








Time Series Analysis

Chapter 4: Seasonality and forecasting: SARIMA, TBATS, Prophet



Daniel Traian PELE

Bucharest University of Economic Studies
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Romanian Academy, Institute for Economic Forecasting
MSCA Digital Finance

Acronyms used in this material (1/2)

- **ACF**: Autocorrelation Function
- **AI**: Artificial Intelligence
- **AIC**: Akaike Information Criterion
- **AR**: AutoRegressive (model)
- **ARCH**: AutoRegressive Conditional Heteroskedasticity
- **ARIMA**: AutoRegressive Integrated Moving Average (model)
- **ARMA**: AutoRegressive Moving Average (model)
- **BIC**: Bayesian Information Criterion
- **CA**: Calendar Adjusted
- **DHR**: Dynamic Harmonic Regression
- **DM**: Diebold–Mariano (test)
- **ENTSO-E**: European Network of Transmission System Operators for Electricity
- **ETS**: Error, Trend, Seasonality (exponential smoothing state space models)
- **EU**: European Union
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **GDP**: Gross Domestic Product
- **HEGY**: Hylleberg–Engle–Granger–Yoo (seasonal unit-root test)
- **HICP**: Harmonised Index of Consumer Prices
- **HLN**: Harvey–Leybourne–Newbold (small-sample correction of the DM test)

Acronyms used in this material (2/2)

- **KPSS**: Kwiatkowski–Phillips–Schmidt–Shin (test)
- **LB**: Ljung–Box (test)
- **MA**: Moving Average
- **MAE**: Mean Absolute Error
- **MAPE**: Mean Absolute Percentage Error
- **MASE**: Mean Absolute Scaled Error
- **MSTL**: Multiple Seasonal-Trend decomposition using Loess
- **NSA**: Not Seasonally Adjusted
- **OCSB**: Osborn–Chui–Smith–Birchenhall (seasonal unit-root test)
- **PACF**: Partial AutoCorrelation Function
- **RMSE**: Root Mean Squared Error
- **SARIMA**: Seasonal ARIMA (model)
- **SCA**: Seasonally and Calendar Adjusted
- **SEATS**: Signal Extraction in ARIMA Time Series
- **STL**: Seasonal and Trend decomposition using Loess
- **TBATS**: Trigonometric seasonality, Box–Cox, ARMA errors, Trend, Seasonal components (model)
- **TRAMO-SEATS**: Time series Regression with ARIMA noise, Missing values and Outliers; Signal Extraction in ARIMA Time Series
- **US**: United States
- **UTC**: Coordinated Universal Time (Temps Universel Coordonné)
- **VAT**: Value Added Tax

Today's question and route

- **Question**: many economic series repeat a pattern every year, week or day; how do we model this pattern, forecast with it, and decide which forecast is better?
  - three answers: **difference** the pattern away (SARIMA), **model** it with regressors (Fourier terms, TBATS, Prophet), or **remove** it (seasonal adjustment)
- **Route** of the chapter
  - seasonal differences and seasonal unit roots (HEGY, Canova–Hansen, OCSB)
  - SARIMA$(p, d, q)(P, D, Q)_s$: identification, estimation, diagnostics, forecasts; the airline model and Romanian GDP
  - seasonal adjustment in practice (X-13ARIMA-SEATS, Eurostat), Fourier terms and calendar effects (Orthodox Easter)
  - multiple seasonality in electricity load: MSTL, dynamic harmonic regression, TBATS, Prophet
  - forecast evaluation (time-series cross-validation, Diebold–Mariano test) and forecast combination
- It builds on Chapter 0 (components, Holt–Winters, MASE) and Chapters 2–3 (ARMA, unit roots, ARIMA): the same ideas, at the seasonal lag s
- Seminar 4 comes before this lecture: it gives the formulas; here we derive them and use them on data

Learning outcomes

- Recognise deterministic and stochastic seasonality and choose seasonal differencing with HEGY, Canova–Hansen and OCSB
- Write, identify, estimate, check and forecast a SARIMA$(p, d, q)(P, D, Q)_s$ model, and expand its polynomials
- Read seasonally adjusted and unadjusted official series correctly, and explain what X-13ARIMA-SEATS does
- Model calendar effects (working days, Orthodox Easter) and multiple seasonality with Fourier terms, MSTL, TBATS and Prophet
- Compare forecasts with time-series cross-validation, MASE and the Diebold–Mariano test, and combine them

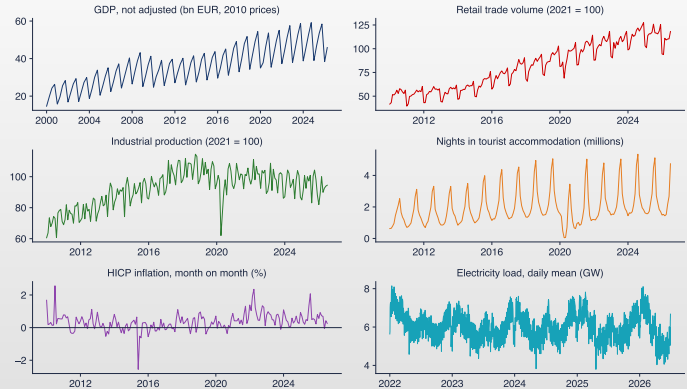
Reading and tools

- Textbook: Huang and Petukhina (2022), Ch. 4–5 (ARIMA and nonstationary models)
  - companion, free online: Hyndman and Athanasopoulos (2021), Ch. 9 (FPP3, Section 9.9), Ch. 10 (FPP3, Section 10.5), Ch. 12 (FPP3, Section 12.1) and Ch. 13
- The classic: Box, Jenkins and Reinsel (2008), Ch. 9 (seasonal models); seasonal adjustment: Ladiray and Quenneville (2001), Dagum and Bianconcini (2016), Eurostat (2024)
- Python Quantlets of this chapter: Quantlets/Ch_04
  - `statsmodels` (`SARIMAX`, `MSTL`, `ExponentialSmoothing`); optional packages `tbats` and `prophet`
- Lecture notebook: open in Google Colab
- Video course: Applied Time Series Analysis with Python

Seasonality: definition and sources

- **Seasonality**: a pattern that repeats with a fixed, known **period** s (observations per cycle)
  - $s = 4$ (quarters), $s = 12$ (months), $s = 7$ (days in a week), $s = 24$ and $s = 168$ (hours in a day and in a week)
  - unlike a **cycle**, whose length is irregular and unknown (Chapter 0)
- **Sources**
  - weather: heating in winter, the beach in August, harvests
  - the calendar: working days, public holidays, **moving holidays** such as Orthodox Easter
  - institutions and habits: the school year, tax deadlines, end-of-year bonuses, the working week
- **Additive** ($y_t = T_t + S_t + R_t$) or **multiplicative** ($y_t = T_t S_t R_t$, additive after logs): Chapter 0; here we work with logs when the swings grow with the level

Seasonality in six Romanian series



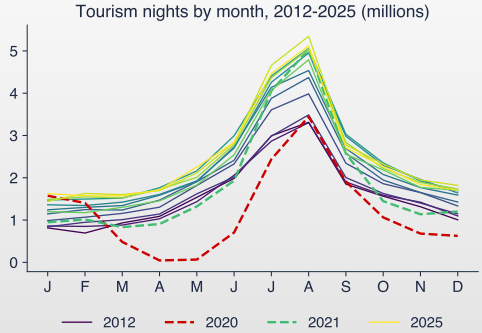
- Eurostat: GDP (quarterly, not adjusted), retail trade, industrial production, tourism nights, HICP; ENTSO-E: daily electricity load

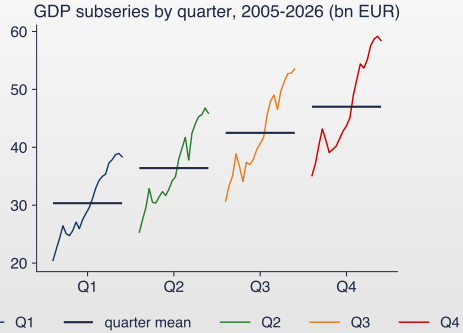
TSA_ch4_seasonal_data

Interpreting the six series

- GDP: the mean of Q4 is 57% above the mean of Q1 (2000–2026); the swing grows with the level: multiplicative, take logs
  - quarter means (bn EUR, 2010 prices): 27.8, 33.6, 39.5, 43.6: construction and agriculture in Q2–Q3
- Tourism: August has 3.7 times the nights of January (2015–2019); the 2020 lockdown breaks the pattern
- HICP: monthly inflation is lower in June (-0.37% on average, 2010–2019, fresh food) and highest in January (0.575%, new prices and excise duties), high again in October (0.49%)
- Electricity: two patterns at once, the week (dense oscillation) and the year (winter peaks): **multiple seasonality**, Section 8

Seasonal plot and subseries plot





⊡ Left: one line per year (seasonal plot); right: each quarter of GDP followed over the years, with its mean (subseries plot)

