








Serii de timp

Capitolul 4: Sezonalitate și prognoză: SARIMA, TBATS, Prophet



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

Acronime folosite în acest material (1/3)

- **ACF**: Autocorrelation Function (funcția de autocorelație)
- **AI**: Artificial Intelligence (inteligență artificială)
- **AIC**: Akaike Information Criterion (criteriul informațional Akaike)
- **AR**: AutoRegressive (model) — model autoregresiv
- **ARCH**: AutoRegressive Conditional Heteroskedasticity (heteroscedasticitate condiționată autoregresivă)
- **ARIMA**: AutoRegressive Integrated Moving Average (model) — model autoregresiv integrat cu medie mobilă
- **ARMA**: AutoRegressive Moving Average (model) — model autoregresiv cu medie mobilă
- **BIC**: Bayesian Information Criterion (criteriul informațional bayesian al lui Schwarz)
- **CA**: Calendar Adjusted (ajustat pentru efectele de calendar)
- **DHR**: Dynamic Harmonic Regression (regresie armonică dinamică)
- **DM**: Diebold–Mariano (test) — testul Diebold–Mariano de comparare a prognozelor
- **ENTSO-E**: European Network of Transmission System Operators for Electricity (rețeaua europeană a operatorilor de transport și de sistem pentru energie electrică)
- **ETS**: Error, Trend, Seasonality (exponential smoothing state space models) — eroare, trend, sezonalitate (modele de netezire exponențială)

Acronime folosite în acest material (2/3)

- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity (heteroscedasticitate condiționată autoregresivă generalizată)
- **HEGY**: Hylleberg–Engle–Granger–Yoo (seasonal unit-root test) — testul Hylleberg–Engle–Granger–Yoo de rădăcină unitară sezonieră
- **HLN**: Harvey–Leybourne–Newbold (small-sample correction of the DM test) — corecția Harvey–Leybourne–Newbold a testului DM pentru eșantioane mici
- **IAPC**: Indicele armonizat al prețurilor de consum
- **KPSS**: Kwiatkowski–Phillips–Schmidt–Shin (test) — testul de staționaritate Kwiatkowski–Phillips–Schmidt–Shin
- **LB**: Ljung–Box (test) — testul de autocorelație Ljung–Box
- **MA**: Moving Average (medie mobilă)
- **MAE**: Mean Absolute Error (eroarea absolută medie)
- **MAPE**: Mean Absolute Percentage Error (eroarea procentuală absolută medie)
- **MASE**: Mean Absolute Scaled Error (eroarea absolută medie scalată)
- **MSTL**: Multiple Seasonal-Trend decomposition using Loess (descompunerea sezonier-trend cu mai multe perioade, prin Loess)
- **NSA**: Not Seasonally Adjusted (neajustat sezonier)

Acronime folosite în acest material (3/3)

- **OCSB**: Osborn–Chui–Smith–Birchenhall (seasonal unit-root test) — testul Osborn–Chui–Smith–Birchenhall de rădăcină unitară sezonieră
- **PACF**: Partial AutoCorrelation Function (funcția de autocorelație parțială)
- **PIB**: Produsul intern brut
- **RMSE**: Root Mean Squared Error (rădăcina erorii pătratice medii)
- **SARIMA**: Seasonal ARIMA (model) — model ARIMA sezonier
- **SCA**: Seasonally and Calendar Adjusted (ajustat sezonier și pentru efectele de calendar)
- **SEATS**: Signal Extraction in ARIMA Time Series (extragerea semnalului din serii de timp ARIMA)
- **STL**: Seasonal and Trend decomposition using Loess (descompunerea în sezonalitate și trend cu Loess)
- **TBATS**: Trigonometric seasonality, Box–Cox, ARMA errors, Trend, Seasonal components (model) — model cu sezonalitate trigonometrică, transformare Box–Cox, erori ARMA, trend și componente sezoniere
- **TRAMO-SEATS**: Time series Regression with ARIMA noise, Missing values and Outliers; Signal Extraction in ARIMA Time Series (metodă de ajustare sezonieră pe baza modelelor ARIMA)
- **TVA**: Taxa pe valoarea adăugată
- **UE**: Uniunea Europeană
- **US**: United States (Statele Unite ale Americii)
- **UTC**: Coordinated Universal Time (Temps Universel Coordonné) — timpul universal coordonat

Întrebarea de azi și traseul

- **Întrebarea**: multe serii economice repetă un tipar în fiecare an, săptămînă sau zi; cum modelăm acest tipar, cum prognozăm cu el și cum decidem care prognoză este mai bună?
  - trei răspunsuri: **eliminăm** tiparul prin diferențiere (SARIMA), îl **modelăm** cu regresori (termeni Fourier, TBATS, Prophet) sau îl **scoatem** din date (ajustarea sezonieră)
- **Traseul** capitolului
  - diferențe sezoniere și rădăcini unitare sezoniere (HEGY, Canova–Hansen, OCSB)
  - SARIMA$(p, d, q)(P, D, Q)_s$: identificare, estimare, diagnosticare, prognoze; modelul airline și PIB-ul României
  - ajustarea sezonieră în practică (X-13ARIMA-SEATS, Eurostat), termeni Fourier și efecte de calendar (Paștele ortodox)
  - sezonalitatea multiplă a consumului de electricitate: MSTL, regresia armonică dinamică, TBATS, Prophet
  - evaluarea prognozelor (validare încrucișată pentru serii de timp, testul Diebold–Mariano) și combinarea prognozelor
- Capitolul se sprijină pe Capitolul 0 (componente, Holt–Winters, MASE) și pe Capitolele 2–3 (ARMA, rădăcini unitare, ARIMA): aceleași idei, la lagul sezonier s
- Seminarul 4 are loc înaintea acestui curs: dă formulele; aici le deducem și le aplicăm pe date

Rezultatele învățării

- Recunoașteți sezonalitatea deterministă și pe cea stochastică și alegeți diferențierea sezonieră cu testele HEGY, Canova–Hansen și OCSB
- Scrieți, identificați, estimați, verificați și prognozați un model SARIMA$(p, d, q)(P, D, Q)_s$ și dezvoltați polinoamele lui
- Interpretați corect seriile oficiale ajustate sezonier și neajustate și explicați ce face X-13ARIMA-SEATS
- Modelați efectele de calendar (zile lucrătoare, Paștele ortodox) și sezonalitatea multiplă cu termeni Fourier, MSTL, TBATS și Prophet
- Comparați prognozele cu validarea încrucișată pentru serii de timp, MASE și testul Diebold–Mariano și combinați-le

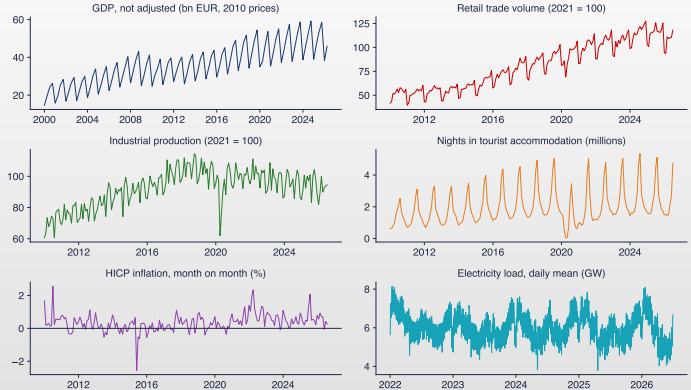
Bibliografie și instrumente

- Manual: Huang și Petukhina (2022), cap. 4–5 (modele ARIMA și modele nestaționare)
  - manual însoțitor, gratuit online: Hyndman și Athanasopoulos (2021), cap. 9 (FPP3, secțiunea 9.9), cap. 10 (FPP3, secțiunea 10.5), cap. 12 (FPP3, secțiunea 12.1) și cap. 13
- Lucrarea clasică: Box, Jenkins și Reinsel (2008), cap. 9 (modele sezoniere); ajustarea sezonieră: Ladiray și Quenneville (2001), Dagum și Bianconcini (2016), Eurostat (2024)
- Quantlet-urile Python ale capitolului: Quantlets/Ch_04
  - `statsmodels` (`SARIMAX`, `MSTL`, `ExponentialSmoothing`); pachetele opționale `tbats` și `prophet`
- Notebook-ul cursului: deschideți în Google Colab
- Curs video: Applied Time Series Analysis with Python

Sezonalitatea: definiție și surse

- **Sezonalitatea**: un tipar care se repetă cu o **perioadă** s fixă și cunoscută (observații pe ciclu)
  - $s = 4$ (trimestre), $s = 12$ (luni), $s = 7$ (zile într-o săptămînă), $s = 24$ și $s = 168$ (ore într-o zi și într-o săptămînă)
  - spre deosebire de **ciclu**, a cărui lungime este neregulată și necunoscută (Capitolul 0)
- **Surse**
  - vremea: încălzirea iarna, litoralul în august, recoltele
  - calendarul: zilele lucrătoare, sărbătorile legale, **sărbătorile mobile**, precum Paștele ortodox
  - instituțiile și obiceiurile: anul școlar, termenele fiscale, primele de sfîrșit de an, săptămîna de lucru
- **Aditivă** ($y_t = T_t + S_t + R_t$) sau **multiplicativă** ($y_t = T_t S_t R_t$, aditivă după logaritmare): Capitolul 0; aici lucrăm cu logaritmi cînd amplitudinea crește cu nivelul

Sezonalitatea în șase serii din România



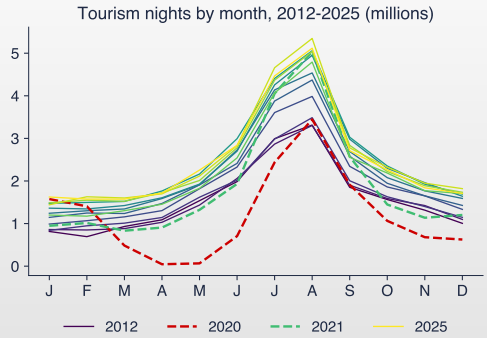
- Eurostat: PIB (trimestrial, neajustat), comerțul cu amănuntul, producția industrială, înnoptările turistice, IAPC; ENTSO-E: consumul zilnic de electricitate

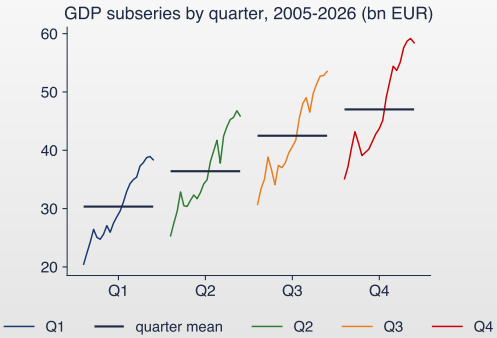
TSA_ch4_seasonal_data

Interpretarea celor șase serii

- PIB: media trimestrului 4 este cu 57% peste media trimestrului 1 (2000–2026); amplitudinea crește cu nivelul: multiplicativă, folosim logaritmi
  - mediile trimestriale (mld. EUR, prețuri 2010): 27,8; 33,6; 39,5; 43,6: construcțiile și agricultura în T2–T3
- Turismul: august are de 3,7 ori mai multe înnoptări decît ianuarie (2015–2019); restricțiile din 2020 (lockdown) întrerup tiparul
- IAPC: inflația lunară este mai mică în iunie (în medie $-0{,}37\%$, 2010–2019, alimentele proaspete) și cea mai mare în ianuarie (0,575%, prețuri și accize noi), din nou ridicată în octombrie (0,49%)
- Electricitatea: două tipare simultan, săptămîna (oscilația deasă) și anul (vîrfurile de iarnă): **sezonalitate multiplă**, secțiunea 8

