Distinguished Lecture Series 2024/2025

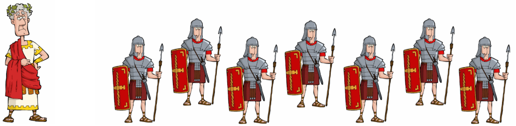
Electricity Price Forecasting II

Rafał Weron*

(with Julia/Python notebooks coded and explained by Arkadiusz Lipiecki)

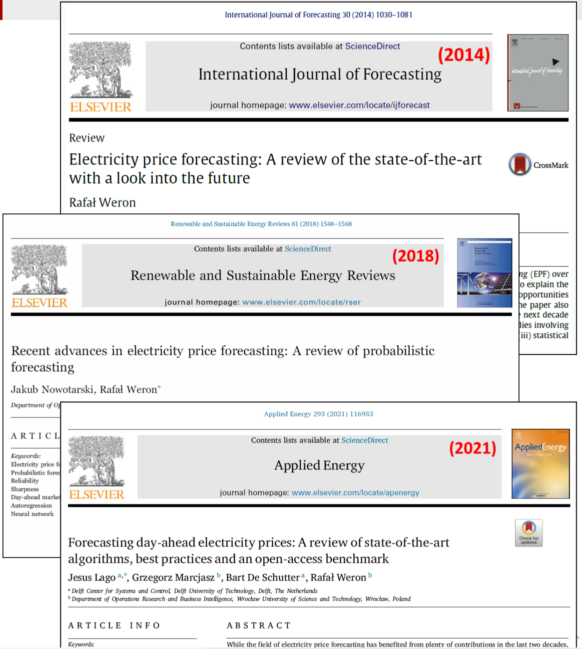
Department of Operations Research and Business Intelligence, Wrocław Tech, Poland

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)

1. Tips and tricks
   - Transformations
   - Seasonal decomposition
   - Combining forecasts
   - Averaging across windows
   - Calibration window selection
2. Lasso and DNN
3. Probabilistic forecasts revisited
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

ABSTRACT

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

Transforming variables

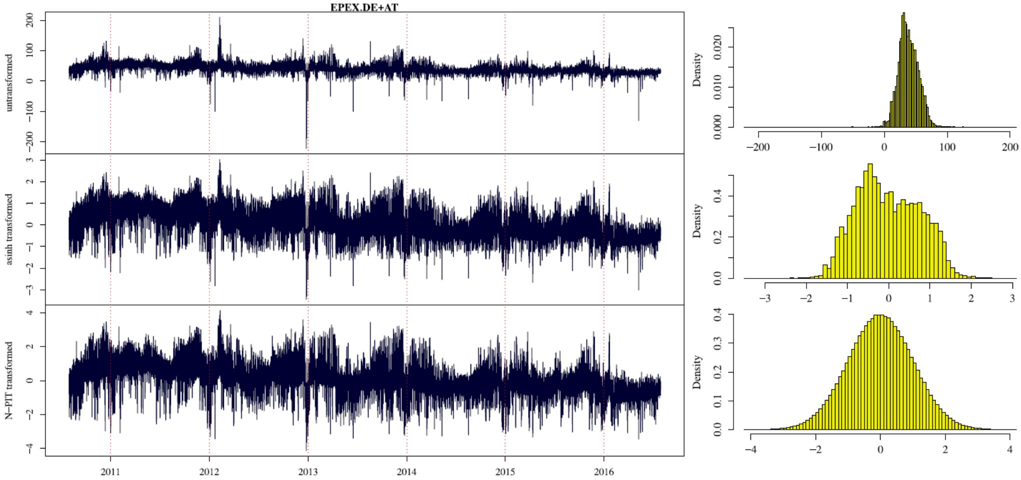
- We can model a nonlinear relationship by transforming y_t and/or x_t's before estimating the model, e.g., using the logarithm:

$$\log y_t = \beta_0 + \beta_1 x_t + \varepsilon$$
$$y_t = \beta_0 + \beta_1 \log x_t + \varepsilon$$
$$\log y_t = \beta_0 + \beta_1 \log x_t + \varepsilon$$

- In the latter model, the slope can be interpreted as elasticity: $\beta_1 =$ average % change in y_t resulting from a 1% increase in x_t
- All three models are still linear in the parameters
- If x_t can be zero or negative, use $\log(x_t + const.)$ or a different transformation

Dealing with asymmetry & heavy tails: ID, asinh and N-PIT

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



A zoo of variance stabilizing transformations (VSTs)

(Uniejewski, Weron & Ziel, 2018, IEEE-TPWRS; Narajewski & Ziel, 2020, J Comm)

- Area hyperbolic sine transformation:

$$y = \operatorname{asinh}(x) \equiv \log\left(x + \sqrt{x^2+1}\right)$$

with inverse $x = \sinh(y)$

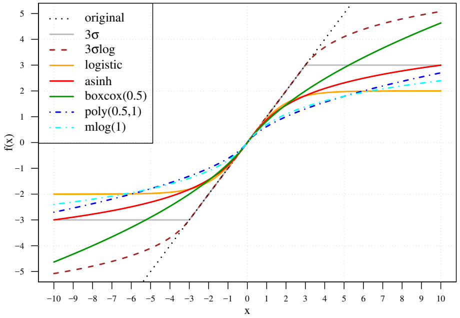
- Mirror-log transform with parameter c $(= \frac{1}{3})$:

$$\operatorname{mlog}(x;c) = \operatorname{sign}(x)\left[\log\left(|x| + \frac{1}{c}\right) + \log(c)\right]$$

with inverse $x = \operatorname{sign}(y)\left[e^{|y|-\log c} - \frac{1}{c}\right]$

- Narajewski & Ziel (2020) suggest an alternative approach to computing back-transformations

asinh & mlog have a log-damping effect and can handle negative values

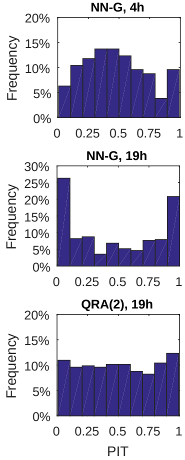


Probability Integral Transform (PIT)
(Dawid, 1984, JRSSA; Nowotarski & Weron, 2018, RSER)

- The Probability Integral Transform is defined as:

$$PIT_t = \hat{F}(x)$$

- If x is $\hat{F}$-distributed then $PIT_t \sim U(0,1)$, see
  https://en.wikipedia.org/wiki/Probability_integral_transform
- A $\cap$-shape histogram $\Rightarrow$ $\hat{F}$ has too fat tails
- A $\cup$-shape histogram $\Rightarrow$ $\hat{F}$ has too light tails
- A flat histogram $\Rightarrow$ $\hat{F}$ is a good estimate of F



A zoo of variance stabilizing transformations cont.

(Diaz & Planas, 2016, IEEE-TPWRS; Uniejewski, Weron & Ziel, 2018, IEEE-TPWRS)

- Define the N-PIT transformation by:

$$y = \Phi^{-1}(PIT_t) = \Phi^{-1}\left(\hat{F}(x)\right)$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal CDF

- Called Nataf by Diaz & Planas (2016)
- If PIT $\sim U(0,1)$ then $y \sim N(0,1)$

- And the t-PIT transformation by:

$$y = G^{-1}(PIT_t) = G^{-1}\left(\hat{F}(x)\right)$$

where $G^{-1}(\cdot)$ is the inverse of the Student-t CDF with $\nu = 8$ degrees of freedom

- If PIT $\sim U(0,1)$ then $y \sim$ Student-t

Note: Transformations can be applied to the prices as well as to exogenous variables

Applying VSTs

(Uniejewski, Weron & Ziel, 2018, IEEE-TPWRS; Narajewski & Ziel, 2020, J Comm)

- All VSTs except PIT-type require 'normalized' prices:

$$P_t \xrightarrow{\text{normalize}} Y_t \xrightarrow{\text{VST}} y_t \xrightarrow{\text{Predict}} \hat{y}_{t+1} \xrightarrow{\text{VST}^{-1}} \hat{Y}_{t+1} \xrightarrow{\text{de-normalize}} \hat{P}_{t+1}$$

- For $(\text{median}, \text{MAD})$ scaling, set for $t \in \mathcal{S}$:
  - $a = \text{median}(P_t)$ and $b = \text{MAD}(P_t)$, i.e., the sample median absolute deviation around a multiplied by $z_{0.75}^{-1} \approx 1.4826$ where $z_{0.75}$ is the 75% quantile of $\text{N}(0,1)$
  - Julia and R use the 1.4826 multiplier by default, but Matlab and Python do not
- Alternatively, use $(\text{mean}, \text{std})$ scaling (z-score, standardization)
- To normalize compute $Y_t = \frac{1}{b}(P_t - a)$, to invert compute $\hat{P}_{t+1} = b\hat{Y}_{t+1} + a$

Dummy (indicator) variables

- Use when a predictor is a categorical variable taking only 2 values
  - Public holiday, World Cup match, nation-wide strike
- If there are more than 2 categories, use several dummies
  - Always one fewer than the number of categories or don't use the intercept
  - Sunday is captured by the intercept (if present) when $d_{1,t} = \ldots = d_{6,t} = 0$

	$d_{1,t}$	$d_{2,t}$	$d_{3,t}$	$d_{4,t}$	$d_{5,t}$	$d_{6,t}$	$d_{7,t}$
Monday	**1**	0	0	0	0	0	0
Tuesday	0	**1**	0	0	0	0	0
Wednesday	0	0	**1**	0	0	0	0
Thursday	0	0	0	**1**	0	0	0
Friday	0	0	0	0	**1**	0	0
Saturday	0	0	0	0	0	**1**	0
Sunday	0	0	0	0	0	0	**?**

Fourier terms (sinusoidal waves)

- An alternative to dummies, esp. for long seasonal (i.e., regular) periods
- If m is the seasonal period, the Fourier terms for $N = 0, 1, 2, \ldots$ are given by:

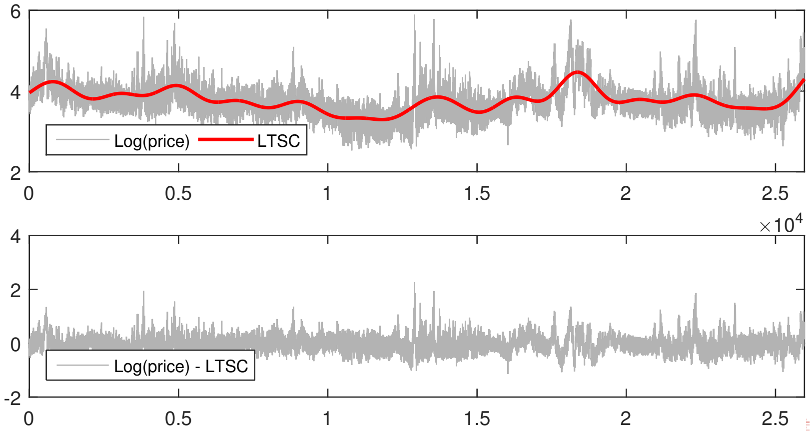
$$x_{2N+1,t} = \sin\left(\frac{2\pi t(N+1)}{m}\right) \quad \text{and} \quad x_{2N+2,t} = \cos\left(\frac{2\pi t(N+1)}{m}\right)$$

- For weekly seasonality, the first 6 of these $\equiv$ 6 dummies
  - Use Fourier terms when m is large, e.g., $m = 365$ days
- Usually fewer Fourier terms are needed than dummy variables
- A model with Fourier terms is often called harmonic regression
  - Successive Fourier terms represent harmonics of $x_{1,t}$ and $x_{2,t}$
  - See `https://en.wikipedia.org/wiki/Harmonic`

What about cyclic behavior and seasonal decomposition?

