugereaza**
Varf ACF doar la lag-ul 1	MA(1) pentru partea non-sezoniera
Varf ACF doar la lag-ul 12	SMA(1) pentru partea sezoniera

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

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.