TSA_ch4_seasonal_data

Interpreting the seasonal plots

- Tourism: one shape every year, a summer peak and a winter trough
  - 2019: 5.06 million nights in August, 1.47 million in January
  - April 2020: 0.045 million (1.76 million in April 2019): an outlier, not a new season
- GDP subseries: all four quarters rise in parallel
  - a stable seasonal shape on top of a common trend
  - if the lines crossed, the seasonal pattern would be changing: stochastic seasonality (next section)



Mamaia, on the Black Sea, September 2013

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

Recap: Seasonality in economic data

- Seasonality repeats with a known period s; a cycle does not
- Weather, the calendar and institutions create it; moving holidays shift it between months
- Romanian GDP, tourism, inflation and electricity load are all seasonal; electricity has several periods at once
- Seasonal and subseries plots show whether the pattern is stable

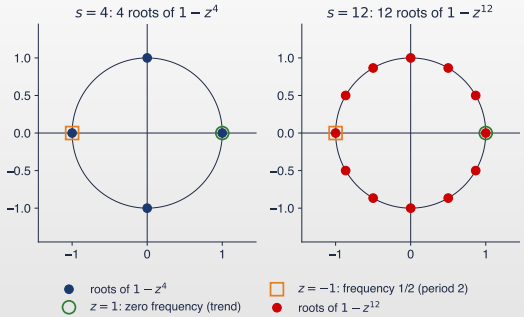
The seasonal difference

- **Seasonal difference**: $\Delta_s y_t = (1 - L^s) y_t = y_t - y_{t-s}$
  - with logs: $100\,\Delta_4 \ln Y_t$ = growth over the same quarter of the previous year, in %
- A **fixed** seasonal pattern $S_t = S_{t-s}$ plus a linear trend: $y_t = a + bt + S_t + \varepsilon_t$
  - a: the intercept; b: the slope of the trend; S_t: the seasonal effect of period t
  - $\Delta_s y_t = bs + \varepsilon_t - \varepsilon_{t-s}$: Δ_s removes the pattern and turns the trend into a constant
  - but it creates the non-invertible MA term $\varepsilon_t - \varepsilon_{t-s}$: **seasonal over-differencing** (as in Chapter 3)
- $\Delta\Delta_s = (1 - L)(1 - L^s) = 1 - L - L^s + L^{s+1}$: needed when the trend is stochastic as well
  - the order of the two differences does not matter: the operators commute

Deterministic and stochastic seasonality

- **Deterministic**: $y_t = \sum_{j=1}^{s} \gamma_j D_{jt} + u_t$, with $D_{jt} = 1$ in season j and u_t stationary
  - the seasonal means γ_j never change; the right tool: seasonal dummies or Fourier terms (Section 7)
- **Stochastic**: the **seasonal random walk** $y_t = y_{t-s} + \varepsilon_t$
  - each season is its own random walk: s separate walks interleaved
  - $\mathrm{Var}(y_t)$ grows without bound; the pattern drifts, and "summer can become winter"
  - the right tool: the seasonal difference Δ_s
- The choice matters: differencing a deterministic pattern over-differences; modelling a stochastic one with dummies leaves a unit root in the residuals

The roots of $1 - z^s$: seasonal unit roots



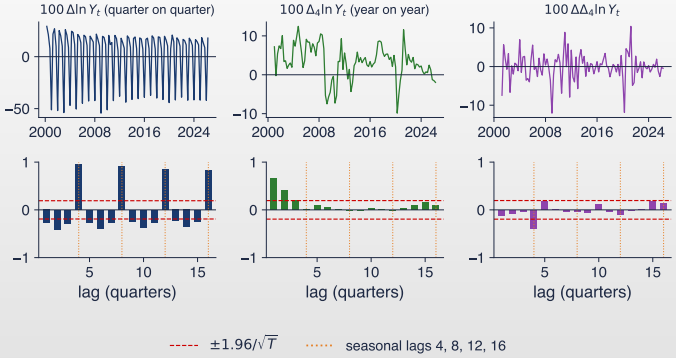
- $1 - z^4 = (1 - z)(1 + z)(1 + z^2)$: roots 1, -1, $\pm i$, all on the unit circle

TSA_ch4_seasonal_roots

Interpreting the seasonal roots

- Δ_s imposes s unit roots at once, one for each **frequency** k/s, $k = 0, \ldots, s-1$
  - quarterly: $z = 1$ (frequency 0, the trend), $z = -1$ (half a cycle per quarter: period 2 quarters), $z = \pm i$ (period 4 quarters: the annual cycle)
- A series may have some of these roots and not others: then Δ_4 over-differences at the frequencies it does not need
- Seasonal unit-root tests ask, frequency by frequency, which roots are present: the HEGY test (next slides)

Romanian GDP: regular and seasonal differences



- ln of real GDP, not adjusted, 106 quarters to 2026Q2; top: three transformations; bottom: their sample ACF up to 16 quarters

TSA_ch4_seasonal_roots

Interpreting the three transformations

- $\Delta \ln Y$: dominated by the season: ACF 0.95 at lag 4 and 0.90 at lag 8, decaying very slowly; sd 26.5%
  - the seasonal analogue of the slow ACF decay of a random walk (Chapter 3)
- $\Delta_4 \ln Y$ (year-on-year growth): sd 4.4%; ACF 0.66 at lag 1, 0.01 at lag 4: an AR-type pattern
- $\Delta\Delta_4 \ln Y$: ACF -0.11 at lag 1 and -0.40 at lag 4 (band ± 0.20)
  - one negative spike at the seasonal lag: a seasonal MA(1), the signature of the airline model (Section 3)

The HEGY test

- Hylleberg, Engle, Granger and Yoo (1990): one regression tests every seasonal root of a quarterly series
  - filters: $y_{1t} = (1 + L + L^2 + L^3)y_t$, $y_{2t} = -(1 - L + L^2 - L^3)y_t$, $y_{3t} = -(1 - L^2)y_t$
- $\Delta_4 y_t = \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{3,t-1} +$ deterministic terms + lags of $\Delta_4 y_t + \varepsilon_t$
  - y_{1t}, y_{2t}, y_{3t}: the series filtered so that each keeps one seasonal frequency; π_k: the coefficient of each frequency
  - H_0: $\pi_1 = 0$ (root 1), t-test; $\pi_2 = 0$ (root -1), t-test; $\pi_3 = \pi_4 = 0$ (roots $\pm i$), F-test
  - non-standard distributions, as for Dickey–Fuller: critical values from tables or by simulation
- Rejecting all three: no seasonal differencing needed; rejecting none: Δ_4 is justified



Robert Engle, a co-author of HEGY, 2017

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

Canova–Hansen and OCSB

- **Canova–Hansen** (Canova and Hansen (1995)): the null is **stable** (deterministic) seasonality
  - the seasonal analogue of KPSS (Chapter 3): rejecting means the pattern changes over time
- **OCSB** (Osborn, Chui, Smith and Birchenhall (1988)): the null is a seasonal unit root; a regression of $\Delta\Delta_s y_t$ on $\Delta_s y_{t-1}$ and Δy_{t-s}
  - the default seasonal test of `auto_arima` in the Python package `pmdarima` (`nsdiffs`)
- As in Chapter 3: tests with opposite nulls answer different questions; use them together
  - and look at the data: seasonal plots, the ACF at lags $s, 2s, 3s$, and the size of $\hat{\Theta}$ after Δ_s ($\hat{\Theta} \approx -1$ signals over-differencing)

Seasonal unit-root tests on Romanian data

HEGY, ln GDP, 2000–2019	statistic	5% critical value	decision
$t(\pi_1)$, root 1	-1.97	-3.39	not rejected
$t(\pi_2)$, root -1	-2.19	-2.77	not rejected
$F(\pi_3, \pi_4)$, roots $\pm i$	6.12	6.64	not rejected

Series (logs, 2005–2019)	D by OCSB	D by Canova–Hansen	d by KPSS
GDP (quarterly)	1	0	1
Retail trade	0	0	1
Industrial production	0	0	1
Tourism nights	0	0	1
HICP	0	0	2

- HEGY: constant, seasonal dummies and trend, 1 lag of $\Delta_4 y$ (AIC), $n = 75$; critical values from 4 000 simulations of $\Delta_4 y_t = \varepsilon_t$
- Interpretation: for GDP, HEGY and OCSB support Δ_4 while Canova–Hansen does not reject a stable pattern; with 20 years of quarters both views fit, so we use Δ_4 and let $\hat{\Theta}$ say how stable the pattern is (Section 5)
- Monthly series: both tests choose $D = 0$, a stable pattern; HICP needs $d = 2$ in logs: inflation itself is close to a unit root (Chapter 3)

Recap: Seasonal differencing and seasonal unit roots

- $\Delta_s = 1 - L^s$ has s unit roots on the circle, one per seasonal frequency
- Deterministic seasonality: dummies or Fourier terms; stochastic seasonality: Δ_s
- HEGY and OCSB test for seasonal unit roots; Canova–Hansen tests for a stable pattern
- Romanian GDP: $\Delta\Delta_4 \ln Y$ leaves one spike at lag 4, a seasonal MA