(Nowotarski & Weron, 2016, ENEECO; Marcjasz et al., 2019, IJF; Chęć et al., 2025, JCM)

Decompose → predict the Long-Term Seasonal Component & residuals → combine



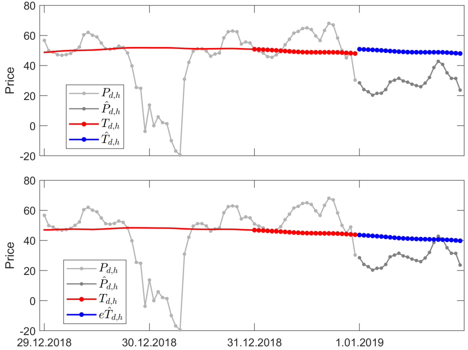
Extrapolating the LTSC to eliminate the ‘jump’

Chęć et al. (2025, JCM) evaluate ARX and LEAR models in the following setup:

$$P_{d,h} \xrightarrow{-\mathrm{LTSC}} Y_{d,h} \xrightarrow{\mathrm{VST}} y_{d,h} \xrightarrow{\mathrm{Predict}} \hat{y}_{d,h} \xrightarrow{\mathrm{VST}^{-1}} \hat{Y}_{d,h} \xrightarrow{+\widehat{\mathrm{LTSC}}} \hat{P}_{d,h}$$

where VST and VST^{-1} include normalization

1. Use ARX or LEAR to compute $\hat{P}_{d+1,h}$ for all h
2. Append these forecasts to $P_{d,h}$ in $\mathcal{S} \rightarrow \mathcal{S}'$
3. Decompose the appended price series
4. Use $\mathrm{LTSC}_{d+1,h}$ as the LTSC forecast
5. Use residuals for $d, h \in \mathcal{S}$ to predict $\hat{y}_{d+1,h}$
6. Combine to yield the final $\hat{P}_{d+1,h}$



Extrapolating the LTSC cont.

(Chęć, Uniejewski & Weron 2025, JCM)

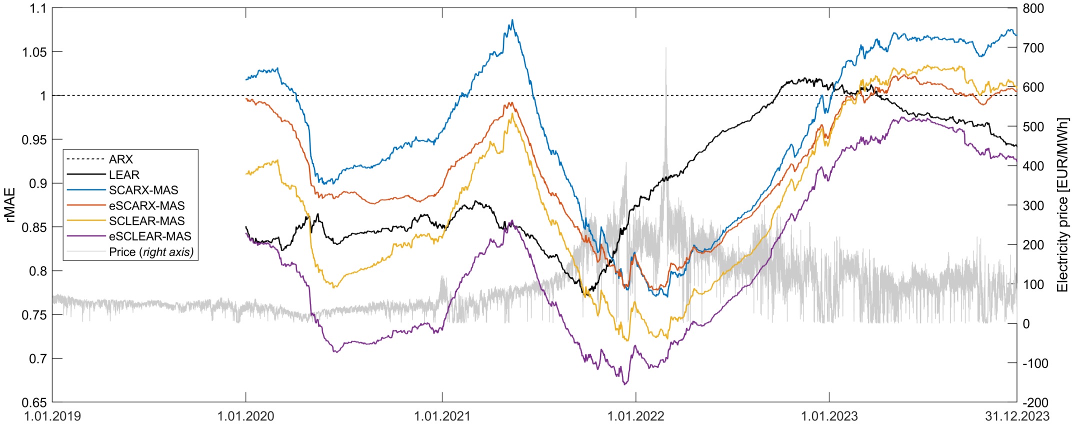
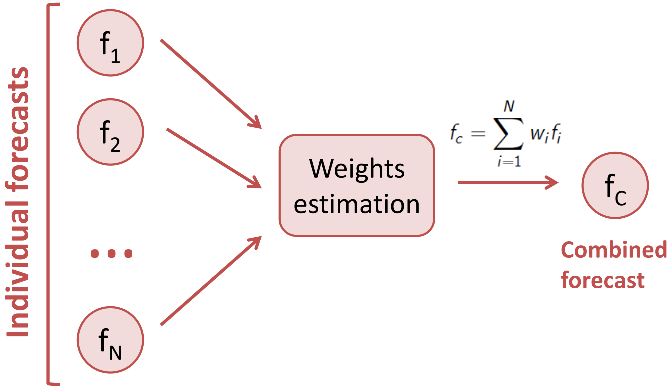


Fig. 6. Rolling 365-day relative mean absolute errors (rMAE) for the Spanish (OMIE) market with respect to MAE of the ARX model; prices in gray

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*

A zoo of (point) forecast combination schemes

(Nowotarski et al., 2014, ENEECO; Weiss et al., 2018, R Journal; Uniejewski & Maciejowska, 2023, IJF)

- Simple combination methods:
  - Simple Average, Trimmed/Winsorized Mean
  - Bates/Granger ($\sim$ IRMSE), Inverse Rank
- Regression-based combinations:
  - OLS regression, Constrained LS regression
  - Least Absolute Deviation (LAD) regression
  - Complete subset regression
- Eigenvector-based combinations ...
- Aggregated Forecast Through Exponential Re-weighting (AFTER)
- Principal Component Analysis (PCA)-based
- LASSO-based, LASSO-PCA

Energy Economics 46 (2014) 395–412
Contents lists available at ScienceDirect
Energy Economics
journal homepage: www.elsevier.com/locate/eneco

An empirical comparison of alternative schemes for combining electricity spot price forecasts

Jakub Nowotarski [a], Eran Raviv [b], Stefan Trück [c], Rafał Weron [a,*]
[a] Institute of Organization and Management, Wrocław University of Technology, Wrocław, Poland
[b] Department of Econometrics, Erasmus University, Rotterdam, The Netherlands
[c] Faculty of Business and Economics, Macquarie University, Sydney, Australia

CONTRIBUTED RESEARCH ARTICLES The R Journal Vol. 10/2, December 2018

Forecast Combinations in R using the ForecastComb Package

by Christoph E. Weiss and Eran Raviv and Gernot Roetzer

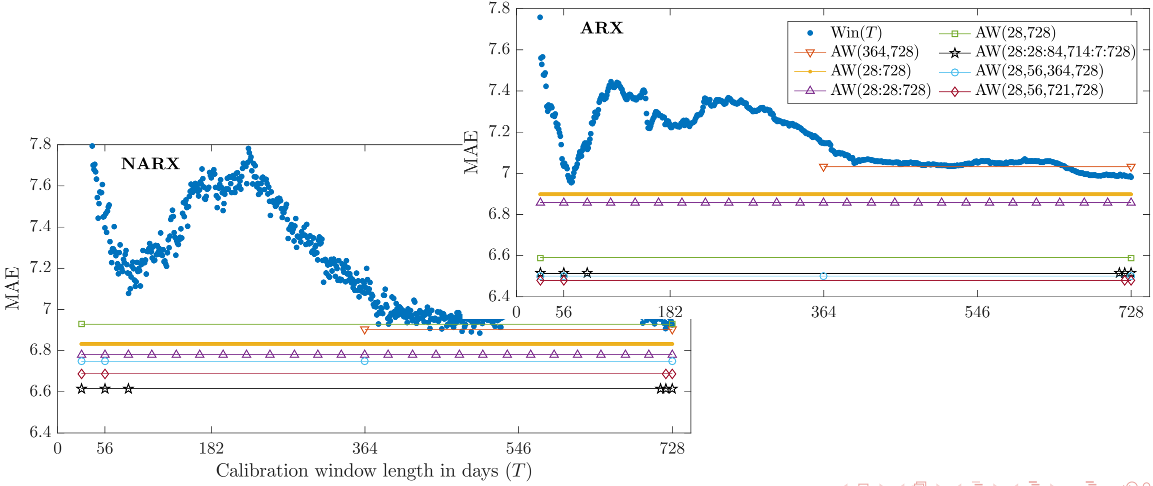
International Journal of Forecasting 39 (2023) 1839–1852
Contents lists available at ScienceDirect
International Journal of Forecasting
journal homepage: www.elsevier.com/locate/ijforecast

LASSO principal component averaging: A fully automated approach for point forecast pooling

Bartosz Uniejewski [*], Katarzyna Maciejowska
Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, 50-370 Wrocław, Poland

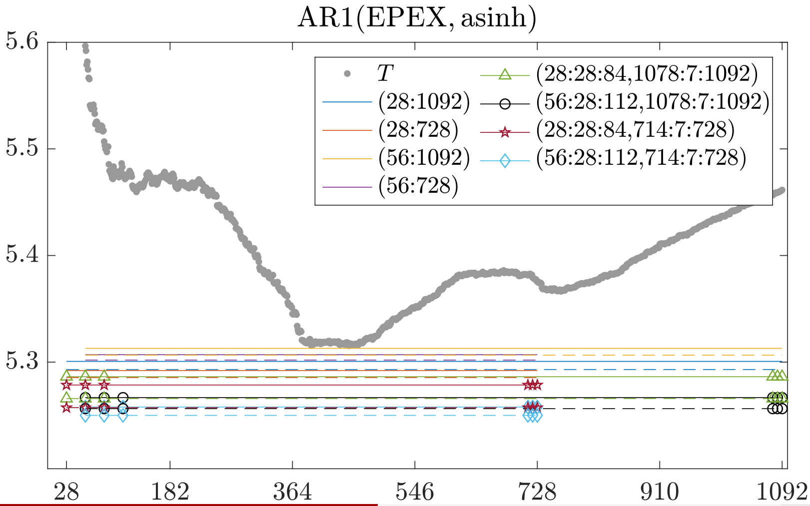
Simple averaging across calibration windows

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



Simple (AW) vs. past performance-weighted (WAW) averaging

(Marcjasz, Serafin & Weron, 2018, Energies)

