

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/360185790>

Forecasting Electricity Prices

Preprint · April 2022

DOI: 10.48550/arXiv.2204.11735

CITATION

1

READS

939

3 authors:



Maciejowska Katarzyna

Wrocław University of Science and Technology

27 PUBLICATIONS 848 CITATIONS

[SEE PROFILE](#)



Bartosz Uniejewski

Wrocław University of Science and Technology

24 PUBLICATIONS 927 CITATIONS

[SEE PROFILE](#)



Rafał Weron

Wrocław University of Science and Technology

240 PUBLICATIONS 12,553 CITATIONS

[SEE PROFILE](#)

Forecasting Electricity Prices

Katarzyna Maciejowska, Bartosz Uniejewski, Rafał Weron

*Department of Operations Research and Business Intelligence,
Wrocław University of Science and Technology, 50-370 Wrocław, Poland*

This version: April 26, 2022

Abstract

Forecasting electricity prices is a challenging task and an active area of research since the 1990s and the deregulation of the traditionally monopolistic and government-controlled power sectors. It is interdisciplinary by nature. It requires expertise in econometrics, statistics or machine learning for developing well-performing predictive models, in finance for understanding market mechanics, and in electrical engineering for comprehension of the fundamentals driving electricity prices.

Although electricity price forecasting aims at predicting both spot and forward prices, the vast majority of research is focused on short-term horizons which exhibit dynamics unlike in any other market. The reason is that power system stability calls for a constant balance between production and consumption, while being weather (both demand and supply) and business activity (demand only) dependent. The recent market innovations do not help in this respect. The rapid expansion of intermittent renewable energy sources is not offset by the costly increase of electricity storage capacities and modernization of the grid infrastructure.

On the methodological side, this leads to three visible trends in electricity price forecasting research as of 2022. Firstly, there is a slow, but more noticeable with every year, tendency to consider not only point but also probabilistic (interval, density) or even path (also called ensemble) forecasts. Secondly, there is a clear shift from the relatively parsimonious econometric (or statistical) models towards more complex and harder to comprehend, but more versatile and eventually more accurate statistical/machine learning approaches. Thirdly, statistical error measures are nowadays regarded as only the first evaluation step. Since they may not necessarily reflect the economic value of reducing prediction errors, more and more often, they are complemented by case studies comparing profits from scheduling or trading strategies based on price forecasts obtained from different models.

Keywords: forecasting, electricity price, day-ahead market, intraday market, variable selection, regularization, regression, quantile regression, neural net, statistical learning, machine learning, deep learning, forecast evaluation, economic value, trading strategy

1 Introduction

Electricity price forecasting (EPF)¹ as a research area of its own appeared in the early 1990s with the liberalization and deregulation of the power sectors in the UK and Scandinavia. The

¹We use EPF when referring to both electricity price forecasting and electricity price forecast(s). The plural form, i.e., forecasts, is abbreviated EPFs.

late 1990s and 2000s were marked by the widespread conversion from the traditionally monopolistic and government-controlled power sectors to competitive power markets in Europe, North America, Australia and eventually in Asia (Mayer and Trück, 2018). Over the years, EPFs have become a fundamental input to companies’ decision-making mechanisms (Weron, 2014). As Hong (2015) estimates, for a medium-sized utility with a 5-gigawatt annual peak load², improving the day-ahead demand forecasts by 1% leads to annual savings of ca. 1.5 million USD. With the additional price forecasts, the savings double. Clearly, the time invested in developing EPF models can pay off.

For newcomers to this research area it is important to realize that the literature has generally focused on horizons of up to 48 hours, since short-term price dynamics is what makes electricity special. In the longer term, prices are averaged across weekly, monthly or annual delivery periods and lose much of their uniqueness. In the short-term, on the other hand, electricity prices exhibit significant seasonality at different levels (daily, weekly and in many markets also annual), short-lived and generally unanticipated price spikes (ranging up to two orders of magnitude), and in some markets even negative values.

However, the “short-term” is not a particular horizon, but a whole spectrum of horizons ranging from a few minutes ahead (real-time, *intraday*, ID; also called “spot” in North America) to a *day-ahead* (DA; called “spot” in Europe). Note that from a financial perspective, both the ID and DA contracts can be regarded as very short-term forwards, with delivery during a particular hourly (half- or quarter-hourly) load period on the same or the next day. Since each day can be divided into a finite number of load periods $h = 1, 2, \dots, H$ with $H = 24, 48$ or 96 , it is common to use double indexing when referring to the electricity price. Here, we denote by $P_{d,h}$ the price for day d and load period h , by $\hat{P}_{d,h}$ its point forecast and by \hat{F}_P or $\hat{F}_{P_{d,h}}$ its predictive distribution.

In what follows, we provide an overview of EPF research, with a particular focus on three current trends:

- #1: increasing popularity of *probabilistic* (interval, density) and *path* (also called *ensemble*) forecasts,
- #2: a visible shift towards *statistical/machine learning* (SL/ML), and
- #3: evaluating the economic value of price predictions.

However, before we start, let us first briefly describe the marketplace and the typical forecasting tasks considered.

2 The Marketplace

As a result of the aforementioned liberalization and deregulation of the power sectors, two basic models for power markets have emerged: power pools – where trading, dispatch and transmission are managed by the *system operator* (SO), and power exchanges – where trading and initial dispatch are managed by an institution independent from the *transmission system operator* (TSO). Participation in power pools is limited to generators and is typically mandatory.

²The annual peak load is the highest electrical power demand in a (calendar) year. The power consumed or generated is measured in multiples of the watt (W). Smaller power plants can generate tens of megawatts (1 MW = 10^6 W), the largest tens of gigawatts (1 GW = 10^9 W). The amount of electricity consumed or generated over a specific period of time is typically measured in megawatt-hours (MWh); it is also the basic unit used in trading electricity.

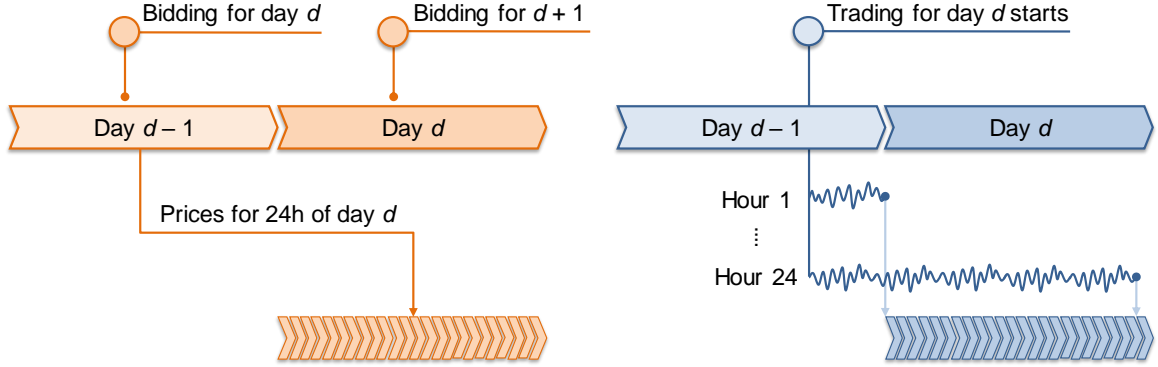


Figure 1: Illustration of bidding and price settlement in auction (*left*) and continuous trading (*right*) power markets. In day-ahead auctions the bids for all load periods (here: hours) of day d can be submitted until a certain hour on day $d - 1$. Intraday markets which admit continuous trading run 24/7 from an afternoon hour on day $d - 1$ up until a few minutes before the delivery on day d .

The *market clearing price* (MCP) is established through a one-sided auction as the intersection of the supply curve constructed from aggregated supply bids of the generators and the demand predicted by the system operator. Often a separate price for each node in the network is calculated, so-called *locational marginal price* (LMP). Such a system was adopted in highly meshed North American networks. On the other hand, in Australia, where the network structure is simpler, zonal pricing was successfully implemented, where for areas without grid limitations a unique price is settled.

In contrast to power pools, participation in power exchanges is – except for some special cases – voluntary and open not only to generators, but also to wholesale consumers and speculators. The price is established either through a two-sided auction (DA, ID) as the intersection of the supply curve constructed from aggregated supply bids and the demand curve constructed from aggregated demand bids or in continuous trading (ID). Most market designs have adopted the uniform-price auction, where buyers who bid at or above the MCP pay that price and sellers who bid at or below the MCP are paid this price. Moreover, in auction markets the bids can be submitted until a certain time – called *gate closure* – which is the same for all load periods, see the left panel in Fig. 1. Hence, auction prices could be viewed as realizations of a multivariate random variable and therefore prices for all load periods should be predicted simultaneously (Ziel and Weron, 2018). On the other hand, some ID markets allow for continuous trading. They run 24/7 from an afternoon hour on day $d - 1$ up until a few minutes before the delivery of electricity during a particular load period on day d , see the right panel in Fig. 1.

In some countries (e.g., Germany, Ireland, Poland) the DA and ID markets are complemented by the so-called balancing market. This technical market is used for pricing differences between the market schedule and actual system demand for very short time horizons before delivery. For instance, the TSO might instruct a generator to increase its output to meet a sudden surge in demand. The producer then receives a premium via the balancing market for the energy generated used to balance the grid.

