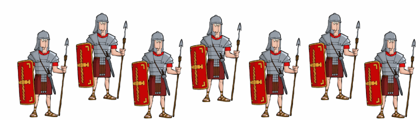
Recent advances in electricity price forecasting: A 2025 perspective

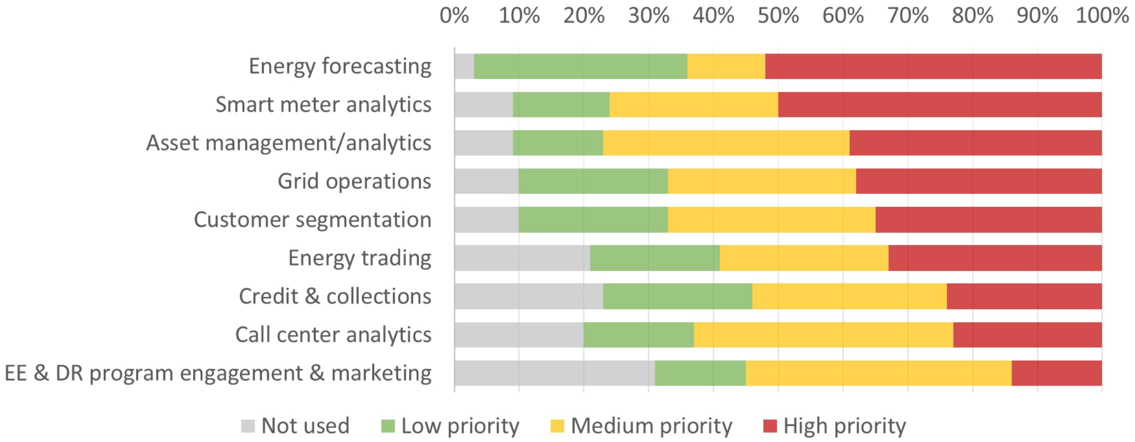
Rafał Weron*,**
Department of Operations Research and Business Intelligence, Wrocław Tech, Poland
https://p.wz.pwr.edu.pl/∼weron.rafal



* Based on joint work with K.Bilińska, Y.Chawla, K.Chęć, A.Lipiecki, K.Maciejowska, W.Nitka, T.Serafin, B.Uniejewski, P.Zaleski (Wrocław Tech), C.Challu, K.Olivares (Nixtla), T.Hong (UNCC), K.Hubicka (UBS), A.Jędrzejewski (U.Aveiro), C.Kath (RWE), J.Lago (Amazon), G.Marcjasz (Alpiq), M.Narajewski (Statkraft), J.Nasiadka (Nokia), J.Nowotarski (BNY Mellon), H.Zareipour (U.Calgary), F.Ziel (U.Duisburg-Essen)
** Julia/Python notebooks coded by A.Lipiecki

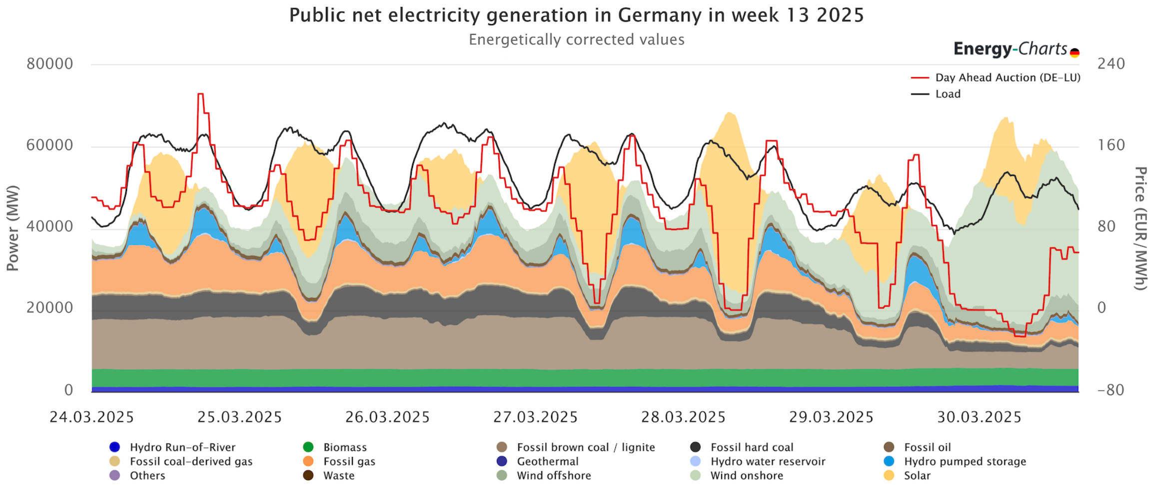
Energy forecasting is a top priority for utilities

(SAS, 2017, White Paper; responses from 136 utilities from 24 countries)




THE WORLD IS CHANGING ...

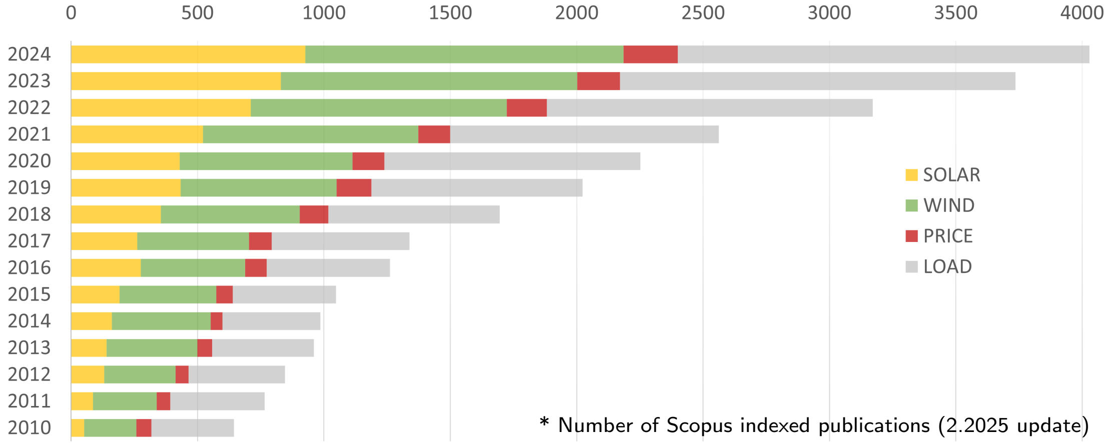
... increased wind and solar penetration take a toll ...



… but the (Academic) Empire

Energy (load, price, wind & solar) forecasting*

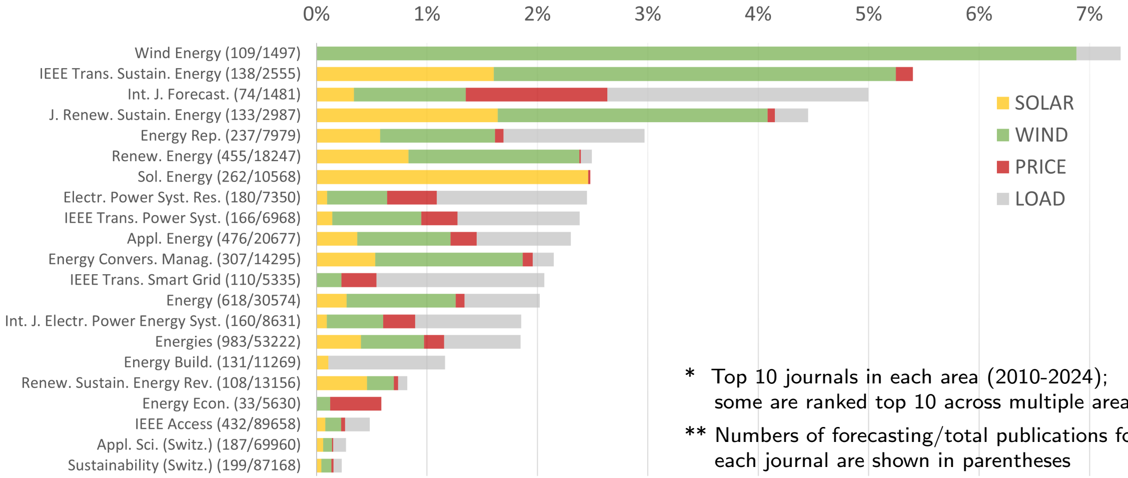
(Hong, Pinson, Wang, Weron, Yang & Zareipour, 2020, IEEE OAJPE)



* Number of Scopus indexed publications (2.2025 update)

Percentage of energy forecasting publications*

(Hong, Pinson, Wang, Weron, Yang & Zareipour, 2020, IEEE OAJPE)

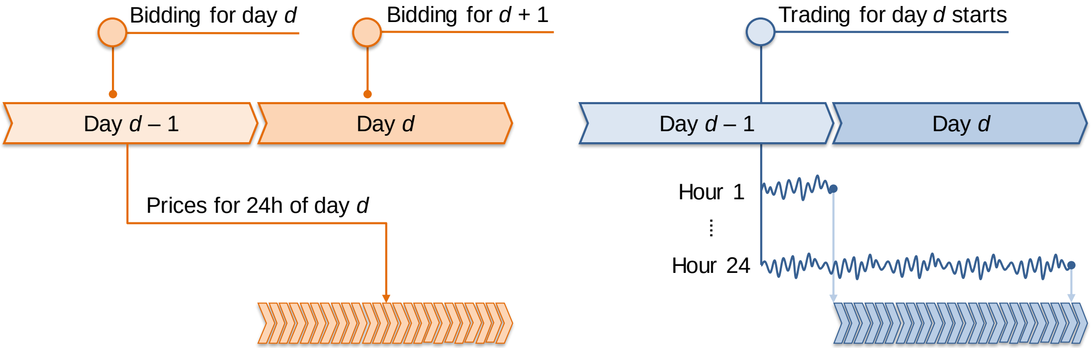


* Top 10 journals in each area (2010-2024); some are ranked top 10 across multiple areas

** Numbers of forecasting/total publications for each journal are shown in parentheses

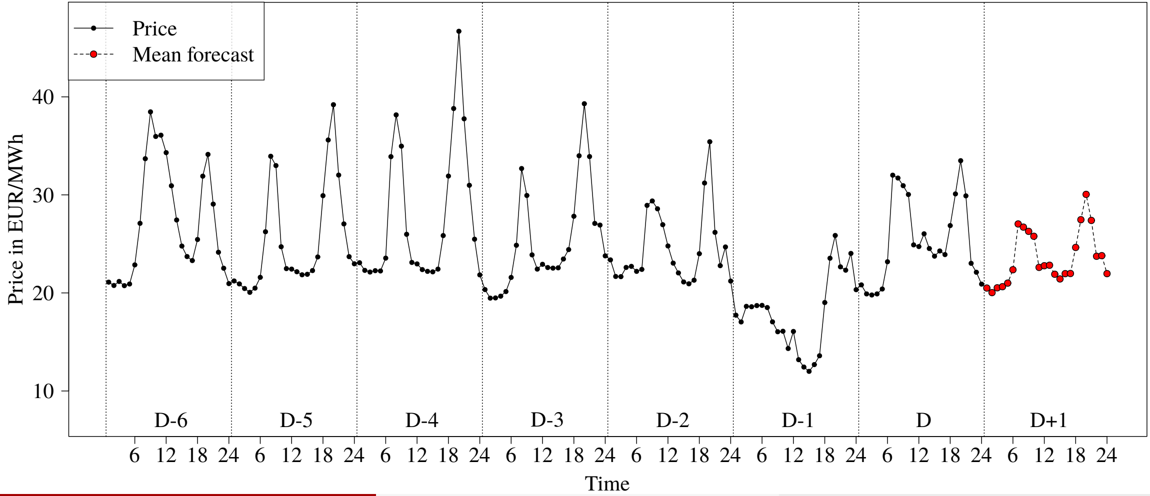
Day-ahead (> 90% of papers) vs. intraday (real-time) markets

(Maciejowska, Uniejewski & Weron, 2023, Oxford Res. Enc.)

