HSC Research Report

HSC/18/08

Electricity price forecasting

Rafał Weron[1]
Florian Ziel[2]

[1] Department of Operations Research, Faculty of Computer Science and Management, Wrocław University of Science and Technology, Poland
[2] House of Energy Markets and Finance, Universität Duisburg-Essen, Campus Essen, Germany

Hugo Steinhaus Center
Wrocław University of Science and Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
http://www.im.pwr.wroc.pl/~hugo/

To appear in: ROUTLEDGE HANDBOOK OF ENERGY ECONOMICS, U.Soytas & R.Sari (eds.), 2018

Electricity price forecasting*

Rafał Weron[a,†] and Florian Ziel[b,‡]

[a]Department of Operations Research, Faculty of Computer Science and Management, Wrocław University of Science and Technology, Wrocław, Poland
[b]House of Energy Markets and Finance, Universität Duisburg-Essen, Campus Essen, Germany

ARTICLE HISTORY
Compiled September 9, 2018

Contents

1 Introduction **2**

2 Horizons **2**
2.1 Intraday and day-ahead . . . 2
2.2 Medium-term . . . 3
2.3 Long-term . . . 4

3 Types of forecasts **4**
3.1 Point . . . 4
3.2 Probabilistic . . . 5
3.3 Ensemble . . . 6

4 Models **6**
4.1 Statistical . . . 6
4.2 Computational intelligence . . . 8
4.3 Reduced-form . . . 8
4.4 Fundamental . . . 9
4.5 Multi-agent . . . 9

5 Forecast evaluation **10**
5.1 Point forecasts . . . 10
5.2 Probabilistic forecasts . . . 11
5.3 Ensemble forecasts . . . 13
5.4 Testing for equal predictive performance . . . 13

References **14**

*Partially supported by the German Research Foundation (DFG, Germany) and the National Science Center (NCN, Poland) through BEETHOVEN grant no. 2016/23/G/HS4/01005.

†E-mail: rafal.weron@pwr.edu.pl
‡E-mail: florian.ziel@uni-due.de

1. Introduction

Electricity price forecasting (EPF) is a branch of energy forecasting on the interface between econometrics/statistics and engineering, which focuses on predicting the spot and forward prices in wholesale electricity markets. Its beginnings can be traced back to the early 1990s, when power sector deregulation led to the introduction of competitive markets in the UK and Scandinavia. The changes quickly spread throughout Europe and North America, and nowadays – in many countries worldwide – electricity is traded under market rules using spot and derivative contracts (Joskow, 2008; Mayer and Trück, 2018).

When modeling and forecasting electricity prices we have to bear in mind that electricity is a very special commodity – it is economically non-storable (hence requires a constant balance between production and consumption) and is dependent on weather (both on the demand and supply sides) and the intensity of business activities (on-peak vs. off-peak hours). The resulting spot price dynamics exhibits seasonality at the daily, weekly and annual levels, and abrupt, short-lived and generally unanticipated price spikes. The forward prices are less volatile due to averaging over weekly, monthly, quarterly or even annual delivery periods.

Over the last 25 years, a variety of methods and ideas have been tried for EPF, with varying degrees of success (Gürtler and Paulsen, 2018; Nowotarski and Weron, 2018; Weron, 2014; Ziel and Steinert, 2018). In this chapter we first briefly discuss the forecasting horizons (Section 2) and the types of forecasts (Section 3), then review the forecasting tools (Section 4) and the evaluation techniques (Section 5) used in the EPF literature. Note, that we use EPF as the abbreviation for both *electricity price forecasting* and *electricity price forecast*. The plural form, i.e., forecasts, is abbreviated EPFs.

2. Horizons

In the EPF literature, it is customary to talk about short-, medium- and long-term predictions, even though there are no strict or even commonly accepted definitions. *Short-term* horizons range from minutes to days; sometimes horizons of less than an hour are referred to as *very short-term*. As a thumb rule, short-term corresponds to horizons for which reliable meteorological forecasts for temperature, wind speed, cloud cover, etc., are available. Short-term predictions are mainly relevant for market operations and system stability Weron (2014).

Medium-term forecasting covers horizons beyond reliable meteorological predictions, but without major impact of political and technological uncertainty, with lead times measured in weeks, months, quarters or even years. The practical relevance arises mainly from maintenance scheduling, resources reallocation, bilateral contracting, derivatives valuation, risk management and budgeting. Finally, *long-term* horizons refer to everything from a few years up to several decades. Such far-reaching forecasts are performed to answer questions about investment planning and policy making (Ziel and Steinert, 2018).

2.1. Intraday and day-ahead

Unlike most other commodity or financial markets, the electricity 'spot market' is typically a *day-ahead* market that does not allow for continuous trading. Agents submit their bids and offers for delivery of electricity during each hour (or blocks of hours) of the next day before a certain market closing time (often called *gate closure*). Then, prices for all *load periods* of the next day are determined at the same time during a uniform-price auction (Weron, 2014).

In electricity markets with zonal pricing (as in Europe), next to the day-ahead markets, so-called *intraday* markets exist. These markets start operating more or less after the day-ahead auction results are announced and run until a few minutes before delivery. Their main purpose is to balance deviations resulting from positions in day-ahead contracts and unexpected changes in demand or generation (Gianfreda et al., 2016; Zaleski and Klimczak, 2015). In most European countries intraday markets operate as continuous trading sessions (e.g., Germany, France, Poland, UK), but in some they are organized in the form of multiple consecutive auctions (e.g., Italy, Spain). However, in the very few intraday EPF publications that exist (Kiesel and Paraschiv, 2017; Wolff and Feuerriegel, 2017; Ziel, 2017), even when markets with continuous trading are considered only an aggregate characteristic, such as the volume weighted intraday price over a certain trading period, is predicted.

Finally, in basically all electricty markets there exist so-called *balancing* (or *real-time*) markets, which are operated by *transmission system operators* (TSOs). These technical markets operate from a few hours before delivery until the delivery itself. Their only purpose is to guarantee system stability Weron (2006). Usually these markets exhibit high price fluctuations, which results in the demand for sophisticated forecasting models. As a consequence, the literature on predicting balancing prices is rather sparse (Klæboe et al., 2015).

2.2. Medium-term

For medium-term horizons, the value of meteorological predictions as well as their impact on EPFs disappears. Still, deterministic demand patterns (e.g., reduced electricity consumption during holiday periods) impact the price dynamics as in the short-term (Ziel et al., 2015). On the other hand, structural variables become important, most notably prices of fuels used for conventional power generation (Maciejowska and Weron, 2016; Mohamed and El-Hawary, 2016). These mainly include natural gas, oil, coal, lignite and uranium, but also CO_2 permits in markets with carbon pricing schemes (like the European Union). Moreover, plans for changes in the power plant portfolio, e.g., maintenance periods, retirement of old power plants or installation of new capacity, become relevant and can be incorporated into mid-term models. The same holds for grid expansions. For instance, when interconnectors between small market zones are implemented, they can have a drastic influence on price dynamics (Ries et al., 2016). Also, prices or complete order books of electricity derivatives can help to forecast spot prices (Steinert and Ziel, 2018).