The timeline of day-ahead and intraday trading activities in selected European countries is illustrated in Fig. 2. As can be seen, the DA and ID markets complement each other. Once the

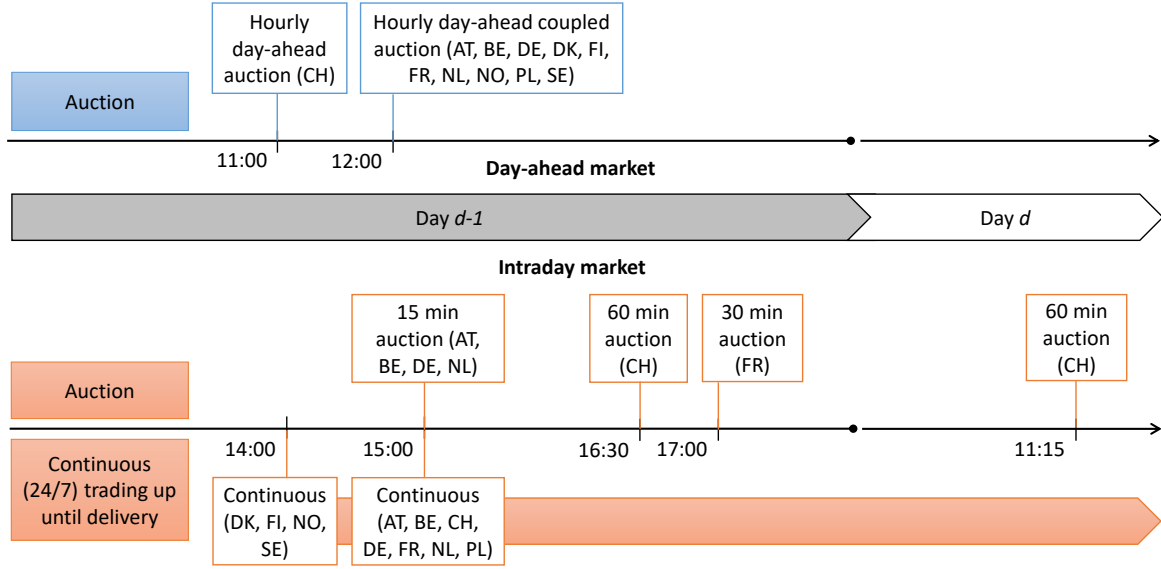


Figure 2: The timeline of day-ahead (*top*) and intraday (*bottom*) trading activities for delivery of electricity on day d in selected European countries: Austria (AT), Belgium (BE), Denmark (DK), Germany (DE), Finland (FI), France (FR), the Netherlands (NL), Norway (NO), Poland (PL), Sweden (SE) and Switzerland (CH).

gate closes for day-ahead bids around noon, various intraday markets open for adjusting these bids. They are particularly important for nondispatchable, stochastic producers such as wind or solar farms, and include both auctions and continuous trading. Note that both the ID and DA contracts can concern delivery during the same load period, only the time the decision has to be made and the bid placed differs.

The presented sequence of events has important implications for study design. In the DA market the forecasting horizons typically range from 12-14 hours for the first load period of the next day to 36-38 hours for the last. However, at the time the predictions are made, i.e., the morning hours of day $d - 1$, the DA prices for all load periods of this day are already known (they were settled around noon on day $d - 2$). Generally, the TSO day-ahead forecasts of the system load (\approx demand) and the system-wide generation from *renewable energy sources* (RES) are also available to market participants at this time.

When the ID market is considered, the selection of the forecasting horizon depends on the research question. Firstly, the predictions can be made on the morning of day $d - 1$, when market participants need to decide how much electricity to bid in the DA market and how much to buy/sell in the ID market or leave for the balancing market. Forecasts of the price spread between DA and ID/balancing markets can provide valuable insights for decision-making (Maciejowska et al., 2019, 2021).

Secondly, the predictions can be used for bidding in ID markets with continuous trading. Although the trading floor opens in the afternoon hours of day $d - 1$, the majority of bids are placed during the last 3-4 hours before the delivery (Narajewski and Ziel, 2020a). Hence, the forecasting horizons considered typically range from a couple of minutes to 4 hours (Janke and Steinke, 2019; Uniejewski et al., 2019b; Narajewski and Ziel, 2020b). Note that different model specifications may be optimal for predicting ID prices for different horizons (Maciejowska

et al., 2020). Since the bidding behavior of market participants is significantly influenced by RES generation forecasts which are available at the time of trading (Kiesel and Paraschiv, 2017; Kulakov and Ziel, 2021), ID price forecasts should not only exploit the short-term price dependencies but also updated predictions of wind and solar power generation. Interestingly, including self-exciting terms in ID models allows to better capture the empirically observed trade clustering (Kramer and Kiesel, 2021).

3 Trend #1: From Point to Probabilistic and Ensemble Forecasts

By far point forecasts are the most popular. Despite a few early attempts, often inspired by developments in wind forecasting (Hong et al., 2020), probabilistic forecasting was not part of the mainstream EPF literature until 2014 and the Global Energy Forecasting Competition (GEFCom2014; Hong et al., 2016). Probabilistic EPF quickly gained momentum and energy analysts have become aware of its importance in energy systems planning and operations. A variety of approaches have been considered, including bootstrapping (Chen et al., 2012), QRA (see Section 3.2), Bayesian statistics (Kostrzewski and Kostrzewska, 2019) and deep learning (Mashlakov et al., 2021; Jędrzejewski et al., 2022). Nevertheless, as of today, no more than 15% of Scopus-indexed articles concern interval or distributional EPF. Path (also called *ensemble*) forecasts, which focus on the multidimensional temporal distribution, are even less popular. Yet, path-dependency is crucial for many optimization problems arising in power plant scheduling, energy storage and trading, and this has been recognized in the recent EPF literature (Janke and Steinke, 2020; Narajewski and Ziel, 2020b).

3.1 Error and Price Distributions

There are two main approaches to probabilistic forecasting: the more elegant one directly considers the distribution of the electricity price, while the more popular one builds on the point forecast and the distribution of errors associated with it. In both cases, the focus can be on prediction intervals, selected quantiles or the whole predictive distribution. For reviews on short- and medium-term probabilistic EPF we refer to Nowotarski and Weron (2018) and Ziel and Steinert (2018), respectively, while for a general treatment to the seminal review of Gneiting and Katzfuss (2014).

If we assume that the point forecast is the expected price³ at a future time point, i.e., $\hat{P}_{d,h} = \mathbb{E}(P_{d,h})$, then we have:

$$P_{d,h} = \hat{P}_{d,h} + \varepsilon_{d,h}, \quad (1)$$

and the distribution F_ε of errors associated with $\hat{P}_{d,h}$ is identical to the distribution F_P of prices, except for a horizontal shift by $\hat{P}_{d,h}$:

$$F_P(x) \equiv \mathbb{P}(P_{d,h} \leq x) = \mathbb{P}(\hat{P}_{d,h} + \varepsilon_{d,h} \leq x) = \mathbb{P}(\varepsilon_{d,h} \leq x - \hat{P}_{d,h}) \equiv F_\varepsilon(x - \hat{P}_{d,h}).$$

This, however, implies that the inverse empirical *cumulative distribution functions* (also called *quantile functions*) satisfy:

$$\hat{F}_P^{-1}(\alpha) = \hat{P}_{d,h} + \hat{F}_\varepsilon^{-1}(\alpha), \quad (2)$$

³Although this is the most common assumption, the point forecast does not have to be the expected value. For instance, it can be the median or any quantile of the predictive distribution.

i.e., they are identical except for a shift by $\hat{P}_{d,h}$, but this time on the vertical axis. Equation (2) provides the basic framework for constructing probabilistic forecasts from prediction errors. If a dense grid of quantiles is considered, e.g., 99 percentiles, then \hat{F}_P can be approximated pretty well (Hong et al., 2016; Nowotarski and Weron, 2018; Uniejewski and Weron, 2021).

If we assume that F_P has a density f_P , then a density forecast \hat{f}_P can be provided as well. However, Ziel and Steinert (2016) argue against using such an approach. Analyzing the fine structure of aggregated supply and demand curves in the German market they found that F_P was multimodal with significant jumps (corresponding to point masses) at certain ‘round’ prices.

3.2 Quantile Regression Averaging

Quantile regression (QR; see Koenker, 2005) is one of the most popular methods for directly modeling the distribution of a random variable. QR approximates the target quantile with a linear function of a set of explanatory variables. In the EPF context, these variables typically contain publicly available market information (load forecasts, generation structure, historical electricity prices, etc.; Bunn et al., 2016; Maciejowska, 2020) and/or point predictions of electricity prices (Weron, 2014). The later case leads to the so-called *Quantile Regression Averaging* (QRA) introduced by Nowotarski and Weron (2015) and originally developed for Team Poland’s participation in the GEFCom2014 competition (Maciejowska and Nowotarski, 2016). It is a forecast combination approach to the computation of quantile forecasts, which bridges the gap between point and probabilistic forecasts. QRA involves applying QR to the point forecasts of a small number of individual forecasting models or experts:

$$q(\alpha|\mathbf{X}_{d,h}) = \mathbf{X}_{d,h}\boldsymbol{\beta}_\alpha, \quad (3)$$

where $q(\alpha|\cdot)$ is the conditional α th quantile of F_P , $\mathbf{X}_{d,h}$ is the vector of point forecasts and $\boldsymbol{\beta}_\alpha$ is the corresponding vector of weights. The latter is estimated by minimizing the following sum of *check functions*:

$$\hat{\boldsymbol{\beta}}_\alpha = \underset{\boldsymbol{\beta}_\alpha}{\operatorname{argmin}} \left\{ \sum_{d,h} \underbrace{(\alpha - \mathbb{1}_{P_{d,h} < \mathbf{X}_{d,h}\boldsymbol{\beta}_\alpha}) (P_{d,h} - \mathbf{X}_{d,h}\boldsymbol{\beta}_\alpha)}_{\text{check function}} \right\}. \quad (4)$$

The very good forecasting performance of QRA has been verified by a number of authors, not only in the area of EPF (Liu et al., 2017; Kostrzewski and Kostrzewska, 2019; Kath and Ziel, 2021; Uniejewski and Weron, 2021). However, its most spectacular success came during the GEFCom2014 competition, when teams using variants of QRA (Gaillard et al., 2016; Maciejowska and Nowotarski, 2016) were ranked in the top two places in the price track.

3.3 Paths and Ensembles

Although the concept of probabilistic EPF is much more general than of point forecasting, it is not sufficient to support operational decisions that depend on future trajectories of electricity prices. For instance, in Germany renewable energy producers can receive less subsidies if the electricity price is negative for 6 hours in a row. Hence, instead of looking at the 24 hourly univariate price distributions $F_{P_{d,1}}, \dots, F_{P_{d,24}}$, we should be focusing on the multidimensional distribution \mathbf{F}_P of the 24-dimensional price vector $\mathbf{P}_d = (P_{d,1}, \dots, P_{d,24})'$. However, many models considered in the literature cannot output such a multidimensional forecast.

Ensemble forecasts provide a practical remedy. An ensemble is a collection of simulated price *paths*, also called *trajectories* or *scenarios*. For a large number of paths the ensemble approximates the underlying distribution \mathbf{F}_P arbitrarily well (Weron and Ziel, 2020). In practice ‘large’

means thousands or millions of paths, which may be a computational challenge (Narajewski and Ziel, 2020b). It should be noted that, on one hand, the same or similar concepts have been used in different disciplines under different names, e.g., *simultaneous prediction intervals*, *prediction bands*, *spatio-temporal trajectories*, *numerical weather prediction ensembles*. On the other, the term ensemble is also used to refer to any averaging of – point or probabilistic – forecasts (Hong et al., 2020).