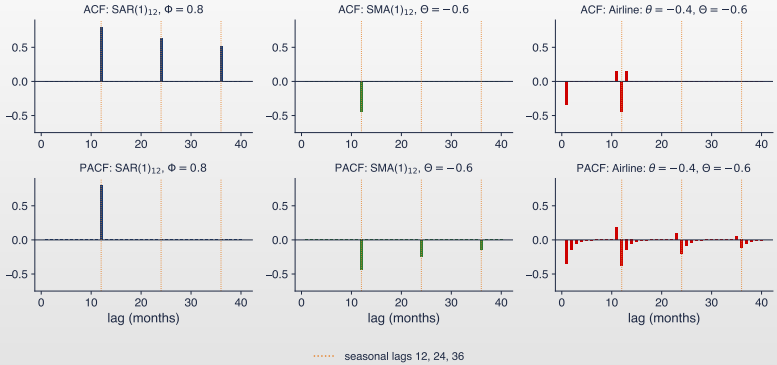
The SARIMA$(p, d, q)(P, D, Q)_s$ model

- $\phi(L)\,\Phi(L^s)\,(1-L)^d(1-L^s)^D\,y_t = c + \theta(L)\,\Theta(L^s)\,\varepsilon_t$, $\varepsilon_t \sim \text{WN}(0, \sigma^2)$
  - regular polynomials $\phi(L)$ (order p), $\theta(L)$ (order q), as in Chapter 2
  - seasonal polynomials in L^s (orders 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: numbers of regular and seasonal differences, usually 0 or 1
- Why multiply? Box, Jenkins and Reinsel (2008): in a monthly series, January depends on December (the regular part) and on last January (the seasonal part)
  - a two-way table: months in the columns, years in the rows; one model along each direction
- Stationarity: the roots of $\phi(z)\Phi(z^s)$ outside the unit circle; invertibility: the roots of $\theta(z)\Theta(z^s)$ outside; for a seasonal AR(1) this means $|\Phi| < 1$

Worked example: the airline model

- SARIMA$(0,1,1)(0,1,1)_{12}$ with $\theta = -0.4$, $\Theta = -0.6$: $(1-L)(1-L^{12})y_t = (1-0.4L)(1-0.6L^{12})\varepsilon_t$
  - left: $y_t - y_{t-1} - y_{t-12} + y_{t-13}$; right: $\varepsilon_t - 0.4\varepsilon_{t-1} - 0.6\varepsilon_{t-12} + 0.24\varepsilon_{t-13}$
  - the coefficient of ε_{t-13}, $\theta\Theta = 0.24$, is not a new parameter: two parameters plus σ^2
- $w_t = \Delta\Delta_{12}y_t$ is an MA(13) with only four nonzero weights; its autocovariances (Chapter 2):
  - $\gamma(0) = \sigma^2(1+\theta^2)(1+\Theta^2) = 1.5776\,\sigma^2$
  - $\rho_1 = \theta/(1+\theta^2) = -0.345$; $\rho_{12} = \Theta/(1+\Theta^2) = -0.441$; $\rho_{11} = \rho_{13} = \rho_1\rho_{12} = 0.152$
  - all other autocorrelations are zero

Theoretical ACF and PACF of three seasonal models



- Monthly, $s = 12$; the airline panel is the stationary part $w_t = \Delta\Delta_{12} y_t$

TSA_ch4_sarima_theory

Interpreting the seasonal patterns

- Seasonal AR(1): the ACF decays geometrically at lags 12, 24, 36 (0.8, 0.64, 0.512); the PACF has one spike at lag 12
  - the AR rule of Chapter 2, read only at multiples of s
- Seasonal MA(1): the ACF cuts off after lag 12; the PACF decays at 12, 24, 36 (-0.242 at lag 24)
- Airline: spikes at 1 (-0.345) and 12 (-0.441), with "satellites" at 11 and 13 (0.152)
  - the satellites are the product $\rho_1\rho_{12}$: the fingerprint of the multiplicative structure

Identification of SARIMA models

After differencing	ACF	PACF
seasonal AR(P)	decays at $s, 2s, \dots$	cuts off after lag Ps
seasonal MA(Q)	cuts off after lag Qs	decays at $s, 2s, \dots$
regular part	lags $1, 2, \dots$ as in Chapter 2	lags $1, 2, \dots$ as in Chapter 2
product of both	satellites at $ks \pm j$	satellites at $ks \pm j$

- **Steps**
  - 1. Stabilise the variance (logs or Box–Cox, Box and Cox (1964)); 2. choose D (seasonal plot, HEGY, OCSB), then d (Chapter 3)
  - 3. Read P, Q at the seasonal lags and p, q at the first lags; 4. keep the orders small: $P, Q \leq 1$ almost always suffice

Estimation and model choice

- **Maximum likelihood** of the differenced series (state-space form, `statsmodels SARIMAX`)
  - $s+1$ observations are lost to $\Delta\Delta_s$; the constant is dropped when $d+D \geq 2$ (it would mean a quadratic trend)
- **Information criteria**: AICc $=$ AIC $+\frac{2k(k+1)}{n-k-1}$ and BIC, compared **only** across models with the same d and D
  - k: the number of estimated parameters; n: the number of observations after differencing
  - differencing changes the data on which the likelihood is computed
- **Automatic SARIMA**: a stepwise search over (p, q, P, Q) by AICc after choosing d and D by tests (Hyndman and Khandakar (2008))
  - in Python: `pmdarima.auto_arima(y, m=12)`; always check the residuals and compare with a benchmark
- **Diagnostics**: Ljung–Box with lags beyond s (e.g. $m = 2s$) and $m-p-q-P-Q$ degrees of freedom (Ljung and Box (1978))

Forecasting with SARIMA

- As for ARIMA (Chapter 3): write the model as an equation for y_t; replace future shocks by 0, future values by their forecasts, past shocks by residuals
  - airline: $\hat{y}_{T+h} = \hat{y}_{T+h-1} + \hat{y}_{T+h-s} - \hat{y}_{T+h-s-1} +$ (MA terms while $h \leq s+1$)
- **Eventual forecast function** of the airline model: a fixed seasonal pattern on a straight line
  - the pattern is an exponentially weighted average of past patterns, with weight $1+\Theta$ on the most recent year
  - $\Theta \to -1$: a fixed (deterministic) pattern; $\Theta = 0$: last year's pattern repeated (seasonal naive)
- Intervals widen with h (two unit roots); for $\ln y$, $\exp(\hat{y}_{T+h})$ is the forecast **median**
  - the mean is $\exp(\hat{y}_{T+h} + \sigma_h^2/2)$, slightly higher; σ_h^2: the forecast error variance at horizon h

Worked example: forecasting a seasonal MA

- Quarterly SARIMA$(0,0,0)(0,1,1)_4$: $y_t = y_{t-4} + \varepsilon_t + \Theta\varepsilon_{t-4}$, $\Theta = -0.6$
  - last year: $y_{T-3}, \ldots, y_T = 100, 120, 130, 140$; residuals $\hat{\varepsilon}_{T-3}, \ldots, \hat{\varepsilon}_T = 2, -1, 0, 3$
- $\hat{y}_{T+1} = y_{T-3} + \Theta\hat{\varepsilon}_{T-3} = 100 - 0.6 \cdot 2 = 98.8$
  - $\hat{y}_{T+2} = 120.6$, $\hat{y}_{T+3} = 130.0$, $\hat{y}_{T+4} = 138.2$
- $\hat{y}_{T+5} = \hat{y}_{T+1} = 98.8$: after one year the MA term is gone and the forecast pattern repeats
  - a large positive surprise last season ($\hat{\varepsilon}_T = 3$) lowers the forecast for that season by $0.6 \cdot 3$: partial correction

Recap: SARIMA models

- SARIMA multiplies regular and seasonal polynomials: $\phi(L)\Phi(L^s)\Delta^d\Delta_s^D y_t = \theta(L)\Theta(L^s)\varepsilon_t$
- Identification: the Chapter 2 rules at lags $s, 2s, \ldots$, plus satellites at $ks \pm j$
- Compare AICc only for the same d and D; check Ljung–Box beyond lag s
- The airline forecast: a seasonal pattern on a line, updated with weight $1 + \Theta$

Box and Jenkins and the airline passengers

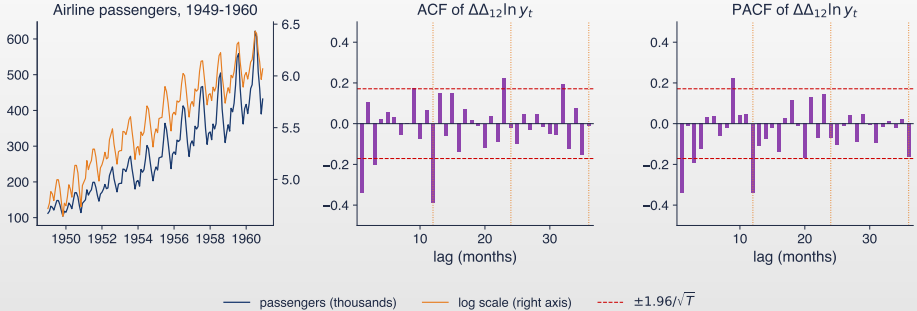
- Box, Jenkins and Reinsel (2008) (first edition 1970) modelled the monthly totals of international airline passengers, 1949–1960 (144 months)
  - growth plus a summer peak whose size grows with the level
  - their model, SARIMA$(0,1,1)(0,1,1)_{\mathbf{12}}$ for $\ln y_t$, has carried the name "airline model" ever since
- Why it lasted
  - two parameters fit many seasonal series: trade, production, retail sales, energy
  - it is the reference model of the regARIMA step of seasonal-adjustment software (Section 6)