The EPF literature suggests that a suitable forecasting setup should either be based on statistical models with a substantial amount of exogenous variables, see Section 4.1, or on structural models, see Section 4.4 (Bello et al., 2016; Weron, 2014; Ziel and Steinert, 2018). Among experts it this field, there is a clear majority opinion that the longer the forecasting horizon the better the performance of structural models compared to statistical approaches. Unfortunately, so far there is no research study available which would support this point of view.

2.3. *Long-term*

For long-term forecasting horizons political/regulatory, technological, economical and social risks come into play. All these components are hard to predict on their own, which makes incorporating them into electricity price models a laborious and an almost impossible task. Complex models must be developed, large teams of experts from different disciplines may be needed and the evaluation becomes tricky, as we require decades of data (not only for electricity prices, but also for all exogenous variables). Consequently, the literature on long-term EPF is almost nonexistent (Ziel and Steinert, 2018). Yet, there are papers which look at electricity prices far into the future (e.g., Lund et al., 2018). These studies consider usually only scenarios or projections, in the sense that they describe future paths conditioned on certain external factors, like political decisions. However, the paths are not associated with probabilities, which would be required for referring to them as forecasts.

3. Types of forecasts

The vast majority of EPF papers are concerned only with point forecasts (Weron, 2014). However, after a decade of limited interest, probabilistic EPF gained momentum with the Global Energy Forecasting Competition (GEFCom2014; Hong et al., 2016), which focused solely on probabilistic energy (load, price, wind and solar) forecasting and used the pinball loss (7) as the only evaluation criterion. Still, probabilistic EPF is not that common. In the last three years only ca. 15% of Scopus-indexed EPF articles concerned interval or distributional predictions (Nowotarski and Weron, 2018; Ziel and Steinert, 2018). Ensemble forecasts, which in contrast to point and probabilistic predictions do not focus on only one point at a time, are even less popular. This is surprising, given that path-dependency is crucial for many optimization problems arising in power plant scheduling, energy storage and trading. But we expect more ensemble EPF papers to be published in the coming years.

3.1. *Point*

The day-ahead price series typically is a result of a conducted once per day (usually around noon) auction, during which all hourly prices for the next day are disclosed at once (Weron, 2014). In the intraday markets that nowadays operate in quite a few countries worldwide, the load periods can be shorter than one hour, e.g., the European Power Exchange (EPEX) also trades half- and quarter-hourly products (Kiesel and Paraschiv, 2017). In either case the day can be divided into a finite number of load periods $h = 1, \ldots, \mathcal{H}$. Hence it is natural to denote by $P_{d,h}$ the price for day d and load period h. A point forecast of $P_{d,h}$, denoted by $\widehat{P}_{d,h}$, is usually understood in the EPF literature – without explicitly saying it – as the expected value (or mean) of the price random variable, i.e., $\mathbb{E}(P_{d,h})$, see the top left panel in Fig. 1. This notion can be easily extended to *quantile* forecasts, which are the building blocks for probabilistic predictions. For instance, $\widehat{q}_{0.5,P}$ could be a point forecast for the median $q_{0.5,P}$ of $P_{d,h}$; recall that the median is the 50% or 0.5-quantile. While the pair $[\widehat{q}_{0.25,P}, \widehat{q}_{0.75,P}]$ yields the interquartile range, i.e., a symmetric 50% prediction interval, illustrated in the top right panel of Fig. 1.

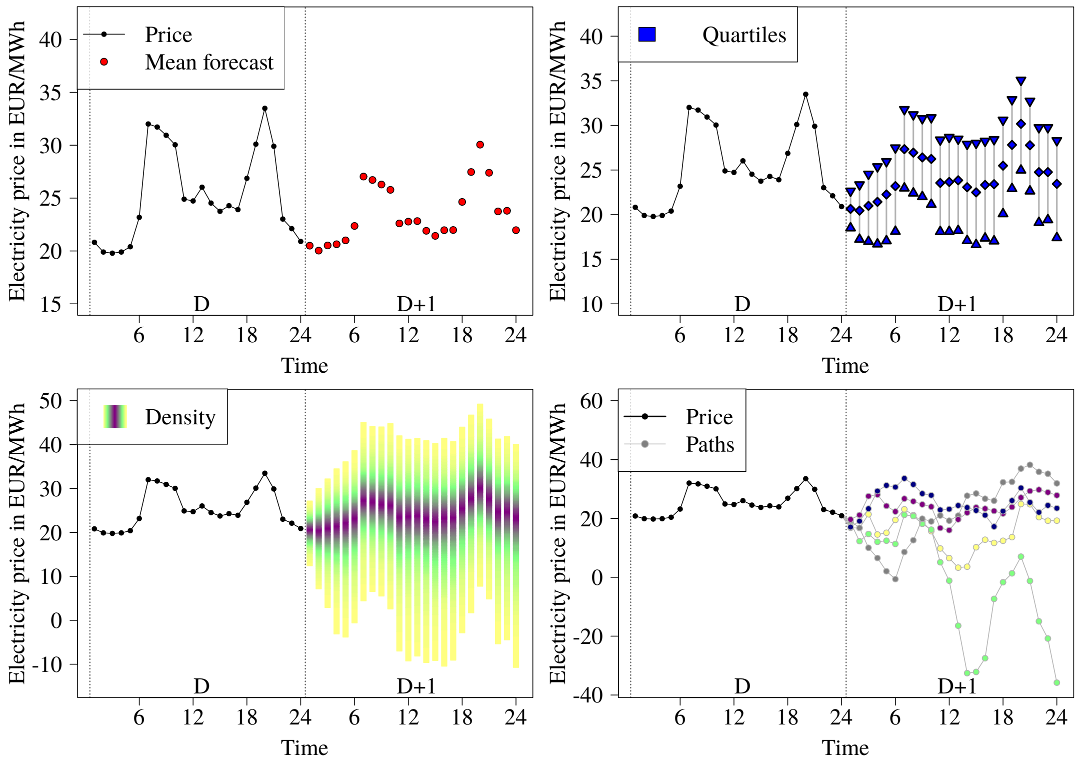

Figure 1. Illustration of the three types of forecasts discussed in Section 3: the mean as the most commonly used *point* prediction (*top left*), quartiles of the predictive distribution (*top right*) and a density forecast (*bottom left*) as examples of *probabilistic* forecasts, and 5 simulated trajectories that form an *ensemble* forecast (*bottom right*).

3.2. Probabilistic

There are two main approaches to probabilistic forecasting. The more popular one builds on the point forecast and the distribution of errors associated with it. The other directly considers the distribution of the spot price and is, for instance, utilized in Quantile Regression Averaging (QRA; see Nowotarski and Weron, 2015). In both cases the focus can be on prediction intervals, selected quantiles – quartiles (as in the top left panel in Fig. 1), deciles, percentiles – or the whole predictive distribution (as in the bottom right panel in Fig. 1).