4 Trend #2: From Regression to Statistical and Machine Learning

Until the mid 2010s, the EPF literature was dominated by relatively parsimonious linear regression and neural network models. They were characterized by a small number – a dozen or two – of *inputs* (also called *features*, *input features*, *explanatory variables*, *regressors*, or *predictors*) and complex data pre-/post-processing:

- replacing outliers, i.e., price spikes, by more ‘normal’ values before estimating the model (Contreras et al., 2003; Bierbrauer et al., 2007; Janczura et al., 2013) or utilizing robust estimation methods (Grossi and Nan, 2019);
- averaging forecasts, both across models (Nan, 2009; Bordignon et al., 2013; Nowotarski et al., 2014) and across calibration windows for the same model (Marcjasz et al., 2018; Hubicka et al., 2019), also in a probabilistic EPF setting (see the QRA approach in Section 3.2 and Serafin et al., 2019);
- using so-called variance stabilizing transformations (VSTs; Schneider, 2011; Diaz and Planas, 2016; Uniejewski et al., 2018; Narajewski and Ziel, 2020a; Shi et al., 2021) to make the marginal distributions less heavy-tailed (Box-Cox family, area hyperbolic sine) or Gaussian (Probability Integral Transform, see Section 5.1.2);
- deseasonalizing the data with respect to the long-term seasonal component (LTSC) before estimating the model (Janczura et al., 2013; Nowotarski and Weron, 2016; Lisi and Pelagatti, 2018; Afanasyev and Fedorova, 2019; Uniejewski et al., 2019a; Marcjasz et al., 2020).

However, as more data and computational power became available, the models became more complex to the extent that expert knowledge was no longer enough to handle them (Jędrzejewski et al., 2022). This paved the way for statistical/machine learning in EPF. Arguably, *statistical learning* (SL) and *machine learning* (ML) are synonyms.⁴ They have just originated in different communities – computer science/artificial intelligence (Mitchell, 1997) or computational statistics (James et al., 2021). Both SL and ML refer to a vast set of (computational, statistical) tools for understanding data, both can improve “automatically” through training. In either case, learning can be supervised or unsupervised. In EPF we are typically interested in supervised learning, which involves building a model for predicting a known output or outputs based on a set of inputs.

⁴Januschowski et al. (2020) even argue that the distinction between *statistical* and *machine learning* forecasting is dubious, as this distinction does not stem from fundamental differences in the methods assigned to either class, but rather is of a “tribal” nature.

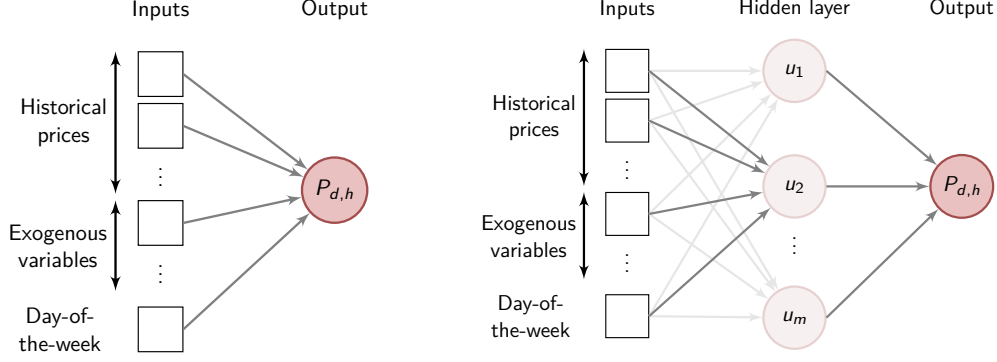


Figure 3: Visualization of a linear regression model (*left*) and a shallow neural network (*right*) with identical inputs and output, i.e., the electricity price for day d and hour h . White squares represent the inputs and u_1, u_2, \dots, u_m the hidden nodes (or neurons). Arrows indicate the flow of information.

4.1 The Expert Benchmark

A class of commonly used EPF benchmarks is based on a parsimonious *autoregressive* (AR) structure with exogenous variables and calendar effects, originally proposed by Misiorek et al. (2006). Since expert knowledge is used to select the regressors, such benchmarks are often called *expert* models (Ziel and Weron, 2018). One of the most popular structures represents the electricity price for day d and hour h by:

$$\begin{aligned}
 P_{d,h} = & \underbrace{\beta_1 P_{d-1,h} + \beta_2 P_{d-2,h} + \beta_3 P_{d-7,h}}_{\text{autoregressive effects}} + \underbrace{\beta_4 P_{d-1,24}}_{\text{end-of-day}} + \underbrace{\beta_5 P_{d-1}^{\max} + \beta_6 P_{d-1}^{\min}}_{\text{non-linear effects}} \\
 & + \underbrace{\beta_7 X_{d,h}^1 + \beta_8 X_{d,h}^2}_{\text{exogenous variables}} + \underbrace{\sum_{j=1}^7 \beta_{h,j+8} D_j}_{\text{weekday dummies}} + \varepsilon_{d,h},
 \end{aligned} \tag{5}$$

where $P_{d-1,h}$, $P_{d-2,h}$ and $P_{d-7,h}$ account for the autoregressive effects and correspond to prices from the same hour h of the previous day, two days before and a week before, $P_{d-1,24}$ is the last known price at the time the prediction is made and provides information about the end-of-day price level, P_{d-1}^{\max} and P_{d-1}^{\min} represent previous day's maximum and minimum prices, $X_{d,h}^1$ and $X_{d,h}^2$ are exogenous variables, D_1, \dots, D_7 are weekday dummies and $\varepsilon_{d,h}$ is the noise term (i.i.d. variables with finite variance). The β_i 's are estimated using *ordinary least squares* (OLS).

Autoregression or more generally linear regression is one of the two most commonly used classes of EPF models (Weron, 2014). The other is the *multi-layer perceptron* (MLP). The simplest neural network, a single-layer perceptron, contains no hidden layers (only inputs and the output) and is equivalent to a linear regression – both represent $P_{d,h}$ by a linear combination of input features, see the left panel in Fig. 3. On the other hand, the MLP includes at least one hidden layer and utilizes a feed-forward architecture – the outputs of the nodes (or neurons) in one layer are inputs to the next one, see the right panel in Fig. 3. Since the output of a node is a weighted sum of all of the inputs transformed by a typically nonlinear activation function, unlike in linear regression, NNs can tackle complex dependence structures encountered in power market data (Keles et al., 2016).

Exogenous variables typically include day-ahead predictions of the system load and RES generation (Lago et al., 2021). Days with high demand and low RES generation are characterized

by relatively high prices. On the other hand, high RES generation pulls prices down; in periods of low demand – holidays and/or at night – even below zero (Zhou et al., 2016). Other fundamental variables may possess explanatory power as well. For instance, fuel and CO₂ allowance prices, especially for medium-term EPF (Maciejowska and Weron, 2016; Ziel and Steinert, 2018). Due to the merit order effect, i.e., dispatching units characterized by the lowest marginal cost of production, the fuel–electricity price relationship changes throughout the day. Natural gas prices impact mainly the peak hours, whereas coal prices influence the off-peak hours. Finally, the day-of-the-week input feature visible in Fig. 3 can be a set of weekday dummies, as in Eqn. (5), or a single multi-valued variable, which is more common in NN models.

4.2 Regularization and the LEAR Model

Selecting regressors is a cumbersome task and expert knowledge does not always identify the relevant ones. In a series of papers in the mid 2010s, Ludwig et al. (2015), Ziel et al. (2015), Gaillard et al. (2016), Uniejewski et al. (2016) and Ziel (2016) introduced the concept of regularization to EPF. In simple terms, the idea behind this approach is to add a penalty term to the *residual sum of squares* (RSS) in OLS regression:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \text{RSS} + \lambda \sum_{i=1}^n |\beta_i|^q \right\}, \quad (6)$$

where λ is the *tuning* or *regularization* hyperparameter. Note, that *hyperparameters* are model parameters that cannot be optimized during the training (estimation) phase, but have to be set or calibrated beforehand, e.g., using cross-validation (James et al., 2021). For $q = 2$ we obtain *ridge regression* (Hoerl and Kennard, 1970) and for $q = 1$ the *least absolute shrinkage and selection operator* (LASSO; Tibshirani, 1996). The latter can shrink β_i ’s not only towards zero but actually to zero itself, thus effectively eliminating some regressors from the model. If we admit both terms in Eqn. (6), i.e., $\lambda_1 \sum |\beta_i| + \lambda_2 \sum \beta_i^2$, then we obtain the so-called *elastic net* (Zou and Hastie, 2015).

In the EPF setting, all three variants were compared in Uniejewski et al. (2016). Ridge regression easily outperformed expert models and stepwise regression techniques, but was significantly worse than the LASSO and the elastic net. At the cost of an additional parameter, the elastic net generally yields more accurate predictions than the LASSO. Nevertheless, the latter has become the golden standard in EPF (Uniejewski and Weron, 2018; Ziel and Weron, 2018; Janke and Steinke, 2019; Narajewski and Ziel, 2020a; Marcjasz, 2020; Zhang et al., 2020; Özen and Yıldırım, 2021). It was even utilized by Lago et al. (2021) to construct a well-performing EPF benchmark – the *LASSO-Estimated AutoRegressive* (LEAR) model. The starting point for

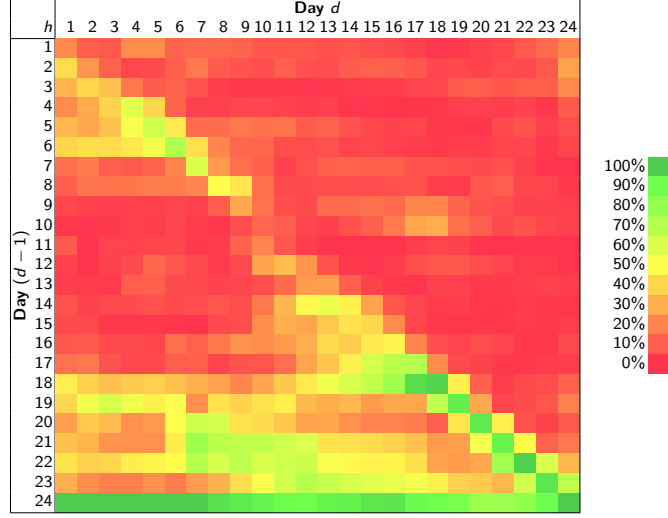


Figure 4: Mean occurrence of non-zero $\beta_{h,i}$'s across datasets from 12 power markets. Columns represent the predicted hours on day d and rows the first 24 variables, i.e., $\sum_{i=1}^{24} \beta_{h,i} P_{d-1,i}$ in Eqn. (7), of a LEAR-type model considered by Ziel and Weron (2018). A heat map is used to indicate more (\rightarrow green) and less (\rightarrow red) commonly-selected variables.

the LEAR model is a parameter-rich regression:

$$\begin{aligned}
P_{d,h} = & \sum_{i=1}^{24} (\beta_{h,i} P_{d-1,i} + \beta_{h,i+24} P_{d-2,i} + \beta_{h,i+48} P_{d-3,i} + \beta_{h,i+72} P_{d-7,i}) \\
& + \sum_{i=1}^{24} (\beta_{h,i+96} X_{d,i}^1 + \beta_{h,i+120} X_{d-1,i}^1 + \beta_{h,i+144} X_{d-7,i}^1) \\
& + \sum_{i=1}^{24} (\beta_{h,i+168} X_{d,i}^2 + \beta_{h,i+192} X_{d-1,i}^2 + \beta_{h,i+216} X_{d-7,i}^2) \\
& + \sum_{k=1}^7 \beta_{h,240+k} D_k + \varepsilon_{d,h},
\end{aligned} \tag{7}$$

which differs from the expert model in Eqn. (5) mainly by allowing for cross-hourly dependencies. In general, the price for hour h may depend on the prices for all 24 hours yesterday, the day before, etc. In practice, only a dozen or two of the potential 247 regressors turn out to be relevant. However, they need not be the ones included in the expert model.

This is visualized in Figure 4 for the first 24 variables, i.e., $\sum_{i=1}^{24} \beta_{h,i} P_{d-1,i}$, of a LEAR-type model considered by Ziel and Weron (2018) and across datasets from 12 power markets.⁵ The yellow-green diagonal indicates that the price for hour h on day $d-1$ is a good predictor of the price for the same hour on day d . The yellow-green bottom rows were a surprising finding

⁵BELPEX price for Belgium, EPEX prices for Switzerland, Germany–Austria and France, EXAA price for Germany–Austria, GEFCOM2014 competition data, Nord Pool prices for West Denmark, East Denmark and the system price, OMIE prices for Spain and Portugal, and OTE price for the Czech Republic. The GEFCOM2014 dataset covers a 3-year period (2011–2013; see Hong et al., 2016), the remaining datasets a 6-year period (July 2010 – July 2016; see Ziel and Weron, 2018).

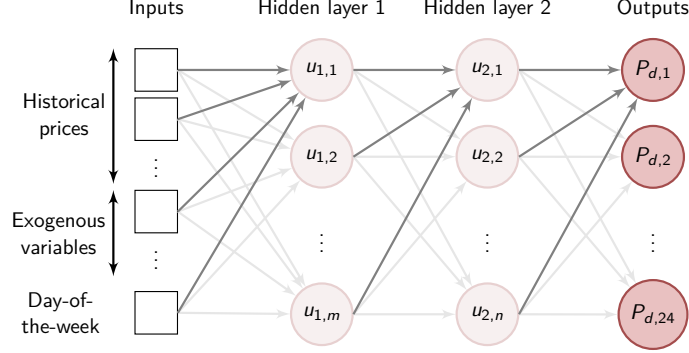


Figure 5: Visualization of the DNN model of Lago et al. (2018), i.e., a feed-forward neural network with two hidden layers and 24 outputs. Like in Figure 3, white squares represent the inputs, $u_{1,1}, \dots, u_{1,m}, u_{2,1}, \dots, u_{2,n}$ the hidden nodes, and arrows indicate the flow of information.

at the time Ziel and Weron (2018) published their paper. They simply mean that late evening prices for day $d - 1$ and particularly the last known price, i.e., $P_{d-1,24}$, are good predictors for all hours of the next day. Since then, terms like $\beta_4 P_{d-1,24}$ in Eqn. (5) have been added to expert models. Interestingly, the performance of LEAR-type models can be further improved by deseasonalizing the data with respect to the long-term seasonal component (LTSC) before estimation (Jędrzejewski et al., 2021), just like in the case of parsimonious regression (Nowotarski and Weron, 2016) and neural network models (Marcjasz et al., 2020).

4.3 Deep Learning and the DNN Model

Starting in the mid 2010s, the EPF research shifted towards models with a larger number of inputs and automatic feature engineering, like the LEAR described in Section 4.2, and architectures that employ *deep learning* (DL) to obtain better hidden data representations. Both families of models are examples of a recent trend called *data-centric* ML, where emphasis is not put on the model, but on input data quality and consistency. Both families use SL/ML methods as means to increase the number of (potential) input features and to reduce the need for human interaction during feature engineering and data processing. The difference is that the second family uses deep architectures, e.g., neural networks with more than one hidden layer (see Goodfellow et al., 2016, for an excellent introduction to DL).

Deep learning EPF models can be traced back to Wang et al. (2017), who proposed an architecture built on stacked *denoising autoencoders* that take a partially corrupted (or noisy) input and are trained to recover the original undistorted input. The DL models that followed were primarily based on the MPL with features modeled as hyperparameters. The most prominent example is probably the DNN model of Lago et al. (2018) that has been shown to improve upon parameter-rich linear regression models estimated via the LASSO.

The DNN is a feed-forward network with two hidden layers of m and n nodes, and 24 outputs, i.e., it jointly predicts 24 hourly prices $P_{d,1}, \dots, P_{d,24}$, see Fig. 5. Its hyperparameters and input features are optimized using the tree-structured Parzen estimator (Bergstra et al., 2011). This is achieved by modeling the features as hyperparameters, with each hyperparameter representing a binary variable that selects whether or not a specific feature is included in the model. Other hyperparameters include the number of neurons per layer, the activation function, the dropout rate, the learning rate, etc. In practice, the model structure can be quite large, Lago et al. (2018)

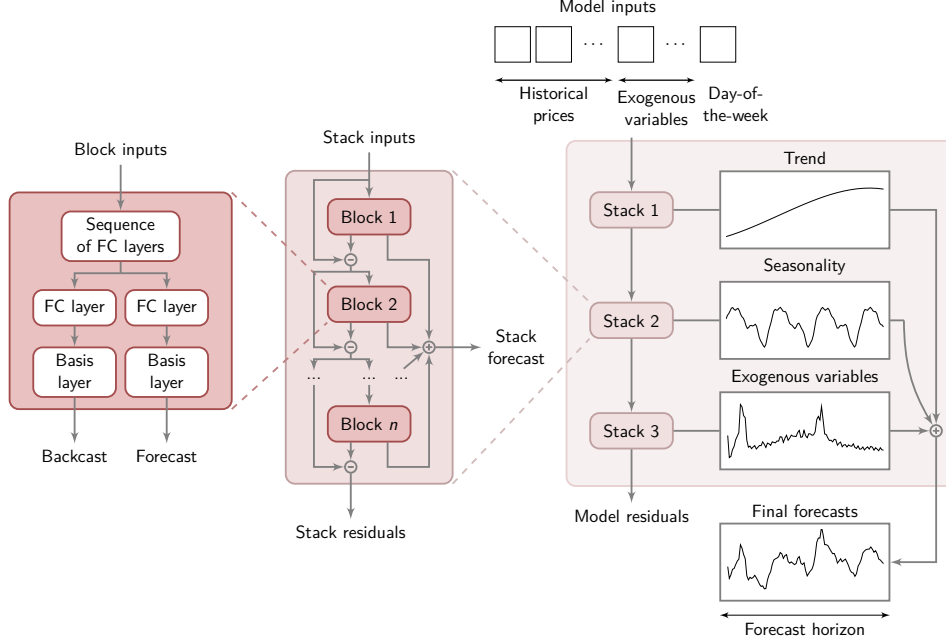


Figure 6: Visualization of the NBEATSx model of Olivares et al. (2022).

report the optimal values in their study of the Belgian market to be $m = 239$ and $n = 162$.

Given the optimal hyperparameters and features, the DNN is recalibrated on a daily basis to provide next day’s electricity price forecasts. Although not strictly mandatory, periodic (e.g., monthly) recalibration of features and hyperparameters can be beneficial. The starting set of input features is the same as for the LEAR model in Eqn. (7), with the only difference that, for the sake of simplicity, the day-of-the-week is modeled with a multi-valued variable, not a set of 7 dummies (Lago et al., 2021). The open-source Python codes for the DNN (and the LEAR) model are available from GitHub (<https://epftoolbox.readthedocs.io>).

4.4 Interpretability and the NBEATSx Model

The architecture of the *neural basis expansion analysis for time series* (NBEATS) model introduced by Oreshkin et al. (2020) has the ability to structurally decompose signals making the outputs easily interpretable. A feature whose absence has made it difficult to apply neural networks in many contexts. Moreover, it has demonstrated state-of-the-art performance on multiple large-scale datasets, including those used in the M4 competition (Makridakis et al., 2020), and it is computationally efficient exhibiting a linear cost with respect to the input size. Recently, it has been successfully applied in mid-term electricity load forecasting (Oreshkin et al., 2021).

In general, the decomposition in the NBEATS model is performed by projecting the objective time series onto basis functions in the fundamental blocks of the network structure. Each fundamental block (the dark red rectangles labeled “block 1”, ..., “block n ” in Fig. 6) consists of two parts: (i) a sequence of fully-connected layers (FC) ended with a fork that returns estimated backward and forward expansion coefficients, and (ii) the backward and forward basis layers that map these coefficients via the basis functions onto two block outputs called the backcast and the forecast. The former is the best estimate of the block inputs given the functional space used in the considered block, whereas the latter is the partial prediction that contributes to the final

forecast.

The blocks are lined up so that the backcast of each is removed from its inputs, and the residuals are passed to the following block as new inputs. Such a residual recursion is performed consecutively over all blocks in the network. The block forecasts, on the other hand, are summed up to produce the final prediction. The NBEATS architecture groups blocks into stacks that specialize in different types of basis functions. Separate stacks can account for the trend and seasonality by modelling these functions as polynomials and harmonic functions, respectively. Consequently, the final forecasts can be decomposed into interpretable components returned by individual stacks.

The NBEATSx model introduced by Olivares et al. (2022) adds to this structure a stack (light red rectangle labeled “stack 3” in Fig. 6) that performs the projection onto exogenous variables. Such an exogenous stack helps to predict the effects induced by holidays and fundamentals (like electric load or RES generation forecasts), and is crucial for EPF. While Olivares et al. report no significant differences between the NBEATS model and the *exponential smoothing recurrent neural network* (ESRNN) of Smyl (2020) that has excelled in the M4 competition, the NBEATSx architecture improves over NBEATS by nearly 20% and up to 5% over the LEAR (Sec. 4.2) and DNN (Sec. 4.3) models.

The hyperparameters and the input features are optimized in the same way as for the DNN model. However, compared to the DNN, the hyperparameter list also includes: the type and the number of stacks, the number of blocks per stack, the degree of trend polynomials, and the number of Fourier bases. The optimization algorithm also selects the best-performing order of stacks. Open-source Python codes are available from PythonRepo (<https://pythonrepo.com/repo/cchallu-nbeatsx-python-deep-learning>).