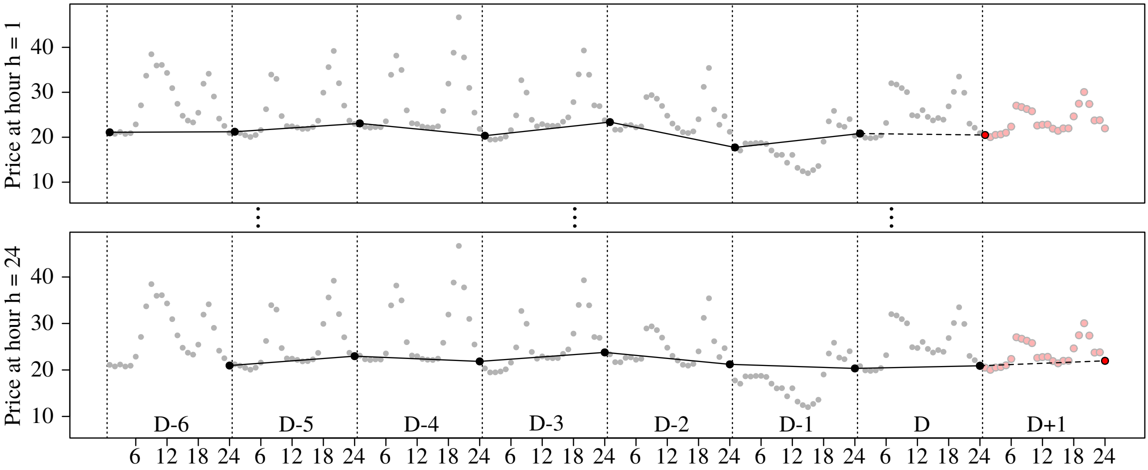


Day-ahead point forecasting: Univariate ...

(Ziel & Weron, 2018, ENEECO)

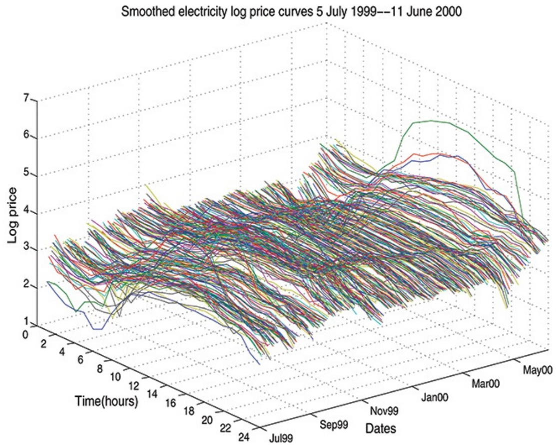


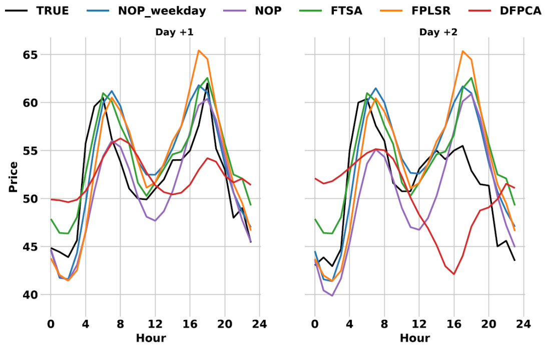
... multivariate ...



... functional (data analysis) ...

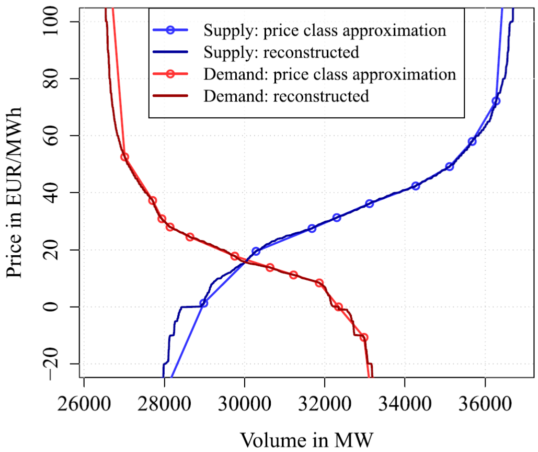
(Chen & Li, 2017, JBES; Chen et al., 2019, Ann.Appl.Stat; Wang & Cao, 2023, Environmetrics)

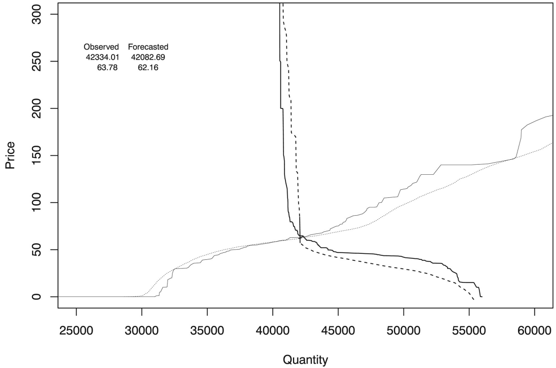




… or supply & demand curves?

(Ziel & Steinert, 2016, ENEECO → 'X-model'; Shah & Lisi, 2020, J.Forecasting)





1 Introduction

2 Model evolution

- (Auto)regressive models
- Averaging across windows
- Shrinkage (regularization)
- LASSO-Estimated AR (LEAR)
- Deep and interpretable ML

3 Beyond point forecasts

4 Financial evaluation

International Journal of Forecasting 30 (2014) 1030–1081

Contents lists available at ScienceDirect **(2014)**

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Review

Electricity price forecasting: A review of the state-of-the-art with a look into the future

Rafał Weron

Renewable and Sustainable Energy Reviews 81 (2018) 1548–1568

Contents lists available at ScienceDirect **(2018)**

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Recent advances in electricity price forecasting: A review of probabilistic forecasting

Jakub Nowotarski, Rafał Weron*

Applied Energy 293 (2021) 116983

Contents lists available at ScienceDirect **(2021)**

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark

Jesus Lago [a,*], Grzegorz Marcjasz [b], Bart De Schutter [a], Rafał Weron [b]

[a] Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands
[b] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Wrocław, Poland

ARTICLE INFO

Keywords:

ABSTRACT

While the field of electricity price forecasting has benefited from plenty of contributions in the last two decades,

Expert ARX-type models

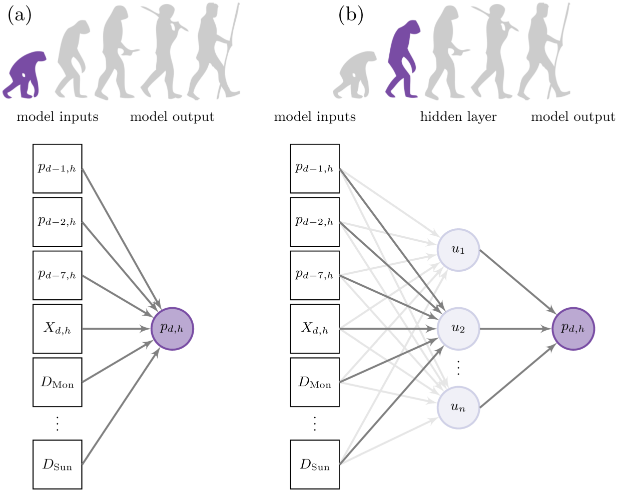
Consider an autoregressive structure with exogenous variables:

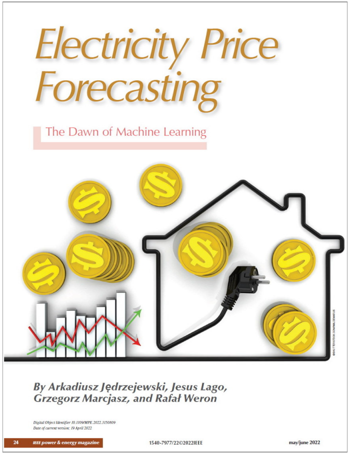
$$P_{d,h} = \beta_0 + \underbrace{\sum_{i=1}^{7} \beta_i P_{d-i,h}}_{\text{AutoRegressive effects}} + \underbrace{\sum_{i=1}^{7} \beta_{i+7} D_i}_{D_1 = \text{Mon}, \ldots} + \underbrace{\sum_{i=1}^{K} \beta_{i+14} X_{d,h}^{(i)}}_{\text{eXogenous variables}} + \varepsilon_{d,h}$$

- There can be no more dummies than categories → set $\beta_0 = 0$ if all D_i's are used
    - Also set $\beta_0 = 0$ if the mean is removed from $P_{d,h}$ beforehand
- Special cases:
    - Naive model $P_{d,h}^{(1)}$ for $\beta_1 = 1$ and $\beta_{i \neq 1} = 0$
    - AR(7), sparse AR(7) with some AR lags missing
    - ARX(7) with $K \geqslant 1$, sparse ARX(7) with some AR lags missing

Linear regression vs. single-output (shallow) neural network

(Jędrzejewski, Lago, Marcjasz & Weron, 2022, IEEE-PEM)




Electricity Price Forecasting

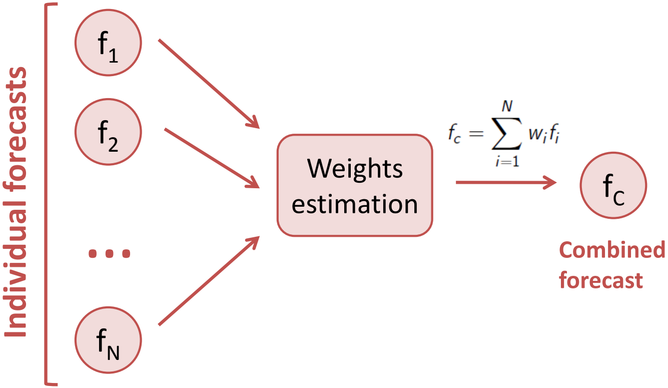
The Dawn of Machine Learning

By Arkadiusz Jędrzejewski, Jesus Lago, Grzegorz Marcjasz, and Rafał Weron

Digital Object Identifier 10.1109/MPE.2022.3150809
Date of current version: 19 April 2022