If we consider the mean price at a future date, i.e, $\widehat{P}_{d,h} = \mathbb{E}(P_{d,h})$, as the 'point forecast' then we can write: $P_{d,h} = \widehat{P}_{d,h} + \varepsilon_{d,h}$, which implies:

$$F_P(x) = F_\varepsilon(x - \widehat{P}_{d,h}), \tag{1}$$

where F_ε is the distribution of errors associated with $\widehat{P}_{d,h}$. This means that the distribution of errors has an identical shape to the distribution of prices, only is shifted by $\widehat{P}_{d,h}$ to the left on the horizontal axis. The corresponding quantile function $\widehat{q}_{\alpha,\varepsilon}$ is also shifted with respect to $\widehat{q}_{\alpha,P}$, but now on the vertical axis and down:

$$\widehat{q}_{\alpha,P} - \widehat{P}_{d,h} = \widehat{q}_{\alpha,\varepsilon}, \tag{2}$$

Equivalently in terms of the inverse empirical *cumulative distribution function* (CDF) we can write:

$$\widehat{F}_P^{-1}(\alpha) = \widehat{P}_{d,h} + \widehat{F}_\varepsilon^{-1}(\alpha). \tag{3}$$

The latter two relationships follow directly from the definition of the inverse empirical CDF and Eqn. (1). They also provide the basic framework for constructing probabilistic forecasts from distributions of prediction errors. Thus, if a dense grid of quantiles is considered (e.g., 99 percentiles: $\alpha_1 = 0.01, \alpha_2 = 0.02, \ldots, \alpha_{99} = 0.99$) then $\widehat{q}_{\alpha_1,P}, \ldots, \widehat{q}_{\alpha_L,P}$ approximate $\widehat{F}_P$ pretty well. If we assume that F_P has a density f_P, i.e., $F_P(x) = \int_{-\infty}^{x} f_P(z)\,dz$, then a density forecast $\widehat{f}_P$ can be provided as well.

3.3. Ensemble

The probabilistic forecasting concept discussed in Section 3.2 seems to be very general. Still, it is not sufficient to answer many problems in energy forecasting. The reason is that $\widehat{P}_{d,h}$ is considered on its own, independently of the forecasts for the neighboring hours. However, instead of looking at the $\mathcal{H}$ univariate price distributions $F_{P_{d,1}}, \ldots, F_{P_{d,\mathcal{H}}}$, we should be focusing on the *$\mathcal{H}$-dimensional distribution* $\boldsymbol{F_P}$ of the $\mathcal{H}$-dimensional price vector $\boldsymbol{P_d} = (P_{d,1}, \ldots, P_{d,\mathcal{H}})'$. Thus, we require a forecast $\widehat{\boldsymbol{F}}_{\boldsymbol{P}_d}$ for the multivariate distribution of $\boldsymbol{P_d}$. Unfortunately, many models cannot provide such a direct distributional forecast.

The solution to the latter problem is to compute an *ensemble forecast*. An ensemble is a collection of M paths $\mathcal{E}_M(\widehat{\boldsymbol{P}}_d) = (\widehat{\boldsymbol{P}}_d^{(1)}, \ldots, \widehat{\boldsymbol{P}}_d^{(M)})$ simulated from a forecasting model, typically using Monte-Carlo. These *paths*, visualized in the bottom right panel of Fig. 1, are also called *trajectories* or *scenarios*. The Glivenko-Cantelli theorem implies that for large M the ensemble approximates the underlying distribution $\boldsymbol{F_P}$ arbitrarily well. For practical applications it is important to remember that M should be chosen as large as possible, i.e., the ensemble should be composed of thousands or millions of paths. Finally, note that although new to EPF, the same or similar concepts have been used in different disciplines under different names: *simultaneous prediction intervals* (SPIs; as opposed to marginal PIs), *prediction bands*, *spatio-temporal trajectories* or *scenarios*, and *numerical weather prediction ensembles* (NWP).

4. Models

Nearly all review and survey publications introduce their own classifications of the techniques used for EPF. Some are better, some are worse, but most have many things in common. Here we review the classification of Weron (2014) with five groups of models, starting from the two most popular classes. Of course, many approaches considered in the literature are hybrid solutions, combining techniques from two or more classes.

4.1. Statistical

Statistical (also called *econometric*) approaches forecast the current price by a weighted combination of the past prices and/or past or current values of exogenous variables (e.g., demand or weather forecasts), typically in a linear regression setting. Autoregressive terms are often used to account for the dependencies between today's prices and those of the previous days, like in the ARX model proposed by Misiorek et al. (2006) and later used in a number of EPF studies (Gaillard et al., 2016; Kristiansen, 2012; Maciejowska et al., 2016; Marcjasz et al., 2018b; Nowotarski et al., 2014; Uniejewski et al., 2016; Weron, 2006):

$$P_{d,h} = \beta_{h,0} + \underbrace{\beta_{h,1} P_{d-1,h} + \beta_{h,2} P_{d-2,h} + \beta_{h,3} P_{d-7,h}}_{\text{autoregressive effects}} + \underbrace{\beta_{h,4} P_{d-1,min}}_{\text{non-linear effect}}$$
$$+ \underbrace{\beta_{h,5} L_{d,h}}_{\text{load forecast}} + \underbrace{\beta_{h,6} D_{Sat} + \beta_{h,7} D_{Sun} + \beta_{h,8} D_{Mon}}_{\text{weekday dummies}} + \varepsilon_{d,h}, \tag{4}$$

where $P_{d,h}$ is the price for day d and hour h, $P_{d-1,min}$ is the minimum of the previous day's 24 hourly prices, $L_{d,h}$ refers to the load forecast for day d and hour h (known on day $d-1$), and the three dummies (D_{Sat}, D_{Sun}, D_{Mon}) model the weekly seasonality. Following Ziel (2016), some authors refer to such parsimonious structures as *expert* models, since they are built on some prior knowledge of experts.

The standard approach to estimate model (4) is ordinary least squares (OLS). The procedure uses electricity prices from the past $\mathcal{D}$ days, i.e., $P_{1,h}, \ldots, P_{\mathcal{D},h}$, to predict the prices for the following day(s), i.e., $P_{\mathcal{D}+1,h}, P_{\mathcal{D}+2,h}$, etc. The value of $\mathcal{D}$ should be chosen with care, so that the estimation sample is 'long enough' to be able to extract patterns, but not 'too long' to give too much weight to distant past. Some authors also argue that $\mathcal{D}$ should be a multiple of the weekly cycle. Overall, there is no industry standard. Many studies consider a 'year' ($\mathcal{D} = 360$, $7 \times 52 = 364$ or 365 days) or 'two years' ($\mathcal{D} = 2 \times 364 = 728$ or $2 \times 365 = 730$) of hourly prices, but some use as short windows as 10-13 days, while other as long as four years (see Marcjasz et al., 2018a, for a review). The problem of optimal window length has been apparently overlooked in the EPF literature. Only very recently a few authors tackle it in a systematic way. Hubicka et al. (2018) propose a novel concept in energy forecasting that combines day-ahead predictions across different calibration windows (from 28 to 728 days) and show that this kind of averaging yields better results than selecting ex-ante only one 'optimal' window length. Marcjasz et al. (2018a) go one step further and introduce a new, well-performing weighting scheme for averaging these forecasts.