5 Trend #3: From Statistical to Economic Evaluation

Over the years a number of authors have criticized the exclusive use of statistical error measures to evaluate and compare forecasts. However, a standardized test ground/procedure for evaluating the economic impact of predictions has not been developed, not only in EPF (Hong et al., 2020), but in forecasting in general (Petropoulos et al., 2022). And this, despite the fact that already three decades ago Murphy (1993) postulated that the “goodness” of a forecast can be assessed in terms of consistency, quality, and value.

While *quality* can be readily quantified by commonly used error metrics, see Section 5.1, the other two characteristics require an explanation. As Murphy (1993) defines it, *consistency* refers to the correspondence between forecasters’ internal, i.e., recorded only in the forecaster’s mind, judgments and their forecasts. Since such judgments are, by definition, unavailable to others, consistency cannot be assessed directly. Yet, some authors explicitly mention using expert knowledge to ex-post correct the results from a statistical or a ML model. For instance, Maciejowska and Nowotarski (2016) ‘manually’ expanded or tightened the PIs in their top performing GEFCom2014 competition approach.

The third characteristic, i.e., *value*, refers to the (incremental) economic and/or other benefits to decision makers from using the predictions. For instance, it may reflect additional revenue resulting from improved forecasts or reduced uncertainty as measured by revenue volatility. As Yardley and Petropoulos (2021) argue, it is a construct that not only incorporates considerations of the utility to the forecaster, which is discussed in Section 5.2 below, but also the computational and opportunity costs. While numerous papers report them, the computational costs are rarely used to compare different methods. One of a few exceptions is an article by Nikolopoulos

and Petropoulos (2018), who study the trade-off between optimal versus suboptimal (but less costly) solutions and find that choosing the latter does not necessarily reduce forecast accuracy. Finally, the opportunity costs reflect the resources wasted on implementing a complex method that eventually is not used, because the decision-makers do not have confidence in a model they do not understand (Green and Armstrong, 2015). Yet, both cited papers do not concern EPF. Hence, in Section 5.1 we will briefly review statistical error metrics, emphasizing their pros and cons, and then – in Section 5.2 – the approaches to evaluating forecast utility, i.e. the main ingredient of the forecast value.

5.1 Statistical Error Measures

5.1.1 Point Forecasts

The most commonly used error metrics for point forecasts include the *mean absolute error* (MAE) and the *root mean squared error* (RMSE), typically across all $H = 24$ hours (48 half-hours or 96 quarter-hours) in the test period:

$$\text{MAE} = \frac{1}{DH} \sum_{d=1}^D \sum_{h=1}^H |\hat{\varepsilon}_{d,h}|, \quad \text{RMSE} = \sqrt{\frac{1}{DH} \sum_{d=1}^D \sum_{h=1}^H \hat{\varepsilon}_{d,h}^2}, \quad (8)$$

where $\hat{\varepsilon}_{d,h} = P_{d,h} - \hat{P}_{d,h}$ and D is the number of days in the test period. It is advised to report both absolute and squared errors, especially if regression and neural network models are compared. The reason is that regression-type models are typically estimated using OLS or its variants, as in Eqn. (6), while NNs are often trained by minimizing absolute errors (Lago et al., 2018; Smyl, 2020; Olivares et al., 2022).

Both MAE and RMSE are scale-dependent and hence hard to compare across different datasets. The often used in other forecasting contexts *mean absolute percentage error* (MAPE) and its “symmetric” variant (sMAPE; see, e.g., Makridakis et al., 2020) are sensitive to values close to zero and may lead to absurd results in EPF. Hyndman and Koehler (2006) advocate using the *mean absolute scaled error* (MASE) which is simply the MAE in Eqn. (8) scaled by the in-sample MAE of a naive⁶ forecast. However, the MASE is not recommended for comparisons of models using different calibration windows, since for each model it will be based on a different scaling factor. Instead, Lago et al. (2021) recommend using relative measures. For instance, the *relative MAE* (rMAE) which normalizes the MAE by the out-of-sample (not in-sample) MAE of a naive forecast.

The significance of differences in EPF accuracy is usually evaluated using the Diebold and Mariano (1995) test for (*unconditional*) *predictive ability* or its generalization – the Giacomini and White (2006) test for *conditional predictive ability*. Both tests can be used for nested and non-nested models, as long as the calibration window does not grow with the sample size, but only the latter accounts for parameter estimation uncertainty. However, energy forecasters are not restricted to these two tests, there is a plethora of available approaches (for a review see, e.g., Section 2.12.6 in Petropoulos et al., 2022).

The Diebold-Mariano (DM) test is an asymptotic z -test of the hypothesis that the mean of the loss differential series is zero. It is based upon the observation that the DM statistic:

$$\text{DM} = \sqrt{DH} \frac{\hat{\mu}}{\hat{\sigma}}, \quad (9)$$

⁶E.g., a random walk forecast. Note that for seasonal time series of period τ , the time lag should be equal to τ . For instance, in EPF it is common to take $\hat{P}_{d,h}^{\text{naive}} = P_{d-\tau,h}$.

is asymptotically standard normal under the assumption of covariance stationarity of the *loss differential* series:

$$\Delta_{d,h} = L_{1,d,h} - L_{2,d,h}, \quad (10)$$

where $L_{i,d,h}$ is the *score* or *loss function* of model i for day d and load period (e.g., hour) h , while $\hat{\mu}$ and $\hat{\sigma}$ are respectively the sample mean and standard deviation of $\Delta_{d,h}$. Covariance stationarity may not be satisfied by forecasts in day-ahead electricity markets, since the H predictions for the next day are made at the same time, using the same information set. Hence, either H independent tests (one for each load period of the day; Bordignon et al., 2013; Nowotarski et al., 2014; Uniejewski et al., 2016; Lago et al., 2018; Gianfreda et al., 2020) or a multivariate variant proposed by Ziel and Weron (2018) are performed (Uniejewski et al., 2018; Hubicka et al., 2019; Marcjasz et al., 2019; Maciejowska et al., 2021; Özen and Yildirim, 2021). The latter jointly tests forecasting accuracy across all H load periods using the ‘daily’ or ‘multivariate’ loss differential series:

$$\Delta_d = \|\varepsilon_{1,d}\|_p - \|\varepsilon_{2,d}\|_p, \quad (11)$$

where $\varepsilon_{i,d}$ is the H -dimensional vector of prediction errors of model i for day d , $\|\varepsilon_{i,d}\|_p = (\sum_{h=1}^H |\varepsilon_{i,d,h}|^p)^{1/p}$ is the p -th norm of that vector, with $p = 1$ for absolute or 2 for squared losses.

Like in the DM test, also in the Giacomini-White (GW) test the object of interest is the loss differential series – univariate or multivariate. We test the null $H_0 : \phi = 0$ in the following regression (here in the multivariate variant):

$$\Delta_d = \phi' X_{d-1} + \epsilon_d, \quad (12)$$

where X_{d-1} contains elements from the information set on day $d - 1$, i.e., a constant and lags of Δ_d , and ϵ_d is an error term. Notice that $\epsilon_d \in R$ is not the 24-dimensional vector $\varepsilon_{i,d}$ of prediction errors from Eq. (11). Sample applications of the GW test in the context of EPF include Marcjasz et al. (2018), Lago et al. (2021) and Olivares et al. (2022).

5.1.2 Probabilistic Forecasts

While defining error measures for point predictions is relatively straightforward, for probabilistic ones this becomes tricky. The problem is that we cannot observe the true price distribution F_P , only a single draw from it, i.e., the observed price $P_{d,h}$. Therefore, evaluation of probabilistic forecasts relies on so-called scoring rules and the notions of reliability, sharpness and resolution. A *scoring rule* – also, as in Eq. (10), called *score* or *loss function* – assigns a numerical score $S(\hat{F}_P, P_{d,h})$ based on the predictive distribution \hat{F}_P and the observed price. A scoring rule is (*strictly*) *proper* if it is (uniquely) optimized in expectation by the true distribution (Gneiting and Raftery, 2007). *Reliability* (also called *calibration* or *unbiasedness*) refers to the statistical consistency between \hat{F}_P and $P_{d,h}$. For instance, a 95% prediction interval (PI) is reliable if it covers exactly 95% of the observed prices. *Sharpness* refers to how concentrated is \hat{F}_P . Finally, *resolution* refers to how much the predicted density varies over time. Since sharpness and resolution are equivalent when probabilistic forecasts have perfect reliability, evaluating probabilistic predictions boils down to “maximizing sharpness subject to reliability” (Gneiting and Katzfuss, 2014; Nowotarski and Weron, 2018).

The most intuitive approach to formally check the reliability of a prediction interval to compute the empirical coverage based on the indicator series of ‘hits and misses’ defined as: $I_{d,h} = 1$ if $P_{d,h} \in \text{PI}$ and zero otherwise. EPF studies typically report the empirical coverage

itself (*PI coverage probability*, PICP) or the *average coverage error*: $\text{ACE} = \text{PICP} - \text{PINC}$, where $\text{PINC} = \alpha$ is the *PI nominal coverage*. To formally check whether $\mathbb{P}(I_{d,h} = 1) = \alpha$, i.e., the so-called *unconditional coverage* (UC), we can use the Kupiec (1995) test, which verifies whether $I_{d,h}$ is i.i.d. Bernoulli with mean α . Since the latter cannot distinguish between randomly distributed and clustered PI exceedances, Christoffersen (1998) introduced the *independence* and *conditional coverage* (CC) tests. The former is tested against a first-order Markov alternative and the latter is a joint test for independence and UC; note, that both can be run for lags larger than one (Berkowitz et al., 2011). In a continuous setting, i.e., when testing \hat{F}_P , not just selected PIs, the most common approach is to use the *Probability Integral Transform*:

$$\text{PIT}_{d,h} = \hat{F}_P(P_{d,h}), \quad (13)$$

which is independent and uniformly distributed if the distributional forecast is perfect. The PIT can be assessed visually (Nowotarski and Weron, 2018) or formally evaluated using the approach of Berkowitz (2001), which jointly tests for independence and normality, i.e., for conditional coverage.

Unlike reliability, sharpness is a property of the forecasts only – the narrower the PI or the more concentrated the predictive distribution the better. Consequently, the PI width itself is a good measure of sharpness. A more elaborate approach relies on proper scoring rules, which actually assess reliability and sharpness simultaneously (Gneiting and Katzfuss, 2014). Among them, arguably the most popular is the *pinball* loss, also known as the *linlin*, *bilinear* or *newsboy* loss (Elliott and Timmermann, 2016) and has become popular in EPF after the Global Energy Forecasting (GEFCom2014) competition (Dudek, 2016; Hong et al., 2016; Maciejowska and Nowotarski, 2016). It is defined by:

$$\text{pinball}^\alpha = \begin{cases} (1 - \alpha) (\hat{P}_{d,h}^\alpha - P_{d,h}), & \text{for } P_{d,h} < \hat{P}_{d,h}^\alpha, \\ \alpha (P_{d,h} - \hat{P}_{d,h}^\alpha), & \text{for } P_{d,h} \geq \hat{P}_{d,h}^\alpha, \end{cases} \quad (14)$$

where $\hat{P}_{d,h}^\alpha$ is the α th quantile of the predictive distribution for day d and load period (e.g., hour) h ; note, that the pinball score is the function minimized in quantile regression, see Section 3.2. The pinball can be averaged across different quantiles, e.g., 99 percentiles, and across load periods of the target day, e.g., 24 hours, to provide the *aggregate pinball score* (APS). If the grid of quantiles is arbitrarily dense, then the average converges to the *Continuous Ranked Probability Score* (Gneiting and Raftery, 2007):

$$\text{CRPS}(\hat{F}_P, P_{d,h}) = \underbrace{\mathbb{E}|\hat{P}_{d,h} - P_{d,h}|}_{\text{reliability}} - \underbrace{\frac{1}{2}\mathbb{E}|\hat{P}_{d,h} - \hat{P}_{d,h}^*|}_{\text{lack of sharpness}}, \quad (15)$$

where random variables $\hat{P}_{d,h}$ and $\hat{P}_{d,h}^*$ are two independent \hat{F}_P -distributed copies. Probabilistic forecasts can be tested for equal predictive performance using the DM and GW tests, just like point forecasts. In this case $L_{i,d,h}$ is replaced by $S_i(\hat{F}_P, P_{d,h})$ in Eq. (10). For sample EPF applications see, e.g., Serafin et al. (2019), Abramova and Bunn (2020), Marcjasz et al. (2020), Muniain and Ziel (2020) and Uniejewski and Weron (2021).

5.1.3 Path Forecasts

Compared to evaluating point or probabilistic predictions, evaluating *path* (also called *ensemble*) forecasts constitutes a challenge – it requires utilizing scoring rules for multivariate distributions

(Scheuerer and Hamill, 2015). The commonly used *Dawid-Sebastiani* and *variogram scores* are not strictly proper in the multivariate setting, while the *log-score* requires forecasts of a multivariate density, which may be not available. Hence, the recommended option is the *energy score* proposed by Gneiting and Raftery (2007), which is a generalization of the pinball and CRPS scores:

$$ES_{d,h} = \underbrace{\frac{1}{M} \sum_{i=1}^M \|P_{d,h}^i - P_{d,h}\|_2}_{\text{distance from the prices}} - \underbrace{\frac{1}{2} \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \|P_{d,h}^i - P_{d,h}^j\|_2}_{\text{distance between paths}}, \quad (16)$$

where $P_{d,h}^i$ for $i = 1, \dots, M$ is the i -th price path forecast and $\|\cdot\|_2$ is the Euclidean norm. When minimizing the energy score we are interested in minimizing the average distance between the simulated paths and the actual price trajectory and at the same time maximizing the average distance between the paths. Its use in EPF is limited, though, probably due to the much higher complexity of the problem (Munian and Ziel, 2020; Narajewski and Ziel, 2020b).

5.2 Economic Measures

There are only a handful of papers which examine the economic impact of EPF errors in a more systematic manner. Interestingly, most of these studies have been published in engineering, not economic or financial journals. The likely reason is that at least a basic knowledge is needed of how power markets, loads and generating units operate. Moreover, as mentioned earlier, there is no standardized test ground/procedure for evaluating the economic impact. Nearly every EPF study considers a different setup. We list them here with the hope of shedding light on this important, but underdeveloped topic.

5.2.1 Supply- and Demand-Side Perspectives

In one of the earlier studies, Delarue et al. (2010) take the supply-side point of view and quantify the *profit loss* that can be expected in a price based unit commitment problem, when incorrect price forecasts are used. Simulations reveal that a combined cycle gas turbine (CCGT) is much more sensitive to EPF errors (the profit can easily lie 20% below the optimal level for a perfect price forecast) than a classic coal fired unit (profit loss rarely exceeds 10%). More interestingly, negatively biased forecasts (i.e., that predict prices lower than actual) typically yield much higher losses than positively biased predictions.

On the other hand, Zareipour et al. (2010) take the demand-side perspective and consider short-term operation scheduling of two typical loads (a process industry owning on-site generation facilities and a municipal water plant with load-shifting capabilities). They introduce the *forecast inaccuracy economic impact* index: $FIEI = [\text{cost}(\hat{P}) - \text{cost}(P)] / \text{cost}(\hat{P})$, so that a positive value of FIEI indicates the percentage of the actual cost of buying electricity attributable to EPF errors. The authors report that a 1% improvement in the MAPE in forecasting accuracy would result in about 0.1%–0.35% cost reductions from short-term EPF, but also conclude that the MAPE is not a good measure.

An interesting concept is considered by Doostmohammadi et al. (2017), who compute the *financial loss/gain* (FLG) time series, defined as the difference between expected profit of a generator and the actual one. Then, based on the day-ahead forecasts of the FLG series, they propose a bidding strategy. However, by doing so, they do not work with the actual profits but with (another) estimate.

Maciejowska et al. (2019, 2021) take the perspective of a small RES utility (e.g., with one wind turbine) which has to decide where to sell 1 MW of electricity during each hour of the next day – in the day-ahead (DA) or the intraday (ID) market. Conditional on the decision, summarized by the *decision variable* based on price forecasts:

$$Y_{d,h} = \begin{cases} 1 & \text{if } \hat{P}_{d,h}^{DA} > \hat{P}_{d,h}^{ID}, \\ 0 & \text{if } \hat{P}_{d,h}^{DA} \leq \hat{P}_{d,h}^{ID}, \end{cases} \quad (17)$$

they compute the additional income over the benchmark, i.e., selling the production in the DA market, as:

$$\pi_{d,h} = Y_{d,h}P_{d,h}^{DA} + (1 - Y_{d,h})P_{d,h}^{ID} - P_{d,h}^{DA}, \quad (18)$$

where $P_{d,h}^{DA}$ and $P_{d,h}^{ID}$ are the electricity prices in the DA and ID markets, respectively. While Maciejowska et al. (2019) utilize the load forecasts published by the German and Polish system operators, Maciejowska et al. (2021) additionally improve the load forecasts for Germany by applying ARX-type models. In both papers, they measure the gains from EPF as the sum of profits in the test period, $\pi = \sum_{d=1}^D \sum_{h=1}^{24} \pi_{d,h}$, and conclude that the statistical measures of forecast accuracy – such as the percent of correct sign classifications of the price spread between the DA and ID markets – do not necessarily coincide with economic benefits.

5.2.2 Trading Strategies

Uniejewski et al. (2018) take a trading perspective (different from the supply- or demand-side point of views in Section 5.2.1) and consider a naive spot-futures trading strategy in the German market. With a perfect day-ahead forecast the buyer could always choose the lower of the two – the day-ahead price (unknown when submitting bids) or the futures price. Since this can never be achieved in reality, the authors bias (or perturb) the ‘crystal-ball’ forecast and show that a 0.20 EUR/MWh decrease in the MAE from using one model instead of another would result in ca. 90,000 EUR profits, for a 1 GW baseload in 2016.

Chitsaz et al. (2018) propose a trading strategy applicable in Ontario’s real-time electricity market. The energy storage operator maximizes profits with optimal scheduling. The schedule is set before the trading period begins, based on the available price forecasts and then it is updated at the end of each hour with a newer price forecasts. The authors conclude that such a strategy yields higher profits when using predictions generated by the proposed ARX model with features selected via the Mutual Information technique (Amjady et al., 2011) – 62% of the potential saving for ‘crystal ball’ predictions, compared with a number of other EPF approaches, e.g., using the so-called Pre-Dispatch Prices (PDPs; publicly available price predictions published by the system operator IESO) – 43% of the potential saving.

Kath and Ziel (2018) propose a multivariate elastic net model (see Section 4.2) for forecasting German quarter-hourly electricity prices. They demonstrate that the “sell in the high and buy in the low market” strategy performs well, leading to substantial benefits for both a net buyer and a net seller. On the other hand, the mean-variance approach does not bring economic benefits, but yields an optimal portfolio in terms of the *Sharpe ratio*:

$$SR = \frac{\bar{\pi}}{\sigma}, \quad (19)$$

where $\bar{\pi}$ denotes the average level of an additional revenue (i.e., $\bar{\pi} = \pi/24D$; see also Eq. (18)) and σ is the standard deviation of the time series of revenues. As such, the Sharpe ratio

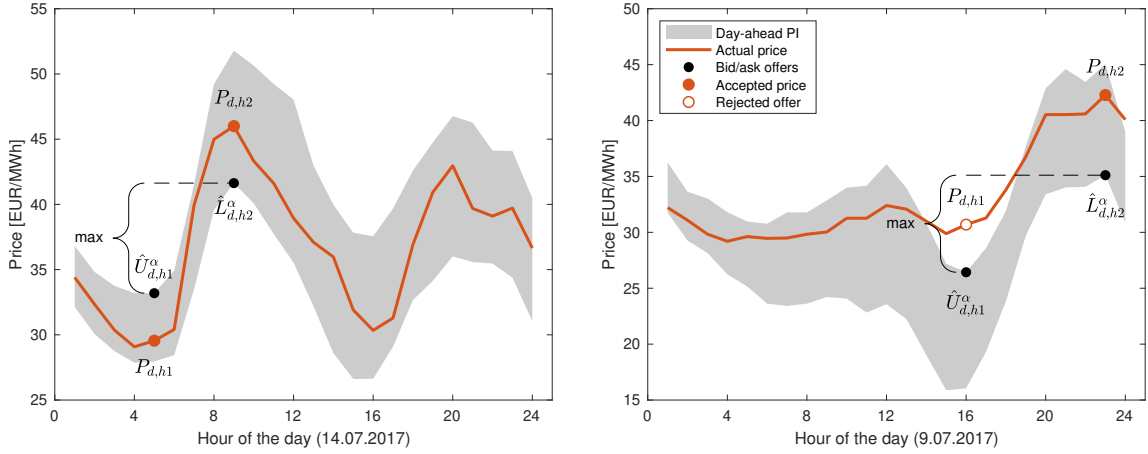


Figure 7: Illustration of the trading strategy considered by Uniejewski and Weron (2021) using German EPEX data for Friday, July 14th (*left panel*) and Sunday, July 9th (*right panel*) 2017. The day-ahead forecast of the PI is plotted in gray, the bids for the selected hours are indicated with black dots and the actual price trajectory is in orange. Note, that on July 9th the buy order is not accepted because $P_{d,h1} > \hat{U}_{d,h1}^\alpha$.

can be used to assess the trade-off between revenue and uncertainty. However, there are more performance measures (Eling and Schuhmacher, 2007; Auer, 2015), including measures based on drawdowns (e.g., Calmar ratio, Sterling ratio), based on partial moments (e.g., omega ratio, Sortino ratio) and based on the Value-at-Risk (VaR; e.g., excess return on VaR, conditional Sharpe ratio). Whether they will turn out to be useful in the EPF context remains yet to be checked.

Uniejewski and Weron (2021) propose a strategy for market participants having access to storage capacity. They consider a realistic setup, inspired by the Virtual Power Plant analyzed in Sikorski et al. (2019), in which the company owns a 1.25 MW battery with an efficiency of 80% per charge and discharge cycle, that cannot be discharged below 20% of the nominal capacity (i.e., 0.25 MW) due to technical limitations. The strategy is straightforward: each day buy energy and charge the battery when the price is low (generally in the early morning hours) and discharge and sell when the price is high (generally in the afternoon hours). Using probabilistic forecasts of the DA prices in the Polish market, the authors determine both the time (buy $h1$ and sell $h2$ hours) and the prices of limit orders submitted to the power exchange. They formulate and solve the following maximization problem:

$$\max_{\{h1, h2\}} \left(0.8 \hat{L}_{d,h2}^\alpha - \hat{U}_{d,h1}^\alpha \right) \quad \text{subject to} \quad h1 < h2. \quad (20)$$

The optimizer selects the lowest price of a given day based on the upper quantile forecast $\hat{U}_{d,h1}^\alpha$ and the highest price based on the lower quantile forecast $\hat{L}_{d,h2}^\alpha$. The company then submits the bid to buy 1 MW for $\hat{U}_{d,h1}^\alpha$ at hour $h1$ and simultaneously the offer to sell 0.8 MW at $\hat{L}_{d,h2}^\alpha$ at hour $h2$; two sample solutions for the German EPEX market are depicted in Fig. 7. If both offers are accepted in the day-ahead market, as in the left panel of Fig. 7, the profit for a given day equals $0.8P_{d,h2} - P_{d,h1}$. However, the probability of each offer to be accepted in the market is equal to $\frac{1-\alpha}{2}$. If one of them is rejected, as in the right panel of Fig. 7 for hour $h1$, the energy has to be bought or sold in the balancing market.

This strategy is further modified in Uniejewski and Weron (2022). Now, all trading is exclusively executed in the day-ahead market. To do so, a twice larger energy storage capacity (2.5 MW) is required to trade the same volume (0.8-1 MW). The idea is to always remain in an intermediate state of the battery, for which both charging and discharging 1 MW is possible. When the bid or the ask is rejected, the authors propose to close the position on the next day by submitting a market order (i.e., with no price limit). They conclude that the Smoothing Quantile Regression Averaging (SQRA) approach they propose outperforms the benchmarks in terms of statistical error metrics (Kupiec test, GW test for the pinball score) in all four considered markets (German EPEX, Scandinavian Nord Pool, Iberian OMIE, North American PJM). However, when the trading strategy is executed, SQRA forecasts lead to higher profits only in two markets (EPEX, PJM). The authors hypothesize that the poor performance for NP and OMIE is due to a twice lower average intraday price spread, i.e., the gap between the maximum and the minimum hourly price for a given day.

6 Further Reading

The first review, published when electricity markets were still in their infancy:

- Bunn, D., 2000. Forecasting loads and prices in competitive power markets. *Proceedings of the IEEE*, 88(2), 163-169.

The first comprehensive EPF review, postulating the need for objective comparative studies and speculating on the future research directions:

- Weron, R., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* 30, 1030-1081. [Open access]

Two thorough treatments of probabilistic EPF, presenting much needed guidelines for the rigorous use of methods, measures and tests, in line with the paradigm of maximizing sharpness subject to reliability:

- Nowotarski, J., Weron, R., 2018. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews* 81, 1548-1568.
- Ziel, F., Steinert, R., 2018. Probabilistic mid- and long-term electricity price forecasting. *Renewable and Sustainable Energy Reviews* 94, 251-266.

A review of energy (load, price, wind and solar generation) forecasting, with a discussion of two challenging problems that deserve rigorous investigation – close-loop forecasting and (economic) valuation of forecasts:

- Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., Zareipour, H., 2020. Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy* 7, 376-388. [Open access]

A recent review with a set of guidelines/best practices for EPF, introducing the `epftoolbox`⁷ with Python codes for two highly competitive benchmark models (LEAR, DNN):

- Lago, J., Marcjasz, G., De Schutter, B., Weron, R., 2021. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy* 293, 116983. [Open access]

⁷Freely available for download from: <https://epftoolbox.readthedocs.io/en/latest>.

A popular science article on the evolution of machine learning models in EPF:

- Jędrzejewski, A., Lago, J., Marcjasz, G., Weron, R., 2022. Electricity price forecasting: The dawn of machine learning. *IEEE Power & Energy Magazine* 20(3), 24-31.

References

- Abramova, E., Bunn, D., 2020. Forecasting the intra-day spread densities of electricity prices. *Energies* 13, 687.
- Afanasyev, D., Fedorova, E., 2019. On the impact of outlier filtering on the electricity price forecasting accuracy. *Applied Energy* 236, 196–210.
- Amjady, N., Keynia, F., Zareipour, H., 2011. Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization. *IEEE Transactions on Sustainable Energy* 2, 265–276.
- Auer, B., 2015. Does the choice of performance measure influence the evaluation of commodity investments? *International Review of Financial Analysis* 38, 142–150.
- Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization, in: *Advances in Neural Information Processing Systems*, pp. 2546–2554.
- Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics* 19, 465–474.
- Berkowitz, J., Christoffersen, P., Pelletier, D., 2011. Evaluating value-at-risk models with desk-level data. *Management Science* 57, 2213–2227.
- Bierbrauer, M., Menn, C., Rachev, S.T., Trück, S., 2007. Spot and derivative pricing in the EEX power market. *Journal of Banking & Finance* 31, 3462–3485.
- Bordignon, S., Bunn, D.W., Lisi, F., Nan, F., 2013. Combining day-ahead forecasts for British electricity prices. *Energy Economics* 35, 88–103.
- Bunn, D., Andresen, A., Chen, D., Westgaard, S., 2016. Analysis and forecasting of electricity price risks with quantile factor models. *Energy Journal* 37, 101–122.
- Bunn, D.W., 2000. Forecasting loads and prices in competitive power markets. *Proceedings of the IEEE* 88, 163–169.
- Chen, X., Dong, Z., Meng, K., Xu, Y., Wong, K., Ngan, H., 2012. Electricity price forecasting with extreme learning machine and bootstrapping. *IEEE Transactions on Power Systems* 27, 2055–2062.
- Chitsaz, H., Zamani-Dehkordi, P., Zareipour, H., Parikh, P.P., 2018. Electricity price forecasting for operational scheduling of behind-the-meter storage systems. *IEEE Transactions on Smart Grid* 9, 6612–6622.
- Christoffersen, P., 1998. Evaluating interval forecasts. *International Economic Review* 39, 841–862.
- Contreras, J., Espínola, R., Nogales, F., Conejo, A., 2003. ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems* 18, 1014–1020.
- Delarue, E., Van Den Bosch, P., D’haeseleer, W., 2010. Effect of the accuracy of price forecasting on profit in a price based unit commitment. *Electric Power Systems Research* 80, 1306–1313.
- Diaz, G., Planas, E., 2016. A note on the normalization of Spanish electricity spot prices. *IEEE Transactions on Power Systems* 31, 2499–2500.
- Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–263.
- Doostmohammadi, A., Amjady, N., Zareipour, H., 2017. Day-ahead financial loss/gain modeling and prediction for a generation company. *IEEE Transactions on Power Systems* 32, 3360–3372.
- Dudek, G., 2016. Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting. *International Journal of Forecasting* 32, 1057–1060.
- Eling, M., Schuhmacher, F., 2007. Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking and Finance* 31, 2632–2647.
- Elliott, G., Timmermann, A., 2016. *Economic Forecasting*. Princeton University Press.

- Gaillard, P., Goude, Y., Nedellec, R., 2016. Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *International Journal of Forecasting* 32, 1038–1050.
- Giacomini, R., White, H., 2006. Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.
- Gianfreda, A., Ravazzolo, F., Rossini, L., 2020. Comparing the forecasting performances of linear models for electricity prices with high RES penetration. *International Journal of Forecasting* 36, 974–986.
- Gneiting, T., Katzfuss, M., 2014. Probabilistic forecasting. *Annual Review of Statistics and Its Application* 1, 125–151.
- Gneiting, T., Raftery, A., 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102, 359–378.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press. Freely available from <http://www.deeplearningbook.org/>.
- Green, K., Armstrong, J., 2015. Simple versus complex forecasting: The evidence. *Journal of Business Research* 68, 1678–1685.
- Grossi, L., Nan, F., 2019. Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources. *Technological Forecasting and Social Change* 141, 305–318.
- Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67.
- Hong, T., 2015. Crystal ball lessons in predictive analytics. *EnergyBiz*, Spring , 35–37.
- Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R.J., 2016. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting* 32, 896–913.
- Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., Zareipour, H., 2020. Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy* 7, 376–388.
- Hubicka, K., Marcjasz, G., Weron, R., 2019. A note on averaging day-ahead electricity price forecasts across calibration windows. *IEEE Transactions on Sustainable Energy* 10, 321–323.
- Hyndman, R., Koehler, A., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22, 679–688.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. *An Introduction to Statistical Learning with Applications in R* (2nd ed.). Springer, New York.
- Janczura, J., Trück, S., Weron, R., Wolff, R., 2013. Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics* 38, 96–110.
- Janke, T., Steinke, F., 2019. Forecasting the price distribution of continuous intraday electricity trading. *Energies* 12, 4262.
- Janke, T., Steinke, F., 2020. Probabilistic multivariate electricity price forecasting using implicit generative ensemble post-processing, in: 2020 International Conference on Probabilistic Methods Applied to Power Systems, PMAAPS 2020 - Proceedings, p. 9183687.
- Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., Callot, L., 2020. Criteria for classifying forecasting methods. *International Journal of Forecasting* 36, 167 – 177.
- Jędrzejewski, A., Lago, J., Marcjasz, G., Weron, R., 2022. Electricity price forecasting: The dawn of machine learning. *IEEE Power & Energy Magazine* 20, 24–31.
- Jędrzejewski, A., Marcjasz, G., Weron, R., 2021. Importance of the long-term seasonal component in day-ahead electricity price forecasting revisited: Parameter-rich models estimated via the lasso. *Energies* 14, 3249.
- Kath, C., Ziel, F., 2018. The value of forecasts: Quantifying the economic gains of accurate quarter-hourly electricity price forecasts. *Energy Economics* 76, 411–423.
- Kath, C., Ziel, F., 2021. Conformal prediction interval estimation and applications to day-ahead and intraday power markets. *International Journal of Forecasting* 37, 777–799.
- Keles, D., Scelle, J., Paraschiv, F., Fichtner, W., 2016. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Applied Energy* 162, 218–230.
- Kiesel, R., Paraschiv, F., 2017. Econometric analysis of 15-minute intraday electricity prices. *Energy Economics* 64, 77–90.
- Koenker, R.W., 2005. *Quantile Regression*. Cambridge University Press.