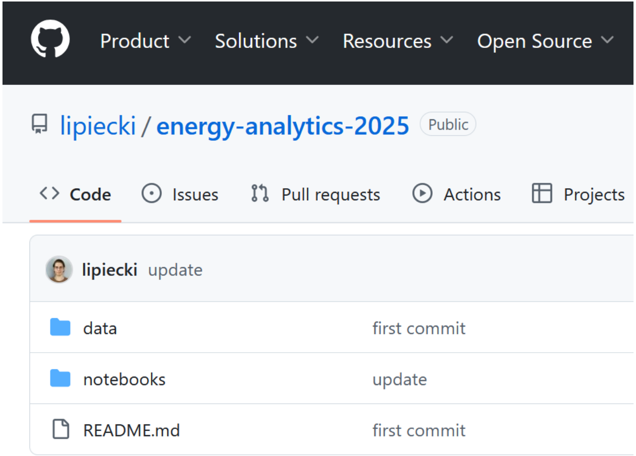


- AR-type model
- *asinh* transformation
- Fitted to German day-ahead prices
- Test period: Aug 2013 – Jul 2016

— **AW**

- - **WAW**

Python snippet: Averaging.ipynb

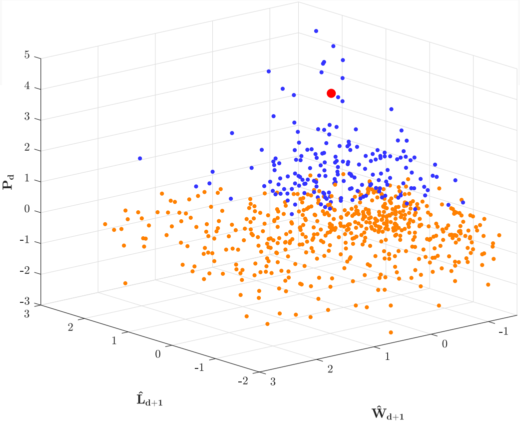


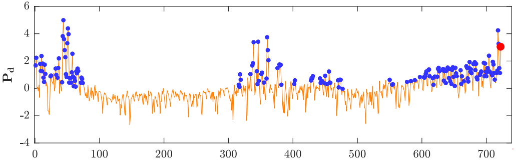
Selecting training subsets: k-NN

(Nitka, Serafin & Sotiros, 2021, ICCS)

- Idea: Use only relevant past information
- Select data for training based on similarity, not the most recent period
- Measure the 'closeness' of today → ●, i.e, $\{P_{d,h}, \hat{L}_{d+1,h}, \hat{W}_{d+1,h}\}$, to the days $\in \mathcal{S}$ with k-Nearest Neighbors (k-NN)
  - $P_{d,h}$ – today's price for hour h
  - $\hat{L}_{d+1,h}$, $\hat{W}_{d+1,h}$ – today's load & wind generation forecasts for tomorrow

Selected for training → ●, discarded → ●

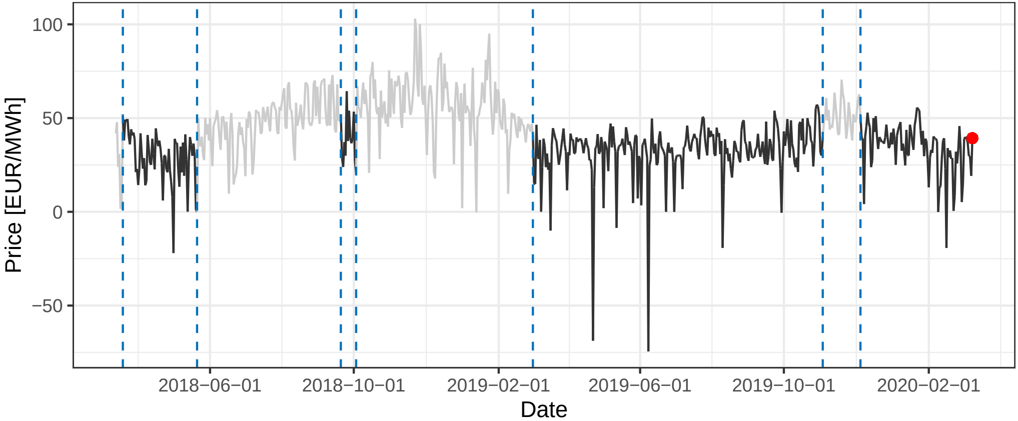




Selecting training subsets: Identifying change/breakpoints

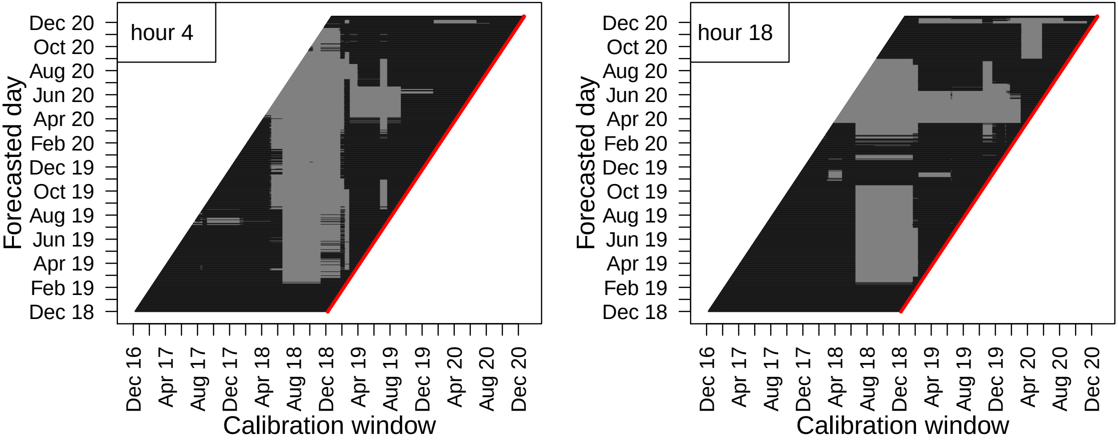
(de Marcos et al., 2020, Energies; Nasiadka et al., 2022, ICCS)

Breakpoints (blue) split $\mathcal{S}$ into similar (**black**) and remaining (= left out; gray) subperiods



EPEX-DE: Narrowest-Over-Threshold (NOT)

NOT: Baranowski et al. (2019, JRSSB); EPF: Nasiadka et al. (2022, ICCS)

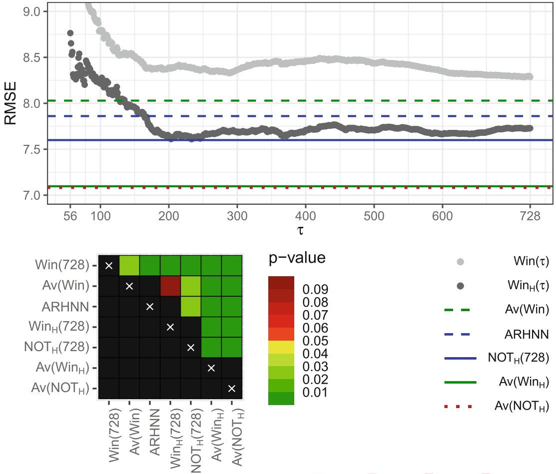


EPEX-DE: RMSE errors and the DM test

(Nasiadka, Nitka & Weron, 2022, ICCS)

ARX (lags 1, 2, 7, min, max) with $\hat{L}$, $\hat{W}$ and $\hat{S}$:

- **Win(τ)** – window of $\tau \in [56, 57, ..., 728]$ days
- **Av(Win)** – Av($\tau = 56, 84, 112, 714, 721, 728$)
- The **ARHNN** model
- **Win$_H$(τ)** = Win(τ) with *asinh*
- **NOT$_H$(728)** – NOT($\tau = 728$) with *asinh*
- **Av(Win$_H$)** = Av(Win) with *asinh*
- **Av(NOT$_H$)** = Av(Win$_H$) with NOT$_H$(728) instead of Win($\tau = 714, 721, 728$)



1. Tips and tricks

2. Lasso and DNN
   - Stepwise regression
   - Shrinkage (regularization)
   - LASSO-Estimated AR (LEAR)
   - Deeper and deeper
   - Interpretable AI

3. Probabilistic forecasts revisited

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

ng (EPF) over
o explain the
pportunities
he paper also
next decade
lies 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

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

Jakub Nowotarski, Rafał Weron*

Department of O

ARTICL

Keywords:
Electricity price f
Probabilistic forec
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

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

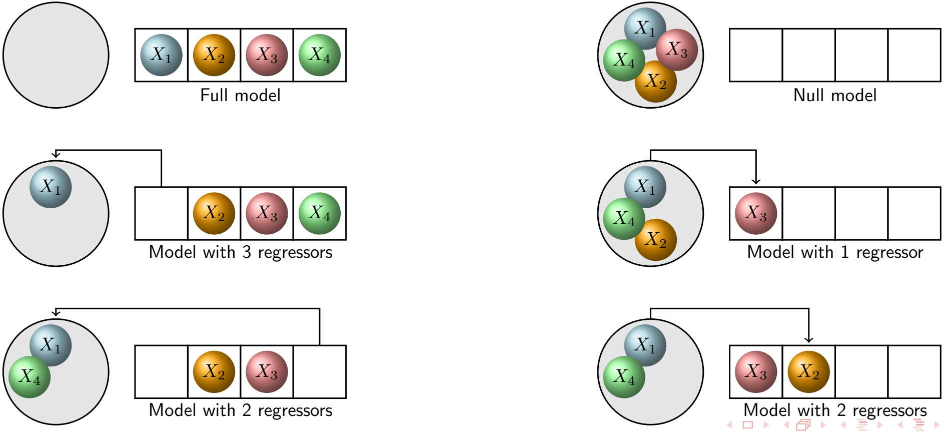
Keywords:

ABSTRACT

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

Stepwise regression: Backward elimination & forward selection

If the number of variables (features, inputs, regressors, predictors) is large, it is not possible to try all possible combinations → variable selection



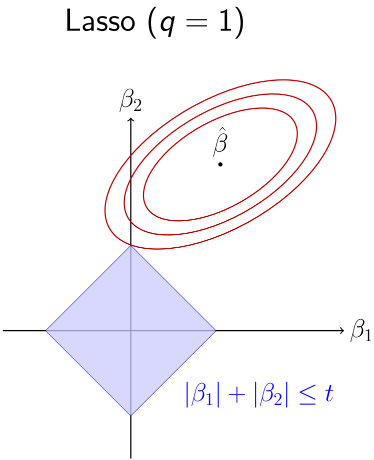
What is shrinkage (regularization)?