While (auto)regression models constitute the largest subset of statistical models, this class also includes:

- similar-day methods, like the *naive* method used by Contreras et al. (2003) and Conejo et al. (2005), which sets $\widehat{P}_{d,h} = P_{d-7,h}$ for Monday, Saturday or Sunday, and $\widehat{P}_{d,h} = P_{d-1,h}$ otherwise,
- exponential smoothing (Cruz et al., 2011; Jonsson et al., 2013),
- threshold models, like Markov regime-switching (MRS; Misiorek et al., 2006), threshold AR (TAR; Misiorek et al., 2006; Weron and Misiorek, 2008) or logistic smooth transition regression (LSTR; Gonzalez et al., 2012),
- models with GARCH innovations, typically in the context of volatility forecasting (Huurman et al., 2012; Koopman et al., 2007),
- shrinkage techniques, primarily LASSO (Ludwig et al., 2015; Ziel, 2016; Ziel and Weron, 2018), but also ridge regression and elastic nets (Uniejewski et al., 2016).

Statistical models are attractive because physical interpretation may be attached to the regressors, thus allowing engineers and system operators to better understand their behavior. The drawback is that – when used on their own – statistical models have problems with representing nonlinear phenomena, even though nonlinear models can be approximated pretty well by linear ones (under some regularity conditions). Nevertheless, nonlinear dependencies can be explicitly included via nonlinear variables, like $P_{d-1,min}$ in Eqn. (4). Alternatively, electricity spot prices (as well as exogenous variables) can be transformed using nonlinear functions, e.g., the area hyperbolic sine or the PIT defined in Eqn. (6), prior to fitting a statistical model. As Marcjasz et al. (2018a) and Uniejewski et al. (2018) report, this usually leads to significantly lower prediction errors.

4.2. Computational intelligence

The second, much more popular in the engineering literature, class of models is that of *computational* (CI) or *artificial intelligence* (AI). It is a very diverse group of nature-inspired tools developed to solve problems which (linear) statistical methods cannot handle efficiently. They combine elements of learning, evolution and fuzziness to create approaches that are capable of adapting to complex dynamic systems (in this sense they may be regarded as 'intelligent'). Artificial neural networks (ANNs), fuzzy systems and evolutionary computation (genetic algorithms, evolutionary programming, swarm intelligence) are unquestionably the main classes of CI (Keller et al., 2016). Some authors additionally include probabilistic reasoning and belief networks (at the intersection with traditional AI), artificial life techniques (at the intersection with biochemistry) and wavelets (at the intersection with digital signal processing), while others associate CI with machine learning, soft computing and data mining (see Weron, 2014, for a discussion).

CI models are flexible and can handle several types of nonlinearity, but at the cost of high computational complexity. Still, they are popular for short-term predictions. Like in load forecasting, ANNs have probably received the most attention (Abedinia et al., 2017; Amjady, 2006; Catalão et al., 2007; Dudek, 2016; Keles et al., 2016). Other techniques – like fuzzy logic, support vector machines (SVM) and evolutionary computation (genetic algorithms, evolutionary programming, swarm intelligence) – have been also applied, but typically in hybrid constructions (see Aggarwal et al., 2009; Weron, 2014, for reviews). Note, however, that the calibration of CI models is sensitive to initial parameters (starting values). A possible remedy is to use committee machines, i.e., averaging forecasts coming from different models or runs of the same model (Marcjasz et al., 2018b), though this increases the computational burden even more.

4.3. Reduced-form

The third class of *reduced-form* (also called *quantitative* or *stochastic*) approaches is not as universal as the previous two. As Weron (2014) argues, these models are not built to provide accurate hourly spot price forecasts, but rather to replicate the main characteristics of electricity prices at the daily time scale. Such models provide a simplified, hence the term 'reduced-form', yet realistic to some extent price dynamics and are commonly used for computationally intensive derivatives pricing (e.g., of electricity futures and options) and Value-at-Risk calculations. The two most popular model classes for the spot price dynamics include:

- *jump-diffusions*, which are a combination of Brownian dynamics and Poisson-type point (jump) processes (Benth et al., 2012; Cartea and Figueroa, 2005; Geman and Roncoroni, 2006; Weron, 2008),

- *regime-switching* models, typically involving a latent process describing the current state of the spot price (e.g., base regime vs. spike regime) and Brownian dynamics in one or more regime (Chen and Bunn, 2014; De Jong, 2006; Janczura et al., 2013; Janczura and Weron, 2010).

The reduced-form models can be either build for the spot or forward prices (Benth et al., 2008; Eydeland and Wolyniec, 2003; Weron, 2006). The main drawback of the former is the difficulty encountered when pricing derivatives, i.e., the identification of the risk premium linking spot and forward prices (Weron and Zator, 2014). On the other hand, forward price models allow for pricing of derivatives in a straightforward manner (but only those written on the forward price of electricity). However, they too have their limitations – scarcity of data for calibration and problems with deriving the properties of spot prices from the predicted forward curves (which include the risk premium).

4.4. Fundamental

Fundamental (also called *structural*) methods constitute the fourth class and include models that try to capture the basic physical and economic factors affecting the production and trading of electricity via a system of functional relationships. The latter are first postulated, then modeled and predicted independently (often via statistical, CI or reduced-form techniques) and aggregated to yield the equilibrium spot price. In general, two subclasses of fundamental models can be identified:

- *parameter rich* fundamental models, often developed as in-house products that use proprietary information (Eydeland and Wolyniec, 2003; Vehviläinen and Pyykkönen, 2005),
- *parsimonious structural* models of supply and demand that are built on publicly available information (Carmona and Coulon, 2014; Kanamura and Ohashi, 2007; Ziel and Steinert, 2016, 2018).

Because of the nature of fundamental data – which is often available only at the weekly or monthly resolution – parameter rich models are better suited for medium-term rather than short-term predictions. Also the parsimonious structural models are rarely used for short-term horizons. They are typically calibrated to daily data and, like reduced-form models, used for derivatives pricing and risk management; the X-model of Ziel and Steinert (2016) is one of the few exceptions. Compared to reduced-form models, they allow for a better description of the market fundamentals and price nonlinearities, in particular price spikes, mainly due to direct modeling of the nonlinear supply curve. This comes, however, at the cost of increasing the computational complexity.

4.5. Multi-agent

Multi-agent models simulate the operation of a system of interacting with each other heterogeneous agents, e.g., generating units, traders. Like fundamental models, they yield the equilibrium spot price by matching the demand and supply in the market. But unlike them, they construct the supply (and sometimes the demand) curve from the bids of the individual agents, not model it at the aggregate level (for a review see Ventosa et al., 2005).

Historically, *cost-based* or *production-cost* models (PCM) were the first members of this class. They were used in the pre-deregulation era with little price uncertainty to predict prices on an hour-by-hour, bus-by-bus level by matching demand forecasts to the supply, obtained by stacking up existing and planned generation units in order of their operating costs. *Equilibrium* (also called *game theoretic*) approaches, including the Nash-Cournot framework, supply function equilibrium and strategic PCMs, may be viewed as generalizations of PCMs that admit strategic bidding. This framework has been used extensively for the analysis of bidding strategies (Borgosz-Koczwara et al., 2009) or market power and market design (Holmberg et al., 2013), but EPF applications are very limited (Ruibal and Mazumdar, 2008). Finally, the increasingly popular adaptive *agent-based simulation* techniques can address features of electricity markets that static equilibrium models ignore, but again, their use for EPF is scarce (Young et al., 2014).