- Kostrzewski, M., Kostrzewska, J., 2019. Probabilistic electricity price forecasting with Bayesian stochastic volatility models. *Energy Economics* 80, 610–620.
- Kramer, A., Kiesel, R., 2021. Exogenous factors for order arrivals on the intraday electricity market. *Energy Economics* 97, 105186.
- Kulakov, S., Ziel, F., 2021. The impact of renewable energy forecasts on intraday electricity prices. *Economics of Energy and Environmental Policy* 10, 79–104.
- Kupiec, P.H., 1995. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* 3, 73–84.
- Lago, J., De Ridder, F., De Schutter, B., 2018. Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy* 221, 386–405.
- Lago, J., Marcjasz, G., De Schutter, B., Weron, R., 2021. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy* 293, 116983.
- Lisi, F., Pelagatti, M., 2018. Component estimation for electricity market data: Deterministic or stochastic? *Energy Economics* 74, 13–37.
- Liu, B., Nowotarski, J., Hong, T., Weron, R., 2017. Probabilistic load forecasting via Quantile Regression Averaging on sister forecasts. *IEEE Transactions on Smart Grid* 8, 730–737.
- Ludwig, N., Feuerriegel, S., Neumann, D., 2015. Putting big data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. *Journal of Decision Systems* 24, 19–36.
- Maciejowska, K., 2020. Assessing the impact of renewable energy sources on the electricity price level and variability – a quantile regression approach. *Energy Economics* 85, 104532.
- Maciejowska, K., Nitka, W., Weron, T., 2019. Day-ahead vs. intraday – forecasting the price spread to maximize economic benefits. *Energies* 12, 631.
- Maciejowska, K., Nitka, W., Weron, T., 2021. Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices. *Energy Economics* 99, 105273.
- Maciejowska, K., Nowotarski, J., 2016. A hybrid model for GEFCom2014 probabilistic electricity price forecasting. *International Journal of Forecasting* 32, 1051–1056.
- Maciejowska, K., Uniejewski, B., Serafin, T., 2020. PCA forecast averaging – predicting day-ahead and intraday electricity prices. *Energies* 13, 3530.
- Maciejowska, K., Weron, R., 2016. Short- and mid-term forecasting of baseload electricity prices in the U.K.: The impact of intra-day price relationships and market fundamentals. *IEEE Transactions on Power Systems* 31, 994–1005.
- Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2020. The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting* 36, 54–74.
- Marcjasz, G., 2020. Forecasting electricity prices using deep neural networks: A robust hyper-parameter selection scheme. *Energies* 13, 13184605.
- Marcjasz, G., Serafin, T., Weron, R., 2018. Selection of calibration windows for day-ahead electricity price forecasting. *Energies* 11, 2364.
- Marcjasz, G., Uniejewski, B., Weron, R., 2019. On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks. *International Journal of Forecasting* 35, 1520–1532.
- Marcjasz, G., Uniejewski, B., Weron, R., 2020. Probabilistic electricity price forecasting with NARX networks: Combine point or probabilistic forecasts? *International Journal of Forecasting* 36, 466–479.
- Mashlakov, A., Kuronen, T., Lensu, L., Kaarna, A., Honkapuro, S., 2021. Assessing the performance of deep learning models for multivariate probabilistic energy forecasting. *Applied Energy* 285, 116405.
- Mayer, K., Trück, S., 2018. Electricity markets around the world. *Journal of Commodity Markets* 9, 77–100.
- Misiorek, A., Trück, S., Weron, R., 2006. Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. *Studies in Nonlinear Dynamics & Econometrics* 10, Article 2.
- Mitchell, T., 1997. *Machine Learning*. McGraw Hill, New York.
- Muniai, P., Ziel, F., 2020. Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices. *International Journal of Forecasting* 36, 1193–1210.

- Murphy, A., 1993. What is a good forecast? an essay on the nature of goodness in weather forecasting. *Weather & Forecasting* 8, 281–293.
- Nan, F., 2009. Forecasting next-day electricity prices: From different models to combination. URL: <http://paduaresearch.cab.unipd.it/2147>. PhD Thesis, Universita degli Studi di Padova, Italy.
- Narajewski, M., Ziel, F., 2020a. Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets* 19, 100107.
- Narajewski, M., Ziel, F., 2020b. Ensemble forecasting for intraday electricity prices: Simulating trajectories. *Applied Energy* 279, 115801.
- Nikolopoulos, K., Petropoulos, F., 2018. Forecasting for big data: Does suboptimality matter? *Computers and Operations Research* 98, 322–329.
- Nowotarski, J., Raviv, E., Trück, S., Weron, R., 2014. An empirical comparison of alternate schemes for combining electricity spot price forecasts. *Energy Economics* 46, 395–412.
- Nowotarski, J., Weron, R., 2015. Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Computational Statistics* 30, 791–803.
- Nowotarski, J., Weron, R., 2016. On the importance of the long-term seasonal component in day-ahead electricity price forecasting. *Energy Economics* 57, 228–235.
- Nowotarski, J., Weron, R., 2018. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews* 81, 1548–1568.
- Olivares, K.G., Challu, C., Marcjasz, G., Weron, R., Dubrawski, A., 2022. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx. *International Journal of Forecasting*, forthcoming.
- Oreshkin, B., Carпов, D., Chapados, N., Bengio, Y., 2020. N-BEATS: neural basis expansion analysis for interpretable time series forecasting, in: 8th International Conference on Learning Representations, ICLR 2020.
- Oreshkin, B., Dudek, G., Pelka, P., Turkina, E., 2021. N-BEATS neural network for mid-term electricity load forecasting. *Applied Energy* 293, 116918.
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., *et al.*, 2022. Forecasting: Theory and practice. *International Journal of Forecasting* (doi: 10.1016/j.ijforecast.2021.11.001).
- Scheuerer, M., Hamill, T.M., 2015. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review* 143, 1321–1334.
- Schneider, S., 2011. Power spot price models with negative prices. *Journal of Energy Markets* 4, 77–102.
- Serafin, T., Uniejewski, B., Weron, R., 2019. Averaging predictive distributions across calibration windows for day-ahead electricity price forecasting. *Energies* 12, 256.
- Shi, W., Wang, Y., Chen, Y., Ma, J., 2021. An effective two-stage electricity price forecasting scheme. *Electric Power Systems Research* 199, 107416.
- Sikorski, T., Jasinski, M., Ropuszynska-Surma, E., Weglarz, M., Kaczorowska, D., Kostyla, P., Leonowicz, Z., Lis, R., Rezmer, J., Rojewski, W., Sobierajski, M., Szymanda, J., Bejmert, D., Janik, P., 2019. A case study on distributed energy resources and energy-storage systems in a virtual power plant concept: Economic aspects. *Energies* 12, 4447.
- Smyl, S., 2020. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting* 36, 75–85.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B* 58, 267–288.
- Uniejewski, B., Marcjasz, G., Weron, R., 2019a. On the importance of the long-term seasonal component in day-ahead electricity price forecasting: Part II – Probabilistic forecasting. *Energy Economics* 79, 171–182.
- Uniejewski, B., Marcjasz, G., Weron, R., 2019b. Understanding intraday electricity markets: Variable selection and very short-term price forecasting using lasso. *International Journal of Forecasting* 35, 1533–1547.
- Uniejewski, B., Nowotarski, J., Weron, R., 2016. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies* 9, 621.
- Uniejewski, B., Weron, R., 2018. Efficient forecasting of electricity spot prices with expert and LASSO models. *Energies* 11, 2039.

- Uniejewski, B., Weron, R., 2021. Regularized quantile regression averaging for probabilistic electricity price forecasting. *Energy Economics* 95, 105121.
- Uniejewski, B., Weron, R., 2022. Smoothing quantile regression averaging for probabilistic electricity price forecasting. *Energy Economics* (submitted).
- Uniejewski, B., Weron, R., Ziel, F., 2018. Variance stabilizing transformations for electricity spot price forecasting. *IEEE Transactions on Power Systems* 33, 2219–2229.
- Wang, L., Zhang, Z., Chen, J., 2017. Short-term electricity price forecasting with stacked denoising autoencoders. *IEEE Transactions on Power Systems* 32, 2673–2681.
- Weron, R., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* 30, 1030–1081.
- Weron, R., Ziel, F., 2020. Electricity price forecasting, in: Soytas, U., Sari, R. (Eds.), *Handbook of Energy Economics*. Routledge, pp. 506–521.
- Yardley, E., Petropoulos, F., 2021. Beyond error measures to the utility and cost of the forecasts. *Foresight* Q4, 36–45.
- Zareipour, H., Canizares, C., Bhattacharya, K., 2010. Economic impact of electricity market price forecasting errors: A demand-side analysis. *IEEE Transactions on Power Systems* 25, 254–262.
- Zhang, R., Li, G., Ma, Z., 2020. A deep learning based hybrid framework for day-ahead electricity price forecasting. *IEEE Access* 8, 143423–143436.
- Zhou, Y., Scheller-Wolf, A., Secomandi, N., Smith, S., 2016. Electricity trading and negative prices: Storage vs. disposal. *Management Science* 62, 880–898.
- Ziel, F., 2016. Forecasting electricity spot prices using LASSO: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems* 31, 4977–4987.
- Ziel, F., Steinert, R., 2016. Electricity price forecasting using sale and purchase curves: The X-model. *Energy Economics* 59, 435–454.
- Ziel, F., Steinert, R., 2018. Probabilistic mid- and long-term electricity price forecasting. *Renewable and Sustainable Energy Reviews* 94, 251–266.
- Ziel, F., Steinert, R., Husmann, S., 2015. Efficient modeling and forecasting of electricity spot prices. *Energy Economics* 47, 89–111.
- Ziel, F., Weron, R., 2018. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics* 70, 396–420.
- Zou, H., Hastie, T., 2015. Regularization and variable selection via the Elastic Nets. *Journal of the Royal Statistical Society B* 67, 301–320.
- Özen, K., Yıldırım, D., 2021. Application of bagging in day-ahead electricity price forecasting and factor augmentation. *Energy Economics* 103, 105573.