The first Boeing 707s for Pan American, Seattle, 1958

Photo: unknown author (1958); public domain; Wikimedia Commons

Identification of the airline model



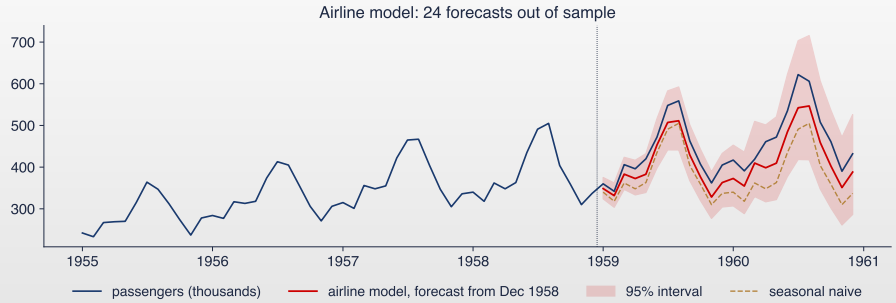
- Left: passengers and their log; middle and right: ACF and PACF of $w_t = \Delta\Delta_{12} \ln y_t$, 131 observations

TSA_ch4_airline

Interpreting the airline correlograms

- The log makes the summer swing constant: the multiplicative pattern becomes additive
- ACF of w_t: -0.34 at lag 1 and -0.39 at lag 12, outside the band ± 0.17
  - a regular MA(1) and a seasonal MA(1); lag 3 (-0.20) is borderline
  - isolated bars at lags 9, 23 and 32 also cross the band: with 36 lags, about two crossings are expected by chance
- The PACF: negative spikes at lags 1, 3 and 12, almost 0 at lag 2, and no clean cut-off after a few lags: consistent with MA terms, not AR terms
- Conclusion: SARIMA$(0,1,1)(0,1,1)_{12}$, exactly the model of Box and Jenkins

Airline model: forecasts out of sample



- Estimated on 1949–1958, forecasts for the 24 months of 1959–1960, back-transformed from $\ln y$

TSA_ch4_airline

Interpreting the airline forecasts

- 1949–1958: $\hat{\theta} = -0.342$ (0.087), $\hat{\Theta} = -0.541$ (0.105); full sample: $\hat{\theta} = -0.402$ (0.073), $\hat{\Theta} = -0.557$ (0.096)
  - full sample: residual sd 3.7% per month; Ljung–Box $Q^*(24)$ on 22 degrees of freedom, p = 0.37
- MAPE over 1959–1960: 8.5% for the airline model, 15.5% for the seasonal naive forecast (which ignores growth)
- December 1960: forecast 388, interval [286, 526], actual 432: inside, but the model under-predicts the 1959–1960 growth
  - the interval width grows from 51 (first month) to 239 (24th month): two unit roots

SARIMA models for Romanian GDP

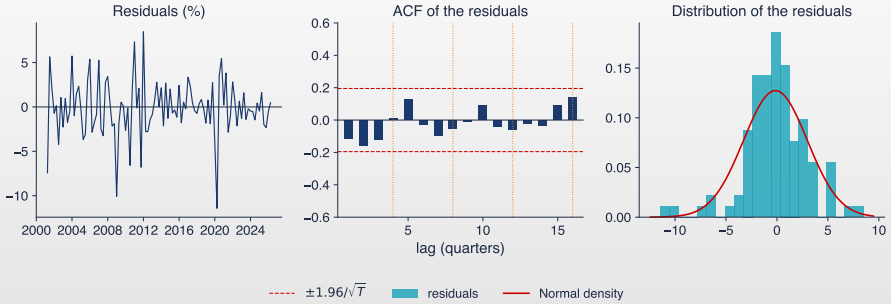
Model	AICc	BIC	LB(8), p	$\hat{\sigma}$ (%)
$(0,1,1)(0,1,1)_4$	520.1	527.6	0.157	3.05
$(0,1,0)(0,1,1)_4$	519.7	**524.8**	0.286	3.08
$(1,1,0)(0,1,1)_4$	520.7	528.3	0.156	3.06
$(1,1,1)(0,1,1)_4$	**518.3**	528.4	0.467	2.98
$(0,1,2)(0,1,1)_4$	518.8	528.8	0.575	3.00
$(0,1,1)(1,1,0)_4$	529.9	537.5	0.003	3.22
$(0,1,1)(1,1,1)_4$	522.2	532.2	0.118	3.05
$(0,1,1)(0,1,0)_4$	545.1	550.2	0.005	3.52

- $y_t = 100 \ln Y_t$, real GDP not adjusted, 2000Q1–2026Q2, $T = 106$; maximum likelihood; all with $d = D = 1$, so the criteria are comparable
- Ljung–Box on 8 lags with $8 - p - q - P - Q$ degrees of freedom

Interpreting the model table

- BIC chooses $(0,1,0)(0,1,1)_4$: only the seasonal MA term; AICc chooses $(1,1,1)(0,1,1)_4$
  - the AICc model has $\hat{\phi} = 0.78$ and $\hat{\theta} = -0.94$: almost a common factor (Chapter 2), little gain
  - the airline model: $\hat{\theta} = -0.171$ (0.109), not significant; $\hat{\Theta} = -0.585$
- A seasonal AR instead of the seasonal MA, or no seasonal term at all: much worse criteria and correlated residuals
- We keep the BIC model: one parameter, $\hat{\Theta} = -0.577$ (0.075), $\hat{\sigma} = 3.08\%$

Diagnostics of the GDP model



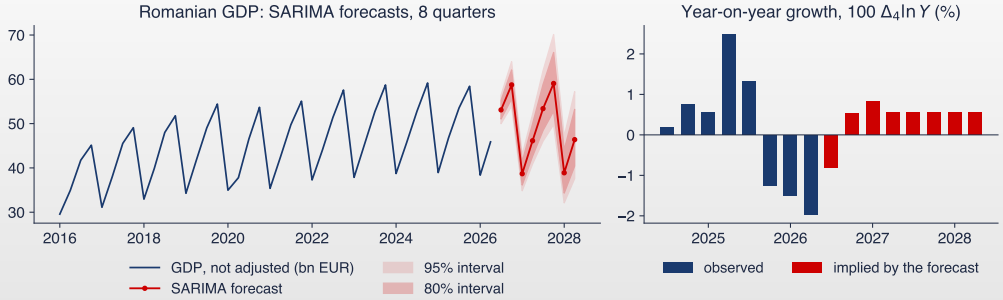
- SARIMA$(0,1,0)(0,1,1)_4$ for $100 \ln Y_t$: residuals, their ACF and their distribution

TSA_ch4_gdp_case

Interpreting the GDP diagnostics

- No autocorrelation left: $Q^*(8) = 8.55$ on 7 degrees of freedom, $p = 0.286$; $Q^*(12)$: $p = 0.526$
  - no spike at the seasonal lags 4, 8, 12, 16: the seasonal structure is captured
- Not Normal: Jarque–Bera 22.0 ($p < 0.001$); the largest residual is -11.4% in 2020Q2 (-3.6 standard deviations): the lockdown
  - without 2020, Jarque–Bera falls to 13.4: an additive-outlier dummy would be the next step

Forecasts of Romanian GDP



- SARIMA$(0,1,0)(0,1,1)_4$, 8 quarters from 2026Q3 to 2028Q2, 80% and 95% intervals; right: implied year-on-year growth

TSA_ch4_gdp_case

Interpreting the GDP forecasts

- Last value: 45.9 bn EUR in 2026Q2; forecasts 53.1, 58.7, 38.7, 46.1 bn EUR for the next four quarters
  - 95% interval for the first quarter [50.0, 56.4], for the eighth [37.6, 57.2]
- Implied year-on-year growth: -0.81%, 0.53%, 0.83%, 0.55%, then a constant 0.55%
  - the model extrapolates the recent weak growth; it knows nothing about fiscal policy or EU funds
- The seasonal pattern of the forecast is a weighted average of past years, with weight $1 + \hat{\Theta} = 0.42$ on the newest year

Recap: Airline passengers and Romanian GDP

- Logs, then $\Delta\Delta_s$, then MA terms at lags 1 and s: the airline recipe works for both series
- For GDP, BIC drops the regular MA term; AICc picks a near-redundant ARMA(1,1)
- Residuals are clean except for the 2020 outlier
- Forecasts repeat a smoothed seasonal pattern on a trend line, with widening intervals

Why official statistics adjust for seasonality

- Users want to know whether the economy **grew this quarter**, not whether Q3 is bigger than Q2
  - 2026Q1: unadjusted GDP changed by -42.1% (log points) from the previous quarter; the adjusted series by -0.06%
  - the first number is pure season; only the second says something about the business cycle
- **Seasonal adjustment**: estimate the seasonal and calendar components and remove them: $y_t^{SA} = y_t - \hat{S}_t - \hat{C}_t$ (or $y_t/(\hat{S}_t\hat{C}_t)$)
  - the classical decomposition and STL of Chapter 0 are simple versions of this
- For forecasting with SARIMA we use the **unadjusted** series: adjustment filters change the autocorrelations and are revised when new data arrive

X-11 and X-13ARIMA-SEATS

- **X-11** (US Census Bureau, 1960s; Ladiray and Quenneville (2001)): iterated moving averages
  - a centred 2×12 moving average for the trend, then 3×3 and 3×5 moving averages of each month across years for the seasonal factors
- **X-12-ARIMA** (Findley et al. (1998)) and **X-13ARIMA-SEATS** add a **regARIMA** pre-adjustment:
  - regression with SARIMA errors (often the airline model) for outliers, working days, Easter
  - ARIMA forecasts extend the series, so the symmetric filters also work at the end: smaller revisions
  - then X-11 filters or SEATS (model-based decomposition of the ARIMA model)
- EU: Eurostat and the national institutes follow Eurostat (2024), with the JDemetra+ software (X-13 and TRAMO-SEATS)



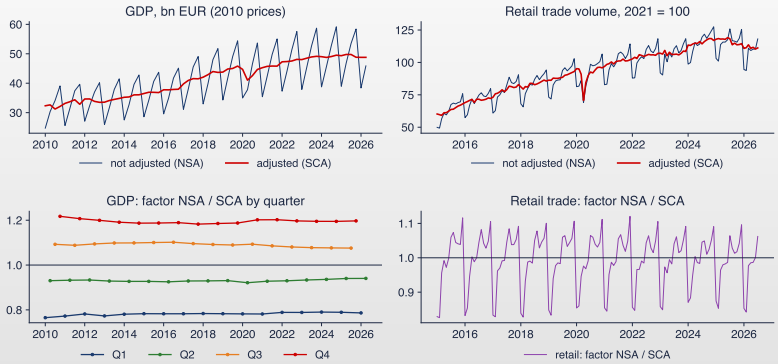
US Census Bureau, Suitland, Maryland: the home of X-11 and X-13

Photo: United States Census Bureau (2007); public domain; Wikimedia Commons

Eurostat series: NSA, CA, SA, SCA

- Each series comes in several versions (dimension `s_adj`):
  - **NSA**: not seasonally adjusted (the raw data)
  - **CA**: calendar adjusted only (working days, holidays)
  - **SA**: seasonally adjusted; **SCA**: seasonally and calendar adjusted
- Which one for which question
  - quarter-on-quarter or month-on-month growth: SCA
  - year-on-year growth: NSA or CA (the seasonal pattern cancels, the calendar does not)
  - SARIMA modelling and forecasting: NSA, with calendar regressors
- The annual totals of NSA and SCA are close (2024: 196.4 and 196.4 bn EUR): adjustment moves output between quarters, it does not create it

Unadjusted and adjusted series



- Eurostat NSA and SCA: GDP (top left) and retail trade (top right); bottom: the implied seasonal and calendar factors NSA/SCA

TSA_ch4_seasonal_data

Interpreting the adjusted series

- GDP factors 2015–2019: Q1 0.78, Q2 0.93, Q3 1.10, Q4 1.19: Q1 output is about a fifth below an average quarter
  - the factors move slowly over the years: the stochastic seasonality that $\hat{\Theta}$ measures
- Retail trade: December factor 1.10, January 0.84 (2016–2019): Christmas shopping, then the January low
- The retail factor also jumps from month to month because of working days and Easter: the calendar part of SCA (Section 7)
- The adjusted GDP shows the 2020 fall and the 2025–2026 stagnation that the raw series hides

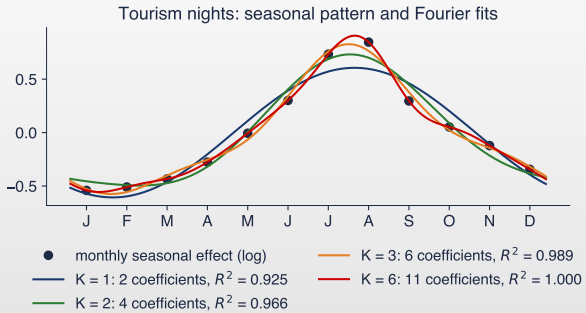
Recap: Seasonal adjustment

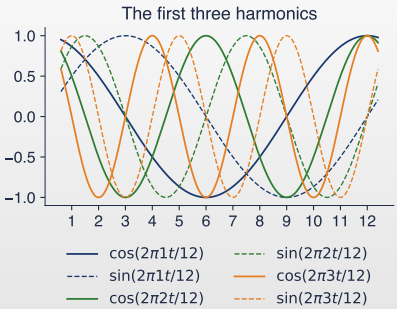
- Seasonal adjustment removes $\hat{S}_t$ and $\hat{C}_t$ to show the trend and the cycle
- X-13ARIMA-SEATS: regARIMA pre-adjustment (outliers, calendar, forecasts), then X-11 or SEATS; in the EU, JDemetra+
- Quarter-on-quarter growth: SCA; year-on-year growth: NSA or CA; SARIMA forecasting: NSA

Seasonal dummies and Fourier terms

- **Dummies**: $y_t = \beta_0 + \sum_{j=2}^{s} \gamma_j D_{jt} + u_t$: $s-1$ coefficients, any shape
  - β_0: the level of season 1; γ_j: the difference between season j and season 1
  - too many for $s = 52$ weeks, $s = 168$ hours or $s = 365$ days; impossible for $s = 365.25$
- **Fourier terms** for a period 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: the coefficients of the k-th harmonic, a wave with k cycles per period; K: the number of sine–cosine pairs
  - $2K$ coefficients; $K = m/2$ reproduces the dummies exactly (for even m); a small K gives a smooth pattern
  - m may be non-integer, and several periods can be combined: e.g. $K = 5$ for $m = 24$ and $K = 4$ for $m = 168$ give 18 regressors instead of $23 + 167$
- K is chosen by AICc or by cross-validation; the same terms are inside TBATS and Prophet (Section 9)

Fourier approximation of a seasonal pattern





- Monthly seasonal effect of ln(tourism nights), 2012–2019, after removing a 13-month centred mean; fits with $K = 1, 2, 3, 6$; for $K = 6$, $\sin(\pi t)$ is 0 at every month and is left out (11 coefficients)

TSA_ch4_calendar_fourier

Interpreting the Fourier fits

- $K = 1$ (two coefficients) already gives $R^2 = 0.925$: the summer peak is close to a single wave
  - $K = 2$: 0.966; $K = 3$: 0.989 with six coefficients instead of eleven
- $K = 5$ (10 coefficients): $R^2 = 0.999$; $K = 6$ adds $\cos(\pi t)$, 11 coefficients, and interpolates the 12 points like 11 dummies ($R^2 = 1.000$)
- Higher harmonics add sharper detail: the August peak needs $K \geq 3$

Calendar effects and Orthodox Easter

- **Working days**: a month with more Monday–Friday non-holiday days produces and sells more
  - known in advance: a regressor that can be used in forecasts
- **Moving holidays**: Orthodox Easter falls between 4 April and 8 May
  - 5 May 2024, 20 April 2025, 12 April 2026, 2 May 2027: shopping shifts between March, April and May
  - in Python: `dateutil.easter.easter(y, EASTER_ORTHODOX)`; Pentecost is 49 days later
- **Easter regressor** (X-13 `easter[w]`): E_t = share of the w days before Easter Sunday that fall in month t
  - $w = 10$: Easter on 20 April gives $E_{\textbf{April}} = 1$; on 5 May, $E_{\textbf{April}} = 0.6$ and $E_{\textbf{May}} = 0.4$



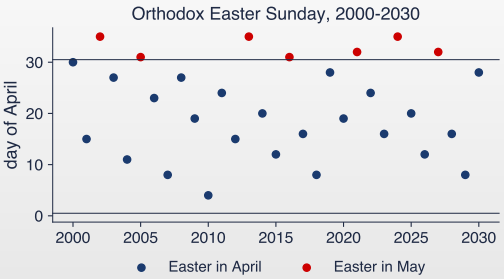
Painted Easter eggs from Romania

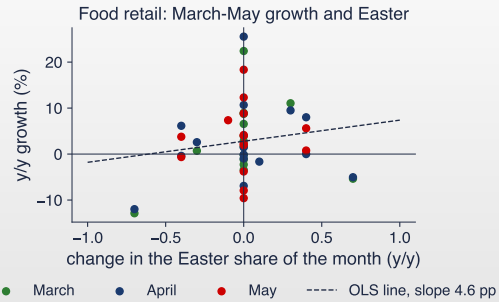
Photo: jackmac34 (Pixabay); CC0; Wikimedia Commons