5. Forecast evaluation

All forecasting methods – point, probabilistic or ensemble – require evaluation with respect to the actual price $P_{d,h}$. While for point predictions it is relatively straightforward, for probabilistic and ensemble forecasts this becomes tricky. Therefore, the concept of a *score* (*scoring rule* or *loss*) comes in handy. In general, this is a function of the forecast and the actual price: $S(\text{`forecast'}, P_{d,h})$. A score is called *strictly proper* with respect to the forecasting target (e.g., the mean or the median) if only the true model optimizes the score; in this case the true model can be identified (Gneiting and Raftery, 2007).

5.1. Point forecasts

In point forecasting, scores are usually based on the so-called *forecast* or *prediction error*, defined as the difference between an observed value and its forecast, i.e., $e_{d,h} \equiv P_{d,h} - \widehat{P}_{d,h}$. The sign does not matter since the most widely used scores of forecast accuracy:

- the *absolute error* defined by $\mathrm{AE}(\widehat{P}_{d,h}, P_{d,h}) = |P_{d,h} - \widehat{P}_{d,h}| = |e_{d,h}|$,
- the *square error* defined by $\mathrm{SE}(\widehat{P}_{d,h}, P_{d,h}) = (P_{d,h} - \widehat{P}_{d,h})^2 = e^2_{d,h}$,

are symmetric. Averaging across the test sample yields the two most popular measures – the *mean absolute error* (MAE) and the *root mean square error* (RMSE):

$$\mathrm{MAE} \equiv \frac{1}{\mathcal{DH}} \sum_{d=1}^{\mathcal{D}} \sum_{h=1}^{\mathcal{H}} |e_{d,h}| \quad \text{and} \quad \mathrm{RMSE} \equiv \sqrt{\frac{1}{\mathcal{DH}} \sum_{d=1}^{\mathcal{D}} \sum_{h=1}^{\mathcal{H}} e^2_{d,h}}, \tag{5}$$

where $\mathcal{D}$ is the number of days in the test sample and $\mathcal{H}$ the number of load periods per day, i.e., $\mathcal{H} = 24$ for hourly periods. The absolute error is a strictly proper score for the median and the square error for the mean. Thus, the MAE is recommended for evaluating median and the RMSE for mean forecasts. However, both measures are commonly used in practice, despite the fact that the vast majority of point forecasts are mean predictions.

Since absolute and square errors are scale-dependent and hence hard to compare between different datasets, many authors use measures based on *percentage errors*, i.e., $e_{d,h}/P_{d,h}$. By far the most popular is the *mean absolute percentage error* (MAPE), which is computed as the mean of $|e_{d,h}|/P_{d,h}$ in the test sample. The MAPE measure works well in load forecasting, but for electricity price trajectories with close to zero or even negative values it may lead to absurd results and other more robust measures are preferred. So-called *scaled errors* were proposed by Hyndman and Koehler (2006) as a robust alternative to percentage errors. The idea is to divide $e_{d,h}$ by the MAE of a naïve forecast in the calibration window (instead of the actual price) and compute the *mean absolute scaled error* (MASE) as the mean of $|e_{d,h}|/\text{MAE}_{\text{naïve}}$ in the test sample. A scaled error is less than one if it arises from a better forecast than the average naïve forecast computed on the training data, and is greater than one if the forecast is worse.

Despite obvious advantages, scaled errors have not been used extensively in energy forecasting, one of the few examples is Jonsson et al. (2013). Alternative normalizations have been proposed instead, e.g., dividing $|e_{d,h}|$ by the average price attained in a given week. This yields the *weekly-weighted* MAE (WMAE; Weron and Misiorek, 2008), also called the *mean weekly error* (MWE; Conejo et al., 2005). As there is no 'industry standard', the error benchmarks used in the EPF literature vary a lot and – what is worse – are not used consistently, e.g., compare the two different definitions of MWE in Contreras et al. (2003) and Conejo et al. (2005).

5.2. Probabilistic forecasts

The main challenge with evaluating probabilistic forecasts is that we never observe the true distribution F_P of the underlying process and hence cannot compare the predictive distribution $\widehat{F}_P$ with the actual one. All we can do is compare $\widehat{F}_P$ with observed past prices. Over the years, a number of ways have been developed to evaluate probabilistic forecasts. The chosen approach will depend on the forecasting target – a quantile forecast requires a different evaluation than a predictive distribution, but sometimes it may also depend on the preference of the forecaster (see Nowotarski and Weron, 2018, for a recent review).

Gneiting and Katzfuss (2014) argue that 'probabilistic forecasting aims to maximize the sharpness of the predictive distributions, subject to reliability', where *reliability* (also called *calibration* or *unbiasedness*) refers to the statistical consistency between the distributional forecasts and the observations, while *sharpness* to how tightly the predicted distribution covers the actual one, i.e., to the concentration of the predictive distributions. For instance, if a 90% prediction interval (PI) covers 90% of the observed prices, then this PI is said to be reliable (Pinson et al., 2007), well calibrated (Gneiting et al., 2007) or unbiased (Taylor, 1999). To obtain the empirical coverage we typically focus on the indicator series of 'hits and misses': $I_{d,h} = 1$ if $P_{d,h} \in [\widehat{L}_{d,h}, \widehat{U}_{d,h}]$ and zero otherwise. Some authors simply report the empirical coverage itself (*PI coverage probability*, PICP), while others subtract it from the nominal level (*PI nominal coverage*, PINC) to obtain the *average coverage error*: ACE = PICP − PINC. Generally, the closer is ACE to zero the better. However, to formally check *unconditional coverage* (UC), i.e., whether $\mathbb{P}(I_{d,h} = 1) = (1 - \alpha)$, we can use the approach of Kupiec (1995), which tests if $I_{d,h}$ is i.i.d. Bernoulli with mean $(1 - \alpha)$.

Since the Kupiec test does not have any power against the alternative that the ones and zeros come clustered together in $I_{d,h}$, Christoffersen (1998) introduced the *independence* and *conditional coverage* (CC) tests. Independence is tested against an explicit first-order Markov alternative and the latter is simply a joint test for independence and UC. In order to capture more than just the first-order dependency, we can conduct the independence test (and consequently the CC test) for any time lag or use the Ljung-Box statistics for a joint test of independence for several lags (Berkowitz et al., 2011).

Testing for the goodness-of-fit of a predictive distribution is, in general, more challenging than evaluating the reliability of a PI. The most common approach is to use the *Probability Integral Transform*:

$$\mathrm{PIT}_{d,h} = \widehat{F}_P(P_{d,h}). \tag{6}$$

If the distributional forecast is perfect, then $\mathrm{PIT}_{d,h}$ is independent and uniformly distributed, which can be assessed using a formal statistical test or graphically (Gneiting and Katzfuss, 2014). Alternatively, we can first apply the inverse of the standard normal CDF to $\mathrm{PIT}_{d,h}$ to yield a Gaussian random variable, then jointly test for independence and normality (i.e., for conditional coverage) using the approach of Berkowitz (2001). The argument behind this is that in finite-samples, tests based on the Gaussian likelihood are more convenient and flexible than tests of uniformity.