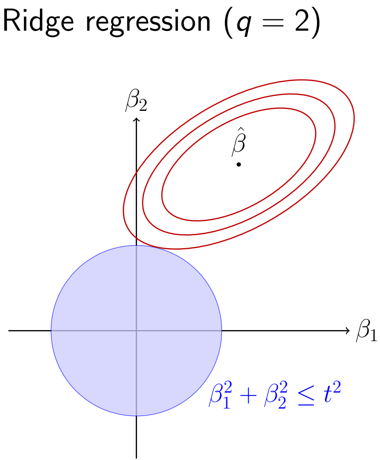
- Minimize the residual sum of squares (RSS $\equiv$ sum of squared errors) $+$ a penalty function of the betas:

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

  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?





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

Ridge regression vs. LASSO

- Ridge regression: RSS penalized by a quadratic shrinkage factor

$$\hat{\beta}^{ridge} = \underset{\beta_j}{\operatorname{argmin}} \left\{ RSS + \lambda \sum_{j=1}^{n} \beta_j^2 \right\}$$

  - For all j, if $\lambda \to \infty$ then $\beta_j \to 0$
  - All β_j's remain non-zero $\to$ all predictors stay in the final model

- LASSO: RSS penalized by a linear shrinkage factor

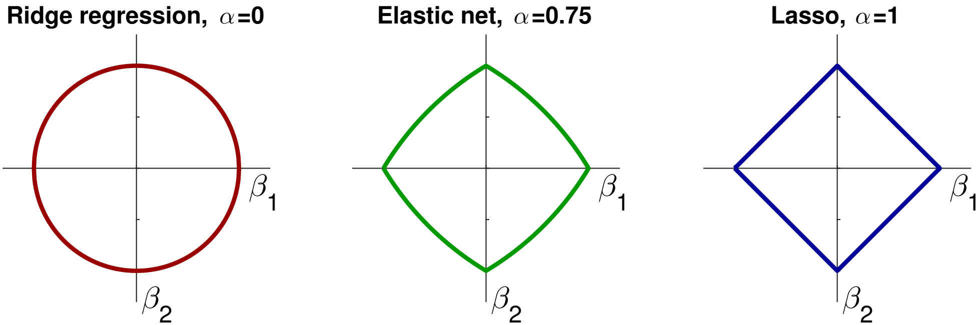
$$\hat{\beta}^{lasso} = \underset{\beta_j}{\operatorname{argmin}} \left\{ RSS + \lambda \sum_{j=1}^{n} |\beta_j| \right\}$$

  - For all j, if $\lambda \geqslant 1$ then $\beta_j = 0$
  - As $\lambda \to 1$, more β_j's become 0 $\to$ the final model does not include all predictors

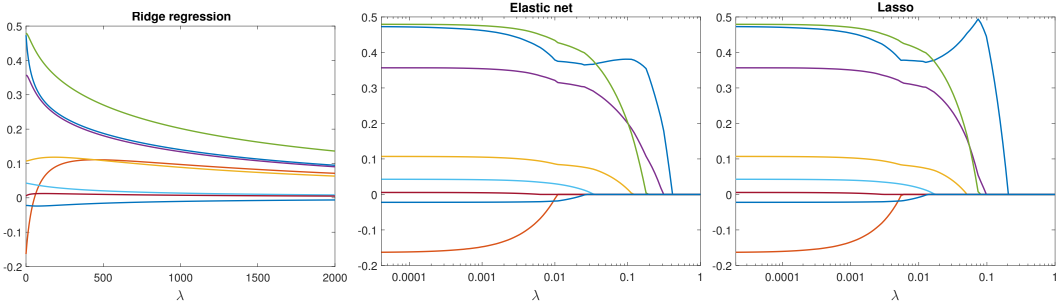
Elastic net (Zou & Hastie, 2015, JRSSB)

- RSS penalized by a mixed quadratic and linear shrinkage factor

$$\hat{\beta}^{EN} = \underset{\beta_j}{\text{argmin}} \left\{ \text{RSS} + \lambda \left(\tfrac{1-\alpha}{2} \sum_{j=1}^{n} \beta_j^2 + \alpha \sum_{j=1}^{n} |\beta_j| \right) \right\}$$



How $\hat{\beta}$'s change when λ increases?



- *Left*: Ridge regression with $\lambda \in (0, 2000)$, linear scale
- *Center*: Elastic net with $\alpha = 0.5$ and $\lambda \in (0, 1)$, log-scale
- *Right*: Lasso with $\lambda \in (0, 1)$, log-scale

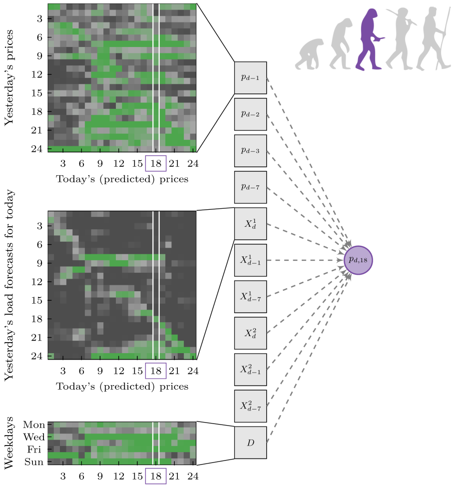
What about performance?

(Uniejewski, Nowotarski & Weron, 2016, Energies)

	ARX-type			AR-type	
	GEFCom	Nord Pool		GEFCom	N2EX (UK)
Naive	14.708	11.141	**Naive**	14.708	9.767
Expert benchmark (AR lags 1,2,7; Mon, Sat, Sun dummies, daily min, load forecast)					
ARX1	11.069	9.739	**AR1**	11.183	8.384
Full ARX model (107 variables)					
fARX	10.911	10.131	**fAR**	12.279	9.724
Selection & shrinkage methods (base model: fARX with 107 variables)					
forwardX	9.876	8.130	**forward**	11.193	8.563
backwardX	10.449	9.421	**backward**	11.968	9.252
RidgeX	9.777	8.972	**Ridge**	10.775	8.237
LassoX	9.476	8.419	**Lasso**	10.722	8.125
EN75X	9.475	**8.056**	**EN75**	10.708	8.124
EN50X	**9.473**	8.287	**EN50**	10.688	8.121
EN25X	9.474	8.529	**EN25**	**10.650**	8.113

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 $h = 18$

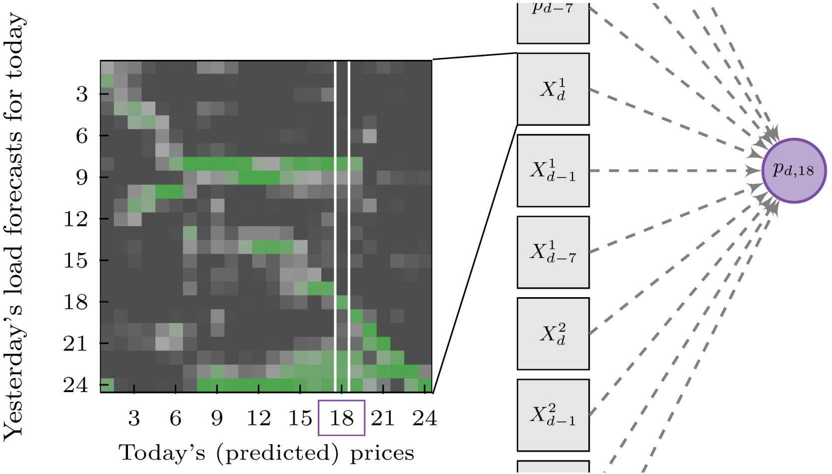
Uniejewski (2024, ORD) compares 10 penalties and finds that only Elastic Net and LQ with $1 < q < 2$:

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

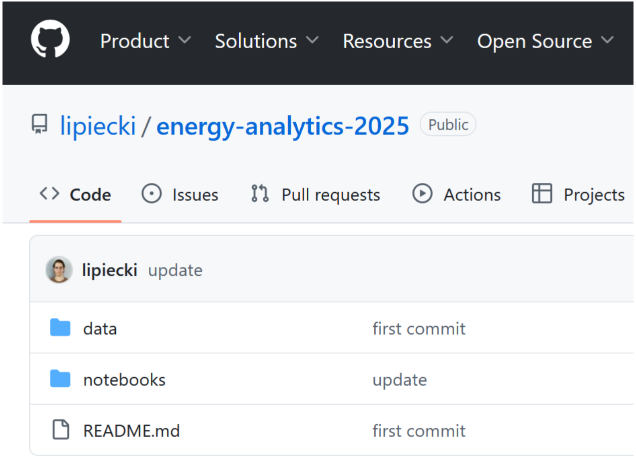
consistently outperform LASSO, but at the cost of one more hyperparameter (α or q)

LASSO-Estimated AR (LEAR): Variable importance

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

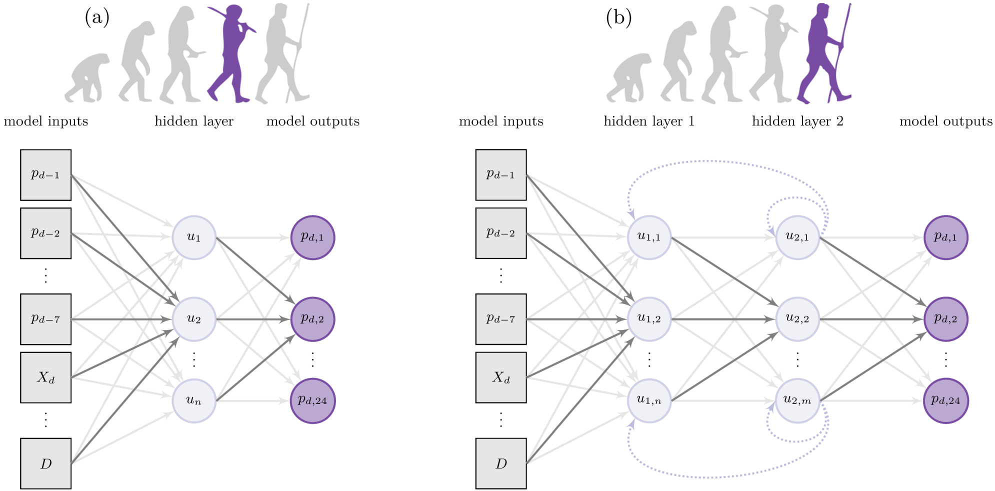


Python snippet: Shrinkage.ipynb



Multi-output shallow and deep neural networks (DNNs)

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



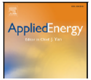
What about performance?