Graficul sezonier și graficul pe subserii





- Stînga: o linie pentru fiecare an (graficul sezonier); dreapta: fiecare trimestru al PIB-ului urmărit de-a lungul anilor, cu media lui (graficul pe subserii)

TSA_ch4_seasonal_data

Interpretarea graficelor sezoniere

- Turismul: aceeași formă în fiecare an, un vîrf vara și un minim iarna
  - 2019: 5,06 milioane de înnoptări în august, 1,47 milioane în ianuarie
  - aprilie 2020: 0,045 milioane (1,76 milioane în aprilie 2019): o valoare aberantă, nu un sezon nou
- Subseriile PIB: toate cele patru trimestre cresc paralel
  - o formă sezonieră stabilă peste un trend comun
  - dacă liniile s-ar intersecta, tiparul sezonier s-ar schimba: sezonalitate stochastică (secțiunea următoare)



Mamaia, pe litoralul Mării Negre, septembrie 2013

Foto: Razvan Socol (2013); CC BY-SA 4.0; Wikimedia Commons

Recapitulare: sezonalitatea în datele economice

- Sezonalitatea se repetă cu o perioadă cunoscută s; ciclul nu
- Vremea, calendarul și instituțiile o creează; sărbătorile mobile o mută între luni
- PIB-ul, turismul, inflația și consumul de electricitate din România sînt toate sezoniere; electricitatea are mai multe perioade simultan
- Graficele sezoniere și pe subserii arată dacă tiparul este stabil

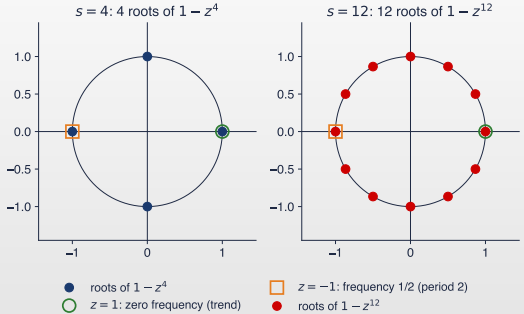
Diferența sezonieră

- **Diferența sezonieră**: $\Delta_s y_t = (1 - L^s) y_t = y_t - y_{t-s}$
  - cu logaritmi: $100\,\Delta_4 \ln Y_t$ = creșterea față de același trimestru al anului anterior, în %
- Un tipar sezonier **fix** $S_t = S_{t-s}$ plus un trend liniar: $y_t = a + bt + S_t + \varepsilon_t$
  - a: termenul liber; b: panta trendului; S_t: efectul sezonier al perioadei t
  - $\Delta_s y_t = bs + \varepsilon_t - \varepsilon_{t-s}$: Δ_s elimină tiparul și transformă trendul într-o constantă
  - dar creează termenul MA neinvertibil $\varepsilon_t - \varepsilon_{t-s}$: **supradiferențiere sezonieră** (ca în Capitolul 3)
- $\Delta\Delta_s = (1 - L)(1 - L^s) = 1 - L - L^s + L^{s+1}$: necesar cînd și trendul este stochastic
  - ordinea celor două diferențe nu contează: operatorii comută

Sezonalitatea deterministă și sezonalitatea stochastică

- **Deterministă**: $y_t = \sum_{j=1}^{s} \gamma_j D_{jt} + u_t$, cu $D_{jt} = 1$ în sezonul j și u_t staționar
  - mediile sezoniere γ_j nu se schimbă niciodată; instrumentul potrivit: variabile dummy sezoniere sau termeni Fourier (secțiunea 7)
- **Stochastică**: **mersul aleator sezonier** $y_t = y_{t-s} + \varepsilon_t$
  - fiecare sezon este propriul mers aleator: s mersuri aleatoare separate, intercalate
  - $\text{Var}(y_t)$ crește nelimitat; tiparul se deplasează, iar „vara poate deveni iarnă"
  - instrumentul potrivit: diferența sezonieră Δ_s
- Alegerea contează: diferențierea unui tipar determinist duce la supradiferențiere; modelarea unuia stochastic cu variabile dummy lasă o rădăcină unitară în reziduuri

Rădăcinile lui $1 - z^s$: rădăcini unitare sezoniere



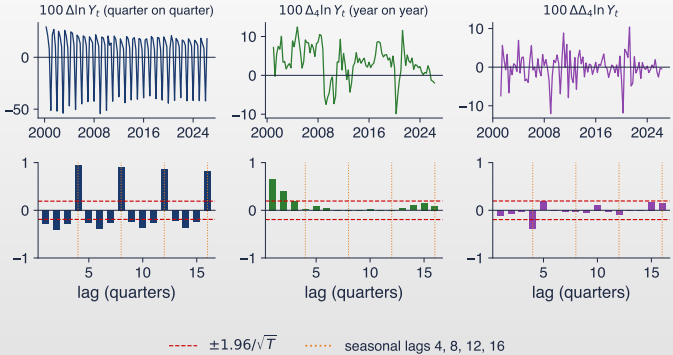
- $1 - z^4 = (1 - z)(1 + z)(1 + z^2)$: rădăcinile 1, -1, $\pm i$, toate pe cercul unitate

TSA_ch4_seasonal_roots

Interpretarea rădăcinilor sezoniere

- Δ_s impune simultan s rădăcini unitare, cîte una pentru fiecare **frecvență** k/s, $k = 0, \ldots, s-1$
  - trimestrial: $z = 1$ (frecvența 0, trendul), $z = -1$ (o jumătate de ciclu pe trimestru: perioada de 2 trimestre), $z = \pm i$ (perioada de 4 trimestre: ciclul anual)
- O serie poate avea unele dintre aceste rădăcini și nu pe altele: atunci Δ_4 supradiferențiază la frecvențele de care nu are nevoie
- Testele de rădăcină unitară sezonieră întreabă, frecvență cu frecvență, ce rădăcini sînt prezente: testul HEGY (slide-urile următoare)

PIB-ul României: diferențe obișnuite și sezoniere



- In din PIB-ul real neajustat, 106 de trimestre pînă în T2 2026; sus: trei transformări; jos: ACF de selecție pînă la 16 trimestre

TSA_ch4_seasonal_roots

Interpretarea celor trei transformări

- $\Delta \ln Y$: dominată de sezon: ACF 0,95 la lagul 4 și 0,90 la lagul 8, cu o descreștere foarte lentă; abaterea standard 26,5%
  - analogul sezonier al descreșterii lente a ACF pentru un mers aleator (Capitolul 3)
- $\Delta_4 \ln Y$ (creșterea anuală): abaterea standard 4,4%; ACF 0,66 la lagul 1, 0,01 la lagul 4: un tipar de tip AR
- $\Delta\Delta_4 \ln Y$: ACF $-0{,}11$ la lagul 1 și $-0{,}40$ la lagul 4 (banda $\pm 0{,}20$)
  - o singură valoare negativă semnificativă la lagul sezonier: un MA(1) sezonier, semnătura modelului airline (secțiunea 3)

Testul HEGY

- Hylleberg, Engle, Granger și Yoo (1990): o singură regresie testează fiecare rădăcină sezonieră a unei serii trimestriale
  - filtre: $y_{1t} = (1 + L + L^2 + L^3)y_t$, $y_{2t} = -(1 - L + L^2 - L^3)y_t$, $y_{3t} = -(1 - L^2)y_t$
- $\Delta_4 y_t = \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{3,t-1} +$ termeni deterministi + laguri ale lui $\Delta_4 y_t + \varepsilon_t$
  - y_{1t}, y_{2t}, y_{3t}: seria filtrată astfel încît fiecare să păstreze o singură frecvență sezonieră; π_k: coeficientul fiecărei frecvențe
  - H_0: $\pi_1 = 0$ (rădăcina 1), test t; $\pi_2 = 0$ (rădăcina -1), test t; $\pi_3 = \pi_4 = 0$ (rădăcinile $\pm i$), test F
  - distribuții nestandard, ca la Dickey–Fuller: valori critice din tabele sau prin simulare
- Respingerea tuturor celor trei: nu este nevoie de diferențiere sezonieră; nicio respingere: Δ_4 este justificată



Robert Engle, coautor al testului HEGY, 2017

Foto: Econterms (2017); CC BY-SA 4.0; Wikimedia Commons

Testele Canova–Hansen și OCSB

- **Canova–Hansen** (Canova și Hansen (1995)): ipoteza nulă este sezonalitatea **stabilă** (deterministă)
  - analogul sezonier al testului KPSS (Capitolul 3): respingerea înseamnă că tiparul se schimbă în timp
- **OCSB** (Osborn, Chui, Smith și Birchenhall (1988)): ipoteza nulă este o rădăcină unitară sezonieră; o regresie a lui $\Delta\Delta_s y_t$ pe $\Delta_s y_{t-1}$ și Δy_{t-s}
  - testul sezonier implicit al funcției `auto_arima` din pachetul Python `pmdarima` (`nsdiffs`)
- Ca în Capitolul 3: testele cu ipoteze nule opuse răspund la întrebări diferite; folosiți-le împreună
  - și priviți datele: graficele sezoniere, ACF la lagurile $s, 2s, 3s$ și mărimea lui $\hat{\Theta}$ după Δ_s ($\hat{\Theta} \approx -1$ semnalează supradiferențierea)

Teste de rădăcină unitară sezonieră pe date din România

HEGY, ln PIB, 2000–2019	statistica	valoarea critică 5%	decizia
$t(\pi_{\mathbf{1}})$, rădăcina 1	$-1{,}97$	$-3{,}39$	nu se respinge
$t(\pi_{\mathbf{2}})$, rădăcina -1	$-2{,}19$	$-2{,}77$	nu se respinge
$F(\pi_{\mathbf{3}}, \pi_{\mathbf{4}})$, rădăcinile $\pm i$	6,12	6,64	nu se respinge

Seria (logaritmi, 2005–2019)	D după OCSB	D după Canova–Hansen	d după KPSS
PIB (trimestrial)	1	0	1
Comerțul cu amănuntul	0	0	1
Producția industrială	0	0	1
Înnoptări turistice	0	0	1
IAPC	0	0	2

- HEGY: constantă, variabile dummy sezoniere și trend, 1 lag al lui $\Delta_{\mathbf{4}} y$ (AIC), $n = 75$; valori critice din 4 000 de simulări ale lui $\Delta_{\mathbf{4}} y_t = \varepsilon_t$
- Interpretare: pentru PIB, HEGY și OCSB susțin $\Delta_{\mathbf{4}}$, în timp ce Canova–Hansen nu respinge un tipar stabil; cu 20 de ani de date trimestriale ambele perspective se potrivesc, deci folosim $\Delta_{\mathbf{4}}$ și lăsăm $\hat{\Theta}$ să arate cît de stabil este tiparul (secțiunea 5)
- Seriile lunare: ambele teste aleg $D = 0$, un tipar stabil; IAPC are nevoie de $d = 2$ în logaritmi: inflația însăși este aproape de o rădăcină unitară (Capitolul 3)

Recapitulare: diferențierea sezonieră și rădăcinile unitare sezoniere

- $\Delta_s = 1 - L^s$ are s rădăcini unitare pe cerc, cîte una pentru fiecare frecvență sezonieră
- Sezonalitatea deterministă: variabile dummy sau termeni Fourier; sezonalitatea stochastică: Δ_s
- HEGY și OCSB testează rădăcinile unitare sezoniere; Canova–Hansen testează un tipar stabil
- PIB-ul României: $\Delta\Delta_4 \ln Y$ lasă o singură valoare semnificativă la lagul 4, un MA sezonier

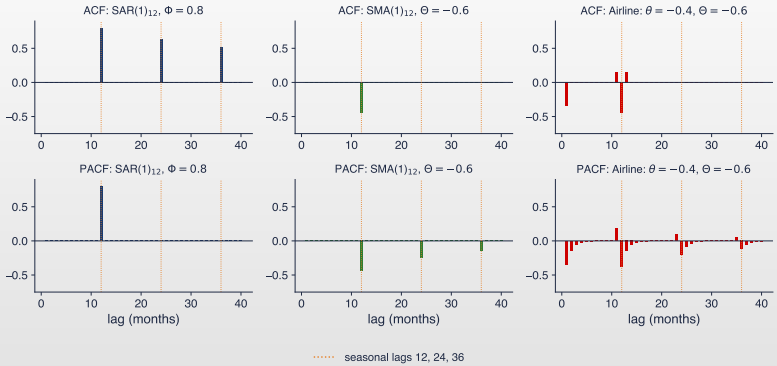
Modelul SARIMA$(p, d, q)(P, D, Q)_s$

- $\phi(L)\,\Phi(L^s)\,(1-L)^d(1-L^s)^D\,y_t = c + \theta(L)\,\Theta(L^s)\,\varepsilon_t$, $\varepsilon_t \sim \text{WN}(0, \sigma^2)$
  - polinoamele obișnuite $\phi(L)$ (ordinul p), $\theta(L)$ (ordinul q), ca în Capitolul 2
  - polinoamele sezoniere în L^s (ordinele P, Q): $\Phi(L^s) = 1 - \Phi_1 L^s - \cdots - \Phi_P L^{Ps}$, $\Theta(L^s) = 1 + \Theta_1 L^s + \cdots + \Theta_Q L^{Qs}$
  - d, D: numerele de diferențe obișnuite și sezoniere, de obicei 0 sau 1
- Motivul produsului (Box, Jenkins și Reinsel (2008)): într-o serie lunară, ianuarie depinde de decembrie (partea obișnuită) și de ianuarie anul trecut (partea sezonieră)
  - un tabel cu două intrări: lunile pe coloane, anii pe rînduri; cîte un model pe fiecare direcție
- Staționaritatea: rădăcinile lui $\phi(z)\Phi(z^s)$ în afara cercului unitate; invertibilitatea: rădăcinile lui $\theta(z)\Theta(z^s)$ în afara lui; pentru un AR(1) sezonier aceasta înseamnă $|\Phi| < 1$

Exemplu rezolvat: modelul airline

- SARIMA$(0,1,1)(0,1,1)_{12}$ cu $\theta = -0{,}4$, $\Theta = -0{,}6$: $(1-L)(1-L^{12})y_t = (1-0{,}4L)(1-0{,}6L^{12})\varepsilon_t$
  - stînga: $y_t - y_{t-1} - y_{t-12} + y_{t-13}$; dreapta: $\varepsilon_t - 0{,}4\varepsilon_{t-1} - 0{,}6\varepsilon_{t-12} + 0{,}24\varepsilon_{t-13}$
  - coeficientul lui ε_{t-13}, $\theta\Theta = 0{,}24$, nu este un parametru nou: doi parametri plus σ^2
- $w_t = \Delta\Delta_{12}y_t$ este un MA(13) cu doar patru ponderi nenule; autocovarianțele lui (Capitolul 2):
  - $\gamma(0) = \sigma^2(1+\theta^2)(1+\Theta^2) = 1{,}5776\,\sigma^2$
  - $\rho_1 = \theta/(1+\theta^2) = -0{,}345$; $\rho_{12} = \Theta/(1+\Theta^2) = -0{,}441$; $\rho_{11} = \rho_{13} = \rho_1\rho_{12} = 0{,}152$
  - toate celelalte autocorelații sînt zero

ACF și PACF teoretice pentru trei modele sezoniere



- Lunar, $s = 12$; panoul airline arată partea staționară $w_t = \Delta\Delta_{\mathbf{12}} y_t$

TSA_ch4_sarima_theory

Interpretarea tiparelor sezoniere

- AR(1) sezonier: ACF descrește geometric la lagurile 12, 24, 36 (0,8; 0,64; 0,512); PACF are o singură valoare nenulă la lagul 12
  - regula AR din Capitolul 2, citită doar la multiplii lui s
- MA(1) sezonier: ACF se anulează după lagul 12; PACF descrește la 12, 24, 36 ($-0{,}242$ la lagul 24)
- Airline: valori nenule la 1 ($-0{,}345$) și 12 ($-0{,}441$), cu „sateliți" la 11 și 13 (0,152)
  - sateliții sînt produsul $\rho_1\rho_{12}$: amprenta structurii multiplicative

Identificarea modelelor SARIMA

După diferențiere	ACF	PACF
AR(P) sezonier	descrește la $s, 2s, \ldots$	se anulează după lagul Ps
MA(Q) sezonier	se anulează după lagul Qs	descrește la $s, 2s, \ldots$
partea obișnuită	lagurile $1, 2, \ldots$ ca în Capitolul 2	lagurile $1, 2, \ldots$ ca în Capitolul 2
produsul celor două	sateliți la $ks \pm j$	sateliți la $ks \pm j$

Pașii

- 1. Stabilizați varianța (logaritmi sau Box–Cox, Box și Cox (1964)); 2. alegeți D (graficul sezonier, HEGY, OCSB), apoi d (Capitolul 3)
- 3. Citiți P, Q la lagurile sezoniere și p, q la primele laguri; 4. păstrați ordinele mici: $P, Q \leq 1$ sînt aproape întotdeauna suficiente

Estimarea și alegerea modelului

- **Verosimilitatea maximă** pentru seria diferențiată (forma în spațiul stărilor, `SARIMAX` din `statsmodels`)
  - se pierd $s+1$ observații prin $\Delta\Delta_s$; constanta lipsește cînd $d+D \geq 2$ (ar însemna un trend pătratic)
- **Criteriile informaționale**: AICc $=$ AIC $+\frac{2k(k+1)}{n-k-1}$ și BIC, comparate **doar** între modele cu aceleași d și D
  - k: numărul parametrilor estimați; n: numărul de observații după diferențiere
  - diferențierea schimbă datele pe care se calculează verosimilitatea
- **SARIMA automat**: o căutare pas cu pas după (p, q, P, Q) cu AICc, după alegerea lui d și D prin teste (Hyndman și Khandakar (2008))
  - în Python: `pmdarima.auto_arima(y, m=12)`; verificați întotdeauna reziduurile și comparați cu un reper
- **Diagnosticarea**: Ljung–Box cu laguri dincolo de s (de exemplu $m = 2s$) și $m - p - q - P - Q$ grade de libertate (Ljung și Box (1978))

Prognoza cu modele SARIMA

- Ca pentru ARIMA (Capitolul 3): scriem modelul ca ecuație pentru y_t; înlocuim șocurile viitoare cu 0, valorile viitoare cu prognozele lor și șocurile trecute cu reziduuri
  - airline: $\hat{y}_{T+h} = \hat{y}_{T+h-1} + \hat{y}_{T+h-s} - \hat{y}_{T+h-s-1} +$ (termeni MA cît timp $h \leq s+1$)
- **Funcția de prognoză pe termen lung** a modelului airline: un tipar sezonier fix pe o dreaptă
  - tiparul este o medie ponderată exponențial a tiparelor trecute, cu ponderea $1+\Theta$ pentru anul cel mai recent
  - $\Theta \to -1$: un tipar fix (determinist); $\Theta = 0$: tiparul de anul trecut repetat (sezonier naiv)
- Intervalele se lărgesc cu h (două rădăcini unitare); pentru $\ln y$, $\exp(\hat{y}_{T+h})$ este **mediana** prognozei
  - media este $\exp(\hat{y}_{T+h} + \sigma_h^2/2)$, puțin mai mare; σ_h^2: varianța erorii de prognoză la orizontul h

Exemplu rezolvat: prognoza unui MA sezonier

- SARIMA$(0,0,0)(0,1,1)_4$ trimestrial: $y_t = y_{t-4} + \varepsilon_t + \Theta\varepsilon_{t-4}$, $\Theta = -0{,}6$
  - ultimul an: $y_{T-3}, \ldots, y_T = 100; 120; 130; 140$; reziduurile $\hat{\varepsilon}_{T-3}, \ldots, \hat{\varepsilon}_T = 2; -1; 0; 3$
- $\hat{y}_{T+1} = y_{T-3} + \Theta\hat{\varepsilon}_{T-3} = 100 - 0{,}6 \cdot 2 = 98{,}8$
  - $\hat{y}_{T+2} = 120{,}6$, $\hat{y}_{T+3} = 130{,}0$, $\hat{y}_{T+4} = 138{,}2$
- $\hat{y}_{T+5} = \hat{y}_{T+1} = 98{,}8$: după un an, termenul MA dispare și tiparul prognozei se repetă
  - o surpriză pozitivă mare în ultimul sezon ($\hat{\varepsilon}_T = 3$) scade prognoza pentru acel sezon cu $0{,}6 \cdot 3$: o corecție parțială

Recapitulare: modele SARIMA

- SARIMA înmulțește polinoame obișnuite și sezoniere: $\phi(L)\Phi(L^s)\Delta^d\Delta_s^D y_t = \theta(L)\Theta(L^s)\varepsilon_t$
- Identificarea: regulile din Capitolul 2 la lagurile $s, 2s, \ldots,$ plus sateliții de la $ks \pm j$
- Comparați AICc doar pentru aceleași d și D; verificați Ljung–Box dincolo de lagul s
- Prognoza airline: un tipar sezonier pe o dreaptă, actualizat cu ponderea $1 + \Theta$

Box, Jenkins și pasagerii companiilor aeriene

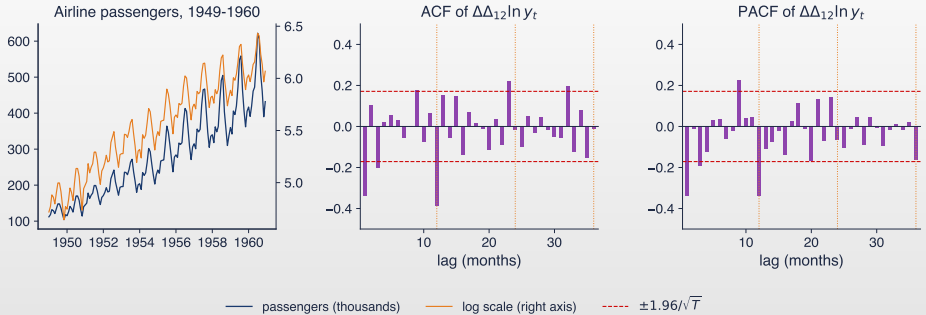
- Box, Jenkins și Reinsel (2008) (prima ediție în 1970) au modelat totalurile lunare ale pasagerilor internaționali, 1949–1960 (144 de luni)
  - creștere plus un vîrf de vară a cărui mărime crește cu nivelul
  - modelul lor, SARIMA$(0,1,1)(0,1,1)_{\mathbf{12}}$ pentru $\ln y_t$, poartă de atunci numele de „model airline"
- Motivele longevității modelului
  - doi parametri se potrivesc multor serii sezoniere: comerț, producție, vînzări cu amănuntul, energie
  - este modelul de referință al etapei regARIMA din programele de ajustare sezonieră (secțiunea 6)