The definition of *sharpness*, on the other hand, derives from the idea that reliable predictive distributions of null width would correspond to perfect point predictions (Gneiting and Raftery, 2007). Unlike reliability, which is a joint property of the predictions and the observations, sharpness is a property of the forecasts only. Sharpness is closely related to the concept of the proper scoring rules. In fact, scoring rules assess reliability and sharpness simultaneously (Gneiting and Katzfuss, 2014).

The *pinball loss* for quantile forecasts and the *continuous ranked probability score* (CRPS) for distributional forecasts are the two most popular proper scoring rules in energy forecasting. The pinball loss is a special case of an *asymmetric piecewise linear loss* function:

$$\mathrm{PL}\left(\widehat{q}_{\alpha,P}, P_{d,h}, \alpha\right) = \begin{cases} (1-\alpha)\left(\widehat{q}_{\alpha,P} - P_{d,h}\right), & \text{for } P_{d,h} < \widehat{q}_{\alpha,P}, \\ \alpha\left(P_{d,h} - \widehat{q}_{\alpha,P}\right), & \text{for } P_{d,h} \geq \widehat{q}_{\alpha,P}, \end{cases} \tag{7}$$

where $\widehat{q}_{\alpha,P}$ is the quantile function and $P_{d,h}$ is the actually observed price. The pinball loss (also known as the *linlin*, *bilinear* or *newsboy* loss; Elliott and Timmermann, 2016) is a strictly proper score for the α-th quantile. To provide an aggregate score, the pinball loss can be averaged across different quantiles, e.g., 99 percentiles (but also across the 24 hours of the target day, as in the GEFCom2014 competition; see Hong et al., 2016). If the α-grid is arbitrarily dense, then this average converges to:

$$\mathrm{CRPS}(\widehat{F}_P, P_{d,h}) = \int_{-\infty}^{\infty} F(z) - \mathbb{I}_{\{z \geq P_{d,h}\}}\, dz = \mathbb{E}|\widehat{P}_{d,h} - P_{d,h}| - \frac{1}{2}\mathbb{E}|\widehat{P}_{d,h} - \widehat{P}'_{d,h}|, \tag{8}$$

where random variables $\widehat{P}_{d,h}$ and $\widehat{P}'_{d,h}$ are two independent copies distributed as $\widehat{F}_P$; the CRPS is a strictly proper score for the true distribution F_P. The latter representation reflects the reliability in the first term and the lack of sharpness in the second. Naturally, a lower pinball or CRPS score indicates a better forecast. Scores for two probabilistic forecasts can be tested for equal predictive performance using one of the methods discussed in Section 5.4.

5.3. Ensemble forecasts

When evaluating ensemble forecasts we consider $\mathcal{E}_M(\widehat{\boldsymbol{P}}_d) = (\widehat{\boldsymbol{P}}_d^{(1)}, \ldots, \widehat{\boldsymbol{P}}_d^{(M)})$, i.e., the ensemble that approximates the multivariate distributional forecast $\widehat{\boldsymbol{F}}_{\boldsymbol{P}}$ of $\boldsymbol{P_d}$, and a scoring rule for multivariate distributions. Out of the rules defined in the literature (Gneiting and Raftery, 2007; Scheuerer and Hamill, 2015), the *Dawid-Sebastiani* and *variogram scores* are not strictly proper for multivariate distributions, while the *log-score* requires forecasts of a multivariate density, which for many models is not available. Only the *energy score* introduced by Gneiting and Raftery (2007) satisfies our criteria and hence is recommended for evaluating ensemble forecasts. It can be regarded as a multivariate version of the CRPS:

$$\mathrm{ES}(\widehat{\boldsymbol{F}}_{\boldsymbol{P}}, \boldsymbol{P_d}) = \mathbb{E}\|\widehat{\boldsymbol{P}}_{\boldsymbol{d}} - \boldsymbol{P_d}\|_2 - \tfrac{1}{2}\mathbb{E}\|\widehat{\boldsymbol{P}}_{\boldsymbol{d}} - \widehat{\boldsymbol{P}}'_{\boldsymbol{d}}\|_2, \tag{9}$$

where $\widehat{\boldsymbol{P}}_{\boldsymbol{d}}$ and $\widehat{\boldsymbol{P}}'_{\boldsymbol{d}}$ are two independent copies distributed as $\widehat{\boldsymbol{F}}_{\boldsymbol{P}}$ and $\|\cdot\|_2$ denotes the Euclidean norm. The energy score can be estimated using the ensemble $\mathcal{E}_M(\widehat{\boldsymbol{P}}_d)$ by replacing the expected values in Eqn. (9) by the sample means across the ensemble. The energy score is based on the energy distance, which is an extremely powerful tool in multivariate statistics (Székely and Rizzo, 2013). For instance, it allows to construct tests for multivariate independence or tests for the equality of two multivariate distributions.

5.4. Testing for equal predictive performance

The methods discussed in Sections 5.1-5.3 can be used to rank the forecasts, but not to draw statistically significant conclusions on the outperformance of the forecasts of one model by those of another. The extremely simple Diebold and Mariano (1995) test can be used for exactly this purpose. It is an asymptotic z-test of the hypothesis that the mean of the *loss differential* series:

$$\delta_{d,h} = S_1(\widehat{F}_P, P_{d,h}) - S_2(\widehat{F}_P, P_{d,h}) \tag{10}$$

is zero, where $S_i(\cdot,\cdot)$ is the score of the forecasts of model i. Although the DM test is much more popular in the point forecasting literature, it is readily applicable to probabilistic and ensemble forecasts (Nowotarski and Weron, 2018). In the point forecasting context, $S_i(\cdot,\cdot)$ will usually be the absolute $|e_{d,h}|$ or squared $e_{d,h}^2$ error, and in the probabilistic or ensemble forecasting context, it may be any strictly proper scoring rule, in particular the pinball loss, the CRPS or the energy score. Given the loss differential series, we compute the statistic:

$$\mathrm{DM} = \sqrt{\mathcal{DH}}\,\frac{\widehat{\mu}(\delta_{d,h})}{\widehat{\sigma}(\delta_{d,h})}, \tag{11}$$

where $\widehat{\mu}(\delta_{d,h})$ and $\widehat{\sigma}(\delta_{d,h})$ are the sample mean and standard deviation of $\delta_{d,h}$, respectively, and $\mathcal{DH}$ is the length of the out-of-sample test period. The key hypothesis of equal predictive accuracy (i.e., equal expected loss) corresponds to $\mathbb{E}(\delta_{d,h}) = 0$, in which case, under the assumption of covariance stationarity of $\delta_{d,h}$, the DM statistic is asymptotically standard normal, and one- or two-sided asymptotic tail probabilities are readily calculated. To avoid a common mistake, we should remember that the DM test compares forecasts of two models, not the models themselves (Diebold, 2015).

Since in day-ahead electricity markets the predictions for all 24 hours of the next day are usually made at the same time using the same information set, forecast errors for a particular day will typically exhibit high serial correlation. Therefore, it is advisable to separately conduct the DM tests for each load period $h = 1, ..., \mathcal{H}$ (as in Bordignon et al., 2013; Nowotarski et al., 2014; Uniejewski et al., 2016) or jointly for all load periods (as suggested by Ziel and Weron, 2018). Obviously, in both approaches the out-of-sample length $\mathcal{DH}$ in Eqn. (11) is reduced to $\mathcal{D}$. Note, that for ensemble forecasting only the latter approach is feasible, so that only one statistic for each pair of models is computed based on the $\mathcal{H}$-dimensional vector of errors for each day.

Alternative forecast comparison test procedures include the *model confidence set* (MCS) approach of Hansen et al. (2011), a test of *forecast encompassing* (Harvey et al., 1998) and the Giacomini and White (2006) test for *conditional predictive ability* (CPA). For two models, the MCS approach is similar to the DM test but estimates the distribution of the test statistic by a bootstrap procedure. In the test of *forecast encompassing*, the null hypothesis is that model 2 encompasses model 1, i.e., that predictions of model 1 do not contain additional information with respect to those of model 2. Finally, in the CPA test the null hypothesis is $H_0 : \boldsymbol{\phi} = 0$ in the regression:

$$\delta_{d,h} = \boldsymbol{\phi}' \mathbb{X}_{d-1,h} + \varepsilon_{d,h}, \tag{12}$$

where $\mathbb{X}_{d-1,h}$ contains elements from the information set on day $d-1$ and load period h (or all load periods for a joint score), i.e., a constant and lags of $\delta_{d,h}$. While both the CPA and DM tests can be used for nested and non-nested models – as long as the calibration window does not grow with the sample size (Giacomini and Rossi, 2013) – only the CPA test accounts for parameter estimation uncertainty (through 'conditioning') and hence is the preferred option.

References

Abedinia, O., Amjady, N., and Zareipour, H. (2017). A new feature selection technique for load and price forecast of electrical power systems. *IEEE Transactions on Power Systems*, 32(1):62–74.

Aggarwal, S., Saini, L., and Kumar, A. (2009). Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power and Energy Systems*, 31:13–22.

Amjady, N. (2006). Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Transactions on Power Systems*, 21(2):887–896.

Bello, A., Bunn, D., Reneses, J., and Muñoz, A. (2016). Parametric density recalibration of a fundamental market model to forecast electricity prices. *Energies*, 9(11):959.

Benth, F., Kiesel, R., and Nazarova, A. (2012). A critical empirical study of three electricity spot price models. *Energy Economics*, 34(5):1589–1616.

Benth, F. E., Benth, J. S., and Koekebakker, S. (2008). *Stochastic modeling of electricity and related markets*. World Scientific.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, 57(12):2213–2227.

Bordignon, S., Bunn, D. W., Lisi, F., and Nan, F. (2013). Combining day-ahead forecasts for British electricity prices. *Energy Economics*, 35:88–103.

Borgosz-Koczwara, M., Weron, A., and Wyłomańska, A. (2009). Stochastic models for bidding strategies on oligopoly electricity market. *Mathematical Methods of Operations Research*, 69(3):579–592.

Carmona, R. and Coulon, M. (2014). A survey of commodity markets and structural models for electricity prices. In Benth, F., Kholodnyi, V., and Laurence, P., editors, *Quantitative Energy Finance: Modeling, Pricing, and Hedging in Energy and Commodity Markets*, pages 41–83. Springer.

Cartea, A. and Figueroa, M. (2005). Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4):313–335.

Catalão, J. P. S., Mariano, S. J. P. S., Mendes, V. M. F., and Ferreira, L. A. F. M. (2007). Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power Systems Research*, 77(10):1297–1304.

Chen, D. and Bunn, D. (2014). The forecasting performance of a finite mixture regime-switching model for daily electricity prices. *Journal of Forecasting*, 33(5):364–375.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Conejo, A. J., Contreras, J., Espínola, R., and Plazas, M. A. (2005). Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting*, 21:435–462.

Contreras, J., Espínola, R., Nogales, F., and Conejo, A. (2003). ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems*, 18(3):1014–1020.

Cruz, A., Muñoz, A., Zamora, J., and Espinola, R. (2011). The effect of wind generation and weekday on Spanish electricity spot price forecasting. *Electric Power Systems Research*, 81(10):1924–1935.

De Jong, C. (2006). The nature of power spikes: A regime-switch approach. *Studies in Nonlinear Dynamics & Econometrics*, 10(3):Article 3.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests. *Journal of Business and Economic Statistics*, 33(1):1–9.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263.

Dudek, G. (2016). Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting. *International Journal of Forecasting*, 32:1057–1060.

Elliott, G. and Timmermann, A. (2016). *Economic Forecasting.* Princeton University Press.

Eydeland, A. and Wolyniec, K. (2003). *Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging.* John Wiley & Sons, Hoboken, NJ.

Gaillard, P., Goude, Y., and Nedellec, R. (2016). Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *International Journal of Forecasting*, 32(3):1038–1050.

Geman, H. and Roncoroni, A. (2006). Understanding the fine structure of electricity prices. *Journal of Business*, 79:1225–1261.

Giacomini, R. and Rossi, B. (2013). Forecasting in macroeconomics. In Hashimzade, N. and Thornton, M., editors, *Handbook of Research Methods and Applications on Empirical Macroeconomics*, pages 381–407. Edward Elgar.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Gianfreda, A., Parisio, L., and Pelagatti, M. (2016). The impact of RES in the Italian day-ahead and balancing markets. *Energy Journal*, 37:161–184.

Gneiting, T., Balabdaoui, F., and Raftery, A. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society B*, 69:243–268.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *The Annual Review of Statistics and Its Application*, 1:125–151.

Gneiting, T. and Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Gonzalez, V., Contreras, J., and Bunn, D. (2012). Forecasting power prices using a hybrid fundamental-econometric model. *IEEE Transactions on Power Systems*, 27(1):363–372.

Gürtler, M. and Paulsen, T. (2018). Forecasting performance of time series models on electricity spot markets: a quasi-meta-analysis. *International Journal of Energy Sector Management*, 12(1):103–129.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79:453–497.

Harvey, D., Leybourne, S., and Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business and Economic Statistics*, 16:254–259.

Holmberg, P., Newbery, D., and Ralph, D. (2013). Supply function equilibria: Step functions and continuous representations. *Journal of Economic Theory*, 148(4):1509–1551.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., and Hyndman, R. J. (2016). Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting*, 32(3):896–913.

Hubicka, K., Marcjasz, G., and Weron, R. (2018). A note on averaging day-ahead electricity price forecasts across calibration windows. *IEEE Transactions on Sustainable Energy*. DOI: 10.1109/TSTE.2018.2869557.

Huurman, C., Ravazzolo, F., and Zhou, C. (2012). The power of weather. *Computational Statistics and Data Analysis*, 56(11):3793–3807.

Hyndman, R. and Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.

Janczura, J., Trück, S., Weron, R., and Wolff, R. (2013). Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics*, 38:96–110.

Janczura, J. and Weron, R. (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy Economics*, 32(5):1059–1073.

Jonsson, T., Pinson, P., Nielsen, H. A., Madsen, H., and Nielsen, T. (2013). Forecasting electricity spot prices accounting for wind power predictions. *IEEE Transactions on Sustainable Energy*, 4(1):210–218.

Joskow, P. (2008). Lessons learned from electricity market liberalization. *Energy Journal*, 29:9–42.

Kanamura, T. and Ohashi, K. (2007). A structural model for electricity prices with spikes: Measurement of spike risk and optimal policies for hydropower plant operation. *Energy Economics*, 29(5):1010–1032.