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]

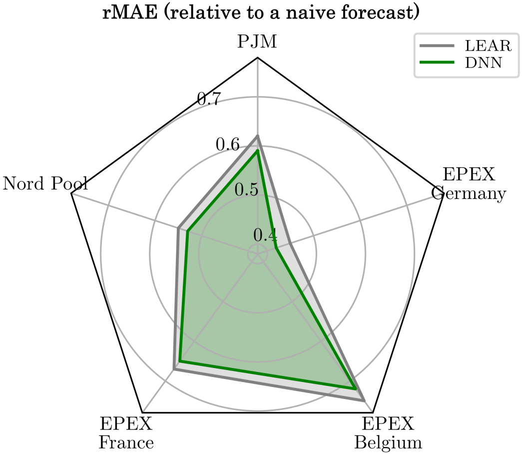
[a] Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands
[b] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Wrocław, Poland

ARTICLE INFO

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

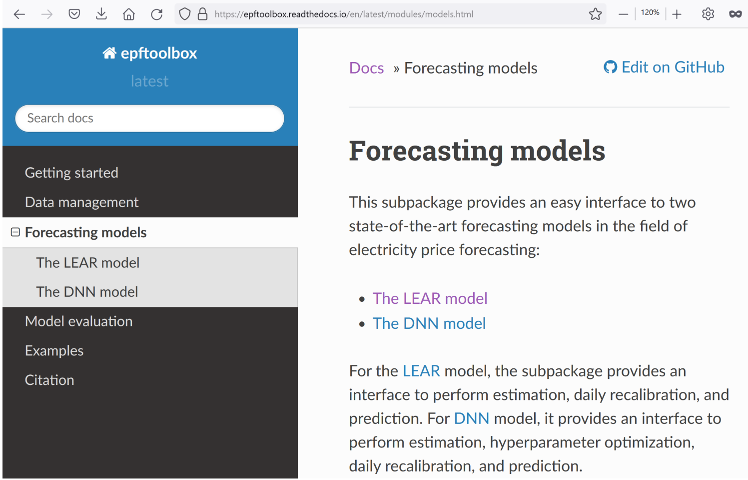
ABSTRACT

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.



epftoolbox: LEAR and DNN Python codes

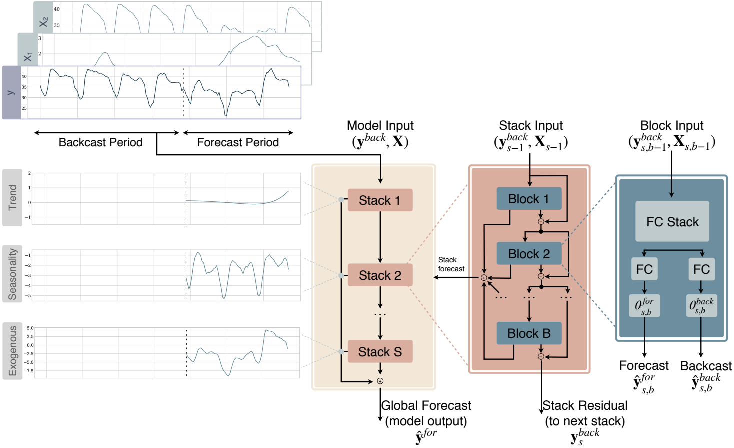
(Lago, Marcjasz, De Schutter & Weron, 2021, APEN)



Experiment yourself:

Interpretable AI: NBEATSx

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



Experiment yourself:

Interpretable AI: 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)

1 Tips and tricks

2 Lasso and DNN

3 Probabilistic forecasts revisited
- Quantile Regression Averaging
- Isotonic Distributional Regression
- PostForecasts.jl
- Combining probabilistic forecasts
- Distributional Deep Neural Nets
- Probabilistic inputs

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
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
lies 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 market
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

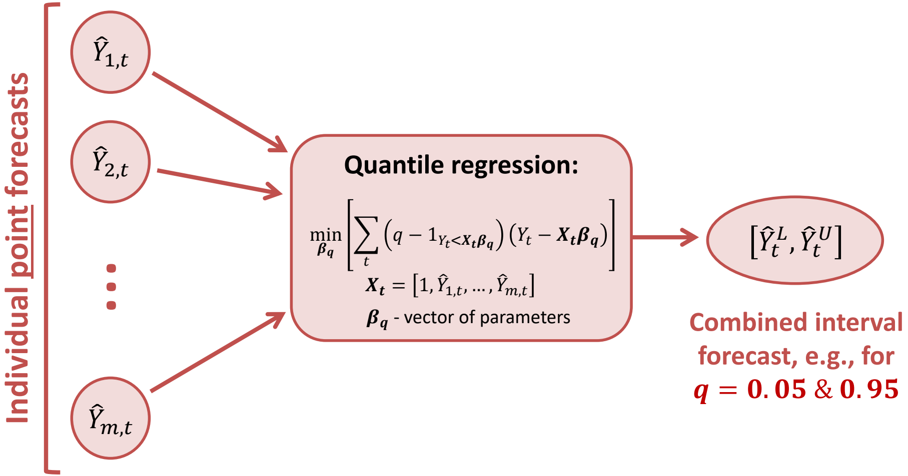
Keywords:

ABSTRACT

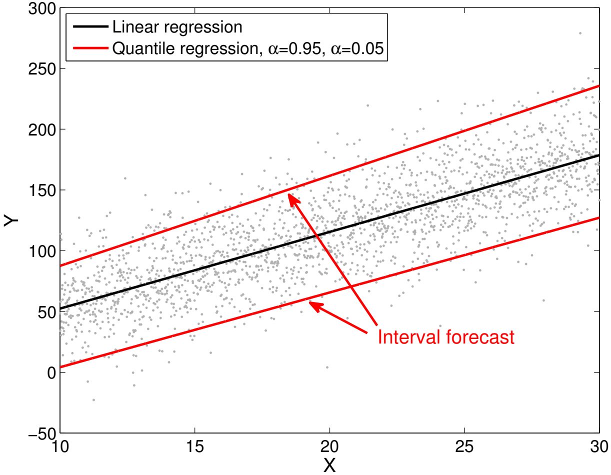
While the field of electricity price forecasting has benefited from plenty of contributions in the last two decades,

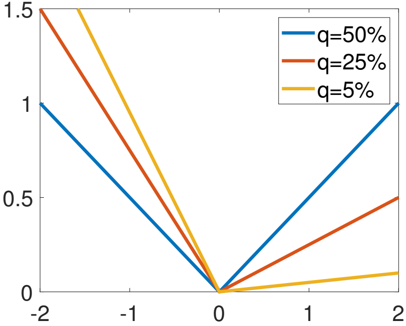
Quantile Regression Averaging (QRA): The idea

(Nowotarski & Weron, 2015, COST)



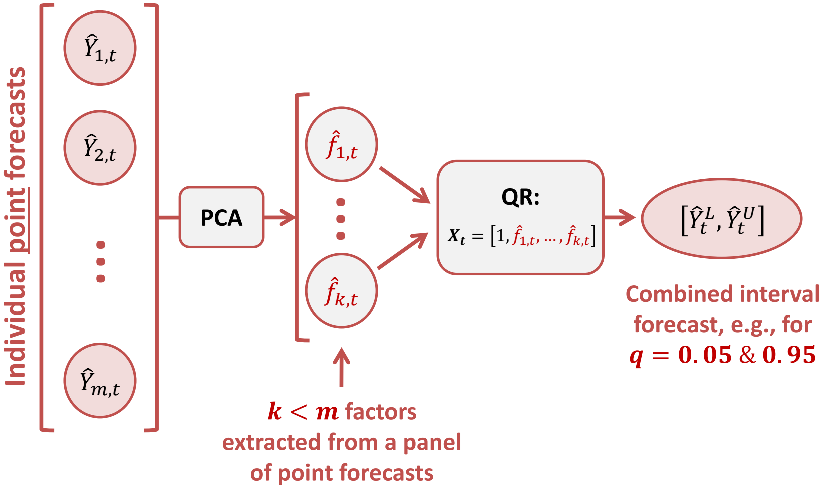
Quantile regression





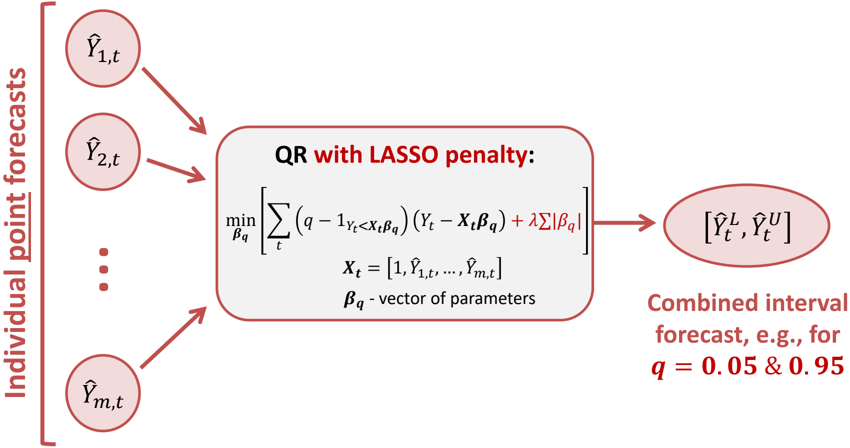
Factor QRA (FQRA): When the number of predictors is large

(Maciejowska, Nowotarski & Weron, 2016, IJF)



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

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



Isotonic Distributional Regression (IDR)

(Henzi et al., 2021, JRSSB; Lipiecki et al., 2024, ENEECO)

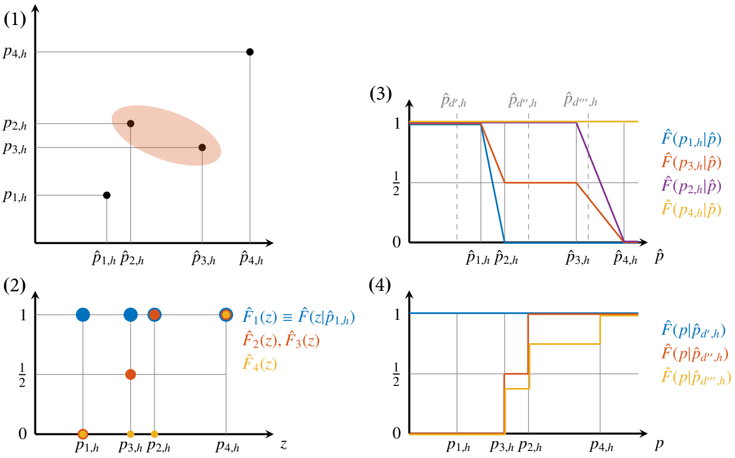
- A nonparametric method for estimating distributions that are isotonic in the regressed variable, i.e., quantiles are nondecreasing wrt the regressor
  - Introduced by Henzi et al. (2021, JRSSB) for weather forecasting
  - Applied by Lipiecki, Uniejewski & Weron (2024, ENEECO) in EPF
- First, all pairs $(\hat{y}_t, y_t)$ are sorted to be ascending in $\hat{y}_t$ for all $t \in \mathcal{S}$: $(\hat{y}_i^{\uparrow}, y_i^{\uparrow})_{i=1}^n$
- Then, the conditional distributions $\hat{F}_i(z) = \hat{F}(z|\hat{y}_i^{\uparrow})$ are obtained by solving the following min-max problem via the abridged pool-adjacent-violators algorithm:

$$\hat{F}_i(z) = \min_{k=1,\dots,i} \max_{j=k,\dots,n} \frac{1}{j-k+1} \sum_{l=k}^{j} \mathbb{1}\{y_l^{\uparrow} < z\}, \text{ where } z \in (y_i^{\uparrow})_{i=1}^n$$

