



Energy Markets Quantitative Analysis

Lecture 2: Forecasting Electricity Prices



Prof. dr. Daniel Traian PELE

Bucharest University of Economic Studies, Department of Statistics and Econometrics
Romanian Academy, Institute for Economic Forecasting
IDA Institute Digital Assets ▪ AI4EFin ▪ MSCA Digital Finance

Learning objectives

- Set up a day-ahead forecast that uses only what is known at 12:00 Central European Time (CET) on the day before delivery.
- Measure accuracy with the mean absolute error (MAE), the root mean squared error (RMSE) and the relative MAE (rMAE), and test differences with the Diebold–Mariano (DM) test.
- Build expert autoregressive models with exogenous variables (ARX), shrink them with the least absolute shrinkage and selection operator (LASSO) and obtain the LASSO-estimated autoregression (LEAR).
- Combine forecasts, preview machine learning (ML) and turn point forecasts into quantiles with quantile regression averaging (QRA).
- Test cointegration of Romania with Hungary, and of power with gas and hydro, as announced in Lecture 1.

Abbreviations (1/2)

ADF	augmented Dickey–Fuller test
AI	artificial intelligence
AIC	Akaike information criterion
ARX	autoregressive model with exogenous variables
AW	averaging across calibration windows
AW-ARX	ARX averaged across calibration windows
AW-LEAR	LEAR averaged across calibration windows
CET	Central European Time
CP	conformal prediction
CRPS	continuous ranked probability score
CVaR	conditional value at risk
DE-LU	Germany–Luxembourg bidding zone
DM	Diebold–Mariano test
DNN	deep neural network
DST	daylight saving time
ECM	error-correction model
ENTSO-E	European Network of Transmission System Operators for Electricity
EPEX	EPEX SPOT power exchange
ETC	exchange-traded commodity
EU	European Union
EUA	EU emission allowance (1 t CO_2)
EUPHEMIA	Pan-European Hybrid Electricity Market Integration Algorithm
EUR	euro
GBM	gradient boosting machine (boosted trees)
HS	historical simulation
HU	Hungary (bidding zone)
HUPX	Hungarian Power Exchange
ISE	Fraunhofer Institute for Solar Energy Systems
LARS	least angle regression

Abbreviations (2/2)

LASSO	least absolute shrinkage and selection operator
LEAR	LASSO-estimated autoregression
LR	likelihood ratio
MAD	median absolute deviation
MAE	mean absolute error
MAPE	mean absolute percentage error
ML	machine learning
MLP	multilayer perceptron (feed-forward neural network)
MTU	market time unit
MW	megawatt
MWh	megawatt-hour
OLS	ordinary least squares
OPCOM	Romanian electricity and gas market operator
PS	pinball score
PZU	Piața pentru Ziua Următoare (day-ahead market)
QRA	quantile regression averaging
ReLU	rectified linear unit
rMAE	relative MAE (model MAE / naive MAE)
RMSE	root mean squared error
RO	Romania (bidding zone)
RSS	residual sum of squares
SDAC	Single Day-Ahead Coupling
sMAPE	symmetric MAPE
TSO	transmission system operator
TTF	Title Transfer Facility (Dutch gas hub)
TWh	terawatt-hour
UC	unconditional coverage
VST	variance-stabilising transformation

Forecast target and data sources

Series	Frequency	Source
Target: day-ahead price of Romania (RO), Piața pentru Ziua Următoare (PZU)	hourly	the Romanian electricity and gas market operator (OPCOM), via energy-charts.info (Fraunhofer ISE) and the European Network of Transmission System Operators for Electricity (ENTSO-E)
Load, wind + solar, hydro, nuclear, cross-border flows	hourly, daily	ENTSO-E (Transelectrica data for Romania), via energy-charts.info
Carbon: European Union (EU) emission allowance (EUA), proxy	daily	Yahoo Finance, SparkChange physical EUA exchange-traded commodity (ETC)
Neighbours: Hungary (HU), Germany–Luxembourg (DE-LU); gas: Title Transfer Facility (TTF)	daily, monthly	Hungarian Power Exchange (HUPX), European Power Exchange (EPEX SPOT), via energy-charts.info; World Bank Pink Sheet (TTF)

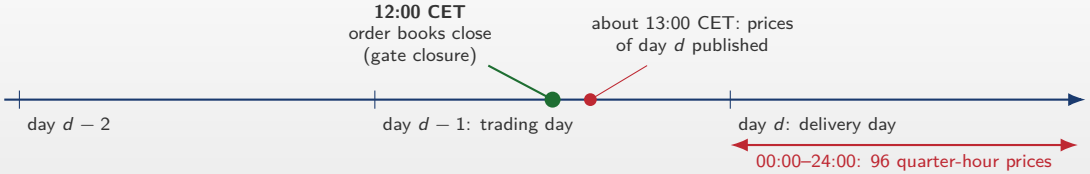
- **Cut-off 31 Oct 2026**, downloaded 4 Nov 2026
  - Panel 1 Jan 2022 – 31 Oct 2026, 1 765 complete days.
- **Test** = the last 365 days (1 Nov 2025 – 31 Oct 2026)
  - **burn-in** = the 182 days before, used only to calibrate the intervals.
- Daylight saving time (DST): 24 columns per day as in Assignment 1
  - The missing spring hour is the mean of its neighbours, the doubled autumn hour is averaged
  - 2 DST days in the test year, kept.

Experiment yourself: notebook §1

Outline

- The forecasting problem and its benchmarks
  - The day-ahead market
  - Information set and horizons
  - Naive benchmarks and error measures
  - Testing forecast accuracy
- Expert models: ARX and LEAR
- Windows, combinations and machine learning
- Probabilistic forecasts and cointegration
- Case studies on recent Romanian data

Day-ahead market: the auction on $d-1$ sets the 96 prices of day d



- **What it is**
  - A single daily auction of the Single Day-Ahead Coupling (SDAC), run in Romania by OPCOM
  - Buyers and sellers submit orders (quantity, price) for each period of the next day
- **How prices are set**
  - At 12:00 CET on day $d-1$ the order books close (gate closure)
  - A common algorithm, the Pan-European Hybrid Electricity Market Integration Algorithm (EUPHEMIA), matches supply and demand; results normally around 13:00 CET
- **Delivery periods**
  - Since 1 October 2025: 96 quarter-hour prices per day (before: 24 hourly)
  - Our data: hourly price = mean of the four quarter-hours, used for the whole test year

(NEMO Committee, 2025, 2026; OPCOM)

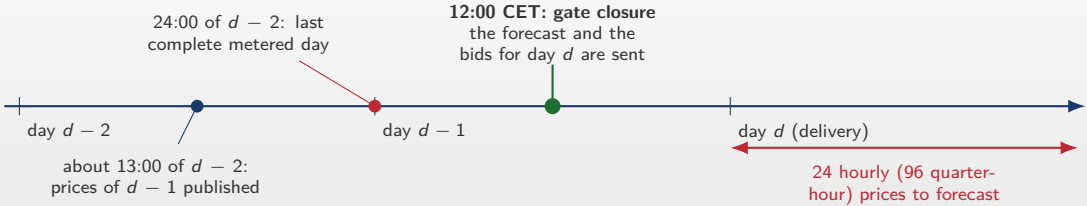
Notation and the information rule for all forecasts

- **Notation**
  - d = delivery day; $d-1$ = the day before, when the auction runs; $d-2$ = two days before
  - $h = 0, \ldots, 23$ = hour of delivery (CET); $P_{d,h}$ = day-ahead price of hour h on day d, in EUR/MWh; $\hat{P}_{d,h}$ = its forecast
- **The rule**
  - The forecast for day d is made before 12:00 CET on day $d-1$, so it may use only what is known at that moment
  - Allowed: all 24 prices of day $d-1$ (fixed by the auction on day $d-2$); load, wind and solar forecasts for day d published by the transmission system operator (TSO) before 12:00
  - Not allowed: any realised value of day d, or values of day $d-1$ that are not yet metered
- **What we forecast and how we score it**
  - The 24 hourly prices of the Romanian PZU, for each day of the test year
  - The error $P_{d,h} - \hat{P}_{d,h}$ is known only after day d is delivered

- **In practice:** the same rule binds a real trading desk
  - A model that breaks it cannot be run at 12:00.

(NEMO Committee, 2025; Lago et al., 2021, APEN)

Information set at 12:00 CET on day $d-1$



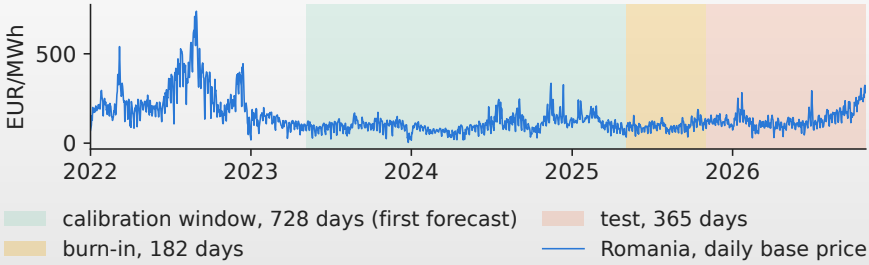
- **Allowed** for the forecast of day d
  - Prices up to $d-1$; realised load, wind, solar, hydro, nuclear and flows up to $d-2$; carbon at the $d-2$ close
- **Not in our data**
  - The day-ahead load and renewable forecasts of the transmission system operator (TSO), which a real desk would use

In practice: a backtest that uses the realised load of $d-1$ looks excellent and cannot be run at 12:00.

(Assignment 1; OPCOM; Lago et al., 2021, APEN)

Experiment yourself: notebook §1

Chronological split into calibration, burn-in and test periods



Source: energy-charts.info (Fraunhofer ISE); OPCOM day-ahead prices for Romania, via ENTSO-E. Data: daily base price, 2022-01-01 to 2026-10-31, CET delivery days.

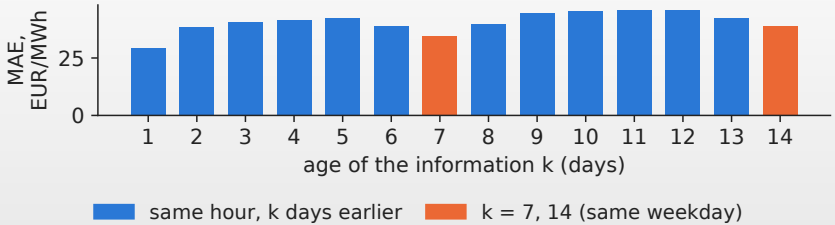
- Models re-estimated every day on the window ending at $d-1$
  - No random K-fold, no random validation split.

- **In practice:** ask any vendor for the rolling backtest, not an in-sample fit.

(Weron, 2014, IJF; Lago, Marcjasz, De Schutter & Weron, 2021, APEN)

EMQA_L2_split_timeline

Forecast error grows with the age of the information, except at the weekly lag



Source: energy-charts.info (Fraunhofer ISE); OPCOM day-ahead prices for Romania, via ENTSO-E. Data: test year 2025-11-01 to 2026-10-31, hourly.

- Forecast $\hat{P}_{d,h} = P_{d-k,h}$
  - MAE 29.37 EUR/MWh for $k = 1$, 38.52 for $k = 2$, back to 34.41 for $k = 7$ (same weekday).

- **In practice:** the day-ahead horizon is 12 to 36 hours
  - The freshest price you hold at 12:00 is that of $d - 1$.

(own computation, test year 1 Nov 2025 – 31 Oct 2026) EMQA_L2_horizon_naive

Point, probabilistic and path forecasts answer different questions

- **In one sentence:** A point forecast gives one number per hour, a probabilistic one its distribution, a path forecast joint scenarios for the 24 hours.

$$\underbrace{\hat{P}_{d,h}}_{\text{point}} \quad \underbrace{\hat{q}^{\tau}_{d,h}:\ \Pr(P_{d,h} \le \hat{q}^{\tau}_{d,h}) = \tau}_{\text{quantile}} \quad \underbrace{\{\hat{P}^{(s)}_{d,0}, \ldots, \hat{P}^{(s)}_{d,23}\}_{s=1}^{S}}_{\text{paths}}$$

- **where**
  - d = delivery day
  - h = hour (CET)
  - $\tau \in (0, 1)$ = quantile level
  - s = scenario
  - S = number of scenarios.

- **On our data**
  - Errors at 12 h and 13 h correlate at 0.96
  - Errors at 3 h and 19 h correlate at only 0.13

- **Who needs what**
  - Supplier buying a block, the point forecast
  - Risk manager, quantiles (Lecture 4)
  - Battery, paths, since it buys in one hour and sells in another

- **In practice:** decide which of the three your decision needs before you buy a forecast.

(Nowotarski & Weron, 2018, RSER)

Experiment yourself: notebook §7

The naive benchmark: yesterday, or last week for Monday and the weekend

- **In one sentence:** Every model must beat the rule a trader would use without any model.

$$\hat{P}^{\text{naive}}_{d,h} = \begin{cases} P_{d-7,h}, & \text{Mon, Sat, Sun} \\ P_{d-1,h}, & \text{Tue–Fri} \end{cases}$$
$$\hat{P}^{(7)}_{d,h} = P_{d-7,h}$$

- **where**
  - $P_{d,h}$ = price of day d, hour h (EUR/MWh)
  - $\hat{P}^{(7)}$ = seasonal naive (same hour last week).

- **Why the switch**, MAE in EUR/MWh
  - Mon, Sat, Sun: $d-7$ 29.87 vs $d-1$ 33.48
  - Tue–Fri: $d-1$ 26.27 vs $d-7$ 37.83
- **Test year**
  - Naive 27.82, seasonal naive 34.41 EUR/MWh
  - Mean price 134.46 EUR/MWh, so the naive error is about one fifth of the price

- **In practice:** ask any forecast vendor for the rMAE against this naive rule on your hours.

(Nogales et al., 2002, IEEE TPWRS; Weron, 2014, IJF)

Experiment yourself: notebook §2

Four error measures and the relative MAE

- **In one sentence:** Absolute errors are in EUR/MWh, squared errors punish spikes, rMAE compares with naive.

$$\text{MAE} = \frac{1}{N}\sum |e_{d,h}|, \quad \text{RMSE} = \sqrt{\frac{1}{N}\sum e_{d,h}^2}$$

$$\text{sMAPE} = \frac{1}{N}\sum \frac{2|e_{d,h}|}{|P_{d,h}|+|\hat{P}_{d,h}|}, \quad \text{rMAE} = \frac{\text{MAE}_{\text{model}}}{\text{MAE}_{\text{naive}}}$$

- **where**
  - $e_{d,h} = P_{d,h} - \hat{P}_{d,h}$
  - N = number of (day, hour) pairs
  - Symmetric mean absolute percentage error (sMAPE), bounded by 200%
  - A value of rMAE below 1 = better than naive.

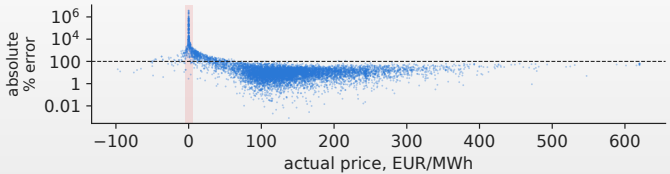
- **On our data**, test year 1 Nov 2025 – 31 Oct 2026
  - Naive: MAE 27.82, RMSE 44.54, sMAPE 34.5%
  - LEAR averaged across calibration windows (AW-LEAR): MAE 22.53, RMSE 36.35, rMAE 0.810
  - Best, ensemble of 4: rMAE 0.795, i.e. 20.5% less error than naive
- MAE rewards the median forecast, RMSE the mean.

- **In practice:** report MAE in EUR/MWh to the desk and rMAE to management, on the same hours.

(Hyndman & Koehler, 2006, IJF; Lago et al., 2021, APEN)

Experiment yourself: notebook §2

Limitations of percentage errors near zero and negative prices



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E load, wind and solar (Transelectrica data for Romania).
Data: test year 2025-11-01 to 2026-10-31, 8760 hours.

- Mean absolute percentage error (MAPE) $= \frac{1}{N}\sum |e_{d,h}|/|P_{d,h}|$
  - For AW-LEAR 5 777.7%, or 32.4% without the 387 hours with $|P| < 5$ EUR/MWh.
- These few hours produce 99.5% of the MAPE.

- **In practice:** never accept a MAPE for power prices
  - Ask for MAE and rMAE.

(Hyndman & Koehler, 2006, IJF; Weron, 2014, IJF)

EMQA_L2_mape_fail

The Diebold–Mariano test of equal predictive accuracy

- **In one sentence:** Compare the daily average error of two models; if the mean difference is large relative to its own noise, one model is really better.

$$\Delta_d = \underbrace{\tfrac{1}{24}\sum_h |e^A_{d,h}|}_{\text{model A}} - \underbrace{\tfrac{1}{24}\sum_h |e^B_{d,h}|}_{\text{model B}}, \qquad \mathrm{DM} = \frac{\bar{\Delta}}{\sqrt{\hat{\sigma}^2_\Delta / n}}$$

- **where**
  - n = test days
  - $\hat{\sigma}^2_\Delta$ = Newey–West variance with 7 lags
  - $p = 1 - \Phi(\mathrm{DM})$
  - Small p = B more accurate than A.

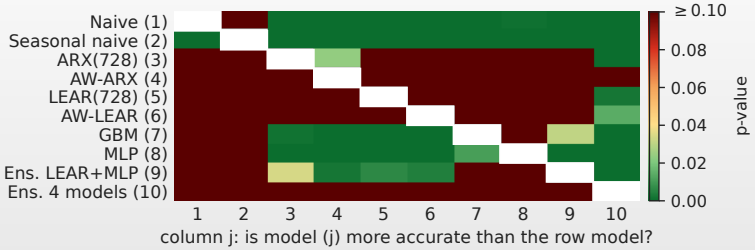
- **On our data**, 365 test days
  - Naive vs AW-LEAR: DM 7.32, $p < 0.001$
  - ARX averaged across windows (AW-ARX) vs AW-LEAR: DM -0.83, $p = 0.797$, no significant difference
  - AW-LEAR vs ensemble of 4: DM 2.21, $p = 0.014$
- Daily averaging removes the correlation between the 24 hours of a day (multivariate version).

- **In practice:** switch models only when the gain is significant on your hours.

(Diebold & Mariano, 1995, JBES; Ziel & Weron, 2018, ENEECO)

Experiment yourself: notebook §2

Pairwise DM tests on the Romanian test year



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E generation and flows (Transelectrica data for Romania); Yahoo Finance (SparkChange EUA ETC). Data: test year 2025-11-01 to 2026-10-31, 365 days; one-sided p-values.

- 8 of the 9 other models beat the naive benchmark at 5%
  - Green cells mark a significant win of the column.

- **In practice:** read a column top to bottom
  - Many green cells mean a model worth paying for.

(Ziel & Weron, 2018, ENEECO; Lago et al., 2021, APEN)

EMQA_L2_dm_matrix

Outline

- The forecasting problem and its benchmarks
- Expert models: ARX and LEAR
  - The expert ARX model
  - Variance-stabilising transformations
  - LASSO-estimated autoregression
- Windows, combinations and machine learning
- Probabilistic forecasts and cointegration
- Case studies on recent Romanian data

The expert ARX model: recent prices, calendar and fundamentals

- **In one sentence:** One linear regression per hour, with the inputs a trader looks at first.

$$Y_{d,h} = \beta_0 + \underbrace{\beta_1 Y_{d-1,h} + \beta_2 Y_{d-2,h} + \beta_3 Y_{d-7,h}}_{\text{autoregression}} + \underbrace{\beta_4 Y_{d-1,23} + \beta_5 Y^{\min}_{d-1} + \beta_6 Y^{\max}_{d-1}}_{\text{yesterday's shape}} + \underbrace{\beta_7 \tilde{L}_{d,h} + \beta_8 \tilde{R}_{d,h}}_{\text{fundamentals of } d-2} + \underbrace{\sum_j \delta_j D^j_d}_{\text{Mon, Sat, Sun}} + \varepsilon_{d,h}$$

- **where**
  - Y = asinh-transformed price
  - $\tilde{L}_{d,h} = L_{d-2,h}$ load, $\tilde{R}_{d,h}$ wind + solar of $d-2$, z-scored.

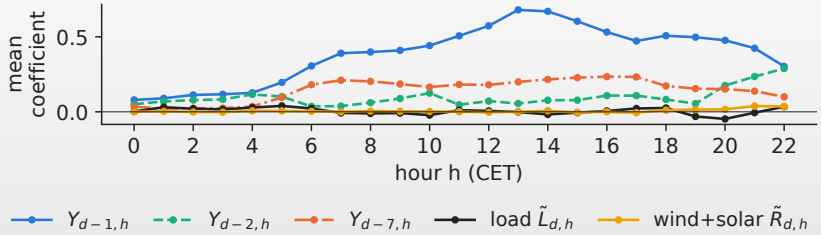
- **Estimation**
  - Ordinary least squares (OLS), separately for each h
  - 12 regressors per hour, as in Assignment 1
- **On our data**, ARX(728)
  - 728-day window, re-estimated daily
  - MAE 22.72 EUR/MWh, rMAE 0.817

- **In practice:** this is the model to beat before anyone sells you a neural network.

(Ziel & Weron, 2018, ENEECO; Weron, 2014, IJF)

Experiment yourself: notebook §3

The ARX coefficients change with the hour of the day



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E load, wind and solar (Transelectrica data for Romania).
Data: ARX(728), mean over the 365 daily fits of the test year.

- $Y_{d-1,h}$ weighs most at hour 13 (0.68) and least at hour 0 (0.08)
  - Night hours lean on the last hour of $d-1$ instead (up to 0.73).
- The load and wind + solar coefficients stay close to zero: at $d-2$ they add little to yesterday's price.

- **In practice:** one model per hour lets the evening peak learn from different drivers than the night.

(own estimation, ARX(728))

EMQA_L2_arx_coef

Small contribution of fundamentals lagged by two days

ARX(728), test year	MAE	RMSE	rMAE
prices and calendar only	22.63	36.49	0.813
+ load, wind + solar of $d-2$	22.72	36.62	0.817
+ hydro, nuclear, net imports, EUA of $d-2$	22.72	36.50	0.817

- Fundamentals missing on a day are carried forward from the past only (nuclear on 29 days).
- EUA = SparkChange ETC close of $d-2$, a proxy of the allowance price.

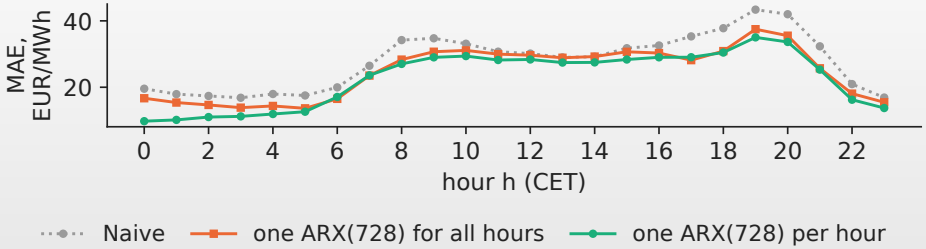
- **Why so little**
  - What happened on $d-2$ is already in the price of $d-1$
  - The outages of the Cernavodă nuclear power plant reach the model two days late
- **What would help**
  - Published day-ahead forecasts of load, wind and solar for day d (ENTSO-E, Transelectrica)
  - They are legal at 12:00 on $d-1$, but not in our free data

- **In practice:** buy the TSO's day-ahead load and renewable forecasts before buying a more complex model.

(own estimation; Lago et al., 2021, APEN)

Experiment yourself: notebook §3

Separate models for each hour are more accurate than a single model



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E load, wind and solar (Transelectrica data for Romania).
Data: test year 2025-11-01 to 2026-10-31.

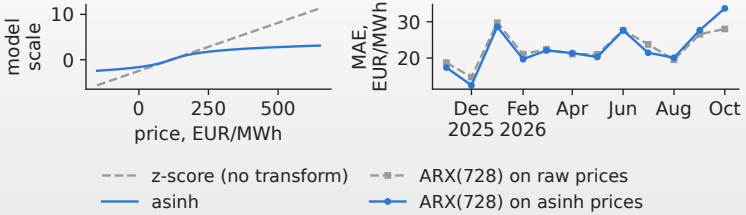
- MAE 22.72 per hour vs 24.54 EUR/MWh pooled (hour intercepts, common slopes)
  - Pooled worse in 21 of 24 hours.

- **In practice:** since 1 Oct 2025 the same choice arises for the 96 quarter-hours.

(Ziel & Weron, 2018, ENEECO)

EMQA_L2_per_hour_single

The asinh transformation for price spikes and negative prices



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E load, wind and solar (Transelectrica data for Romania).
Data: left, median and scale of the last window; right, test year 2025-11-01 to 2026-10-31.

- Variance-stabilising transformation (VST) $Y = \operatorname{asinh}((P - m)/s)$, m = median, $s = 1.4826 \times$ median absolute deviation (MAD), from the window only.
- ARX(728) MAE 22.86 raw vs 22.72 with asinh
  - RMSE goes the other way (35.85 vs 36.62).

- **In practice:** asinh works below zero, where the logarithm cannot be used.

(Uniejewski, Weron & Ziel, 2018, IEEE TPWRS)

EMQA_L2_vst

LEAR: LASSO selection among 247 candidate regressors

- **In one sentence:** A penalty on the size of the coefficients sets the useless ones exactly to zero, so the data choose the regressors.

$$\hat{\beta}_h = \arg\min_{\beta} \left\{ \underbrace{\sum_{d \in \text{cal}} (Y_{d,h} - \mathbf{x}_d^\top \beta)^2}_{\text{fit}} + \underbrace{\lambda_h \sum_j |\beta_j|}_{\text{penalty}} \right\}$$

- **where**
  - $\mathbf{x}_d = 247$ regressors, all 24 prices of $d-1, d-2, d-3, d-7$ (96), load and wind + solar of $d-2, d-3, d-7$ (144), 7 weekday dummies
  - λ_h chosen every day by the Akaike information criterion (AIC) inside the window.

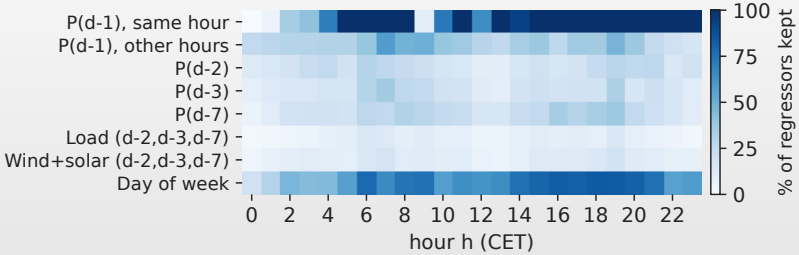
- **On our data**, LEAR(728)
  - On average 42 of 247 regressors kept (17%), from 8 to 121
  - MAE 22.62; averaged over 364, 546, 728 days (AW-LEAR) 22.53 EUR/MWh
- **Choosing λ honestly**
  - By an information criterion, or on a validation block that ends before the forecast day
  - Never on the test year

- **In practice:** a LASSO model explains itself, since the kept regressors show what drives each hour.

(Tibshirani, 1996; Uniejewski, Nowotarski & Weron, 2016; Lago et al., 2021)

Experiment yourself: notebook §4

Regressors selected by LEAR: yesterday's same hour and the weekday, rarely the lagged fundamentals



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E load, wind and solar (Transelectrica data for Romania).
Data: LEAR(728), 365 daily fits of the test year.

- Kept on average, $P_{d-1,h}$ 79%, weekday 64%, load 9%, wind + solar 11% of the regressors.

- **In practice:** real gains need forecasts of the fundamentals for day d.

(own estimation, LEAR(728), test year 1 Nov 2025 – 31 Oct 2026)

EMQA_L2_lear_sel

Outline

- The forecasting problem and its benchmarks
- Expert models: ARX and LEAR
- Windows, combinations and machine learning
  - Calibration windows
  - Combining forecasts
  - Machine learning: a preview of Lecture 7
- Probabilistic forecasts and cointegration
- Case studies on recent Romanian data

Averaging forecasts across calibration windows

- **In one sentence:** Instead of guessing the best window length, estimate the model on several windows and average the forecasts.

$$\hat{P}^{\text{AW}}_{d,h} = \frac{1}{|\mathcal{T}|} \sum_{T \in \mathcal{T}} \underbrace{\hat{P}^{(T)}_{d,h}}_{\text{window of } T \text{ days}}$$

- **where**
  - Rolling window = last T days, re-estimated daily
  - ARX: $\mathcal{T} = \{56, 84, 112, 714, 721, 728\}$
  - LEAR: $\mathcal{T} = \{364, 546, 728\}$
  - Averaging across calibration windows (AW).

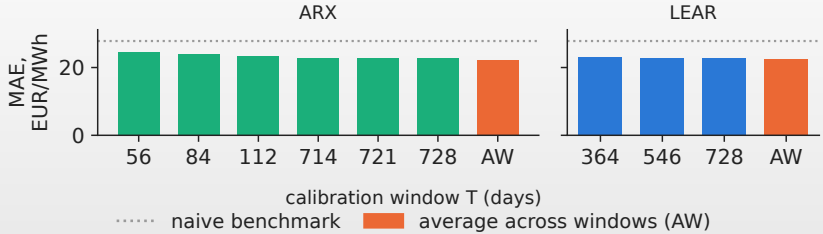
- **On our data**, MAE in EUR/MWh
  - ARX(56) 24.44, ARX(112) 23.41, ARX(728) 22.72
  - AW-ARX 22.30, better than every single window
  - LEAR(364) 22.96, LEAR(728) 22.62, AW-LEAR 22.53
- Mixing short (56–112) and long (714–728) windows gives both adaptation and stability.

- **In practice:** when the market changes regime (2022, 2024), short windows react first
  - Keep them in the average.

(Hubicka, Marcjasz & Weron, 2019; Marcjasz, Serafin & Weron, 2018)

Experiment yourself: notebook §5

Averaging across windows is more accurate than the best single window



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E load, wind and solar (Transelectrica data for Romania).
Data: test year 2025-11-01 to 2026-10-31.

- AW-ARX 22.30 vs the best single window ARX(714) 22.68
  - Dotted line = naive 27.82 EUR/MWh.

- **In practice:** you do not need to know the right window in advance
  - Average a short and a long one.

(Hubicka, Marcjasz & Weron, 2019, IEEE TSTE)

EMQA_L2_windows

Combining different models: errors that do not coincide cancel out

- **In one sentence:** If two models are wrong on different days, their simple average is wrong less often than either.

$$\hat{P}^{\text{ens}}_{d,h} = \frac{1}{K}\sum_{k=1}^{K} \hat{P}^{(k)}_{d,h}, \quad \text{Var}(\bar{e}) = \frac{\bar{\sigma}^2}{K}\left(1 + (K-1)\bar{\rho}\right)$$

- **where**
  - K = number of models
  - $\bar{\sigma}^2$ = average error variance
  - $\bar{\rho}$ = average correlation of the errors
  - The lower $\bar{\rho}$, the larger the gain
  - Equal weights need no estimation.

- **On our data**, ensemble of 4
  - AW-ARX, AW-LEAR, a gradient boosting machine (GBM) and a multilayer perceptron (MLP), equal weights
  - MAE 22.13 EUR/MWh, rMAE 0.795, the best of all models
  - Better than AW-LEAR in 8 of 12 months and in 15 of 24 hours
  - DM test against AW-LEAR: $p = 0.014$
- The two weaker members (GBM, MLP) still help, because their errors differ.

- **In practice:** ask vendors for forecasts you can average with yours, not for a single best model.

(Bates & Granger, 1969; Nowotarski & Weron, 2015)

Experiment yourself: notebook §5

Accuracy of all models on the Romanian test year

Model	MAE	RMSE	sMAPE, %	rMAE
Naive	27.82	44.54	34.5	1.000
Seasonal naive	34.41	55.48	39.4	1.237
ARX(56)	24.44	40.09	30.0	0.879
ARX(728)	22.72	36.62	27.8	0.817
AW-ARX	22.30	36.34	27.6	0.802
ARX(728), one model	24.54	37.97	29.5	0.882
ARX(728), no VST	22.86	35.85	28.4	0.822
LEAR(364)	22.96	36.80	28.1	0.825
LEAR(728)	22.62	36.36	27.7	0.813
AW-LEAR	22.53	36.35	27.7	0.810
GBM	24.05	37.73	28.7	0.864
MLP	25.54	40.05	29.8	0.918
Ensemble LEAR+MLP	23.12	36.95	28.1	0.831
Ensemble of 4	22.13	35.76	27.4	0.795

- Test year 1 Nov 2025 – 31 Oct 2026, 8 760 hours; green = better, red = worse, per column.
- Without the 2 DST days the MAE of AW-LEAR is 22.48; the ranking does not change.

(Weron, 2014, IJF; Lago, Marcjasz, De Schutter & Weron, 2021, APEN)

Experiment yourself: notebook §5

Gradient boosting: many small trees, each correcting the previous ones

- **In one sentence:** Trees capture thresholds and interactions (e.g. low wind and high load in the evening) that a linear model misses.

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \nu\, f_m(\mathbf{x})$$

$$f_m = \arg\min_f \sum_i \left(r_i^{(m)} - f(\mathbf{x}_i)\right)^2$$

- **where**
  - f_m = small regression tree
  - $r_i^{(m)} = Y_i - F_{m-1}(\mathbf{x}_i)$ = current residual
  - $\nu = 0.05$ = learning rate
  - 500 trees fixed in advance.

- **Our gradient boosting machine (GBM)**
  - One model for all hours, with the hour as a feature; 20 features, all available at 12:00 on $d-1$
  - Refitted at the start of each of the 12 test months on all days before it (expanding window)
- **Result**
  - MAE 24.05, rMAE 0.864; behind AW-LEAR over the year
  - But the best single model in the stress week (MAE 34.35)

- **In practice:** trees are worth adding to an ensemble even when they do not win alone.

(Friedman, 2001, Ann. Stat.; Lago et al., 2021, APEN)

Experiment yourself: notebook §6

Neural networks: 24 prices out of one input vector

- **In one sentence:** A multi-output network forecasts the whole day at once, like the deep neural network (DNN) of Lago et al.

$$\hat{\mathbf{Y}}_d = W_3\,\phi\Big(W_2\,\phi(W_1\mathbf{x}_d + b_1) + b_2\Big) + b_3 \in \mathbb{R}^{24}$$

- **where**
  - $\mathbf{x}_d$ = the 247 LEAR regressors + 4 fundamentals of $d-2$
  - ϕ = rectified linear unit (ReLU)
  - Layers of 128 and 64 neurons
  - Multilayer perceptron (MLP).

- **Time-series discipline**
  - Epochs chosen on the last 15% of the training days, in time order (median 20)
  - 4 seeds averaged; refitted monthly, scaling fitted on training days only

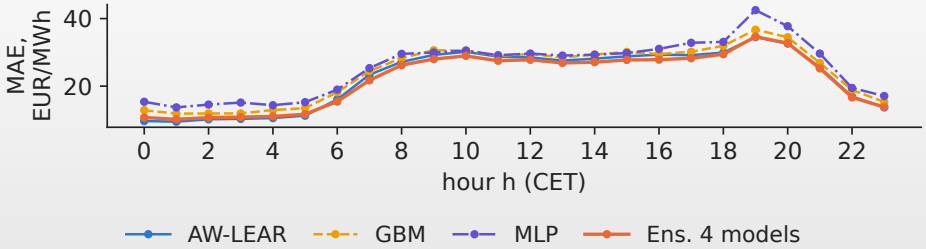
- **Result**
  - MAE 25.54, rMAE 0.918; a small untuned network does not beat LEAR
  - The tuned DNN of Lago et al. (2021) does better (Lecture 7)

- **In practice:** a neural network needs years of data and careful tuning, or you keep LEAR.

(Lago, Marcjasz, De Schutter & Weron, 2021, APEN)

Experiment yourself: notebook §6

Accuracy by hour: machine-learning models alone and in the ensemble



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E generation and flows (Transelectrica data for Romania); Yahoo Finance (SparkChange EUA ETC). Data: test year 2025-11-01 to 2026-10-31; ML refitted monthly.

- Better than AW-LEAR in 0 hours (GBM), 0 (MLP) and 15 of 24 (ensemble of 4).
- The hardest hour is 19 h CET, the evening ramp (AW-LEAR MAE 34.52 EUR/MWh).

- **In practice:** put your best analyst on the evening hours, where the money is lost.

(own computation, test year 1 Nov 2025 – 31 Oct 2026)

EMQA_L2_ml_hour

Research from our group on Romanian load: machine learning is most accurate for smooth series

- **The study** (Proșcanu, Proșcanu, Pele & Costea, 2025)
  - Hourly national load of Romania, 2022–2024; classical seasonal models against neural networks
- **Our replication** (studiu de caz C2, same gate-closure rule)
  - 217 test days from 27 May 2024; load metered up to $d - 2$
  - MAPE: gradient boosting 4.09%, Chronos-Bolt foundation model 4.16%, MLP 4.58%
  - Linear regression 4.85%, seasonal autoregressive model 8.03%, naive two days earlier 10.17%
- **Why the contrast with prices**
  - Load is smooth and never near zero, so MAPE is meaningful and nonlinear models have patterns to learn
  - Prices jump and turn negative; there, a well-built linear model is hard to beat
- C2 figures are computed by the C2 case notebook on our data; the paper's own results are in the case brief.

- **In practice:** forecast your own load with ML and the price with LEAR plus an ensemble.

(Proșcanu, Proșcanu, Pele & Costea, 2025, RJEF; case study C2)

Outline

- The forecasting problem and its benchmarks
- Expert models: ARX and LEAR
- Windows, combinations and machine learning
- Probabilistic forecasts and cointegration
  - Quantile regression averaging
  - Cointegration: prices that move together
- Case studies on recent Romanian data

Quantile regression averaging: from point forecasts to quantiles

- **In one sentence:** Regress the actual price on several point forecasts, once per quantile level, and the forecasts become a distribution.

$$\hat{q}^{\tau}_{d,h} = \beta^{\tau}_{0,h} + \beta^{\tau}_{1,h}\hat{P}^{\text{AW-ARX}}_{d,h} + \beta^{\tau}_{2,h}\hat{P}^{\text{AW-LEAR}}_{d,h}$$

$$\text{PS}_{\tau}(q, y) = \big(\mathbb{1}\{y < q\} - \tau\big)(q - y)$$

- **where**
  - β^{τ}_{h} minimises the pinball score (PS) on the last 182 days of forecasts, re-estimated weekly
  - $\tau = 0.01, \ldots, 0.99$.

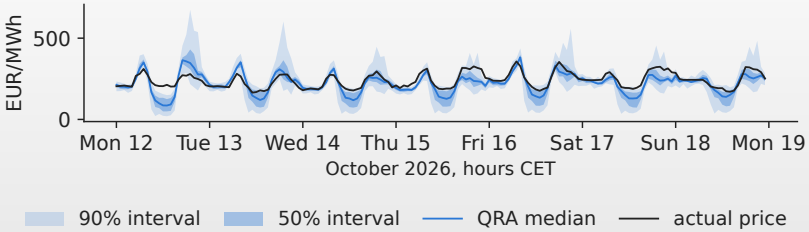
- **On our data**, average pinball score
  - QRA 8.33; conformal prediction (CP) 8.47; historical simulation (HS) 8.41
  - CP and HS use the AW-LEAR errors of the last 112 days, as in Assignment 1
- **Coverage of the 90% interval**
  - QRA 85.7%, CP 87.0%, HS 85.8%; mean QRA width 93.0 EUR/MWh
  - All below 90%; the Kupiec test rejects correct coverage

- **In practice:** a 90% band that holds 86% of the prices understates the risk of the open position.

(Koenker & Bassett, 1978; Nowotarski & Weron, 2015)

Experiment yourself: notebook §7

A week of QRA intervals in Romania: wide in the evening, narrow at night



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E load, wind and solar (Transelectrica data for Romania).
Data: week 2026-10-12 to 2026-10-18, hourly.

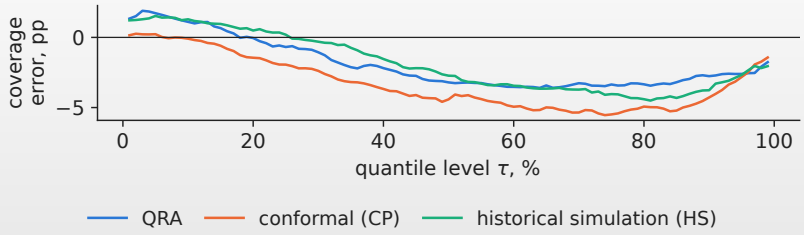
- Week 12 Oct 2026 – 18 Oct 2026
  - The band widens where the evening price is uncertain and narrows in the calm night hours.
- The QRA median has MAE 22.14 EUR/MWh, as good as the best point forecast.

- **In practice:** bid the 50% band for volumes you must buy, the tails for what you can shift.

(Nowotarski & Weron, 2015, Comput. Stat.)

EMQA_L2_qra_week

All three methods are too narrow in the tails



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E load, wind and solar (Transelectrica data for Romania).
Data: test year 2025-11-01 to 2026-10-31, 99 quantiles.

- Coverage error = share of prices below the τ-quantile minus τ (0 = calibrated)
  - Above 0 on the left and below 0 on the right = tails too narrow.
- Kupiec statistic for the 90% interval: QRA 162.9, CP 83.3, HS 154.1 (5% critical value 3.84).

- **In practice:** widen the tails before using quantiles for collateral (Lecture 4).

(Kupiec, 1995, J. Derivatives; Gneiting & Raftery, 2007, JASA)

EMQA_L2_prob_rel

Two trending prices are cointegrated when their spread keeps returning

- **In one sentence:** Each price may wander, but if a combination is stationary the gap keeps closing and the regression is not spurious.

$$\underbrace{P_t^{\mathrm{RO}} = \alpha + \beta P_t^{\mathrm{HU}} + u_t}_{\text{step 1, OLS}}$$

$$\underbrace{\Delta \hat{u}_t = \rho\, \hat{u}_{t-1} + \sum\nolimits_j \gamma_j \Delta \hat{u}_{t-j} + \eta_t}_{\text{step 2, unit root in } \hat{u}}$$

- **where**
  - Engle–Granger, H_0 no cointegration ($\rho = 0$), MacKinnon critical values
  - Johansen, trace test on the rank r of a vector error-correction model.

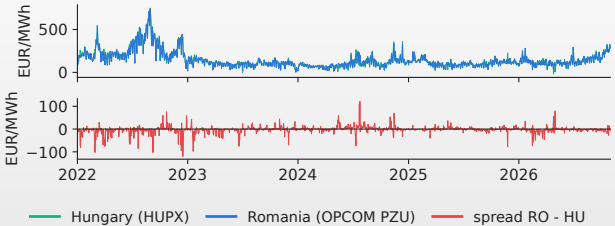
- **Romania vs Hungary**, daily, since 1 Jan 2022
  - $\hat{\beta} = 0.963$; Engle–Granger statistic -10.58 (5% critical value -3.34): cointegrated
  - A shock to the spread halves in 0.8 days
- **Johansen trace test**
  - $r = 0$: 139.7 vs 15.49; $r \leq 1$: 6.21 vs 3.84
  - Rank 2, the levels are nearly stationary since 2022 (augmented Dickey–Fuller (ADF) $p = 0.064$)

- **In practice:** a RO–HU spread trade closes within days, not months.

(Engle & Granger, 1987; Johansen, 1991; MacKinnon, 1996)

Experiment yourself: notebook §8

Romania and Hungary: cointegrated prices in coupled markets



Source: energy-charts.info (Fraunhofer ISE); prices of OPCOM (RO), HUPX (HU) and EPEX SPOT (DE-LU), via ENTSO-E. Data: daily base prices, 2022-01-01 to 2026-10-31.

- Spread RO − HU: mean −2.07, standard deviation 15.08 EUR/MWh
  - Romania vs Germany–Luxembourg: 21.06 and 37.79.
- In the test year RO and HU prices were identical in 10.1% of the hours
  - The rest reflects congested borders.

- **In practice:** use the Hungarian price as a regressor or a check on your Romanian forecast.

(Engle & Granger, 1987, Econometrica)

EMQA_L2_coint_daily

Power, gas and hydro: an error-correction model on monthly data

- **In one sentence:** If price and gas are cointegrated, the price is pulled back to its gas-implied level each month, and hydro shifts it in between.

$$\Delta P_t = c + \underbrace{\alpha\,(P_{t-1} - a - \beta\,\mathrm{TTF}_{t-1})}_{\text{error correction}} + \gamma\,\Delta\mathrm{TTF}_t + \delta\,H_t + \varepsilon_t$$

- **where**
  - P_t = monthly mean price of Romania
  - TTF in EUR/MWh
  - H_t = hydro output minus its calendar-month mean (TWh), stationary
  - $\alpha < 0$ = speed of return.

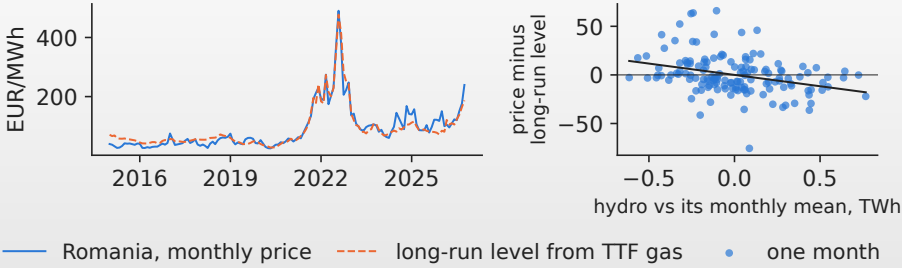
- **On our data**, 2015-01 – 2026-10
  - Engle–Granger -5.64 (5% critical value -3.38); Johansen rank 1
  - $\hat{\beta} = 1.95$, as if a gas plant of about 51% efficiency set the price
- **Error-correction model (ECM)**
  - $\hat{\alpha} = -0.452$, half-life 1.2 months
  - $\hat{\delta} = -12.80$ EUR/MWh per TWh of extra hydro ($t = -3.10$)

- **In practice:** a hydro producer in a dry month can expect prices above the gas-implied level for one or two months.

(Engle & Granger, 1987; Johansen, 1991, Econometrica)

Experiment yourself: notebook §8

Power, gas and hydro: the Lecture 1 regression is not spurious



Source: energy-charts.info (OPCOM prices; ENTSO-E hydro generation, Transelectrica data); World Bank Pink Sheet (TTF). Data: monthly, 2015-01 to 2026-10.

- Left, the monthly price follows its gas-implied level
  - Right, months with more hydro than usual sit below it.

- **In practice:** read a dry hydro year as a price premium over gas, not as a new price level.

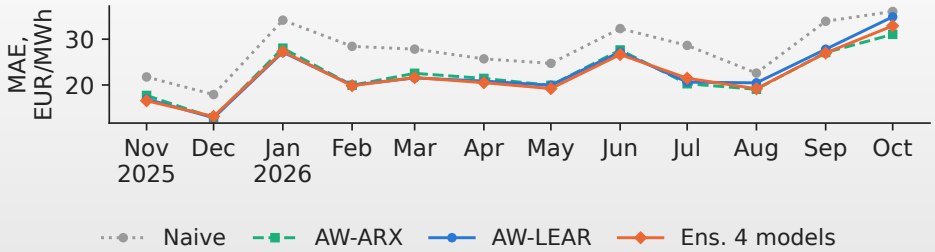
(Lecture 1, hydro case; own estimation)

EMQA_L2_coint_monthly

Outline

- The forecasting problem and its benchmarks
- Expert models: ARX and LEAR
- Windows, combinations and machine learning
- Probabilistic forecasts and cointegration
- Case studies on recent Romanian data
  - (a) A year of forecasts
  - (b) A stress week
  - (c) The 15-minute market time unit
  - (d) Economic value of the forecast
  - Research and practice: AIDA 2026

Case study (a): a year of forecasts, month by month



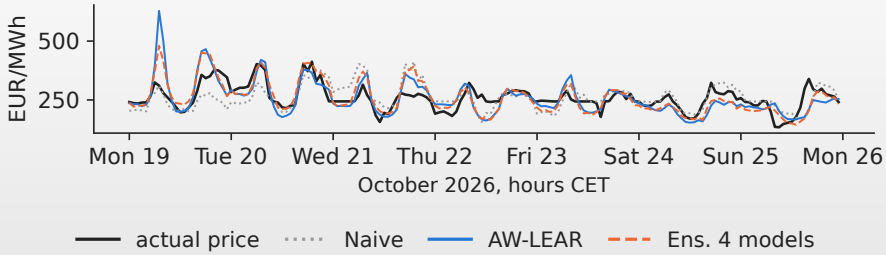
Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E generation and flows (Transelectrica data for Romania); Yahoo Finance (SparkChange EUA ETC). Data: test year 2025-11-01 to 2026-10-31.

- AW-LEAR beats naive in 12 of 12 months
  - MAE from 12.84 (Dec 2025) to 34.88 EUR/MWh (Oct 2026).

In practice: report accuracy by month, since a good annual MAE can hide a bad autumn.

(own computation, test year 1 Nov 2025 – 31 Oct 2026) EMQA_L2_monthly_mae

Case study (b): the stress week of 19 Oct 2026



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E generation and flows (Transelectrica data for Romania); Yahoo Finance (SparkChange EUA ETC). Data: week 2026-10-19 to 2026-10-25, hourly.

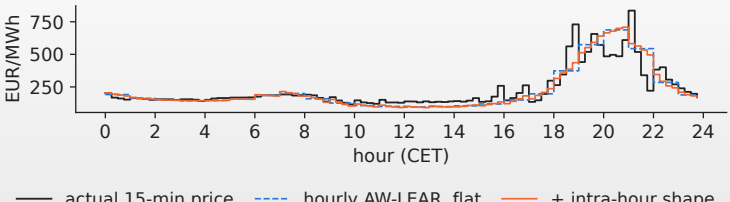
- Mean 258.59, peak 412.35 EUR/MWh
  - MAE naive 40.19, AW-LEAR 36.54, GBM 34.35, ensemble 34.95.

In practice: every model lags a jump in level by a day, so hedge the evening hours.

(own computation; OPCOM)

EMQA_L2_stress_week

Case study (c): forecasting 96 quarter-hours since 1 October 2025



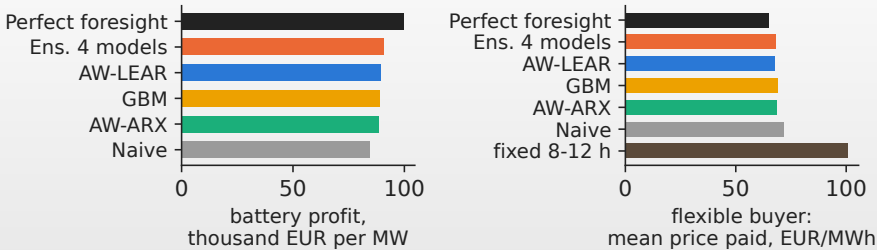
Source: energy-charts.info (Fraunhofer ISE); OPCOM 15-minute day-ahead prices for Romania, via ENTSO-E. Data: 2026-07-01, 96 quarter-hours (CET).

- The market time unit (MTU) is 15 minutes
  - Forecast = hourly AW-LEAR + the mean intra-hour shape of the 28 previous days.
- 363 test days without DST, quarter-hour MAE 26.02 flat vs 24.53 EUR/MWh shaped.

- **In practice:** a battery bids per quarter-hour
  - Give it the intra-hour shape, not a flat hour.

(own computation; OPCOM; SDAC 15-minute go-live, 1 Oct 2025) EMQA_L2_mtu15

Case study (d): economic value of a more accurate forecast



Source: energy-charts.info (Fraunhofer ISE); OPCOM prices; ENTSO-E generation and flows (Transelectrica data for Romania); Yahoo Finance (SparkChange EUA ETC). Data: test year 2025-11-01 to 2026-10-31; 1 MW / 2 MWh, round-trip efficiency 90%.

- Battery 1 MW / 2 MWh, one daily cycle, 363 days without DST
  - 90 738 EUR per MW with the ensemble (91.0% of perfect foresight), 84 623 with naive.
- Flexible 4-hour load: 67.87 EUR/MWh with AW-LEAR vs 100.59 for the fixed 8–12 h CET block.

- **In practice:** translate rMAE into EUR before paying for a forecast.

(own computation; round-trip efficiency 90% assumed)

EMQA_L2_battery

Forecasting checklist and a test for the use of future information

- **Before you trust a backtest**
  - Every input public before 12:00 CET on $d-1$
  - Scalers, transforms, λ and thresholds fitted in the window only
  - Chronological validation; the test year touched once
- **Then report**
  - MAE and rMAE against naive, DM p-values, coverage of the intervals
  - Results by month and by hour, not only one annual number

- **Our self-check** (analysis script, section 7)
  - 9 364 timestamp checks, each input at least 12 h before the cut-off
  - 17 forecast series recomputed on 14 days, GBM and MLP on the first day of all 12 refits
  - Later data deleted or replaced by nonsense; forecasts identical up to rounding
  - Status: passed

- **In practice:** make the leakage check part of every model review, as Assignment 1 does.

(Assignment 1; Lago et al., 2021, APEN)

Experiment yourself: notebook §10

AIDA 2026: economic evaluation of probabilistic forecasts

- **R. Weron, keynote** (4 June 2026, slides)
  - One model per hour (as our ARX, LEAR) or all 24 hours jointly; averaging across windows and LEAR stay the benchmarks to beat
- **What the cited work found**
  - Trading on quantiles earned 19–34% more than trading on point forecasts (Uniejewski & Weron, 2021)
  - Test year 2025: continuous ranked probability score (CRPS) 3.56 for the pretrained transformer TabPFN 3.0 vs 4.71 for LEAR + CP
- **A. Petukhina**, with G. Mazzon, D. Marazzina (3 June 2026, slides)
  - Great Britain, outside SDAC, runs two hourly day-ahead auctions (EPEX SPOT, Nord Pool): buy in one, sell in the other
  - One joint forecast of both auction prices and the system price (copula model); test Mar – Sep 2022
- **What it found**
  - EPEX auction: MAE 47.884 vs 27.926 for LASSO + QRA, but CRPS 16.852 vs 32.605
  - Trading on mean and conditional value at risk (CVaR): 22 001 EUR vs 6 283.3 and −3 005.1 EUR for the LASSO + QRA models

- **In practice:** judge a forecast by the decision it feeds: in Great Britain the model with the larger MAE earned more.

(as printed on the AIDA 2026 slides, Bucharest; papers: Uniejewski & Weron, 2021, ENEECO; Agakishiev et al., 2025, ENEECO)

AIDA 2026: a data-driven merit order forecasts the price distribution

- **P. Ghelasi**, with J. Browell, F. Ziel (3 June 2026, slides)
  - Price = merit order (plants sorted by variable cost) at the load: a fundamental, interpretable model
  - Efficiencies and CO_2 intensities estimated from past prices, within expert bounds
- **Germany, point forecasts**
  - Train 1 Oct 2022 – 1 Oct 2023, test 1 Oct 2023 – 1 Oct 2024, no recalibration
  - MAE, EUR/MWh: 11.83 for this model; 13.11 LASSO, 13.96 boosted trees, 16.36 classical merit order, 24.96 naive
- **Probabilistic merit order**
  - Forecast the distribution of each input (load, solar, wind, fuels); plug each sampled scenario into the merit order
  - Parameters chosen to minimise the CRPS of the resulting prices
- **What it found**
  - CRPS 13.92 vs 14.04 for a Johnson S_U distribution model and 20.46 for naive
  - MAE 17.28 vs 18.21 and 25.98; RMSE 25.81 vs 29.19 and 41.47
  - Simulated prices form non-parametric distributions, e.g. a cluster near zero at midday

In practice: a fundamental model estimated like a statistical one gives forecasts you can explain plant by plant.

(talk by P. Ghelasi at AIDA 2026, Bucharest, 3 June 2026: slides; paper: Ghelasi & Ziel, 2026, ENEECO)

Quiz 1: the naive rule on a Monday

- It is Sunday, 11:30 CET
  - You must forecast Monday, hour 19:00, with the naive benchmark
  - Which price do you use?

Answer

- The price of the previous Monday at 19:00 ($d - 7$)
  - On our test year this beats Sunday's price on Mon, Sat and Sun (MAE 29.87 vs 33.48 EUR/MWh).

Quiz 2: a suspicious backtest

- A vendor reports rMAE 0.55 for Romania and uses the realised load of day $d-1$ as a regressor
  - Do you buy?

Answer

- No: at 12:00 on $d-1$ that day is not yet metered, so the backtest leaks information
  - Ask for a rerun with data up to $d-2$ (or published day-ahead forecasts)
  - Our honest best is rMAE 0.795.

Quiz 3: reading a DM test

- AW-ARX or AW-LEAR? The DM test gives $p = 0.797$
  - Which model goes into production?

Answer

- Neither wins significantly
  - Take the simpler one, or better their average, which is part of the best ensemble (rMAE 0.795).

Quiz 4: an error measure for the board

- Your report shows MAPE 5 778% for the Romanian price forecast
  - The board is alarmed
  - What do you say?

Answer

- MAPE divides by prices near zero (387 hours below 5 EUR/MWh produce 99.5% of it)
  - Report MAE 22.53 EUR/MWh and rMAE 0.810 instead.

Quiz 5: cointegration in practice

- The RO–HU spread is cointegrated with a half-life of 0.8 days
  - Today Romania is 40 EUR/MWh above Hungary
  - What do you expect?

Answer

- The gap should shrink by about half within a day, unless the border stays congested
  - It is a signal for cross-border trading, not a guarantee.

Wrap-up

- The forecasting problem and its benchmarks
  - The day-ahead market
  - Information set and horizons
  - Naive benchmarks and error measures
  - Testing forecast accuracy
- Expert models: ARX and LEAR
  - The expert ARX model
  - Variance-stabilising transformations
  - LASSO-estimated autoregression
- Windows, combinations and machine learning
  - Calibration windows
  - Combining forecasts
  - Machine learning: a preview of Lecture 7
- Probabilistic forecasts and cointegration
  - Quantile regression averaging
  - Cointegration: prices that move together
- Case studies on recent Romanian data
  - (a) A year of forecasts
  - (b) A stress week
  - (c) The 15-minute market time unit
  - (d) Economic value of the forecast
  - Research and practice: AIDA 2026

- Honest setup first: information at 12:00 CET on $d-1$, rolling windows, naive benchmark, DM tests.
- On the test year the best model, an ensemble of 4, cuts the naive error by 20.5%
  - Next: volatility (Lecture 3).

Research from our group on forecasting Romanian power

- **Benchmarking load forecasts**: classical seasonal models against neural networks on Romania's national consumption
  - In the paper a neural network has a clearly lower error than the seasonal autoregressive benchmark (Proșcanu, Proșcanu, Pele & Costea, 2025, RJEF) — case study C2.
- **Four generations of price models**: a comparative evaluation of four generations of forecasting methods for energy prices, a reference for choosing a model family (Andrei, Velev, Toma, Pele & Lessmann, 2024, arXiv 2411.03372).
- **Explaining price forecasts across horizons**: which drivers matter at which horizon and in which price regime, for Romanian hourly prices (Andrei, Pele & Cramer, 2026, PICBE).
- **Early warning of price spikes**: an explainable artificial intelligence (AI) system in which the previous price and the residual load drive the spikes (Andrei, Pele, Cramer, Amza & Diaconu, 2026, PICBE) — preview of Lecture 7.
- Findings as stated in the papers; numbers on the other slides come from our own data.

(Bucharest University of Economic Studies; Romanian Academy, Institute for Economic Forecasting)

Further reading and resources

- **R. Weron, IIF Distinguished Lectures on electricity price forecasting** (2025): slides part I, part II, notebooks.
- **epftoolbox**, LEAR and DNN with the open benchmark of Lago et al. (2021): github.com/jeslago/epftoolbox.
- **Forecasting – Principles and Practice**, 3rd ed., otexts.com/fpp3.
- **This course**: code, data, notebooks — github.com/danpele/EMQA.
- **Assignment 1** applies ARX, LEAR, DM and the HS/CP intervals to your own test window.
- **Experiment yourself**: the notebook reproduces every chart and number of this lecture — notebooks/.

References (1/4)

- Andrei, A.-V., Pele, D.-T. & Cramer, A.-A. (2026). Multi-horizon explainability in energy price forecasting: Disentangling market dynamics and renewable intermittency in Romania's liberalized electricity market. *Proceedings of the International Conference on Business Excellence*, 20(1), 104–117. doi:10.2478/picbe-2026-0014
- Andrei, A.-V., Pele, D.-T., Cramer, A.-A., Amza, A. & Diaconu, D. (2026). Explainable AI-based early warning system for electricity price spikes: Evidence from the Romanian market. *Proceedings of the International Conference on Business Excellence*, 20(1), 4820–4836. doi:10.2478/picbe-2026-0352
- Andrei, A.-V., Velev, G., Toma, F.-M., Pele, D. T. & Lessmann, S. (2024). Energy price modelling: A comparative evaluation of four generations of forecasting methods. arXiv:2411.03372. arxiv.org/abs/2411.03372
- Agakishiev, I., Härdle, W. K., Kopa, M., Kozmík, K. & Petukhina, A. (2025). Multivariate probabilistic forecasting of electricity prices with trading applications. *Energy Economics*, 141, 108008. doi:10.1016/j.eneco.2024.108008
- Bates, J. M. & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, 20(4), 451–468. doi:10.1057/jors.1969.103
- 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
- Engle, R. F. & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2), 251–276. doi:10.2307/1913236
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. doi:10.1214/aos/1013203451
- Ghelasi, P., Browell, J. & Ziel, F. (2026). A probabilistic merit order for electricity price forecasting. AIDA Conference, Bucharest University of Economic Studies, 3 June 2026. github.com/danpele/EMQA/resources/aida2026

References (2/4)

- Ghelasi, P. & Ziel, F. (2026). A data-driven merit order: Learning a fundamental electricity price model. *Energy Economics*, 154, 109114. doi:10.1016/j.eneco.2025.109114
- Gneiting, T. & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. doi:10.1198/016214506000001437
- Hubicka, K., Marcjasz, G. & Weron, R. (2019). A note on averaging day-ahead electricity price forecasts across calibration windows. *IEEE Transactions on Sustainable Energy*, 10(1), 321–323. doi:10.1109/TSTE.2018.2869557
- 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
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6), 1551–1580. doi:10.2307/2938278
- Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50. doi:10.2307/1913643
- Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), 73–84. doi:10.3905/jod.1995.407942
- Lago, J., Marcjasz, G., De Schutter, B. & Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293, 116983. doi:10.1016/j.apenergy.2021.116983
- NEMO Committee (2025). 15-Minute MTU in SDAC was implemented. SDAC press release, 1 October 2025. nemo-committee.eu

References (3/4)

- NEMO Committee (2026). Single Day-Ahead Coupling (SDAC) and the EUPHEMIA algorithm. nemo-committee.eu/sdac
- MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, 11(6), 601–618. doi:10.1002/(SICI)1099-1255(199611)11:6<601::AID-JAE417>3.0.CO;2-T
- Marcjasz, G., Serafin, T. & Weron, R. (2018). Selection of calibration windows for day-ahead electricity price forecasting. *Energies*, 11(9), 2364. doi:10.3390/en11092364
- Mazzon, G., Petukhina, A. & Marazzina, D. (2026). Multivariate probabilistic day-ahead electricity price forecasting and cross-auction trading in the UK market. AIDA Conference, Bucharest University of Economic Studies, 3 June 2026. github.com/danpele/EMQA/resources/aida2026
- Nogales, F. J., Contreras, J., Conejo, A. J. & Espínola, R. (2002). Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, 17(2), 342–348. doi:10.1109/TPWRS.2002.1007902
- Nowotarski, J. & Weron, R. (2015). Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Computational Statistics*, 30(3), 791–803. doi:10.1007/s00180-014-0523-0
- Nowotarski, J. & Weron, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81, 1548–1568. doi:10.1016/j.rser.2017.05.234
- Proșcanu, C. A., Proșcanu, M. E., Pele, D. T. & Costea, A. (2025). Watt's next? Benchmarking time series models on Romania's national electricity consumption. *Romanian Journal of Economic Forecasting*, 28(3), 146–163. RePEc.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

References (4/4)

- Uniejewski, B., Nowotarski, J. & Weron, R. (2016). Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies*, 9(8), 621. doi:10.3390/en9080621
- Uniejewski, B., Weron, R. & Ziel, F. (2018). Variance stabilizing transformations for electricity spot price forecasting. *IEEE Transactions on Power Systems*, 33(2), 2219–2229. doi:10.1109/TPWRS.2017.2734563
- Uniejewski, B. & Weron, R. (2021). Regularized quantile regression averaging for probabilistic electricity price forecasting. *Energy Economics*, 95, 105121. doi:10.1016/j.eneco.2021.105121
- Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), 1030–1081. doi:10.1016/j.ijforecast.2014.08.008
- Weron, R. (2026). Machine learning for electricity price forecasting. Keynote, AIDA Conference, Bucharest University of Economic Studies, 4 June 2026. github.com/danpele/EMQA/resources/aida2026
- Ziel, F. & Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, 70, 396–420. doi:10.1016/j.eneco.2017.12.016

Appendix A: LEAR in full

$$\mathrm{AIC}(\lambda) = n \ln \frac{\mathrm{RSS}(\lambda)}{n} + 2\,\mathrm{df}(\lambda)$$

- Residual sum of squares (RSS)
  - The degrees of freedom $\mathrm{df}(\lambda)$ = number of non-zero coefficients
  - Path by least angle regression (LARS), in scikit-learn.
- Price block: m, s = median and 1.4826·MAD of the prices of rows $d - T - 7, \ldots, d - 1$, pooled over 24 hours.
- Load and wind + solar: own asinh transform from rows up to $d - 2$
  - Then every column z-scored on the calibration rows.

Block	Columns
prices $d-1$, all 24 hours	24
prices $d-2$, $d-3$, $d-7$	72
load $d-2$, $d-3$, $d-7$	72
wind + solar $d-2$, $d-3$, $d-7$	72
day of week	7
total	**247**

(Uniejewski, Nowotarski & Weron, 2016, Energies; Lago et al., 2021, APEN)

Appendix B: MAE by quarter of the test year

Model	2025 Q4	2026 Q1	2026 Q2	2026 Q3	2026 Q4
naive	19.80	30.19	27.55	28.32	36.05
AW-ARX	15.24	23.62	22.97	22.09	31.05
AW-LEAR	14.86	22.99	22.64	22.91	34.88
GBM	16.25	24.71	22.98	24.97	37.88
MLP	17.27	25.49	24.89	26.30	41.58
ensemble of 4	14.82	23.02	22.09	22.50	32.91

- Days per quarter: 61, 90, 91, 92, 31 (the test year starts on 1 Nov 2025).
- Green = best in that quarter, red = worst.

(Weron, 2014, IJF; Lago, Marcjasz, De Schutter & Weron, 2021, APEN)

Appendix C: probabilistic scores and the Kupiec test

$$\mathrm{LR}_{\mathrm{UC}} = -2\ln\frac{(1-p_0)^{n-x}p_0^{x}}{(1-\hat{p})^{n-x}\hat{p}^{x}} \sim \chi^2_1, \quad \hat{p} = \frac{x}{n}$$

- Likelihood ratio (LR) for unconditional coverage (UC)
  - $x =$ hours outside the 90% interval out of $n =$ 8 760
  - $p_0 = 0.10$.
- Hours of the same day are not independent, so the test is indicative
  - A test on the daily miss share with Newey–West variance gives $p = <0.001$ for QRA.
- Pinball score averaged over 99 quantiles
  - The continuous ranked probability score $\approx 2\times$ this score.

(Kupiec, 1995; Gneiting & Raftery, 2007, JASA)

Method	Pinball	Cov. 50%	Cov. 90%	Kupiec
QRA	8.33	47.3	85.7	162.9
CP	8.47	46.5	87.0	83.3
HS	8.41	45.8	85.8	154.1

Appendix D: unit-root and cointegration tests in full

Series / pair	Test	Statistic	p / crit. 5%
RO, HU, DE-LU daily levels	ADF p	–	0.064, 0.052, 0.015
RO–HU	Engle–Granger	-10.58	-3.34
RO–HU	Johansen $r = 0$ / $r \leq 1$	139.7 / 6.21	15.49 / 3.84
RO–DE-LU	Engle–Granger	-5.79	-3.34
RO price, TTF, monthly levels	ADF p	–	0.295, 0.249
hydro minus monthly mean	ADF p	–	0.002
price–TTF	Engle–Granger	-5.64	-3.38
price–TTF–hydro	Engle–Granger	-5.95	-3.80
price–TTF	Johansen $r = 0$ / $r \leq 1$	21.32 / 3.74	15.49 / 3.84

- ADF with AIC lags; Johansen with a constant, 6 lagged differences (daily) or 2 (monthly).
- ECM with heteroskedasticity- and autocorrelation-robust standard errors (6 lags), 141 months, $R^2 = 0.778$.

(Engle & Granger, 1987; Johansen, 1991; MacKinnon, 1996)