Keles, D., Scelle, J., Paraschiv, F., and Fichtner, W. (2016). Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Applied Energy*, 162:218–230.

Keller, J., Liu, D., and Fogel, D. (2016). *Fundamentals of Computational Intelligence: Neural Networks, Fuzzy Systems, and Evolutionary Computation*. Wiley.

Kiesel, R. and Paraschiv, F. (2017). Econometric analysis of 15-minute intraday electricity prices. *Energy Economics*, 64:77–90.

Klæboe, G., Eriksrud, A. L., and Fleten, S.-E. (2015). Benchmarking time series based forecasting models for electricity balancing market prices. *Energy Systems*, 6(1):43–61.

Koopman, S., Ooms, M., and Carnero, A. (2007). Periodic seasonal Reg-ARFIMA-GARCH models for daily electricity spot prices. *Journal of the American Statistical Association*, 102:16–27.

Kristiansen, T. (2012). Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy*, 49:328–332.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2).

Ludwig, N., Feuerriegel, S., and 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(1):19–36.

Lund, H., Sorknæs, P., Mathiesen, B. V., and Hansen, K. (2018). Beyond sensitivity analysis:

A methodology to handle fuel and electricity prices when designing energy scenarios. *Energy Research & Social Science*, 39:108–116.

Maciejowska, K., Nowotarski, J., and Weron, R. (2016). Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. *International Journal of Forecasting*, 32(3):957–965.

Maciejowska, K. and 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(2):994–1005.

Marcjasz, G., Serafin, T., and Weron, R. (2018a). Selection of calibration windows for day-ahead electricity price forecasting. *Energies*, 11(9):2364.

Marcjasz, G., Uniejewski, B., and Weron, R. (2018b). On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks. *International Journal of Forecasting*. DOI: 10.1016/j.ijforecast.2017.11.009.

Mayer, K. and Trück, S. (2018). Electricity markets around the world. *Journal of Commodity Markets*, 9:77–100.

Misiorek, A., Trück, S., and Weron, R. (2006). Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. *Studies in Nonlinear Dynamics & Econometrics*, 10(3):Article 2.

Mohamed, A. and El-Hawary, M. (2016). Mid-term electricity price forecasting using SVM. Canadian Conference on Electrical and Computer Engineering, Art. no. 7726765.

Nowotarski, J., Raviv, E., Trück, S., and Weron, R. (2014). An empirical comparison of alternate schemes for combining electricity spot price forecasts. *Energy Economics*, 46:395–412.

Nowotarski, J. and Weron, R. (2015). Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Computational Statistics*, 30(3):791–803.

Nowotarski, J. and Weron, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548–1568.

Pinson, P., Nielsen, H., Moller, J., Madsen, H., and Kariniotakis, G. (2007). Non-parametric probabilistic forecasts of wind power: Required properties and evaluation. *Wind Energy*, 10(6):497–516.

Ries, J., Gaudard, L., and Romerio, F. (2016). Interconnecting an isolated electricity system to the European market: The case of Malta. *Utilities Policy*, 40:1–14.

Ruibal, C. and Mazumdar, M. (2008). Forecasting the mean and the variance of electricity prices in deregulated markets. *IEEE Transactions on Power Systems*, 23(1):25–32.

Scheuerer, M. and Hamill, T. M. (2015). Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4):1321–1334.

Steinert, R. and Ziel, F. (2018). Short-to mid-term day-ahead electricity price forecasting using futures. Working Paper, arXiv:1801.10583. To appear in: The Energy Journal.

Székely, G. J. and Rizzo, M. L. (2013). Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272.

Taylor, J. W. (1999). Evaluating volatility and interval forecasts. *Journal of Forecasting*, 18:111–128.

Uniejewski, B., Nowotarski, J., and Weron, R. (2016). Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies*, 9:621.

Uniejewski, B., Weron, R., and Ziel, F. (2018). Variance stabilizing transformations for electricity spot price forecasting. *IEEE Transactions on Power Systems*, 33(2):2219–2229.

Vehviläinen, I. and Pyykkönen, T. (2005). Stochastic factor model for electricity spot price – The case of the Nordic market. *Energy Economics*, 27(2):351–367.

Ventosa, M., Baillo, A., Ramos, A., and Rivier, M. (2005). Electricity market modeling trends. *Energy Policy*, 33(7):897–913.

Weron, R. (2006). *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach.* John Wiley & Sons, Chichester.

Weron, R. (2008). Market price of risk implied by Asian-style electricity options and futures. *Energy Economics*, 30:1098–1115.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081.

Weron, R. and Misiorek, A. (2008). Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, 24:744–763.

Weron, R. and Zator, M. (2014). Revisiting the relationship between spot and futures prices in the Nord Pool electricity market. *Energy Economics*, 44:178–190.

Wolff, G. and Feuerriegel, S. (2017). Short-term dynamics of day-ahead and intraday electricity prices. *International Journal of Energy Sector Management*, 11(4):557–573.

Young, D., Poletti, S., and Browne, O. (2014). Can agent-based models forecast spot prices in electricity markets? Evidence from the New Zealand electricity market. *Energy Economics*, 45:419–434.

Zaleski, P. and Klimczak, D. (2015). Prospects for the rise of renewable sources of energy in Poland. Balancing renewables on the intra-day market. In Zamasz, K., editor, *Capacity Market in Contemporary Economic Policy*, pages 124–138. Difin.

Ziel, F. (2016). Forecasting electricity spot prices using LASSO: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems*, 31(6):4977–4987.

Ziel, F. (2017). Modeling the impact of wind and solar power forecasting errors on intraday electricity prices. IEEE Conference Proceedings - EEM17, DOI: 10.1109/EEM.2017.7981900.

Ziel, F. and Steinert, R. (2016). Electricity price forecasting using sale and purchase curves: The X-model. *Energy Economics*, 59:435–454.

Ziel, F. and Steinert, R. (2018). Probabilistic mid- and long-term electricity price forecasting. *Renewable and Sustainable Energy Reviews*, 94:251–266.

Ziel, F., Steinert, R., and Husmann, S. (2015). Efficient modeling and forecasting of electricity spot prices. *Energy Economics*, 47:89–111.

Ziel, F. and Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, 70:396–420.

HSC Research Report Series 2018

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html

01 *An empirical analysis of green energy adoption among residential consumers in Poland* by Anna Kowalska-Pyzalska
02 *Efficient forecasting of electricity spot prices with expert and LASSO models* by Bartosz Uniejewski and Rafał Weron
03 *A note on averaging day-ahead electricity price forecasts across calibration windows* by Katarzyna Hubicka, Grzegorz Marcjasz and Rafał Weron
04 *Household willingness to pay for green electricity in Poland* by Anna Kowalska-Pyzalska and David Ramsey
05 *Probabilistic electricity price forecasting with NARX networks: Combine point or probabilistic forecasts?* by Grzegorz Marcjasz, Bartosz Uniejewski and Rafał Weron
06 *Selection of calibration windows for day-ahead electricity price forecasting* by Grzegorz Marcjasz, Tomasz Serafin and Rafał Weron
07 *Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO* by Bartosz Uniejewski, Grzegorz Marcjasz and Rafał Weron
08 *Electricity price forecasting* by Rafał Weron and Florian Ziel