24 IEEE power & energy magazine 1540-7977/22©2022IEEE may/june 2022

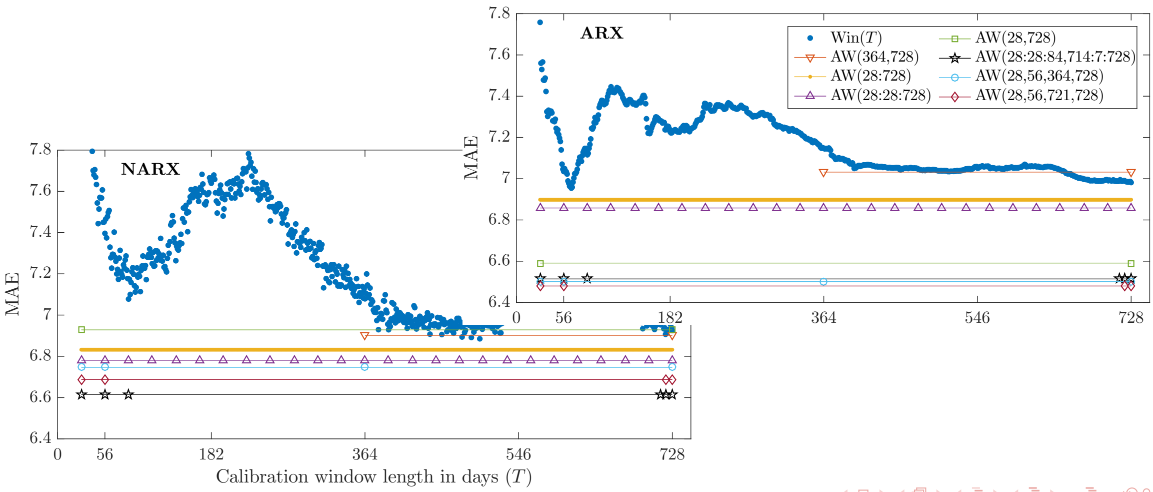
Point forecast averaging: The idea



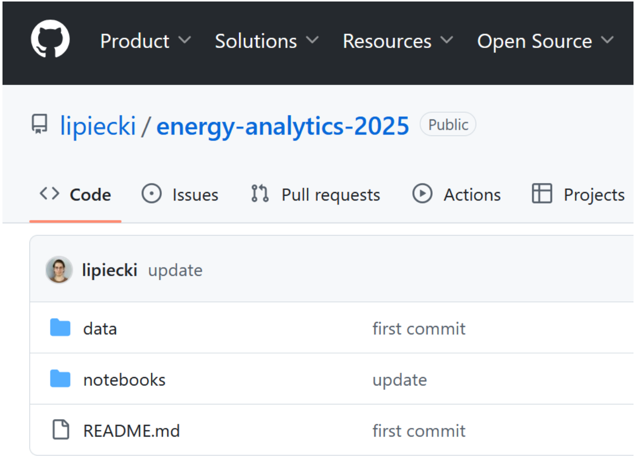
- Dates back to the works of Bates, Crane, Crotty & Granger (1960s)
- 'Statistical/Econometric world': *combining forecasts*, *forecast averaging* (not *model averaging*)
- 'AI/Engineering world': *committee machines*, *ensemble averaging*, *expert aggregation*

Simple averaging across calibration windows

Structural breaks: Pesaran, Pick (2011, JBES); EPF: Hubicka et al. (2019, IEEE-TSE)



Python snippet: Averaging.ipynb



What is shrinkage (regularization)?

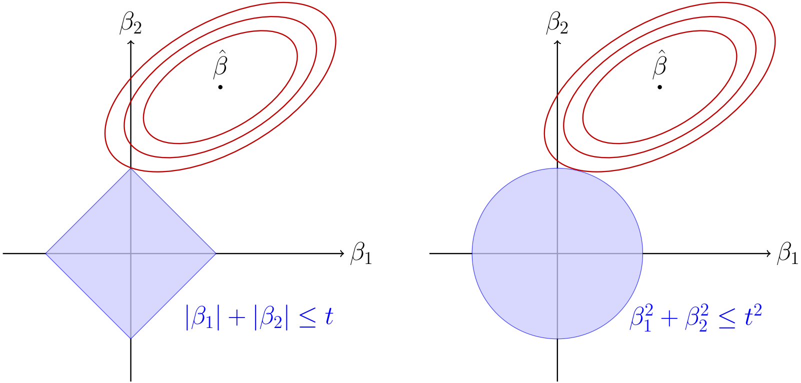
- Minimize the residual sum of squares (RSS $\equiv$ sum of squared errors) $+$ a penalty function of the betas:

$$\hat{\beta} = \underset{\beta_j}{\text{argmin}} \left\{ \underbrace{\sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} \beta_j x_{i,j} \right)^2}_{RSS} + \lambda \underbrace{\sum_{j=1}^{n} |\beta_j|^q}_{penalty} \right\}$$

  where $\lambda \geqslant 0$ is a tuning (or regularization) parameter
- Most popular:
  - $q = 2 \rightarrow$ Ridge regression (Hoerl & Kennard, 1970, Technometrics)
  - $q = 1 \rightarrow$ Least absolute shrinkage & selection operator (Tibshirani, 1996, JRSSB)

How does it work?

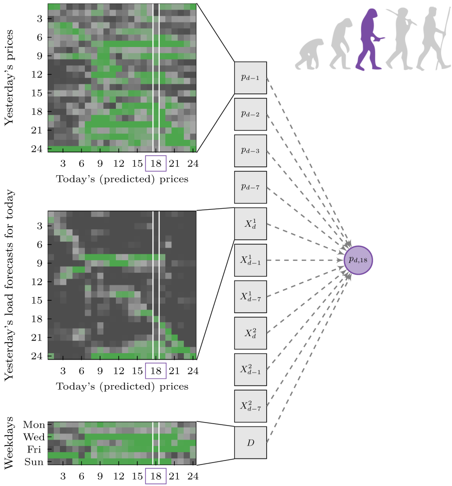
Lasso ($q = 1$) Ridge regression ($q = 2$)



- Blue areas
  constraint regions
- Red ellipses
  contours of the least squares error function

LASSO-Estimated AR (LEAR)

(Uniejewski et al., 2016; Ziel, 2016; Ziel & Weron, 2018; Jędrzejewski et al., 2022)



Note: Whole 24-hourly vectors are inputs, not only values for the target hour $h = 18$

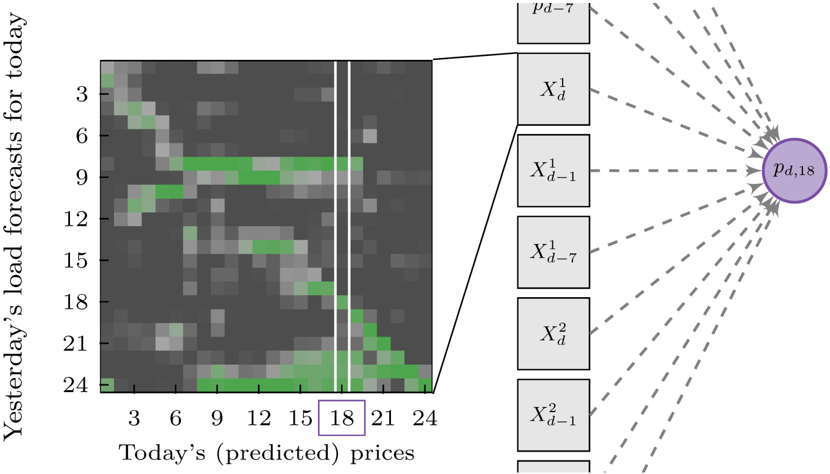
Uniejewski (2024, ORD) compares 10 penalties and finds that only Elastic Net and LQ with $1 < q < 2$:

$$\hat{\beta}^{LQ} = \underset{\beta_j}{\operatorname{argmin}} \left\{ RSS + \lambda \sum_{j=1}^{n} |\beta_j|^q \right\}$$

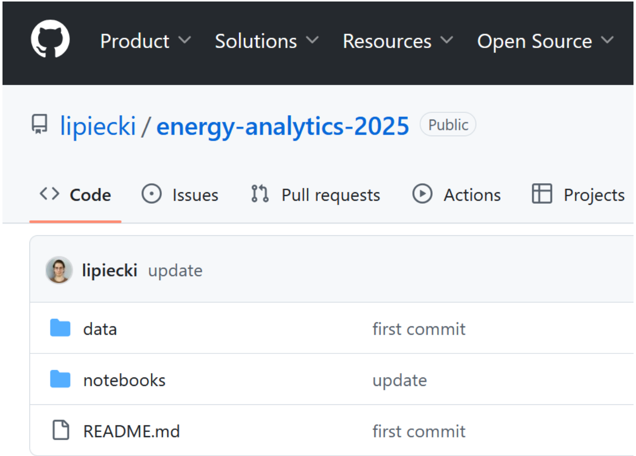
consistently outperform LASSO, but at the cost of one more hyperparameter (α or q)

LASSO-Estimated AR (LEAR): Variable importance

(Uniejewski et al., 2016; Ziel, 2016; Ziel & Weron, 2018; Jędrzejewski et al., 2022; Uniejewski, 2024)

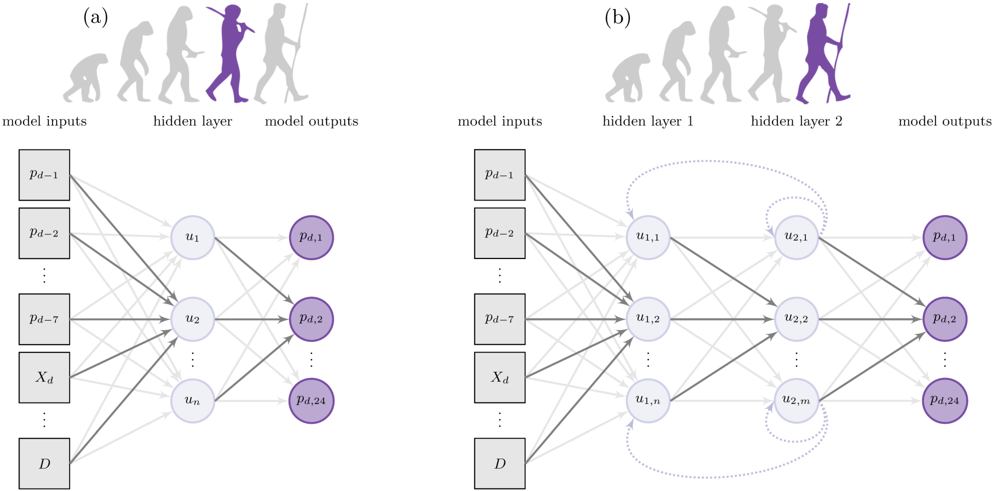


Python snippet: Shrinkage.ipynb



Multi-output shallow and deep neural networks (DNNs)

(Lago et al., 2021, APEN; Jędrzejewski et al., 2022, IEEE-PEM)



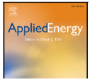
What about performance?

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

(2021)



Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark

Jesus Lago [a,*], Grzegorz Marcjasz [b], Bart De Schutter [a], Rafał Weron [b]

[a] Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands
[b] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Wrocław, Poland

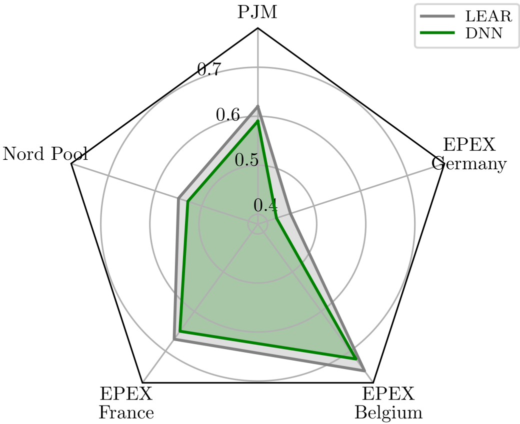
ARTICLE INFO

Keywords:
Electricity price forecasting
Regression model
Deep learning
Open-access benchmark
Forecast evaluation
Best practices

ABSTRACT

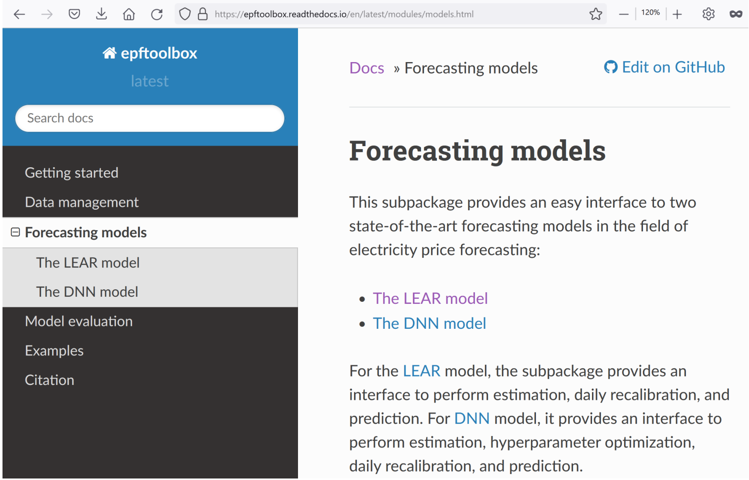
While the field of electricity price forecasting has benefited from plenty of c
it arguably lacks a rigorous approach to evaluating new predictive algorit
using unique, not publicly available datasets and across too short and l
The proposed new methods are rarely benchmarked against well establ
models, the accuracy metrics are sometimes inadequate and testing the sig
performance is seldom conducted. Consequently, it is not clear which met
best practices when forecasting electricity prices. In this paper, we tack
of-the-art statistical and deep learning methods across multiple years and
set of best practices. In addition, we make available the considered datas
models, and a specifically designed python toolbox, so that new algorit
future studies.

rMAE (relative to a naive forecast)



epftoolbox: LEAR and DNN Python codes

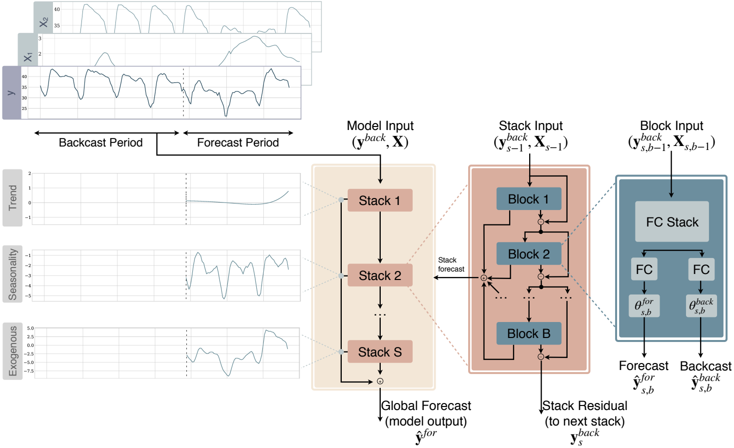
(Lago, Marcjasz, De Schutter & Weron, 2021, APEN)



Experiment yourself:

Interpretable AI: NBEATSx

(Olivares, Challu, Marcjasz, Weron & Dubrawski, 2023, IJF)



Experiment yourself:

Interpretable ML: What about performance?

(Olivares, Challu, Marcjasz, Weron & Dubrawski, 2023, IJF)

Tab. 3. Accuracy measures for day-ahead price forecasts (test sample of 2 years)

		ARx1	LEARx	DNN	NBEATSx
NP	MAE	2.01	1.74	1.68	1.62
	rMAE	0.63	0.55	0.53	0.51
	sMAPE	5.84	5.01	4.88	4.70
	RMSE	3.71	3.36	3.32	3.27
PJM	MAE	3.53	3.01	2.86	2.90
	rMAE	0.73	0.62	0.59	0.60
	sMAPE	13.64	11.98	11.33	11.61
	RMSE	5.74	5.13	5.04	4.84
EPEX-DE	MAE	4.36	3.61	3.41	3.29
	rMAE	0.54	0.45	0.42	0.41
	sMAPE	17.73	14.74	14.08	13.99
	RMSE	7.38	6.51	5.93	5.65
Daily recalibration [s]		—	18.57	50.65	81.61

LEARx, **DNN** – LEAR and DNN models from Lago et al. (2021, APEN), after Erratum
NBEATSx – NBEATSx-I (interpretable configuration) from Olivares et al. (2023, IJF)

International Journal of Forecasting 30 (2014) 1030–1081

Contents lists available at ScienceDirect

(2014)

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Review

Electricity price forecasting: A review of the state-of-the-art with a look into the future

Rafał Weron

Renewable and Sustainable Energy Reviews 81 (2018) 1548–1568

Contents lists available at ScienceDirect

(2018)

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Recent advances in electricity price forecasting: A review of probabilistic forecasting

Jakub Nowotarski, Rafał Weron*

Department of O

ARTICL

Keywords:
Electricity price f
Probabilistic fore
Reliability
Sharpness
Day-ahead marke
Autoregression
Neural network

ng (EPF) over
o explain the
pportunities
he paper also
next decade
dies involving
iii) statistical

Applied Energy 293 (2021) 116983

Contents lists available at ScienceDirect

(2021)

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark

Jesus Lago [a,*], Grzegorz Marcjasz [b], Bart De Schutter [a], Rafał Weron [b]

[a] Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands
[b] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Wrocław, Poland

ARTICLE INFO

Keywords:

ABSTRACT

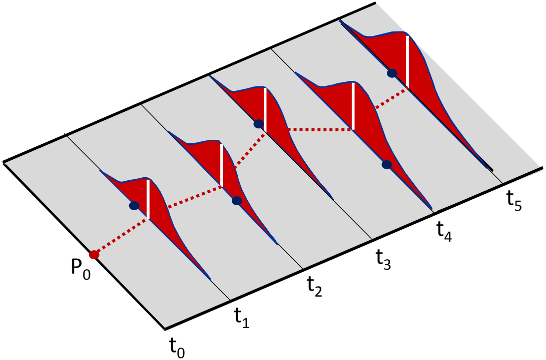
While the field of electricity price forecasting has benefited from plenty of contributions in the last two decades,

Probabilistic (interval, density) forecasting

(Gneiting & Katzfuss, 2014, Annu Rev)

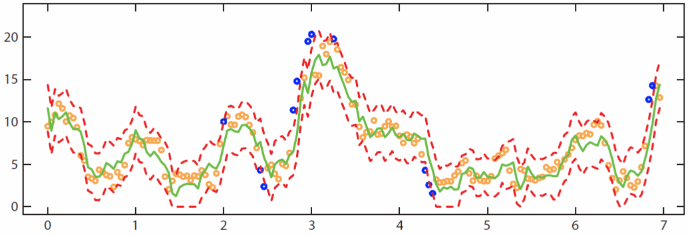
We cannot observe the true underlying distribution $\Rightarrow$ we cannot compare the *predictive distribution* $\hat{F}$ with the actual one F ... only with past observations

Gneiting et al. (2007a, 2007b, 2014) argue that probabilistic forecasting aims to '*maximize the sharpness of the predictive distributions, subject to reliability*'



Reliability (calibration, unbiasedness)

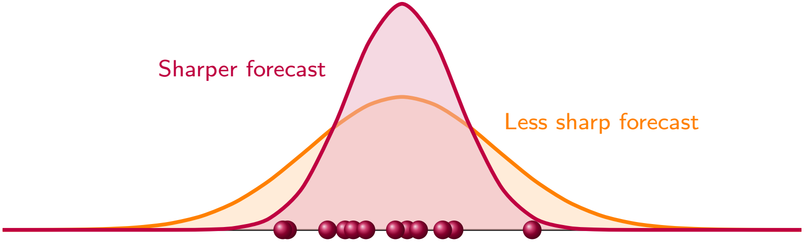
- Refers to the statistical consistency between $\hat{F}$ and the observations
- If a 90% PI covers 90% of the observed prices, then this PI is said to be:
  - reliable (Pinson et al., 2007; Pinson & Kariniotakis, 2010)
  - well calibrated (Gneiting et al., 2007a, 2007b, 2014)
  - unbiased (Taylor, 1999)
- Example: 13 ∘ or 'misses' and 155 ∘ or 'hits' $\longrightarrow$ the coverage is $\frac{155}{168} \approx 92\%$



Sharpness

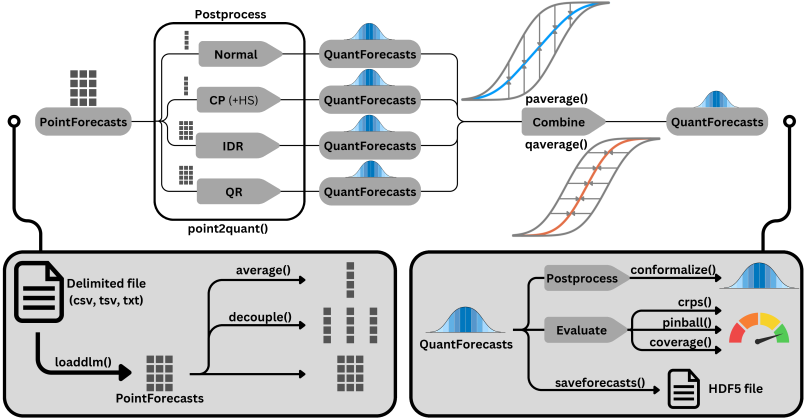
(Pinson et al., 2007, Wind En; Gneiting & Raftery, 2007, JASA; Gneiting & Katzfuss, 2014, Annu Rev)

- Refers to the concentration or tightness of the predictive distributions
  - Derives from the idea that reliable predictive distributions of null width correspond to perfect point predictions
  - Reliability is a joint property of the predictions and the observations
  - Sharpness is a property of the forecasts only



Postprocessing point forecasts

(Vannitsem et al., 2018, Elsevier; Chen et al., 2024, Ann Appl Stat; Lipiecki et al., 2024, ENEECO)



Experiment yourself:

The 'normal' benchmark

- Assume that the prediction errors are $N(\mu, \sigma^2)$-distributed
- Training corresponds to estimating $\hat{\mu}$ and $\hat{\sigma}$ of $\varepsilon_t = y_t - \hat{y}_t$ for $t \in \mathcal{S}$
  - $\mathcal{S}$ is the training set (or calibration window)
- The τ-th quantile conditional on $\hat{y}_t$ is obtained via:

$$\hat{q}_{\tau|\hat{y}_t} = \hat{y}_t + \hat{\mu} + \hat{\sigma} F_N^{-1}(\tau)$$

where $F_N^{-1}(\tau)$ is the inverse of the standard normal CDF, i.e., with $\mu = 0, \sigma = 1$

Historical simulation

(Hendricks, 1996, EPR; Alexander, 2008, Wiley; Nowotarski & Weron, 2018, RSER)

- A model-independent approach that computes

$$\hat{q}_{\tau|\hat{y}_t} = \hat{y}_t + Q_\tau(\varepsilon_t)$$

where $Q_\tau(\varepsilon_t)$ is the sample τ-quantile of $\varepsilon_t = y_t - \hat{y}_t$ for $t \in \mathcal{S}$

- The term historical simulation (HS) can be traced back to the beginnings of VaR
- Similar to bootstrapped residuals (see, e.g., Hyndman & Athanasopoulos, 2021, FPP3), but each ε_t is sampled exactly once

FEDERAL RESERVE BANK *of* NEW YORK

ECONOMIC POLICY REVIEW

Evaluation of Value-at-Risk Models Using Historical Data

April 1996Volume 2, Number 1

JEL classification: G11, G15, G28

Author: Darryll Hendricks

Recent studies have underscored the need for market participants to develop reliable methods of measuring risk. One increasingly popular technique is the use of "value-at-risk" models, which convey estimates of market risk for an entire portfolio in one number. The author explores how well these models actually perform by applying twelve value-at-risk approaches to 1,000 randomly chosen foreign exchange portfolios. Using nine criteria to evaluate model performance, he finds that the approaches generally capture the risk that they set out to assess and tend to produce risk estimates that are similar in average size. No approach, however, appears to be superior by every measure.

Conformal prediction

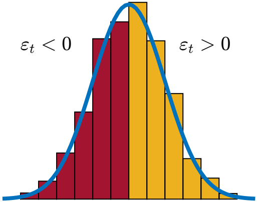
(Vovk et al., 2005, Springer; Kath & Ziel, 2021, IJF; Lipiecki et al., 2024, ENEECO)

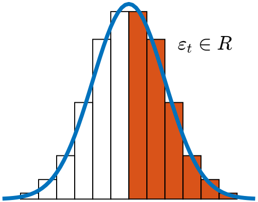
- For $t \in \mathcal{S}$ calculate the so-called non-conformity scores $\lambda_t = |\varepsilon_t| = |y_t - \hat{y}_t|$, then compute

$$\hat{q}_{\tau|haty_t} = \hat{y}_t - \mathbb{1}_{\tau \leqslant 0.5} Q_{2\tau}(\lambda) + \mathbb{1}_{\tau \geqslant 0.5} Q_{2(1-\tau)}(\lambda)$$

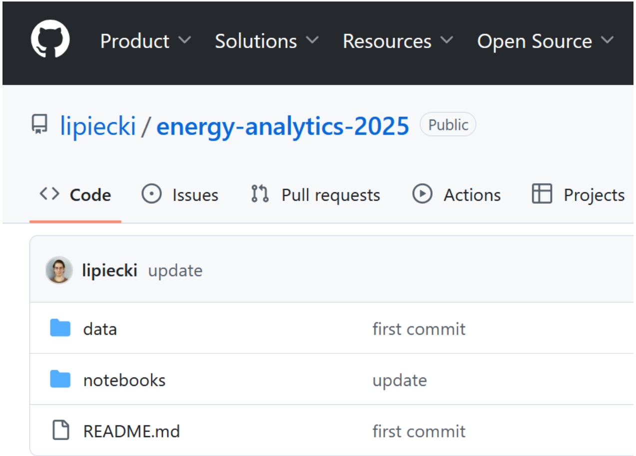
  where $Q_\tau(\lambda)$ is the τ-th sample quantile of λ_t

- This version is called inductive or split CP, however, a 'split' is not needed if $\hat{y}_t$'s are already available
- HS works with ε_t's, CP with $|\varepsilon_t|$'s $\longrightarrow$ symmetric $\hat{F}$
- Using $\varepsilon_1, \ldots, \varepsilon_T$, HS approximates the whole distribution, CP only the positive half $\longrightarrow$ smoother $\hat{F}$

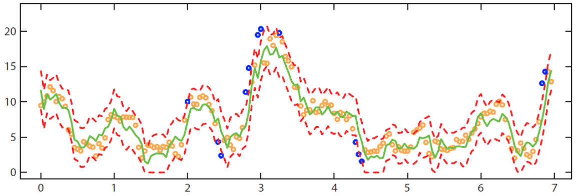




Python snippet: UncertaintyQuantification.ipynb



Unconditional coverage (UC)



- (Empirical) coverage is measured by

$$I_t = \begin{cases} 1 & \text{if } P_t \in \mathrm{PI} \rightarrow \circ \text{ 'hit'} \\ 0 & \text{if } P_t \notin \mathrm{PI} \rightarrow \circ \text{ 'miss'} \end{cases}$$

and should match the nominal rate

$$\mathbb{P}(P_t \in \mathrm{PI}) = \mathbb{P}(I_t = 1) = (1 - \alpha)$$

- Some studies report only the so-called PI Coverage Probability

$$\mathbf{PICP} = \frac{1}{T}\sum_{t=1}^{T} I_t \cdot 100\%$$

- Other subtract it from the nominal coverage to obtain the so-called Average Coverage Error

$$\mathbf{ACE} = \mathrm{PICP} - \mathrm{PINC}$$

where PINC – PI Nominal Coverage

UC and the Kupiec test

(Kupiec, 1995, J Derivatives)

- Checks whether ACE = 0 or $\mathbb{P}(I_t = 1) = (1-\alpha)$, given that ∘ are independent
  - Equivalent to testing that I_t is i.i.d. Bernoulli with mean $(1-\alpha)$
  - Rejects the null ('good PI') if the percent of misses is statistically different from α
- The likelihood ratio statistics for unconditional coverage:

$$LR_{UC} = -2\log\left\{\frac{(1-c)^{n_0}c^{n_1}}{(1-\pi)^{n_0}\pi^{n_1}}\right\} \sim \chi^2(1)$$

  - $c = (1-\alpha)$ is the nominal and $\pi = \frac{n_1}{n_0+n_1}$ the empirical coverage rate
  - n_0 and n_1 are respectively the number of 0's and 1's in I_t
- Independence, conditional coverage → Christoffersen (1998, IER) test

Continuous Ranked Probability Score (CRPS)

(Gneiting & Raftery, 2007, JASA; Gneiting & Katzfuss, 2014, Annu Rev; Nitka & Weron, 2023, ORD)

- The CRPS is the standard metric for evaluating probabilistic forecasts:

$$\text{CRPS}(\hat{F}, x) = \int_{-\infty}^{\infty} \left(\hat{F}(y) - \mathbb{1}_{\{x \leqslant y\}}\right)^2 dy$$

  where $\hat{F}$ is the predictive distribution and x is the observation, e.g., electricity price
- It is a proper scoring rule, i.e., quoting the true distribution as the forecast is an optimal strategy in expectation
- Problem: in practice we often work with a finite set of quantile forecasts

CRPS and the pinball score

(Gneiting & Raftery, 2007, JASA; Nowotarski & Weron, 2018, RSER; Nitka & Weron, 2023, ORD)

- The CRPS can be approximated by:

$$\mathrm{CRPS}(\hat{F}, x) \approx \frac{2}{M} \sum_{i=1}^{M} \mathrm{PS}\left(\hat{q}, x, q_i\right)$$

  where
  - $q_1 < \ldots < q_M$ is an equidistant dense grid of probabilities, e.g., 99 percentiles
  - $\hat{q} \equiv \hat{F}^{-1}(q)$ is the quantile forecast for quantile level $q \in (0, 1)$

  and the pinball score is defined as:

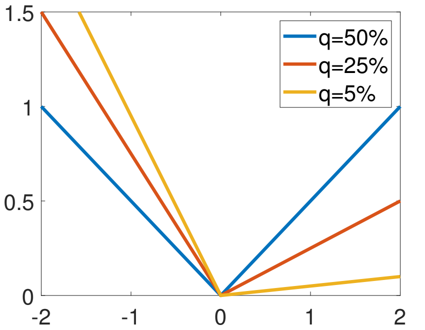
$$\mathrm{PS}(\hat{q}, x, q) = \left(\mathbb{1}_{\{x<\hat{q}\}} - q\right)(\hat{q} - x)$$

- Note: The scaling factor of 2 is usually omitted in practice

Pinball score (or loss) in more detail

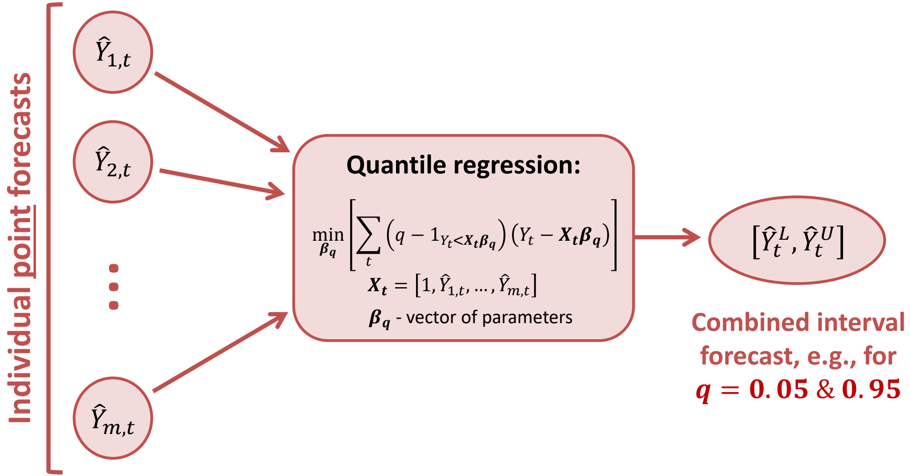
$$\text{PS}\left(\hat{q}, x, q\right) = \left(\mathbb{1}_{\{x<\hat{q}\}} - q\right)\left(\hat{q} - x\right) = \begin{cases} \left(1 - q\right)\left(\hat{q} - x\right) & \text{for } x < \hat{q} \\ q\left(x - \hat{q}\right) & \text{for } x \geqslant \hat{q} \end{cases}$$

- Also known as the quantile score, check function or the linlin/bilinear/newsboy loss
- For an Aggregate PS (or APS) average:
  - across all t in the test period
  - across all quantiles $\rightarrow$ CRPS

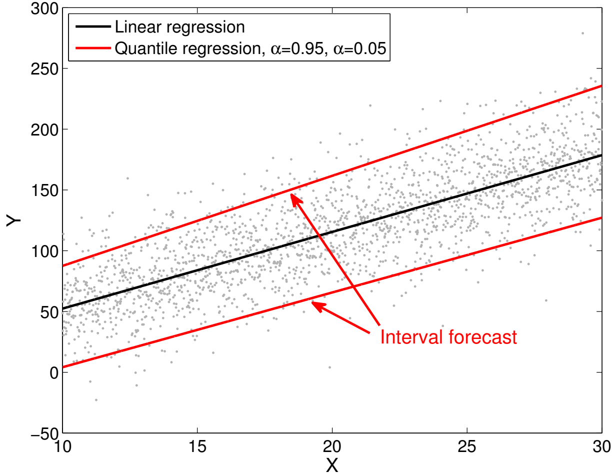


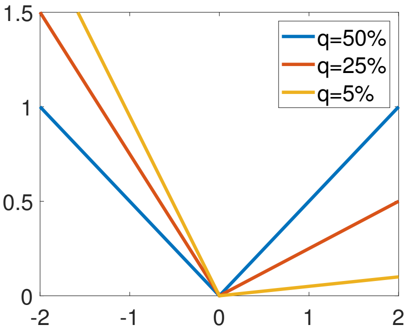
Quantile Regression Averaging (QRA): The idea

(Nowotarski & Weron, 2015, COST)



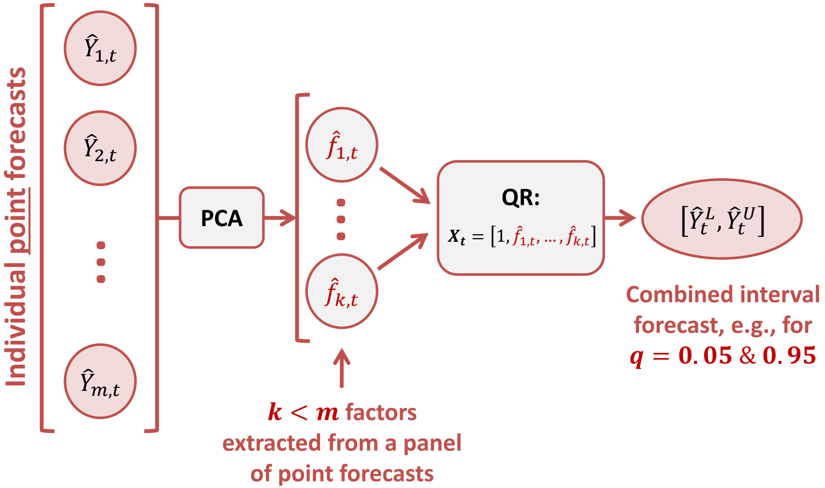
Quantile regression





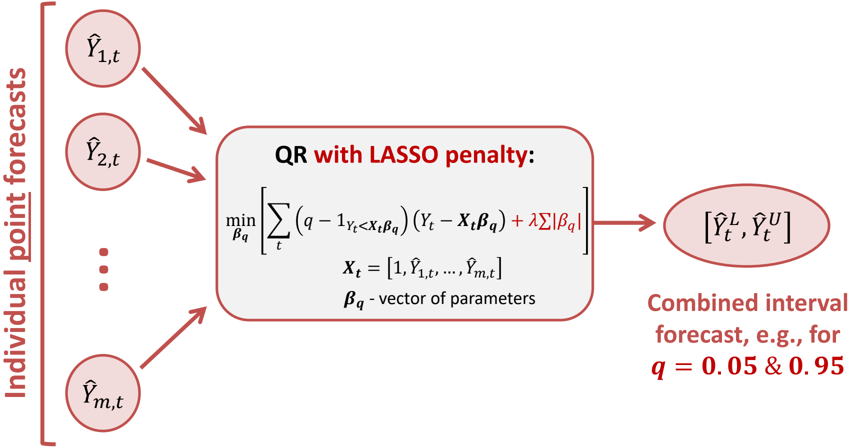
Factor QRA (FQRA): When the number of predictors is large

(Maciejowska, Nowotarski & Weron, 2016, IJF)



LASSO QRA (LQRA): When the number of predictors is large

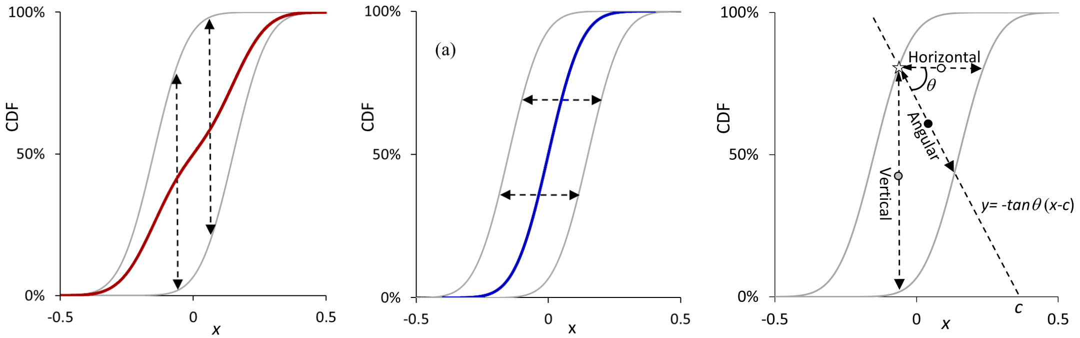
(Uniejewski & Weron, 2021, ENEECO) $\sim 5\%$ improvement over QRA



Combining probabilistic forecasts is tricky ...

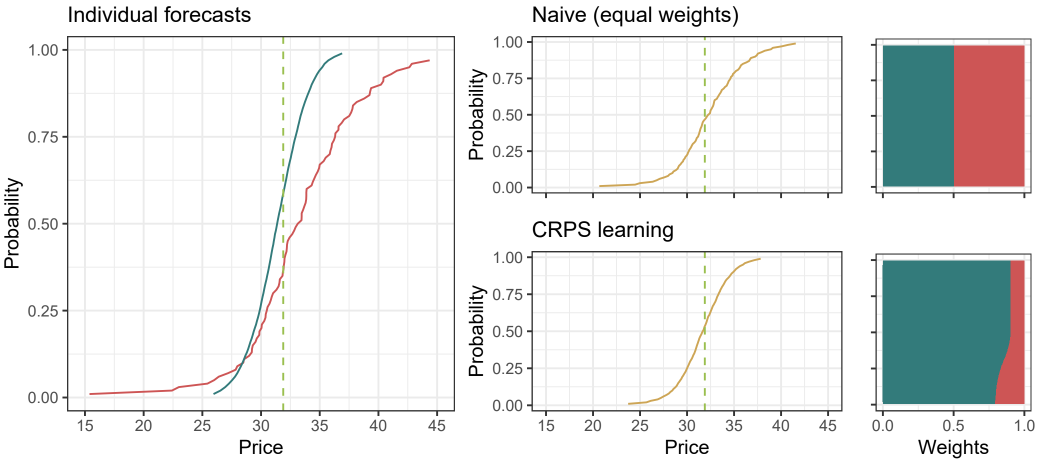
(Lichtendahl et al., 2013, Mgnt Sci; Uniejewski et al., 2019, ENEECO; Taylor & Meng, 2023, arXiv)

Vertical (over probabilities), **horizontal** (quantiles → sharper CDF) ... or at any angle



... can use unequal weights (here: CRPS learning) ...

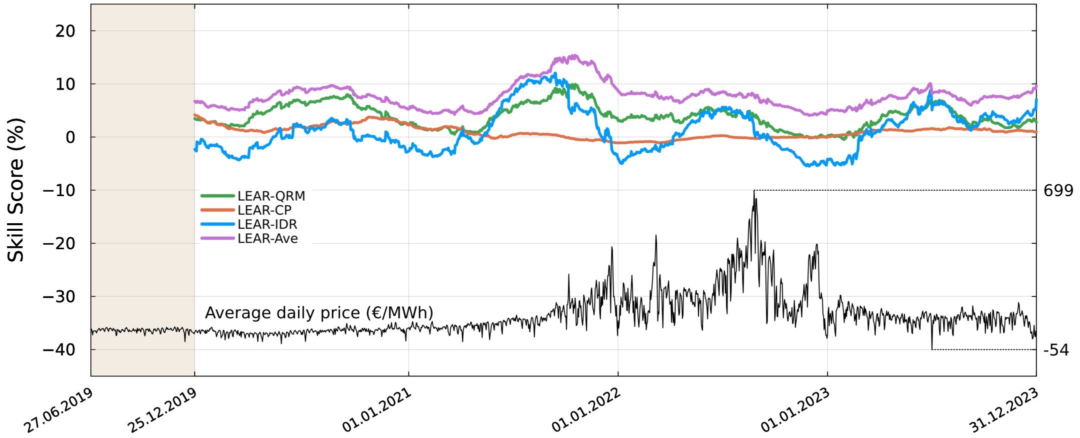
(Berrisch & Ziel, 2023, JEconometrics; Nitka & Weron, 2023, ORD)



... but can be beneficial

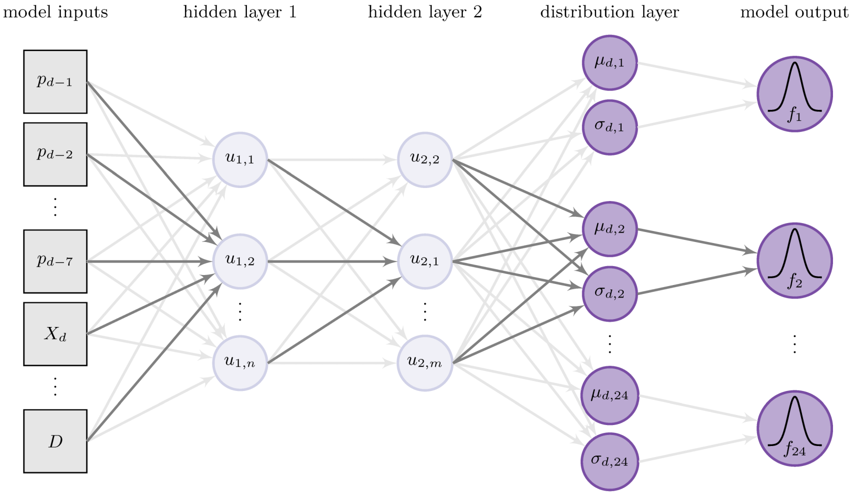
(Lipiecki, Uniejewski & Weron, 2024, ENEECO; illustrated here using EPEX-DE data)

Skill score: % difference wrt the CRPS of $N(0,\hat{\sigma}(\varepsilon_{d,h}))$ innovations; $\varepsilon_{d,h}$ – prediction errors of the LEAR model



Distributional Deep Neural Nets (DDNN)

(Jędrzejewski et al., 2022, IEEE-PEM; Marcjasz, Narajewski, Weron & Ziel, 2023, ENEECO)



DDNNs can perform very well ...

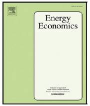
(Marcjasz, Narajewski, Weron & Ziel, 2023, ENEECO)

Energy Economics 125 (2023) 106843

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco



Distributional neural networks for electricity price forecasting

Grzegorz Marcjasz [a,*], Michał Narajewski [b], Rafał Weron [a], Florian Ziel [b]

[a] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, 50-370 Wrocław, Poland
[b] House of Energy Markets and Finance, University of Duisburg–Essen, 45141 Essen, Germany

A R T I C L E I N F O

JEL classification:
C44
C45
C46
C22
C53
Q47

Keywords:
Distributional neural network
Probabilistic forecasting

A B S T R A C T

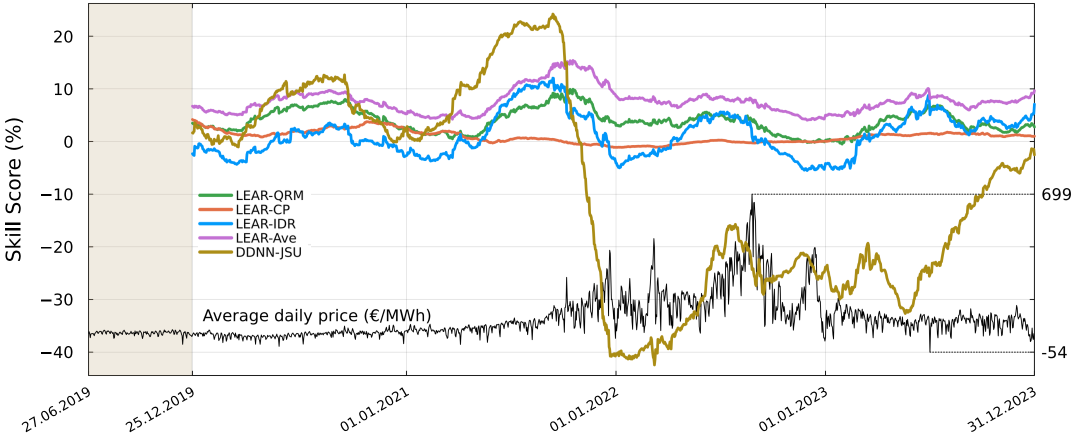
We present a novel approach to probabilistic electricity price forecasting which utilizes distributional neural networks. The model structure is based on a deep neural network containing a so-called probability layer, i.e., the outputs of the network are parameters of the normal or Johnson's SU distribution. To validate our approach, we conduct a comprehensive forecasting study complemented by a realistic trading simulation with day-ahead electricity prices in the German market. The proposed distributional deep neural network outperforms state-of-the-art benchmarks by over 7% in terms of the continuous ranked probability score and by 8% in terms of the per-transaction profits. The obtained results not only emphasize the importance of higher moments when modeling volatile electricity prices, but also – given that probabilistic forecasting is the essence of risk management – provide important implications for managing portfolios in the power sector.

Experiment yourself:

... but can fail miserably during volatile periods

(Lipiecki, Uniejewski & Weron, 2024, ENEECO; illustrated here using EPEX-DE data)

Skill score: % difference wrt the CRPS of N(0,$\hat{\sigma}(\varepsilon_{d,h})$) innovations; $\varepsilon_{d,h}$ – prediction errors of the LEAR model



What about probabilistic inputs?

(Uniejewski & Ziel, 2025, arXiv)

Expert model, i.e., ARX-type, for the day-ahead electricity price

$$P_{d,h} = \beta_1 P_{d-1,h} + \beta_2 P_{d-2,h} + \beta_3 P_{d-7,h} + \beta_4 P_{d-1,24} + \beta_5 P_{d-1}^{min} + \beta_6 P_{d-1}^{max} + \beta_7 \widehat{L}_{d,h} + \beta_8 \widehat{S}_{d,h}$$
$$+\beta_9 \widehat{W}_{d,h} + \beta_{10} EUA_{d-2} + \beta_{11} Gas_{d-2} + \beta_{12} Oil_{d-2} + \beta_{13} Coal_{d-2} + \sum_{i=1}^{7} \beta_{13+i} D_i + \varepsilon_{d,h}$$

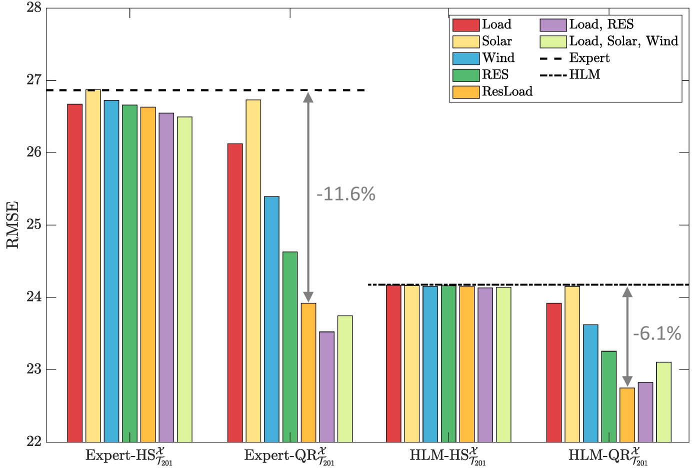
High-dimensional linear model (HLM), i.e., LEAR-type

$$P_{d,h} = \sum_{i=1}^{24} (\beta_i P_{d-1,i} + \beta_{24+i} P_{d-7,i}) + \beta_{49} P_{d-1}^{\text{min}} + \beta_{50} P_{d-1}^{\text{max}} + \sum_{i=1}^{24} \left(\beta_{50+i} \widehat{L}_{d,i} + \beta_{74+i} \widehat{L}_{d-1,i}\right)$$
$$+ \sum_{i=1}^{24} \left(\beta_{98+i} \widehat{S}_{d,i} + \beta_{122+i} \widehat{S}_{d-1,i}\right) + \sum_{i=1}^{24} \left(\beta_{146+i} \widehat{W}_{d,i} + \beta_{170+i} \widehat{W}_{d-1,i}\right)$$
$$+ \beta_{195} \text{Coal}_{d-2} + \beta_{196} \text{Gas}_{d-2} + \beta_{197} \text{Oil}_{d-2} + \beta_{198} \text{EUA}_{d-2} + \sum_{i=1}^{7} \beta_{198+i} D_i + \varepsilon_{d,h}$$

What if instead of using point forecasts we use the predictive distributions of load and RES generation?

Probabilistic inputs increase accuracy

(Uniejewski & Ziel, 2025, arXiv)



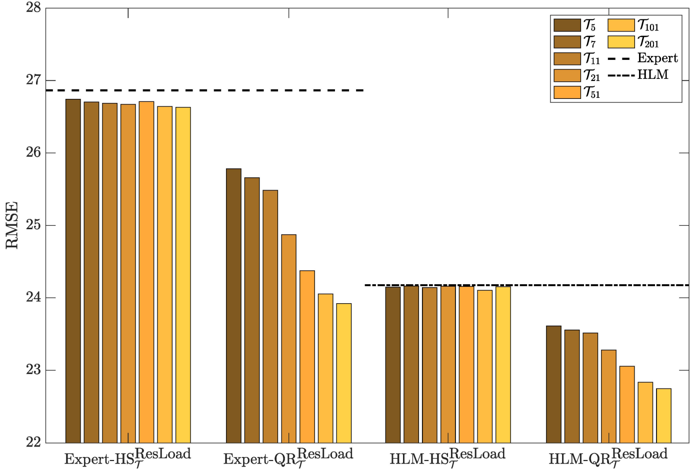
- RMSE for Germany (2018-23)
- Models with 201 quantiles ($\mathcal{T}_{201}$) approximating $\hat{F}_X$

$$X = \{L, S, W, RES = S + W, L - RES, \{L, RES\}, \{L, S, W\}\}$$

- Postprocessing schemes: HS and 'QRA' (QR on X)
- Note the 11.6% and 6.1% increases in accuracy

The finer the grid of quantiles the better the forecast

(Uniejewski & Ziel, 2025, arXiv)



- A finer grid of quantiles $\mathcal{T}_5 \longrightarrow \mathcal{T}_{201}$ generally improves the accuracy of $\hat{p}_{d,h}$
- Changes are marginal for HS
- But substantial for QRA

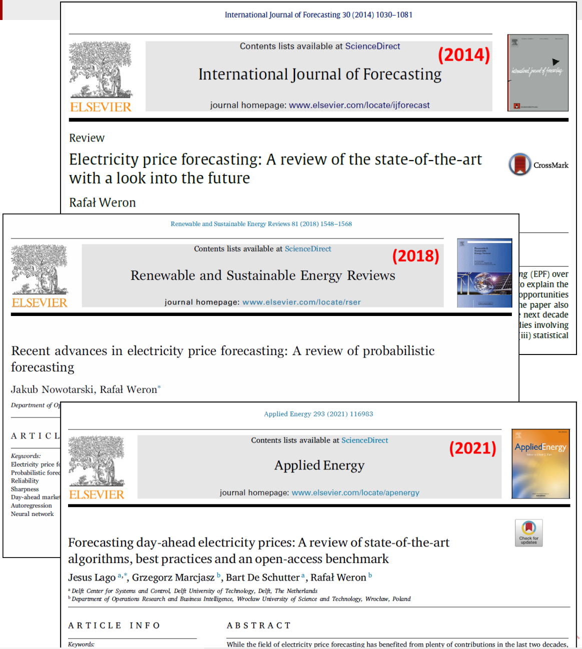
1 Introduction

2 Model evolution

3 Beyond point forecasts

4 Financial evaluation

- Day-ahead bidding with BESS
- Which loss function to minimize?