Regression with SARIMA errors

- $100 \ln y_t = \beta_E E_t + \beta_W \mathrm{WD}_t + \text{outlier dummies} + u_t$, with $u_t \sim \mathrm{SARIMA}(0,1,1)(0,1,1)_{12}$
  - E_t: the Easter regressor; WD_t: the number of working days in month t; β_E, β_W: their effects, in %
  - the "regARIMA" model of X-13; estimated in one step by maximum likelihood (`SARIMAX` with exog)
  - ordinary least squares with SARIMA errors ignored would give wrong standard errors (Chapter 3, spurious regression)
- Data: Eurostat retail trade volume, Romania, not adjusted, January 2010–July 2026: food (G47_FOOD) and all retail (G47)
  - dummies for April and May 2020 (lockdown)

Easter and Romanian retail trade





- Left: Orthodox Easter Sunday as a day of April (above 30: May); right: year-on-year growth of food retail in March–May against the change in the Easter share (2020–2021 left out)

TSA_ch4_calendar_fourier

Interpreting the Easter effect

- Food retail: $\hat{\beta}_E = 5.9\%$ (se 2.3, $t = 2.6$): a month that holds the 10 days before Easter sells about 5.9% more food
  - one working day more: $+0.59\%$ (se 0.17); April 2020: -15%
- All retail: Easter 0.8% (se 1.9), not significant; working days $+1.02\%$ (se 0.15)
  - the Easter effect is in food, the working-day effect in everything
- The raw scatter is noisy (slope 4.6 percentage points): VAT changes and prices also move year-on-year growth; the regression separates the effects
- Easter fell in May in 2002, 2005, 2013, 2016, 2021, 2024: a forecaster who ignores this mistakes April and May

Recap: Fourier terms and calendar effects

- Fourier terms describe a seasonal pattern with $2K$ coefficients, for any period, even non-integer
- Working days and moving holidays are known in advance: model them as regressors
- Romanian food retail rises about 5.9% in the Easter month; total retail follows working days
- Regression with SARIMA errors: estimate regressors and dynamics together

Electricity load in Romania

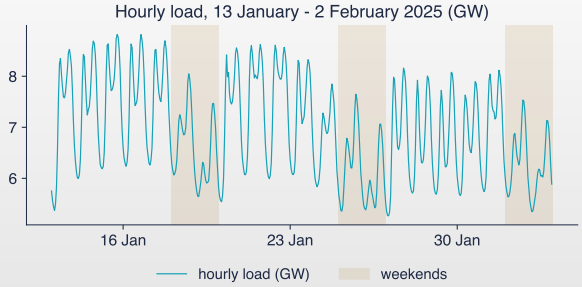
- **Load**: the electricity consumed in the system, in MW (hourly averages); it must be matched by generation at every moment
  - forecasts are needed for every hour of the next day: day-ahead markets, reserves, grid planning
- Data: ENTSO-E (the European network of transmission system operators), "monthly hourly load values", Romania
  - 39 406 hours, 2022–2026, Romanian winter time (UTC+2) to avoid the clock changes
- Load forecasting has its own competitions: GEFCom2014 (Hong et al. (2016))

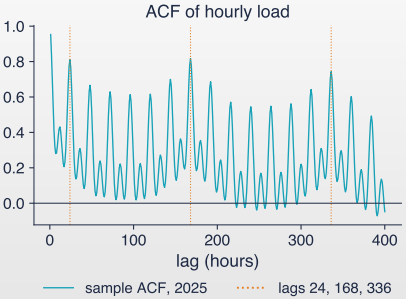


Iron Gates II hydropower plant on the Danube

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

Hourly load: the day and the week





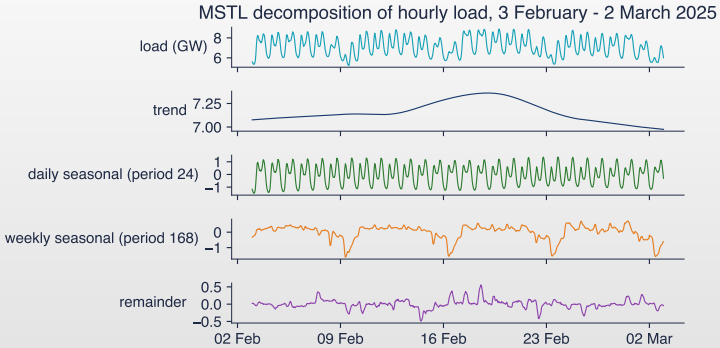
- Left: three weeks of January 2025; right: sample ACF of all hours of 2025 up to lag 400

TSA_ch4_multiple_seasonality

Interpreting the hourly load

- Daily cycle: lowest around 3:00 (5.3 GW on average in 2025), highest around 19:00 (7.3 GW)
  - ACF 0.81 at lag 24, 0.43 at lag 12
- Weekly cycle: weekends are lower; ACF 0.82 at lag 168 (one week), above the lag-24 value
  - the same hour one week ago is the best single predictor: Monday is like last Monday, not like Sunday
- And an annual cycle on top (heating and air conditioning): three periods, 24, 168 and about 8766 hours
- SARIMA is not enough: it has one integer period s, while here $s = 24$, 168 and 8766 act at once
  - a polynomial in L^{168} is slow and unstable, 365.25 is not an integer, and holidays break the weekly pattern
  - answers: MSTL, the dynamic harmonic regression (DHR), TBATS, Prophet

MSTL: one seasonal component per period



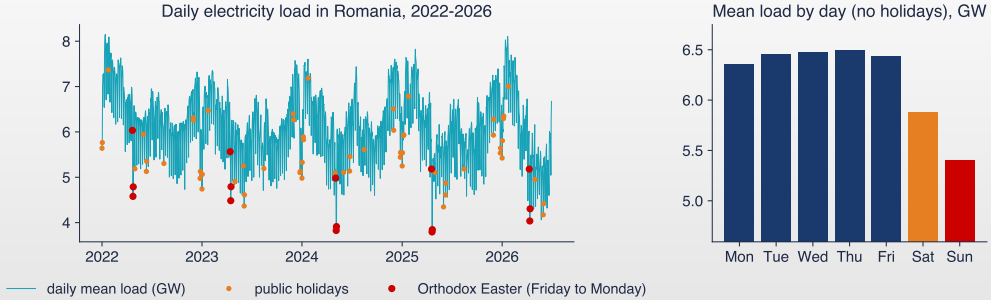
- MSTL (Bandara, Hyndman and Bergmeir (2025)) applies STL (Chapter 0) repeatedly, once for each period (24 and 168 hours), until the components settle

TSA_ch4_multiple_seasonality

Interpreting the MSTL decomposition

- Of the variance around the trend: 70% daily component, 29% weekly component, 2% remainder
  - range of the daily component in the first week: 2.8 GW; of the weekly one: 2.1 GW
- The weekly component is flat Monday to Friday and drops on Saturday and more on Sunday
- The remainder is small but not white noise: cold spells and holidays; a forecasting model adds dynamics and regressors

Daily load and public holidays



- Daily mean load, 2022–2026, 1 642 days; Romanian public holidays marked; right: mean by day of the week, holidays excluded

TSA_ch4_multiple_seasonality

Interpreting the daily load

- Week: Saturday 9% and Sunday 17% below Wednesday
  - Year: January 25% above May; a smaller summer peak (air conditioning)
- Holidays: the lowest day of the sample is Easter Sunday, 20 April 2025 (3.79 GW, mean 6.18 GW)
  - the four Easter days (Friday to Monday) average 4.62 GW; Christmas to New Year and Pentecost are also low (Pentecost Sunday was the lowest day of 2023)
- The level falls over 2022–2026; possible causes: the price shock of 2022 and rooftop solar panels that cover part of the demand behind the meter

A dynamic harmonic regression for daily load

- $y_t = \sum_{k=1}^{4} \left[\alpha_k \sin \frac{2\pi kt}{365.25} + \beta_k \cos \frac{2\pi kt}{365.25}\right] + \sum_j \delta_j D_{jt}^{\text{day}} + \lambda_1 H_t + \lambda_2 E_t + \lambda_3 X_t + u_t$
  - α_k, β_k: annual Fourier coefficients; δ_j: effect of weekday j against Sunday; $\lambda_1, \lambda_2, \lambda_3$: effects of the special days, in GW
  - annual Fourier terms; weekday dummies (Sunday is the base); H public holiday, E Easter (Friday to Monday), X 24 December – 2 January
  - $u_t \sim$ ARIMA(1,1,1), chosen by AIC: **dynamic harmonic regression** (DHR, FPP3, Section 10.5)
- Estimates, data to the first forecast origin (30 June 2025), in GW:
  - Wednesday +1.08, Monday +0.88, Saturday +0.50 relative to Sunday
  - public holiday −0.38 (se 0.02), Easter −0.58 more, Christmas–New Year −0.61
  - errors: AR 0.80, MA −0.98, $\hat{\sigma} = 186$ MW; Ljung–Box(14) p < 0.001: some dynamics left

Recap: Multiple seasonality

- Hourly load has daily, weekly and annual cycles; daily load has weekly and annual cycles
- SARIMA handles one integer period; MSTL decomposes several
- Holidays, Easter most of all, are the largest deviations: they need regressors
- DHR = Fourier terms + calendar dummies + ARIMA errors: flexible and transparent

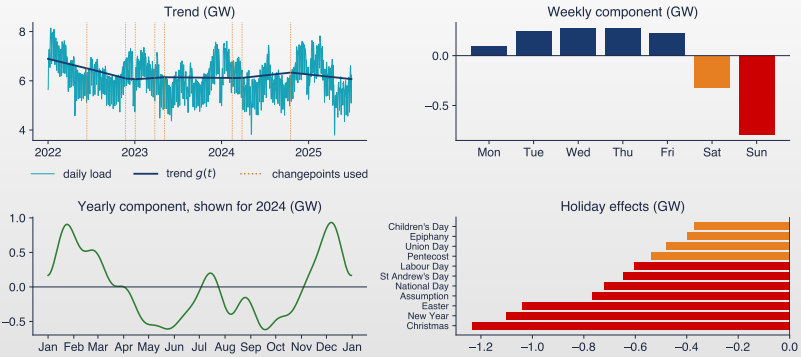
TBATS

- De Livera, Hyndman and Snyder (2011): exponential smoothing (ETS, Chapter 0) for complex seasonality; the name lists its parts
  - trigonometric seasonality (T), Box–Cox transformation (B), ARMA errors (A), trend (T), seasonal components (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); level and trend updated as in Holt's method
  - $y_t^{(\omega)}$: the series after a Box–Cox transformation with parameter ω; ℓ_t: level; b_t: slope; ϕ: damping factor; $s_t^{(i)}$: seasonal component i
  - each period m_i has K_i harmonics $s_{j,t}^{(i)}$, each a rotating pair (cosine, sine) updated by the errors: a Fourier pattern that can evolve
- Automatic choices by AIC: Box–Cox or not, trend or damped trend or none, the K_i, the ARMA orders
  - daily load to June 2025, periods 7 and 365.25 (no Box–Cox, no trend): $K = 3$ weekly and $K = 9$ annual harmonics, ARMA(1,2), 28 states
- Limits: no regressors, hence no holidays or Easter; slow on long series; Python package `tbats`

Prophet

- Taylor and Letham (2018) (Meta): a regression in time, $y(t) = g(t) + s(t) + h(t) + \varepsilon_t$
  - $g(t)$: piecewise-linear (or logistic) trend whose slope may change at **changepoints**; a Laplace prior shrinks most changes to zero
  - $s(t)$: Fourier terms, by default $K = 10$ for the year and $K = 3$ for the week
  - $h(t)$: one effect per holiday (with optional windows around it), from a list of dates
- Estimated by Stan (maximum a posteriori); intervals simulate future changepoints and the noise
  - no autoregressive part: today's surprise does not move tomorrow's forecast
- Built for many business series with strong calendar effects and missing data; Python package `prophet`

Prophet components for daily load



- Prophet with Romanian public holidays, daily load 2022 – June 2025

TSA_ch4_tbats_prophet

Interpreting the Prophet components

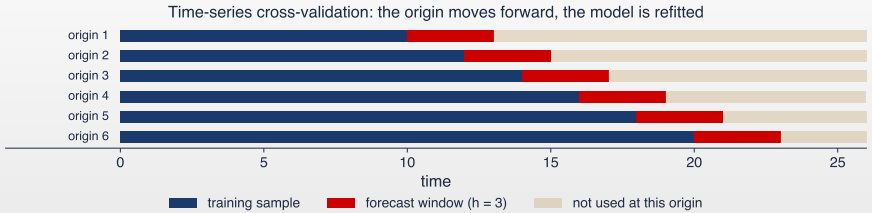
- Trend from 6.89 to 6.07 GW, flatter after 2023; 25 candidate changepoints, most shrunk to zero
  - weekly: Wednesday $+0.27$, Saturday -0.32, Sunday -0.80 GW
- Yearly: highest in early December ($+0.93$ GW, 7 December 2024), lowest in mid-September (-0.62 GW, 18 September 2024), shown for 2024
- Holidays: Christmas -1.24, New Year -1.10, Easter -1.04 GW; Children's Day only -0.37 GW
- Each component is readable and can be checked against what we know about the economy

Recap: DHR, TBATS and Prophet side by side

	DHR	TBATS	Prophet
Seasonality	Fourier, fixed	Fourier, evolving	Fourier, fixed
Several periods	yes	yes	yes
Holidays, regressors	yes	no	yes
Short-run dynamics	ARMA errors	ARMA errors	none
Trend	through the errors (ARIMA)	local (Holt)	piecewise linear
Speed	fast	slow	fast

- All three are regressions on sines and cosines at heart; they differ in what else they model
- Which forecasts best is an empirical question: the next section answers it out of sample

Time-series cross-validation



- Rolling forecast origin (Tashman (2000); FPP3, Section 5.10): fit on the data up to the origin, forecast h steps, move the origin, refit
- Expanding window (shown) or sliding window of fixed length; never use future data to fit

TSA_ch4_evaluation

Interpreting the design of our experiment

- Daily load: 26 origins every 14 days, from 30 June 2025 to 15 June 2026; horizon $h = 14$ days
  - one year of out-of-sample forecasts, every season and every holiday, Easter 2026 included
- Seven methods, all refitted at each origin
  - seasonal naive (period 7); ETS (Holt–Winters, damped trend, period 7); SARIMA(2,0,1)$(0,1,1)_7$; DHR; TBATS; Prophet
  - Combination: the simple average of the five model forecasts (all except the seasonal naive)
- The orders of SARIMA and of the DHR errors are chosen once, by AIC, on the data before the first origin: no peeking at the test period
- Cross-validation that ignores time order (random folds) is valid only for pure autoregressions with white-noise errors (Bergmeir, Hyndman and Koo (2018))

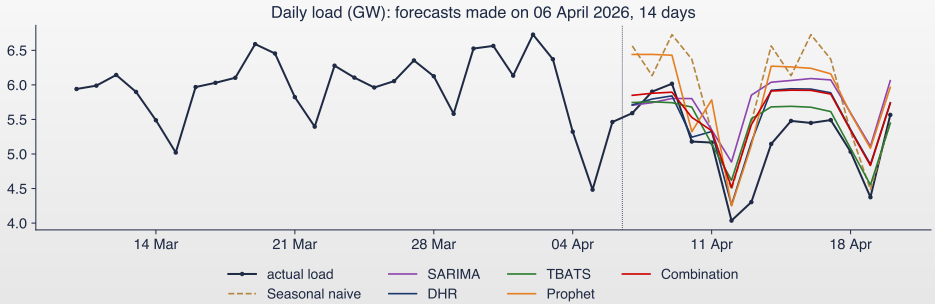
The Diebold–Mariano test

- Diebold and Mariano (1995): is the difference in accuracy between two forecasts larger than chance?
  - e_{1t}, e_{2t}: the errors of the two forecasts; L: the loss function
  - loss differential $d_t = L(e_{1t}) - L(e_{2t})$, $L(e) = e^2$ or $|e|$; H_0: $E[d_t] = 0$
- $\mathrm{DM} = \dfrac{\bar{d}}{\sqrt{\hat{V}/n}}$, $\hat{V} = \hat{\gamma}_d(0) + 2\sum_{k=1}^{h-1} \hat{\gamma}_d(k)$
  - $\bar{d}$: the mean of the n differentials; $\hat{\gamma}_d(k)$: their autocovariance at lag k; a negative DM favours forecast 1
  - h-step errors overlap, so d_t is autocorrelated up to lag $h-1$
  - asymptotically $N(0,1)$; small samples: $\mathrm{HLN} = \sqrt{\frac{n+1-2h+h(h-1)/n}{n}}$ DM, compared with $t(n-1)$ (Harvey, Leybourne and Newbold (1997))
- Our design: one value per origin, d_j = average loss difference over the 14-day window; the windows do not overlap, so $h = 1$ in the formula
  - the test compares forecasts, not models: re-estimation noise is part of what is tested

Worked example: a Diebold–Mariano test

- Five windows; window-average differences in squared error, method 1 minus method 2: $-0.30, 0.10, -0.50, -0.20, -0.40$
  - $\bar{d} = -0.26$: method 1 has the smaller losses in four windows out of five
- $\hat{\gamma}_d(0) = \frac{1}{5}\sum(d_j - \bar{d})^2 = 0.042$
  - $\mathrm{DM} = -0.26/\sqrt{0.042/5} = -2.82$; $\mathrm{HLN} = \sqrt{4/5} \cdot \mathrm{DM} = -2.53$
- $|-2.53| < t_{0.975}(4) = 2.78$ ($\mathrm{p} = 0.065$): equal accuracy is not rejected at 5%
  - the normal critical value 1.96 would have rejected: with few windows, the correction matters

Forecasts from one origin, across Easter 2026



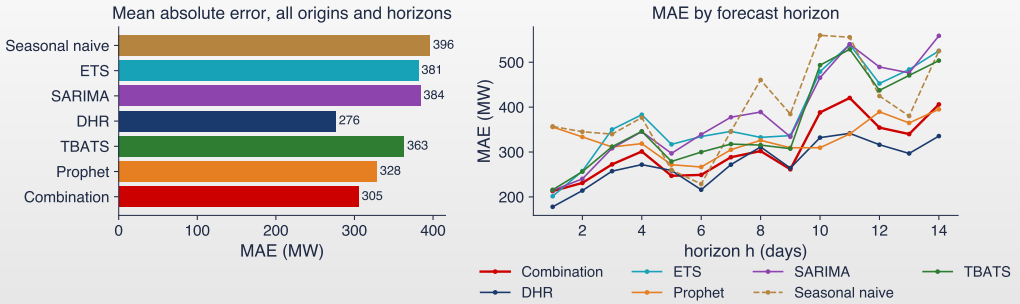
- Origin 6 April 2026, 14 days ahead; Orthodox Easter Sunday was 12 April 2026; ETS is not drawn (its MAE is on the next slide)

TSA_ch4_evaluation

Interpreting one forecast window

- MAE in this window (MW): ETS 299, TBATS 309, DHR 344, Combination 394, SARIMA 584, Prophet 616, seasonal naive 686
  - the seasonal naive method copies the previous normal week into Easter week
- Every model sees the Sunday dip; DHR and Prophet also lower Good Friday and Easter Monday, which the weekly models treat as working days
- After Easter the load stays below every forecast for the rest of the week; ETS and TBATS, whose level follows the latest days, lose least here
- One window is anecdote; the next slides average over all 26 windows

Cross-validation results for daily load



- 26 origins × 14 horizons for each method; left: overall MAE; right: MAE by horizon

TSA_ch4_evaluation

Interpreting the cross-validation results

Method	MAE	RMSE	MASE	DM against s. naive	DM against DHR
Seasonal naive	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)
Combination	305	380	0.99	-3.82 (<0.001)	1.89 (0.070)

- MAE and RMSE in MW; MASE (Hyndman and Koehler (2006)) = MAE / 309 MW, the in-sample MAE of the weekly seasonal naive method; DM: HLN statistic (p-value) on window-average squared errors, row method minus column method
- DHR wins; with the combination, it is the only method with MASE below 1; Prophet is the second single model: the two that know the holidays
- ETS, SARIMA and TBATS are not significantly better than the seasonal naive method

A second test: Romanian GDP

Method	MAE, $h = 1$	MAE, $h = 4$	without 2019–2020, $h = 1$	without 2019–2020, $h = 4$
SARIMA (airline)	1.69	2.70	1.32	1.89
ETS	1.76	2.84	1.41	2.44
Seasonal naive	3.81	3.74	3.80	3.49
Combination (SARIMA, ETS)	1.70	2.65	1.34	2.02

- $100 \ln Y$, errors in %; 51 quarterly origins, 2012Q4–2025Q2, expanding window; 43 origins without 2019–2020
- DM (HLN), SARIMA against ETS: $h = 1$: -1.46 ($p = 0.150$); $h = 4$: 0.50 ($p = 0.620$)
- SARIMA against the seasonal naive method: $h = 1$: -4.00 ($p < 0.001$); $h = 4$: -0.54 ($p = 0.591$)
- One year ahead, the pandemic errors dominate and no difference is significant; without them, SARIMA has the smallest MAE at $h = 4$

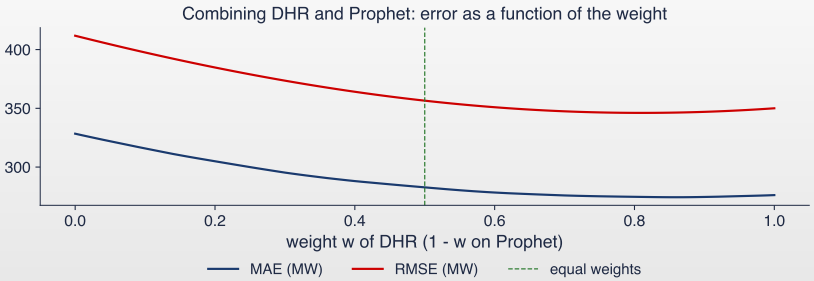
Recap: Forecast evaluation

- Evaluate out of sample with rolling origins; choose model orders before the test period
- MASE compares with the seasonal naive method and across series
- The DM test (with the HLN correction) says whether a gain is more than chance
- Daily load: the models that know the calendar (DHR, Prophet) win; for GDP, SARIMA and ETS are close

Why combining forecasts works

- Bates and Granger (1969): a weighted average of two unbiased forecasts, $f_c = wf_1 + (1-w)f_2$
  - w: the weight of forecast 1; σ_1, σ_2: the standard deviations of the two errors; ρ: their correlation
  - $\text{MSE}(w) = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho\sigma_1\sigma_2$
  - best weight $w^* = \dfrac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$; the gain is large when ρ is small
- Example: $\sigma_1 = 1$, $\sigma_2 = 1.2$, $\rho = 0.75$: $\text{MSE}(0.5) = 1.060$; $w^* = 0.84$, $\text{MSE}(w^*) = 0.984$
  - highly correlated errors: the equal-weight average is worse than the better forecast ($\sigma_1^2 = 1$), and even the best weight gains little
- The **forecast combination puzzle** (Stock and Watson (2004); Smith and Wallis (2009)): estimated weights rarely beat equal weights
  - the weights must be estimated, and their sampling error eats the theoretical gain

Combining DHR and Prophet



- Cross-validation errors of the daily load; w is the weight of DHR

TSA_ch4_evaluation

Interpreting the combination

- MAE: Prophet alone (w = 0) 328 MW, DHR alone (w = 1) 276 MW, equal weights 283 MW
  - best in hindsight: $w = 0.86$, MAE 274 MW, a small gain chosen after seeing the test data
- The errors of the two methods are strongly correlated ($\rho = 0.75$): both miss the same cold spells
- The five-model average (MAE 305) is second only to DHR and better than four of its five members: cheap insurance when the winner is not known in advance
- In the M4 competition (Makridakis, Spiliotis and Assimakopoulos (2020)), most of the best methods were combinations; review: Wang et al. (2023)

Possible contribution of AI

- **Code**: a first draft of a rolling-origin loop that refits SARIMA, DHR, TBATS and Prophet and stores the errors by horizon
- **Calendars**: lists of holidays and moving feasts for other countries, to be checked against official sources
- **Explanation**: a second derivation of the ACF of the airline model, or of the eventual forecast function
- Example 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.`

Checks you must run

- The Easter date: Orthodox, not Western (`EASTER_ORTHODOX`); in 2024 the two differ by five weeks
- No look-ahead: features, model orders and holiday lists must be available at each origin
- The sign conventions of θ and Θ in the software, and the degrees of freedom of Ljung–Box
- Adjusted or unadjusted data: SARIMA on NSA; growth rates on the right version
- That the DM test uses losses from the same windows for both methods, and the HLN correction for small n
- Every cited reference: it must exist; check the DOI

Key takeaways

- Seasonality repeats with a known period; it can be deterministic (dummies, Fourier) or stochastic (Δ_s)
- SARIMA multiplies regular and seasonal polynomials; the airline model fits many series with two parameters
- Official data: SCA for short-run growth, NSA with calendar regressors for SARIMA forecasting
- Calendar effects (working days, Orthodox Easter, holidays) are known in advance and often matter more than the model class
- Multiple seasonality: MSTL, DHR, TBATS, Prophet; on Romanian daily load, DHR wins out of sample
- Judge forecasts by rolling origins, MASE and the DM test; combine them when the winner is uncertain

Key formulas

Quantity	Formula
Seasonal difference	$\Delta_s y_t = (1-L^s)y_t, \quad \Delta\Delta_s = 1 - L - L^s + L^{s+1}$
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	$\bar{d}/\sqrt{\hat{V}/n}, \quad d_t = L(e_{1t}) - L(e_{2t})$
Combination	$w^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}$

Self-assessment

- **Question**: how many parameters does SARIMA$(1,1,1)(0,1,1)_{\mathbf{12}}$ have, and at which lags does ε appear?
  - **Answer**: $\phi, \theta, \Theta, \sigma^{\mathbf{2}}$, four; ε at lags 0, 1, 12, 13 (the last with coefficient $\theta\Theta$)
- **Question**: unadjusted GDP falls by 40% from Q4 to Q1. Is this a recession?
  - **Answer**: no; it is the seasonal pattern; look at the SCA series or at year-on-year growth
- **Question**: a DM test with 26 windows gives HLN $= -1.93$. Is the first forecast significantly better at 5%?
  - **Answer**: no: $|-1.93| < t_{\mathbf{0.975}}(25) = 2.06$; it is better on average, but not significantly
- **Project idea**: forecast Romanian hourly electricity load with a DHR (daily, weekly and annual Fourier terms, holidays, temperature) against TBATS, Prophet and a combination, by rolling origins and DM tests
  - Next: Chapter 5, conditional volatility: ARCH and GARCH models

References (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

References (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

References (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