- $\hat{F}(z|\hat{y})$ for any $\hat{y} \in \mathbb{R}$ is obtained by interpolating between $\hat{F}_i(z)$'s

Isotonic Distributional Regression (IDR)

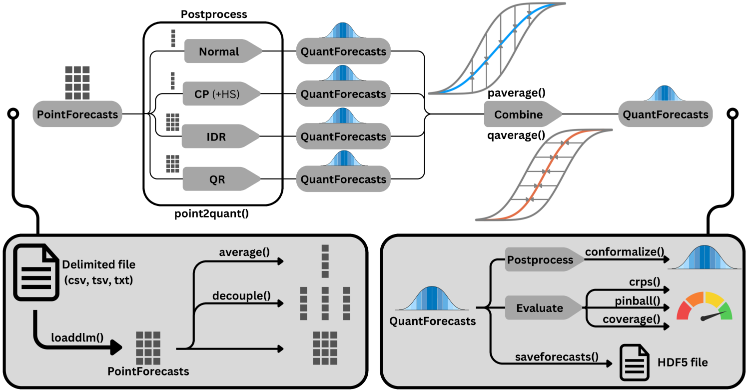
(Lipiecki et al., 2024, ENEECO)



1. $\mathcal{S}$ of $m = 4$ days for hour h, with price predictions $\hat{p}_{1,h} \leqslant \ldots \leqslant \hat{p}_{4,h}$ and respective prices $p_{1,h}, \ldots, p_{4,h}$
2. Conditional CDFs after pooling points in the orange ellipse
3. Interpolated $\hat{F}(z|\hat{p})$ as a function of $\hat{p}$; $\hat{F}_p$ for the next day is obtained from the intersections of $\hat{F}(p_{i,h}|\hat{p})$ and a vertical line at the next day's point forecast
4. $\hat{F}_p$ corresponding to three hypothetical next day's forecasts: $\hat{p}_{d',h}, \ldots, \hat{p}_{d''',h}$

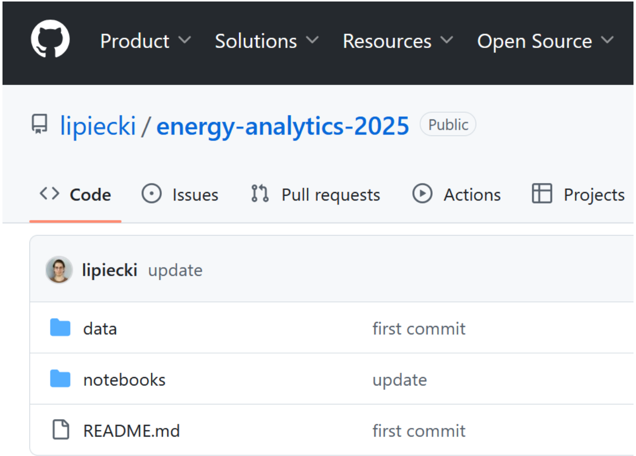
PostForecasts.jl: Julia package for postprocessing forecasts

(Lipiecki & Weron, 2025, WP + GitHub)



Experiment yourself:

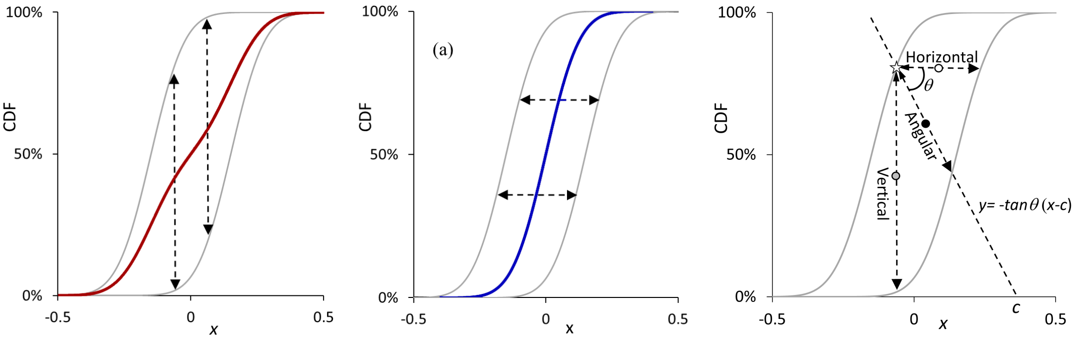
Julia snippet: PostForecasts.ipynb



Combining probabilistic forecasts is tricky ...

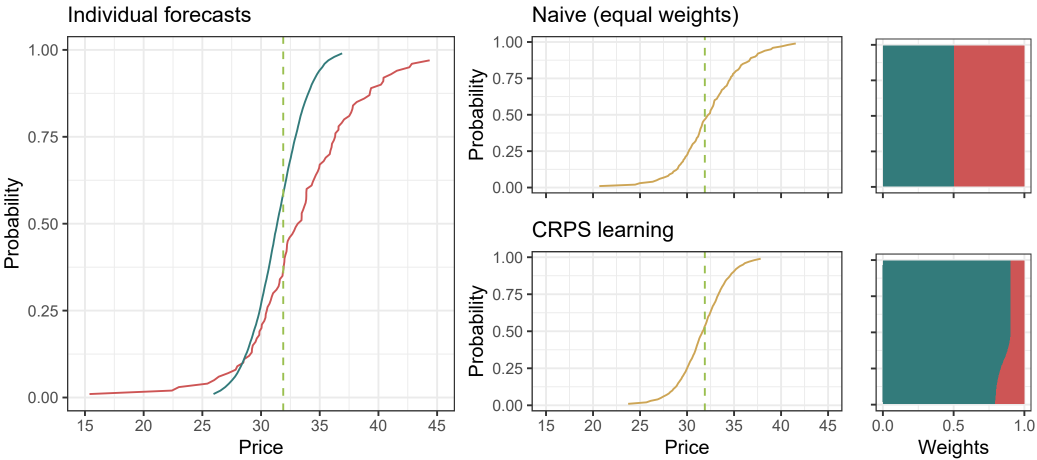
(Lichtendahl et al., 2013, Mgnt Sci; Uniejewski et al., 2019, ENEECO; Taylor & Meng, 2023, arXiv)

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



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

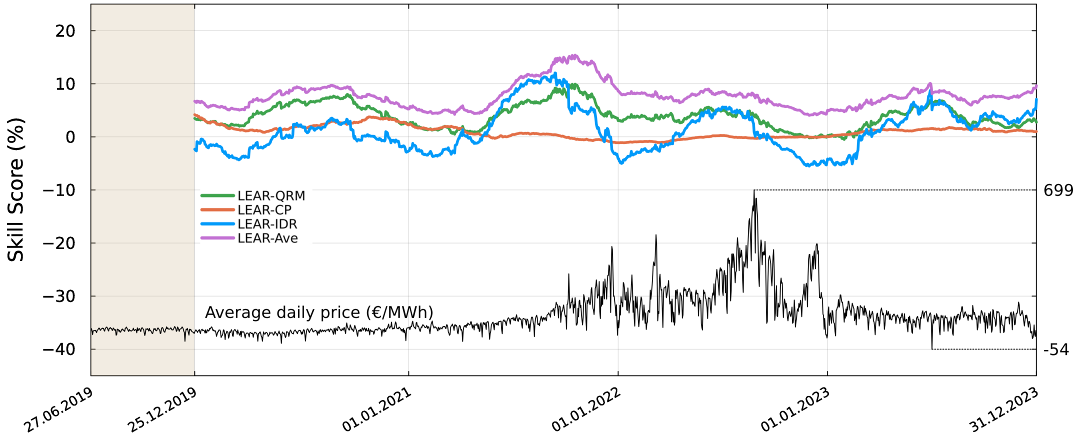
(Berrisch & Ziel, 2023, JEconometrics; Nitka & Weron, 2023, ORD)



... but can be beneficial

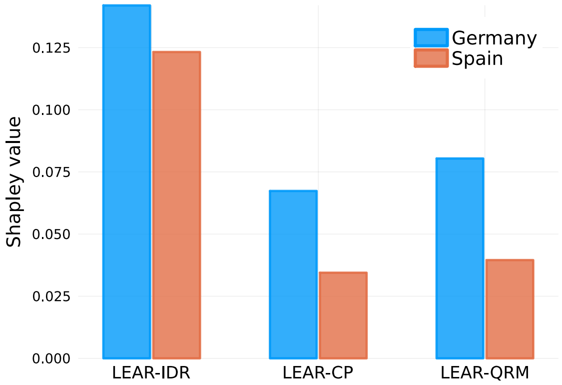
(Lipiecki, Uniejewski & Weron, 2024, ENEECO; illustrated here using EPEX-DE data)

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



Contribution of CP, QRM and IDR to the ensemble

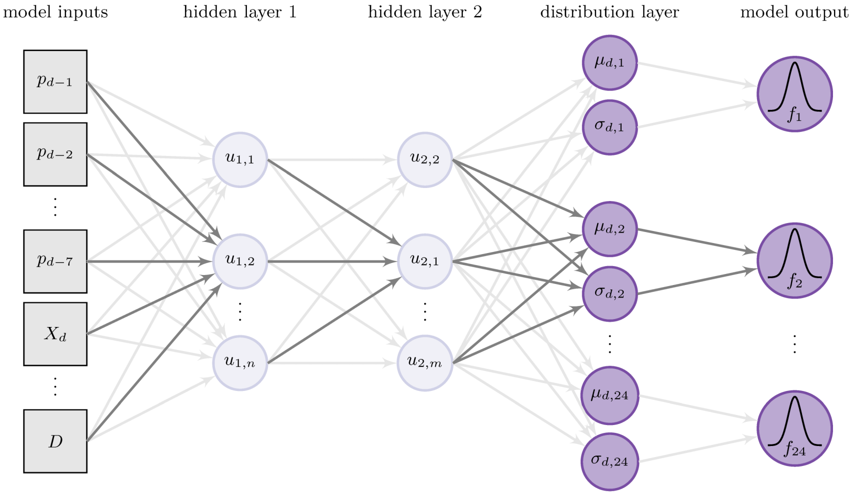
(Lipiecki, Uniejewski & Weron, 2024, ENEECO)