International Journal of Forecasting 30 (2014) 1030–1081

Contents lists available at ScienceDirect

(2014)

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

ELSEVIER

Review

Electricity price forecasting: A review of the state-of-the-art with a look into the future

Rafał Weron

ng (EPF) over
o explain the
pportunities
he paper also
next decade
dies involving
iii) statistical

Renewable and Sustainable Energy Reviews 81 (2018) 1548–1568

Contents lists available at ScienceDirect

(2018)

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

ELSEVIER

Recent advances in electricity price forecasting: A review of probabilistic forecasting

Jakub Nowotarski, Rafał Weron*

Department of Op

ARTICL

Keywords:
Electricity price f
Probabilistic fore
Reliability
Sharpness
Day-ahead marke
Autoregression
Neural network

Applied Energy 293 (2021) 116983

Contents lists available at ScienceDirect

(2021)

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

ELSEVIER

Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark

Jesus Lago [a,*], Grzegorz Marcjasz [b], Bart De Schutter [a], Rafał Weron [b]

[a] Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands
[b] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Wrocław, Poland

ARTICLE INFO

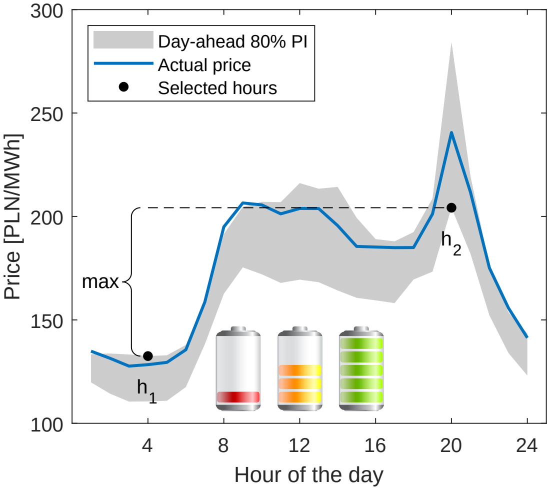
Keywords:

ABSTRACT

While the field of electricity price forecasting has benefited from plenty of contributions in the last two decades,

Quantile-based trading strategies

(Uniejewski & Weron, 2021, ENEECO)



Simultaneously bid day-ahead:

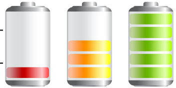
- buy 1 MW for $\hat{P}^{1-\alpha}_{d,h_1}$ at hour h_1
- sell 0.8 MW at $\hat{P}^{\alpha}_{d,h_2}$ at hour h_2

$\leftarrow$ With $\alpha = 10\%$ the spread is maximized for $\hat{P}^{90\%}_{d,4} = 133$ and $\hat{P}^{10\%}_{d,20} = 204$ PLN/MWh

- If bid(s) not accepted, settle in the balancing market

Are quantile-based trading strategies more profitable?

(Uniejewski & Weron, 2021, ENEECO)



Strategy			**Profit**		
Naive (4am–12pm)			33 065.29		
Point forecasts-based			33 722.39		
Quantile-based	1-99%	5-95%	10-90%	20-80%	25-75%
Q-Ave	41 317.92	43 328.89	43 432.31	43 289.09	43 033.88
F-Ave	39 848.26	43 369.44	44 052.04	44 088.11	43 130.34
QRM	41 163.29	43 054.28	43 124.12	43 731.54	42 240.25
LQRA(77)	42 360.05	44 135.49	44 713.52	44 684.40	43 624.65
LQRA(BIC)	42 886.10	43 993.23	44 502.81	45 396.21	42 741.57
LQRA(CV)	41 693.80	43 971.88	44 238.45	45 073.19	43 103.88

Quantile-based trading yields 19-34% higher profits than point forecasts-based!

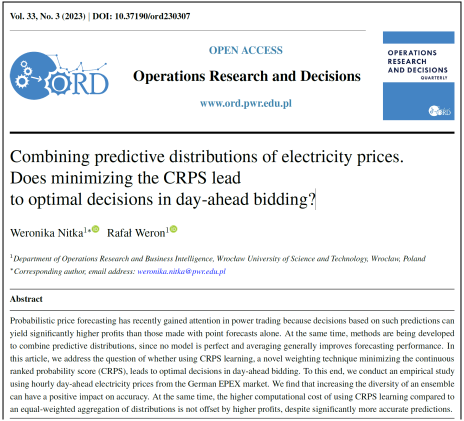
Does CRPS learning lead to higher trading profits?

(Nitka & Weron, 2023, ORD)

% change in the CRPS and profits per trade when using CRPS learning (instead of naive averaging)

	CRPS	Profits/trade (EUR/MWh) 50% PI	70% PI	90% PI
Ensemble #1	0.0%	-1.1%	-0.8%	-3.6%
Ensemble #2	-0.1%	-0.6%	-0.8%	-2.5%
Ensemble #3	-0.7%	-1.7%	-1.5%	-6.3%
Ensemble #4	-0.4%	-2.1%	-1.4%	-2.8%
Ensemble #5	0.2%	-0.5%	-0.1%	-5.0%
Ensemble #6	-0.8%	-0.4%	-1.0%	-4.3%

- CRPS learning yields **lower** CRPS → **good**
- CRPS learning yields **lower** profits → **bad**


Vol. 33, No. 3 (2023) | DOI: 10.37190/ord230307

OPEN ACCESS

Operations Research and Decisions

www.ord.pwr.edu.pl

Combining predictive distributions of electricity prices. Does minimizing the CRPS lead to optimal decisions in day-ahead bidding?

Weronika Nitka[1*] Rafał Weron[1]

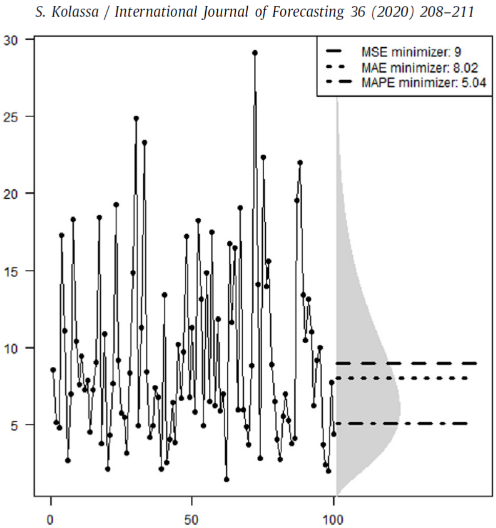
[1]Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Wrocław, Poland

*Corresponding author, email address: weronika.nitka@pwr.edu.pl

Abstract

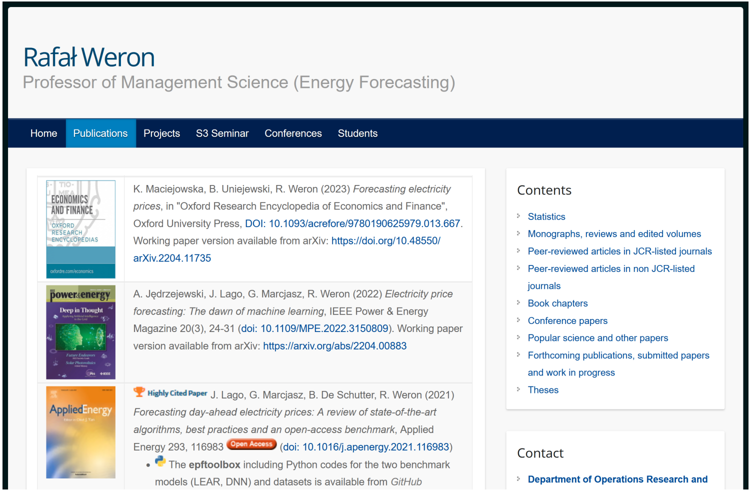
Probabilistic price forecasting has recently gained attention in power trading because decisions based on such predictions can yield significantly higher profits than those made with point forecasts alone. At the same time, methods are being developed to combine predictive distributions, since no model is perfect and averaging generally improves forecasting performance. In this article, we address the question of whether using CRPS learning, a novel weighting technique minimizing the continuous ranked probability score (CRPS), leads to optimal decisions in day-ahead bidding. To this end, we conduct an empirical study using hourly day-ahead electricity prices from the German EPEX market. We find that increasing the diversity of an ensemble can have a positive impact on accuracy. At the same time, the higher computational cost of using CRPS learning compared to an equal-weighted aggregation of distributions is not offset by higher profits, despite significantly more accurate predictions.

What is the 'best' (point) forecast?



- Gneiting (2011, IJF; 2022, JASA) and Kolassa (2020, IJF) argue that there is not a single 'best point forecast'
- The different error measures evaluate 'from different angles':
  - The mean minimizes ε_t^2, hence MSE, RMSE, etc.
  - The median minimizes $|\varepsilon_t|$, hence MAE, MASE, rMAE, etc.
  - The τ-quantile minimizes Pinball Score, hence the CRPS
- But what does maximize the profits?

Articles & working papers on https://p.wz.pwr.edu.pl/∼weron.rafal/Publ



Rafał Weron

Professor of Management Science (Energy Forecasting)

Home | Publications | Projects | S3 Seminar | Conferences | Students

K. Maciejowska, B. Uniejewski, R. Weron (2023) *Forecasting electricity prices*, in "Oxford Research Encyclopedia of Economics and Finance", Oxford University Press, DOI: 10.1093/acrefore/9780190625979.013.667. Working paper version available from arXiv: https://doi.org/10.48550/arXiv.2204.11735

A. Jędrzejewski, J. Lago, G. Marcjasz, R. Weron (2022) *Electricity price forecasting: The dawn of machine learning*, IEEE Power & Energy Magazine 20(3), 24-31 (doi: 10.1109/MPE.2022.3150809). Working paper version available from arXiv: https://arxiv.org/abs/2204.00883

Highly Cited Paper J. Lago, G. Marcjasz, B. De Schutter, R. Weron (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 Open Access (doi: 10.1016/j.apenergy.2021.116983)

- The **epftoolbox** including Python codes for the two benchmark models (LEAR, DNN) and datasets is available from *GitHub*

Contents

Contact

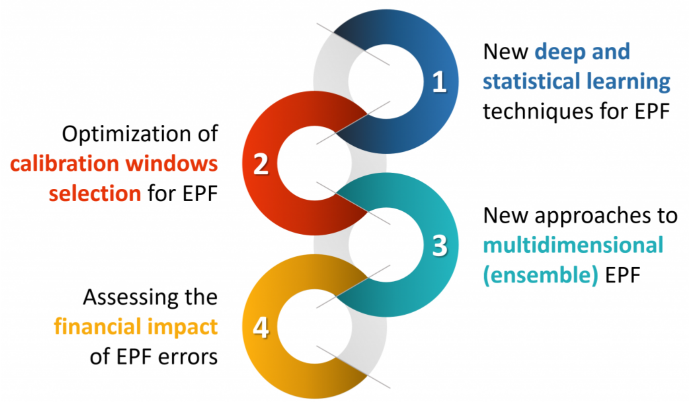
- **Department of Operations Research and**

See yourself:

CrossFIT (2019-2025)

Crossing Frontiers in electricity prIce forecasTing

(https://kbo.pwr.edu.pl/en/research/research-projects/ncn-2018-30-a-hs4-00444)



PRIORITY (2023-2026)

PRobabilistic mid- and long-term prIce fORecasting In electriciTY markets
(https://kbo.pwr.edu.pl/en/research/research-projects/dfg-ncn-2021-43-i-hs4-02578-2)