Primele avioane Boeing 707 pentru Pan American, Seattle, 1958

Foto: autor necunoscut (1958); domeniu public; Wikimedia Commons

Identificarea modelului airline



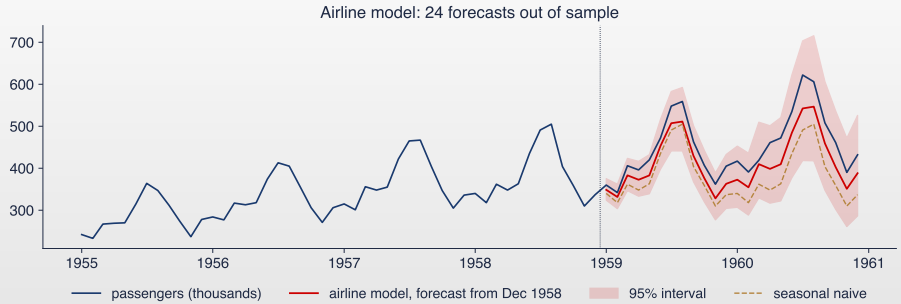
- Stînga: pasagerii și logaritmul lor; mijloc și dreapta: ACF și PACF pentru $w_t = \Delta\Delta_{12} \ln y_t$, 131 de observații

TSA_ch4_airline

Interpretarea corelogramelor airline

- Logaritmul face constantă amplitudinea verii: tiparul multiplicativ devine aditiv
- ACF pentru w_t: $-0{,}34$ la lagul 1 și $-0{,}39$ la lagul 12, în afara benzii $\pm 0{,}17$
  - un MA(1) obișnuit și un MA(1) sezonier; lagul 3 ($-0{,}20$) este la limită
  - barele izolate de la lagurile 9, 23 și 32 depășesc și ele banda: cu 36 de laguri, aproximativ două depășiri sînt așteptate din întîmplare
- PACF: valori negative mari la lagurile 1, 3 și 12, aproape 0 la lagul 2 și nicio anulare clară după cîteva laguri: în acord cu termeni MA, nu cu termeni AR
- Concluzia: SARIMA$(0,1,1)(0,1,1)_{12}$, exact modelul lui Box și Jenkins

Modelul airline: prognoze în afara eșantionului



- Estimat pe 1949–1958, prognoze pentru cele 24 de luni din 1959–1960, transformate înapoi din $\ln y$

TSA_ch4_airline

Interpretarea prognozelor airline

- 1949–1958: $\hat{\theta} = -0{,}342$ (0,087), $\hat{\Theta} = -0{,}541$ (0,105); întregul eșantion: $\hat{\theta} = -0{,}402$ (0,073), $\hat{\Theta} = -0{,}557$ (0,096)
  - întregul eșantion: abaterea standard a reziduurilor 3,7% pe lună; Ljung–Box $Q^*(24)$ cu 22 de grade de libertate, p = 0,37
- MAPE pe 1959–1960: 8,5% pentru modelul airline, 15,5% pentru prognoza sezonieră naivă (care ignoră creșterea)
- Decembrie 1960: prognoza 388, intervalul [286; 526], valoarea reală 432: în interior, dar modelul subestimează creșterea din 1959–1960
  - lățimea intervalului crește de la 51 (prima lună) la 239 (luna 24): două rădăcini unitare

Modele SARIMA pentru PIB-ul României

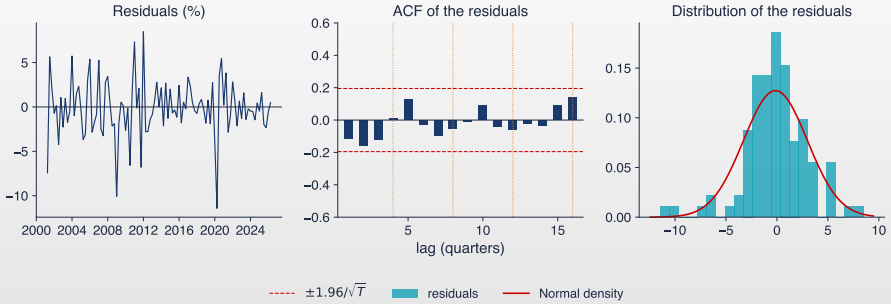
Modelul	AICc	BIC	LB(8), p	$\hat{\sigma}$ (%)
$(0,1,1)(0,1,1)_4$	520,1	527,6	0,157	3,05
$(0,1,0)(0,1,1)_4$	519,7	**524,8**	0,286	3,08
$(1,1,0)(0,1,1)_4$	520,7	528,3	0,156	3,06
$(1,1,1)(0,1,1)_4$	**518,3**	528,4	0,467	2,98
$(0,1,2)(0,1,1)_4$	518,8	528,8	0,575	3,00
$(0,1,1)(1,1,0)_4$	529,9	537,5	0,003	3,22
$(0,1,1)(1,1,1)_4$	522,2	532,2	0,118	3,05
$(0,1,1)(0,1,0)_4$	545,1	550,2	0,005	3,52

- $y_t = 100 \ln Y_t$, PIB real neajustat, T1 2000–T2 2026, $T = 106$; verosimilitate maximă; toate cu $d = D = 1$, deci criteriile sînt comparabile
- Ljung–Box pe 8 laguri, cu $8 - p - q - P - Q$ grade de libertate

Interpretarea tabelului de modele

- BIC alege $(0,1,0)(0,1,1)_4$: doar termenul MA sezonier; AICc alege $(1,1,1)(0,1,1)_4$
  - modelul AICc are $\hat{\phi} = 0{,}78$ și $\hat{\theta} = -0{,}94$: aproape un factor comun (Capitolul 2), cu un cîștig mic
  - modelul airline: $\hat{\theta} = -0{,}171$ (0,109), nesemnificativ; $\hat{\Theta} = -0{,}585$
- Un AR sezonier în locul MA sezonier sau niciun termen sezonier: criterii mult mai slabe și reziduuri corelate
- Păstrăm modelul BIC: un singur parametru, $\hat{\Theta} = -0{,}577$ (0,075), $\hat{\sigma} = 3{,}08\%$

Diagnosticarea modelului pentru PIB



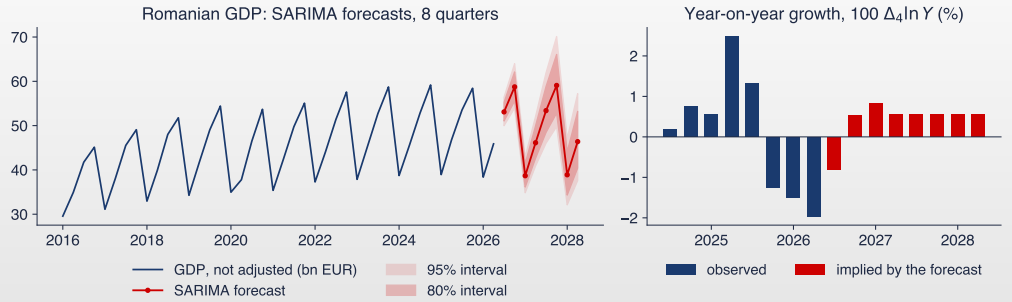
- SARIMA$(0,1,0)(0,1,1)_4$ pentru $100 \ln Y_t$: reziduurile, ACF a lor și distribuția lor

TSA_ch4_gdp_case

Interpretarea diagnosticării PIB

- Nu rămîne autocorelație: $Q^*(8) = 8{,}55$ cu 7 grade de libertate, $p = 0{,}286$; $Q^*(12)$: $p = 0{,}526$
  - nicio valoare semnificativă la lagurile sezoniere 4, 8, 12, 16: structura sezonieră este surprinsă
- Nu este distribuția Normală: Jarque–Bera 22,0 ($p < 0{,}001$); cel mai mare reziduu este $-11{,}4\%$ în T2 2020 ($-3{,}6$ abateri standard): perioada de lockdown
  - fără 2020, Jarque–Bera scade la 13,4: o variabilă dummy pentru valoarea aberantă ar fi pasul următor

Prognozele PIB-ului României



- SARIMA$(0,1,0)(0,1,1)_4$, 8 trimestre, din T3 2026 pînă în T2 2028, intervale de 80% și 95%; dreapta: creșterea anuală implicată

TSA_ch4_gdp_case

Interpretarea prognozelor PIB

- Ultima valoare: 45,9 mld. EUR în T2 2026; prognozele 53,1; 58,7; 38,7; 46,1 mld. EUR pentru următoarele patru trimestre
  - intervalul de 95% pentru primul trimestru [50,0; 56,4], pentru al optulea [37,6; 57,2]
- Creșterea anuală implicată: $-0{,}81\%$; $0{,}53\%$; $0{,}83\%$; $0{,}55\%$, apoi constant $0{,}55\%$
  - modelul extrapolează creșterea slabă recentă; nu conține informații despre politica fiscală sau fondurile europene
- Tiparul sezonier al prognozei este o medie ponderată a anilor trecuți, cu ponderea $1 + \hat{\Theta} = 0{,}42$ pentru anul cel mai recent

Recapitulare: pasagerii aerieni și PIB-ul României

- Logaritmi, apoi $\Delta\Delta_s$, apoi termeni MA la lagurile 1 și s: rețeta airline funcționează pentru ambele serii
- Pentru PIB, BIC renunță la termenul MA obișnuit; AICc alege un ARMA(1,1) aproape redundant
- Reziduurile sînt curate, cu excepția valorii aberante din 2020
- Prognozele repetă un tipar sezonier netezit pe o linie de trend, cu intervale tot mai largi

Rostul ajustării sezoniere în statistica oficială

- Utilizatorii vor să știe dacă economia **a crescut în acest trimestru**, nu dacă T3 este mai mare decît T2
  - T1 2026: PIB-ul neajustat s-a modificat cu $-42{,}1\%$ (puncte logaritmice) față de trimestrul anterior; seria ajustată cu $-0{,}06\%$
  - prima cifră este doar sezon; doar a doua spune ceva despre ciclul economic
- **Ajustarea sezonieră**: estimăm componentele sezonieră și de calendar și le eliminăm: $y_t^{SA} = y_t - \hat{S}_t - \hat{C}_t$ (sau $y_t/(\hat{S}_t\hat{C}_t)$)
  - descompunerea clasică și STL din Capitolul 0 sînt variante simple ale acestei idei
- Pentru prognoza cu SARIMA folosim seria **neajustată**: filtrele de ajustare schimbă autocorelațiile și sînt revizuite cînd apar date noi

X-11 și X-13ARIMA-SEATS

- **X-11** (US Census Bureau, anii 1960; Ladiray și Quenneville (2001)): medii mobile aplicate iterativ
  - o medie mobilă centrată 2×12 pentru trend, apoi medii mobile 3×3 și 3×5 ale fiecărei luni de-a lungul anilor pentru factorii sezonieri
- **X-12-ARIMA** (Findley et al. (1998)) și **X-13ARIMA-SEATS** adaugă o **preajustare regARIMA**:
  - regresie cu erori SARIMA (adesea modelul airline) pentru valori aberante, zile lucrătoare, Paște
  - prognozele ARIMA prelungesc seria, astfel încît filtrele simetrice funcționează și la capăt: revizuiri mai mici
  - apoi filtrele X-11 sau SEATS (descompunerea pe baza modelului ARIMA)
- UE: Eurostat și institutele naționale urmează Eurostat (2024), cu programul JDemetra+ (X-13 și TRAMO-SEATS)



US Census Bureau, Suitland, Maryland: locul de naștere al X-11 și X-13

Foto: United States Census Bureau (2007); domeniu public; Wikimedia Commons

Seriile Eurostat: NSA, CA, SA, SCA

- Fiecare serie apare în mai multe variante (dimensiunea `s_adj`):
  - **NSA**: neajustată sezonier (datele brute)
  - **CA**: ajustată doar pentru efectele de calendar (zile lucrătoare, sărbători)
  - **SA**: ajustată sezonier; **SCA**: ajustată sezonier și pentru efectele de calendar
- Varianta potrivită pentru fiecare întrebare
  - creșterea față de trimestrul sau luna anterioară: SCA
  - creșterea față de anul anterior: NSA sau CA (tiparul sezonier se anulează, calendarul nu)
  - modelarea și prognoza SARIMA: NSA, cu regresori de calendar
- Totalurile anuale NSA și SCA sînt apropiate (2024: 196,4 și 196,4 mld. EUR): ajustarea mută producția între trimestre, nu o creează

Serii neajustate și ajustate



- Eurostat NSA și SCA: PIB (stînga sus) și comerțul cu amănuntul (dreapta sus); jos: factorii sezonieri și de calendar implicați, NSA/SCA

TSA_ch4_seasonal_data

Interpretarea seriilor ajustate

- Factorii PIB 2015–2019: T1 0,78, T2 0,93, T3 1,10, T4 1,19: producția din T1 este cu aproximativ o cincime sub un trimestru mediu
  - factorii se mișcă lent de-a lungul anilor: sezonalitatea stochastică măsurată de $\hat{\Theta}$
- Comerțul cu amănuntul: factorul pentru decembrie 1,10, pentru ianuarie 0,84 (2016–2019): cumpărăturile de Crăciun, apoi minimul din ianuarie
- Factorul pentru comerț sare și de la o lună la alta din cauza zilelor lucrătoare și a Paștelui: partea de calendar a SCA (secțiunea 7)
- PIB-ul ajustat arată scăderea din 2020 și stagnarea din 2025–2026 pe care seria brută le ascunde

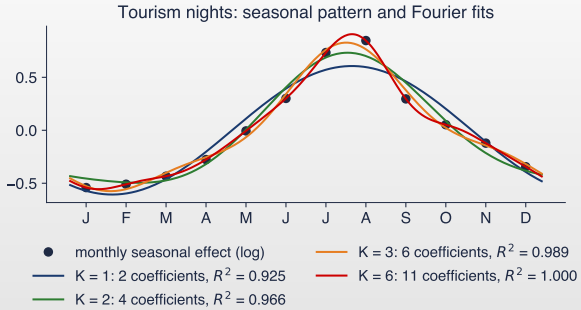
Recapitulare: ajustarea sezonieră

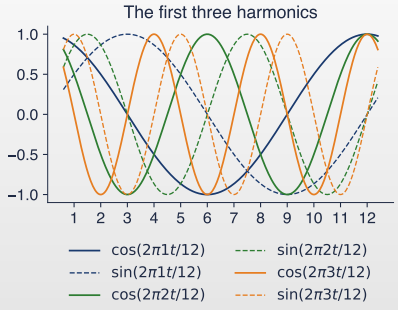
- Ajustarea sezonieră elimină $\hat{S}_t$ și $\hat{C}_t$ pentru a arăta trendul și ciclul
- X-13ARIMA-SEATS: preajustare regARIMA (valori aberante, calendar, prognoze), apoi X-11 sau SEATS; în UE, JDemetra+
- Creșterea față de trimestrul anterior: SCA; creșterea anuală: NSA sau CA; prognoza SARIMA: NSA

Variabile dummy sezoniere și termeni Fourier

- **Variabile dummy**: $y_t = \beta_0 + \sum_{j=2}^{s} \gamma_j D_{jt} + u_t$: $s-1$ coeficienți, orice formă
  - β_0: nivelul sezonului 1; γ_j: diferența dintre sezonul j și sezonul 1
  - prea mulți pentru $s = 52$ de săptămîni, $s = 168$ ore sau $s = 365$ de zile; imposibil pentru $s = 365{,}25$
- **Termeni Fourier** pentru o perioadă m: $\sum_{k=1}^{K} \left[\alpha_k \sin \frac{2\pi kt}{m} + \beta_k \cos \frac{2\pi kt}{m}\right]$
  - α_k, β_k: coeficienții armonicei k, o undă cu k cicluri pe perioadă; K: numărul de perechi sinus–cosinus
  - $2K$ coeficienți; $K = m/2$ reproduce exact variabilele dummy (pentru m par); un K mic dă un tipar neted
  - m poate fi neîntreg și se pot combina mai multe perioade: de exemplu $K = 5$ pentru $m = 24$ și $K = 4$ pentru $m = 168$ dau 18 regresori în locul a $23 + 167$
- K se alege cu AICc sau prin validare încrucișată; aceiași termeni se află în TBATS și Prophet (secțiunea 9)

Aproximarea Fourier a unui tipar sezonier





- Efectul sezonier lunar al ln(înnoptări turistice), 2012–2019, după eliminarea unei medii centrate pe 13 luni; ajustări cu $K = 1, 2, 3, 6$; pentru $K = 6$, $\sin(\pi t)$ este 0 în fiecare lună și este omis (11 coeficienți)

TSA_ch4_calendar_fourier

Interpretarea ajustărilor Fourier

- $K = 1$ (doi coeficienți) dă deja $R^2 = 0{,}925$: vîrful de vară este aproape o singură undă
  - $K = 2$: 0,966; $K = 3$: 0,989 cu șase coeficienți în loc de unsprezece
- $K = 5$ (10 coeficienți): $R^2 = 0{,}999$; $K = 6$ adaugă $\cos(\pi t)$, 11 coeficienți, și trece prin cele 12 puncte, ca 11 variabile dummy ($R^2 = 1{,}000$)
- Armonicele superioare adaugă detalii mai fine: vîrful din august cere $K \geq 3$

Efecte de calendar și Paștele ortodox

- **Zilele lucrătoare**: o lună cu mai multe zile de luni pînă vineri care nu sînt sărbători produce și vinde mai mult
  - cunoscute dinainte: un regresor care poate fi folosit în prognoze
- **Sărbătorile mobile**: Paștele ortodox cade între 4 aprilie și 8 mai
  - 5 mai 2024, 20 aprilie 2025, 12 aprilie 2026, 2 mai 2027: cumpărăturile se mută între martie, aprilie și mai
  - în Python: `dateutil.easter.easter(y, EASTER_ORTHODOX)`; Rusaliile sînt cu 49 de zile mai tîrziu
- **Regresorul Paște** (`easter[w]` din X-13): E_t = partea din cele w zile dinaintea Duminicii Paștelui care cade în luna t
  - $w = 10$: Paștele pe 20 aprilie dă $E_{\textbf{aprilie}} = 1$; pe 5 mai, $E_{\textbf{aprilie}} = 0{,}6$ și $E_{\textbf{mai}} = 0{,}4$



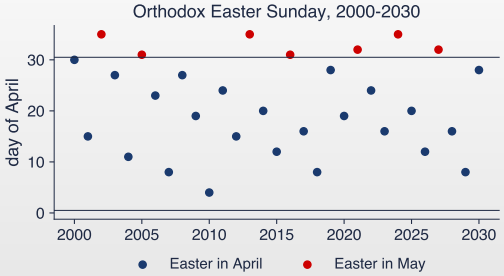
Ouă încondeiate de Paște din România

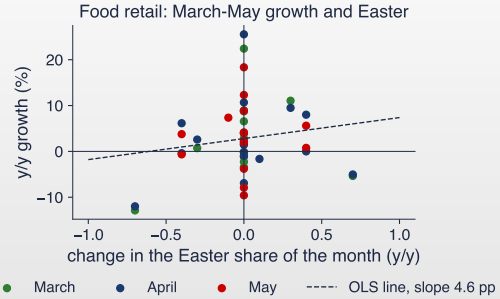
Foto: jackmac34 (Pixabay); CC0; Wikimedia Commons

Regresia cu erori SARIMA

- $100 \ln y_t = \beta_E E_t + \beta_W \mathrm{WD}_t +$ variabile dummy pentru valori aberante $+ u_t$, cu $u_t \sim \mathrm{SARIMA}(0,1,1)(0,1,1)_{12}$
  - E_t: regresorul Paște; WD_t: numărul de zile lucrătoare din luna t; β_E, β_W: efectele lor, în %
  - modelul „regARIMA" din X-13; estimat într-un singur pas prin verosimilitate maximă (`SARIMAX` cu exog)
  - metoda celor mai mici pătrate care ignoră erorile SARIMA ar da erori standard greșite (Capitolul 3, regresia falsă)
- Datele: volumul comerțului cu amănuntul, Eurostat, România, neajustat, ianuarie 2010–iulie 2026: alimente (G47_FOOD) și total (G47)
  - variabile dummy pentru aprilie și mai 2020 (lockdown)

Paștele și comerțul cu amănuntul din România





- Stînga: Duminica Paștelui ortodox ca zi din aprilie (peste 30: mai); dreapta: creșterea anuală a vînzărilor de alimente în martie–mai față de modificarea ponderii Paștelui (fără 2020–2021)

TSA_ch4_calendar_fourier

Interpretarea efectului Paștelui

- Alimente: $\hat{\beta}_E = 5{,}9\%$ (eroarea standard 2,3, $t = 2{,}6$): o lună care cuprinde cele 10 zile dinaintea Paștelui vinde cu aproximativ 5,9% mai multe alimente
  - o zi lucrătoare în plus: $+0{,}59\%$ (eroarea standard 0,17); aprilie 2020: -15%
- Totalul comerțului: Paștele 0,8% (eroarea standard 1,9), nesemnificativ; zilele lucrătoare $+1{,}02\%$ (eroarea standard 0,15)
  - efectul Paștelui se află în alimente, efectul zilelor lucrătoare în toate vînzările
- Norul de puncte brut este zgomotos (panta 4,6 puncte procentuale): modificările TVA și prețurile mișcă și ele creșterea anuală; regresia separă efectele
- Paștele a căzut în mai în 2002; 2005; 2013; 2016; 2021; 2024: cine ignoră acest lucru greșește prognozele pentru aprilie și mai

Recapitulare: termeni Fourier și efecte de calendar

- Termenii Fourier descriu un tipar sezonier cu $2K$ coeficienți, pentru orice perioadă, chiar neîntreagă
- Zilele lucrătoare și sărbătorile mobile sînt cunoscute dinainte: le modelăm ca regresori
- Vînzările de alimente din România cresc cu aproximativ 5,9% în luna Paștelui; totalul comerțului urmează zilele lucrătoare
- Regresia cu erori SARIMA: estimăm împreună regresorii și dinamica

Consumul de electricitate din România

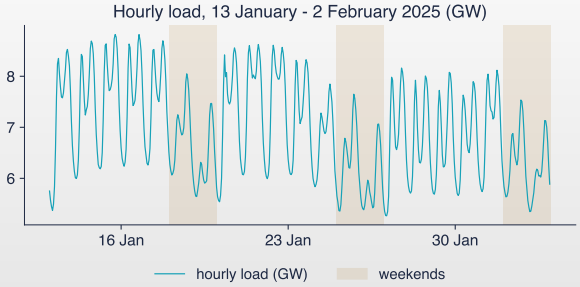
- **Consumul** (sarcina): energia electrică consumată în sistem, în MW (medii orare); trebuie egalat de producție în fiecare moment
  - prognozele sînt necesare pentru fiecare oră a zilei următoare: piața pentru ziua următoare, rezerve, planificarea rețelei
- Datele: ENTSO-E (rețeaua europeană a operatorilor de transport și de sistem), „monthly hourly load values", România
  - 39 406 de ore, 2022–2026, ora de iarnă a României (UTC+2), pentru a evita schimbările de oră
- Prognoza consumului are propriile competiții: GEFCom2014 (Hong et al. (2016))

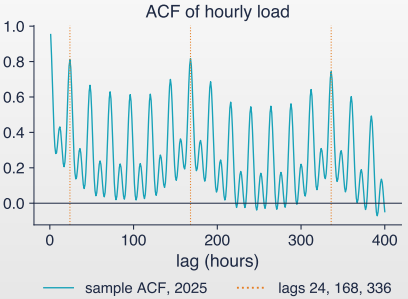


Hidrocentrala Porțile de Fier II, pe Dunăre

Foto: Nenea hartia (2016); CC BY-SA 4.0; Wikimedia Commons

Consumul orar: ziua și săptămîna





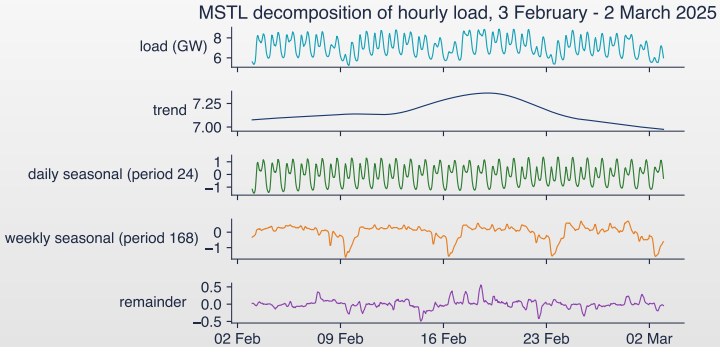
⊡ Stînga: trei săptămîni din ianuarie 2025; dreapta: ACF de selecție pentru toate orele din 2025, pînă la lagul 400

TSA_ch4_multiple_seasonality

Interpretarea consumului orar

- Ciclul zilnic: minimul în jurul orei 3:00 (în medie 5,3 GW în 2025), maximul în jurul orei 19:00 (7,3 GW)
  - ACF 0,81 la lagul 24, 0,43 la lagul 12
- Ciclul săptămînal: weekendurile sînt mai joase; ACF 0,82 la lagul 168 (o săptămînă), peste valoarea de la lagul 24
  - aceeași oră de acum o săptămînă este cel mai bun predictor individual: lunea seamănă cu lunea trecută, nu cu duminica
- Și un ciclu anual peste acestea (încălzire și aer condiționat): trei perioade, 24, 168 și aproximativ 8766 de ore
- SARIMA nu este suficient: are o singură perioadă întreagă s, în timp ce aici acționează simultan $s = 24$, 168 și 8766
  - un polinom în L^{168} este lent și instabil, 365,25 nu este un număr întreg, iar sărbătorile rup tiparul săptămînal
  - răspunsuri: MSTL, regresia armonică dinamică (DHR), TBATS, Prophet

MSTL: cîte o componentă sezonieră pentru fiecare perioadă



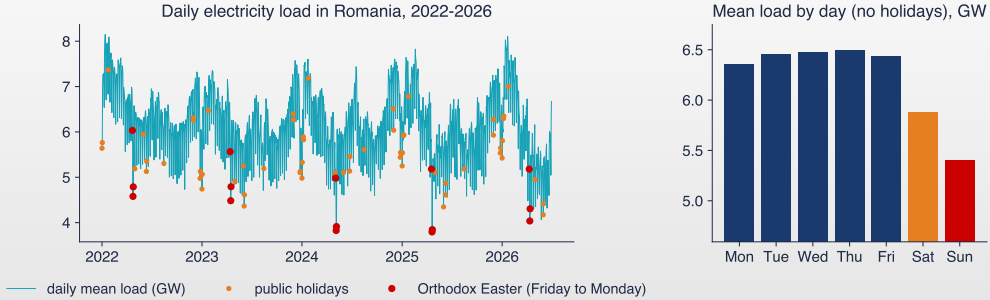
- MSTL (Bandara, Hyndman și Bergmeir (2025)) aplică STL (Capitolul 0) în mod repetat, cîte o dată pentru fiecare perioadă (24 și 168 de ore), pînă cînd componentele se stabilizează

TSA_ch4_multiple_seasonality

Interpretarea descompunerii MSTL

- Din varianța în jurul trendului: 70% componenta zilnică, 29% componenta săptămînală, 2% componenta neregulată
  - amplitudinea componentei zilnice în prima săptămînă: 2,8 GW; a celei săptămînale: 2,1 GW
- Componenta săptămînală este plată de luni pînă vineri și scade sîmbăta și mai mult duminica
- Componenta neregulată este mică, dar nu este zgomot alb: valuri de frig și sărbători; un model de prognoză adaugă dinamică și regresori

Consumul zilnic și sărbătorile legale



- Consumul mediu zilnic, 2022–2026, 1 642 de zile; sărbătorile legale marcate; dreapta: media pe zile ale săptămînii, fără sărbători

TSA_ch4_multiple_seasonality

Interpretarea consumului zilnic

- Săptămîna: sîmbăta cu 9%, iar duminica cu 17% sub miercuri
  - Anul: ianuarie cu 25% peste mai; un vîrf mai mic vara (aerul condiționat)
- Sărbătorile: cea mai joasă zi din eșantion este Duminica Paștelui, 20 aprilie 2025 (3,79 GW, media 6,18 GW)
  - cele patru zile de Paște (vineri–luni) au în medie 4,62 GW; perioada Crăciun–Anul Nou și Rusaliile sînt și ele joase (Duminica Rusaliilor a fost cea mai joasă zi din 2023)
- Nivelul scade în 2022–2026; cauze posibile: șocul de preț din 2022 și panourile solare de pe acoperișuri, care acoperă o parte din cerere fără a trece prin rețea

O regresie armonică dinamică pentru consumul zilnic

- $y_t = \sum_{k=1}^{4} \left[\alpha_k \sin \frac{2\pi kt}{365{,}25} + \beta_k \cos \frac{2\pi kt}{365{,}25}\right] + \sum_j \delta_j D_{jt}^{\text{zi}} + \lambda_1 H_t + \lambda_2 E_t + \lambda_3 X_t + u_t$
  - α_k, β_k: coeficienții Fourier anuali; δ_j: efectul zilei j față de duminică; $\lambda_1, \lambda_2, \lambda_3$: efectele zilelor speciale, în GW
  - termeni Fourier anuali; variabile dummy pentru zilele săptămînii (duminica este baza); H sărbătoare legală, E Paște (vineri–luni), X 24 decembrie – 2 ianuarie
  - $u_t \sim$ ARIMA(1,1,1), ales cu AIC: **regresia armonică dinamică** (DHR, FPP3, secțiunea 10.5)
- Estimările, cu datele pînă la prima origine de prognoză (30 iunie 2025), în GW:
  - miercuri +1,08, luni +0,88, sîmbătă +0,50 față de duminică
  - sărbătoare legală −0,38 (eroarea standard 0,02), Paștele încă −0,58, Crăciun–Anul Nou −0,61
  - erori: AR 0,80, MA −0,98, $\hat{\sigma} = 186$ MW; Ljung–Box(14) p $< 0{,}001$: rămîne puțină dinamică

Recapitulare: sezonalitatea multiplă

- Consumul orar are cicluri zilnice, săptămînale și anuale; consumul zilnic are cicluri săptămînale și anuale
- SARIMA tratează o singură perioadă întreagă; MSTL descompune mai multe
- Sărbătorile, mai ales Paștele, sînt cele mai mari abateri: au nevoie de regresori
- DHR = termeni Fourier + variabile dummy de calendar + erori ARIMA: flexibilă și transparentă

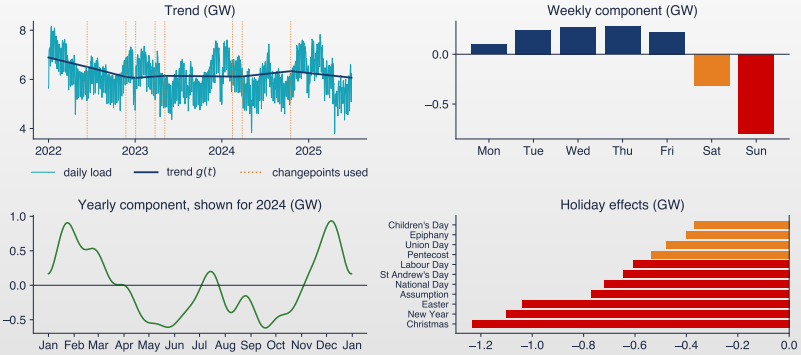
TBATS

- De Livera, Hyndman și Snyder (2011): netezire exponențială (ETS, Capitolul 0) pentru sezonalitate complexă; numele enumeră părțile modelului
  - sezonalitate trigonometrică (T), transformare Box–Cox (B), erori ARMA (A), trend (T), componente sezoniere (S)
- $y_t^{(\omega)} = \ell_{t-\mathbf{1}} + \phi b_{t-\mathbf{1}} + \sum_i s_{t-\mathbf{1}}^{(i)} + d_t$, d_t ARMA(p, q); nivelul și trendul se actualizează ca în metoda Holt
  - $y_t^{(\omega)}$: seria după o transformare Box–Cox cu parametrul ω; ℓ_t: nivelul; b_t: panta; ϕ: factorul de amortizare; $s_t^{(i)}$: componenta sezonieră i
  - fiecare perioadă m_i are K_i armonici $s_{j,t}^{(i)}$, fiecare o pereche rotitoare (cosinus, sinus) actualizată de erori: un tipar Fourier care poate evolua
- Alegeri automate cu AIC: Box–Cox sau nu, trend, trend amortizat sau fără trend, valorile K_i, ordinele ARMA
  - consumul zilnic pînă în iunie 2025, perioadele 7 și 365,25 (fără Box–Cox, fără trend): $K = 3$ armonici săptămînale și $K = 9$ anuale, ARMA(1,2), 28 de stări
- Limite: fără regresori, deci fără sărbători sau Paște; lent pe serii lungi; pachetul Python `tbats`

Prophet

- Taylor și Letham (2018) (Meta): o regresie în timp, $y(t) = g(t) + s(t) + h(t) + \varepsilon_t$
  - $g(t)$: trend liniar pe porțiuni (sau logistic) a cărui pantă se poate schimba în **puncte de schimbare**; o distribuție a priori Laplace aduce cele mai multe schimbări la zero
  - $s(t)$: termeni Fourier, implicit $K = 10$ pentru an și $K = 3$ pentru săptămînă
  - $h(t)$: un efect pentru fiecare sărbătoare (cu ferestre opționale în jurul ei), dintr-o listă de date
- Estimat cu Stan (maximul a posteriori); intervalele simulează punctele de schimbare viitoare și zgomotul
  - fără parte autoregresivă: eroarea de azi nu modifică prognoza de mîine
- Construit pentru multe serii economice cu efecte puternice de calendar și date lipsă; pachetul Python `prophet`

Componentele Prophet pentru consumul zilnic



⊡ Prophet cu sărbătorile legale din România, consumul zilnic 2022 – iunie 2025

TSA_ch4_tbats_prophet

Interpretarea componentelor Prophet

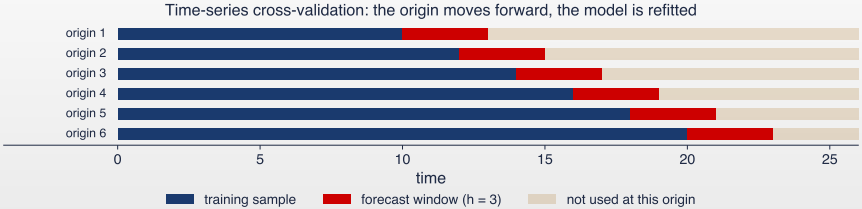
- Trendul de la 6,89 la 6,07 GW, mai plat după 2023; 25 de puncte de schimbare candidate, cele mai multe aduse la zero
  - săptămînal: miercuri $+0{,}27$, sîmbătă $-0{,}32$, duminică $-0{,}80$ GW
- Anual: maximul la începutul lui decembrie ($+0{,}93$ GW, 7 decembrie 2024), minimul la mijlocul lui septembrie ($-0{,}62$ GW, 18 septembrie 2024), arătat pentru 2024
- Sărbătorile: Crăciunul $-1{,}24$, Anul Nou $-1{,}10$, Paștele $-1{,}04$ GW; Ziua Copilului doar $-0{,}37$ GW
- Fiecare componentă poate fi citită și verificată cu ceea ce știm despre economie

Recapitulare: DHR, TBATS și Prophet comparate

	DHR	TBATS	Prophet
Sezonalitatea	Fourier, fixă	Fourier, evolutivă	Fourier, fixă
Mai multe perioade	da	da	da
Sărbători, regresori	da	nu	da
Dinamica pe termen scurt	erori ARMA	erori ARMA	niciuna
Trendul	prin erori (ARIMA)	local (Holt)	liniar pe porțiuni
Viteza	rapidă	lentă	rapidă

- În esență, toate trei sînt regresii pe sinusuri și cosinusuri; diferă prin ceea ce mai modelează
- Care dintre ele prognozează mai bine este o întrebare empirică: secțiunea următoare îi răspunde în afara eșantionului

Validarea încrucișată pentru serii de timp



- Origine de prognoză mobilă (Tashman (2000); FPP3, secțiunea 5.10): estimăm pe datele pînă la origine, prognozăm h pași, mutăm originea, reestimăm
- Fereastră care se extinde (în figură) sau fereastră glisantă de lungime fixă; nu folosim niciodată date viitoare la estimare

TSA_ch4_evaluation

Interpretarea schemei de evaluare

- Consumul zilnic: 26 de origini la 14 zile, de la 30 iunie 2025 la 15 iunie 2026; orizontul $h = 14$ zile
  - un an de prognoze în afara eșantionului, toate anotimpurile și toate sărbătorile, inclusiv Paștele din 2026
- Șapte metode, toate reestimate la fiecare origine
  - sezonieră naivă (perioada 7); ETS (Holt–Winters, trend amortizat, perioada 7); SARIMA$(2,0,1)(0,1,1)_7$; DHR; TBATS; Prophet
  - Combinarea: media simplă a prognozelor celor cinci modele (toate, cu excepția celei sezoniere naive)
- Ordinele SARIMA și ale erorilor DHR se aleg o singură dată, cu AIC, pe datele dinaintea primei origini: fără a privi perioada de test
- Validarea încrucișată care ignoră ordinea în timp (grupuri aleatoare) este validă doar pentru autoregresii pure cu erori de tip zgomot alb (Bergmeir, Hyndman și Koo (2018))

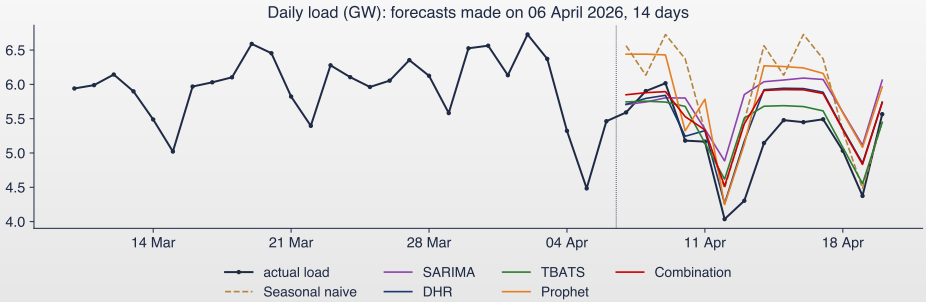
Testul Diebold–Mariano

- Diebold și Mariano (1995): este diferența de acuratețe dintre două prognoze mai mare decît ar rezulta din întîmplare?
  - $e_{\mathbf{1}t}$, $e_{\mathbf{2}t}$: erorile celor două prognoze; L: funcția de pierdere
  - diferența de pierdere $d_t = L(e_{\mathbf{1}t}) - L(e_{\mathbf{2}t})$, $L(e) = e^{\mathbf{2}}$ sau $|e|$; $H_{\mathbf{0}}$: $E[d_t] = 0$
- $\mathrm{DM} = \dfrac{\bar{d}}{\sqrt{\hat{V}/n}}$, $\hat{V} = \hat{\gamma}_d(0) + 2\sum_{k=\mathbf{1}}^{h-\mathbf{1}} \hat{\gamma}_d(k)$
  - $\bar{d}$: media celor n diferențe; $\hat{\gamma}_d(k)$: autocovarianța lor la lagul k; un DM negativ favorizează prognoza 1
  - erorile la h pași se suprapun, deci d_t este autocorelat pînă la lagul $h-1$
  - asimptotic $N(0,1)$; eșantioane mici: $\mathrm{HLN} = \sqrt{\frac{n+\mathbf{1}-\mathbf{2}h+h(h-\mathbf{1})/n}{n}}$ DM, comparat cu $t(n-1)$ (Harvey, Leybourne și Newbold (1997))
- Schema noastră: o valoare pe origine, d_j = diferența medie de pierdere în fereastra de 14 zile; ferestrele nu se suprapun, deci $h = 1$ în formulă
  - testul compară prognoze, nu modele: zgomotul reestimării face parte din ceea ce se testează

Exemplu rezolvat: un test Diebold–Mariano

- Cinci ferestre; diferențele medii ale erorilor pătratice, metoda 1 minus metoda 2: $-0{,}30; 0{,}10; -0{,}50; -0{,}20; -0{,}40$
  - $\bar{d} = -0{,}26$: metoda 1 are pierderi mai mici în patru ferestre din cinci
- $\hat{\gamma}_d(0) = \frac{1}{5}\sum(d_j - \bar{d})^2 = 0{,}042$
  - $\mathrm{DM} = -0{,}26/\sqrt{0{,}042/5} = -2{,}82$; $\mathrm{HLN} = \sqrt{4/5} \cdot \mathrm{DM} = -2{,}53$
- $|-2{,}53| < t_{0{,}975}(4) = 2{,}78$ ($\mathrm{p} = 0{,}065$): acuratețea egală nu se respinge la 5%
  - valoarea critică a distribuției Normale, 1,96, ar fi dus la respingere: cu puține ferestre, corecția contează

Prognoze dintr-o singură origine, peste Paștele din 2026



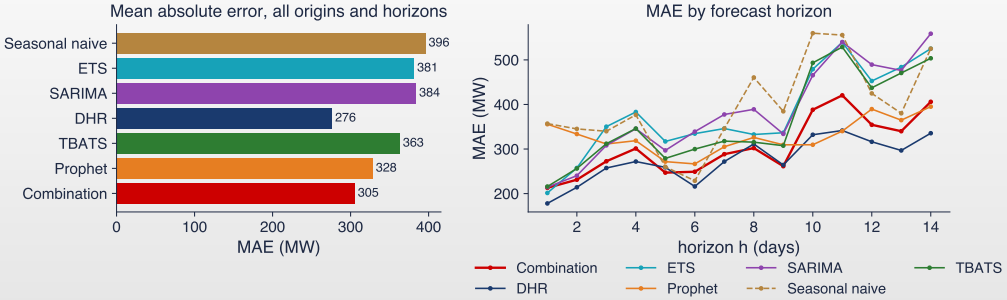
- Originea 6 aprilie 2026, 14 zile înainte; Duminica Paștelui ortodox a fost pe 12 aprilie 2026; ETS nu este desenat (MAE-ul lui se află pe slide-ul următor)

TSA_ch4_evaluation

Interpretarea unei ferestre de prognoză

- MAE în această fereastră (MW): ETS 299, TBATS 309, DHR 344, Combinarea 394, SARIMA 584, Prophet 616, sezonieră naivă 686
  - metoda sezonieră naivă copiază săptămîna normală anterioară în săptămîna Paștelui
- Toate modelele văd scăderea de duminică; DHR și Prophet coboară și Vinerea Mare și a doua zi de Paște, pe care modelele săptămînale le tratează ca zile lucrătoare
- După Paște, consumul rămîne sub toate prognozele pînă la sfîrșitul săptămînii; ETS și TBATS, al căror nivel urmează ultimele zile, au aici erorile cele mai mici
- O singură fereastră este un caz izolat; slide-urile următoare fac media pe toate cele 26 de ferestre

Rezultatele validării încrucișate pentru consumul zilnic



- 26 de origini × 14 orizonturi pentru fiecare metodă; stînga: MAE totală; dreapta: MAE pe orizonturi

TSA_ch4_evaluation

Interpretarea rezultatelor validării încrucișate

Metoda	MAE	RMSE	MASE	DM față de s. naivă	DM față de DHR
Sezonieră naivă	396	529	1,28	–	4,42 ($<0,001$)
ETS	381	492	1,23	$-0,80$ (0,431)	3,08 (0,005)
SARIMA	384	477	1,24	$-1,23$ (0,231)	3,29 (0,003)
DHR	**276**	**350**	**0,89**	$-4,42$ ($<0,001$)	–
TBATS	363	468	1,18	$-1,50$ (0,145)	3,02 (0,006)
Prophet	328	412	1,06	$-2,99$ (0,006)	2,76 (0,011)
Combinarea	305	380	0,99	$-3,82$ ($<0,001$)	1,89 (0,070)

- MAE și RMSE în MW; MASE (Hyndman și Koehler (2006)) = MAE / 309 MW, MAE în eșantion a metodei sezoniere naive săptămînale; DM: statistica HLN (p-value-ul) pe media erorilor pătratice din fiecare fereastră, metoda de pe rînd minus metoda de pe coloană
- DHR cîștigă; împreună cu combinarea, este singura metodă cu MASE sub 1; Prophet este al doilea model individual: singurele două care includ sărbătorile
- ETS, SARIMA și TBATS nu sînt semnificativ mai bune decît metoda sezonieră naivă

Un al doilea test: PIB-ul României

Metoda	MAE, $h = 1$	MAE, $h = 4$	fără 2019–2020, $h = 1$	fără 2019–2020, $h = 4$
SARIMA (airline)	1,69	2,70	1,32	1,89
ETS	1,76	2,84	1,41	2,44
Sezonieră naivă	3,81	3,74	3,80	3,49
Combinarea (SARIMA, ETS)	1,70	2,65	1,34	2,02

- $100 \ln Y$, erori în %; 51 de origini trimestriale, T4 2012–T2 2025, fereastră care se extinde; 43 de origini fără 2019–2020
- DM (HLN), SARIMA față de ETS: $h = 1$: $-1{,}46$ ($p = 0{,}150$); $h = 4$: $0{,}50$ ($p = 0{,}620$)
- SARIMA față de metoda sezonieră naivă: $h = 1$: $-4{,}00$ ($p < 0{,}001$); $h = 4$: $-0{,}54$ ($p = 0{,}591$)
- La un an înainte, erorile din pandemie domină și nicio diferență nu este semnificativă; fără ele, SARIMA are cea mai mică MAE la $h = 4$

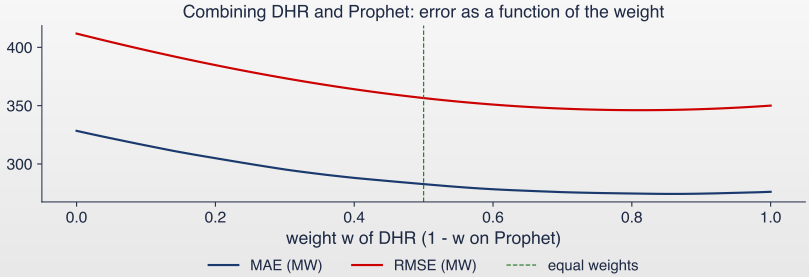
Recapitulare: evaluarea prognozelor

- Evaluați în afara eșantionului, cu origini mobile; alegeți ordinele modelelor înaintea perioadei de test
- MASE compară cu metoda sezonieră naivă și între serii
- Testul DM (cu corecția HLN) arată dacă un cîștig este mai mult decît un efect al întîmplării
- Consumul zilnic: modelele care includ calendarul (DHR, Prophet) cîștigă; pentru PIB, SARIMA și ETS sînt apropiate

Avantajul combinării prognozelor

- Bates și Granger (1969): o medie ponderată a două prognoze nedeplasate, $f_c = wf_1 + (1-w)f_2$
  - w: ponderea prognozei 1; σ_1, σ_2: abaterile standard ale celor două erori; ρ: corelația lor
  - $\text{MSE}(w) = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho\sigma_1\sigma_2$
  - ponderea optimă $w^* = \dfrac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$; cîștigul este mare cînd ρ este mic
- Exemplu: $\sigma_1 = 1$, $\sigma_2 = 1{,}2$, $\rho = 0{,}75$: $\text{MSE}(0{,}5) = 1{,}060$; $w^* = 0{,}84$, $\text{MSE}(w^*) = 0{,}984$
  - erori puternic corelate: media cu ponderi egale este mai slabă decît prognoza mai bună ($\sigma_1^2 = 1$), iar chiar ponderea optimă cîștigă puțin
- **Paradoxul combinării prognozelor** (Stock și Watson (2004); Smith și Wallis (2009)): ponderile estimate sînt rareori mai bune decît ponderile egale
  - ponderile trebuie estimate, iar eroarea lor de estimare consumă cîștigul teoretic

Combinarea DHR și Prophet



⊡ Erorile din validarea încrucișată pentru consumul zilnic; w este ponderea DHR

TSA_ch4_evaluation

Interpretarea combinării

- MAE: doar Prophet (w = 0) 328 MW, doar DHR (w = 1) 276 MW, ponderi egale 283 MW
  - cea mai bună retrospectiv: $w = 0{,}86$, MAE 274 MW, un cîștig mic ales după ce am văzut datele de test
- Erorile celor două metode sînt puternic corelate ($\rho = 0{,}75$): ambele ratează aceleași valuri de frig
- Media celor cinci modele (MAE 305) este a doua, după DHR, și mai bună decît patru dintre cele cinci componente: o asigurare ieftină cînd cîștigătorul nu este cunoscut dinainte
- În competiția M4 (Makridakis, Spiliotis și Assimakopoulos (2020)), cele mai multe dintre metodele de top au fost combinații; o sinteză: Wang et al. (2023)

Contribuția posibilă a AI

- **Cod**: o primă versiune a unei bucle cu origini mobile care reestimează SARIMA, DHR, TBATS și Prophet și păstrează erorile pe orizonturi
- **Calendare**: liste de sărbători și sărbători mobile pentru alte țări, care trebuie verificate cu sursele oficiale
- **Explicații**: o a doua deducere a ACF pentru modelul airline sau a funcției de prognoză pe termen lung
- Exemplu de prompt
  - `Write Python code that forecasts daily electricity load 14 days ahead with a regression on annual Fourier terms, weekday dummies and Romanian public holidays (Orthodox Easter from dateutil) with ARMA errors, and evaluates it against SARIMA by rolling origins and a Diebold-Mariano test.`

Verificări necesare

- Data Paștelui: ortodox, nu occidental (`EASTER_ORTHODOX`); în 2024 cele două diferă cu cinci săptămîni
- Fără informații din viitor: variabilele, ordinele modelelor și listele de sărbători trebuie să fie disponibile la fiecare origine
- Convențiile de semn pentru θ și Θ în program și gradele de libertate ale testului Ljung–Box
- Date ajustate sau neajustate: SARIMA pe NSA; ratele de creștere pe varianta potrivită
- Că testul DM folosește pierderile din aceleași ferestre pentru ambele metode și corecția HLN pentru n mic
- Fiecare referință citată: trebuie să existe; verificați DOI-ul

Idei de reținut

- Sezonalitatea se repetă cu o perioadă cunoscută; poate fi deterministă (variabile dummy, Fourier) sau stochastică (Δ_s)
- SARIMA înmulțește polinoame obișnuite și sezoniere; modelul airline se potrivește multor serii, cu doi parametri
- Datele oficiale: SCA pentru creșterea pe termen scurt, NSA cu regresori de calendar pentru prognoza SARIMA
- Efectele de calendar (zile lucrătoare, Paștele ortodox, sărbători) sînt cunoscute dinainte și contează adesea mai mult decît clasa de modele
- Sezonalitatea multiplă: MSTL, DHR, TBATS, Prophet; pentru consumul zilnic din România, DHR cîștigă în afara eșantionului
- Judecați prognozele cu origini mobile, MASE și testul DM; combinați-le cînd cîștigătorul este incert

Formule de reținut

Mărimea	Formula
Diferența sezonieră	$\Delta_s y_t = (1-L^s)y_t, \quad \Delta\Delta_s = 1 - L - L^s + L^{s+1}$
SARIMA	$\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D y_t = \theta(L)\Theta(L^s)\varepsilon_t$
Airline	$\rho_1 = \frac{\theta}{1+\theta^2}, \quad \rho_s = \frac{\Theta}{1+\Theta^2}, \quad \rho_{s\pm1} = \rho_1\rho_s$
Fourier	$\sum_{k=1}^{K}[\alpha_k \sin(2\pi kt/m) + \beta_k \cos(2\pi kt/m)]$
Prophet	$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$
MASE	$\frac{1}{h}\sum \lvert e_{T+j}\rvert \,/\, \frac{1}{T-m}\sum \lvert y_t - y_{t-m}\rvert$
DM	$\bar{d}/\sqrt{\hat{V}/n}, \quad d_t = L(e_{1t}) - L(e_{2t})$
Combinarea	$w^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}$

Autoevaluare

- **Întrebare**: cîți parametri are SARIMA$(1,1,1)(0,1,1)_{\mathbf{12}}$ și la ce laguri apare ε?
  - **Răspuns**: $\phi, \theta, \Theta, \sigma^{\mathbf{2}}$, patru; ε la lagurile 0, 1, 12, 13 (ultimul cu coeficientul $\theta\Theta$)
- **Întrebare**: PIB-ul neajustat scade cu 40% din T4 în T1. Este o recesiune?
  - **Răspuns**: nu; este tiparul sezonier; priviți seria SCA sau creșterea față de anul anterior
- **Întrebare**: un test DM cu 26 de ferestre dă HLN $= -1{,}93$. Este prima prognoză semnificativ mai bună la 5%?
  - **Răspuns**: nu: $|-1{,}93| < t_{\mathbf{0{,}975}}(25) = 2{,}06$; este mai bună în medie, dar nu semnificativ
- **Idee de proiect**: prognozați consumul orar de electricitate din România cu o DHR (termeni Fourier zilnici, săptămînali și anuali, sărbători, temperatură), comparată cu TBATS, Prophet și o combinare, prin origini mobile și teste DM
  - Urmează: Capitolul 5, volatilitatea condiționată: modelele ARCH și GARCH

Bibliografie (1/3)

- Bandara, K., Hyndman, R. J., & Bergmeir, C. (2025). MSTL: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns. *International Journal of Operational Research*, 52(1), 79–98. doi:10.1504/ijor.2025.143957
- Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4), 451–468. doi:10.1057/jors.1969.103
- Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83. doi:10.1016/j.csda.2017.11.003
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26(2), 211–243. doi:10.1111/j.2517-6161.1964.tb00553.x
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control* (4th ed.). Wiley. doi:10.1002/9781118619193
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (3rd ed.). Springer. doi:10.1007/978-3-319-29854-2
- Canova, F., & Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business & Economic Statistics*, 13(3), 237–252. doi:10.1080/07350015.1995.10524598
- Dagum, E. B., & Bianconcini, S. (2016). *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation*. Springer. doi:10.1007/978-3-319-31822-6
- De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496), 1513–1527. doi:10.1198/jasa.2011.tm09771
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. doi:10.1080/07350015.1995.10524599

Bibliografie (2/3)

- Eurostat (2024). *ESS Guidelines on Seasonal Adjustment, 2024 edition*. Publications Office of the European Union. ec.europa.eu/eurostat
- Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2), 127–152. doi:10.1080/07350015.1998.10524743
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291. doi:10.1016/S0169-2070(96)00719-4
- Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting*, 32(3), 896–913. doi:10.1016/j.ijforecast.2016.02.001
- Huang, C., & Petukhina, A. (2022). *Applied Time Series Analysis and Forecasting with Python*. Springer. doi:10.1007/978-3-031-13584-2
- 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. doi:10.1016/0304-4076(90)90080-D
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. otexts.com/fpp3
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1–22. doi:10.18637/jss.v027.i03
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. doi:10.1016/j.ijforecast.2006.03.001

Bibliografie (3/3)

- Ladiray, D., & Quenneville, B. (2001). *Seasonal Adjustment with the X-11 Method*. Lecture Notes in Statistics 158. Springer. doi:10.1007/978-1-4613-0175-2
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303. doi:10.1093/biomet/65.2.297
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74. doi:10.1016/j.ijforecast.2019.04.014
- Osborn, D. R., Chui, A. P. L., Smith, J. P., & Birchenhall, C. R. (1988). Seasonality and the order of integration for consumption. *Oxford Bulletin of Economics and Statistics*, 50(4), 361–377. doi:10.1111/j.1468-0084.1988.mp50004002.x
- Petropoulos, F., et al. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, 38(3), 705–871. doi:10.1016/j.ijforecast.2021.11.001
- Smith, J., & Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71(3), 331–355. doi:10.1111/j.1468-0084.2008.00541.x
- Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6), 405–430. doi:10.1002/for.928
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4), 437–450. doi:10.1016/S0169-2070(00)00065-0
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. doi:10.1080/00031305.2017.1380080
- Wang, X., Hyndman, R. J., Li, F., & Kang, Y. (2023). Forecast combinations: An over 50-year review. *International Journal of Forecasting*, 39(4), 1518–1547. doi:10.1016/j.ijforecast.2022.11.005