- Shapley contributions, i.e., Shapley values without the empty coalition, for 4.5-year test sets
- IDR contributes the most despite being the worst of the three
- In 2023 the contribution of IDR exceeds 75% in both markets
- Combining IDR-generated $\hat{F}$ with those of (better performing) QRA and CP significantly improves the accuracy

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*

ARTICLE INFO

JEL classification:
C44
C45
C46
C22
C53
Q47

Keywords:
Distributional neural network
Probabilistic forecasting

ABSTRACT

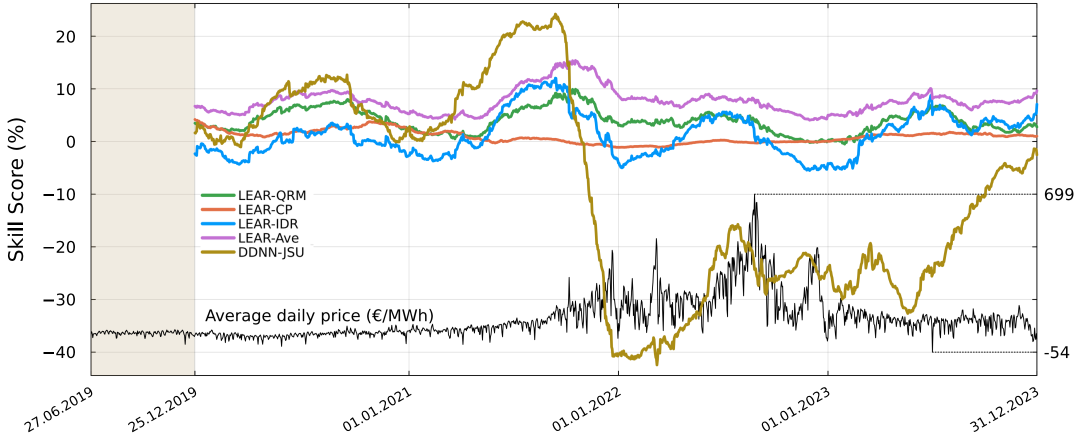
We present a novel approach to probabilistic electricity price forecasting which utilizes distributional neural networks. The model structure is based on a deep neural network containing a so-called probability layer, i.e., the outputs of the network are parameters of the normal or Johnson's SU distribution. 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

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

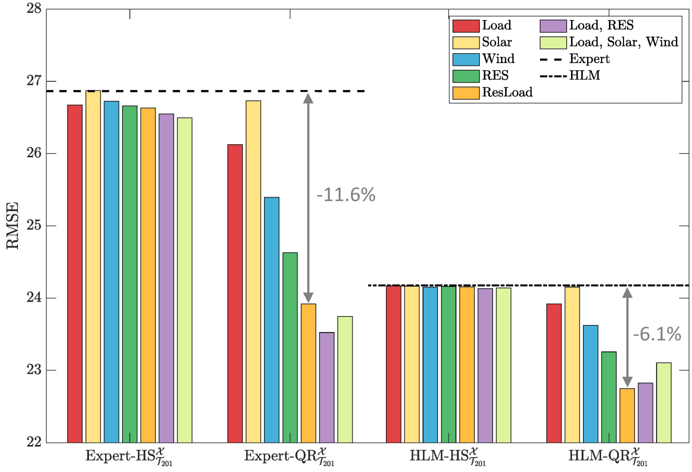
High-dimensional linear model (HLM), i.e., LEAR-type

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

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)



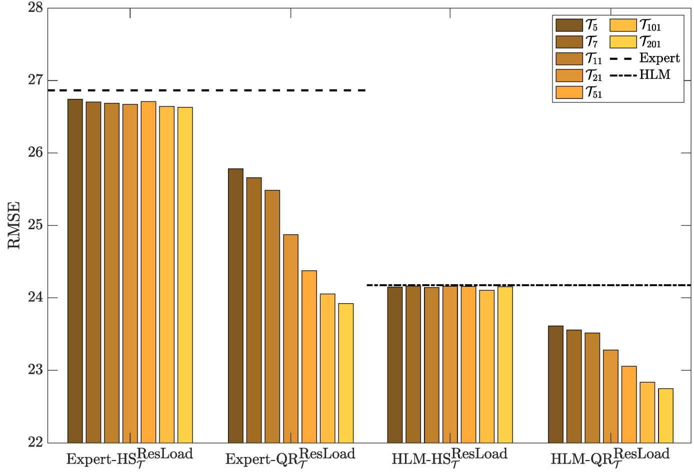
- RMSE for Germany (2018-23)
- Models with 201 quantiles ($\mathcal{T}_{201}$) approximating $\hat{F}_X$

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

- 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

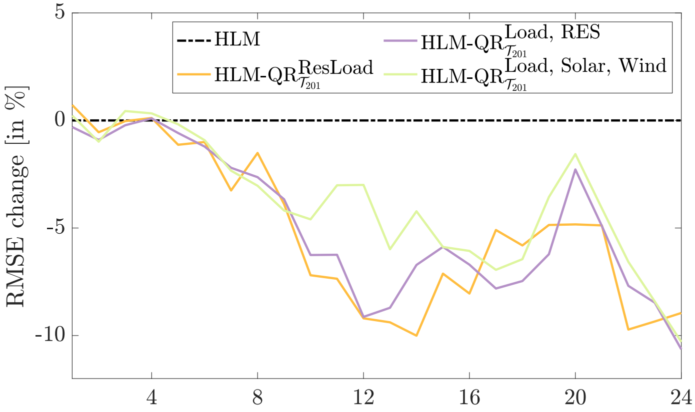
(Uniejewski & Ziel, 2025, arXiv)



- 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

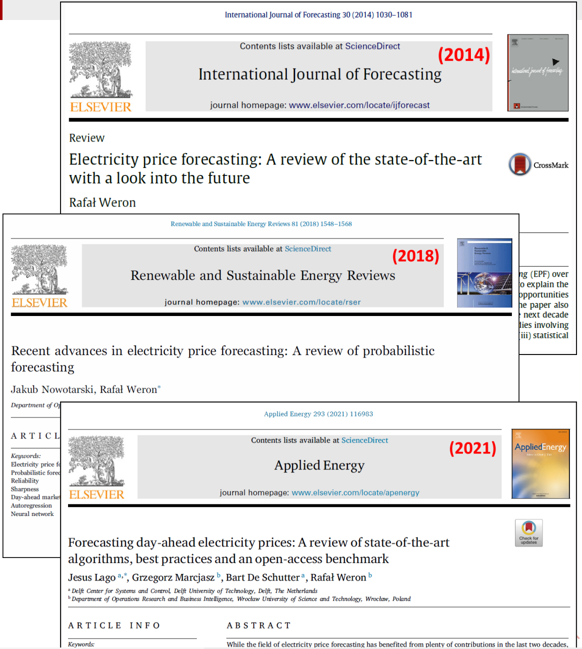
Wrap up: A simple technique to improve (point) price forecasts

(Uniejewski & Ziel, 2025, arXiv)



- The gains from probabilistic inputs are higher for afternoon and late night hours
- For these hours probabilistic inputs are selected more often
- Extreme quantile forecasts usually remain in the final LASSO-estimated model
- Overall, the residual load, i.e., $L - RES$, is the best fundamental predictor

1. Tips and tricks
2. Lasso and DNN
3. Probabilistic forecasts revisited
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

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

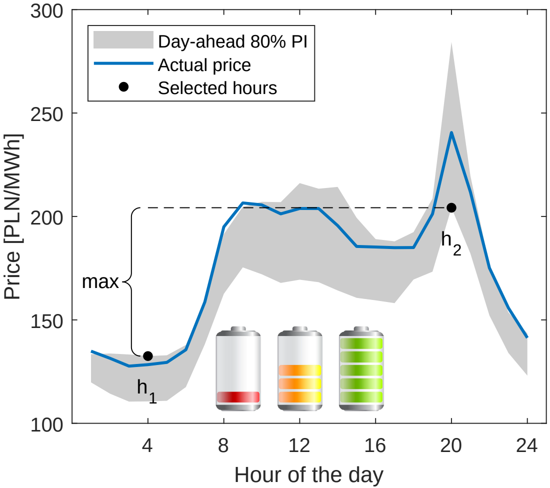
Keywords:

ABSTRACT

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

Quantile-based trading strategies

(Uniejewski & Weron, 2021, ENEECO)



Simultaneously bid day-ahead:

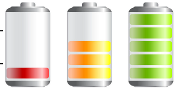
- buy 1 MW for $\hat{P}^{1-\alpha}_{d,h_1}$ at hour h_1
- sell 0.8 MW at $\hat{P}^{\alpha}_{d,h_2}$ at hour h_2

$\leftarrow$ 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

OPERATIONS RESEARCH AND DECISIONS QUARTERLY

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

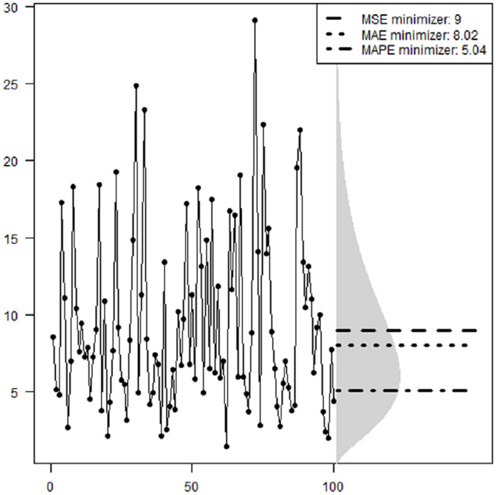
[*] Corresponding author, email address: weronika.nitka@pwr.edu.pl

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?

S. Kolassa / International Journal of Forecasting 36 (2020) 208–211



- 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?

Loss functions in regression models

(Serafin & Weron, 2024, WORMS)

ARX-type models with *asinh*-transformed regressors, one for the day-ahead price:

$$DA_{d,h} = \underbrace{\sum_{p=1}^{7} \beta_p DA_{d-p,h}}_{\text{AR-type effects}} + \underbrace{\beta_8 \overline{DA}_{d-1} + \beta_9 \underline{DA}_{d-1}}_{\text{yesterday's max \& min prices}} + \underbrace{\beta_{10} \hat{L}_{d,h} + \beta_{11} \hat{W}_{d,h}}_{\text{load \& wind prod. forecasts}}$$

$$+ \underbrace{\beta_{12} API2_{d-2} + \beta_{13} TTF_{d-2}}_{\text{coal \& NG prices}} + \underbrace{\sum_{p=1}^{7} \beta_{p+13} D_p}_{\text{daily dummies}} + \varepsilon_{d,h}$$

and one for the day-ahead price spread:

$$\Delta DA_{d,h_1,h_2} = \sum_{p=1}^{7} \beta_p \Delta DA_{d-p,h_1,h_2} + \beta_8 \left(\overline{DA}_{d-1} - \underline{DA}_{d-1}\right) + \beta_9 \Delta \hat{L}_{d,h_1,h_2}$$

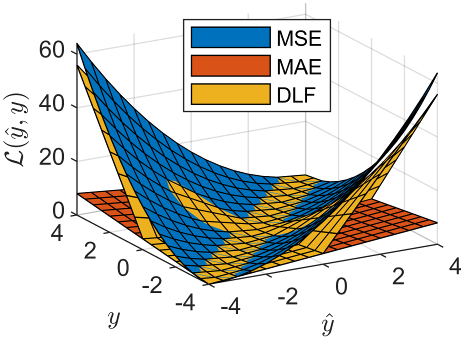
$$+ \beta_{10} \Delta \hat{W}_{d,h_1,h_2} + \beta_{11} API2_{d-2} + \beta_{12} TTF_{d-2} + \sum_{p=1}^{7} \beta_{p+12} D_p + \varepsilon_{d,h_1,h_2}$$

Loss functions in regression models cont.

(Serafin & Weron, 2024, WORMS)

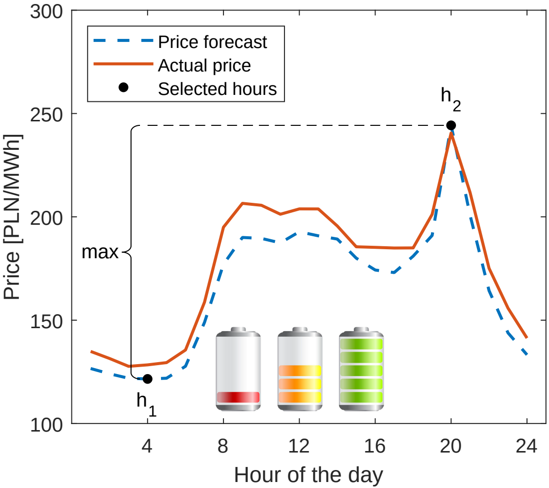
Consider the following loss functions:

- MSE: $\mathcal{L}(\hat{y}_t, y_t) = \text{mean}(y_t - \hat{y}_t)^2$
- MAE: $\mathcal{L}(\hat{y}_t, y_t) = \text{mean}|y_t - \hat{y}_t|$
- sMAPE: $\mathcal{L}(\hat{y}_t, y_t) = \text{mean}\frac{|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|}$
- PS for $q = 10\%, 20\%, \ldots, 90\%$:
  $\mathcal{L}(\hat{y}_t, y_t) = \text{mean}(\mathbb{1}_{y \leqslant \hat{y}_t} - q)(y_t - \hat{y}_t)$
- Directional Loss Function (DLF) which linearly penalizes $\widehat{\Delta DA}$ of the same sign, but adds a quadratic cost to predictions of wrong sign:
  $\mathcal{L}(\hat{y}_t, y_t) = \text{mean}\Big(|y_t - \hat{y}_t| + \max(0, -3 y_t \hat{y}_t)\Big)$



Point forecasts-based trading strategies

BESS operating cost of 100 EUR per charge-discharge cycle (Serafin & Weron, 2024, WORMS)



- Compared to Uniejewski & Weron (2021, ENEECO), threshold T limits trading to the most profitable days
- After selecting (h_1, h_2) check if the expected profit exceeds T:

$$\widehat{\Delta DA}_{d,h_1,h_2} > T$$

- If met, place price-taker buy and sell orders, otherwise do not trade

Point forecasts-based trading strategies cont.

(Serafin & Weron, 2024, WORMS)

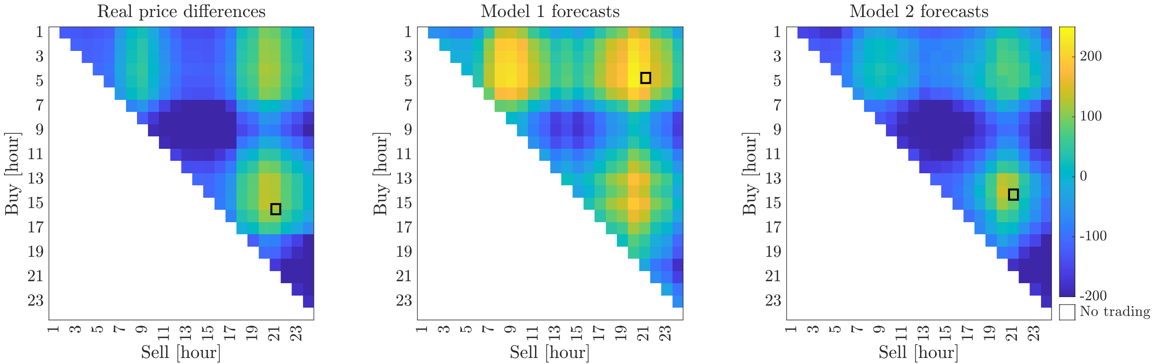


Figure 2: Real price differences, differences of predicted prices from Model 1 and predicted price spreads from Model 2 on 09.12.2023. Forecasts are obtained by minimizing the MAE. Black rectangles – pair of hours with the highest price difference (*from left to right*): (15, 20), (5, 20) and (14, 20)

German market data from ENTSO-E Transparency

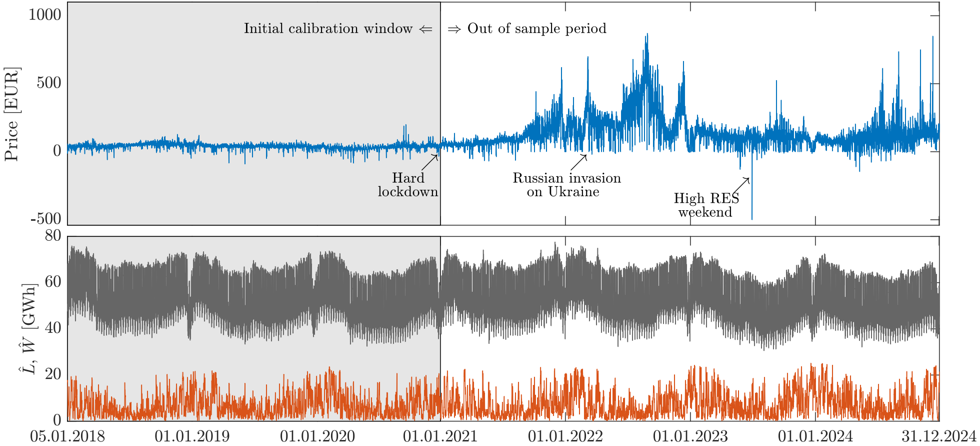
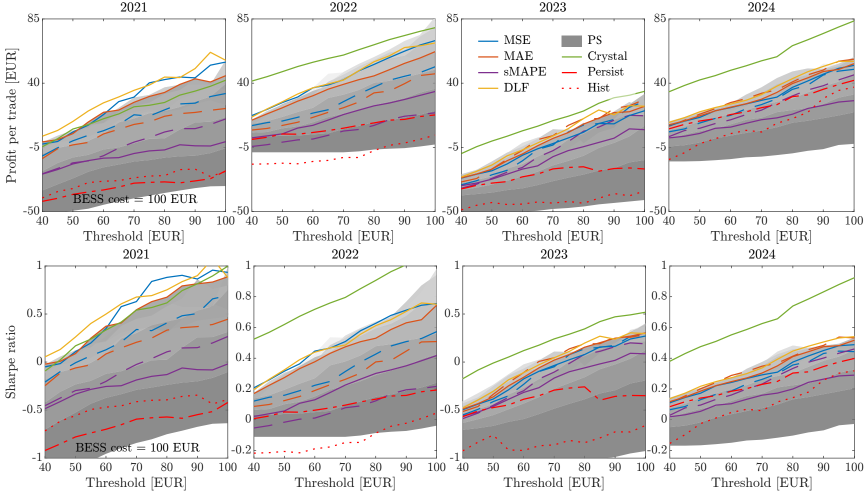


Figure 1: Day-ahead prices , load forecasts and wind generation forecasts. Events: hard lockdown in Germany (15.12.2020), Russian invasion of Ukraine (24.02.2022), and a high RES generation weekend (on Sunday 02.07.2023 at 3pm the price reached −500 EUR/MWh)

Profits per trade and Sharpe ratios

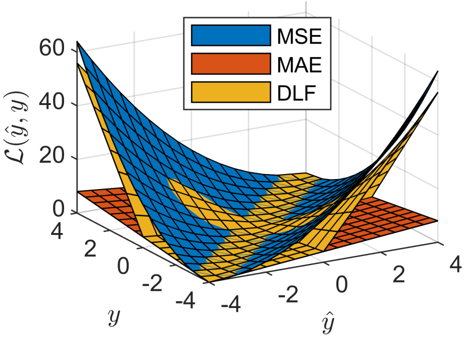
DLF and PS (→ light gray) for low q are the best-performing models



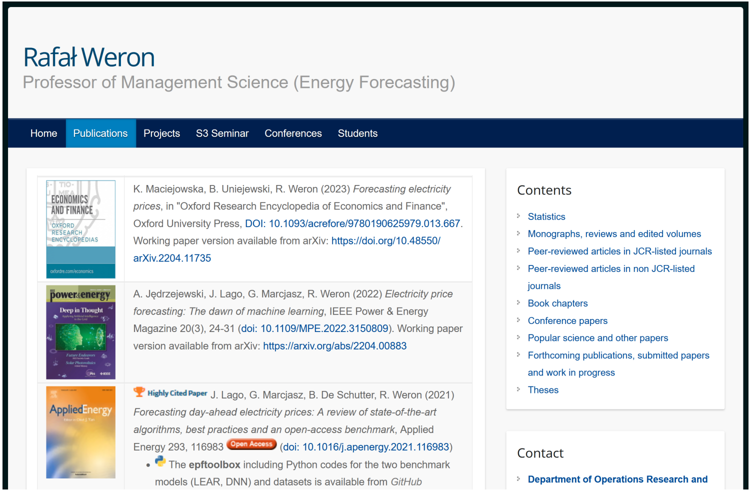
Loss functions in regression models: Conclusions

(Serafin & Weron, 2024, WORMS)

- DLF and PS for low q (here: 10%) exhibit the most consistent performance
- They are the best-performing models in terms of profits per trade and Sharpe ratios
- The total profit is commonly used in the literature, but
  - It favors frequent trading
  - Can lead to misleading conclusions, esp. when the BESS operating costs are ignored



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

- Statistics
- Monographs, reviews and edited volumes
- Peer-reviewed articles in JCR-listed journals
- Peer-reviewed articles in non JCR-listed journals
- Book chapters
- Conference papers
- Popular science and other papers
- Forthcoming publications, submitted papers and work in progress
- Theses

Contact

- **Department of Operations Research and**

See yourself: