



Contents lists available at ScienceDirect

International Journal of Forecasting

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

Review

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

Rafał Weron

Institute of Organization and Management, Wrocław University of Technology, Wrocław, Poland

ARTICLE INFO

Keywords:

Electricity price forecasting
Day-ahead market
Seasonality
Autoregression
Neural network
Factor model
Forecast combination
Probabilistic forecast

ABSTRACT

A variety of methods and ideas have been tried for *electricity price forecasting* (EPF) over the last 15 years, with varying degrees of success. This review article aims to explain the complexity of available solutions, their strengths and weaknesses, and the opportunities and threats that the forecasting tools offer or that may be encountered. The paper also looks ahead and speculates on the directions EPF will or should take in the next decade or so. In particular, it postulates the need for objective comparative EPF studies involving (i) the same datasets, (ii) the same robust error evaluation procedures, and (iii) statistical testing of the significance of one model's outperformance of another.

© 2014 The Author. Published by Elsevier B.V. on behalf of International Institute of Forecasters.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Contents

1. Introduction.....	2
2. Literature query.....	3
2.1. Bibliometrics of 'electricity price forecasting'	3
2.2. Major review and survey publications	5
3. What and how are we forecasting?	7
3.1. The electricity 'spot' price	7
3.2. Forecasting horizons.....	9
3.3. Evaluating point forecasts.....	9
3.4. Overview of modeling approaches	10
3.5. Multi-agent models	11
3.5.1. Nash-Cournot framework	11
3.5.2. Supply function equilibrium	11
3.5.3. Strategic production-cost models	12
3.5.4. Agent-based simulation models	12
3.5.5. Strengths and weaknesses	13
3.6. Fundamental models	13
3.6.1. Parameter-rich fundamental models	14
3.6.2. Parsimonious structural models	14
3.6.3. Strengths and weaknesses	15

E-mail address: rafal.weron@pwr.wroc.pl.

<http://dx.doi.org/10.1016/j.ijforecast.2014.08.008>

0169-2070/© 2014 The Author. Published by Elsevier B.V. on behalf of International Institute of Forecasters. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

3.7.	Reduced-form models	15
3.7.1.	Jump-diffusion models.....	16
3.7.2.	Markov regime-switching models.....	18
3.7.3.	Strengths and weaknesses	20
3.8.	Statistical models.....	20
3.8.1.	Similar-day and exponential smoothing methods.....	20
3.8.2.	Regression models.....	21
3.8.3.	AR-type time series models	22
3.8.4.	ARX-type time series models.....	23
3.8.5.	Threshold autoregressive models.....	25
3.8.6.	Heteroskedasticity and GARCH-type models	26
3.8.7.	Strengths and weaknesses	27
3.9.	Computational intelligence models.....	27
3.9.1.	Taxonomy of neural networks.....	28
3.9.2.	Feed-forward neural networks	28
3.9.3.	Recurrent neural networks	30
3.9.4.	Fuzzy neural networks	31
3.9.5.	Support vector machines	31
3.9.6.	Strengths and weaknesses	31
4.	A look into the future of 'electricity price forecasting'	32
4.1.	Fundamental price drivers and input variables.....	32
4.1.1.	Modeling and forecasting the trend-seasonal components.....	32
4.1.2.	Spike forecasting and the reserve margin.....	33
4.2.	Beyond point forecasts	36
4.2.1.	Interval forecasts	36
4.2.2.	Density forecasts.....	37
4.2.3.	Threshold forecasting.....	38
4.3.	Combining forecasts	38
4.3.1.	Point forecasts.....	39
4.3.2.	Probabilistic forecasts.....	41
4.4.	Multivariate factor models.....	42
4.5.	The need for an EPF-competition	44
4.5.1.	A universal test ground	44
4.5.2.	Guidelines for evaluating forecasts	45
4.6.	Final word.....	46
	Acknowledgments	46
	References.....	46

1. Introduction

Since the early 1990s, the process of deregulation and the introduction of competitive markets have been reshaping the landscape of the traditionally monopolistic and government-controlled power sectors. In many countries worldwide, electricity is now traded under market rules using spot and derivative contracts. However, electricity is a very special commodity. It is economically non-storable, and power system stability requires a constant balance between production and consumption (Kaminski, 2013; Shahidehpour, Yamin, & Li, 2002). At the same time, electricity demand depends on weather (temperature, wind speed, precipitation, etc.) and the intensity of business and everyday activities (on-peak vs. off-peak hours, weekdays vs. weekends, holidays and near-holidays, etc.). On the one hand, these unique and specific characteristics lead to price dynamics not observed in any other market, exhibiting seasonality at the daily, weekly and annual levels, and abrupt, short-lived and generally unanticipated price spikes. On the other hand, they have encouraged researchers to intensify their efforts in the development of better forecasting techniques.

At the corporate level, electricity price forecasts have become a fundamental input to energy companies' decision-making mechanisms (Bunn, 2004; Eydeland & Wolyniec,

2003; Weron, 2006). As the California crisis of 2000–2001 showed, electric utilities are the most vulnerable, since they generally cannot pass their costs on to the retail consumers (Joskow, 2001). The costs of over-/under-contracting and then selling/buying power in the balancing (or real-time) market are typically so high that they can lead to huge financial losses or even bankruptcy. Extreme price volatility, which can be up to two orders of magnitude higher than that of any other commodity or financial asset, has forced market participants to hedge not only against volume risk but also against price movements. Price forecasts from a few hours to a few months ahead have become of particular interest to power portfolio managers. A generator, utility company or large industrial consumer who is able to forecast the volatile wholesale prices with a reasonable level of accuracy can adjust its bidding strategy and its own production or consumption schedule in order to reduce the risk or maximize the profits in day-ahead trading.

A variety of methods and ideas have been tried for *electricity price forecasting* (EPF), with varying degrees of success. This review article aims to explain the complexity of the available solutions, with a special emphasis on the strengths and weaknesses of the individual methods. In an attempt to determine which approaches are the most popular, In Section 2 we provide an overview of the

existing literature on EPF, including a bibliometric study of the Web of Science and Scopus databases, and a brief summary of the review/survey publications on this topic. In Section 3, we explain the mechanics of price formation in electricity markets and define the main object of interest: the day-ahead electricity price. Next, following Weron (2006), we classify the techniques in terms of both the planning horizon's duration and the applied methodology, and review the most interesting approaches. We look back over the last 15 years of EPF, in an attempt to systematize the rapidly growing literature. Then, in Section 4, we look ahead and speculate on the directions EPF will or should take in the next decade or so. In particular, we propose a universal test ground that all forecasters should use in order to allow for direct comparisons between the different studies, stress the importance of seasonality and fundamentals in EPF, and highlight some recent trends – interval and density forecasting, the ‘forgotten art’ of combining forecasts, and the increasing popularity of multivariate factor models.

2. Literature query

There are essentially two ways to learn about a new research area. One is to perform a literature query using one of the established databases and find the ‘hot topics’, the highly cited papers (hoping that they are ‘the influential ones’), and the publishing trends. The other is to read a couple of review/survey papers, trusting that they are unbiased, wide in scope and relatively up-to-date. To help a newcomer to the field of *electricity price forecasting* (EPF), we have performed both a bibliometric analysis (Section 2.1) and a critical review of the review/survey publications that are out there (Section 2.2).

2.1. Bibliometrics of ‘electricity price forecasting’

In this section, we report on the bibliometric analysis we performed on 10 May 2014 using two well-established and generally acknowledged databases: *Web of Science* (WoS) and *Scopus*. The results do differ quantitatively, as the collections of publications indexed by WoS and Scopus are not the same, but do not differ qualitatively. Generally, WoS is a subset of Scopus, meaning that we could limit our analysis to WoS only. However, the Scopus search engine is more user-friendly and allows for more refined queries. If we limit our search to journal articles published in English only, then the differences between the databases are not that significant. We will first present general results for both databases, then more specialized queries for Scopus only. We should also note that the choice of these two databases has its limitations, most notably the fact that some of the newer journals, like the *Journal of Energy Markets*, are not indexed in these systems.

In Fig. 1, we plot the numbers of WoS- and Scopus-indexed EPF publications in the years 1989–2013.¹ The overall numbers of publications are 304 for WoS and

497 for Scopus, of which 136 (45%) and 206 (41%), respectively, are journal articles. Articles indexed within the Web of Science refer to journals listed in the Journal Citation Reports only, while the collection of Scopus-listed journals is much richer. Both databases are constantly being expanded to cover more volumes of proceedings, but the numbers are still much less representative of the true number of conference papers than is the case for journals and journal articles. The Scopus-indexed collection of reviews, conference reviews, books and book chapters is even less complete. Hence, in what follows, we will concentrate mostly on journal articles.

Except for a few isolated cases, EPF publications did not appear in the literature before the year 2000. The next major breakthrough occurred in the years 2005 and 2006, when the number of publications first doubled, then tripled with respect to 2002–2004 figures. Initially, this increased inflow of EPF publications was due mostly to proceedings (WoS terminology) or conference (Scopus terminology) papers; journal articles followed with a delay. The overall publication rate increased until 2009/2010, then dropped to 2006–2008 levels because of a reduced number of conference papers. As of 2013, the topic seems to have saturated the research community, although the number of citations is still increasing, as can be seen in Fig. 2. Possibly a new fundamental impulse – like the deregulation of the late 1990s or the increased volatility of electricity spot prices in the mid-2000s – is needed in order to propel electricity price forecasting to a new level of publication intensity.

As far as subject categories are concerned, most of the articles have appeared in journals classified by Scopus as *Engineering* or *Energy*, followed by *Computer Science*, *Mathematics*, *Business*, *Management & Accounting* and *Economics, Econometrics & Finance*. It is also interesting to see which outlets are the most popular for EPF articles. Clearly, the number one journal is *IEEE Transactions on Power Systems*, with 33 publications (out of 206 indexed by Scopus), see Fig. 3. Interestingly, the share of ‘neural network’-type (more generally: artificial or computational intelligence) methods and statistical time series models is equal in this collection: nine ‘neural network’ papers, nine ‘statistical time series’ papers, four papers where both approaches have been used and 11 papers where neither ‘neural network’ nor ‘statistical time series’ methods have been used. It should be noted that the classification was automatic and may include some errors. For ‘neural network’-type papers, the Scopus query given in footnote 1 was modified to include

query: TS=((('forecasting electricity" OR 'predicting electricity") AND (('electricity spot" OR 'electricity day-ahead" OR 'electricity price')) OR (('price forecasting" OR 'price prediction" OR 'forecasting price" OR 'predicting price" OR 'forecasting spikes" OR 'forecasting VAR')) AND (('electricity spot price" OR 'electricity price" OR 'electricity market" OR 'day-ahead market" OR 'power market'))); and the equivalent Scopus query: TITLE-ABS-KEY(. . .). All look-ups have been refined further to exclude non-English language texts or to include only specific document types.

¹ To search publication titles, abstracts and keywords for ‘electricity price forecasting’-related phrases, we have used the following WoS

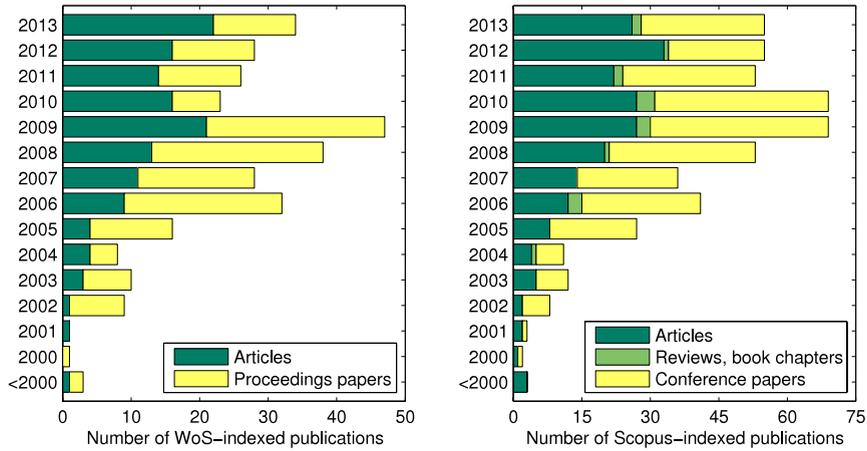


Fig. 1. The numbers of WoS- (left panel) and Scopus-indexed (right panel) *electricity price forecasting* (EPF) publications in the years 1989–2013. All publications prior to the year 2000 (three for WoS, three for Scopus) have been aggregated into one category, '<2000'.

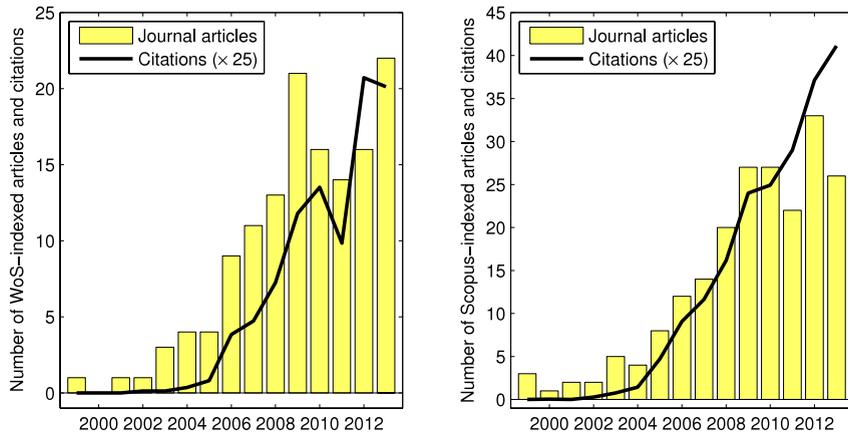


Fig. 2. The numbers of WoS- (left panel) and Scopus-indexed (right panel) EPF journal articles and citations of those articles in the years 1989–2013. All articles prior to the year 2000 (one for WoS, three for Scopus) have been aggregated into one category, '<2000'; i.e., the first bin. Note that the numbers of citations are roughly 25 times higher than the numbers of articles.

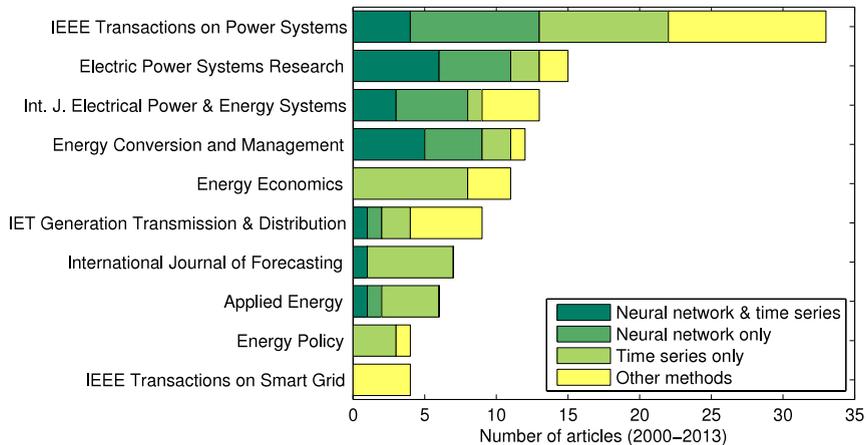


Fig. 3. The numbers of Scopus-indexed EPF articles published in the years 2000–2013 in the ten most popular journals. 'Neural network'-type models are more often published in electrical engineering journals, while statistical time series models tend to be published in *Energy Economics*, *International Journal of Forecasting*, *Applied Energy* and *Energy Policy*.

(AND ‘neural network’), and yielded 91 articles. Adding (OR ‘artificial intelligence’) or (OR ‘fuzzy’) increased the count by one to four articles only; neither of these modifications was used. The look-up for ‘statistical time series’ methods is more complicated, as there is no single most popular phrase. We used a logical search string which included most of the commonly used keywords or phrases in such articles.² The overall count obtained is 100 articles out of the 206 indexed by Scopus. If we consider four disjoint sets: (i) ‘neural network’ papers, (ii) ‘statistical time series’ papers, (iii) papers where both approaches were used and (iv) papers where neither ‘neural network’ nor ‘statistical time series’ methods were used, then the overall counts are 46, 55, 45 and 60, respectively. Looking again at Fig. 3, we can conclude that ‘neural network’-type models are published more often in electrical engineering journals, while statistical time series models tend to appear in *Energy Economics*, *International Journal of Forecasting*, *Applied Energy* (which is somewhat surprising, as this journal is classified by Scopus in the *Energy and Engineering* subject areas) and *Energy Policy*.

A probable reason for the latter situation is the difference between the educational training of electrical engineers and econometricians (statisticians), which constitute the two main groups of authors who submit papers to these two journal classes. In the late 1990s, computational intelligence (CI) methods – and neural networks in particular – were a ‘hot topic’ among engineers, and engineering faculties offered many such courses.³ On the other hand, the typical background of an electrical engineer educated in the 1990s and 2000s did not include much of statistics. The situation is quite the opposite among econometricians and statisticians. CI classes were not (and still are not) part of the typical curriculum. These differences in educational training have their consequences in the quality of the research. Typically, ‘electrical engineering’ papers consider sophisticated CI tools and relatively simple (or not properly applied) statistical models, and when the two are compared, the former tend to perform better. On the other hand, ‘econometric’ or ‘statistical’ papers usually show that (advanced) statistical models outperform (simple) CI techniques. In addition, given that electrical engineers typically have no training or experience in a statistically sound validation of the model performance, there is definitely room for improvement and closer cooperation between the two communities.

To end this section, let us comment briefly on the most popular outlets for proceedings papers. Definitely the number one are the numerous IEEE conferences (on *Power Engineering*, *Power Systems*, *Man & Cybernetics*, and *Neural Networks*). Next in line are *Lecture Notes in Computer Science*, the proceedings of the *European Electricity Market (EEM) Conference* and the proceedings of the *Chinese*

Control and Decision Conference. The overall count of Scopus-indexed conference papers is 274 (in the years 2000–2013), compared to 206 journal articles (in the years 1989–2013).

2.2. Major review and survey publications

The publication trends discussed in Section 2.1 suggest that electricity price forecasting has saturated the research community. On the other hand, the small numbers of books and review articles on this topic indicate that this research area is not very mature yet. To the best of our knowledge, there are essentially only three books which address EPF:

- **Shahidehpour et al. (2002, Chapter 3, pp. 57–113)** discuss the basics of electricity pricing and forecasting (price formation, volatility, exogenous variables), describe a price forecasting module based on neural networks, and comment on performance evaluation.
- **Weron (2006, Chapter 4, pp. 101–155)** provides an overview of modeling approaches, then concentrates on practical applications of statistical methods for day-ahead forecasting (ARMA-type, ARMAX, GARCH-type, regime-switching), discusses interval forecasts, and moves on to quantitative stochastic models for derivatives pricing (jump-diffusion models and Markov regime-switching).
- **Zareipour (2008, Chapters 3–4; pages 52–105 in the author’s Ph.D. Thesis from 2006, on which the book is based)** begins by reviewing linear time series models (ARIMA, ARX, ARMAX) and nonlinear models (regression splines, neural networks), then uses them for forecasting hourly prices in the Ontario power market.

There are a few more books which touch upon the topic of electricity price forecasting, but they generally concentrate on modeling the stochastic price dynamics for risk management and derivatives valuation, rather than on day-ahead price forecasting; see for example **Benth, Benth, and Koekebakker (2008)**; **Bunn (2004)**; **Burger, Graeber, and Schindlmayr (2007)**; **Eydeland and Wolyniec (2003)**; **Fiorenzani (2006)**; **Huisman (2009)**; **Keppeler, Bourbonnais and Girod (2007)**; **Lewis (2005)**, and **Weber (2006)**. There is also a recent monograph by **Yan and Chowdhury (2010a)**, based on the master’s thesis of the first author, but it considers only mid-term electricity price forecasting, with a time frame of between one and six months. Although mid-term EPF is important for resource reallocation, maintenance scheduling, bilateral contracting, budgeting and planning purposes, it is beyond the few hours to few days ahead forecasting horizons that are typically considered in the EPF literature.

Regarding review and survey articles, the situation looks a little better: the first review papers were already being published in the early 2000s. In an invited paper that appeared in the *Proceedings of the IEEE*, **Bunn (2000)** reviews some of the main methodological issues and techniques which are related to the forecasting of daily loads and prices in competitive power markets. He concludes that “the forecasting of loads and prices are mutually intertwined activities” and that game theory and the economic perspective cannot be “an accurate basis for daily

² The Scopus query given in footnote 1 was modified to include: AND (‘AR’ OR ‘ARMA’ OR ‘ARIMA’ OR ‘GARCH’ OR ‘VAR’ OR ‘time series model’ OR ‘regression’ OR ‘autoregressive’ OR ‘autoregression’ OR ‘volatility’).

³ Thanks to Tao Hong for pointing this out.

forecasts". He advocates the use of methods which involve variable segmentation (separate models for each load period), neural techniques (that are able to model the non-linear behavior) and forecast combinations. Surprisingly, the latter approach has been neglected – at least in the context of EPF – until very recently, see Section 4.3. In Chapter 5 of Skantze and Ilic (2001), the authors classify electricity price models and existing relevant publications into six groups, and discuss them very briefly in terms of their objectives, characteristics, advantages and disadvantages. One of the model classes mentioned is that of equilibrium (multi-agent; see Section 3.5) models, which have been reviewed more extensively in the Ph.D. thesis of Batlle (2002) and the review article of Ventosa, Baillo, Ramos, and Rivier (2005) in *Energy Policy*. Both of these publications discuss the different approaches to modeling strategic bidding behavior in power markets, including the Nash-Cournot framework and the supply function equilibrium approach. Although such models cannot generally provide accurate daily or hourly price forecasts, as was observed by Bunn (2000), there have been some attempts for the Spanish market (see e.g., Garcia-Alcalde, Ventosa, Rivier, Ramos, & Relanó, 2002).

In an *IEEE Power & Energy Magazine* discussion article on real-world market representation with agent-based models, Koritarov (2004) argues that the purpose of ABM is not necessarily to predict the outcome of a system; rather, it is to reveal and explain the complex and aggregate system behaviors that emerge from the interactions of the heterogeneous individual entities. At the same time, he concludes that the ABM approach is positioned well for performing short- and long-term electricity price forecasting, resource forecasting and asset valuation. Unfortunately, he does not provide any examples of EPF applications of ABM. Weidlich and Veit (2008) also fail to find any examples of EPF in a survey of agent-based wholesale electricity market models in *Energy Economics*.

In another *IEEE Power & Energy Magazine* discussion article, Amjady and Hemmati (2006) explain the need for short-term price forecasts, review problems related to EPF, and put forward proposals for such predictions. They argue that time series techniques (AR, ARIMA, GARCH) are generally only successful "in the areas where the frequency of the data is low, such as weekly patterns ...", which is contradicted by the empirical evidence presented in Section 3.8. Furthermore, they advocate the use of artificial (or computational) intelligence and hybrid approaches (neural networks, fuzzy regression, fuzzy neural networks, cascaded architecture of neural networks, and committee machines), which are "capable of tracking the hard nonlinear behaviors of hourly load and especially price signals". In a later publication, Amjady (2012, Chapter 4) briefly reviews EPF methods, then focuses again on artificial intelligence-based methods, and in particular feature selection techniques and hybrid forecast engines. He also discusses forecast error measures, the fine tuning of model parameters, and price spike predictions.

In the year 2009, two similar survey articles, co-authored by the same three researchers, appeared in parallel in the *International Journal of Electrical Power and Energy Systems* and the *International Journal of Energy Sector*

Management. Aggarwal, Saini, and Kumar (2009a) review 47 'time series' and 'neural network' papers published between 1997 and 2006 in terms of the model type and architecture, forecast horizon(s), model input and output variables, preprocessing and datasets used. They conclude that "there is no systematic evidence of out-performance of one model over the other models on a consistent basis", which may be attributed to "the large differences in price developments (...) in different power markets". In a more recent – in terms of the publications reviewed – article, Aggarwal, Saini, and Kumar (2009b) also compare 'time series' and 'neural network' papers. They classify EPF models as falling into one of three categories (although differently from Aggarwal et al., 2009a): heuristics (naïve, moving average), simulations (production cost and game theoretical) and statistical models, where the last category – somewhat surprisingly – includes both time series (regression) and artificial intelligence models. They expand the analysis to include quantitative comparisons of (i) the forecasting accuracy and (ii) the computational speed of different forecasting techniques. In our opinion, the value of (i) is disputable. Even if the forecasting accuracy is reported for the same market and the same out-of-sample (forecasting) test period, the errors of the individual methods are not truly comparable if different in-sample (calibration) periods are used. Moreover, the implementation of the algorithms differs between software packages, and is generally very sensitive to the initial conditions in the case of nonlinear or multi-parameter models. It may be impossible to replicate the results, even given the exact model structure, as was reported by Weron (2006) for the case of the multi-parameter transfer function (ARMAX) model of Nogales, Contreras, Conejo, and Espinola (2002). On the other hand, a table with the computation speeds of different forecasting techniques is interesting. Unfortunately, though, it cannot be used to draw quantitative conclusions, due to the differences in processors used, software implementations, calibration periods, etc. Finally, Aggarwal et al. (2009b) conclude that "there is no hard evidence of out-performance of one model over all other models on a consistent basis" and that longer "test periods of one to two years should be used". We cannot argue with these conclusions.

In a recent survey article published in the *IEEE Signal Processing Magazine*, Chan et al. (2012) review neural networks, support vector machines, time series models (ARMA, ARMAX, GARCH), and functional principal component analysis (FPCA) models for electricity prices/load, wind and solar forecasting. They advocate the use of multivariate factor models, and especially of the robust FPCA, which is shown to outperform both the standard FPCA and an AR model with a time varying mean in a limited forecasting study.

In a chapter in the *Wiley Encyclopedia of Electrical and Electronics Engineering*, Garcia-Martos and Conejo (2013) review short- and medium-term EPF, with a focus on time series models. Specifically, they consider ARIMA and seasonal ARIMA models calibrated to hourly prices for day-ahead predictions, and vector ARIMA (essentially VAR) and unobserved component (i.e., factor) models for

medium-term horizons. Sadly, in the most novel part on factor models, the authors limit the discussion to their own approach (García-Martos, Rodríguez, & Sanchez, 2011, 2012), and neither review nor compare other relevant publications (see Section 4.4). Interestingly, though, the chapter includes an introduction to the computation of prediction intervals, a topic which is addressed very rarely in the EPF literature.

In a short review article, Hong (2014) briefly discusses spatial load forecasting, short-term load forecasting, EPF, and two ‘smart grid era’ research areas: demand-response and renewable-generation forecasting. He classifies EPF models into three groups: simulation methods (which require a mathematical model of the electricity market, load forecasts, outage information, and bids from market participants), statistical methods, and AI methods. Perhaps the most important contribution of the paper is that the author emphasizes the need for rigorous out-of-sample testing of the different methods proposed in the literature. We will return to this issue in Section 4.5.

In the most recent survey of structural models, published as a chapter in the book *Quantitative Energy Finance*, Carmona and Coulon (2014) present a detailed analysis of the structural approach for electricity modeling, emphasizing its merits relative to traditional reduced-form models. Building on several recent articles, they advocate a broad and flexible structural framework for spot prices, incorporating demand, capacity and fuel prices in several ways, while calculating closed-form forward prices throughout.

The above-mentioned articles, book chapters and Ph.D. theses are complemented by a few survey conference papers of varying quality. Niimura (2006) studies over 100 papers and classifies them as either simulation models (production cost and game theoretical) or statistical models (which again include time series, regression, and artificial intelligence models). Hagi and Tafreshi (2007) construct a different classification in which they categorize ‘time series’ models as either ‘stationary’ (including ARIMA, ARIMA-Wavelet, ARX and ARMAX models) or ‘non-stationary’ (including neural networks, regime-switching models, GARCH, jump-diffusions and mean-reversion models). This is a very confusing classification, as some of the ‘stationary’ models are non-stationary in a statistical sense (for instance, ARIMA), while some of the ‘non-stationary’ models are stationary (for instance, mean-reversion models)! Daneshi and Daneshi (2008) consider over 100 papers and classify them as time series models, neural networks, fuzzy set models, fuzzy neural networks and other techniques. Similar in scope are the papers of Hu, Taylor, Wan, and Irving (2009) and Negnevitsky, Mandal, and Srivastava (2009), together with the more recent survey of Cerjan, Krzelj, Vidak, and Delimar (2013).

3. What and how are we forecasting?

3.1. The electricity ‘spot’ price

Unlike most other commodity or financial markets, the electricity ‘spot market’ is typically a *day-ahead market* that does not allow for continuous trading. This is a result

of system operators requiring advance notice in order to verify that the schedule is feasible and falls within transmission constraints. In a day-ahead market, agents submit their bids and offers for the delivery of electricity during each hour (or a shorter *load period*) of the next day before a certain market closing time, see Fig. 4. Thus, when dealing with the modeling and forecasting of intraday electricity prices, it is important to remember that, in most markets, prices for all contracts of the next day are determined at the same time using the same available information (Huisman, Huurman, & Mahieu, 2007; Peña, 2012).

The genuine role of an organized market for electricity (like a *power exchange* or a *power pool*) is to match the supply and demand of electricity so as to determine the *market clearing price* (MCP). Typically, the MCP is established in an auction, conducted once per day, as the intersection between the supply curve (constructed from aggregated supply bids) and the demand curve (constructed from aggregated demand bids) or the system operator estimated demand (in one-sided auction markets, like in Australia or Spain), for each of the load periods; see Fig. 5. Buy (sell) orders are accepted in order of increasing (decreasing) prices until the total demand (supply) is met. Note that bids with negative prices are allowed in many markets, potentially leading to negative prices when the demand is very low (the costs of shutting down and ramping up a power plant unit can exceed the loss from accepting negative prices) or the production from renewable sources is very high (most notably from wind), see e.g. Cutler, Boerema, MacGill, and Outhred (2011), Fanone, Gamba, and Prokopczuk (2013), Keles, Genoese, Möst, and Fichtner (2012). Recall that in a *uniform-price* (or *marginal*) auction market, buyers with bids above (or equal to) the MCP pay that price, and suppliers with offers below (or equal to) the MCP are paid the same price. Hence, on 3.1.2014, hour 18–19, a supplier would have been paid 30.94 EUR/MWh for the quantity sold in the day-ahead Elspot market at Nord Pool, regardless of his actual bid (and his marginal costs), as long as it was at or below 30.94 EUR/MWh; see the left panel in Fig. 5. In contrast, in a *pay-as-bid* (or *discriminatory*) auction, a supplier would be paid exactly the price he bid for the quantity transacted; in effect, he would be paid an amount that corresponds more closely to his marginal costs. Both approaches have both pros and cons, and the choice between them is not obvious. However, most market designs have adopted the uniform-price auction, with the UK under NETA being one of the few exceptions.

When there is no transmission congestion, the MCP is the only price for the entire system. However, when there is congestion, *locational marginal prices* (LMP) or *zonal clearing prices* differ from the system price and from each other. For smaller and medium-sized markets (like the German EEX, Polish GEE, Scandinavian Nord Pool or Spanish OMEL), the system price is usually established, but for larger markets (like the North American PJM), zonal prices or prices for major market hubs are computed. Interestingly, transmission congestion itself can be predicted in the short-term, as was shown by Løland, Ferkingstad, and Wilhelmssen (2012) for the South Norway (NO1) price area of the Nord Pool system.

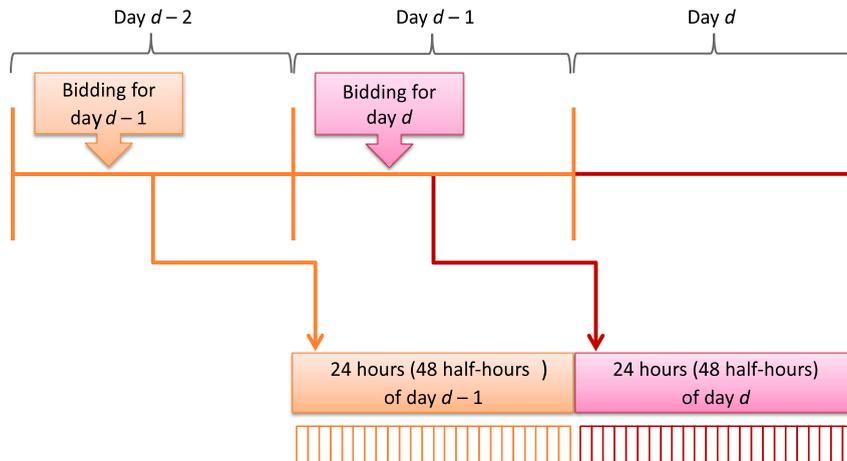


Fig. 4. The spot electricity market is typically a day-ahead auction market that does not allow for continuous trading. Before a certain market closing time on day $d - 1$, agents must submit their bids and offers for the delivery of electricity during each hour (or half-hour) of day d .

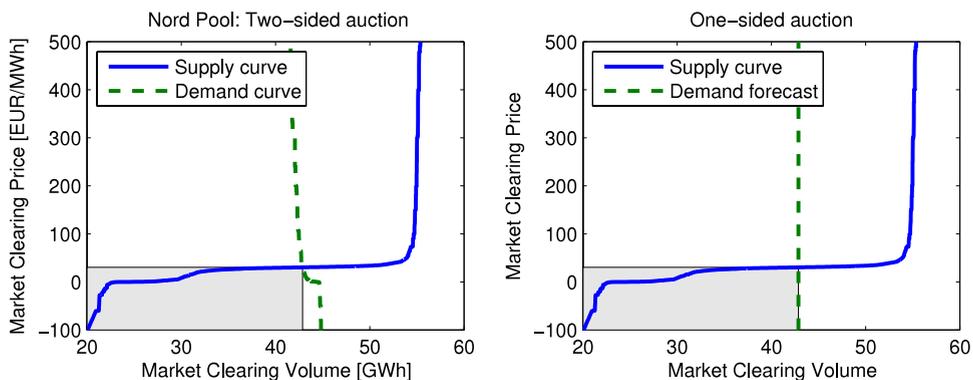


Fig. 5. *Left panel:* In a power exchange, like the Scandinavian Nord Pool, the market clearing price (MCP) is established through a two-sided auction as the intersection between the supply curve (constructed from aggregated supply bids) and the demand curve (constructed from aggregated demand bids). Here, the MCP is 30.94 EUR/MWh for Friday, 3.1.2014, hour 18–19. *Right panel:* A hypothetical market cross for a one-sided auction (power pool). Note that bids with negative prices are allowed in many markets, potentially leading to negative prices – a behavior which is not generally observed in other financial or commodity markets.

Nodal prices are the sum of generation marginal costs and transmission congestion costs, and can be different for different buses (or nodes), even within a local area. They are the ideal reference, because the electricity value is based on where it is generated and delivered. However, they generally lead to higher transaction costs and a greater complexity of the pricing mechanism (Weron, 2006). On the other hand, zonal prices may differ between different zones or areas, but are the same within a zone, i.e., a portion of the grid within which congestion is expected to occur infrequently or has relatively low congestion-management costs. Nodal (locational) pricing developed in the highly meshed North American networks, where transmission lines criss-cross the electricity system. In Australia, where the network structure is simpler, zonal pricing was implemented successfully. Although the European network is rather complex, it is evolving into a zonal market, often with an entire country constituting a zone.

For very short time horizons before delivery, the (transmission) system operator (TSO, SO) operates the

so-called *balancing* (or *real-time*) market. This technical market is used to price deviations in supply and demand from day-ahead or long-term contracts. The TSO needs to be able to call in extra production at very short notice, since the deviations must be corrected on a continuous basis in order to ensure system balance. It should be noted that the balancing market is not the only technical market. To minimize the reaction time in the case of deviations in supply and demand, the system operator runs an *ancillary services* market, which typically includes the down regulation service, the spinning and non-spinning reserve services, and the responsive reserve service. Day-ahead, balancing and ancillary services markets serve different purposes and are complementary. The modeling and forecasting of prices from the latter two markets is rather rare in the literature, but there are some exceptions. For instance, Ma, Luh, Kasiviswanathan, and Ni (2004) develop neural network models for forecasting real-time LMP before and after the day-ahead market is cleared, and test them using data from the PJM and New England markets; Olsson and Soder (2008)

build a model for balancing prices at Nord Pool using combined seasonal ARIMA and discrete Markov processes; and Czapaj, Tomasik, and Lubicki (2009) forecast balancing market and power exchange day-ahead prices jointly in Poland using a neural network. More recently, recognizing the fact that emerging smart grid technologies and the large-scale integration of variable resources into the grid have led to a growth of the market for ancillary services, Wang, Zareipour, and Rosehart (2014) investigate the application of reduced-form approaches (MRJD and MRS, see Section 3.7) for modeling the behaviors of operating reserve and regulation prices in the Ontario and New York markets. The patterns and characteristics of the prices of ancillary services differ considerably from those of day-ahead electricity prices, with the particular features of a lower price level, higher variability and more frequent and extreme spikes. The last feature in particular makes the prices for ancillary services more difficult to predict.

Some markets – like the Australian National Electricity Market (NEM) and the Ontario Electricity Market (OEM) – follow a single settlement real-time structure (Zareipour, 2008). In such a system, bids must be submitted to the market operator on the pre-dispatch day, but the volume can then be revised up to 5 (NEM) or 10 (OEM) minutes prior to the dispatch time without any restriction. The prices are set by the market operator each 5 min, and the spot prices are then determined in half-hourly (NEM) or hourly (OEM) trading intervals, as an average over the 5-min prices. As was pointed out by Higgs and Worthington (2008), Janczura, Trueck, Weron, and Wolff (2013) and Zareipour, Bhattacharya, and Canizares (2007), the Australian and Ontario electricity markets are significantly more volatile and spike-prone than most other markets. This has been confirmed by various short-term price and spike forecasting studies for the Australian (Amjady & Keynia, 2009; Becker, Hurn, & Pavlov, 2007; Christensen, Hurn, & Lindsay, 2012; Dong, Wang, Jiang, & Wu, 2011) and Ontario (Aggarwal, Saini, & Kumar, 2008; Lei & Feng, 2012; Mandal, Haque, Meng, Martinez, & Srivastava, 2012; Rodriguez & Anders, 2004) power markets. It is no surprise that Aggarwal et al. (2009b) conclude in their review paper that the accuracy levels achieved by the various models for day-ahead forecasts are higher than those achieved for real-time forecasts.

Finally, it should be noted that although we use the terms *spot* and *day-ahead* interchangeably here, the former need not necessarily refer to the day-ahead market. The European convention is to refer to the day-ahead price as the *spot price*. However, in the US, the term *spot price* is typically reserved for the intra-day real-time market, while the day-ahead price is called the *forward price* (see e.g. Longstaff & Wang, 2004). Nowadays, some markets in Europe (e.g., in the UK) also allow continuous trading for individual load periods, up to a few hours before delivery. With the shifting of volume from the day-ahead to balancing markets, the term *spot* is also being used more and more often in Europe to refer to the real-time market. The average of the 24 hourly (or 48 half-hourly) prices is called the *daily price*, the *daily spot price* or the *baseload price*. The average of prices for the on-peak hours (typically 8 am to 8 pm) is called the *peakload price*. These ‘daily’ price

conventions generally refer to day-ahead prices. In single settlement real-time markets, the averages are computed for real-time prices.

3.2. Forecasting horizons

It is customary to talk about short-, medium- and long-term electricity price forecasting, but there is no consensus in the literature as to what the thresholds should actually be. Short-term EPF generally involves forecasts from a few minutes up to a few days ahead, and is of prime importance in day-to-day market operations, as was discussed in Section 3.1. Medium-term time horizons, from a few days to a few months ahead, are generally preferred for balance sheet calculations, risk management and derivatives pricing. In many cases, evaluation is based not on the actual point forecasts, but on the distributions of prices over certain future time periods. As this type of modeling has a long-standing tradition in finance, an inflow of ‘finance solutions’ is observed readily (see Section 3.7). Finally, the main objective of long-term EPF – with lead times measured in months, quarters or even years – is investment profitability analysis and planning, such as determining the future sites or fuel sources of power plants. As Ventosa et al. (2005) remark, capacity-investment decisions are the main variables, and unit-commitment decisions are usually neglected in this context. While similar tools and techniques can be used for short- and medium-term horizons, long-term horizons generally require a totally different approach (which is beyond the scope of this review).

3.3. Evaluating point forecasts

The vast majority of EPF papers are concerned only with point forecasts (see Sections 4.2 and 4.5.2 for a discussion of interval and density forecasts). The most widely used measures of accuracy are those based on *absolute errors*: $AE_h = |P_h - \hat{P}_h|$, where P_h is the actual and \hat{P}_h the predicted price for load period h . In particular, for hourly point forecasts, the *daily/weekly mean absolute error* (MAE) is computed as the mean of $T = 24$ or 168 absolute errors. Since absolute errors are hard to compare between different datasets, many authors use measures based on *absolute percentage errors*: $APE_h = AE_h/P_h$. By far the most popular is the *mean absolute percentage error* (MAPE), which is computed as the mean of T absolute percentage errors. The MAPE measure works well in load forecasting, since load values are significantly higher than zero, but MAPE can be misleading when applied to electricity prices. In particular, when electricity prices are close to zero, MAPE values become very large, regardless of the actual absolute errors. On the other hand, when electricity prices spike, the resulting MAPE values are small, irrespective of the absolute differences. Moreover, for negative spot prices, they become negative and hard to interpret.

In a more general point forecasting context, Hyndman and Koehler (2006) compare a number of popular measures of accuracy and find them to be degenerate in commonly occurring situations. They advocate the use of *scaled errors* as a robust alternative to using percentage errors

when comparing forecast accuracies across series on different scales. For a non-seasonal time series, a scaled error uses one-step-ahead naïve forecasts (based on the most recent observation; $m = 1$ in Eq. (1)). However, for seasonal time series, a scaled error should be defined using seasonal naïve forecasts instead (Hyndman & Athanasopoulos, 2013). The resulting (seasonal) *mean absolute scaled error* is defined as:

$$\text{MASE}_{T,m} = \frac{1}{T} \sum_{h=1}^T \frac{|P_h - \hat{P}_h|}{\frac{1}{T-m} \sum_{h=m+1}^T |P_h - P_{h-m}|}, \quad (1)$$

where m is the length of the cycle; see also Section 3.8.1 for a discussion of similar-day forecasts in electricity markets. When working with hourly electricity prices, we can set $m = 24$ and $T = 168$ to obtain a *weekly MASE*. However, if we want to take the weekday/weekend effect into account, we have to set $m = 168$ and T significantly greater than 168. A scaled error has the nice interpretation that it is less than one if it arises from a better forecast than the average m -step-ahead naïve forecast computed in-sample; conversely, if the forecast is worse than the naïve forecast, it is greater than one.

Scaled errors have not been used extensively in energy economics thus far. To the best of our knowledge, only Garcia-Ascanio and Mate (2010) and Jonsson, Pinson, Nielsen, Madsen, and Nielsen (2013) utilize absolute or squared scaled errors in the EPF context. Alternative normalizations have been proposed instead, see for example Misiorek, Trück, and Weron (2006); Nogales and Conejo (2006); Shahidehpour et al. (2002); Weron and Misiorek (2008), and the references in the paragraphs below. Probably the most common approach is to normalize the absolute error by the average price obtained in the evaluation interval (e.g. a day, a week). This yields the *daily- or weekly-weighted mean absolute errors* (DMAE, WMAE; also known as the *mean daily/weekly errors*, MDE, MWE):

$$\begin{aligned} \text{DMAE}_{(T=24)}, \text{WMAE}_{(T=168)} &= \frac{1}{\bar{P}_T} \text{MAE}_T \\ &= \frac{1}{T} \sum_{h=1}^T \frac{|P_h - \hat{P}_h|}{\bar{P}_T}, \end{aligned} \quad (2)$$

where $\bar{P}_T = \frac{1}{T} \sum_{h=1}^T P_h$ is the mean price in the time interval T .

Apart from l^1 -type norms, square or l^2 -type norms are also used, usually in the more 'econometric' papers. Perhaps the most popular are the daily and weekly *root mean square errors* (RMSE; sometimes denoted by DRMSE and WRMSE, see e.g. Weron, 2006), calculated as the square root of the average of squared differences between the predicted and actual prices:

$$\text{RMSE}_{(T=24 \text{ or } 168)} = \sqrt{\frac{1}{T} \sum_{h=1}^T (P_h - \hat{P}_h)^2}. \quad (3)$$

Like in the absolute error-based measures, the squared differences $(P_h - \hat{P}_h)^2$ in the above formula can also be

normalized by the square of the current actual price to yield the *root mean square percentage error* (RMSPE; see Hyndman & Koehler, 2006), or by the square of the mean daily (or weekly) price to yield the *daily- or weekly-weighted root mean square errors* (DRMSE, WRMSE), or by $\frac{1}{T-24} \sum_{h=25}^T (P_h - P_{h-24})^2$ to yield the (seasonal) *root mean square scaled error* (RMSSE; see Jonsson et al., 2013).

Finally, we have to note that there is no 'industry standard', and the error benchmarks used in the literature vary a lot. As Weron (2006) observes, this may lead to confusion, since the names are not used consistently. For instance, Contreras, Espínola, Nogales, and Conejo (2003); Garcia, Contreras, van Akkeren, and Garcia (2005) and Nogales et al. (2002) define the 'mean weekly error' as the weekly MAPE (literally, as the average of the seven daily 'average prediction errors', i.e., daily MAPE values), while Conejo, Contreras, Espinola, and Plazas (2005) and Conejo, Plazas, Espínola, and Molina (2005) use Eq. (2) with $T = 168$. Likewise, in the latter three papers, the weekly RMSE, denoted by $\sqrt{\text{FMSE}}$, is computed using Eq. (3) with $T = 168$, while in the former two articles the normalization by $\sqrt{1/168}$ is missing. As a result, laborious multi-paper comparisons, like that performed by Aggarwal et al. (2009b), have to be treated with caution and a dose of skepticism. In particular, neither Conejo, Contreras et al. (2005) nor Conejo, Plazas et al. (2005) use the MAPE measure, as was suggested by Aggarwal et al. in their Tables III and IV.

3.4. Overview of modeling approaches

Nearly all of the review and survey publications discussed in Section 2.2 offer their own classifications of the various approaches that have been developed for analyzing and predicting electricity prices. Some of them are better, some are worse, but all have many things in common. Without loss of generality, we take the classification of Weron (2006) as a starting point, with six groups of models. We then alter it by combining the first two groups into one larger class (due to the decreasing popularity of production-cost models and the increasing use of simulation models):

- *Multi-agent (multi-agent simulation, equilibrium, game theoretic)* models, which simulate the operation of a system of heterogeneous agents (generating units, companies) interacting with each other, and build the price process by matching the demand and supply in the market.
- *Fundamental (structural)* methods, which describe the price dynamics by modeling the impacts of important physical and economic factors on the price of electricity.
- *Reduced-form (quantitative, stochastic)* models, which characterize the statistical properties of electricity prices over time, with the ultimate objective of derivatives evaluation and risk management.
- *Statistical (econometric, technical analysis)* approaches, which are either direct applications of the statistical techniques of load forecasting or power market implementations of econometric models.

- *Computational intelligence* (artificial intelligence-based, non-parametric, non-linear statistical) techniques, which combine elements of learning, evolution and fuzziness to create approaches that are capable of adapting to complex dynamic systems, and may be regarded as 'intelligent' in this sense.

Finally, we should mention that many of the modeling and price forecasting approaches considered in the literature are *hybrid* solutions, combining techniques from two or more of the groups listed above. Their classification is non-trivial, if indeed it is even possible. We illustrate the proposed taxonomy in Fig. 6. The main model types will be reviewed in Sections 3.5–3.9.

3.5. Multi-agent models

Forecasting wholesale electricity prices used to be a straightforward, though laborious, task. It generally concerned medium- and long-term time horizons, and involved matching demand estimates to the supply, obtained by stacking up existing and planned generation units in order of their operating costs. These *cost-based* models (*production-cost models*, PCM) had the capability to forecast prices on an hour-by-hour, bus-by-bus level (see for example Wood & Wollenberg, 1996, for a comprehensive discussion). However, they ignored strategic bidding practices, including the execution of market power. They were appropriate for regulated markets with little price uncertainty, a stable structure and no gaming, but are not suitable for competitive electricity markets. *Equilibrium (game theoretic)* approaches may be viewed as generalizations of cost-based models, amended with strategic bidding considerations. These models are especially useful in predicting expected price levels in markets with no price history, but known supply costs and market concentration. On the other hand, the increasingly popular adaptive *agent-based simulation* techniques can address features of electricity markets that static equilibrium models ignore.

In an excellent review paper, Ventosa et al. (2005) identify three main electricity market modeling trends: optimization, equilibrium and simulation models. In their classification, optimization models focus on the profit maximization problem for one of the firms competing in the market. As such, they are not useful in the EPF context, and will not be reviewed here. The equilibrium models discussed below (Nash-Cournot framework, supply function equilibrium) represent the overall market behavior, taking into consideration competition among all participants. Finally, simulation models are an alternative to equilibrium models when the problem under consideration is too complex to be addressed within a formal equilibrium framework. Since the equilibrium and simulation models defined by Ventosa et al. share many common features, we have decided to consider them jointly in one wide *multi-agent* class.

3.5.1. Nash-Cournot framework

In the *Nash-Cournot framework*, electricity is treated as a homogeneous good, and the market equilibrium is deter-

mined through the capacity setting decisions of the suppliers. Unfortunately, these models tend to provide prices higher than those observed in reality. Researchers have addressed this problem by introducing the concept of *conjectural variations*, see for example Day, Hobbs, and Pang (2002), Garcia-Alcalde et al. (2002) and Vives (1999), which aims to represent the fact that rivals react to high electricity prices by producing more. For sample applications of the Nash-Cournot framework, see Borenstein, Bushnell, and Knittel (1999); Cabero et al. (2005); Rubin and Babcock (2013) and Sapio and Wyłomańska (2008). Although their approach is hybrid in nature, Ruibal and Mazumdar (2008) provide one of the very few applications of this framework to EPF. A fundamental bid-based stochastic model is proposed for predicting electricity hourly prices and average prices in a given period. Two sources of uncertainty are considered: the availability of the generating units and demand. The results show that as the number of firms in the market decreases, the expected values of prices increase by a significant amount. The variances for the Cournot model also increase, but those for the SFE model (see Section 3.5.2) decrease. Ruibal and Mazumdar also demonstrate that an accurate temperature forecast can reduce the prediction error of the electricity price forecasts significantly.

3.5.2. Supply function equilibrium

The second approach models the price as the equilibrium of companies bidding with supply (and possibly demand) curves into the wholesale market. Calculating the *supply function equilibrium* (SFE) requires a set of differential equations to be solved, rather than the typical set of algebraic equations that arises in the Nash-Cournot framework. Thus, these models have considerable limitations concerning their numerical tractability. To speed up computations, the demand can be aggregated into blocks. This in turn leaves the extreme values out of the analysis, which we are not prepared to accept when focusing on EPF or risk management. Furthermore, as Bolle (2001) emphasizes, supply curve bidding will only lead to results which differ from the Nash-Cournot equilibrium if the demand uncertainty (or another source of uncertainty) leads to an ex-ante undetermined equilibrium. Otherwise, the supply bidding collapses to one point, which corresponds to the Nash-Cournot equilibrium.

For decreasing the numerical complexity of general SFE models, linear SFE models have been proposed. In such models, the demand is 'linear' (or, more precisely, 'affine'; at each moment in time the demand as a function of price has a non-zero intercept and a constant negative slope, see Baldick, Grant, & Kahn, 2004), marginal costs are linear or affine, and SFE can be obtained in terms of either linear or affine supply functions. The market clearing condition, yielding the price at time t , is

$$\sum_{j=1}^m q_j(p_t) = D_t,$$

assuming that a solution exists. The bid curve $q_j : [P_{min}, P_{max}] \rightarrow [0, U_j]$ is defined by

$$q_j = q_j(p_t) = \beta_j(p_t - \alpha_j),$$

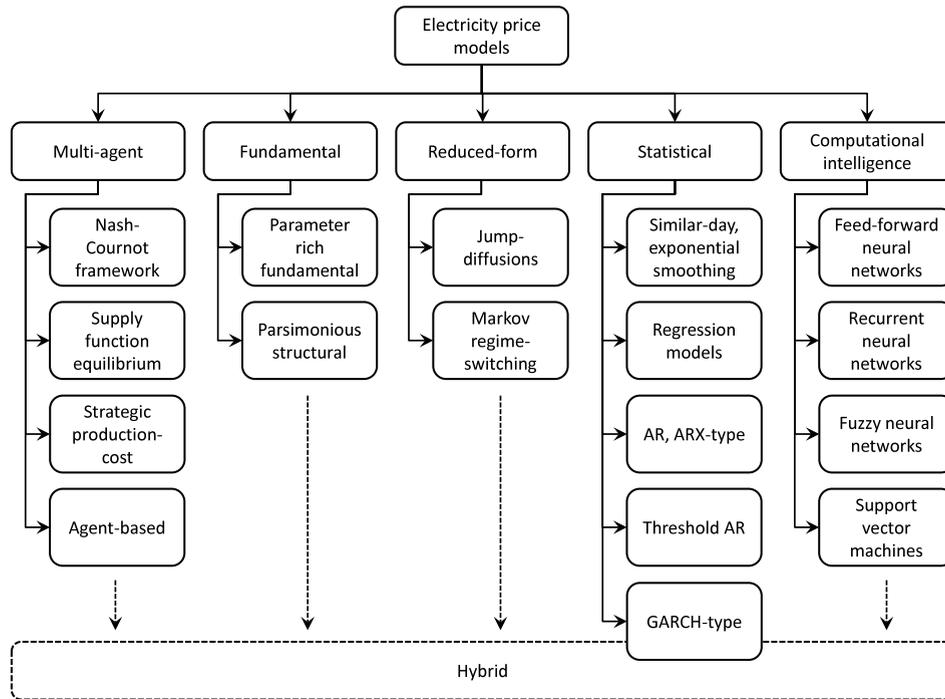


Fig. 6. A taxonomy of electricity spot price modeling approaches. The main model types are reviewed in Sections 3.5–3.9, with a special emphasis on their forecasting capabilities.

where α_j is the intercept, β_j is the slope of the supply function for the j th firm, U_j is the generation capacity for this firm, and the system demand curve $D(p_t)$ is assumed to be ‘linear’ in p_t . All firms receive the marginal clearing price for their supply. Since the supply functions are non-decreasing and the market clearing price is the same for all players, this market clearing condition maximizes the (revealed) social welfare when there is no transmission congestion. This framework has been used extensively for the analysis of bidding strategies (Borgosz-Koczwara, Weron, & Wyłomańska, 2009; Niu, Baldick, & Zhu, 2005), market power and market design (Baldick et al., 2004; Holmberg, Newbery, & Ralph, 2013), and congestion management (Hobbs, Metzler, & Pang, 2000); but electricity price forecasting applications have been very limited (see e.g. Ruibal & Mazumdar, 2008).

3.5.3. Strategic production-cost models

A third, less popular static equilibrium approach has been proposed by Batlle (2002) and Batlle and Barquín (2005) as a modification of the traditional production-cost models. The *strategic* PCM (SPCM) takes agents’ bidding strategies into account, based on conjectural variation. Each agent tries to maximize its own profits, taking into account its cost structures and the expected behaviors of its competitors, modeled through a strategic parameter, which represents the slope of the residual demand function for each production level of the generator. When simulating the supply curve building process, the SPCM assumes that the firm just knows its costs and its conjecture about the derivative of its residual demand function. As no iterations are made, firms do not have the

chance to refine their bids and take into account rivals’ reactions (as in SFE models). Compared with the Nash-Cournot and SFE models, the main advantage of the SPCM is its computational speed, which makes it suitable for real-time analysis.

3.5.4. Agent-based simulation models

The static equilibrium models discussed above are based on a formal definition of equilibrium, expressed in the form of a system of algebraic or differential equations. Even if the set of equations has a solution, it is often very hard to find, and the modeler has to resort to heuristics to ‘solve’ the problem (Day et al., 2002; Ventosa et al., 2005). Moreover, such modeling approaches have limitations in the way in which the competition between participants can be represented. On the other hand, agent-based simulation models do not have these limitations, while being not much harder to ‘solve’.

Over the last two decades, *agent-based computational economics* (ACE) has become a widely accepted approach to solving both theoretical and practical problems in energy economics (see e.g. Guerci, Rastegar, & Cincotti, 2010; Kowalska-Pyzalska, Maciejowska, Suszczyński, Sznajd-Weron, & Weron, 2014; Sun & Tesfatsion, 2007; Weidlich & Veit, 2008). The basic tool of ACE – an *agent-based model* (ABM; sometimes referred to as a *multi-agent system* or a *multi-agent simulation*) – is a class of computational structures and rules for simulating the actions and interactions of autonomous agents (whether individuals or collective entities, such as organizations or groups), with the ultimate objective being to assess their effects on the system as a whole.

One of the first applications of ACE to modeling the strategic behavior observed in electricity markets was described in the paper by [Bower and Bunn \(2000\)](#), who test a number of market designs which are relevant for the changes that have taken place in the England and Wales market. They conclude that daily bidding, together with uniform pricing, yields the lowest prices, while hourly bidding under the pay-as-bid system yields the highest prices. In a similar context, [Day and Bunn \(2001\)](#) propose a simulation model for analyzing the potential for market power. This agent-free simulation approach is similar to the SFE scheme, but it provides a more flexible framework that allows for a consideration of actual marginal cost data and asymmetric firms.

In a review article, [Koritarov \(2004\)](#) argues that the purpose of ABM is not necessarily to predict the outcome of a system, but rather to reveal and explain the complex and aggregate system behaviors that emerge from the interactions of the heterogeneous agents. Indeed, if the Scopus query given in footnote 1 is appended with AND (‘‘agent-based’’ OR ‘‘multi-agent’’), it yields five publications, only three of which are related to EPF. This did not prevent [Koritarov](#) from concluding that the ABM approach is positioned well for the performance of short- and long-term electricity price forecasting. Perhaps with the development of more powerful processors and cloud computing, ABM will someday provide efficient tools for EPF.

Currently, ABM are merely elements of complex hybrid EPF systems, rather than being the source of price forecasts themselves. For instance, [Gao, Bompard, Napoli, and Zhou \(2008\)](#) present a monitoring system which consists of two units: a price forecast module, which delivers input variables to the multi-agent market simulator. The two units cooperate to build a monitoring system for predicting future power market scenarios and to deliver market clearing and production schedule information. [Guerci, Ivaldi, and Cincotti \(2008\)](#) develop an artificial power exchange, called the Genoa market, and are able to obtain simulated price trajectories with properties observed for peak- and off-peak prices in the Italian market. However, they do not focus on forecasting. Similar in spirit is the work by [Jabłońska and Kauranne \(2011\)](#), who build two multi-agent models based on a Capasso-Morale-type population dynamics approach and use them to reproduce the statistical features of Nord Pool spot prices.

[Chatzidimitriou, Chrysopoulos, Symeonidis, and Mitkas \(2012\)](#) use Cassandra, a dynamic platform for the development of multi-agent systems, to generate load and price predictions for the day-ahead market in Greece. They propose a hybrid scheme in which autonomously adaptive recurrent neural networks (see Section 3.9.3) are encapsulated into Cassandra agents. [Sousa, Pinto, Vale, Praca, and Morais \(2012\)](#) present another hybrid ABM-based method that aims to provide market players with strategic bidding capabilities, thus allowing them to achieve the highest possible gains in the market. Their method uses a neural network as an auxiliary forecasting tool for predicting electricity market prices. Through the analysis of prediction error patterns, the simulation method predicts the expected error for the next forecast, and uses it to adapt

the actual forecast. In a very recent paper, [Ladjici, Tiguercha, and Boudour \(2014\)](#) investigate the use of competitive co-evolutionary algorithms to calculate suppliers' optimal strategies in a deregulated electricity market. In their model, agents can take part in both spot and forward transactions, and act strategically in order to maximize their overall profit. The strategic interactions of market agents are modeled as a non-cooperative game, and a competitive co-evolutionary algorithm is used to calculate the Nash equilibrium strategies, thus ensuring the best outcome for each agent.

3.5.5. Strengths and weaknesses

On the one hand, multi-agent models – and agent-based models in particular – are a class of extremely flexible tools for the analysis of strategic behavior in electricity markets. On the other hand, this freedom is also a weakness, as it requires the assumptions embedded in the simulation to be justified, both theoretically and empirically. A number of components have to be defined: the players, their potential strategies, the ways in which they interact, and the set of payoffs. Obviously, a substantial modeling risk is present. While in classical power pools the sellers are generators, and their characteristics are identifiable through their assets directly, in power exchanges every type of market participant can be a seller. For instance, a distribution company that has over-contracted in the bilateral market can be a seller in the power exchange's spot market. Thus, the problem of identifying the relevant market players and their strategies becomes highly nontrivial.

Moreover, despite the few forecasting applications discussed above, multi-agent models generally focus on qualitative issues rather than quantitative results. They may provide insights as to whether or not prices will be above marginal costs, and how this might influence the players' outcomes. However, they pose problems if more quantitative conclusions have to be drawn, particularly if electricity prices have to be predicted with a high level of precision.

3.6. Fundamental models

The next class of models, known as *fundamental* or *structural* models, tries to capture the basic physical and economic relationships which are present in the production and trading of electricity. The functional associations between fundamental drivers (loads, weather conditions, system parameters, etc.) are postulated, and the fundamental inputs are modeled and predicted independently, often via statistical, reduced-form or computational intelligence techniques. Moreover, many of the EPF approaches considered in the literature are hybrid solutions with time series, regression and neural network models using fundamental factors – like loads, fuel prices, wind power or temperature – as input variables, see e.g. [Gonzalez, Contreras, and Bunn \(2012\)](#); [Karakatsani and Bunn \(2008\)](#); [Kristiansen \(2012\)](#); [Liebl \(2013\)](#); and [Weron and Misiorek \(2008\)](#). In general, two subclasses of fundamental models can be identified: parameter rich models and parsimonious structural models of supply and demand. For a very good introduction to the ‘fundamentals behind fundamental models’, we refer to [Burger et al. \(2007, Chapter 4\)](#).

3.6.1. Parameter-rich fundamental models

Models from the first subclass are often developed as proprietary, in-house products, and therefore, their details are not disclosed publicly. Most of the results published relate to hydro-dominant power markets. In particular, [Johnsen \(2001\)](#) presents a supply–demand model for the Norwegian power market from a time before the common Nordic market had started. He uses hydro inflow, snow and temperature conditions to explain spot price formation. [Eydeland and Wolyniec \(2003\)](#) develop a hybrid fundamental model and calibrate it to data from ERCOT, NYPOOL and PJM. They start with the processes for the primary drivers (such as fuels, outages and temperature/demand), then construct the bid stack transformation and obtain electricity prices. The simulated price processes exhibit spikes, mean reversion, fat tails of the price distributions, and a correct forward price volatility structure.

[Vahviläinen and Pyykkönen \(2005\)](#) build an even more parameter-rich fundamental model for the Nordic market. Considering stochastic climate factors like temperature and precipitation, they model the hydrological inflow and snow-pack development that affect hydro power generation, the major source of electricity in Scandinavia. Using 27 scalar parameters (13 climate, 4 demand and 10 supply parameters) and 29 formulas defining the relationships between the fundamental variables, they arrive at the spot price formula: the production volume weighted average of the supply price of condensing power and the supply price of hydro-power. The weight is a sum of the amount of condensing production and the amount of regulated hydro-production. [Vahviläinen and Pyykkönen](#) show that their model is able to capture the observed fundamentally motivated market price movements on a monthly scale.

3.6.2. Parsimonious structural models

The subclass of much simpler structural models can be traced back to [Barlow \(2002\)](#). Starting from an empirical analysis of market supply and demand curves, he builds the spot price process by applying the inverse of the Box–Cox transformation (which includes an exponential function as a special case) to an Ornstein–Uhlenbeck process, see Eq. (5) below. As a result, [Barlow](#) obtains a jumpless spot price model which can exhibit spikes, and calibrates it to data from the Alberta and California markets.

In the same spirit, [Kanamura and Ōhashi \(2007\)](#) define a hockey-stick shaped supply curve (see Fig. 7) that matches the empirically observed curves better than the inverse of the Box–Cox transformation:

$$P_t = f(S_t) = \begin{cases} \alpha_1 + \beta_1 D_t & \text{for } D_t \leq z - s, \\ a + bD_t + cD_t^2 & \text{for } D_t \in (z - s, z + s), \\ \alpha_2 + \beta_2 D_t & \text{for } D_t \geq z + s, \end{cases} \quad (4)$$

where z is the mid-point of the domain of the quadratic curve stretched between $z - s$ and $z + s$, $\alpha_{1,2}$ and $\beta_{1,2}$ are the intercepts and slopes, respectively, of the linear parts of the supply curve (to the left and right of the quadratic regime), and a , b and c are the coefficients of the quadratic curve.

Then, combine this with an inelastic vertical demand curve with horizontal stochastic deviations $X_t = D_t - \bar{D}_t$ driven by a mean-reverting process of the form:

$$dX_t = (\alpha - \beta X_t)dt + \sigma dW_t, \quad (5)$$

where β is the speed of mean-reversion, $\frac{\alpha}{\beta}$ is the long term mean-reversion level, σ is the volatility and dW_t are the increments of a standard Wiener process (i.e., Brownian motion). The above stochastic differential equation is known in mathematics as the Ornstein–Uhlenbeck process, and was introduced to finance by [Vasicek \(1977\)](#), originally for modeling interest rate dynamics. It is the backbone of all reduced-form models for commodity prices, see Section 3.7. [Kanamura and Ōhashi](#) fit their model to PJM price and demand data, and show that it can generate electricity price spikes (see the bottom panel of Fig. 7), and fits the observed data better than a jump-diffusion model (see Section 3.7.1). This is mainly because this simple structural model incorporates the sudden and large increase in the slope of the supply curve by using a hockey-stick shaped function. In the second part of the paper, the authors then use it to model the optimal operation policy for a pumped-storage hydropower generator. In a follow-up paper, [Kanamura and Ōhashi \(2008\)](#) use this model to show that the transition probabilities of electricity prices cannot be constant, and depend on both the current demand level relative to the supply capacity and the trends of demand fluctuation. Independently, [Boogert and Dupont \(2008\)](#) use a similar supply–demand framework to model the hourly day-ahead price of electricity in the Dutch APX market, and are quite successful at predicting spot price movements 24 h ahead. One of their main findings is that the reserve margin should be included in a spot electricity model in order to enhance the performance, see also the discussion in Section 4.1.2.

[Coulon and Howison \(2009\)](#) develop a fundamental model for spot electricity prices, based on stochastic processes for the underlying factors (fuel prices, power demand and generation capacity availability), as well as a parametric form for the bid stack function that maps these price drivers to the power price. Using observed bid data, they find high correlations between the movements of bids and the corresponding fuel prices. Using a stochastic model of the bid stack, [Carmona, Coulon, and Schwarz \(2013\)](#) translate the demand for power and the prices of generating fuels into electricity spot prices. The stack structure allows for a range of generator efficiencies per fuel type and for the possibility of future changes in the merit order of the fuels. The derived spot price process captures important stylized facts of historical electricity prices, including both spikes and the complex dependence upon its underlying supply and demand drivers. Furthermore, under mild assumptions on the distributions of the input factors, they obtain closed-form formulas for electricity forward contracts and for spark and dark spread options. In a similar context, [Aid, Campi, and Langrené \(2013\)](#) develop a structural risk-neutral model in which a scarcity function is introduced to allow for deviations of the spot price from the marginal fuel price, thus leading to price spikes. Like [Carmona et al. \(2013\)](#), they focus on pricing and hedging electricity derivatives, and show that, when far from delivery, electricity futures behave like a basket of futures on fuels.

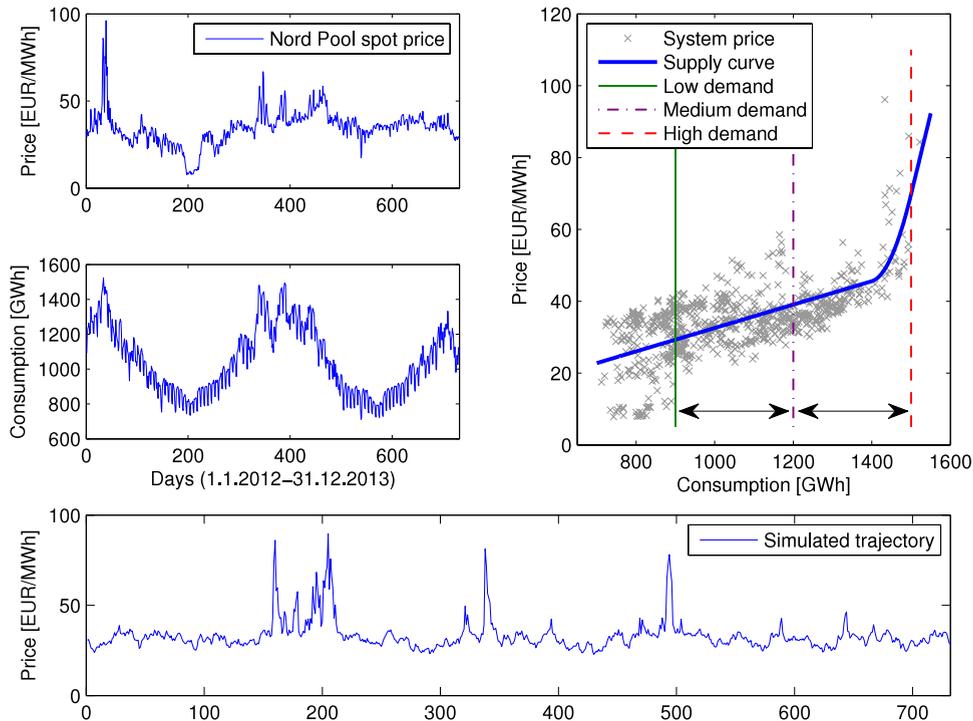


Fig. 7. Daily Nord Pool spot price (top left panel) and consumption (mid left panel) in the Nordic region (Denmark, Finland, Norway, Sweden) over the period 1.1.2012–31.12.2013. The Kanamura and Ōhashi (2007) supply function, see Eq. (4), is fitted here to the Nordic consumption–price data (top right panel) and combined with a stochastic mean-reverting demand level, see Eq. (5), yielding a relatively spiky simulated spot price trajectory (bottom panel). Note that the latter is a pure stochastic spot price component, without either weekly seasonality or the long-term cyclic component.

3.6.3. Strengths and weaknesses

Two major challenges arise in the practical implementation of fundamental models. The first one is data availability. Depending on the market, more or less information on plant capacities and costs, demand patterns and transmission capacities is available to the researcher or practitioner for constructing such a model. Because of the nature of fundamental data (which is often collected over longer time intervals, such as weekly or monthly), pure fundamental models are more suitable for medium-term predictions than short-term. This is also true for the parsimonious structural models. They are typically calibrated to daily data, and ignore the fine relationships at the hourly resolution. Their application, like that of the reduced-form models (see Section 3.7), is generally limited to risk management and derivatives pricing. In fact, they can be seen as direct competitors of the former, allowing for a better description of the market fundamentals, though at the cost of an increased complexity of the analytical calculations and calibration procedures. For an extended discussion, see the very recent review by Carmona and Coulon (2014).

The second challenge is the incorporation of stochastic fluctuations of the fundamental drivers. In building the model, we make specific assumptions about physical and economic relationships in the marketplace, and therefore the price projections generated by the models are very sensitive to violations of these assumptions. Moreover, the more detailed the model is, the more effort is needed to adjust the parameters. Consequently, there

exists a significant modeling risk in the application of the fundamental approach.

3.7. Reduced-form models

A common feature of the finance-inspired *reduced-form* (quantitative, stochastic) models of price dynamics is that their main intention is not to provide accurate hourly price forecasts, but rather to replicate the main characteristics of daily electricity prices, like marginal distributions at future time points, price dynamics, and correlations between commodity prices. Such models lie at the heart of derivatives pricing and risk management systems. If the price process chosen is not appropriate for capturing the main properties of electricity prices, the results from the model are likely to be unreliable. At the same time, if the model is too complex, the computational burden will prevent its use on-line in trading departments (Weron, 2006). On the one hand, the tools that are applied have their roots in methods that have been developed for modeling other energy commodities or interest rates (because of the mean-reversion property; see e.g. Burger et al., 2007); on the other hand, they integrate actuarial (claim arrival processes; see e.g. Čížek, Härdle & Weron, 2011) or econometric (abrupt changes in prices; see e.g. Hamilton, 2008) mechanisms. In a way, the jump-diffusion models that are reviewed next and the Markov regime-switching models discussed in Section 3.7.2 offer the best of the two worlds: they are trade-offs between

model parsimony and adequacy to capture the unique characteristics of electricity prices.

Depending on the type of market under consideration, stochastic techniques can be divided into two main classes: spot and forward price models. The former provide a proper representation of the dynamics of spot prices, which, in the wake of the deregulation of power markets, becomes a necessary tool for trading purposes. Their main drawback is the problem of pricing derivatives, i.e., the identification of the risk premium linking spot and forward prices (or those of other derivatives); for a discussion, see the recent review by Weron and Zator (2014a). On the other hand, forward price models allow for the pricing of derivatives in a straightforward manner (but only of those written on the forward price of electricity). However, they too have their limitations; most importantly, the lack of data that can be used for calibration and the inability to derive the properties of spot prices from the analysis of forward curves. In this review, we focus on spot price models. Forward price models are the domain of mathematical finance, and we refer to Benth et al. (2008) and Eydeland and Wolyniec (2003) for extended discussions. As Borak and Weron (2008) and Fleten and Lemming (2003) show, constructing smooth forward price curves in electricity markets can be a tedious and challenging exercise; however, the benefits of doing it are the readily available medium-term price forecasts for multiple horizons. These forecasts can be biased, though, and include the risk premium. Care should be taken when using them, see for example Gjolberg and Brattested (2011), Kristiansen (2007) and Ronn and Wimschulte (2009).

3.7.1. Jump-diffusion models

The various jump-diffusion models found in the energy economics literature can be obtained as special cases of the following general stochastic differential equation (SDE) for the increment of the (deseasonalized and detrended) spot electricity price X_t :

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t + dq(X_t, t), \quad (6)$$

where dW_t are the increments of a standard Wiener process (i.e., Brownian motion) and $dq(X_t, t)$ are the increments of a pure jump process.

If the drift term is such that it forces mean reversion to a stochastic or deterministic long-term mean at a constant rate, then the resulting process is called a *mean-reverting jump-diffusion* (MRJD). Quite often, the drift takes the following form: $\mu(X_t, t) = (\alpha - \beta X_t)$ (i.e., as in Eq. (5)), but other specifications are also used. For instance, Cartea and Figueroa (2005) use a geometric MRJD process where $\mu(X_t, t) = \beta(\rho(t) - \ln X_t)X_t$ and $\rho(t)$ is a time-dependent mean reverting level – a function of a deterministic sinusoidal seasonality and the time-dependent volatility $\sigma(t)$. In one of the first publications on modeling electricity spot prices, Kaminski (1997) utilizes Merton's jump-diffusion model, which is a combination of a geometric Brownian motion (GBM; i.e., with $\mu(X_t, t) = \mu X_t$ and $\sigma(X_t, t) = \sigma X_t$) and a jump process. Its main drawback is that it ignores mean-reversion to the 'normal' price regime. If a price spike occurred, GBM would 'assume' that the

new price level is a normal event, and would proceed randomly via a continuous diffusion process, dW_t , with no consideration of prior price levels, and only a small chance of returning to the pre-spike level. More recently, Albanese, Lo, and Tompaidis (2012) present a numerical algorithm for pricing derivatives on electricity prices, and study its rate of convergence for the case of the Merton jump-diffusion model. However, they then use the algorithm to calculate the prices and sensitivities of both European and Bermudan electricity derivatives within the more realistic jump-diffusion model of Geman and Roncoroni (2006).

For the sake of simplicity, the volatility term $\sigma(X_t, t)$ is usually set to a constant. However, the empirical evidence suggests that electricity prices exhibit heteroskedasticity (Bhar, Colwell, & Xiao, 2013; Karakatsani & Bunn, 2010; Keles et al., 2012). To circumvent this, inspired by the interest rate modeling literature, Janczura and Weron (2009) utilize the square root process of Cox, Ingersoll, and Ross (1985), while Janczura and Weron (2010) use a more general form of the volatility term: $\sigma(X_t, t) = \sigma X_t^\gamma$, with γ being a scalar parameter of the model. On the other hand, Cartea and Figueroa (2005) use a time-dependent volatility $\sigma(t)$ in their geometric MRJD model.

The process $q(X_t, t)$ is a pure jump process (typically independent of W_t) with a given intensity and severity, e.g., a compound Poisson process (Čížek, Härdle & Weron, 2011). For the sake of simplicity, one often sets $q(X_t, t) = Jdq(t)$, where J is a normal or log-normal random variable and $dq(t)$ are increments of a homogeneous Poisson process (HPP) with constant intensity λ . However, the empirical data suggest that the HPP may not be the best choice for the jump component. Price spikes are seasonal; they typically show up in higher-price seasons, like winter in Scandinavia and summer in the central US. Using a non-homogeneous Poisson process (NHPP) with a (deterministic) periodic intensity function $\lambda(t)$ may be more reasonable, as was suggested by Weron (2008), for example. However, the scarcity of jumps on the daily scale can make the identification of any adequate periodic function problematic in some markets. For instance, Geman and Roncoroni (2006) use a highly convex, two-parameter periodic intensity function to ensure that the price jump occurrences cluster around the peak dates and rapidly fade away. However, they estimate the parameters using only 6, 16 and 27 (for the COB, PJM and ECAR markets, respectively) spike occurrences, which makes the calibration results highly questionable, especially for COB. Bhar et al. (2013) propose a jump-diffusion model with the intensity being the sum of four seasonal dummies. They calibrate the model to PJM prices from a more recent period (2004–2009), and conclude that the Winter and Summer intensities are almost twice as high as those in Spring and Fall. Studying German EEX spot prices, Seifert and Uhrig-Homburg (2007) find that Poisson jump and Poisson spike processes (i.e., with the 'bounce back' effect introduced by Weron, Simonsen, & Wilman, 2004) with constant intensities are unable to model electricity price spike patterns correctly, and the clustering of spikes in particular. They suggest using a stochastic jump intensity, which provides more flexibility.

After a jump, the price is forced back to its normal level by the mean reversion mechanism. However, a high

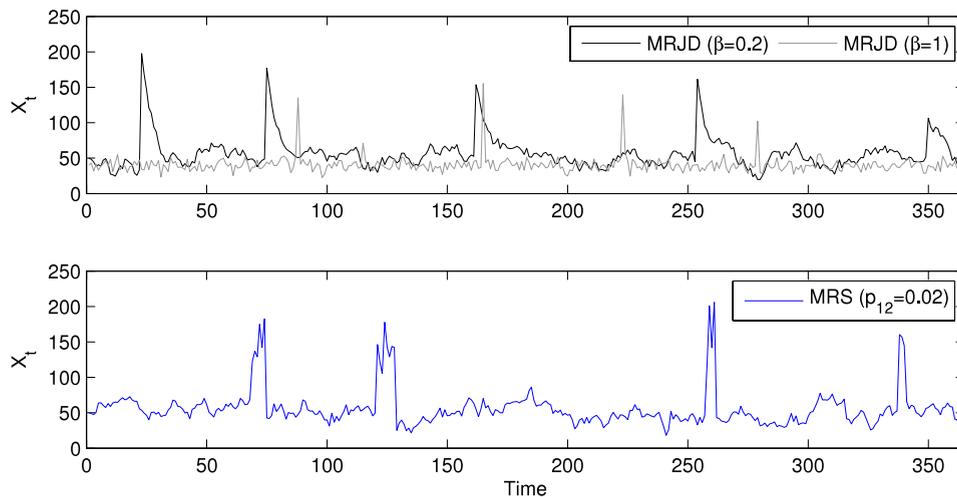


Fig. 8. *Top panel:* Two sample trajectories of the standard MRJD process, see Eq. (7), for two different speeds of mean-reversion: $\beta = 0.2$ and $\beta = 1$. The remaining parameters are the same for both trajectories: $\frac{\alpha}{\beta} = 40$, $\sigma = 6$, $\mu = 100$, $\gamma = 30$, and $\lambda = 0.02$. Clearly, the low rate of mean reversion yields more realistic dynamics in the ‘base regime’, but is too slow to force the price back to its normal level after a jump. On the other hand, a high speed of mean reversion leads to unrealistic dynamics in the ‘base regime’, but a reasonable price behavior after a jump. *Bottom panel:* A sample trajectory of a 2-state MRS process with independent regimes, see Eqs (10)–(11), having characteristics similar to the MRJD with $\beta = 0.2$, i.e., with base regime parameters $\frac{\alpha}{\beta} = 40$, $\beta = 0.2$ and $\sigma = 6$, Gaussian spikes with $\mu = \frac{\alpha}{\beta} + 100$ and $\gamma = 30$, and the transition matrix defined by $p_{11} = 0.98$ and $p_{22} = 0.88$. Observe that MRS models allow for consecutive spikes in a very natural way.

rate of mean reversion, such as is required to force the price back to its normal level after a jump, would lead to a highly overestimated value of this parameter for prices outside the ‘spike regime’; see the top panel of Fig. 8. To circumvent this, Escribano, Pena, and Villaplana (2002) allow for signed jumps. However, if these follow each other randomly, the spike shape obviously has a very low probability of being generated. Geman and Roncoroni (2006) suggest using mean reversion coupled with upward and downward jumps, with the direction of a jump being dependent on the current price level. Weron, Bierbrauer, and Trück (2004) and Weron, Simonsen et al. (2004) assume that a positive jump is always followed by a negative jump of (approximately) the same size, in order to capture the rapid decline of electricity prices after a spike; these are referred to by Seifert and Uhrig-Homburg (2007) as ‘Poisson spike processes’. At the daily level, i.e., when analyzing average daily prices, this approach is a good enough approximation for some less spiky markets. Benth, Kallsen, and Meyer-Brandis (2007) model the spot electricity price using a sum of non-Gaussian Ornstein–Uhlenbeck processes, each of which reverts to the mean at a different speed, and having pure jump processes with only positive jumps as sources of randomness. In an empirical study utilizing German EEX market data, Benth, Kiesel, and Nazarova (2012) compare the ‘factor’ model of Benth et al. (2007), the MRJD of Cartea and Figueroa (2005) and the ‘threshold’ model of Geman and Roncoroni (2006), and conclude that the mean-reversion parameters for both the MRJD and ‘threshold’ models are unable to distinguish between spikes and the base signal, thus leading to an overly slow mean reversion for the spikes and an overly fast mean-reversion for the base signal. These two models try to compensate for this by a very high volatility, meaning that the pathwise properties of the EEX price dynamics are

not captured well. On the other hand, the ‘factor’ model captures mean-reversion very well, but does not capture the variability of the paths appropriately.

The problem of calibrating jump-diffusion models is related to the more general problem of estimating the parameters of continuous-time jump processes from discretely sampled data; refer to Cont and Tankov (2003) for an excellent review. Of particular interest are the estimation procedures that involve the characteristic function: the maximum likelihood (ML) and partial ML estimation based on a Fourier inversion of the conditional characteristic function (CCF), and the quasi-ML estimation based on conditional moments computed from the derivatives of the CCF evaluated at zero (Asai, McAleer, & Yu, 2006; Singleton, 2001).

In some cases, it is easier to work with the discrete-time version of the SDE that governs the price dynamics. For instance, the mean-reverting diffusion defined in Eq. (5) can be discretized as an autoregressive time series of order one, i.e. AR(1), see Section 3.8.3. Similarly, a MRJD is equivalent to a set of two AR(1) processes with different noise terms. The second, ‘jump’ AR(1) process is chosen with a probability equal to the intensity of the Poisson component. However, even after discretization, the discontinuities inherent in the jump-diffusion processes cause problems. The likelihood function includes an infinite sum over all possible numbers of jump occurrences in a given time interval, and has to be either approximated or truncated in order to allow for a numerical computation of ML estimates (Cont & Tankov, 2003; Huisman, 2009). One popular approach is to approximate the likelihood function by a mixture of normal distributions. For instance, given a standard MRJD model

$$dX_t = (\alpha - \beta X_t)dt + \sigma dW_t + Jdq, \quad (7)$$

where $J \sim N(\mu, \gamma^2)$ is a normal random variable and $dq(t)$ are increments of a HPP with constant intensity λ . Ball and Torous (1983) suggest discretizing the process ($dt \rightarrow \Delta t$; for simplicity let $\Delta t = 1$) and assuming that λ is small, so that the arrival rate for two jumps within one period (e.g., a day) is negligible. Then, the Poisson process is approximated well by a simple binary probability of a jump, $\lambda \Delta t = \lambda$, and of no jump, $(1 - \lambda) \Delta t = (1 - \lambda)$. The MRJD model can be written as an AR(1) process, with the mean and variance of the Gaussian noise term being conditional on the arrival of a jump in a given time interval. More explicitly,

$$x_t = \alpha + (1 - \beta)x_{t-1} + \varepsilon_{t,i},$$

where the subscript i can be either 1 (if no jump occurred in this time period) or 2 (if there was a jump), $\varepsilon_{t,1} \sim N(0, \sigma^2)$ and $\varepsilon_{t,2} \sim N(\mu, \sigma^2 + \gamma^2)$. Then, the model can be estimated by ML, with the likelihood function being a product of densities of a mixture of two normals (Matlab code is available from <http://ideas.repec.org/c/boc/bocode/m429008.html>).

One potentially undesirable empirical property of ML-type methods of calibrating jump-diffusion processes is that they tend to converge to the smallest and most frequent jump component of the actual data, though we would prefer to capture the lower frequency, large jump component. Instead of following the statistically sound 'maximum likelihood' route, many practitioners use a hybrid or stepwise approach (see e.g. Cartea & Figueroa, 2005; Weron, 2008). First, the jumps (spikes) are filtered from the mean-reverting diffusion (for a discussion, see e.g. Janczura et al., 2013; Ullrich, 2012), then their frequency (intensity) can be extracted by simple counting, and the distributional parameters describing the severity of the jumps can be obtained by standard identification techniques. Next, the mean-reverting 'jump-free' diffusion is calibrated to the filtered series. With a similar goal in mind, Chan, Gray, and van Campen (2008) explore a recently developed method of separating the total variation into jump and non-jump components, and, using quadratic variation theory, non-parametrically estimate jump parameters for five zones of the Australian power market. They find that, while a large proportion of the total realized variation is attributable to the continuous part of the price process, jumps make a significant contribution: up to 11% of the total variation in some zones. From a forecasting perspective, the realized variation is shown to be highly persistent, and a modest increase in forecast accuracy can be attained by dividing the total variation into its jump and non-jump components.

3.7.2. Markov regime-switching models

One of the major weaknesses of jump-diffusion models is that they cannot exhibit consecutive spikes at the frequency found in market data. Also, spike clustering can be observed on the daily time scale as well as the hourly time scale (as can be seen in Fig. 9; for more empirical evidence, see e.g. Christensen, Hurn, & Lindsay, 2009). In contrast, Markov regime-switching (MRS) models allow for consecutive spikes in a very natural way. The return of prices to the 'normal' regime after a spike is

straightforward too, as the regime-switching mechanism admits temporal changes in the model dynamics; see the bottom panel in Fig. 8.

MRS models are also more versatile than the *hidden Markov models* (HMM; in the strict sense, see e.g. Cappe, Moulines, & Ryden, 2005) that are more popular in signal processing, since they allow for temporary dependence within the regimes, and in particular, for mean reversion. As the latter is a characteristic feature of electricity spot prices, it is important to have a model that captures this phenomenon. Indeed, the base regime is typically modeled by a mean-reverting diffusion model (for reviews, see Huisman, 2009; Janczura & Weron, 2010), which is sometimes heteroskedastic (Janczura & Weron, 2009). For the spike regime(s), on the other hand, a number of different specifications have been suggested in the literature, ranging from mean-reverting diffusions (Karakatsani & Bunn, 2008), to Gaussian (Huisman & de Jong, 2003; Liebl, 2013), lognormal (Weron, Bierbrauer et al., 2004), exponential (Bierbrauer, Menn, Rachev, & Trück, 2007), heavy tailed (Bierbrauer, Trück, & Weron, 2004; Weron, 2009) and non-parametric (Eichler & Türk, 2013) random variables, to mean-reverting diffusions with Poisson jumps (Arvesen, Medbø, Fleten, Tomasgard, & Westgaard, 2013; De Jong, 2006; Keles et al., 2012; Mari, 2008).

The idea underlying Markov regime-switching is to represent the observed stochastic behavior of a (deseasonalized and detrended) spot price process X_t by L separate states or regimes with different underlying stochastic processes $X_{t,j}$, $j = 1, \dots, L$. The switching mechanism between the states is assumed to be an unobserved (latent) Markov chain R_t governed by the transition matrix \mathbf{P} containing the probabilities $p_{ij} = P(R_{t+1} = j \mid R_t = i)$ of switching from regime i at time t to regime j at time $t + 1$:

$$\mathbf{P} = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1L} \\ p_{21} & p_{22} & \dots & p_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ p_{L1} & p_{L2} & \dots & p_{LL} \end{pmatrix},$$

with $p_{ii} = 1 - \sum_{j \neq i} p_{ij}$. (8)

Because of the Markov property, the current state R_t at time t depends on the past only through the most recent value R_{t-1} . In general, L regime models can be considered. However, two or three regimes are typically enough to model the dynamics of electricity spot prices adequately (Janczura & Weron, 2010; Karakatsani & Bunn, 2010).

There are essentially two popular classes of MRS models that are used in the energy economics literature. Both are based on a discretized version of the mean-reverting diffusion process defined in Eq. (5), sometimes with a more general, heteroskedastic volatility term: $\sigma(X_t, t) = \sigma |X_t|^\gamma$. They differ in the type of dependence between the regimes. In the first specification, only the model parameters change depending on the state process values, while in the second, the individual regimes are driven by independent processes.

Dependent regimes with the same random noise process in all regimes (but different parameters; an

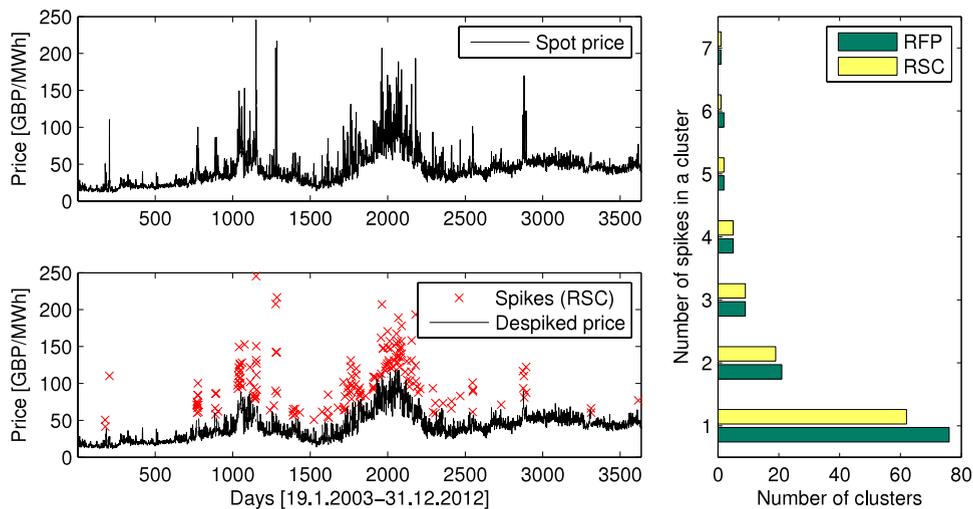


Fig. 9. Upper left panel: APX-UK average daily spot price over the period 19.1.2003–31.12.2012. Lower left panel: Spikes (identified using a regime-switching classification, RSC) and spike-filtered prices. Right panel: Histogram of the size of spike clusters for two spike filtering methods – RSC and recursive filter on prices (RFP). In particular, out of 170 (194 for RFP) spikes identified, there was one cluster of seven spikes, one (two for RFP) cluster(s) of six spikes, and two clusters of five spikes.

approach dating back to the seminal work of Hamilton, 1989) lead to computationally simpler models, where the observed process X_t is described by a parameter-switching time series of the form:

$$X_t = \alpha_{R_t} + (1 - \beta_{R_t})X_{t-1} + \sigma_{R_t}|X_{t-1}|^{\gamma_{R_t}}\varepsilon_t, \quad (9)$$

sharing the same set of random innovations in the L regimes; the ε_t s are assumed to be $N(0, 1)$ -distributed. Sample applications of this approach include those of Bordignon, Bunn, Lisi, and Nan (2013); Karakatsani and Bunn (2008); Kosater and Mosler (2006) and Mount, Ning, and Cai (2006).

On the other hand, independent regimes (introduced by Huisman & de Jong, 2003) allow for a greater flexibility and admit qualitatively different dynamics in each regime. They seem to be a more natural choice for electricity spot price processes, which can exhibit moderately volatile behaviors in the base regime and very volatile behaviors in the spike regime (because of the change in the slope of the demand function, see Fig. 7). Such models have been used by Arvesen et al. (2013); Bierbrauer et al. (2004, 2007); Eichler and Türk (2013); Janczura (2014); Kosater and Mosler (2006); Liebl (2013); Mari (2008); and Weron (2009), among others. The independent regime process X_t is defined as:

$$X_t = \begin{cases} X_{t,1} & \text{if } R_t = 1, \\ \vdots & \vdots \\ X_{t,L} & \text{if } R_t = L, \end{cases} \quad (10)$$

where at least one regime $i = 1, \dots, L$ is given by:

$$X_{t,i} = \alpha_i + (1 - \beta_i)X_{t-1,i} + \sigma_i|X_{t-1,i}|^{\gamma_i}\varepsilon_{t,i}. \quad (11)$$

The other regimes are modeled by independent and identically distributed (i.i.d.) random variables. For instance, in the three-regime model advocated by Janczura and Weron (2010), the second regime ($R_t = 2$) represents the sudden

price spikes that are caused by unexpected supply shortages, and is given by i.i.d. random variables from the shifted log-normal distribution: $\log(X_{t,2} - q_2) \sim N(\mu_2, \sigma_2^2)$, for $X_{t,2} > q_2$. The same assumption that observations from the spike regime should not be smaller than some threshold is also used by Eichler and Türk (2013). The third regime ($R_t = 3$) is responsible for sudden price drops (and possibly negative prices), and is governed by the shifted 'inverse log-normal' law: $\log(-X_{t,3} + q_3) \sim N(\mu_3, \sigma_3^2)$, for $X_{t,3} < q_3$. The values q_i in the above formulas can be either optimized numerically as in Janczura and Weron (2014) or chosen arbitrarily, e.g., let q_2 be the third quartile and q_3 the first quartile of the (deseasonalized) dataset; for many datasets, this choice is close to the optimal values. Such a specification of the spike and drop regime distributions ensures that observations below (above) the third (first) quartile will not be classified as spikes (drops). It should be noted that, once estimated, the values q_2 and q_3 are treated as constant parameters of the model.

The calibration of regime-switching models with an observable state process (like Threshold AR models, see Section 3.8.5), boils down to the problem of estimating the parameters in each regime independently. In case of MRS models, however, the calibration process is not straightforward, since the state process is latent and not observable directly. We have to infer the parameters and state process values at the same time. The most popular is probably the Expectation-Maximization (EM) algorithm, which was first used for estimating MRS models by Hamilton (1990), and was later refined by Kim (1994). It is a two-step iterative procedure, reaching a local maximum of the likelihood function. First, the conditional probabilities of the process being in regime j at time t , the so-called *smoothed inferences*, are computed for a parameter vector θ . Next, new and more exact maximum likelihood (ML) estimates of θ are calculated using the likelihood function, weighted with the smoothed inferences from the previous step. Note that the

introduction of independent regimes results in a significantly increased computational burden. See Janczura and Weron (2012) for an efficient modification of the algorithm to overcome this problem (Matlab code is available from <http://ideas.repec.org/s/wuu/hscode.html>). It allows for calibration that is 100 to over 1000 times faster than the competing approach of Huisman and de Jong (2002), utilizing the probabilities of the last 10 observations.

Note that, as a byproduct of calibrating a MRS model to deseasonalized and detrended data, we obtain the conditional probabilities of the process being in a certain regime at a given time. All prices with probabilities of being in one of the extreme regimes which exceed a certain threshold, say 50%, may be classified as outliers. For instance, if we calibrate a two-state MRS with an independent lognormal spike regime and mean-reverting base regime dynamics (see Eq. (11)), with spike cutoff $q_2 = 95\%$, to APX-UK average daily spot prices from the period 19.1.2003–31.12.2012, then we will identify 170 spikes, as in the lower left panel of Fig. 9. The other spike-filtering technique used in this figure – the recursive filter on prices (RFP) – classifies as spikes all prices that exceed the mean price level by three standard deviations, with the outlying observations being removed one by one in a ‘recursive filter’ fashion (for details, see Janczura et al., 2013).

3.7.3. Strengths and weaknesses

Reduced-form models are generally not expected to forecast hourly prices accurately, but are expected to recover the main characteristics of electricity spot prices, typically at the daily time scale. Such models provide a simplified, yet reasonably realistic picture of the price dynamics, and are commonly used for derivatives pricing and risk analysis (for reviews, see e.g. Benth et al., 2008; Eydeland & Wolyniec, 2003). Interestingly, when it comes to volatility or price spike forecasts, reduced-form models have been reported to perform reasonably well, see Section 4.1.2.

The few known attempts to use either mean-reverting jump-diffusions (Weron & Misiorek, 2008) or Markov regime-switching models (Misiorek et al., 2006) for forecasting the next day’s hourly prices have generally confirmed their poor performance in this context. These results are in line with earlier reports by Bessec and Bouabdallah (2005) and Dacco and Satchell (1999), who question the adequacy of MRS models for forecasting in general. On the other hand, Kosater and Mosler (2006) reach opposite conclusions, at least for medium-term forecasts of average daily prices from the German EEX market. They compare parameter switching (see Eq. (9)) and independent regime (see Eqs (10)–(11)) MRS specifications to a mean-reverting diffusion (an AR(1) process in discrete time), and find that the regime-switching models are slightly more accurate for 30- to 80-day-ahead forecasts. In contrast, for UK data, Heydari and Siddiqui (2010) find that their regime-switching model is unlikely to capture electricity price behaviors in the medium-term, and their non-linear model with stochastic volatility for logarithms of electricity prices performs better than either the linear or regime-switching models, in terms of valuing a gas-fired power plant. Similarly, Liebl

(2013) observes a poor performance of the MRS model proposed by Huisman and de Jong (2003) for one- to 20-day-ahead forecasts of daily EEX spot prices, relative to three factor models (see Section 4.4). However, the combination of MRS and vector autoregressions (as was proposed by Lanne, Lütkepohl, & Maciejowska, 2010, in a macroeconomic context) may potentially turn out to be a useful approach in EPF as well.

3.8. Statistical models

Reduced-form models excel at derivatives valuation and risk analytics. However, when forecasting day-ahead electricity prices, the models’ simplicity and analytical tractability are no longer an advantage. In fact, a model’s simplicity can be a serious limitation. Historically, the first inflow of statistical EPF techniques consisted chiefly of statistical methods of load forecasting. By a simple substitution of prices for loads (and possibly loads for temperatures), the researchers were able to obtain EPF models. As time passed, more and more contemporary statistical, econometric or signal processing techniques were introduced to this area.

Statistical (econometric, technical analysis) methods forecast the current price by using a mathematical combination of the previous prices and/or previous or current values of exogenous factors, typically consumption and production figures, or weather variables. The two most important categories are additive and multiplicative models. They differ in whether the predicted price is the sum (additive) of a number of components or the product (multiplicative) of a number of factors. The former are far more popular. Note, however, that the two are closely related: a multiplicative model for prices can be transformed into an additive model for log-prices.

Statistical models are attractive because some physical interpretation may be attached to their components, thus allowing engineers and system operators to understand their behavior. They are often criticized for their limited ability to model the (usually) nonlinear behavior of electricity prices and related fundamental variables; however, in practical applications, their performances are comparable to those of their non-linear alternatives (discussed in Section 3.9).

3.8.1. Similar-day and exponential smoothing methods

A very popular benchmark model in EPF is the *similar-day* method. It is based on searching historical data for days with characteristics similar to the predicted day, and taking those historical values as forecasts of future prices (Shahidehpour et al., 2002; Weron, 2006). Similar characteristics may include the day of the week, day of the year, holiday type, and weather or consumption figures. Instead of a single similar-day price, the forecast may be a linear combination or a regression procedure that can include several similar days.

One of the more common implementations of the similar-day approach, which was probably introduced to EPF by Nogales et al. (2002) and is dubbed the *naïve method*, proceeds as follows. A Monday is similar to the Monday of the previous week, and the same rule applies for Saturdays

and Sundays. A Tuesday is similar to the previous Monday, and the same rule applies for Wednesdays, Thursdays and Fridays. As was argued by Conejo, Contreras et al. (2005), Contreras et al. (2003) and Nogales et al. (2002), forecasting procedures that are not calibrated carefully fail to pass this 'naïve test' surprisingly often.

Another relatively simple benchmark, which is very popular in load forecasting (see e.g. Taylor, 2010) but less popular in EPF, is *exponential smoothing*. It is a pragmatic approach to forecasting, whereby the prediction is constructed from an exponentially weighted average of past observations:

$$\hat{x}_t = s_t = \alpha x_t + (1 - \alpha)s_{t-1}. \quad (12)$$

Each smoothed value s_t is the weighted average of the previous observations, where the weights decrease exponentially depending on the value of parameter $\alpha \in (0, 1)$. More complex models have been developed to accommodate time series with seasonal and trend components. The general idea here is that forecasts are not computed from consecutive previous observations alone, but an independent (smoothed) trend and seasonal component can be added. For reviews of point and interval forecasting using exponential smoothing, we refer to Gardner (2006) and Hyndman, Koehler, Ord, and Snyder (2008).

An interesting variant of exponential smoothing is the so-called THETA method of Assimakopoulos and Nikolopoulos (2000). Hyndman and Billah (2003) demonstrate that it is equivalent to simple exponential smoothing with drift, where the drift is half the value of the slope of a linear regression fitted to the data. As such, the THETA method provides a form of shrinkage which limits the ability of the model to produce extremely inaccurate forecasts. The method performed very well in the M3 forecasting competition (Makridakis & Hibon, 2000). However, it should be noted that a vast majority of the test samples included data sampled at a monthly or lower frequency. It remains an open question as to whether the THETA method would perform well for daily or hourly electricity prices.

Summing up, to the best of our knowledge, only one article has used exponential smoothing as a method for EPF (though exponential smoothing is sometimes used as a component of a larger model, see e.g. Jonsson et al., 2013). Cruz, Muñoz, Zamora, and Espinola (2011) utilize double seasonal exponential smoothing as a benchmark for more sophisticated models. In their study, exponential smoothing performs slightly better than ARIMA, and both outperform the naïve method for hourly spot prices from the Spanish market. However, all three benchmarks are worse than either dynamic regression models (i.e., ARX) or a neural network. Interestingly, exponential smoothing outperforms all other methods for hour 22.

3.8.2. Regression models

Regression is one of the most widely used statistical techniques. The general purpose of multiple regression is to learn more about the relationships between several independent or predictor variables and a dependent or criterion variable. Multiple regression is based on least squares: the model is fitted such that the sum-of-squares

of the differences between observed and predicted values is minimized. In its classical form, multiple regression assumes that the relationship between variables is linear:

$$P_t = \mathbf{B}\mathbf{X}_t + \varepsilon_t = b_1X_t^{(1)} + \dots + b_kX_t^{(k)} + \varepsilon_t, \quad (13)$$

where \mathbf{B} is a $1 \times k$ vector of constant coefficients, \mathbf{X}_t is the $k \times 1$ vector of regressors (some or all of which may be transformed beforehand, e.g., by applying the Box-Cox or a polynomial transformation) and ε_t is an error term. The regressors are selected in-sample among the explanatory variables considered, which are assumed to be correlated with the electricity price P_t . In such a standard case, estimation can be performed using maximum likelihood methods. A *time-varying regression* (TVR) model allows for price driver effects that evolve continuously:

$$P_t = \mathbf{B}_t\mathbf{X}_t + \varepsilon_t = b_{1,t}X_t^{(1)} + \dots + b_{k,t}X_t^{(k)} + \varepsilon_t, \quad (14)$$

where \mathbf{B}_t is now a $1 \times k$ vector of time-varying coefficients. TVR model parameters can be estimated using state space methods and the Kalman filter (see e.g. Durbin & Koopman, 2001).

Despite the large number of alternatives, linear regression models are still among the most popular EPF approaches. However, in most papers they are combined with other, typically more sophisticated methods; various interesting applications are discussed in the following paragraphs. Moreover, it is often hard to separate regression and autoregression approaches, as many of them are called 'regression models' but include lagged electricity prices as regressors. Such models could just as well be called autoregressions with exogenous variables (see Section 3.8.4).

In one of the early applications of regression models, Kim, Yu, and Song (2002) utilize wavelet decomposition coupled with multiple regression. That is, the regression coefficients are calculated using the wavelet decomposition detail series and the predicted demand. The day-ahead price forecast is then given by the previous day's low frequency and the predicted high frequency components. A similar forecasting technique is applied by Conejo, Contreras et al. (2005) to hourly PJM data. Also, Schmutz and Elkuch (2004) use multiple regression with gas prices, available nuclear capacity, temperatures and rainfall as regressors, and a mean-reverting stochastic process for the residuals.

Koopman, Ooms, and Carnero (2007) consider general seasonal periodic regression models with ARIMA, ARFIMA (also known as Fractional ARIMA or FARIMA) and GARCH disturbances for the analysis of daily spot prices of electricity. The regressors capture yearly cycles, holiday effects, and possible interventions in the mean and variance. The authors conclude that for the Nord Pool market (but not for other European markets), a long memory model with periodic coefficients is required in order to model daily spot prices effectively. However, the models' forecasting performances are not evaluated. Karakatsani and Bunn (2008) build a fundamental 'regression model' for each of the 48 half-hourly load periods in the British market, and compare its day-ahead forecasting performance to those of TVR and regime-switching regression models. They conclude that models which invoke market fundamentals and time-varying coefficients exhibit the best

predictive performances among various alternatives. Bordinon et al. (2013) use similar linear regression and TVR models in their evaluation of different forecast combination schemes, see Section 4.3.1.

Azadeh, Moghaddam, Mahdi, and Seyedmahmoudi (2013) propose an algorithm which switches between the predictions of different models (neural networks, fuzzy regressions and a standard regression) based on some pre-specified rules, and use it for long-term (annual time scale) EPF. Jonsson et al. (2013) introduce a two-step methodology for EPF, with a focus on the impact of the predicted system load and wind power generation. The nonlinear and nonstationary influences of these explanatory variables are accommodated in a nonparametric and TVR model. In a second step, an AR model and exponential smoothing are applied to account for residual autocorrelation and seasonal dynamics. Empirical day-ahead forecasting results for the Western Danish price area of Nord Pool demonstrate the practical benefits of accounting for these explanatory variables.

3.8.3. AR-type time series models

The standard time series model that takes into account the random nature and time correlations of the phenomenon under study is the *AutoRegressive Moving Average* model. In the ARMA(p, q) model, the current value of the price X_t is expressed linearly in terms of its p past values (autoregressive part), and in terms of q previous values of the noise (moving average part):

$$\phi(B)X_t = \theta(B)\varepsilon_t. \quad (15)$$

Here, B is the backward shift operator, i.e., $B^h X_t \equiv X_{t-h}$; $\phi(B)$ is a shorthand notation for $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$; and $\theta(B)$ is a shorthand notation for $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$, where ϕ_1, \dots, ϕ_p and $\theta_1, \dots, \theta_q$ are the coefficients of autoregressive and moving average polynomials, respectively. Note that some authors and computer software packages (e.g., SAS) use a different definition of the second polynomial: $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$. Finally, ε_t is i.i.d. noise (or *white noise*) with zero mean and finite variance, which is often denoted by $WN(0, \sigma^2)$. For $q = 0$, we obtain the well-known *AutoRegressive* AR(p) model, and for $p = 0$, we get the *Moving Average* MA(q) model.

The ARMA modeling approach assumes that the time series under study is (weakly) stationary. If it is not, then a transformation of the series to the stationary form has to be done first. One of the simplest ways to achieve this is to perform differencing. Box and Jenkins (1976) introduced a general model that contained both AR and MA parts, and explicitly included differencing in the formulation. The *AutoRegressive Integrated Moving Average* (ARIMA) or *Box-Jenkins model* has three types of parameters: the autoregressive parameters (ϕ_1, \dots, ϕ_p), the number of differencing passes at lag-one (d), and the moving average parameters ($\theta_1, \dots, \theta_q$). A series that needs to be differenced d times at lag-1 and afterward has orders p and q of the AR and MA components, respectively, is denoted by ARIMA(p, d, q), and can be written conveniently as:

$$\phi(B)\nabla^d X_t = \theta(B)\varepsilon_t, \quad (16)$$

where $\nabla x_t \equiv (1 - B)x_t$ is the lag-1 differencing operator, which is a special case of the more general lag- h differencing operator: $\nabla_h x_t \equiv (1 - B^h)x_t \equiv x_t - x_{t-h}$. Note that ARIMA($p, 0, q$) is simply an ARMA(p, q) process.

Sometimes simple differencing at lag-1, even repeated many times, is not enough to make the series stationary. In particular, seasonal signals of period greater than one, like electricity loads or prices, require differencing at longer lags. Such processes are known as *seasonal ARIMA* (SARIMA) models. The general notation for the order of a seasonal ARIMA model with both seasonal and nonseasonal factors is ARIMA(p, d, q) \times (P, D, Q) $_s$. The term (p, d, q) represents the order of the nonseasonal part, while (P, D, Q) $_s$ represents the order of the seasonal part. The value of s is the number of observations in the seasonal pattern, e.g., seven for daily series with weekly periodicity, 24 for hourly series with daily periodicity, etc. The SARIMA model can be written compactly as:

$$\phi(B)\Phi(B^s)\nabla^d \nabla_s^D X_t = \theta(B)\Theta(B^s)\varepsilon_t. \quad (17)$$

Note that every SARIMA model can be transformed into an ordinary, though long, ARMA model in the variable $\tilde{X}_t \equiv \nabla^d \nabla_s^D X_t$. As a consequence, the estimation of ARIMA and SARIMA model parameters is analogous to that for ARMA processes. The latter generally consist of two steps: model identification (using information criteria to compensate for the effect of the improvement in fit at the cost of model complexity), and estimation of the coefficients (e.g., by least squares regression, recursive least squares, maximum likelihood, or the prediction error method). The forecasting of ARMA-type models can be conducted via the Durbin–Levinson algorithm or the innovations algorithm, or by using the Kalman filter for models specified in state space form. For reviews, we refer to Brockwell and Davis (1996); Ljung (1999); Shumway and Stoffer (2006), and the very recent open access e-book by Hyndman and Athanasopoulos (2013).

AR-type models provide the backbone of all time series models of electricity prices. There have been some EPF applications of (S)AR(IMA) models, but the majority of papers propose and use time series models with exogenous variables (load, temperature, wind). These will be discussed in Section 3.8.4.

Cuaresma, Hlouskova, Kossmeier, and Obersteiner (2004) apply variants of AR(1) and general ARMA processes (including ARMA with jumps) to short-term EPF in the German EEX market. They conclude that specifications in which each hour of the day is modeled separately present uniformly better forecasting properties than specifications for the whole time series, and that the inclusion of simple probabilistic processes for the arrival of extreme price events (jumps) could lead to improvements in the forecasting abilities of univariate models for electricity spot prices.

In a related study, Weron and Misiorek (2005) use various autoregression schemes for modeling and forecasting prices in the California market. They observe that an AR model with lags of 24, 48 and 168 h, where each hour of the day is modeled separately, performs better than the single large (S)ARIMA specification for all hours proposed by Contreras et al. (2003). The reduction in WMAE, see

Eq. (2), even reaches 30% for a normal, non-spiky out-of-sample test period (first week of April 2000). Misiolek et al. (2006) find that this simple AR model structure, when expanded to include a load forecast of the system operator, is a tough competitor among the AR(X)-GARCH, TAR(X) and MRS models. ARX turns out to be the best in a relatively calm period in the California market (April to mid-June, 2000), and second best (after TARX) in a more volatile period (second half of 2000). Also, Jonsson et al. (2013) successfully use a similarly simple AR model (with lags of 24, 48 and 168 h) to account for residual autocorrelation and seasonal dynamics, and use it for short-term EPF.

Conejo, Plazas et al. (2005) propose a wavelet-ARIMA technique, which consists of (i) decomposing the price series using a discrete wavelet transform (DWT), (ii) modeling the resulting detail and approximation series using ARIMA processes to obtain 24 hourly predicted values, and (iii) applying the inverse wavelet transform, to yield the predicted prices for the next 24 h. The performance of the wavelet-ARIMA technique is generally better than that of a standard ARIMA process. In all four weekly test samples (Spanish market, year 2002), the mean weekly errors are reduced; for the winter week, the error is reduced by 25%.

In the same spirit, Shafie-Khah, Moghaddam, and Sheikh-El-Eslami (2011) propose a hybrid method for forecasting day-ahead electricity prices, in which a wavelet transform provides a set of 'better-behaved' time series, an ARIMA model is used to generate a linear forecast, and then a radial basis function (RBF) network (see Section 3.9.2) is used to correct the estimation error of the wavelet-ARIMA forecast. Following Huang, Huang, and Wang (2005), a particle swarm optimization is used to optimize the network structure. The results for the Spanish market show that the proposed hybrid method can provide an improvement in forecasting accuracy over a standard ARIMA model, the wavelet-ARIMA model of Conejo, Plazas et al. (2005), the fuzzy neural network of Amjady (2006), and the neural network of Catalão, Mariano, Mendes, and Ferreira (2007), and also over the 'mixed model' of Garcia-Martos, Rodriguez, and Sanchez (2007) in three test periods out of four. The last of these is a set of 24 hourly ARIMA models for weekdays (which are calibrated only to weekday prices) and a set of 24 hourly ARIMA models for weekends (which are calibrated to weekday and weekend prices). Consequently, the model of Garcia-Martos et al. (2007) may be treated as a generalization of the approach advocated by Cuaresma et al. (2004) and Misiolek et al. (2006), where each hour of the day is modeled by a separate AR-type model.

In a more 'econometric' application, Haldrup and Nielsen (2006) observe that there seems to be a strong support for long memory and fractional integration in Nord Pool area prices over the period 2000–2003. One possible explanation for this is the fact that a significant amount of the electricity supply in Nord Pool is from hydropower plants, and it is a classical empirical finding that river flows and water reservoir levels exhibit long memory. Consequently, Haldrup and Nielsen calibrate seasonal ARFIMA models to Nord Pool area prices and use them for forecasting. Lagarto, De Sousa, Martins,

and Ferrão (2012) describe an interesting methodology which combines elements of time series and multi-agent modeling. They forecast the next day's 24 hourly prices using an ARIMA model applied to the conjectural variations (see Section 3.5.1) of the firms participating in the Spanish power market. They find that the conjectural variations price forecast performs better than the naïve method, and slightly better than a pure ARIMA model. Further applications of (S)AR(IMA) models in EPF include the studies by Amjady and Hemmati (2009); Che and Wang (2010); Cruz et al. (2011); and Tan, Zhang, Wang, and Xu (2010). In these papers, they are used as benchmarks for more complicated models or hybrid constructions involving neural networks, support vector machines or GARCH components.

3.8.4. ARX-type time series models

The time series models discussed in Section 3.8.3 relate the signal under study to its own past, and do not explicitly use the information contained in other related time series. However, as has already been discussed extensively in Section 3.6, electricity prices are also influenced by the present and past values of various exogenous factors, most notably the generation capacity, load profiles and ambient weather conditions. To capture the relationship between prices and these fundamental variables, time series models with *eXogenous* or *input* variables can be used. These models do not constitute a new class; rather, they can be viewed as generalizations of existing classes. For instance, ARX, ARMAX, ARIMAX and SARIMAX are generalized counterparts of AR, ARMA, ARIMA and SARIMA, respectively. Models with input variables are also known as *transfer function*, *dynamic regression*, *Box-Tiao*, *intervention* or *interrupted time series* models. Some authors distinguish among them, while others use the names interchangeably, thus causing a lot of confusion in the literature (for a discussion in the context of electricity markets, see Weron, 2006). Moreover, as was noted in Section 3.8.2, it is often hard to distinguish between regression and ARX-type models. If the number of fundamental regressors is large, then they are typically called regression models; if the autoregressive structure is complex, then they should be classified as ARX-type models instead.

The mechanism for including exogenous variables is analogous for all ARMA-type models. We will now describe the ARMAX model without loss of generality. In this model, the current value of the spot price X_t is expressed linearly in terms of its past values, in terms of previous values of the noise, and, additionally, in terms of present and past values of the exogenous variable(s). The *AutoRegressive Moving Average model with eXogenous variables* $V^{(1)}, \dots, V^{(k)}$, or $\text{ARMAX}(p, q, r_1, \dots, r_k)$, can be written compactly as:

$$\phi(B)X_t = \theta(B)\varepsilon_t + \sum_{i=1}^k \psi^i(B)V_t^{(i)}, \quad (18)$$

where r_i are the orders of the exogenous factors and $\psi^i(B)$ is a shorthand notation for $\psi^i(B) = \psi_0^i + \psi_1^i B + \dots + \psi_{r_i}^i B^{r_i}$, with the ψ_j^i 's being the corresponding coefficients.

Alternatively, the ARMAX model is often defined in a transfer function form:

$$X_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t + \sum_{i=1}^k \tilde{\psi}^i(B) V_t^{(i)}, \quad (19)$$

where $\tilde{\psi}^i$ are the appropriate coefficient polynomials. For $\theta(B) \equiv 1$, Eq. (19) yields the dynamic regression form of the ARX model.

Typically, the estimation of ARX models is conducted using either least squares or *instrumental variables* techniques. The former minimizes the sum of squares of the right-hand side minus the left-hand side of Eq. (18), with respect to ϕ and the ψ^i 's ($\theta(B) \equiv 1$ for the ARX model). The latter determines ϕ and the ψ^i 's so that the error between the right- and left-hand sides becomes uncorrelated with certain linear combinations of the inputs. For the calibration of ARMAX coefficients, maximum likelihood (ML) or the *prediction error method* is typically used. In the latter, the parameters of the model are chosen so that the difference between the model's (predicted) output and the measured output is minimized. For Gaussian disturbances, it coincides with ML estimation. Like ML, the prediction error method typically involves an iterative, numerical search for the best fit (see Ljung, 1999, and Matlab's System Identification Toolbox). Other calibration techniques have also been proposed, such as a weighted recursive least squares algorithm (Fan & McDonald, 1994), evolutionary programming (Yang, Huang, & Huang, 1996), and particle swarm optimization (Huang et al., 2005).

Time series models with exogenous variables have been applied extensively to short-term EPF. Nogales et al. (2002) utilize ARMAX and ARX models (which they call 'transfer function', TF, and 'dynamic regression', DR, respectively) for predicting hourly prices in California and Spain. The two models perform comparably, with the weekly MAPE (note that Nogales et al. call it the 'Mean Weekly Error', see also Section 3.3) being just below 3% for the first week of April 2000 in California and around 5% for the third weeks of August and November 2000 in Spain. The results are significantly better than for the ARIMA and ARIMA-E (ARIMA with load as an explanatory variable, i.e., ARIMAX) models proposed by Contreras et al. (2003). Somewhat surprisingly, however, the TF and DR models – which utilize one common multi-parameter specification for all hours – outperform the ARIMA-E model by more than 40%. Both the TF and ARIMA-E models use the same variables. This may be related to the ways in which the load data are included in the two methods. In ARIMA-E, it is just an explanatory variable, but in the TF specification, it is bundled with the autoregressive part of the model. What is even more surprising is that the performance of ARIMA is comparable to that of ARIMA-E, even though the latter additionally uses an important exogenous variable.

Nogales and Conejo (2006) repeat their earlier study for 2003 PJM market data. Again, the TF model performs better than a standard ARIMA process; however, only an 18% reduction in MAPE value is observed for the test period (July–August 2003) this time. In a related study, Conejo, Contreras et al. (2005) compare different methods of short-term EPF: three time series specifications (ARIMA, TF and

DR), a wavelet multivariate regression technique, and a multilayer perceptron (MLP; see Section 3.9.2) with one hidden layer. For a dataset comprising PJM prices from the year 2002, the ARIMA model is worse than the time series models with exogenous variables (more than 75% worse for the last week of July 2002), but better than the MLP.

Instead of considering a single time series specification for all hours, Weron and Misiorek (2005) and Misiorek et al. (2006) use a set of 24 relatively small ARX models, one for each hour of the day, with the CAISO day-ahead load forecast as the exogenous variable and three dummies for recovering the weekly seasonality. They conclude that these models perform much better than the single large (S)ARIMA specification for all hours proposed by Contreras et al. (2003), and slightly worse than the TF and DR models of Nogales et al. (2002). However, only the results for the first week of April 2000 in the California power market are comparable, as this is the only common test sample used in all four papers. Moreover, the TF and DR models are calibrated to spike preprocessed data (though the procedure is not disclosed), while the ARX models are calibrated to raw data. In Case Study 4.3.8, Weron (2006) calibrates ARX models to spike preprocessed California electricity spot prices and observes that the results improve (and are comparable to the other models), though only for the first two weeks of April. Later, when the prices become more volatile, spike preprocessing turns out to be suboptimal. This may imply that the spike preprocessed TF and DR models are particularly good for the calm, first week of April 2000, but not in general.

Nittel and Roberts (2005) consider various econometric models for day-ahead EPF in the California market, including mean-reverting diffusions and jump diffusions, a seasonal ARMA process (called 'ARMAX'), an AR-EGARCH specification (allowing for asymmetry in heteroskedasticity), and a seasonal ARMA model with the temperature, squared temperature and cubed temperature as explanatory variables. They find all temperature variables to be highly statistically significant during the pre-crisis period (April 1, 1998–April 30, 2000); however, the price-temperature relationship breaks down during the crisis period (May 1, 2000–August 31, 2000). The weekly RMSE is also the lowest of all models examined, though the difference from the seasonal ARMA process is small.

Zareipour, Canizares, Bhattacharya, and Thomson (2006) evaluate the usefulness of publicly available electricity market information in forecasting the hourly Ontario energy price (HOEP). Two forecasting horizons are considered, 3 h and 24 h, and the forecasting performances of transfer function (i.e., ARMAX) and dynamic regression (i.e., ARX) models are compared with those of ARIMA models. The authors find that the publicly available information (before the 'real-time') can be used to improve the HOEP forecast accuracy to some extent, but that unusually high or low prices remain unpredictable.

Weron and Misiorek (2008) compare the accuracies of 12 relatively parsimonious time series methods for day-ahead EPF: AR models (using the same specification as Misiorek et al., 2006) and their extensions – spike preprocessed, threshold (see Section 3.8.5) and semiparametric autoregressions (i.e., AR models with nonparametric innovations) – as well as mean-reverting jump diffusions. The

methods are compared using a time series of hourly spot prices and system-wide loads from California, and a series of hourly spot prices and air temperatures from the Nordic market. The authors find evidence that (i) models with the system load as the exogenous variable generally perform better than pure price models, but that this is not necessarily the case when the air temperature is considered as the exogenous variable; and that (ii) semiparametric models, and the *smoothed nonparametric ARX* (SNARX) model in particular, generally lead to better point and interval (see Section 4.2.1) forecasts than their competitors, and also, more importantly, they have the potential to perform well under diverse market conditions. The motivation for using semiparametric models stems from the fact that a nonparametric kernel density estimator will generally yield a better fit to any empirical data (the model error terms in particular) than any parametric distribution. The semiparametric ARX models relax the normality assumption needed for the maximum likelihood estimation in the ARX model. They have the same functional form as the considered ARX model, but the parameter estimates are obtained from a numerical maximization of the empirical likelihood, as was suggested by Cao, Hart, and Saavedra (2003).

Lira, Muñoz, Nuñez, and Cipriano (2009) evaluate the efficiency of Takagi–Sugeno–Kang and ARMAX models (identified by means of a Kalman filter) for day-ahead EPF in the Colombian market. The models include exogenous variables such as reservoir levels and load. The results show that a segmentation of prices into three intervals, based on load behavior, contributes to a significantly better fit. Yan and Chowdhury (2010b) present a hybrid mid-term (on a time frame of between one and six months) EPF model combining both a least squares support vector machine (LSSVM) and ARMAX models. The model shows an improved forecasting accuracy for PJM data compared to a forecasting model using a single LSSVM. Cruz et al. (2011) compare the predictive accuracies of a set of methods (SARIMA, double seasonal exponential smoothing, dynamic regression and a feed-forward neural network), and find evidence that their predictive accuracies can be outperformed significantly by taking into account the system operator's wind generation forecasts.

More recent applications of ARX-type time series models include those of Kristiansen (2012), who modifies the model of Weron and Misiorek (2008) to include Nordic demand and Danish wind power as exogenous variables and models prices jointly across all hours (rather than separately for each hour of the day); Caihong and Wenheng (2012), who present a new method for the system identification of multi-input, single output ARMAX models using the CPSO algorithm, and test it on data from the California power market; and Bordignon et al. (2013), who use an ARMAX(1, 1, 1) model in their evaluation of different forecast combination schemes (see Section 4.3).

3.8.5. Threshold autoregressive models

Roughly speaking, two main classes of *regime-switching models* can be distinguished: those where the regime can be determined by an observable variable (and,

consequently, the regimes that have occurred in the past and present are known with certainty) and those where the regime is determined by an unobservable, latent variable (i.e., the MRS models discussed in Section 3.7.2). In the latter case, we can never be certain that a particular regime has occurred at a particular point in time, but can only assign or estimate probabilities of their occurrences.

The most prominent member of the first class is the *Threshold Autoregressive* (TAR) model originally proposed by Tong and Lim (1980). It assumes that the regime is specified by the value of an observable variable v_t relative to a threshold value T :

$$\phi_1(B)X_t\mathbb{I}_{(v_t \geq T)} + \phi_2(B)X_t\mathbb{I}_{(v_t < T)} = \varepsilon_t, \quad (20)$$

where $\phi_i(B)$ is a shorthand notation for $\phi_i(B) = 1 - \phi_{i,1}B - \dots - \phi_{i,p}B^p$, $i = 1, 2$; B is the backward shift operator; $\mathbb{I}_{(\cdot)}$ denotes the indicator function; and X_t is the spot electricity price. To simplify the exposition, we have specified a two-regime model only; however, a generalization to multi-regime models is straightforward. The inclusion of exogenous (fundamental) variables is also possible: AR processes are simply replaced by ARX processes in Eq. (20), leading to the TARX model.

The *Self Exciting TAR* (SETAR) model arises when the threshold variable is taken as the lagged value of the price series itself, i.e., $v_t = X_{t-d}$; see Tong (1990) for an overview and Lucheroni (2012) for an alternative construction in the context of electricity markets. The model may also be modified further by allowing for a gradual transition between the regimes, leading to the *Smooth Transition AR* (STAR) model. A popular choice for the transition function is the logistic function:

$$G(X_{t-d}; \gamma, T) = [1 + \exp\{-\gamma(X_{t-d} - T)\}]^{-1},$$

where d is the lag and γ determines the smoothness of the transition. The resulting model is known as the *Logistic STAR* (LSTAR) model.

There are a few documented applications of regime-switching TAR-type models to electricity prices. Robinson (2000) fits an LSTAR model to prices in the England and Wales wholesale electricity pool, and shows that its performance is superior to that of a linear autoregressive alternative. Stevenson (2001) calibrates AR and TAR processes to wavelet filtered half-hourly data from the New South Wales (Australia) market, and concludes that the TAR specification (with v_t being the change in demand and $T = 0$) outperforms the AR alternative in forecasting performance. Rambharat, Brockwell, and Seppi (2005) introduce a SETAR-type model with an exogenous variable (temperature recorded at the same time as the maximum price of the day) and a gamma distributed jump component. A common threshold level is used for determining both the AR coefficients and the jump intensities. The authors estimate the model using a Markov chain Monte Carlo with three years of daily data from Allegheny County, Pennsylvania, and find it to be superior (both in-sample and out-of-sample) to a jump-diffusion model.

Weron and Misiorek (2006) calibrate various time series specifications, including TAR and TARX (with the system-wide load as the exogenous variable) models,

and evaluate their predictive power in the California market. The TAR(X) models use the price for hour 24 on the previous day as the threshold variable v_t , and the threshold level is estimated for every hour in a multi-step optimization procedure with ten equally spaced starting points spanning the entire parameter space. During the calm pre-crisis period, the out-of-sample forecasting results are well below acceptable levels, and the models even fail to outperform the naïve approach. Later in the test sample, when the regime switches are more common and the price stays in the spiky regime for longer periods of time, the models (TARX in particular) yield much better forecasts. However, their performances are still disappointing. In a related study, Misiolek et al. (2006) expand the range of threshold variables tested, and find that a value of v_t that is equal to the difference between the mean prices for yesterday and eight days ago leads to a much better forecasting performance. The resulting TAR(X) models are comparable in point forecasting accuracy to their respective linear specifications. Weron and Misiolek (2008) use the same TAR(X) specifications, but for Nord Pool data from two periods: 1998–1999 and 2003–2004. They find that, in terms of point forecasts, the TAR(X) models have relatively large numbers of best forecasts, but their mean errors are (nearly) the worst in the more regular and less spiky 1998–1999 period, indicating that when they are wrong, they miss the actual spot price by a large amount. Also, the prediction intervals (PI) are of very poor quality for both periods.

Using logistic smooth transition regression (LSTR) as an estimation framework, Chen and Bunn (2010) test the proposition that electricity spot price dynamics present a pattern of varying intra-day nonlinear functions of its key fundamental variables. For three distinct periods of the day (off-peak, morning peak and evening peak), they identify quite different models. The main transitional variables identified for regime switching at these times are the carbon price for the off-peak, when coal is the marginal technology, reserve margin for the morning peak, when the load is increasing most quickly, and market concentration for the evening peak, when market power effects are most exercisable. In a follow-up study, Gonzalez et al. (2012) investigate the performances of two hybrid forecasting models for predicting the next-day spot electricity prices on the APX-UK power exchange: (i) a conventional hybrid approach which combines a fundamental model, formulated with supply stack modeling, with an econometric model using data on price drivers, and (ii) an extended variant of this model which includes LSTR to represent regime-switching for periods of structural change. The out-of-sample point forecasts of both hybrid approaches (especially of the hybrid-LSTR) compare favorably to those of non-hybrid SARMA, SARMAX and LSTR models. The quality of the PIs is evaluated by comparing the nominal coverages of the models to the true coverage (no formal tests are performed). The LSTR model gives the best results, closely followed by the hybrid-ARX and SARMAX models. For the hybrid-LSTR model, the observed number of exceeding prices is significantly higher than the theoretical number, due to the overly narrow PIs.

3.8.6. Heteroskedasticity and GARCH-type models

The linear AR(X)-type models assume *homoskedasticity*, i.e., a constant variance and covariance function. From an empirical point of view, financial time series – including electricity spot prices – exhibit various forms of non-linear dynamics, with the crucial one being the strong dependence of the variability of the series on its own past. Some of the non-linearities of these series relate to a non-constant conditional variance, and they are characterized in general by the clustering of large shocks, or *heteroskedasticity*.

The *AutoRegressive Conditional Heteroskedastic* (ARCH) model of Engle (1982) was the first formal model which successfully addressed the problem of heteroskedasticity. In this model, the conditional variance of the time series is represented by an autoregressive process, namely a weighted sum of squared preceding observations. In practical applications, the order of the calibrated model turns out to be rather large. On the other hand, if we let the conditional variance depend not only on the past values of the time series, but also on a moving average of past conditional variances, the resulting model allows for a more parsimonious representation of the data. The *Generalized AutoRegressive Conditional Heteroskedastic* GARCH(p, q) model of Bollerslev (1986) is defined as:

$$X_t = \varepsilon_t \sigma_t, \quad \text{with } \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i h_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (21)$$

where ε_t are i.i.d. with zero mean and finite variance, and the coefficients have to satisfy $\alpha_i, \beta_j \geq 0, \alpha_0 > 0$ in order to ensure that the conditional variance is strictly positive. The identification and estimation of GARCH models is performed analogously to that of (S)AR(IMA) models; ML is the preferred algorithm. By itself, the GARCH model is not attractive for short-term EPF; however, when coupled with an AR-type model, it presents an interesting alternative: the (S)AR(IMA)-GARCH model, where the residuals of the regression part are modeled further with a GARCH process. Although electricity prices exhibit heteroskedasticity, the general experience with GARCH-type components in EPF models is mixed. There are cases where modeling heteroskedasticity is advantageous, but there are at least as many examples where such models perform poorly.

In one of the first applications of GARCH models to electricity markets, Knittel and Roberts (2005) evaluate an AR-EGARCH specification and find it to be superior to five other time series models during the crisis period (May 1, 2000–August 31, 2000) in California. However, during the pre-crisis period (April 1, 1998–April 30, 2000), the AR-EGARCH process yields the worst forecasts of all models examined. A similar result is obtained by Garcia et al. (2005), who study ARIMA models with GARCH residuals and conclude that ARIMA-GARCH outperforms a generic ARIMA model, but only when high volatility and price spikes are present.

Diongue, Guegan, and Vignal (2009) investigate conditional mean and conditional variance forecasts using a dynamic model following a k -factor GIGARCH process, and apply this method to the German EEX prices in the years 2000–2002. The forecasting performance of the model (up

to one month ahead) is compared with that of a SARIMA-GARCH benchmark model, and the empirical evidence shows that the proposed model outperforms the benchmark.

In an extensive empirical study, Karakatsani and Bunn (2010) apply three complementary modeling approaches in order to uncover the fundamental and behavioral drivers of the electricity price volatility both over time and across intra-day trading periods. They attribute the residual volatility to regular, non-linear agent reactions to market fundamentals (covariates of heteroskedasticity), the adaptation of price formation due to substantial agent learning (time-varying effects), and the transient extreme pricing in periods of scarcity (regime-switching dynamics). Considering a number of GARCH-type models, they find that (i) GARCH effects diminish when each of the above sources of volatility is accounted for, and (ii) allowing for the time-varying responses of prices to fundamentals can yield more precise volatility estimates than an explicit GARCH specification.

Tan et al. (2010) use a wavelet transform to decompose historical price series, then predict each subseries separately using either an ARIMA-GARCH model (for the approximation series) or a GARCH model (for three detail series). This method is examined in the Spanish and PJM electricity markets and compared to various other methods, including the fuzzy neural network of Amjady (2006). In a related paper, Wu and Shahidehpour (2010) present a hybrid ARMAX-GARCH adaptive wavelet neural network model, and test it using PJM market data. The ARMAX model is used to catch the linear relationship between the price return series and the explanatory variable (load); the GARCH model is used to unveil the heteroskedastic character of residuals; and the wavelet neural network is used to present the nonlinear, nonstationary impact of load series on electricity prices.

Gianfreda and Grossi (2012) investigate the impact of technologies, market concentration, congestions and volumes on price dynamics in the Italian power market. Implementing the Reg-ARFIMA-GARCH models of Koopman et al. (2007), they assess the forecasting performances of selected models and show that the models perform better when these factors are considered. In a related study, Hurman, Ravazzolo, and Zhou (2012) consider GARCH-type time-varying volatility models. They find that models augmented with weather forecasts statistically outperform specifications which ignore this information in the density forecasting of Scandinavian day-ahead electricity prices.

3.8.7. Strengths and weaknesses

Statistical methods forecast the current price by using a mathematical combination of previous prices and/or previous or current values of exogenous factors. The forecasting accuracy depends not only on the numerical efficiency of the algorithms employed, but also on the quality of the data analyzed, and the ability to incorporate important fundamental factors, such as historical demand, demand and consumption forecasts, weather forecasts or fuel prices.

Some authors classify statistical models as *technical analysis* tools. Technical analysts do not attempt to

measure an asset's intrinsic or fundamental value; instead, they look at price charts for patterns and indicators that will determine an asset's future performance. While the efficiency and usefulness of technical analysis in financial markets is often questioned, the methods stand a better chance in power markets, because of the seasonality prevailing in electricity price processes during normal, non-spiky periods.

In the presence of spikes, however, statistical methods perform rather poorly. This is especially true for price-only models, but models with fundamental variables do not perform well either. While it is clear that price spikes should be captured using an adequate stochastic model, the literature does not agree as to whether or not these observations have to be included in the estimation process of statistical models. In a recent extensive simulation study, Janczura et al. (2013) show that a better in-sample fit can be achieved by filtering average daily prices with some 'reasonable' procedure for outlier detection, then calibrating the seasonal and stochastic components of the model to spike-filtered data. In the context of forecasting hourly day-ahead prices, some authors also recommend filtering out spikes before calibrating AR-type or neural network models, see e.g. Conejo, Contreras et al. (2005), Contreras et al. (2003), Nogales et al. (2002), Shahidehpour et al. (2002) and Weron and Misiorek (2008).

A list of 'reasonable' spike detection methods includes recursive filters (Cartea & Figueroa, 2005; Weron, 2008), variable price thresholds (Trück, Weron, & Wolff, 2007), fixed price change thresholds (Bierbrauer et al., 2004), regime-switching classification (RSC; Janczura et al., 2013), and wavelet filtering (Stevenson, 2001; Weron, 2006). Only fixed price thresholds (see e.g. Boogert & Dupont, 2008; Fanone et al., 2013) are not recommended, because they ignore the long-term trend-seasonal behavior of electricity prices. Once the spikes have been identified, they have to be replaced by 'normal', less spiky values. A non-exhaustive list of solutions includes replacing spikes with a chosen threshold (Shahidehpour et al., 2002), the mean of the two neighboring prices (Weron, 2008), one of the neighboring prices (Geman & Roncoroni, 2006), or 'similar day' values, e.g., the median of all prices having the same weekday and month (Bierbrauer et al., 2007). If a long-term trend-seasonal component (LTSC) is estimated, Janczura et al. (2013) suggest replacing spikes with the LTSC itself. Doing this is like replacing the extraordinary conditions leading to a spike with the typical or 'normal' conditions on that day of the week and season of the year. The replacement of a particular spike may be interpreted as a low marginal cost power plant replacing a very high marginal cost power plant on the marginal cost curve on that day, or the replacement of a day exhibiting an extreme and unanticipated demand with a typical load profile for that day.

3.9. Computational intelligence models

Computational intelligence (CI) is hard to define. As Duch (2007) puts it, CI is "a new buzzword that means different things to different people". We like to think of CI as a very diverse group of nature-inspired computational

techniques that have been developed to solve problems which traditional methods (e.g., statistical) cannot handle efficiently. CI combines elements of learning, evolution and fuzziness to create approaches that are capable of adapting to complex dynamic systems, and may be regarded as 'intelligent' in this sense. Some authors use the term computational intelligence as a synonym for *artificial intelligence* (AI), see e.g. Poole, Mackworth, and Goebel (1998) and nearly the entire EPF literature. Others see it as an offshoot of AI (Konar, 2005; Rutkowski, 2008). We identify more with the latter approach, and use the term computational intelligence throughout the remainder of the article. We should note that other names for CI techniques may be encountered in the literature as well, such as *non-parametric* or *non-linear statistical*. However, these terms are too narrow or conflict with other classes of methods. For instance, there are both non-parametric (e.g., kernel density estimator) and non-linear (e.g., threshold AR) techniques that are generally classified as belonging to the group of statistical methods, see Section 3.8.

Artificial neural networks, fuzzy systems, support vector machines (SVM) and evolutionary computation (genetic algorithms, evolutionary programming, swarm intelligence) are unquestionably the main classes of CI techniques. Some authors also 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). CI can also be associated with *soft computing*, *machine learning*, *data mining* and *cybernetics* (Madani, Correia, Rosa & Filipe, 2011; Wang & Fu, 2005).

CI models are flexible and can handle complexity and non-linearity. This makes them promising for short-term predictions, and a number of authors have reported their excellent performance in EPF. As in load forecasting, artificial neural networks have probably received the most attention (Aggarwal et al., 2009a,b; Weron, 2006). Other non-parametric techniques, such as fuzzy logic, genetic algorithms, evolutionary programming and swarm intelligence, have also been applied, but typically in hybrid constructions.

3.9.1. Taxonomy of neural networks

Every *artificial neural network* (ANN, NN) model can be classified in terms of its architecture and learning algorithm. The architecture (or topology) describes the neural connections, and the learning (or training) algorithm provides information on how the ANN adapts its weight for every training vector. In the EPF context, ANN models may also be classified depending on the number of output nodes. The first group includes those that have only one output node, and are used to forecast the next hour's price (see e.g. Gonzalez, San Roque, & Garcia-Gonzalez, 2005; Mandal, Senjyu, & Funabashi, 2006), the price h hours ahead (see e.g. Amjady, 2006; Hu et al., 2008; Rodriguez & Anders, 2004), the next day's peak price (see e.g. Areekul, Senju, Toyama, Chakraborty, Yona, & Urasaki, 2010), the next day's average on-peak price (see e.g. Guo & Luh, 2004; Zhang & Luh, 2005), or the next day's average

baseload price (see e.g. Pao, 2006). The second, less popular, group includes those that have several output nodes and forecast a vector of prices, typically 24 (or 48) nodes for forecasting the next day's complete price profile (see e.g. Yamin, Shahidehpour, & Li, 2004).

Network nodes (or 'neurons') are arranged in a relatively small number of connected layers of elements between network inputs and outputs, see Fig. 10. The outputs are linear or non-linear functions of the inputs. The inputs may be the outputs of other network elements, as well as actual network inputs. In terms of architecture, ANNs may be classified into two main categories: (i) *feed-forward networks*, which have no loops, and (ii) *recurrent (or feedback) networks*, in which loops occur because of feedback connections. The feed-forward networks are generally preferred for forecasting, whereas recurrent networks excel in pattern classification and categorization (Jain, Mao, & Mohiuddin, 1996; Rutkowski, 2008).

ANN models can be used to obtain not only point forecasts but also prediction intervals (PI, i.e., interval forecasts). Note that many publications mistakenly refer to PIs as 'confidence intervals', see De Gooijer and Hyndman (2006), Hyndman (2013), and Section 4.2.1. There are five main approaches to computing PIs in the ANN literature: resampling (or bootstrapping; this is the most popular), parameter perturbation, 'delta' (which interprets the ANN as a nonlinear regression model and applies asymptotic theories for the construction of PIs), mean-variance estimation (MVE; this estimates the variance using a dedicated ANN) and Bayesian inference. For reviews and discussions, see e.g. Khosravi, Nahavandi, Creighton, and Atiya (2011) and Zhang and Luh (2005).

3.9.2. Feed-forward neural networks

The simplest network, a *single-layer perceptron*, contains no hidden layers and is equivalent to a linear regression. The forecasts are obtained by a linear combination of the inputs. The weights (corresponding to the coefficients of the regression) are selected using a 'learning algorithm' that minimizes some cost function, e.g., the mean squared error (Hyndman & Athanasopoulos, 2013). By adding an intermediate layer with hidden nodes, we obtain the non-linear *multi-layer perceptron* (MLP). This most common family of feed-forward networks has neurons organized into layers that have unidirectional connections between them; that is, the outputs of the nodes in one layer are inputs to the next layer. The *radial basis function* (RBF) network is a special class of feed-forward networks. It has two layers: each node in the hidden layer employs a radial basis function (with the most common being a Gaussian kernel, see Fig. 10) as the activation function. In contrast, the activation functions of MLP are typically piecewise linear or sigmoid. Amjady and Hemmati (2006) note that RBF networks are effective in exploiting local data characteristics, while MLP networks are good at capturing global data trends.

Back-propagation, which may be regarded as a gradient steepest descent method, is by far the most popular training algorithm for the MLP (Zhang, Patuwo, & Hu, 1998), including in EPF applications (Aggarwal et al., 2009a). It uses continuously valued functions and supervised learning.

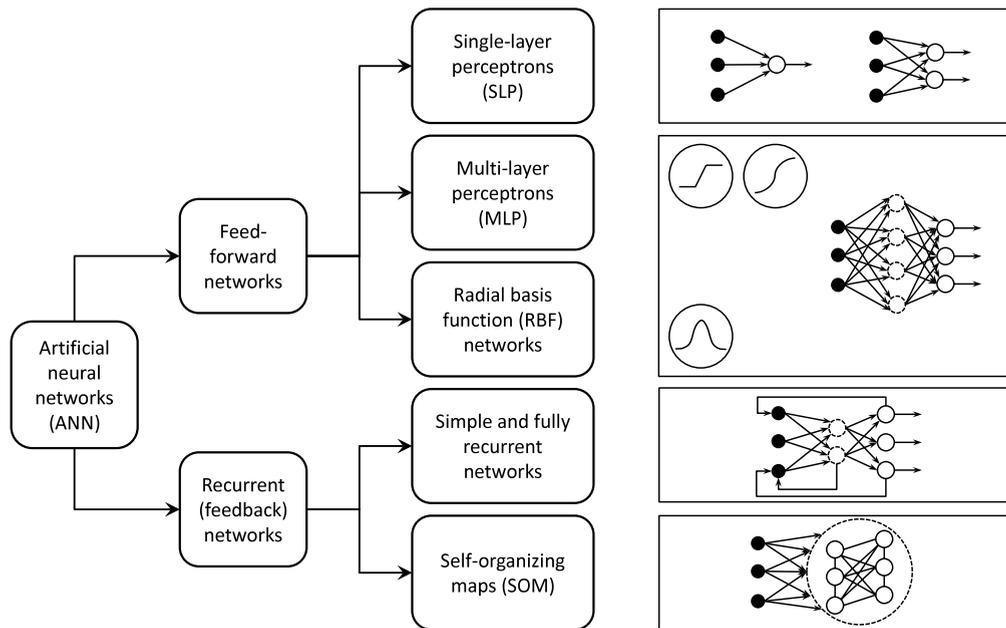


Fig. 10. A taxonomy of the network architectures that are most popular in EPF. Input nodes are denoted by filled circles, output nodes by empty circles, and nodes in the hidden layer by empty circles with a dashed outline. The activation functions for RBF networks are radial basis functions, like a Gaussian kernel, while MLP typically use piecewise linear or sigmoid activation functions.

The Levenberg–Marquardt algorithm is the second most popular training procedure; for sample applications in EPF, see e.g. Catalão et al. (2007); Pindoriya, Singh, and Singh (2008); and Rodriguez and Anders (2004). Amjady (2007) argues that it trains a network 10–100 times faster than back-propagation. However, alternative procedures have also been suggested. For instance, Amjady and Hemmati (2009) propose a hybrid system in which a real-coded genetic algorithm (RCGA) with an enhanced stochastic search capability is used to train a MLP, while cross-validation, repetitive training and archiving techniques enhance its generalization capability. They show that the method can provide more accurate results for the Spanish market than a standard ARIMA model, a wavelet-ARIMA model or a fuzzy ANN (see Section 3.9.4). Pao (2006) employs a generalized delta learning rule, while Zhang and Luh (2005) use the Kalman filter.

The most common training algorithm for the RBF network is a *two-step hybrid learning* algorithm: first, kernel positions and kernel widths are estimated using an unsupervised clustering algorithm, then a supervised least mean square algorithm is employed to determine the connection weights between the hidden layer and the output layer. This hybrid algorithm converges much faster than the back-propagation. However, for many problems, the RBF network often involves a larger number of hidden units than a corresponding MLP, and the final efficiencies of the two ANN structures are problem-dependent (Jain et al., 1996; Rutkowski, 2008).

In a pre-operational training period, the weights assigned to neuron connections are determined by matching historical time, weather, fuel and demand data to historical electricity prices. However, more complex constructions are also used. For instance, Garetta, Romeo, and Gil

(2006) use a combination of univariate MLP networks, in which three auxiliary networks forecast maximum, minimum and medium values of the price, and then this information is fed to five principal MLP networks in order to forecast the electricity price. Hu et al. (2008) use a market concentration index – a measure of the oligopolistic structure of the power market – as an input variable for a MLP, and show that it has a considerable impact on the forecasts.

The MLP architecture has been used by Chen, Dong, Meng, Xu, Wong, and Nagan (2012); Cruz et al. (2011); Garcia-Ascanio and Mate (2010); Garetta et al. (2006); Mandal et al. (2006); Pindoriya et al. (2008); and Yamin et al. (2004), among others; while the less popular RBF architecture has been used by Guo and Luh (2003); Lin, Gow, and Tsai (2010); Pindoriya et al. (2008); and Yao, Song, Zhang, and Cheng (2000), among others. It should be noted, however, that the standard MLP and RBF networks are generally used as benchmarks for other more sophisticated techniques, or as elements of hybrid structures. For instance, Gonzalez et al. (2005) propose a hybrid MLP input–output hidden Markov model (IOHMM; see also Section 3.7.2), in which a conditional probability transition matrix governs the probabilities of remaining in the same state, or switching to another. Mori and Awata (2007) combine regression trees (for evaluating if–then rules and classifying input data into clusters) with normalized RBF networks to calculate more accurate one-step-ahead electricity price forecasts.

Keynia and Amjady (2008) use a hybrid MLP-type model that involves wavelet decomposition, a mixed data model that includes time- and wavelet-domain features, a relief algorithm for feature selection, and a MLP for forecasting and cross-validation. The new algorithm compares favorably with three other MLP

models for PJM data. Amjady and Keynia (2009) propose a MLP in which the numbers of hidden and input neurons are adjusted based on an iterative procedure, after which an evolutionary algorithm is used to make further adjustments to the weights of the network in the neighborhood of the weights found initially. Chaâbane (2014a) models the residuals of an ARFIMA model using a MLP with past prices as inputs (which can be treated as a special case of the recurrent NARX network, see Section 3.9.3). Shafie-Khah et al. (2011) construct a hybrid wavelet-ARIMA-RBF network, in which a RBF network corrects the estimation error of the wavelet-ARIMA forecast. Like in Huang et al. (2005), a particle swarm optimization is used to optimize the network structure. Finally, Guo and Luh (2004) use a ‘committee machine’ composed of one MLP and one RBF network to alleviate the problem of the input–output data misrepresentation by a single ANN. This approach resembles combining forecasts, which will be discussed in Section 4.3.

3.9.3. Recurrent neural networks

Feed-forward networks are classified as static in the sense that they produce only one set of output values, not a sequence of values from a given input. They are also memoryless: their response to an input is independent of the previous network state. On the other hand, *recurrent* (or *feedback*) networks are dynamic systems. When a new input pattern is presented, the neuron outputs are computed. Because of the feedback, the inputs to each neuron are modified, which leads the network to enter a new state.

Simple recurrent networks include Elman and Jordan networks as special cases (see e.g. Jacobsson, 2005). The Elman ANN is a three-layer network with the addition of a set of ‘context units’. There are connections from the hidden (middle) layer to these context units; they have fixed weights (e.g., one) and do not have to be updated during training. As a result, each of the neurons in the hidden layer processes both the external input signals and signals from feedback, but the signals from the output layer are not subject to the feedback operation. In the Jordan networks, the context units (also called ‘the state layer’) are fed from the output layer instead of the hidden layer, and have a recurrent connection to themselves. A more general class is that of *fully recurrent networks*, also known as *real-time recurrent networks* (RTRN). In such structures, the outputs of all neurons are connected recurrently to all neurons in the network. Simple and fully recurrent networks can be trained using gradient algorithms; however, these take a more complex form than is the case of network learning without feedback (Rutkowski, 2008).

Sharma and Srinivasan (2013) combine a FitzHugh–Nagumo model, for mimicking the spiky price behavior, with an Elman network, for regulating the latter, and a feed-forward ANN, for modelling the residuals. The hybrid model thus developed is used for point and interval forecasting in markets in Australia, Ontario, Spain and California. Note that the FitzHugh–Nagumo model had been used previously for the same purpose by Lucheroni (2012). Anbazhagan and Kumarappan (2013) use Elman networks to obtain short-term price forecasts in the market of mainland Spain. They conclude that their

network performs better than a number of other EPF approaches, including ARIMA, wavelet-ARIMA, MLP, fuzzy ANN and wavelet-ARIMA-RBF networks. However, simple recurrent networks are inherently weak in learning time series with long-term dependencies using gradient based algorithms. This ‘forgetting behavior’ (Frasconi, Gori, & Soda, 1992) is due to the so-called *vanishing gradient* property, where, under certain conditions, the fraction of the error gradient that is due to information h time steps in the past decreases exponentially as h increases.

To overcome the vanishing gradient problem, *nonlinear autoregressive models with exogenous inputs* (NARX) have been proposed by Lin, Horne, Tino, and Giles (1996). These recurrent networks also have very good learning capabilities and generalization performances. A typical NARX network is a three-layer feed-forward architecture, with sigmoid activation functions in the hidden layer, linear activation functions in the output layer, and delay lines for storing previous values of the predicted time series, x_t , and the exogenous variables, z_t . The output of the NARX network, x_t , is fed back to the input of the network (through delays: x_{t-1}, \dots, x_{t-p}). In a way, a NARX architecture resembles a Jordan network. At the same time, it is also a neural network (nonlinear) variant of the well-known ARX time series model, see Section 3.8.4. Surprisingly, NARX networks were not used for EPF until very recently, despite the fact that various statistical software packages, like Matlab, offer ready-to-use functions and user-friendly interfaces. To the best of our knowledge, only one paper on EPF has applied an explicit NARX architecture; see also the empirical results discussed in Section 4.3.1. Specifically, Andalib and Atry (2009) use a NARX model to forecast hourly Ontario energy prices (HOEP), where both the lagged values of HOEP and the lagged values of hourly demand are considered as explanatory variables. However, a similar effect is achieved if the inputs to a feed-forward network (e.g., a standard MLP) are past prices. Chaâbane (2014a) even calls the network he uses NAR: a nonlinear autoregressive model.

The networks reviewed thus far can be trained using either supervised (‘with a teacher’, with known answers) learning for pattern classification and forecasting, or unsupervised (‘without a teacher’) learning for data analysis and clustering. *Self-organizing maps* (SOM) are trained using only the latter approach: the learning sequence is made only of input values, without the desired output signal. One of the more popular architectures, known as Kohonen’s SOM, consists of a two-dimensional array of nodes, each of which is connected to all input nodes. It can be used for the projection of multivariate data, density approximation, and clustering. SOM networks have not been used extensively in EPF, but there are examples of applications in hybrid structures. For instance, Fan, Mao, and Chen (2007) and Niu, Liu, and Wu (2010) use SOM classifiers to cluster hourly electricity price data according to their similarities (to resolve the problem of insufficient training data), and then employ support vector machines (see Section 3.9.5) to predict the prices within each subset.

3.9.4. Fuzzy neural networks

Fuzzy logic is a generalization of the usual Boolean logic, in that, instead of an input taking a value of 0 or 1, it has certain qualitative ranges associated with it. For example, a temperature may be low, medium or high. Fuzzy logic allows outputs to be deduced from fuzzy or noisy inputs, and, importantly, there is no need to specify a precise mapping of inputs to outputs. Following the logical processing of fuzzy inputs, a *defuzzification* process may be used in order to produce precise outputs (e.g., prices for particular hours). *Fuzzy neural networks* (FNN) combine the learning and computational power of traditional ANNs with fuzzy logic (Konar, 2005; Rutkowski, 2008). A considerable amount of research attention has been devoted to rule generation using various FNN structures; for reviews in soft computing, see e.g. Mitra and Hayashi (2000) and Wang and Fu (2005).

One of the first applications of fuzzy logic to EPF was performed by Hong and Hsiao (2002), who utilize fuzzy-c-means for classifying historical data into three clusters (peak, medium and off-peak), and then employ a recurrent network for forecasting. Vahidinasab, Jadid, and Kazemi (2008) take a similar approach, but use a MLP for price forecasting. Rodriguez and Anders (2004) build an adaptive-network-based fuzzy inference system (ANFIS), which combines an adaptive mechanism with Sugeno-type rules and uses a combination of the least squares method and back-propagation for training the membership function and the linear combination parameters. They show that the ANFIS performs better than a MLP. Amjady (2006) proposes a FNN which has an inter-layer and a feed-forward architecture and uses a new hypercubic training mechanism. The method is shown to predict Spanish hourly day-ahead electricity prices better than ARIMA, wavelet-ARIMA, MLP or a RBF network.

More recently, Meng, Dong, and Wong (2009) train a RBF network using fuzzy-c-means, and differential evolution is used to auto-configure the network structure and to obtain model parameters. Furthermore, a moving window wavelet de-noising technique is introduced so as to improve the network performance in forecasting Queensland (Australia) electricity prices. Catalão, Pousinho, and Mendes (2011) propose a hybrid approach, which combines a wavelet transform, particle swarm optimization and an adaptive-network-based fuzzy inference system. Finally, Azadeh et al. (2013) present an integrated, multi-step algorithm which combines three ANNs, seven fuzzy regressions (see e.g. Gładysz & Kuchta, 2011) and one standard regression model to provide a joint framework for long-term (annual time scale) EPF. The algorithm switches between the predictions of the different models based on some pre-specified rules. The results indicate that the standard and fuzzy regressions considerably outperform ANNs.

3.9.5. Support vector machines

The *support vector machine* (SVM) is a classification and regression tool that has its roots in Vapnik's (1995) statistical learning theory. In contrast to ANNs, which try to define complex functions of the input space, SVM performs a non-linear mapping of the data into a high dimensional space, then uses simple linear functions to create linear

decision boundaries in the new space. An attractive feature of SVM is that it gives a single solution that is characterized by the global minimum of the optimized functional, rather than multiple solutions associated with local minima, as do ANNs. Furthermore, they rely less heavily on heuristics (i.e., an arbitrary choice of the model) and have a more flexible structure (Čížek, Härdle & Weron, 2011). SVM has been applied widely to pattern classification problems and non-linear regressions. After SVM classifiers have been trained, they can be used to predict future trends. As Wang and Fu (2005) note, the meaning of the term prediction is different in the context of SVM. Here, 'prediction' means a supervised classification that involves two steps: first, a SVM is trained as a classifier using a part of the data, then this classifier is used to classify ('predict') the rest of the data in the data set. The classification may be improved further by introducing individual penalty parameters for each sample and using an AdaBoost-like algorithm in the training phase (Zięba, Tomczak, Lubicz, & Świątek, 2014).

The applications of SVM in electricity price forecasting are typically those of elements in hybrid systems; however, in one of the first papers on this topic, Sansom, Downs, and Saha (2002) compare a MLP and a SVM with the same inputs, and conclude that the SVM produces more consistent forecasts and requires less time for optimal training. Also, Zhao, Dong, Xu, and Wong (2008) employ a SVM to forecast the value of the spot price.

Fan et al. (2007) and Niu et al. (2010) use SOM classifiers to cluster hourly electricity price data according to their similarity, then employ SVM to predict the prices within each subset. Che and Wang (2010) propose a hybrid model called SVRARIMA that combines both support vector regression (SVR; to capture the nonlinear patterns) and ARIMA models. The results demonstrate that the SVRARIMA model outperforms some of the existing ANN approaches and traditional ARIMA models. Yan and Chowdhury (2010b) present a hybrid mid-term (on a time frame between one and six months) EPF model combining least-squares SVM and ARMAX models. The model shows an improved forecasting accuracy for PJM data compared to a forecasting model using a single least-squares SVM. Chaăbane (2014b) proposes a new hybrid model, which exploits the features of ARFIMA and least-squares SVM, and shows that, for Nord Pool data, it outperforms the two individual models when applied separately.

3.9.6. Strengths and weaknesses

The major strength of computational intelligence tools is their ability to handle complexity and non-linearity. In general, CI methods are better at modeling these features of electricity prices than the statistical techniques discussed in Section 3.8. At the same time, this flexibility is also their major weakness. The ability to adapt to non-linear, spiky behaviors may not necessarily result in better point forecasts. This is similar to the case of Markov regime-switching models, which have the potential to model the highly volatile and non-linear price processes, but have been reported to perform poorly in forecasting in general (Bessec & Bouabdallah, 2005; Dacco & Satchell, 1999). The non-linear models have another potential

advantage, though: they should be able to provide better interval and density forecasts than the linear models. However, this has not been investigated extensively to date, see Section 4.2.

Moreover, the pool of available CI tools is so diverse and rich that it is hard to find an optimal solution. Worse still, it is hard to compare the different CI methods thoroughly. Even if the forecasting accuracy is reported for the same market and the same out-of-sample (forecasting) test period, the errors of the individual methods are not truly comparable unless identical in-sample (calibration) periods are used as well, and therefore they cannot be used to formulate general statements about a method's efficiency unless such is the case. Instead, conclusions can only be drawn about the performance of a given implementation of a method, with certain initial conditions (parameters) and for a certain calibration dataset. Although this critique is not limited to CI techniques, it is particularly true in their case because of their non-linearity and their multi-parameter specifications.

4. A look into the future of 'electricity price forecasting'

In the previous sections, we have looked back at the last 15 years of electricity price forecasting, in an attempt to systematize the rapidly growing body of literature and the overwhelming diversity of methods. Now, it is time to look ahead and speculate on the directions EPF will or should take over the next decade or so. In Sections 4.1–4.5, we discuss five main topics, which have been indicated, either explicitly or implicitly, in the preceding sections.

4.1. Fundamental price drivers and input variables

A key point in EPF is the appropriate selection of input variables. On the one hand, the electricity price exhibits seasonality at the daily and weekly levels, and the annual level to some extent. In the short term, the latter may be ignored, but the daily and weekly seasonalities have to be taken into account. In the mid-term, the daily profile becomes irrelevant (and most EPF models work with average daily prices), but the annual seasonality (if present), or a longer-term trend-cycle component, plays a crucial role. Finally, in the long term, when the time horizon is measured in years, the daily, weekly and even annual seasonality may be ignored, and long-term trends dominate.

On the other hand, as has been discussed in previous sections, the electricity spot price is dependent on a large set of fundamental drivers, including system loads (demand, consumption figures), weather variables (temperatures, wind speed, precipitation, solar radiation), fuel costs (oil and natural gas, and to a lesser extent coal), reserve margin (surplus generation, i.e., available generation minus/over predicted demand), and the scheduled maintenance or forced outages of important power grid components. Their historical (past) values and (market or expert) predictions for the forecasting horizon considered are valuable for the construction and proper calibration of the models. Care should be taken, however, as in some periods or markets their influence on the spot price may be

very limited. For instance, Maciejowska (2014) reports for the UK market that fundamental drivers (wind generation, demand, gas price) played a minor role, while speculative or spot price shocks were responsible for up to 95% of the price volatility in 2011 and 2012.

As Amjady and Hemmati (2006) observe, most papers select a combination of these fundamental drivers, based on the heuristics and experience of the forecaster. The model category (multi-agent, fundamental, reduced form, statistical or computational intelligence) and data availability are the other important decision variables. Although 'pure price' models are sometimes encountered in EPF, they are in the minority in the most common day-ahead forecasting scenario. Thus, some input features have to be selected, but their optimal choice remains an open question. The development of an objective method of selecting a minimum set of the most effective input variables would be very valuable. We doubt, however, that one universal set can be found for all power markets.

4.1.1. Modeling and forecasting the trend-seasonal components

In the standard approach to seasonal decomposition, a time series – say, the electricity spot price P_t – is decomposed into the long-term trend-seasonal component (LTSC) T_t , the short-term seasonal component (STSC) s_t , and the remaining variability, error or stochastic component X_t , in either an additive (i.e., $P_t = T_t + s_t + X_t$) or a multiplicative fashion (i.e., $P_t = T_t \cdot s_t \cdot X_t$), see also Section 3.8. Note that in time series analysis, a distinction is drawn between *seasonal* patterns of a fixed period and *cyclic* patterns that exhibit rises and falls that are not of a fixed period (Hyndman & Athanasopoulos, 2013).

The hourly and weekly seasonality – which is due generally to the variable intensity of business activities throughout the week – is usually captured by a combination of the autoregressive structure of the models (i.e., lagged prices are input variables) and dummy variables. The forecasting of such a seasonal pattern is straightforward. To simplify this task even more, some studies perform the forecasts separately across the hours, thus eliminating the need for explicit modeling of the daily price profile, but leading to 24 (or 48) sets of parameters (see e.g. Karakatsani & Bunn, 2008, 2010; Misiorek et al., 2006; Raviv, Bouwman, & van Dijk, 2013). The rationale comes from (i) the demand forecasting literature, which has generally favored the multi-model specification for short-term predictions (Bunn, 2000; Shahidehpour et al., 2002), (ii) an argument that each hour displays a rather distinct price profile, reflecting the daily variation of demand, costs and operational constraints (Karakatsani & Bunn, 2008; Weron, 2006), and (iii) the day-ahead market structure, where the delivery of electricity during a particular hour is a different contract from delivery in the next hour (see Section 3.1). The weekly dummies typically do not cover the whole week but are restricted to the more distinct days, e.g., Monday, Saturday and Sunday (Weron & Misiorek, 2008) or Monday, Friday, Saturday and Sunday (Kristiansen, 2012).

The annual seasonality is present in electricity spot prices (due to changing weather conditions throughout the year), but in most cases it is dominated by a more irregular

cyclic component that depends on macroeconomic variables (e.g., fuel prices, economic growth) and long-term weather trends (e.g., lower than historical precipitation or temperatures). In the time series literature, this would be called a *trend-cycle* component; in electricity price modeling, it is referred to instead as a *trend-seasonal* component, to reflect the underlying annual seasonality. There are essentially three approaches to modeling the LTSC in electricity spot prices:

- piecewise constant functions or dummies, possibly combined with a linear trend (Fanone et al., 2013; Fleten, Heggedal, & Siddiqui, 2011; Gianfreda & Grossi, 2012; Haugom & Ullrich, 2012; Higgs & Worthington, 2008; Knittel & Roberts, 2005);
- sinusoidal functions or sums of sinusoidal functions of different frequencies (Benth et al., 2012; Bierbrauer et al., 2007; Carrea & Figueroa, 2005; De Jong, 2006; Geman & Roncoroni, 2006; Seifert & Uhrig-Homburg, 2007; Weron, 2008);
- wavelets (Conejo, Contreras et al., 2005; Janczura & Weron, 2010, 2012; Schlueter, 2010; Stevenson, 2001; Stevenson, Amaral, & Peat, 2006; Weron, 2006, 2009; Weron, Bierbrauer et al., 2004; Weron, Simonsen et al., 2004) or other nonparametric smoothing techniques like Friedman's supersmoother, the Hodrick–Prescott filter, spline functions, empirical mode decomposition, and singular spectrum analysis (Bordignon et al., 2013; Lisi & Nan, 2014; Weron & Zator, 2014b).

When building stochastic models for EPF in the mid-term, the problem that is of the utmost importance is the estimation and consequent forecasting of the trend-seasonal components in the data. While the STSC is less important for derivatives valuation and risk management applications, the LTSC is crucial for the accuracy of the simulation-based spot price models. A misspecification of the LTSC can introduce biases or artificial price variability. This may result in a bad estimate of the mean reversion level or of the price spike intensity and severity, and consequently, in underestimating the risk, and even in incurring financial losses (Janczura et al., 2013; Trück et al., 2007). For instance, consider Nord Pool spot prices for the evening peak hour (5 pm–6 pm) over the two-year period 1.1.2012–31.12.2013. If we fit a wavelet-based LTSC (here using six levels of decomposition, S_6 , and the Daubechies wavelet of order 12; for details, see Nowotarski, Tomczyk, & Weron, 2013), a sine (of variable period, amplitude and phase shift), and monthly dummies, and subtract them from the prices (together with the weekly dummies), we will obtain three different stochastic components: $X_t^{(i)} = P_t - T_t^{(i)} - s_t^{(i)}$, where $i =$ 'wavelet', 'sine' or 'monthly dummies', see Fig. 11. Next, if we calibrate a stochastic model – here, for simplicity, a MRJD defined by Eq. (7) – we will obtain different parameters, potentially leading to significantly different sample trajectories, as in Fig. 12. In this example, only the wavelet-based LTSC yields a reasonable stochastic model, with the other two approaches underestimating the mean jump size and overestimating the spike occurrence. Apparently the jump component tries to correct for deviations from the mean-reverting behavior of the sine and monthly dummies-implied stochastic components.

Forecasting a piecewise constant or a sinusoidal LTSC is straightforward, but the in-sample fit is generally poor, yielding a sub-optimal model for the stochastic component. On the other hand, forecasting a nonparametric seasonal component is particularly troublesome, and some authors only actually evaluate the out-of-sample prediction of the stochastic part X_t , without considering the LTSC (see e.g. Bordignon et al., 2013). In a large simulation study, Nowotarski et al. (2013) consider a battery of over 300 models (including monthly dummies and models based on Fourier or wavelet decompositions, combined with linear or exponential decay) and find that the wavelet-based models are significantly better than the commonly used monthly dummies and sine-based models, in terms of forecasting spot prices up to a year ahead. This result calls into question the validity and usefulness of stochastic models of spot electricity prices built on the latter two types of LTSC models.

The overall impression is that the issue of seasonality has been downplayed in the EPF literature. In our opinion, this is a serious shortcoming, and efforts should be made to address it properly in future research (see also Janczura et al., 2013; Lisi & Nan, 2014; Nowotarski, Raviv, Trück, & Weron, 2014). While an inadequate treatment of seasonality will only lead to worse forecasts in the day-ahead context, for longer-term predictions it may result in a critical flaw in the constructed EPF model.

4.1.2. Spike forecasting and the reserve margin

When it comes to volatility or price spike forecasts, the reduced-form models discussed in Section 3.7 have been reported to perform reasonably well. For instance, Becker et al. (2007) demonstrate that a time-varying probability regime-switching model can help to predict price spikes in Queensland, Australia. Chan et al. (2008) find that, while a large proportion of the total realized spot price variation is attributable to the continuous ('base regime') part of the price process, a modest increase in the volatility forecast accuracy can be obtained by dividing the total variation into its jump and non-jump components in a jump-diffusion framework. On the other hand, Christensen et al. (2012) take an approach that is 'orthogonal' to the rest of the EPF literature and consider the time series of price spikes, not the time series of spot prices. They study half-hourly data from the extremely spiky Australian market; it is probable that data from other markets would not contain enough spikes to calibrate the models. The authors treat the time series of spikes as a discrete-time point process and represent it as a nonlinear variant of the autoregressive conditional hazard (ACH) model. They compute one-step-ahead forecasts of the probability of a price spike for each half hour in the forecast period (July–September 2007), and conclude that the ACH model performs better than the benchmark logit model. Finally, Christensen et al. (2012) explore the profitability of an informal trading strategy utilizing electricity futures contracts and spike forecasts from the two models. They conclude that using the futures market as a hedge based on the forecasts of the ACH model has the potential to provide significant returns: more than 20% in the out-of-sample period considered for the NSW and Victoria markets. However, transaction

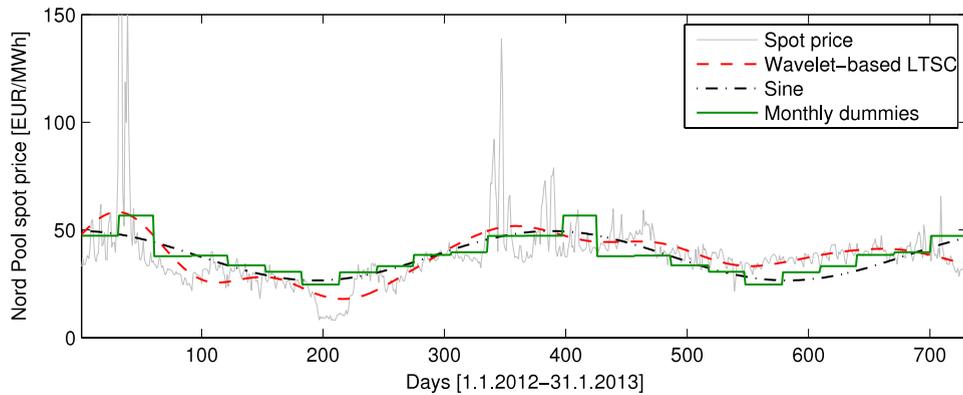


Fig. 11. Nord Pool system spot prices for the evening peak hour (5 pm–6 pm) over the two-year period 1.1.2012–31.12.2013, together with three estimated LTSC: wavelet-based LTSC (here using six levels of decomposition, S_6 , and the Daubechies wavelet of order 12; for details, see Nowotarski et al., 2013), a sine (here: $f(t) = 11.46 \sin(1.88t + 1.60)$) and monthly dummies.

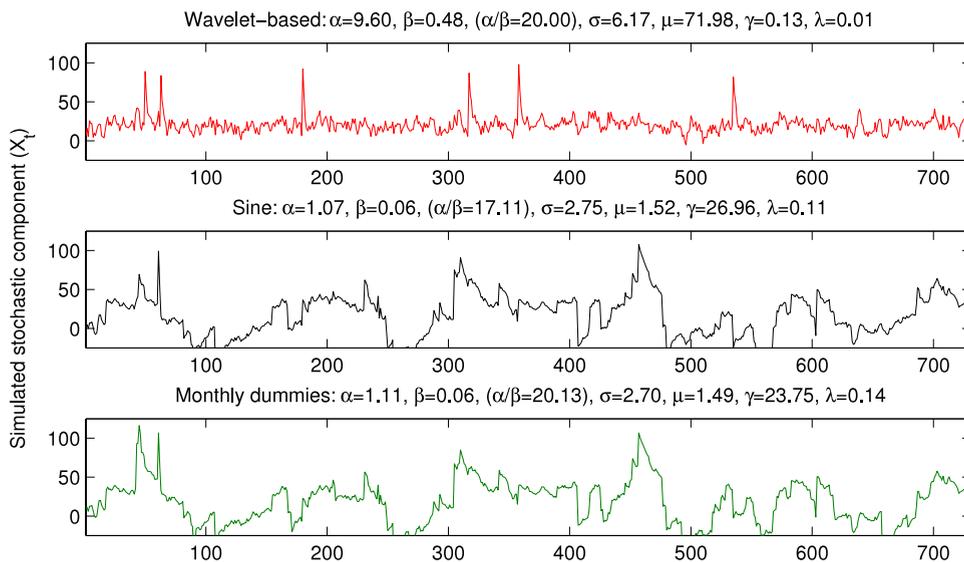


Fig. 12. Sample simulated trajectories of a MRJD model fitted to the stochastic components obtained by subtracting each of the LTSC (wavelet-based, sine, monthly dummies) from the Nord Pool spot price, see Fig. 11. Note the significant differences in the parameters of the MRJD model; for parameter definitions, see Eq. (7). All three trajectories were obtained using the same set of random numbers.

costs are not taken into account and synthetic contracts are priced artificially (due to the unavailability of intra-day futures prices).

One may wonder whether spike forecasting could be improved further by considering fundamental variables. Indeed it could. One of the most influential fundamental variables, especially when it comes to predicting spike occurrences or spot price volatility, is the *reserve margin*, also called *surplus generation*. It relates the available capacity (generation, supply), C_t , to the demand (load), D_t , at a given moment in time t . The traditional engineering notion of the reserve margin defines it as the difference between the two, i.e., $RM = C_t - D_t$ (see e.g. Eydeland & Wolyniec, 2003; Harris, 2006). However, some authors prefer to work with dimensionless ratios, $\rho_t = \frac{D_t}{C_t}$ (Anderson & Davison, 2008; Cartea, Figueroa, & Geman, 2009; Davison, Anderson, Marcus, & Anderson, 2002; Maryniak, 2013; Maryniak & Weron, 2014), $R_t = \frac{C_t}{D_t} - 1$ (Mount et al., 2006; Zareipour et al., 2006; Zareipour, Janjani, Leung, Motamedi,

& Schellenberg, 2011), or the so-called *capacity utilization* $CU = 1 - \frac{D_t}{C_t}$ (Boogert & Dupont, 2008).

The reserve margin has seen some limited application in electricity spot price modeling and forecasting. For instance, Zareipour et al. (2006) evaluate the usefulness of publicly available electricity market information in forecasting the hourly Ontario energy price (HOEP), and find that the reserve margin is a useful indicator. Anderson and Davison (2008) and Davison et al. (2002) propose a functional form for the relationship between the reserve margin and the probability of a spike. Burger, Klar, Müller, and Schindlmayr (2004) incorporate a function of the demand to capacity ('relative availability of power plants') ratio ρ_t into their spot price model. Boogert and Dupont (2008) assume that the spot price is a function of capacity utilization (which they call 'reserve margin'), and estimate its empirical form for Dutch electricity prices. Mount et al. (2006) propose a MRS model (see Section 3.7.2) where the switching probabilities and the conditional mean of the

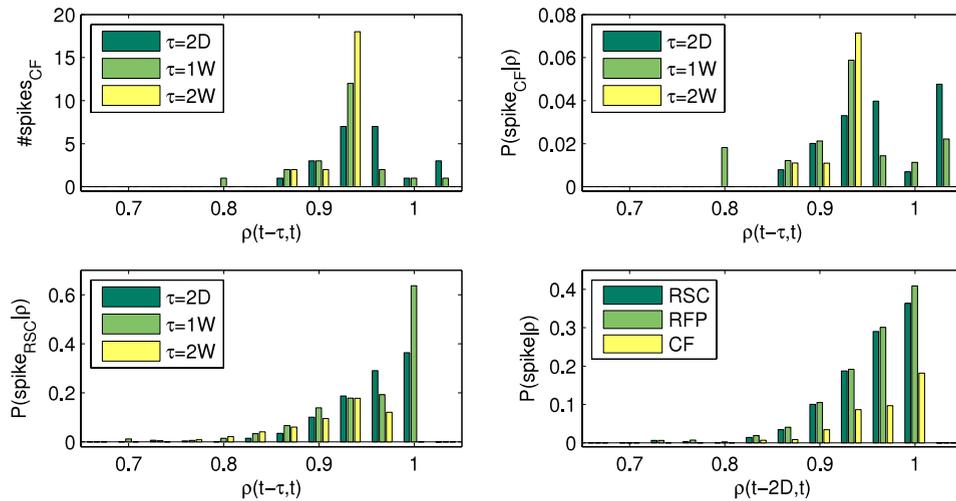


Fig. 13. Upper left: The number of spikes against the demand-to-capacity ratio, i.e., $\rho(t - \tau, t)$, for $\tau =$ two days, one week and two weeks in the period 1.6.2003–31.3.2006. Note that for $\tau =$ one and two weeks, most spikes are observed near $\rho = 0.93$, as was reported by Cartea et al. (2009). The spikes are identified using the approach of Cartea and Figueroa (2005). Upper right: The probability of observing a spike $P(\text{spike}_{CF}|\rho)$ for a given ρ in the same period. Lower left: The probability of observing a spike $P(\text{spike}_{RSC}|\rho)$ for a given ρ in the more recent period 1.1.2006–31.12.2012. The spikes are identified using the RSC method (see Janczura et al., 2013, and Fig. 9). Lower right: The probability of observing a spike $P(\text{spike}|\rho)$ for a given $\rho(t - 2D, t)$ in the same period. The spikes identified using the RSC, RFP (see Fig. 9) and CF methods. Note that the dark green bars in the lower panels illustrate the same data, only the scale is different.

spot price in each regime vary with both time and the reserve margin.

While it is beyond doubt that the reserve margin is a valuable explanatory variable, it remains an open question as to how such data can be obtained and used for forecasting. An interesting approach is taken by Cartea et al. (2009), who work with publicly available forecasts for the UK market (see www.bmreports.com), and consider a variant of the demand-to-capacity ratio:

$$\rho(t_1, t_2) = \frac{D(t_1, t_2)}{C(t_1, t_2)}, \quad (22)$$

where $D(t_1, t_2)$ is the National Demand Forecast (also referred to as Indicated Demand) and $C(t_1, t_2)$ is the predicted Generation Capacity (also referred to as Indicated Generation), and both are calculated at time t_1 (e.g., today) for an upcoming period t_2 . The period t_2 may be a day or a week, and the forecast horizon ranges from two days to 52 weeks. Although it is unlikely, the demand-to-capacity ratio (Eq. (22)) can take values that are higher than unity because it is based on forecasts, not actual values. Such situations have indeed occurred in the British market in the period considered by Cartea et al., i.e., June 2003–March 2006. Analyzing $\rho(t - 1W, t)$ ratios, i.e., forecasts for week t available one week earlier, they find that, except in a few cases, all spikes appear when $\rho \in [0.908, 0.960]$; see the upper left panel in Fig. 13. This is surprising, given that much higher values of the ratio have been observed: up to 1.097 for 2-day-ahead, 1.069 for 1-week-ahead and 1.031 for 2-week-ahead forecasts. It is as if, once the demand-to-capacity ratio exceeds a certain, very high level, the supply (and perhaps the generation) side(s) of the market do everything they can to prevent spikes, while for high but not extremely high values of ρ they are not very concerned with the situation and make no serious attempt to prevent them.

In follow-up studies, Maryniak (2013) and Maryniak and Weron (2014) look at more recent data (up to December 2012) and check how the results vary over time and how they depend on the definition of a spike. The dataset used in those papers, and also here, covers the period 19.1.2003–31.12.2012, and consists of (i) APX-UK average daily spot prices (see the upper left panel in Fig. 9), (ii) National Demand Forecasts (the forecasts are published daily for 2–14 days ahead, and once a week for 2–52 weeks ahead), and (iii) surplus forecasts (i.e., reserve margin forecasts; 2- to 14-day-ahead forecasts published on weekdays, and 2- to 52-week-ahead forecasts once a week). The latter two sets were obtained from Elexon (www.elexon.co.uk), a company that runs the British balancing market. Since not all forecasts are available on a daily basis, we use the most recent available value as a proxy for that day's forecast.

If we plot the number of spikes against the demand-to-capacity ratio, i.e., $\rho(t - \tau, t)$, for $\tau =$ two days, one week and two weeks, then we can observe that most spikes cluster near $\rho = 0.93$, which coincides with the results of Cartea et al. (2009). The time period considered is the same, i.e., 1.6.2003–31.3.2006, but the number of spikes (i.e., 22) is larger than in their Figure 2 and Table 2 (i.e., 13). Hence, our results are not as clear-cut as theirs. The difference stems from the fact that, while we use the same approach (denoted by CF; see also Cartea & Figueroa, 2005) as they do in the calibration of their regime-switching, reserve margin-dependent model, Cartea et al. only identify as 'spikes' in their Figure 2 and Table 2 those prices which correspond to the 'peaks' of the multi-day spikes. Moreover, if we plot the empirical probability of observing a spike $P(\text{spike}_{CF}|\rho)$, $\rho = 0.93$ no longer seems so special, especially for $\tau =$ two days, see the upper right panel in Fig. 13.

The changes that took place in 2005 have had a substantial impact on the structure and behavior of the British power market. In April 2005, the NETA system was replaced by BETTA, which covered not only England and Wales, but also Scotland. As a result of investments in generation, the supply side has seen a further increase in capacity in the years since, leading to a larger reserve margin and fewer spikes. In the lower panels of Fig. 13, we plot the empirical probabilities of observing spikes in the more recent period 1.1.2006–31.12.2012. In the lower left panel, we show the probability of observing a spike $P(\text{spike}_{\text{RSC}}|\rho)$ for a given ρ , with the spikes being identified using a regime-switching classification (RSC; see Janczura et al., 2013, and Fig. 9). The lower right panel shows the probability of observing a spike $P(\text{spike}|\rho)$ for a given $\rho(t - 2D, t)$, with the spikes being identified using three methods: RSC, recursive filter on prices (RFP; see Janczura et al., 2013, and Fig. 9), and the CF method of Cartea and Figueroa (2005). Clearly, irrespective of the spike identification method, the probability of a spike increases with an increased demand-to-capacity ratio, at least for the shorter forecasting horizons ($\tau =$ two days and one week). It seems that for the two-week-ahead forecasts, there is still ample time to take appropriate countermeasures in the case of very high values of ρ , so as to reduce the probability of a spike to zero (see the lower left panel).

Interestingly, the results obtained are in line with the ‘industrial standard’ of 85% for the demand-to-capacity ratio that warrants a safe functioning of the power system (Anderson & Davison, 2008). The probability of spikes $\rho(t - 2D, t)$ is below 2% for $\rho < 85\%$, and well below 1% for $\rho < 82\%$. On the other hand, it is substantially higher for values of ρ above this threshold: up to 40% for $\tau =$ two days and up to 60% for $\tau =$ one week! This is clear indication that the reserve margin has a huge potential for explaining the spike probability, as was conjectured by Christensen et al. (2009). Its rare application in EPF can be justified only by the difficulty of obtaining good quality reserve margin data. Given that more and more system operators are disclosing such information nowadays, reserve margin data should be playing a significant role in EPF in the near future.

4.2. Beyond point forecasts

According to the comprehensive review study by De Gooijer and Hyndman (2006) on forecasting time series, the use of prediction intervals and densities, or *probabilistic forecasting*, has become much more common over the past three decades, as practitioners have come to understand the limitations of point forecasts. This does not seem to be the case in EPF. The EPF Scopus query, see footnote 1, when modified to include AND (‘prediction interval’ OR ‘interval forecast’ OR ‘confidence interval’) yielded only 16 articles and conference papers (out of 480 EPF publications, see Section 2.1). Density forecasts are even less popular: the same Scopus query modified to include AND ‘density forecast’ returned only one article. However, as Amjady and Hemmati (2006) remark, electrical engineers are aware that high-quality market clearing price prediction intervals (PI) would help utilities to submit effective bids with low risks.

4.2.1. Interval forecasts

It should be noted that, as in the general forecasting literature, some authors use the term *confidence interval* instead of *prediction interval* (PI). A PI is associated with a random variable (e.g., electricity price) that is yet to be observed, while a confidence interval is associated with a parameter of a model, see Hyndman (2013) for a discussion. In most forecasting applications we are interested in PIs, i.e., intervals which contain the true values of future observations with a specified probability, not in confidence intervals.

When forecasting one step ahead, which is definitely the most common setup in EPF, the standard deviation of the forecast distribution is the same as (if there are no parameters to be estimated, as in the naïve method, see Sections 3.3 and 3.8.1), or slightly larger than (because of the uncertainty associated with model selection and parameter estimation), the residual standard deviation, see Hyndman and Athanasopoulos (2013). This difference is often ignored, including in multi-step-ahead forecasts, meaning that many model-based PIs are too narrow. One way to address this problem is to use bootstrapping, see e.g. Cao (1999) and De Gooijer and Hyndman (2006). See also Hansen (2006), who constructs asymptotic forecast intervals that incorporate the uncertainty due to parameter estimation, by incorporating a simple proportional adjustment of the interval endpoints which depends on the asymptotic variance of the interval estimates.

In one of the first publications on interval EPF, Zhang, Luh, and Kasiviswanathan (2003) develop an algorithm for obtaining the PIs (which they call ‘confidence intervals’) from a cascaded ANN model by using the Quasi-Newton method. In a follow-up paper, Zhang and Luh (2005) present a modified U-D factorization method within the decoupled extended Kalman filter framework. The computational speed and numerical stability of this method are improved significantly relative to the earlier method. The new method also provides smaller PIs.

Misiorek et al. (2006) compare the accuracies of seven relatively parsimonious time series methods for day-ahead EPF (see also Section 3.8.5), and evaluate their performances in terms of one-step-ahead point (for all models) and interval (for four models) forecasts. The latter (called ‘confidence intervals’) are determined analytically as quantiles of the error term density (for ARX, ARX-GARCH and TARX models), or using Monte Carlo simulations (for the MRS model). Misiorek et al. evaluate the quality of the PIs by comparing the nominal coverages of the models to the true coverage, and conclude that TARX models outperform their competitors in both point and interval forecasting.

In a follow-up study, Weron and Misiorek (2008) compare the accuracies of 12 time series models (for a discussion, see Section 3.8.4), and evaluate their performances in terms of one-step-ahead point and interval forecasts. Two types of PIs are computed: distribution-based and empirical. The method of calculating empirical PIs resembles the estimation of the Value-at-Risk via historical simulation, and consists of computing sample quantiles of the empirical distribution of the one-step-ahead prediction errors. The distribution-based PIs are computed as quantiles of

the error term density: Gaussian for AR-type models and kernel estimator-implied for the semiparametric models. Then, Weron and Misiorek use the conditional coverage test of Christoffersen (1998) to evaluate the quality of the PIs, and find that the semiparametric models, and SNARX in particular, generally lead to better interval forecasts than their competitors, and also, more importantly, have the potential to perform well under diverse market conditions.

Zhao et al. (2008) propose a data mining-based approach in order to achieve two major objectives: to forecast the electricity spot price and to estimate the respective PIs. In the proposed approach, a support vector machine (SVM) is employed to forecast the value of the spot price. To forecast the PIs, the authors construct a statistical model by introducing a heteroskedastic variance equation to the SVM. Their empirical results show that the proposed method is highly effective relative to existing methods such as GARCH models.

Serinaldi (2011) introduces the class of Generalized Additive Models for Location, Scale and Shape (GAMLSS) for forecasting the dynamically varying distribution of electricity prices. The PIs (called ‘confidence intervals’) are obtained as the time-varying quantiles of the density forecasts. Like in Misiorek et al. (2006), the accuracy of the PIs is checked by comparing the nominal coverage with the actual one. Somewhat surprisingly, the density forecasts themselves are not analyzed.

Garcia-Martos et al. (2011) construct PIs based on one-day-ahead forecasts of the common volatility factors in the proposed GARCH-SeaDFA factor model, but do not either evaluate or test their efficiency. Also in a multivariate context, Wu, Chan, Tsui, and Hou (2013) propose a recursive dynamic factor analysis (RDFA) algorithm, where the principal components (PC) are tracked recursively using a subspace tracking algorithm, while the PC scores are tracked further and predicted recursively via the Kalman filter. From the latter, the covariance, and hence the interval, of the predicted electricity price is estimated. The accuracy of the PIs is checked by comparing the nominal coverage with the actual one (called ‘calibration bias’ here and by computing the ‘interval score’ (also known as the ‘Winkler score’, see Gneiting & Raftery, 2007; Maciejowska, Nowotarski, & Weron, 2014; Winkler, 1972), which favors narrow PIs and penalizes observations that do not lie within the PIs according to the nominal proportions.

Gonzalez et al. (2012) investigate the performances of two hybrid forecasting models for predicting the next-day spot electricity prices in the APX-UK power exchange: (i) a hybrid approach which combines a fundamental model, formulated using supply stack modeling, with an econometric model using data on price drivers, and (ii) an extended variant of this model which includes logistic smooth transition regression (LSTR) to represent regime-switching for periods of structural change. The out-of-sample point forecasts of the two hybrid approaches (and of the hybrid-LSTR in particular) compare favorably to those of non-hybrid SARMA, SARMAX and LSTR models. The quality of the PIs is evaluated by comparing the nominal coverage of the models to the true coverage (no formal tests are performed). The LSTR model gives the best results, followed closely by the hybrid-ARX and

SARMAX models. For the hybrid-LSTR model, the number of exceeding prices observed is significantly higher than the theoretical number, due to the overly narrow PIs.

Chen et al. (2012) combine an extreme learning machine (ELM; a learning algorithm for a single hidden layer MLP which can overcome the problems caused by gradient descent type methods) with a wild (or external) bootstrap approach, and use them to compute point and interval forecasts of half-hourly spot prices in the Australian electricity market. The uncertainty of data noise is not considered in the construction of the PIs, and the accuracy of the PIs is only checked by comparing the nominal coverage with the actual one. In a follow-up paper, Wan, Xu, Wang, Dong, and Wong (2014) first use ELM to obtain point forecasts of half-hourly Australian spot prices. Then, to compute PIs, they use a complex – though over 100 times faster than a traditional bootstrap-based ANN approach – procedure involving $N + 1$ additional neural networks. They (i) construct $N = 100$ bootstrapped samples from the residuals of the point forecasts, (ii) calibrate N new MLPs (using ELM), and (iii) use MLE to train yet another MLP for the residuals noise variance approximation. This time, the PI accuracy is evaluated based on both the nominal coverage (called ‘reliability’) and the PI width (called ‘sharpness’), by computing the ‘interval score’.

Khosravi, Nahavandi, and Creighton (2013) propose a hybrid method for the construction of PIs, which uses moving block bootstrapped neural networks and GARCH models for forecasting electricity prices. Rather than employing the traditional ML estimation, the parameters of the GARCH model are adjusted via the minimization of a PI-based cost function. The method is tested on hourly electricity prices from Australian and New York markets. The authors claim that the proposed method generates narrow PIs with a large coverage probability. However, the accuracy measure they use – the so-called Coverage Width-based Criterion (CWC) – possesses serious flaws and as Pinson and Tasty (2014) argue should be avoided in PI evaluation. Khosravi et al. (2013) also do not conduct formal statistical tests for coverage. In fact, except for Weron and Misiorek (2008), none of the papers discussed in this Section perform such tests. There is certainly a need for the techniques reviewed in Section 4.5.2 to be introduced to the EPF literature.

4.2.2. Density forecasts

Obviously, it is more useful for a modeler to know the entire forecast density than a single PI. However, this is, or at least seems to be, a more difficult task. For a comprehensive review of the computation of density forecasts, we refer to Tay and Wallis (2000). This topic has barely been touched upon in the EPF literature. As was mentioned above, Serinaldi (2011) forecasts the distribution of electricity prices using the GAMLSS approach, but computes and discusses only the PIs (obtained as quantiles of the density forecasts).

Huurman et al. (2012) consider GARCH-type time-varying volatility models, and find that models that are augmented with weather forecasts statistically outperform specifications which ignore this information in the

density forecasting of Scandinavian day-ahead electricity prices. Like Diebold, Gunther, and Tay (1998), they utilize the *probability integral transform* (PIT) scores of the realization of the variable with respect to the forecast densities, and use the Berkowitz (2001) likelihood ratio test for the zero mean, unit variance and independence of the PIT scores to infer the goodness-of-fit. Huurman et al. also measure the relative predictive accuracy by applying the Kullback–Leibler Information Criterion (KLIC; see Bao, Lee, & Saltoglu, 2007).

In a recent paper, Jonsson, Pinson, Madsen, and Nielsen (2014) develop a semi-parametric methodology for generating prediction densities of day-ahead electricity prices in Western Denmark (Nord Pool), comprising a time-adaptive quantile regression model for the 5%–95% quantiles and a description of the distribution tails by exponential distribution. They evaluate the quality of the forecasts by computing the average Continuous Ranked Probability Score (CRPS) and the related Continuous Ranked Probability Skill Score (CRPSS). Jonsson et al. do not perform formal statistical tests, but Gneiting, Balabdaoui, and Raftery (2007) argue that the null hypothesis of no difference in predictive performances can be tested easily, given the CRPS values.

4.2.3. Threshold forecasting

Before we conclude Section 4.2, let us mention a recent approach to EPF that is not yet well known in the literature, but may become popular in the near future, especially in industry. On the one hand, it may be treated as a generalization of spike occurrence forecasting (see Section 4.1.2), where the number of regimes is more than two, as in a three-state (or more) MRS model (see Section 3.7.2). On the other hand, it could be considered as interval forecasting where, instead of constructing a PI around a point forecast, a future price is allocated to one of a few prespecified price intervals spanning the entire range of attainable prices. The rationale for *threshold forecasting* comes from the fact that applications like demand-side management do not require exact values of future spot prices, but instead use specific price thresholds as the basis for making scheduling decisions. For instance, an industrial consumer may decide to shut down a production line if prices exceed a certain threshold.

To the best of our knowledge, the first paper to utilize this approach was that of Zareipour et al. (2011). The authors use two SVM-based models to classify future electricity prices in the Ontario and Alberta markets with respect to prespecified price thresholds. For both markets, the prices are classified into three groups: (i) from the price floor (defined by the applicable market rules in Ontario and Alberta: –2000 and 0 respectively) to the average price in the year 2008 (50 and 90 respectively), (ii) from the average price to twice the average price, and (iii) from twice the average price to the price cap (2000 and 1000 respectively). The authors find that the proposed models provide significantly (not in a statistical sense, as no formal testing is conducted) more accurate results than the three price forecasting models (ARIMA, ARX, ARMAX) used by Zareipour et al. (2006), a mixed similar-day and ANN predictor, or the pre-dispatch price

forecasts published by the ISO (available for Ontario only). Interestingly, they show that the demand is not as useful for price classification as for price forecasting, though it leads to a slightly better classification on average. Hence, in a follow-up paper, Huang, Zareipour, Rosehart, and Amjadi (2012) limit the initial feature (input) set to lagged prices, and concentrate on finding a better classifier than SVM.

Threshold forecasting seems to be particularly important for volatile markets, where using the predicted prices (e.g., those obtained from time series or computational intelligence models) is likely to lead to a worse classification. However, it should be noted that there is a cost associated with the higher classification accuracy attained within threshold forecasting, namely the loss of exact price values, which are obviously available in classical EPF. Mixing the two approaches may not be the best idea, as they can lead to contradictory forecasts.

Finally, note that threshold forecasting is somewhat related to the concept of the *critical load level*, see Bo and Li (2009). The authors look at LMPs from the system level perspective and focus on the phenomenon of the step-wise price variation as the load increases, i.e., they consider, not prespecified price intervals, but a set of discrete price levels. Under a certain assumed probability distribution of the actual load, they propose to consider probabilistic LMPs and formulate the probability mass function of this random variable. Although the approach is illustrated only for test networks (a modified PJM five-bus system and the IEEE 118-bus system), their concept is general, and may be used for analyzing the integration of renewables into today's electricity markets and demand response activities.

4.3. Combining forecasts

The idea of combining forecasts goes back to the late 1960s and the seminal papers of Bates and Granger (1969) and Crane and Crotty (1967). Since then, many authors have suggested the superiority of *forecast combinations* (also referred to as *combining forecasts*, *forecast averaging* or *model averaging*) over the use of individual models, see e.g. Clemen (1989); de Menezes, Bunn, and Taylor (2000); Timmermann (2006); and Wallis (2011); and references therein. Given the abundance of averaging schemes, Hibon and Evgeniou (2005) propose a criterion for selecting among forecasts, and show that the accuracy of the selected combinations is significantly better than those of the selected individual forecasts using this criterion, and that the selected combinations are less variable. They also make the important comment that “the advantage of combining is not that the best possible combinations perform better than the best possible individual forecasts” (i.e., ex-post), but that “it is less risky in practice to combine forecasts than to select an individual forecasting method” (i.e., ex-ante).

Despite this popularity, the combination of forecasts has not been discussed extensively in the context of electricity markets to date. There is some limited evidence on the adequacy of combining forecasts of electricity demand (dating back to the 1980s, see Bunn, 1985a; Bunn & Farmer, 1985; Smith, 1989; Taylor, 2010; Taylor & Majithia, 2000) or transmission congestion (Løland et al.,

2012); however, apart from the unpublished Ph.D. thesis of Nan (2009), it was only very recently that Bordignon et al. (2013); Maciejowska et al. (2014); Nowotarski et al. (2014); Nowotarski and Weron (2014a,b) and Raviv et al. (2013) provided empirical support for the benefits of combining forecasts in obtaining better predictions of electricity spot prices.

We should mention here that combining forecasts is related to the concept of *committee machines* (Haykin, 1998), which is also referred to as *ensemble averaging*. A committee machine is composed of multiple networks. The individual ANNs are trained, perform predictions and then are updated in such a way as if they were stand-alone (\rightarrow individual forecasts). Then, a 'weight calculator' generates weighting coefficients by which individual predictions are combined linearly in a 'combiner' neuron (\rightarrow combined forecast). To the best of our knowledge, only Guo and Luh (2004) use committee machines for EPF. They combine a RBF network, which uses 23 inputs and six clusters, and a MLP, which uses 55 inputs and eight hidden neurons, to compute daily average on-peak electricity prices for New England. They consider three committee machines: (i) one with simple arithmetic averaging, (ii) one where the correlation matrix used to determine weighting coefficients is re-calculated whenever new prediction errors become available, and (iii) one newly developed combiner. Interestingly, this promising approach involving committee machines has not been used in more recent publications. What is even more surprising is that the two approaches – forecast combinations and committee machines – seem to be evolving independently, with researchers from the two groups being unaware of the parallel developments.

4.3.1. Point forecasts

Numerous combining methods have been proposed in the literature. Among them, *simple averaging* (i.e., the arithmetic mean of individual forecasts) stands out as the most popular and surprisingly robust approach (Bunn, 1985b; Clemen, 1989; Genre, Kenny, Meyler, & Timmermann, 2013; Stock & Watson, 2004). Ordinary Least Squares regression or *OLS averaging* is another easy-to-implement approach. The idea was first described by Crane and Crotty (1967), but it was the influential paper of Granger and Ramanathan (1984) that inspired further research efforts in this direction. In OLS averaging, the combined forecast is determined using the following regression:

$$P_t = w_{0t} + \sum_{i=1}^M w_{it} \hat{P}_{it} + e_t, \quad (23)$$

where P_t is the actual electricity spot price at time t , $\hat{P}_{1t}, \dots, \hat{P}_{Mt}$ are the M individual price forecasts calculated for time t , and w_{it} is the weight assigned to forecast i at time t . This approach has the advantage of generating unbiased combined forecasts without the need to worry about the bias of the individual models. However, the OLS estimates of the weights are inefficient, due to the possible presence of serial correlation in the combined forecast errors. The vector of estimated weights \mathbf{w}_t is

likely to exhibit an unstable behavior, a problem that has sometimes been dubbed 'bouncing betas'. As a result, minor fluctuations in the sample can cause major shifts of the weight vector. And electricity spot prices definitely are volatile!

To address this issue, Aksu and Gunter (1992) consider variants of OLS averaging with either non-negative weights (Nonnegativity Restricted Least Squares, NRLS) or weights that are restricted to sum to unity (Equality Restricted Least Squares, ERLS). They find that NRLS and simple averaging almost always outperform ERLS, which – without a constant term – on average produces more accurate forecasts than OLS averaging. Raviv et al. (2013) combine the two restrictions to yield *CLS averaging*, i.e., constrained least squares, with positive weights that sum to unity. An alternative variant of OLS averaging that is more robust to outliers is proposed by Nowotarski et al. (2014). They apply the absolute loss function instead of the quadratic one in Eq. (23), to yield the Least Absolute Deviation regression or *LAD averaging*. This method may be viewed as a special case of quantile regression, with the quantile being equal to 0.5 (Koenker, 2005).

To illustrate the power of combining forecasts, let us consider the hourly electricity prices from the Nord Pool market over the period 8.8.2012–31.12.2013; see the upper panel in Fig. 14. The period 8.8.2012–7.5.2013 (273 days = 39 weeks) is used only for the calibration of the individual models, and hence, the first forecast is made for the 24 h of 8.5.2013. Six 'pure-price' models are considered: AR, TAR, SNAR, MRJD (all four models as in Weron & Misiorek, 2008), NAR (i.e., a recurrent network with the same inputs as the first three models, estimated using the Levenberg–Marquardt algorithm; see also Section 3.9.3), and a multivariate three-factor model (FM; see Eqs (25)–(26) and the description in Section 4.4). Like in Weron and Misiorek (2008), the calibration window is expanded by one day after the 24 hourly forecasts have been made for the next day. Three combining schemes are used, namely simple, CLS and LAD averaging, and calibrated on a rolling window of 28 days (which turned out to yield better combined forecasts than an expanding window). The first forecast of the averaging models is made for the 24 h of 5.6.2013. All models are evaluated in terms of the WMAE, see Eq. (2), in the 30-week period 5.6.2013–31.12.2013. The results are illustrated in the lower panel of Fig. 14 and summarized in Table 1. Clearly, combining is advantageous. The simple and CLS averaging lead to forecasts that are the best on average (in terms of WMAE) and most often overall (out of those obtained from the nine individual and combined models), and that deviate the least from the best possible forecast (in terms of WMAE). In particular, CLS averaging stands out as the optimal approach, yielding the lowest WMAE and *m.d.f.b.* statistics, see Table 1. It is also the only averaging approach that is able to provide a reasonable forecast (WMAE = 10.48) in the most spiky third week of the test period, only slightly worse than the best-performing model in this week, the recurrent network model (WMAE = 9.21). Apart from LAD (WMAE = 13.65), all other models yield extremely large errors, ranging from 17.25 (Simple) to 21.14 (TAR).

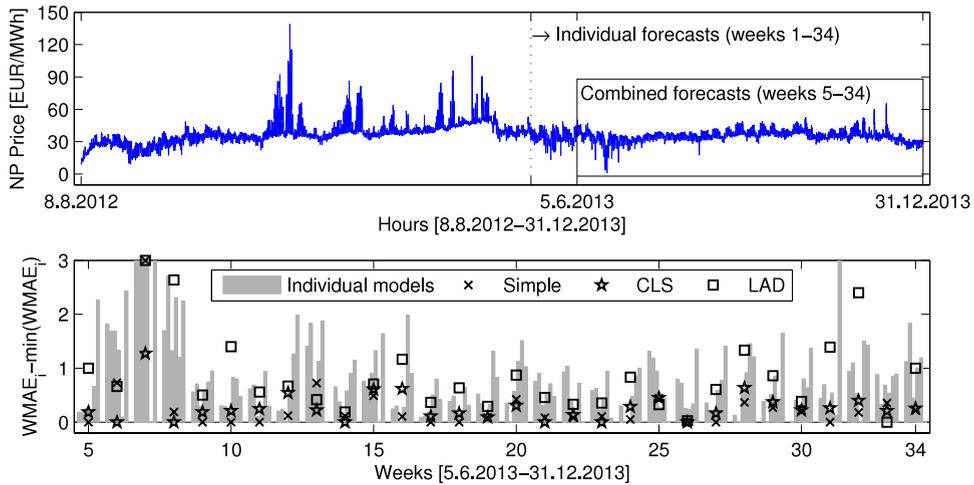


Fig. 14. Top panel: Nord Pool hourly system spot price in the period 8.8.2012–31.12.2013. The out-of-sample 30-week test period is indicated by a rectangle, while the vertical dotted line represents the beginning of the individual models’ forecasts and the calibration window for the forecast averaging methods (four weeks prior to the test period). Bottom panel: A plot illustrating the deviation of a particular model’s WMAE, see Eq. (2), with respect to the WMAE of the best model in week i , i.e., $WMAE_i - \min(WMAE_i)$. All values exceeding three are set to three.

Table 1
Summary statistics for the six individual and three averaging methods.

	Individual models						Forecast combinations		
	AR	TAR	SNAR	MRJD	NAR	FM	Simple	CLS	LAD
WMAE	5.03 (3.40)	5.07 (3.53)	4.77 (3.26)	4.98 (3.17)	4.88 (1.62)	5.36 (3.17)	4.47 (2.87)	4.29 (1.88)	4.92 (2.41)
# best	1	3	4	1	2	4	8	6	1
m.d.f.b.	1.01	1.05	0.75	0.96	0.86	1.34	0.45	0.27	0.89

Notes: WMAE is the mean value of WMAE for a given model (with standard deviation in parentheses), # best is the number of weeks in which a given averaging method performs best in terms of WMAE, and finally m.d.f.b. is the mean deviation from the best model in each week. The best values in each row are emphasized in bold. The out-of-sample test period covers 30 weeks (5.6.2013–31.12.2013).

As the literature on combining forecasts “is now voluminous and rather repetitive” (Wallis, 2011), we do not attempt to review all or even most methods. Instead, we only mention briefly three other approaches that have been applied in EPF. One is to choose the weights for each model based on the inverse of the Root Mean Squared Errors (IRMSE). Clearly, models that produce smaller errors will be assigned larger weights than models with higher errors, an approach dating back to Bates and Granger (1969), and later adopted by Diebold and Pauly (1987), Stock and Watson (2004) and Timmermann (2006), among others. Interestingly, for two different sets of individual models, Nowotarski et al. (2014) and Raviv et al. (2013) observe that, in the case of electricity prices, IRMSE averaging leads to nearly the same predictions as simple averaging. This is due to the fact that the RMSE errors of the individual models tend to be large compared to the differences between them. Hence, the IRMSE weights are different from each other but very close to the equal weights of simple averaging. A potential remedy would be to subtract a certain value, say half of the lowest RMSE value, from the errors, and then apply the algorithm.

The second approach is to use adaptive weights. In the simplest case, any of the models discussed so far can be reestimated at every time step (using either an expanding or a rolling window), meaning that the weights would become adaptive. A more sophisticated adaptive approach

is, for instance, Aggregated Forecast Through Exponential Re-weighting (AFTER; see Sanchez, 2008; Zou & Yang, 2004). Finally, the third approach is to use Bayesian Model Averaging (BMA) to avoid the a priori decision to use all models (Madigan & Raftery, 1994); see also Geweke and Amisano (2010); Geweke and Whiteman (2006); Hoogerheide, Kleijn, Ravazzolo, Van Dijk, and Verbeek (2010); and Koop and Potter (2004) for more recent variants and applications. The model weights for BMA are given by Bayes’ theorem, according to which we compute the posterior probabilities for each of the possible individual model combination options $m_l, l = 1, \dots, 2^M$, not the M individual models. Once the weights are set, the conditional expectation of the forecast is calculated for each of the options considered, and the resulting forecast combination is given by $\hat{P}_t^c = \sum_{l=1}^{2^M} w_l \mathbb{E}(P_t | m_l, \theta_l)$, where θ_l is the collection of parameters required for combination option l (R code is available from <http://cran.r-project.org/web/packages/BMA>).

In the first paper in EPF to consider forecast averaging explicitly, Nan (2009) evaluates three averaging schemes (simple and two variants of IRMSE-type averaging) on a dataset comprising 2005–2006 British day-ahead electricity prices for four half-hourly load periods. The author finds that combinations only work better during the Spring season for load period six, which is a very calm period, and

argues that the reason for such a disappointing performance is that the 19 individual models introduce too much variation in the combinations, as some models perform very poorly during particular seasons and/or for particular hours. Nan (2009) then applies the model confidence set (MCS) and forecast encompassing techniques (see Section 4.5.2) to select subsets of two to four models for combining, which differ for each season and each load period, and is able to outperform the individual predictors in most cases. Interestingly, Nowotarski et al. (2014) do not confirm the need to select subsets of individual models for combining, and argue that the problem faced by Nan (2009) is due, not to an overabundance of individual models, but to their similarity – they are all variants of four base specifications: ARMAX, linear regression, TVR and a MRS model.

This is confirmed to some extent by the approach taken in a follow-up article by Bordignon et al. (2013), who combine forecasts obtained from only five individual models (the fifth is a variant of the MRS model estimated on a rolling window of six months, not an expanding one). Five combining methods are considered, including simple, IRMSE-type and AFTER averaging. The authors examine whether forecast combinations outperform individual methods, from both an ex-post (i.e., using full sample information) and an ex-ante (i.e., using only information available at the time the forecast is made) perspective. In the more realistic ex-ante perspective, they find that combined forecasts perform better than individual forecasts, with the difference being significant in 33% of cases (they apply the DM test, see Section 4.5.2, and consider five half-hourly load periods). On the other hand, the individual forecasts are significantly less accurate than the combined forecasts in only 1% of cases.

Raviv et al. (2013) model the hourly prices by considering the intra-day relationships between the individual hours in the Nord Pool spot market. For the univariate analysis, they use heterogeneous autoregressive (HAR) and dynamic ARX models. For the multivariate analysis, they use VAR-type, Bayesian VAR, reduced rank regression (RRR), principal component regression and reduced rank Bayesian VAR models. The authors' focus is not on investigating the usefulness of averaging forecasts, but their empirical application finds that additional gains are achieved by using forecast combinations of individual models: even the best individual model is outperformed by forecast averaging (though not by a huge margin).

In an extensive empirical study, involving the 12 individual models used by Weron and Misiorek (2008), four datasets from three major European and North American markets, and seven averaging schemes (simple, OLS, NRLS, CLS, LAD, IRMSE, BMA), Nowotarski et al. (2014) find that the performances are not uniform across the markets considered. While their findings also show the additional benefits of combining forecasts for deriving more accurate predictions ex-ante, they are not as clear-cut as those of Bordignon et al. (2013). The authors find that four forecast averaging methods out of seven (namely simple averaging, NRLS, CLS and IRMSE) clearly outperform the benchmark ARX model and the best individual (BI) ex-ante scheme (a selection

scheme which picks one of the models that performed best in the past). However, one of the four, NRLS, is outperformed significantly (with respect to the DM test) by the benchmark ARX model and the BI selection scheme roughly as often as it outperforms them. Nowotarski et al. also remark that methods like OLS, NRLS and BMA, which allow for unconstrained weights, perform poorly and should be avoided in EPF. On the other hand, they recommend CLS averaging as a choice which may not be optimal, but will not worsen the prediction accuracy significantly compared to the BI ex-ante model; note that CLS averaging is also best in Table 1. Finally, Nowotarski et al. (2014) find that, while simple averaging and IRMSE are significantly more accurate than the benchmark ARX model in 50% of cases, and significantly less accurate in only 1% of cases, they suffer from a sensitivity to a consistent divergence in the performances of the individual forecasts, as is demonstrated by the poor performance for one of the four datasets.

4.3.2. Probabilistic forecasts

Although the literature on the combination of point forecasts is very rich, the topic of combining *probabilistic* (i.e., interval and density) forecasts is not so popular. Moreover, to the best of our knowledge, prior to three very recent papers, there had not been a single publication on the combination of interval or density forecasts in EPF. Nowotarski and Weron (2014a) examine possible accuracy gains from forecast averaging in the context of interval forecasts. They propose a new method for constructing PIs – dubbed Quantile Regression Averaging (QRA; Matlab code is available from <http://ideas.repec.org/s/wuu/hrcode.html>) – which utilizes the concept of quantile regression (QR; see e.g. Koenker, 2005) and a pool of point forecasts of individual (i.e., not combined) time series models. Using the conditional coverage test of Christoffersen (1998), they reach the conclusion that, while the empirical PIs (see Section 4.2.1) from combined forecasts do not provide significant gains for the PJM dataset considered, the QRA-based PIs are found to be more accurate than those of the best individual (SNAR) and benchmark (AR) models. In a follow-up paper, Nowotarski and Weron (2014b) consider a different calibration scheme and a more spiky (in the out-of-sample test period) Nord Pool dataset, and again confirm the superiority of the QRA-based PIs. Maciejowska et al. (2014) further extend the QRA approach and use principal component analysis to automate the selection process from among a large set of individual forecasting models available for averaging. In terms of unconditional coverage, conditional coverage and the Winkler score, the resulting Factor QRA (or FQRA) approach performs significantly better than the benchmark ARX model and moderately better than QRA (for data from the British power market over the period 1.7.2010–31.12.2012).

In the general forecasting context, there have been very few papers that have dealt explicitly with the combination of interval forecasts (note that the latter can be obtained as the quantiles of density forecasts). Luckily, there has been some progress in the area of density forecasts in the last decade, which will hopefully infiltrate the

EPF literature in the coming years. For instance, Wallis (2005) proposes a finite mixture distribution as an appropriate statistical model for a combined density forecast, then discusses its implications for combining interval forecasts. Hall and Mitchell (2007) propose a data-driven approach to the direct combination of density forecasts by taking a weighted linear combination of the competing density forecasts. The combination weights are chosen to minimize the 'distance', as measured by the Kullback–Leibler information criterion, between the predicted and true but unknown density. Mitchell and Wallis (2011) review current density forecast evaluation procedures and introduce a new test of density forecast efficiency. Kociecki, Kolasa, and Rubaszek (2012) introduce a formal method of combining expert and model density forecasts when the sample of past forecasts is unavailable. Finally, Billio, Casarin, Ravazzolo, and Van Dijk (2013) propose a Bayesian combination approach for multivariate predictive densities which relies upon a distributional state space representation of the combination weights.

4.4. Multivariate factor models

As was discussed in Sections 3.6–3.9, the literature on forecasting daily electricity prices has concentrated largely on models that use only information at the aggregated (i.e., daily) level. On the other hand, the very rich body of literature on forecasting intra-day prices has used disaggregated data (i.e., hourly or half-hourly), but generally has not explored the complex dependence structure of the multivariate price series. A notable exception is a working paper from 1997, published by Wolak (2000), in which *principal component analysis* (PCA) is applied to hourly or half-hourly prices from the UK, Scandinavia, Australia and New Zealand, in order to gain an understanding of the price formation mechanism and measure the relative predictability of the daily vector of prices in each country.

A decade passed before the multivariate context of spot electricity prices was picked up again by Huisman et al. (2007) and Panagiotelis and Smith (2008). In the first paper, hourly data from The Netherlands, Germany and France are expressed in the form of a panel, and the authors use seemingly unrelated regressions (SUR); they find that the prices in peak and off-peak hours are correlated highly among each other, but that there is much less correlation between the two groups. In the second, a first order vector autoregressive model with exogenous effects and skew t distributed innovations is used, and the authors uncover strong diurnal variation in many of the parameters.

The *vector autoregressive* (VAR) structure is a good starting point for multivariate factor models; for an excellent introduction to multivariate time series models, see Lütkepohl (2005). Let us first represent the hourly (half-hourly load periods can be considered analogously) spot price as a set of 24 univariate AR processes:

$$P_{kt} = \alpha_k \mathbf{D}_t + \sum_{i=1}^q \beta_{ik} P_{k,t-i} + \varepsilon_{kt}, \quad (24)$$

where $k = 1, \dots, 24$, α_k is a vector of parameters, and \mathbf{D}_t is a vector of exogenous, deterministic variables. This

can be interpreted as a restricted VAR(q) model, with diagonal parameter matrices \mathbf{B}_i and uncorrelated residuals ε_t , i.e., $\mathbf{P}_t = \mathbf{A}\mathbf{D}_t + \sum_{i=1}^q \mathbf{B}_i \mathbf{P}_{t-i} + \varepsilon_t$, where $\mathbf{P}_t = [P_{1t}, \dots, P_{24t}]'$, $\varepsilon_t = [\varepsilon_{1t}, \dots, \varepsilon_{24t}]'$, \mathbf{A} is a vector of deterministic parameters and \mathbf{B}_i are 24×24 matrices of autoregressive parameters. The restricted VAR model uses information about hourly prices, but does not explore the intra-day correlation structure. Since all hours during the day are correlated with each other, or at least within the peak and off-peak hours (Huisman et al., 2007), it seems reasonable to model them jointly. However, if we do so, the large number of parameters needing estimation ($1 + 24q$ in each equation) may result in over-fitting, yielding small in-sample residuals but large out-of-sample errors.

If we want to explore the structure of intra-day electricity prices, we need to use dimension reduction methods; for instance, factor models with factors estimated as *principal components* (PC). PC estimation is consistent for large dimensional models where both of the dimensions – time and the number of series – tend to infinity (Bai, 2003; Bai & Ng, 2002; Stock & Watson, 2002). When considering hourly data for one location, the panel consists of 24 variables. However, when multiple locations are considered, like the 20 PJM locations studied by Maciejowska and Weron (forthcoming), the panel should be sufficiently large to approximate the true factors.

The main assumption of the factor models is that all variables P_{kt} , $k = 1, \dots, 24$, co-move, and depend on a small set of common factors $\mathbf{F}_t = [F_{1t}, \dots, F_{Nt}]'$. The individual series P_{kt} can be modeled as a linear function of N principal components \mathbf{F}_t and stochastic residuals v_{kt} :

$$P_{kt} = \Lambda_k \mathbf{F}_t + v_{kt}, \quad (25)$$

where the *loads* (or *loadings*) $\Lambda_k = [\Lambda_{k1}, \dots, \Lambda_{kN}]$ describe the relationship between the factors \mathbf{F}_t and the panel variables P_{kt} . Note that these loads are not 'power system loads', but model parameters (as in Bai, 2003). The eigenvectors corresponding to the N largest eigenvalues of the matrix $\mathbf{P}\mathbf{P}$ multiplied by \sqrt{T} are consistent estimators of the common factors \mathbf{F}_t (see e.g. Stock & Watson, 2002). The number of common factors can be chosen on the basis of information criteria (like IC_2 and IC_3 , proposed by Bai & Ng, 2002) or the fraction of the total variability explained.

In order to be able to predict future values of P_{kt} , we need to forecast both the common factors F_{nt} and the idiosyncratic components v_{kt} . Although the factors are contemporaneously orthogonal, they may still be intertemporally correlated, due to normalization assumptions. Hence, it seems reasonable to model them jointly. Moreover, they may depend on some other variables, such as the deterministic variables (\mathbf{D}_t). At the same time, the idiosyncratic components can only be correlated weakly across periods, and can therefore be modeled separately for each hour. Moreover, they cannot have the same seasonal pattern, because all of the co-movement between hours is captured by the factors. It is natural (see e.g. Maciejowska & Weron, forthcoming) to assume that the common factors follow a VAR(p) model:

$$\mathbf{F}_t = \Phi \mathbf{D}_t + \sum_{i=1}^p \Theta_i \mathbf{F}_{t-i} + \zeta_t, \quad (26)$$

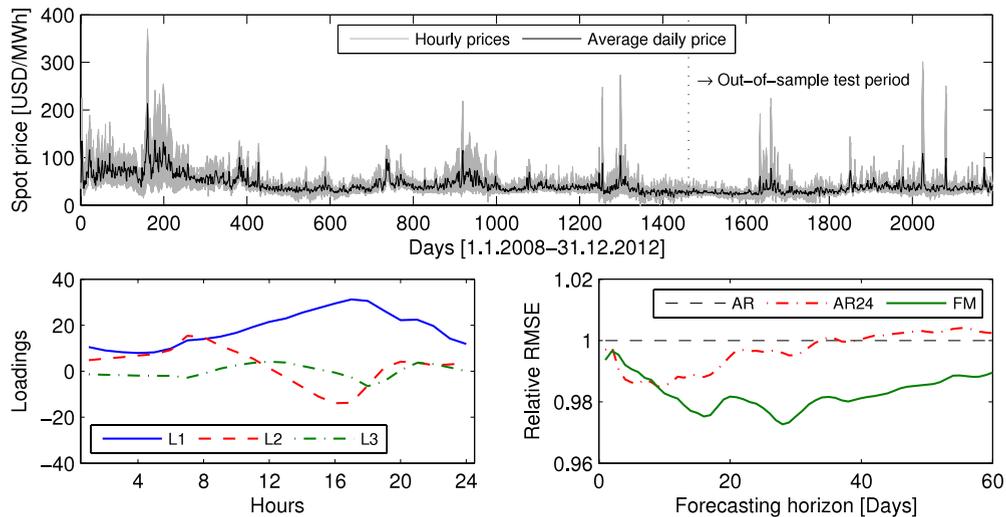


Fig. 15. Upper panel: PJM Dominion Hub hourly (gray) and average daily (black) system spot prices in the period 1.1.2008–31.12.2012. The vertical dotted line represents the beginning of the two-year out-of-sample test period. Lower left panel: The loadings obtained for a three-factor model, see Eqs. (25)–(26). Lower right panel: The relative RMSE of average daily price forecasts with respect to the forecasts of the benchmark univariate AR model. Clearly, the factor model (FM) outperforms the benchmark for all forecast horizons.

where Φ denotes a $N \times M$ matrix of deterministic coefficients, M is the number of deterministic variables, and Θ_i are $N \times N$ matrices of autoregressive parameters. To describe and forecast the idiosyncratic components v_{kt} , we can use AR models, independently for each k .

To illustrate the gains from developing factor models, let us consider the hourly electricity prices for the Dominion Hub in the PJM power market (US) over the period 1.1.2008–31.12.2013, see the upper panel of Fig. 15. The first three years are used for calibration only (we use a rolling calibration window), and the last two for out-of-sample testing. Three models are evaluated: a benchmark univariate AR(7) model, a restricted VAR(7) model (see Eq. (24)), and a three-factor model (see Eq. (25)), with the factors given by a VAR(7) model (see Eq. (26)). The factor loadings obtained are depicted in the lower left panel. The first loading may be interpreted as the level with an afternoon peak profile, the second as the morning peak, and the third as the mid-day peak. The relative RMSEs of average daily price forecasts with respect to the forecasts of the benchmark univariate AR model, i.e., $RMSE_i/RMSE_{AR}$, are plotted in the lower right panel for forecasting horizons of 1 to 60 days. Clearly, the factor model (FM) outperforms the benchmark for all forecast horizons. The difference is significant at the 5% level (according to the Diebold & Mariano, 1995, test; see Section 4.5.2) for horizons of four days or more. On the other hand, the restricted VAR model is better than the benchmark in the short-term, but worse in the mid-term.

In contrast to the relatively long history of using univariate models for EPF (see Sections 3.6–3.9), applications of multivariate models have appeared in the literature only within the last six years. Chen, Deng, and Huo (2008) use manifold learning (an extension of PCA) to remove intra-day and intra-week seasonality from hourly electricity prices, and predict them using three techniques. Their approach compares favorably to those of ARIMA, ARX and naïve methods in one day, one week and one month

ahead forecasting of hourly NYISO prices. Härdle and Trück (2010) use dynamic semiparametric factor models (DSFM) for EPF of hourly electricity prices in the German EEX market. They find that a model with three factors is able to explain up to 80% of the variation in hourly prices; however, the explanatory power decreases significantly for periods with higher numbers of price spikes.

Over the last two years, an increased inflow of ‘multivariate EPF papers’ can be observed. In particular, Peña (2012) analyzes hourly electricity prices in three day-ahead markets using a periodic panel model, and finds that, when all hourly prices are modeled jointly as a panel, autoregressive periodic components models describe the data better than standard non-periodic models. Garcia-Martos et al. (2012) propose to extract common factors from hourly prices and use them for one-day-ahead forecasting within a dynamic factor model (DFM) framework. They also report some preliminary results showing the usefulness of factor models for mid- and long-term predictions. Vilar, Cao, and Aneiros (2012) use a nonparametric regression technique with functional explanatory data and a semi-functional partial linear (SFPL) model to forecast hourly day-ahead prices in the Spanish market, and find it to be superior to the ARIMA and naïve approaches.

Elattar (2013) proposes to combine kernel PCA (for extracting features of the inputs) with a Bayesian local informative vector machine (for making the predictions), and finds the resulting technique to be superior to 12 other methods, including ARIMA and ANN, for short-term price forecasting in the Spanish market in 2002. Miranian, Abdollahzade, and Hassani (2013) apply the singular spectrum analysis (SSA; which is somewhat similar to PCA) to obtain extremely accurate one-step-ahead predictions of the hourly day-ahead prices in the Australian and Spanish power markets. Their results are controversial, however, as their method is roughly three times more accurate than the competitors (ARIMA, MLP and RBF networks), and is presumably able to predict irregularly

appearing price spikes almost perfectly for a test week in January 2006, even in the extremely spiky Australian market. Wu et al. (2013) propose an RDFA algorithm (see Section 4.2.1) and show that it outperforms functional PCA, AR with a time varying mean, and SVR models in predicting hourly day-ahead prices in the Australian and New England markets.

There are also a few articles which exploit the idea of using disaggregated data for the forecasting of aggregated variables, an approach with roots in macroeconometrics (Bermingham & D'Agostino, 2014; Hendry & Hubrich, 2011). For instance, Liebl (2013) proposes the modelling and prediction of electricity spot prices by first finding the functional relationship between prices and demand in terms of daily price-demand functions, then parametrizing the series of daily price-demand functions using a functional factor model. He demonstrates the power of this approach by comparing aggregated daily price forecasts for 1 to 20 days ahead from the model with those from two simple univariate time series models for daily prices (AR and MRS) and two alternative functional data models for hourly prices (DSFM and SFPL). Maciejowska and Weron (2013) use half-hourly data from the UK power market to forecast the average daily spot prices both directly (via ARX and vector ARX models) and indirectly (via factor models). The results indicate that there are forecast improvements from incorporating disaggregated (i.e., hourly) data, especially when the forecast horizon exceeds one week. Raviv et al. (2013) exploit the information embedded in the cross correlation of Nord Pool hourly price series in order to obtain more accurate one-step-ahead average daily price forecasts for Scandinavia. Finally, Maciejowska and Weron (forthcoming) evaluate the forecasting performances of four multivariate models (a restricted VAR and three factor models) calibrated to hourly and/or zonal day-ahead prices in the PJM market, and compare them with that of a univariate AR model, which uses only average daily data for the PJM Dominion Hub. The results indicate that there are forecast improvements from incorporating the additional information, essentially for all forecast horizons considered, from one day to two months, but only when the correlation structure of prices across locations and hours is modeled using factor models.

As the literature review in this section suggests, there is definitely potential in using the multivariate modeling approach. With the increase of computational power, the real-time calibration of these complex models will become feasible (Chan et al., 2012). We expect to see more EPF applications of the multivariate framework in the coming years.

4.5. The need for an EPF-competition

All major review publications (see Section 2.2) have concluded that there are problems with comparing the methods developed and used in the EPF literature. This is due mainly to the use of different datasets, different software implementations of the forecasting models and different error measures, but also to the lack of statistical rigor in many studies.

As a result, many of the published results seem to contradict each other. For instance, Misiorek et al. (2006) report a very poor forecasting performance of a MRS model, while Kosater and Mosler (2006) reach the opposite conclusion for a similar MRS model but a different market and mid-term forecasting horizons. On the other hand, Heydari and Siddiqui (2010) find that a regime-switching model does not capture price behaviors correctly in the mid-term. The cross-category comparisons are even less conclusive and more biased. Typically, advanced statistical techniques are compared with simple CI methods (see e.g. Conejo, Contreras et al., 2005), and vice versa (see e.g. Amjadi, 2006). However, our impression is that sophisticated, fine-tuned representatives of the two groups should be competitive if on equal terms. Moreover, at least at this stage, it seems unlikely that there exists any one model that would systematically outperform other models on a consistent basis.

4.5.1. A universal test ground

All of this calls for a comprehensive, thorough study involving (i) the same datasets, (ii) the same robust error evaluation procedures, and (iii) statistical testing of the significance of one model's outperformance of another. The time has come for an 'M-Competition' in EPF.⁴ The major advantage of such a comprehensive forecasting competition is that it assures objectivity, while guaranteeing expert knowledge.

We agree with Aggarwal et al. (2009b) that the forecasting test periods used in most EPF studies are too short to yield conclusive results. Test samples of carefully selected one-week periods, even if taken from different seasons of the year, generally ignore the problem of special days (holidays, near-holidays). Only longer test samples of several months to over a year should be considered. Moreover, while the number of test series considered in the most recent M-Competition (i.e., 3003) is by far too large for an EPF-Competition, there should be sufficient data available to enable such a competition to be conducted, given that electricity markets have existed for over a decade in many countries now.

Some power exchanges and data vendors openly provide high-frequency (hourly, half-hourly) time series of electricity prices on their web pages. For instance, Nord Pool publishes price and other fundamental power market data for the most recent two-year period;⁵ Elexon, the British market operator, publishes all kinds of balancing market data (including reserve margin forecasts);⁶ electricity prices for the UK can be downloaded from the APX power exchange website;⁷ and GDF Suez

⁴ The Makridakis or M-Competitions were empirical studies that compared the performances of large numbers of major time series methods using recognized experts who provided the forecasts for their own method of expertise; see Makridakis and Hibon (2000) for a discussion of the results.

⁵ See www.nordpoolspot.com/Market-data1/Downloads/Historical-Data-Download1/Data-Download-Page.

⁶ See www.bmreports.com.

⁷ See www.apxgroup.com/market-results/apx-power-uk/ukpx-rpd-historical-data.

provides hourly prices for five US markets, including the world's largest power market – the Pennsylvania-New Jersey-Maryland Interconnection (PJM).⁸ Perhaps some of these entities would be interested in participating in an EPF-competition and maintaining a database of electricity market time series which could form a universal testing ground for all EPF experts.

Finally, let us note that since submitting the first version of this article to IJF, we have learned that GEFCom2014 (see www.gefcom.org) will include a track on electricity price forecasting! The Global Energy Forecasting Competition is an initiative of the IEEE Working Group on Energy Forecasting. The first event, GEFCom2012, included only two tracks – load and wind power forecasting – but attracted more than 200 teams which submitted more than two thousand entries (Hong, Pinson, & Fan, 2014). The second competition, to be launched on 15 August 2014, will probably attract even more participants. Hopefully, the organizers will take into account some of the suggestions put forward in this article.

4.5.2. Guidelines for evaluating forecasts

Error measures for point forecasts were discussed in Section 3.3. A selection of the better-performing measures (weighted-MAE, seasonal MASE or RMSSE) should be used either exclusively or in conjunction with the more popular ones (MAPE, RMSE). One issue in relation to error measures that has apparently been downplayed in the EPF literature is that of statistical testing for the significance of the differences in forecasting accuracies of the models. In econometrics, the most popular approach is the Diebold and Mariano (1995) test; see Diebold (2013) for a recent discussion of its uses and abuses. The DM test is simply an asymptotic z-test of the hypothesis that the mean of the loss differential series, i.e., $d_t = L(\varepsilon_{1,t}) - L(\varepsilon_{2,t})$, is zero. In applications, $L(\varepsilon_{i,t})$ is typically taken to be the absolute $|\varepsilon_{i,t}|$ or square $\varepsilon_{i,t}^2$ loss, and $\varepsilon_{i,t} = X_t - \hat{X}_{i,t}$ is the forecast error for model $i = 1, 2$. The test statistic is then calculated as: $DM = \bar{d}/\hat{\sigma}_d$, where \bar{d} is the sample mean of the loss differential and $\hat{\sigma}_d$ is a consistent estimate of the standard deviation. Since forecast errors, and hence loss differentials, may be serially correlated, $\hat{\sigma}_d$ has to be calculated robustly. The DM test statistic is $N(0,1)$ -distributed, and one- or two-sided tests can be constructed easily. Nowadays, many statistical computing environments, like Matlab or R, include the DM test in the standard releases or as an add-in.

Alternative forecast comparison test procedures include the *model confidence set* approach of Hansen, Lunde, and Nason (2011), which is similar to the DM test for two models but estimates the distribution of the test statistic using a bootstrap procedure, and a test of forecast encompassing, whose null hypothesis is that the predictions from model 1 do not contain additional information relative to those of model 2 (if this is the case, we say that model 2 encompasses model 1; see Harvey, Leybourne, & Newbold, 1998). In one of the few applications in EPF, Bordignon et al. (2013) perform all three tests for evaluating

combined point forecasts, see Section 4.3. Moreover, Cruz et al. (2011); Cuaresma et al. (2004); Diongue et al. (2009); Gianfreda and Grossi (2012); Hong and Wu (2012); Maciejowska and Weron (forthcoming) and Nowotarski et al. (2014) perform the DM test.

Evaluating interval and density forecasts is more tricky. While there are numerous methods for calculating interval forecasts, only a few studies have proposed appropriate validation methods. One of the main exceptions is the seminal paper of Christoffersen (1998), which develops a model-independent approach based on the concept of PI violations. Three tests are carried out in the likelihood ratio (LR) framework, for the unconditional coverage, independence and conditional coverage. The LR statistics corresponding to the former two tests are distributed asymptotically as $\chi^2(1)$, and those corresponding to the latter as $\chi^2(2)$. Moreover, if we condition on the first observation, then the conditional coverage LR test statistic is the sum of the other two (Matlab code is available from <http://ideas.repec.org/s/wuu/hocode.html>). The unconditional coverage test compares the nominal coverages of the models to the true coverage, and is also known in the risk management (Value-at-Risk backtesting) literature as the Kupiec (1995) test. The independence test checks that the PI violations do not cluster. Finally, the conditional coverage test is a combination of the two. Christoffersen's tests have been applied by Maciejowska et al. (2014), Nowotarski and Weron (2014a,b), Sharma and Srinivasan (2013) and Weron and Misiorek (2008) for evaluating electricity spot price PIs, and by Chan and Gray (2006) and Cifter (2013) in the context of computing the Value-at-Risk for daily electricity spot prices, i.e., PIs for spot price returns. It should be noted that the independence test and hence the conditional coverage test are typically conducted only with respect to the first order dependency of exceedances. However, as Clements and Taylor (2003) show, the test can be easily modified to measure higher order dependency; see e.g. Maciejowska et al. (2014) for a sample application of this approach in the context of EPF. Berkowitz, Christoffersen, and Pelletier (2011) go one step further and using the Ljung-Box statistic jointly test for independence in the first m lags.

Wallis (2003) recasts Christoffersen's tests in the framework of χ^2 statistics, and considers their extension to density forecasts. The use of the contingency tables framework increases these methods' accessibility to users, and allows the incorporation of a more informative decomposition of the χ^2 goodness-of-fit statistic and the calculation of exact small-sample distributions. More recently, Dumitrescu, Hurlin, and Madkour (2013) propose a generalized method of moments (GMM) approach for testing PIs using discrete polynomials. The series of PI violations is split into blocks of size N . The sum of violations within each block follows a binomial distribution, and the proposed approach involves testing that the series of sums is indeed an i.i.d. sequence of random variables that are binomially distributed. Candelson, Colletaz, Hurlin, and Tokpavi (2011) use a similar approach in the context of Value-at-Risk backtesting. See also Berkowitz et al. (2011) for a review of autocorrelation-based, duration-based and

⁸ See <http://www.gdfsuezenergyresources.com/index.php?id=712>.

spectral density-based tests for clustering of Value-at-Risk exceedances.

Regarding density forecasts, a good starting point is the comprehensive review of Tay and Wallis (2000); see also Wallis (2003), who proposes χ^2 tests for both intervals and densities, and Berkowitz (2001), who suggests an approach to the evaluation of density forecasts that is now popular in the Value-at-Risk backtesting literature. Finally, Bao et al. (2007) compare various density forecasting models using the Kullback–Leibler Information Criterion (KLIC) of a candidate density forecast model with respect to the true density, and discuss how this KLIC is related to the KLIC based on the *probability integral transform* (PIT) in the framework of Diebold et al. (1998). They find that the two approaches are asymptotically equivalent, but that the PIT-based KLIC is better for evaluating the adequacy of each density forecasting model and the original KLIC is better for comparing competing models.

4.6. Final word

We hope that the methods, problems and suggestions discussed in this section and in the article as a whole will encourage researchers working in the area of electricity price forecasting to develop more efficient and better-grounded models and techniques. We also hope that this review will provide an impetus for those working in other areas of forecasting to move into the exciting, unique, and largely unexplored world of wholesale electricity markets.

Acknowledgments

This paper has benefited from conversations with the participants at the Conferences on Energy Finance (EF2012, EF2013), the ‘European Energy Market’ (EEM12, EEM14) Conferences, the Energy Finance Christmas Workshops (EFC12, EFC13), and the seminars at Macquarie University, National University of Singapore (NUS), Norwegian University of Science and Technology (NTNU), University of Sydney, University of Verona and Wrocław University of Technology. Critical comments and suggestions from Tao Hong, Rob Hyndman, Pierre Pinson and two anonymous referees are gratefully acknowledged. Special thanks for feedback on earlier versions of the manuscript and computational assistance go to Katarzyna Maciejowska, Paweł Maryniak and Jakub Nowotarski. This work was supported by funds from the National Science Centre (NCN, Poland) through grant no. 2011/01/B/HS4/01077.

References

Albanese, C., Lo, H., & Tompaidis, S. (2012). A numerical algorithm for pricing electricity derivatives for jump-diffusion processes based on continuous time lattices. *European Journal of Operational Research*, 222(2), 361–368.

Aggarwal, S. K., Saini, L. M., & Kumar, A. (2008). Electricity price forecasting in Ontario electricity market using wavelet transform in artificial neural network based model. *International Journal of Control, Automation and Systems*, 6(5), 639–650.

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

Aggarwal, S. K., Saini, L. M., & Kumar, A. (2009b). Short term price forecasting in deregulated electricity markets. A review of statistical models and key issues. *International Journal of Energy Sector Management*, 3(4), 333–358.

Aïd, R., Campi, L., & Langrené, N. (2013). A structural risk-neutral model for pricing and hedging power derivatives. *Mathematical Finance*, 23(3), 387–438.

Aksu, C., & Gunter, S. I. (1992). An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts. *International Journal of Forecasting*, 8(1), 27–43.

Amjady, N. (2006). Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Transactions on Power Systems*, 21, 887–996.

Amjady, N. (2007). Short-term bus load forecasting of power systems by a new hybrid method. *IEEE Transactions on Power Systems*, 22, 333–341.

Amjady, N. (2012). Short-term electricity price forecasting. In J. P. S. Catalão (Ed.), *Electric power systems: advanced forecasting techniques and optimal generation scheduling*. CRC Press.

Amjady, N., & Hemmati, M. (2006). Energy price forecasting. *IEEE Power and Energy Magazine*, March/April, 20–29.

Amjady, N., & Hemmati, M. (2009). Day-ahead price forecasting of electricity markets by a hybrid intelligent system. *European Transactions on Electrical Power*, 19(1), 89–102.

Amjady, N., & Keynia, F. (2009). Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique. *Energy Conversion and Management*, 50(12), 2976–2982.

Anbazhagan, S., & Kumarappan, N. (2013). Day-ahead deregulated electricity market price forecasting using recurrent neural network. *IEEE Systems Journal*, 7, 866–872.

Andalib, A., & Atry, F. (2009). Multi-step ahead forecasts for electricity prices using NARX: a new approach, a critical analysis of one-step ahead forecasts. *Energy Conversion and Management*, 50, 739–747.

Anderson, C. L., & Davison, M. (2008). A hybrid system-econometric model for electricity spot prices: Considering spike sensitivity to forced outage distributions. *IEEE Transactions on Power Systems*, 23(3), 927–937.

Areekul, P., Senju, T., Toyama, H., Chakraborty, S., Yona, A., Urasaki, N., et al. (2010). A new method for next-day price forecasting for PJM electricity market. *International Journal of Emerging Electric Power Systems*, 11(2), art. no. 3.

Arvesen, T., Medbø, V., Fleten, S.-E., Tomasgard, A., & Westgaard, S. (2013). Linepack storage valuation under price uncertainty. *Energy*, 52, 155–164.

Asai, M., McAleer, M., & Yu, J. (2006). Multivariate stochastic volatility: a review. *Econometric Reviews*, 25, 145–175.

Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, 16, 521–530.

Azadeh, A., Moghaddam, M., Mahdi, M., & Seyedmahmoudi, S. H. (2013). Optimum long-term electricity price forecasting in noisy and complex environments. *Energy Sources, Part B: Economics, Planning and Policy*, 8(3), 235–244.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1), 135–171.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.

Baldick, R., Grant, R., & Kahn, E. (2004). Theory and application of linear supply function equilibrium in electricity markets. *Journal of Regulatory Economics*, 25(2), 143–167.

Ball, C. A., & Torous, W. N. (1983). A simplified jump process for common stock returns. *Journal of Finance and Quantitative Analysis*, 18(1), 53–65.

Bao, Y., Lee, T.-H., & Saltoglu, B. (2007). Comparing density forecast models. *Journal of Forecasting*, 26, 203–225.

Barlow, M. (2002). A diffusion model for electricity prices. *Mathematical Finance*, 12, 287–298.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Operations Research Quarterly*, 20(4), 451–468.

Battle, C. (2002). *A model for electricity generation risk analysis*. Ph.D. thesis, Madrid: Universidad Pontificia de Comillas.

Battle, C., & Barquin, J. (2005). A strategic production costing model for electricity market price analysis. *IEEE Transactions on Power Systems*, 20(1), 67–74.

Becker, R., Hurn, S., & Pavlov, V. (2007). Modelling spikes in electricity prices. *The Economic Record*, 83(263), 371–382.

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

Benth, F. E., Kallsen, J., & Meyer-Brandis, T. (2007). A non-Gaussian Ornstein–Uhlenbeck process for electricity spot price modeling and derivatives pricing. *Applied Mathematical Finance*, 14(2), 153–169.

- Benth, F. E., Kiesel, R., & Nazarova, A. (2012). A critical empirical study of three electricity spot price models. *Energy Economics*, 34(5), 1589–1616.
- Berkowitz, J. (2001). Testing density forecasts with applications to risk management. *Journal of Business and Economic Statistics*, 19, 465–474.
- Berkowitz, J., Christoffersen, P., & Pelletier, D. (2011). Evaluating Value-at-Risk models with desk-level data. *Management Science*, 57(12), 2213–2227.
- Bertingham, C., & D'Agostino, A. (2014). Understanding and forecasting aggregate and disaggregate price dynamics. *Empirical Economics*, 46, 765–788.
- Bessec, M., & Bouabdallah, O. (2005). What causes the forecasting failure of Markov-switching models? A Monte Carlo study. *Studies in Nonlinear Dynamics and Econometrics*, 9(2), Article 6.
- Bhar, R., Colwell, D. B., & Xiao, Y. (2013). A jump diffusion model for spot electricity prices and market price of risk. *Physica A*, 392(15), 3213–3222.
- Bierbrauer, M., Menn, C., Rachev, S. T., & Trück, S. (2007). Spot and derivative pricing in the EEX power market. *Journal of Banking and Finance*, 31, 3462–3485.
- Bierbrauer, M., Trück, S., & Weron, R. (2004). Modeling electricity prices with regime switching models. *Lecture Notes in Computer Science*, 3039, 859–867.
- Billio, M., Casarin, R., Ravazzolo, F., & Van Dijk, H. K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, 177(2), 213–232.
- Bo, R., & Li, F. (2009). Probabilistic LMP forecasting considering load uncertainty. *IEEE Transactions on Power Systems*, 24(3), 1279–1289.
- Bolle, F. (2001). Competition with supply and demand functions. *Energy Economics*, 23, 253–277.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31, 307–327.
- Boogert, A., & Dupont, D. (2008). When supply meets demand: the case of hourly spot electricity prices. *IEEE Transactions on Power Systems*, 23(2), 389–398.
- Borak, S., & Weron, R. (2008). A semiparametric factor model for electricity forward curve dynamics. *Journal of Energy Markets*, 1(3), 3–16.
- Bordignon, S., Bunn, D. W., Lisi, F., & Nan, F. (2013). Combining day-ahead forecasts for British electricity prices. *Energy Economics*, 35, 88–103.
- Borenstein, S., Bushnell, J., & Knittel, C. R. (1999). Market power in electricity markets: beyond concentration measures. *The Energy Journal*, 20, 65–88.
- Borgosz-Koczvara, M., Weron, A., & Wylomańska, A. (2009). Stochastic models for bidding strategies on oligopoly electricity market. *Mathematical Methods of Operations Research*, 69(3), 579–592.
- Bower, J., & Bunn, W. (2000). Model based comparison of pool and bilateral markets for electricity. *The Energy Journal*, 21(3), 1–29.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. San Francisco: Holden-Day.
- Brockwell, P. J., & Davis, R. A. (1996). *Introduction to time series and forecasting* (2nd ed.). New York: Springer-Verlag.
- Bunn, D. W. (1985a). Forecasting electric loads with multiple predictors. *Energy*, 10(6), 727–732.
- Bunn, D. W. (1985b). Statistical efficiency in the linear combination of forecasts. *International Journal of Forecasting*, 1, 151–163.
- Bunn, D. W. (2000). Forecasting loads and prices in competitive power markets. *Proceedings of the IEEE*, 88(2), 163–169.
- Bunn, D. W. (Ed.) (2004). *Modelling prices in competitive electricity markets*. Chichester: Wiley.
- Bunn, D. W., & Farmer, E. D. (Eds.) (1985). *Comparative models for electrical load forecasting*. Wiley.
- Burger, M., Graeber, B., & Schindlmayr, G. (2007). *Managing energy risk: an integrated view on power and other energy markets*. Wiley.
- Burger, M., Klar, B., Müller, A., & Schindlmayr, G. (2004). A spot market model for pricing derivatives in electricity markets. *Quantitative Finance*, 4(1), 109–122.
- Cabero, J., Baíllo, Á., Cerisola, S., Ventosa, M., García-Alcalde, A., Perán, F., & Relación, G. (2005). A medium-term integrated risk management model for a hydrothermal generation company. *IEEE Transactions on Power Systems*, 20(3), 1379–1388.
- Caihong, L., & Wenheng, S. (2012). The study on electricity price forecasting method based on time series ARMAX model and chaotic particle swarm optimization. *International Journal of Advancements in Computing Technology*, 4(15), 198–205.
- Candelon, B., Colletaz, G., Hurlin, C., & Tokpavi, S. (2011). Backtesting value-at-risk: a GMM duration-based test. *Journal of Financial Econometrics*, 9, 314–343.
- Cao, R. (1999). An overview of bootstrap methods for estimating and predicting time series. *Test*, 8(1), 95–116.
- Cao, R., Hart, J. D., & Saavedra, A. (2003). Nonparametric maximum likelihood estimators for AR and MA time series. *Journal of Statistical Computation and Simulation*, 73(5), 347–360.
- Cappe, O., Moulines, E., & Ryden, T. (2005). *Inference in hidden Markov models*. Springer.
- Carmona, R., & Coulon, M. (2014). A survey of commodity markets and structural models for electricity prices. In F. E. Benth, V. Kholodnyi, & P. Laurence (Eds.), *Quantitative energy finance: modeling, pricing, and hedging in energy and commodity markets*. Springer.
- Carmona, R., Coulon, M., & Schwarz, D. (2013). Electricity price modeling and asset valuation: a multi-fuel structural approach. *Mathematics and Financial Economics*, 7(2), 167–202.
- Cartea, A., & Figueroa, M. (2005). Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4), 313–335.
- Cartea, A., Figueroa, M., & Geman, H. (2009). Modelling electricity prices with forward looking capacity constraints. *Applied Mathematical Finance*, 16(2), 103–122.
- Catalão, J. P. S., Mariano, S. J. P. S., Mendes, V. M. F., & Ferreira, L. A. F. M. (2007). Short-term electricity prices forecasting in a competitive market: a neural network approach. *Electric Power Systems Research*, 77, 1297–1304.
- Catalão, J. P. S., Pousinho, H. M. I., & Mendes, V. M. F. (2011). Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. *IEEE Transactions on Power Systems*, 26(1), 137–144.
- Cerjan, M., Krzelj, I., Vidak, M., & Delimar, M. (2013). A literature review with statistical analysis of electricity price forecasting methods. In *Proceedings of EuroCon 2013* (pp. 756–763).
- Chaăbane, N. (2014a). A hybrid ARFIMA and neural network model for electricity price prediction. *International Journal of Electrical Power and Energy Systems*, 55, 187–194.
- Chaăbane, N. (2014b). A novel auto-regressive fractionally integrated moving average-least-squares support vector machine model for electricity spot prices prediction. *Journal of Applied Statistics*, 41(3), 635–651.
- Chan, K. F., & Gray, P. (2006). Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting*, 22, 283–300.
- Chan, K. F., Gray, P., & van Campen, B. (2008). A new approach to characterizing and forecasting electricity price volatility. *International Journal of Forecasting*, 24(4), 728–743.
- Chan, S. C., Tsui, K. M., Wu, H. C., Hou, Y., Wu, Y.-C., & Wu, F. F. (2012). Load/price forecasting and managing demand response for smart grids. *IEEE Signal Processing Magazine*, September, 68–85.
- Chatzidimitriou, K. C., Chrysopoulos, A. C., Symeonidis, A. L., & Mitkas, P. A. (2012). Enhancing agent intelligence through evolving reservoir networks for predictions in power stock markets. In *Lecture Notes in Computer Science: Vol. 7103* (pp. 228–247). LNAI.
- Che, J., & Wang, J. (2010). Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. *Energy Conversion and Management*, 51(10), 1911–1917.
- Chen, D., & Bunn, D. W. (2010). Analysis of the nonlinear response of electricity prices to fundamental and strategic factors. *IEEE Transactions on Power Systems*, 25, 595–606.
- Chen, J., Deng, S.-J., & Huo, X. (2008). Electricity price curve modeling and forecasting by manifold learning. *IEEE Transactions on Power Systems*, 23(3), 877–888.
- Chen, X., Dong, Z. Y., Meng, K., Xu, Y., Wong, K. P., & Ngan, H. W. (2012). Electricity price forecasting with extreme learning machine and bootstrapping. *IEEE Transactions on Power Systems*, 27(4), 2055–2062.
- Christensen, T., Hurn, S., & Lindsay, K. (2009). It never rains but it pours: modeling the persistence of spikes in electricity prices. *The Energy Journal*, 30(1), 25–48.
- Christensen, T., Hurn, S., & Lindsay, K. (2012). Forecasting spikes in electricity prices. *International Journal of Forecasting*, 28, 400–411.
- Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862.
- Čížek, P., Härdle, W., & Weron, R. (Eds.) (2011). *Statistical tools for finance and insurance* (2nd ed.). Berlin: Springer.
- Clemen, R. T. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting*, 5, 559–583.
- Clements, M. P., & Taylor, N. (2003). Evaluating interval forecasts of high-frequency financial data. *Journal of Applied Econometrics*, 18, 445–456.
- Cifter, A. (2013). Forecasting electricity price volatility with the Markov-switching GARCH model: evidence from the Nordic electric power market. *Electric Power Systems Research*, 102, 61–67.
- Conejo, A. J., Contreras, J., Espinola, R., & Plazas, M. A. (2005). Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting*, 21(3), 435–462.

- Conejo, A. J., Plazas, M. A., Espinola, R., & Molina, A. B. (2005). Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Transactions on Power Systems*, 20(2), 1035–1042.
- Cont, R., & Tankov, P. (2003). *Financial modelling with jump processes*. Chapman & Hall / CRC Press.
- Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems*, 18(3), 1014–1020.
- Coulon, M., & Howison, S. (2009). Stochastic behaviour of the electricity bid stack: from fundamental drivers to power prices. *Journal of Energy Markets*, 2(1), 29–69.
- Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53, 385–407.
- Crane, D. B., & Crotty, J. R. (1967). A two-stage forecasting model: exponential smoothing and multiple regression. *Management Science*, 6(13), B501–B507.
- Cruz, A., Muñoz, A., Zamora, J. L., & Espinola, R. (2011). The effect of wind generation and weekend on Spanish electricity spot price forecasting. *Electric Power Systems Research*, 81(10), 1924–1935.
- Cuaresma, J. C., Hlouskova, J., Kossmeier, S., & Obersteiner, M. (2004). Forecasting electricity spot prices using linear univariate time-series models. *Applied Energy*, 77, 87–106.
- Cutler, N. J., Boerema, N. D., MacGill, I. F., & Outhred, H. R. (2011). High penetration wind generation impacts on spot prices in the Australian national electricity market. *Energy Policy*, 39(10), 5939–5949.
- Czapaj, R., Tomasiak, G., & Lubicki, T. (2009). On the possibility of short-term electricity prices forecasting on Polish parquets considering the German EEX AG exchange. *Przegląd Elektrotechniczny*, 85(3), 140–143.
- Dacco, R., & Satchell, C. (1999). Why do regime-switching models forecast so badly? *Journal of Forecasting*, 18(1), 1–16.
- Daneshi, H., & Daneshi, A. (2008). Price forecasting in deregulated electricity markets – a bibliographical survey. In *Proceedings of DRPT 2008* (pp. 657–661).
- Davison, M., Anderson, C. L., Marcus, B., & Anderson, K. (2002). Development of a hybrid model for electrical power spot prices. *IEEE Transactions on Power Systems*, 17(2), 257–264.
- Day, C., & Bunn, D. (2001). Divestiture of generation assets in the electricity pool of England and Wales: a computational approach to analyzing market power. *Journal of Regulatory Economics*, 19(2), 123–141.
- Day, C. J., Hobbs, B. F., & Pang, J.-S. (2002). Oligopolistic competition in power networks: a conjectured supply function approach. *IEEE Transactions on Power Systems*, 17(3), 597–607.
- De Gooijer, J. G., & Hyndman, R. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22, 443–473.
- De Jong, C. (2006). The nature of power spikes: A regime-switch approach. *Studies in Nonlinear Dynamics and Econometrics*, 10(3), Article 3.
- Diebold, F. X. (2013). *Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of Diebold–Mariano tests*. Working Paper, Department of Economics, University of Pennsylvania.
- Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39, 863–883.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253–263.
- Diebold, F. X., & Pauly, P. (1987). Structural change and the combination of forecasts. *Journal of Forecasting*, 6, 21–40.
- Diongue, A. K., Guegan, D., & Vignal, B. (2009). Forecasting electricity spot market prices with a k-factor GIGARCH process. *Applied Energy*, 86(4), 505–510.
- Dong, Y., Wang, J., Jiang, H., & Wu, J. (2011). Short-term electricity price forecast based on the improved hybrid model. *Energy Conversion and Management*, 52, 2987–2995.
- Duch, W. (2007). What is computational intelligence and where is it going? In W. Duch, & J. Mandziuk (Eds.), *Springer studies in computational intelligence: Vol. 63. Challenges for computational intelligence* (pp. 1–13).
- Dumitrescu, E.-I., Hurlin, C., & Madkour, J. (2013). Testing interval forecasts: a GMM-based approach. *Journal of Forecasting*, 32, 97–110.
- Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford University Press.
- Eichler, M., & Türk, D. (2013). Fitting semiparametric Markov regime-switching models to electricity spot prices. *Energy Economics*, 36, 614–624.
- Elattar, E. E. (2013). Day-ahead price forecasting of electricity markets based on local informative vector machine. *IET Generation, Transmission and Distribution*, 7(10), 1063–1071.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987–1007.
- Escribano, A., Pena, J. I., & Villaplana, P. (2002). *Modelling electricity prices: International evidence*. Working Paper 02-27, Universidad Carlos III de Madrid.
- Eydeland, A., & Wolyniec, K. (2003). *Energy and power risk management*. Hoboken, NJ: Wiley.
- Fan, J. Y., & McDonald, J. D. (1994). A real-time implementation of short-term load forecasting for distribution power systems. *IEEE Transactions on Power Systems*, 9, 988–994.
- Fan, S., Mao, C., & Chen, L. (2007). Next-day electricity-price forecasting using a hybrid network. *IET Proceedings of Generation, Transmission and Distribution*, 1(1), 176–182.
- Fanone, E., Gamba, A., & Prokopczuk, M. (2013). The case of negative day-ahead electricity prices. *Energy Economics*, 35, 22–34.
- Fiorenzani, S. (2006). *Quantitative methods for electricity trading and risk management: advanced mathematical and statistical methods for energy finance*. Palgrave Macmillan.
- Fleten, S.-E., Heggedal, A. M., & Siddiqui, A. (2011). Transmission capacity between Norway and Germany: a real options analysis. *Journal of Energy Markets*, 4(1), 121–147.
- Fleten, S. E., & Lemming, J. (2003). Constructing forward price curves in electricity markets. *Energy Economics*, 25, 409–424.
- Frasconi, P., Gori, M., & Soda, G. (1992). Local feedback multilayered networks. *Neural Computation*, 4, 120–130.
- Gao, C., Bompard, E., Napoli, R., & Zhou, J. (2008). Design of the electricity market monitoring system. *Proceedings of DRPT 2008* (pp. 99–106), art. no. 4523386.
- García, R. C., Contreras, J., van Akkeren, M., & García, J. B. (2005). A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems*, 20(2), 867–874.
- García-Ascanio, C., & Mate, C. (2010). Electric power demand forecasting using interval time series: A comparison between VAR and iMLP. *Energy Policy*, 38(2), 715–725.
- García-Alcalde, A., Ventosa, M., Rivier, M., Ramos, A., & Relanó, G. (2002). Fitting electricity market models. A conjectural variations approach. *Proceedings of the 14th PSCC conference*, Seville.
- García-Martos, C., & Conejo, A. J. (2013). Price forecasting techniques in power systems. In *Wiley encyclopedia of electrical and electronics engineering* (pp. 1–23). <http://dx.doi.org/10.1002/047134608X.W8188>.
- García-Martos, C., Rodríguez, J., & Sánchez, M. J. (2007). Mixed models for short-run forecasting of electricity prices: application for the Spanish market. *IEEE Transactions on Power Systems*, 22, 544–551.
- García-Martos, C., Rodríguez, J., & Sánchez, M. J. (2011). Forecasting electricity prices and their volatilities using unobserved components. *Energy Economics*, 33(6), 1227–1239.
- García-Martos, C., Rodríguez, J., & Sánchez, M. J. (2012). Forecasting electricity prices by extracting dynamic common factors: application to the Iberian market. *IET Generation, Transmission and Distribution*, 6(1), 11–20.
- Gardner, E. S., Jr. (2006). Exponential smoothing: The state of the art – Part II. *International Journal of Forecasting*, 22, 637–666.
- Gareta, R., Romeo, L. M., & Gil, A. (2006). Forecasting of electricity prices with neural networks. *Energy Conversion and Management*, 47, 1770–1778.
- Geman, H., & Roncoroni, A. (2006). Understanding the fine structure of electricity prices. *Journal of Business*, 79, 1225–1261.
- Genre, V., Kenny, G., Meyler, A., & Timmermann, A. (2013). Combining expert forecasts: can anything beat the simple average? *International Journal of Forecasting*, 29(1), 108–121.
- Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26, 216–230.
- Geweke, J., & Whiteman, C. (2006). Bayesian forecasting. In G. Elliott, C. W. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (pp. 3–80). Elsevier.
- Gianfreda, A., & Grossi, L. (2012). Forecasting Italian electricity zonal prices with exogenous variables. *Energy Economics*, 34(6), 2228–2239.
- Gjølberg, O., & Brattested, T.-L. (2011). The biased short-term futures price at Nord Pool: can it really be a risk premium? *Journal of Energy Markets*, 4(1), 3–19.
- Gładysz, B., & Kuchta, D. (2011). A method of variable selection for fuzzy regression – the possibility approach. *Operations Research and Decisions*, 2/2011, 5–15.
- Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 69(2), 243–268.
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.

- Gonzalez, A. M., San Roque, A. M., & Garcia-Gonzalez, J. (2005). Modeling and forecasting electricity prices with input/output hidden Markov models. *IEEE Transactions on Power Systems*, 20(1), 13–24.
- Gonzalez, V., Contreras, J., & Bunn, D. W. (2012). Forecasting power prices using a hybrid fundamental-econometric model. *IEEE Transactions on Power Systems*, 27(1), 363–372.
- Granger, C. W., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3, 197–204.
- Guerci, E., Ivaldi, S., & Cincotti, S. (2008). Learning agents in an artificial power exchange: tacit collusion, market power and efficiency of two double-auction mechanisms. *Computational Economics*, 32, 73–98.
- Guerci, E., Rastegar, M. A., & Cincotti, S. (2010). Agent-based modeling and simulation of competitive wholesale electricity markets. In S. Rebennack, et al. (Eds.), *Handbook of power systems II – energy systems* (pp. 241–286). Springer.
- Guo, J.-J., & Luh, P. B. (2003). Selecting input factors for clusters of Gaussian radial basis function networks to improve market clearing price prediction. *IEEE Transactions on Power Systems*, 18(2), 665–672.
- Guo, J.-J., & Luh, P. B. (2004). Improving market clearing price prediction by using a committee machine of neural networks. *IEEE Transactions on Power Systems*, 19(4), 1867–1876.
- Haghi, H. V., & Tafreshi, S. M. M. (2007). Modeling and forecasting of energy prices using non-stationary Markov models versus stationary hybrid models including a survey of all methods. In *Proceedings of IEEE Canada EPC 2007* (pp. 429–434).
- Haldrup, N., & Nielsen, M. Ø. (2006). A regime switching long memory model for electricity prices. *Journal of Econometrics*, 135, 349–376.
- Hall, S. G., & Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23(1), 1–13.
- Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57, 357–384.
- Hamilton, J. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45, 39–70.
- Hamilton, J. (2008). Regime switching models. In *The new Palgrave dictionary of economics* (2nd edn.). London: Macmillan.
- Hansen, B. E. (2006). Interval forecasts and parameter uncertainty. *Journal of Econometrics*, 135, 377–398.
- Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, 79, 453–497.
- Harris, C. (2006). *Electricity markets: pricing, structures and economics*. Chichester: Wiley.
- Harvey, D., Leybourne, S., & Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business and Economic Statistics*, 16, 254–259.
- Haugom, E., & Ullrich, C. J. (2012). Forecasting spot price volatility using the short-term forward curve. *Energy Economics*, 34, 1826–1833.
- Haykin, S. (1998). *Neural networks: a comprehensive foundation* (2nd ed.). Prentice-Hall.
- Härdle, W., & Trück, S. (2010). *The dynamics of hourly electricity prices*. SFB 649 Discussion Paper 2010-013.
- Hendry, D. F., & Hubrich, K. (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business and Economic Statistics*, 29, 216–227.
- Heydari, S., & Siddiqui, A. (2010). Valuing a gas-fired power plant: A comparison of ordinary linear models, regime-switching approaches, and models with stochastic volatility. *Energy Economics*, 32, 709–725.
- Hibon, M., & Evgeniou, T. (2005). To combine or not to combine: Selecting among forecasts and their combinations. *International Journal of Forecasting*, 21, 15–24.
- Higgs, H., & Worthington, A. (2008). Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: the Australian wholesale spot electricity market. *Energy Economics*, 30, 3172–3185.
- Hobbs, B. F., Metzler, C. B., & Pang, J. S. (2000). Strategic gaming analysis for electric power systems: an MPEC approach. *IEEE Transactions on Power Systems*, 15, 638–645.
- Holmberg, P., Newbery, D., & Ralph, D. (2013). Supply function equilibria: Step functions and continuous representations. *Journal of Economic Theory*, 148(4), 1509–1551.
- Hong, T. (2014). Energy forecasting: past, present, and future. *Foresight, Winter*, 43–48.
- Hong, T., Pinson, P., & Fan, S. (2014). Global Energy Forecasting Competition 2012. *International Journal of Forecasting*, 30(2), 357–363.
- Hong, Y.-Y., & Hsiao, C.-Y. (2002). Locational marginal price forecasting in deregulated electricity markets using artificial intelligence. *IEE Proceedings: Generation, Transmission and Distribution*, 149(5), 621–626.
- Hong, Y.-Y., & Wu, C.-P. (2012). Day-ahead electricity price forecasting using a hybrid principal component analysis network. *Energies*, 5(11), 4711–4725.
- Hoogerheide, L., Kleijn, R., Ravazzolo, F., Van Dijk, H. K., & Verbeek, M. (2010). Forecast accuracy and economic gains from Bayesian model averaging using time-varying weights. *Journal of Forecasting*, 29, 251–269.
- Hu, L., Taylor, G., Wan, H.-B., & Irving, M. (2009). A review of short-term electricity price forecasting techniques in deregulated electricity markets. In *Proceedings of the universities PEC*, art. no. 5429485.
- Hu, Z., Yang, L., Wang, Z., Gan, D., Sun, W., & Wang, K. (2008). A game-theoretic model for electricity markets with tight capacity constraints. *International Journal of Electrical Power and Energy Systems*, 30, 207–215.
- Huang, C.-M., Huang, C.-J., & Wang, M.-L. (2005). A particle swarm optimization to identifying the ARMAX model for short-term load forecasting. *IEEE Transactions on Power Systems*, 20, 1126–1133.
- Huang, D., Zareipour, H., Rosehart, W. D., & Amjady, N. (2012). Data mining for electricity price classification and the application to demand-side management. *IEEE Transactions on Smart Grid*, 3(2), 808–817.
- Huisman, R. (2009). *An introduction to models for the energy markets*. Risk Books.
- Huisman, R., & de Jong, C. (2002). *Option formulas for mean-reverting power prices with spikes*. ERIM Report Series Reference No. ERS-2002-96-F&A.
- Huisman, R., & de Jong, C. (2003). Option pricing for power prices with spikes. *Energy Power Risk Management*, 7(11), 12–16.
- Huisman, R., Huurman, C., & Mahieu, R. (2007). Hourly electricity prices in day-ahead markets. *Energy Economics*, 29, 240–248.
- Huurman, C., Ravazzolo, F., & Zhou, C. (2012). The power of weather. *Computational Statistics and Data Analysis*, 56(11), 3793–3807.
- Hyndman, R. (2013). *The difference between prediction intervals and confidence intervals*. Hyndsight Blog (13 March 2013), <http://robjhyndman.com/hyndsight/intervals>.
- Hyndman, R., & Athanasopoulos, G. (2013). *Forecasting: principles and practice*. Online at <http://otexts.org/fpp/>.
- Hyndman, R., & Billah, B. (2003). Unmasking the theta method. *International Journal of Forecasting*, 19, 287–290.
- Hyndman, R., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679–688.
- Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer.
- Jabłońska, M., & Kauranne, T. (2011). Multi-agent stochastic simulation for the electricity spot market price. *Lecture Notes in Economics and Mathematical Systems*, 652, 3–14.
- Jacobsson, H. (2005). Rule extraction from recurrent neural networks: A taxonomy and review. *Neural Computation*, 17(6), 1223–1263.
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: a tutorial. *Computer*, 29(3), 31–44.
- Janczura, J. (2014). Pricing electricity derivatives within a Markov regime-switching model: a risk premium approach. *Mathematical Methods of Operations Research*, 79(1), 1–30.
- Janczura, J., Trueck, 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.
- Janczura, J., & Weron, R. (2009). Regime switching models for electricity spot prices: Introducing heteroskedastic base regime dynamics and shifted spike distributions. *IEEE conference proceedings – EEM09*, <http://dx.doi.org/10.1109/EEM.2009.5207175>.
- Janczura, J., & Weron, R. (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy Economics*, 32, 1059–1073.
- Janczura, J., & Weron, R. (2012). Efficient estimation of Markov regime-switching models: an application to electricity spot prices. *AStA – Advances in Statistical Analysis*, 96(3), 385–407.
- Janczura, J., & Weron, R. (2014). Inference for Markov-regime switching models of electricity spot prices. In F. E. Benth, P. Laurence, & V. Kholodnyi (Eds.), *Quantitative energy finance* (pp. 137–155). Springer.
- Johnsen, T. A. (2001). Demand, generation and price in the Norwegian market for electric power. *Energy Economics*, 23(3), 227–251.
- Jonsson, T., Pinson, P., Madsen, H., & Nielsen, H. A. (2014). Predictive densities for day-ahead electricity prices using time-adaptive quantile regression. *Energies*, 7(9), 5523–5547. <http://dx.doi.org/10.3390/en7095523>.
- Jonsson, T., Pinson, P., Nielsen, H. A., Madsen, H., & Nielsen, T. S. (2013). Forecasting electricity spot prices accounting for wind power predictions. *IEEE Transactions on Sustainable Energy*, 4(1), 210–218.
- Joskow, P. L. (2001). California's electricity crisis. *Oxford Review of Economic Policy*, 17(3), 365–388.
- Kaminski, V. (1997). The challenge of pricing and risk managing electricity derivatives. In *The US power market*. London: Risk Books.
- Kaminski, V. (2013). *Energy markets*. Risk Books.

- Kanamura, T., & Ōhashi, K. (2007). A structural model for electricity prices with spikes: measurement of spike risk and optimal policies for hydropower plant operation. *Energy Economics*, 29, 1010–1032.
- Kanamura, T., & Ōhashi, K. (2008). On transition probabilities of regime switching in electricity prices. *Energy Economics*, 30, 1158–1172.
- Karakatsani, N. V., & Bunn, D. W. (2008). Forecasting electricity prices: the impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, 24(4), 764–785.
- Karakatsani, N. V., & Bunn, D. W. (2010). Fundamental and behavioural drivers of electricity price volatility. *Studies in Nonlinear Dynamics and Econometrics*, 14(4), art. no. 4.
- Keles, D., Genoese, M., Möst, D., & Fichtner, W. (2012). Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. *Energy Economics*, 34(4), 1012–1032.
- Kepler, J. H., Bourbonnais, R., & Girod, J. (Eds.) (2007). *The econometrics of energy systems*. Palgrave Macmillan.
- Keynia, F., & Amjadi, N. (2008). Electricity price forecasting with a new feature selection algorithm. *Journal of Energy Markets*, 1(4), 47–63.
- Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on Neural Networks*, 22(9), 1341–1356.
- Khosravi, A., Nahavandi, S., & Creighton, D. (2013). A neural network-GARCH-based method for construction of prediction intervals. *Electric Power Systems Research*, 96, 185–193.
- Kim, C.-I., Yu, I.-K., & Song, Y. H. (2002). Prediction of system marginal price of electricity using wavelet transform analysis. *Energy Conversion and Management*, 43, 1839–1851.
- Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60, 1–22.
- Knittel, C. R., & Roberts, M. R. (2005). An empirical examination of restructured electricity prices. *Energy Economics*, 27, 791–817.
- Kociecki, A., Kolasa, M., & Rubaszek, M. (2012). A Bayesian method of combining judgmental and model-based density forecasts. *Economic Modelling*, 29(4), 1349–1355.
- Koerner, R. (2005). *Quantile regression*. Cambridge University Press.
- Konar, A. (2005). *Computational intelligence: principles, techniques and applications*. Springer.
- Koop, G., & Potter, S. (2004). Forecasting in dynamic factor models using Bayesian model averaging. *The Econometrics Journal*, 7, 550–565.
- Koopman, S. J., Ooms, M., & Carnero, M. A. (2007). Periodic seasonal reg-ARFIMA-GARCH models for daily electricity spot prices. *Journal of the American Statistical Association*, 102(477), 16–27.
- Koritarov, V. S. (2004). Real-world market representation with agents. *IEEE Power and Energy Magazine*, 2(4), 39–46.
- Kosater, P., & Mosler, K. (2006). Can Markov regime-switching models improve power-price forecasts? Evidence from German daily power prices. *Applied Energy*, 83, 943–958.
- Kowalska-Pyzalska, A., Maciejowska, K., Suszczyński, K., Sznajd-Weron, K., & Weron, R. (2014). Turning green: Agent-based modeling of the adoption of dynamic electricity tariffs. *Energy Policy*, 72, 164–174.
- Kristiansen, T. (2007). Pricing of monthly forward contracts in the Nord Pool market. *Energy Policy*, 35, 307–316.
- Kristiansen, T. (2012). Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy*, 49, 328–332.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3(2), 73–84.
- Lajdici, A. A., Tiguercha, A., & Boudour, M. (2014). Nash equilibrium in a two-settlement electricity market using competitive coevolutionary algorithms. *International Journal of Electrical Power and Energy Systems*, 57, 148–155.
- Lagarto, J., De Sousa, J., Martins, A., & Ferrão, P. (2012). Price forecasting in the day-ahead Iberian electricity market using a conjunctural variations ARIMA model. *IEEE Conference Proceedings – EEM12*, art. no. 6254734.
- Lanne, M., Lütkepohl, H., & Maciejowska, K. (2010). Structural vector autoregressions with Markov switching. *Journal of Economic Dynamics and Control*, 34(2), 121–131.
- Lei, M., & Feng, Z. (2012). A proposed grey model for short-term electricity price forecasting in competitive power markets. *International Journal of Electrical Power and Energy Systems*, 43(1), 531–538.
- Lewis, N. (2005). *Energy risk modeling: applied modeling methods for risk managers*. Palgrave Macmillan.
- Liebl, D. (2013). Modeling and forecasting electricity spot prices: a functional data perspective. *Annals of Applied Statistics*, 7(3), 1562–1592.
- Lisi, F., & Nan, F. (2014). Component estimation for electricity prices: procedures and comparisons. *Energy Economics*, 44, 143–159.
- Lin, T. N., Horne, B. G., Tino, P., & Giles, C. L. (1996). Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks*, 7(6), 1329–1337.
- Lin, W.-M., Gow, H.-J., & Tsai, M.-T. (2010). An enhanced radial basis function network for short-term electricity price forecasting. *Applied Energy*, 87(10), 3226–3234.
- Lira, F., Muñoz, C., Nuñez, F., & Cipriano, A. (2009). Short-term forecasting of electricity prices in the Colombian electricity market. *IET Generation, Transmission and Distribution*, 3(11), 980–986.
- Ljung, L. (1999). *System identification – theory for the user* (2nd ed.). Prentice Hall: Upper Saddle River.
- Longstaff, F. A., & Wang, A. W. (2004). Electricity forward prices: a high-frequency empirical analysis. *Journal of Finance*, 59(4), 1877–1900.
- Løland, A., Ferkingstad, E., & Wilhelmsen, M. (2012). Forecasting transmission congestion. *Journal of Energy Markets*, 5(3), 65–83.
- Lucheroni, C. (2012). A hybrid SETARX model for spikes in tight electricity markets. *Operations Research and Decisions*, 1/2012, 13–49.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: Springer-Verlag.
- Ma, Y., Luh, P. B., Kasiviswanathan, K., & Ni, E. (2004). A neural network-based method for forecasting zonal locational marginal prices. *Proceedings of IEEE PES 2004* (pp. 296–302).
- Maciejowska, K. (2014). Fundamental and speculative shocks, what drives electricity prices? *IEEE conference proceedings - EEM14* <http://dx.doi.org/10.1109/EEM.2014.6861289>.
- Maciejowska, K., & Weron, R. (2013). Forecasting of daily electricity spot prices by incorporating intra-day relationships: Evidence from the UK power market. *IEEE Conference Proceedings – EEM13*. <http://dx.doi.org/10.1109/EEM.2013.6607314>.
- Maciejowska, K., Nowotarski, J., & Weron, R. (2014). Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. *International Journal of Forecasting* (submitted for publication). Working paper version available from RePEc: <http://ideas.repec.org/p/wuu/wpaper/hsc1409.html>.
- Maciejowska, K., & Weron, R. (2014). Forecasting of daily electricity prices with factor models: utilizing intra-day and inter-zone relationships. *Computational Statistics*, <http://dx.doi.org/10.1007/s00180-014-0531-0>.
- Madani, K., Correia, A. D., Rosa, A., & Filipe, J. (Eds.) (2011). *Computational intelligence*. Springer.
- Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, 89, 1535–1546.
- Makridakis, S., & Hibon, M. (2000). The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16, 451–476.
- Mari, C. (2008). Random movements of power prices in competitive markets: a hybrid model approach. *Journal of Energy Markets*, 1(2), 87–103.
- Mandal, P., Haque, A. U., Meng, J., Martinez, R., & Srivastava, A. K. (2012). A hybrid intelligent algorithm for short-term energy price forecasting in the Ontario market. *Proceedings of IEEE PES 2012*, art. no. 6345461.
- Mandal, P., Senjyu, T., & Funabashi, T. (2006). Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market. *Energy Conversion and Management*, 47, 2128–2142.
- Maryniak, P. (2013). *Using indicated demand and generation data to predict price spikes in the UK power market*. M.Sc. thesis, Wrocław University of Technology.
- Maryniak, P., & Weron, R. (2014). *Forecasting the occurrence of electricity price spikes in the UK power market*. *Energy Economics* (submitted for publication). Working paper version available from RePEc: <http://ideas.repec.org/p/wuu/wpaper/hsc1411.html>.
- de Menezes, L. M., Bunn, D. W., & Taylor, J. W. (2000). Review of guidelines for the use of combined forecasts. *European Journal of Operations Research*, 120, 190–204.
- Meng, K., Dong, Z. Y., & Wong, K. P. (2009). Self-adaptive radial basis function neural network for short-term electricity price forecasting. *IET Generation, Transmission and Distribution*, 3(4), 325–335.
- Miranian, A., Abdollahzade, M., & Hassani, H. (2013). Day-ahead electricity price analysis and forecasting by singular spectrum analysis. *IET Generation, Transmission and Distribution*, 7(4), 337–346.
- Misorek, 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 and Econometrics*, 10(3), Article 2.
- Mitchell, J., & Wallis, K. F. (2011). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, 26(6), 1023–1040.
- Mitra, S., & Hayashi, Y. (2000). Neuro-fuzzy rule generation: survey in soft computing framework. *IEEE Transactions on Neural Networks*, 11, 748–768.

- Mori, H., & Awata, A. (2007). Data mining of electricity price forecasting with regression tree and normalized radial basis function network. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, art. no. 4414228.
- Mount, T. D., Ning, Y., & Cai, X. (2006). Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics*, 28, 62–80.
- Nan, F. (2009). *Forecasting next-day electricity prices: from different models to combination*. Ph.D. thesis, Università degli Studi di Padova, <http://paduaresearch.cab.unipd.it/2147/>.
- Negnevitsky, M., Mandal, P., & Srivastava, A.K. (2009). An overview of forecasting problems and techniques in power systems. In *Proceedings of IEEE PES 2009*, <http://dx.doi.org/10.1109/PES.2009.5275480>.
- Niimura, T. (2006). Forecasting techniques for deregulated electricity market prices – extended survey. In *Proceedings of IEEE PSCE2006* (pp. 51–56).
- Niu, H., Baldick, R., & Zhu, G. (2005). Supply function equilibrium bidding strategies with fixed forward contracts. *IEEE Transactions on Power Systems*, 20(4), 1859–1867.
- Niu, D., Liu, D., & Wu, D. D. (2010). A soft computing system for day-ahead electricity price forecasting. *Applied Soft Computing Journal*, 10(3), 868–875.
- Nogales, F. J., & Conejo, A. J. (2006). Electricity price forecasting through transfer function models. *Journal of the Operational Research Society*, 57, 350–356.
- Nogales, F. J., Contreras, J., Conejo, A. J., & Espinola, R. (2002). Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, 17, 342–348.
- Nowotarski, J., Raviv, E., Trück, S., & Weron, R. (2014). An empirical comparison of alternate schemes for combining electricity spot price forecasts. *Energy Economics*, <http://dx.doi.org/10.1016/j.eneco.2014.07.014>.
- Nowotarski, J., Tomczyk, J., & Weron, R. (2013). Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. *Energy Economics*, 39, 13–27.
- Nowotarski, J., & Weron, R. (2014a). Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Computational Statistics*, <http://dx.doi.org/10.1007/s00180-014-0523-0>.
- Nowotarski, J., & Weron, R. (2014b). Merging quantile regression with forecast averaging to obtain more accurate interval forecasts of Nord Pool spot prices. *IEEE Conference Proceedings – EEM14*. <http://dx.doi.org/10.1109/EEM.2014.6861285>.
- Olsson, M., & Soder, L. (2008). Modeling real-time balancing power market prices using combined SARIMA and Markov processes. *IEEE Transactions on Power Systems*, 23(2), 443–450.
- Panagiotelis, A., & Smith, M. (2008). Bayesian forecasting of intraday electricity prices using multivariate skew-elliptical distributions. *International Journal of Forecasting*, 24, 710–727.
- Pao, H.-T. (2006). A neural network approach to m-daily-ahead electricity price prediction. *Lecture Notes in Computer Science*, 3972, 1284–1289.
- Peña, J. I. (2012). A note on panel hourly electricity prices. *Journal of Energy Markets*, 5(4), 81–97.
- Pindoriya, N. M., Singh, S. N., & Singh, S. K. (2008). An adaptive wavelet neural network-based energy price forecasting in electricity markets. *IEEE Transactions on Power Systems*, 23(3), 1423–1432.
- Pinson, P., & Tasty, J. (2014). Discussion of “Prediction intervals for short-term wind farm generation forecasts” and “Combined nonparametric prediction intervals for wind power generation”. *IEEE Transactions on Sustainable Energy*, 5(3), 1019–1020.
- Poole, D., Mackworth, A., & Goebel, R. (1998). *Computational intelligence: a logical approach*. Oxford University Press.
- Rambharat, B. R., Brockwell, A. E., & Seppi, D. J. (2005). A threshold autoregressive model for wholesale electricity prices. *Journal of the Royal Statistical Society, Series C*, 54(2), 287–300.
- Raviv, E., Bouwman, K. E., & van Dijk, D. (2013). *Forecasting day-ahead electricity prices: utilizing hourly prices*. Tinbergen Institute Discussion Paper 13-068/III. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.2266312>.
- Robinson, T. A. (2000). Electricity pool prices: a case study in nonlinear time-series modelling. *Applied Economics*, 32(5), 527–532.
- Rodriguez, C. P., & Anders, G. J. (2004). Energy price forecasting in the Ontario competitive power system market. *IEEE Transactions on Power Systems*, 19(1), 366–374.
- Ronn, E. I., & Wimschulte, J. (2009). Intra-day risk premia in European electricity forward markets. *Journal of Energy Markets*, 2(4), 71–98.
- Rubin, O. D., & Babcock, B. A. (2013). The impact of expansion of wind power capacity and pricing methods on the efficiency of deregulated electricity markets. *Energy*, 59(15), 676–688.
- Ruibal, C. M., & Mazumdar, M. (2008). Forecasting the mean and the variance of electricity prices in deregulated markets. *IEEE Transactions on Power Systems*, 23(1), 25–32.
- Rutkowski, L. (2008). *Computational intelligence: methods and techniques*. Springer.
- Sanchez, I. (2008). Adaptive combination of forecasts with application to wind energy. *International Journal of Forecasting*, 24, 679–693.
- Sansou, D. C., Downs, T., & Saha, T. K. (2002). Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants. *Journal of Electrical and Electronics Engineering, Australia*, 22(3), 227–233.
- Sapio, S., & Wylomańska, A. (2008). The impact of forward trading on the spot power price volatility with Cournot competition. *IEEE Conference Proceedings – EEM08*, art. no. 4579013.
- Schluerer, S. (2010). A long-term/short-term model for daily electricity prices with dynamic volatility. *Energy Economics*, 32, 1074–1081.
- Schmutz, A., & Elkuch, P. (2004). Electricity price forecasting: application and experience in the European power markets. In *Proceedings of the 6th IAAE European Conference*, Zürich.
- Seifert, J., & Uhrig-Homburg, M. (2007). Modelling jumps in electricity prices: theory and empirical evidence. *Review of Derivatives Research*, 10, 59–85.
- Serinaldi, F. (2011). Distributional modeling and short-term forecasting of electricity prices by generalized additive models for location, scale and shape. *Energy Economics*, 33(6), 1216–1226.
- Shafie-Khah, M., Moghaddam, M. P., & Sheikh-El-Eslami, M. K. (2011). Price forecasting of day-ahead electricity markets using a hybrid forecast method. *Energy Conversion and Management*, 52(5), 2165–2169.
- Shahidehpour, M., Yamin, H., & Li, Z. (2002). *Market operations in electric power systems: forecasting, scheduling, and risk management*. Wiley.
- Sharma, V., & Srinivasan, D. (2013). A hybrid intelligent model based on recurrent neural networks and excitable dynamics for price prediction in deregulated electricity market. *Engineering Applications of Artificial Intelligence*, 26(5–6), 1562–1574.
- Shumway, R. H., & Stoffer, D. S. (2006). *Time series analysis and its applications* (2nd ed.). Springer.
- Singleton, K. J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, 102, 111–141.
- Skantze, P. L., & Ilic, M. D. (2001). *Valuation, hedging and speculation in competitive electricity markets: a fundamental approach*. Kluwer Academic Publishers.
- Smith, D. G. (1989). Combination of forecasts in electricity demand prediction. *Journal of Forecasting*, 8, 349–356.
- Sousa, T. M., Pinto, T., Vale, Z., Praca, I., & Morais, H. (2012). Adaptive learning in multiagent systems: a forecasting methodology based on error analysis. *Advances in Intelligent and Soft Computing*, 156, 349–357.
- Stevenson, M. (2001). *Filtering and forecasting spot electricity prices in the increasingly deregulated Australian electricity market*. QFRC Research Paper No 63, UTS.
- Stevenson, M. J., Amaral, J. F. M., & Peat, M. (2006). Risk management and the role of spot price predictions in the Australian retail electricity market. *Studies in Nonlinear Dynamics and Econometrics*, 10(3), Article 4.
- Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460), 1167–1179.
- Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23, 405–430.
- Sun, J., & Tesfatsion, L. (2007). Dynamic testing of wholesale power market designs: an open-source agent-based framework. *Computational Economics*, 30, 291–327.
- Tan, Z., Zhang, J., Wang, J., & Xu, J. (2010). Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Applied Energy*, 87(11), 3606–3610.
- Tay, A. S., & Wallis, K. F. (2000). Density forecasting: a survey. *Journal of Forecasting*, 19, 235–254.
- Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operations Research*, 204, 139–152.
- Taylor, J. W., & Majithia, S. (2000). Using combined forecasts with changing weights for electricity demand profiling. *Journal of the Operational Research Society*, 51, 72–82.
- Timmermann, A. G. (2006). Forecast combinations. In G. Elliott, C. W. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (pp. 135–196). Elsevier.
- Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford University Press.

- Tong, H., & Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B*, 42, 245–292.
- Trück, S., Weron, R., & Wolff, R. (2007). Outlier treatment and robust approaches for modeling electricity spot prices. *Proceedings of the 56th Session of the ISI*. Available at MPRA: <http://mpra.ub.uni-muenchen.de/4711/>.
- Ullrich, C. J. (2012). Realized volatility and price spikes in electricity markets: the importance of observation frequency. *Energy Economics*, 34(6), 1809–1818.
- Vahidinab, V., Jadid, S., & Kazemi, A. (2008). Day-ahead price forecasting in restructured power systems using artificial neural networks. *Electric Power Systems Research*, 78(8), 1332–1342.
- Vahviläinen, I., & Pyykkönen, T. (2005). Stochastic factor model for electricity spot price – the case of the Nordic market. *Energy Economics*, 27(2), 351–367.
- Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5, 177–188.
- Ventosa, M., Baillo, Á., Ramos, A., & Rivier, M. (2005). Electricity market modeling trends. *Energy Policy*, 33(7), 897–913.
- Vilar, J. M., Cao, R., & Aneiros, G. (2012). Forecasting next-day electricity demand and price using nonparametric functional methods. *Electrical Power and Energy Systems*, 39, 48–55.
- Vives, X. (1999). *Oligopoly pricing*. Cambridge, MA: MIT Press.
- Wallis, K. F. (2003). Chi-squared tests of interval and density forecasts, and the Bank of England fan charts. *International Journal of Forecasting*, 19, 165–175.
- Wallis, K. F. (2005). Combining density and interval forecasts: a modest proposal. *Oxford Bulletin of Economics and Statistics*, 67, 983–994.
- Wallis, K. F. (2011). Combining forecasts – forty years later. *Applied Financial Economics*, 21, 33–41.
- Wan, C., Xu, Z., Wang, Y., Dong, Z. Y., & Wong, S. K. P. (2014). A hybrid approach for probabilistic forecasting of electricity price. *IEEE Transactions on Smart Grid*, 5(1), 463–470.
- Wang, L., & Fu, X. (2005). *Data mining with computational intelligence*. Springer.
- Wang, P., Zareipour, H., & Rosehart, W. D. (2014). Descriptive models for reserve and regulation prices in competitive electricity markets. *IEEE Transactions on Smart Grid*, 5(1), 471–479.
- Weber, R. (2006). *Uncertainty in the electric power industry*. Springer.
- Weidlich, A., & Veit, D. (2008). A critical survey of agent-based wholesale electricity market models. *Energy Economics*, 30, 1728–1759.
- Weron, R. (2006). *Modeling and forecasting electricity loads and prices: a statistical approach*. Chichester: Wiley.
- Weron, R. (2008). Market price of risk implied by Asian-style electricity options and futures. *Energy Economics*, 30, 1098–1115.
- Weron, R. (2009). Heavy-tails and regime-switching in electricity prices. *Mathematical Methods of Operations Research*, 69(3), 457–473.
- Weron, R., Bierbrauer, M., & Trück, S. (2004). Modeling electricity prices: jump diffusion and regime switching. *Physica A*, 336, 39–48.
- Weron, R., Simonsen, I., & Wilman, P. (2004). Modeling highly volatile and seasonal markets: evidence from the Nord Pool electricity market. In H. Takayasu (Ed.), *The application of econophysics* (pp. 182–191). Tokyo: Springer.
- Weron, R., & Misiorek, A. (2005). Forecasting spot electricity prices with time series models. *IEEE Conference Proceedings – EEM05* (pp. 133–141).
- Weron, R., & Misiorek, A. (2006). Short-term electricity price forecasting with time series models: A review and evaluation. In W. Mielczarski (Ed.), *Complex electricity markets* (pp. 231–254). Łódź: IEPL& SEP.
- Weron, R., & 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., & Zator, M. (2014a). Revisiting the relationship between spot and futures prices in the Nord Pool electricity market. *Energy Economics*, 44, 178–190.
- Weron, R., & Zator, M. (2014b). *A note on using the Hodrick–Prescott filter in electricity markets*. Working paper version available from RePEc: <http://ideas.repec.org/p/wuu/wpaper/hsc1404.html>.
- Winkler, R. L. (1972). A decision-theoretic approach to interval estimation. *Journal of the American Statistical Association*, 67(337), 187–191.
- Wolak, F. A. (2000). Market design and price behavior in restructured electricity markets: an international comparison. In T. Ito & A. O. Krueger (Eds.), *Deregulation and interdependence in the Asia-Pacific Region, NBER-EASE: Vol. 8* (pp. 79–137). University of Chicago Press.
- Wood, A. J., & Wollenberg, B. F. (1996). *Power generation, operation and control*. New York: Wiley.
- Wu, H. C., Chan, S. C., Tsui, K. M., & Hou, Y. (2013). A new recursive dynamic factor analysis for point and interval forecast of electricity price. *IEEE Transactions on Power Systems*, 28(3), 2352–2365.
- Wu, L., & Shahidehpour, M. (2010). A hybrid model for day-ahead price forecasting. *IEEE Transactions on Power Systems*, 25(3), 1519–1530.
- Yamin, H. Y., Shahidehpour, S. M., & Li, Z. (2004). Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets. *International Journal of Electrical Power and Energy Systems*, 26, 571–581.
- Yan, X., & Chowdhury, N. A. (2010a). *Electricity market clearing price forecasting in a deregulated market: a neural network approach*. VDM Verlag Dr. Müller.
- Yan, X., & Chowdhury, N. A. (2010b). Mid-term electricity market clearing price forecasting: a hybrid LSSVM and ARMAX approach. *International Journal of Electrical Power and Energy Systems*, 53(1), 20–26.
- Yang, H. T., Huang, C. M., & Huang, C. L. (1996). Identification of ARMAX model for short term load forecasting: an evolutionary programming approach. *IEEE Transactions on Power Systems*, 11, 403–408.
- Yao, S. J., Song, Y. H., Zhang, L. Z., & Cheng, X. Y. (2000). Prediction of system marginal price by wavelet transform and neural network. *Electric Machines and Power Systems*, 28(10), 983–993.
- Zareipour, H. (2008). *Price-based energy management in competitive electricity markets*. VDM Verlag Dr. Müller.
- Zareipour, H., Bhattacharya, K., & Canizares, C. A. (2007). Electricity market price volatility: the case of Ontario. *Energy Policy*, 35, 4739–4748.
- Zareipour, H., Canizares, C. A., Bhattacharya, K., & Thomson, J. (2006). Application of public-domain market information to forecast Ontario's wholesale electricity prices. *IEEE Transactions on Power Systems*, 21(4), 1707–1717.
- Zareipour, H., Janjani, A., Leung, H., Motamedi, A., & Schellenberg, A. (2011). Classification of future electricity market prices. *IEEE Transactions on Power Systems*, 26(1), 165–173.
- Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, 14, 35–62.
- Zhang, L., & Luh, P. B. (2005). Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method. *IEEE Transactions on Power Systems*, 20(1), 59–66.
- Zhang, L., Luh, P. B., & Kasiviswanathan, K. (2003). Energy clearing price prediction and confidence interval estimation with cascaded neural networks. *IEEE Transactions on Power Systems*, 18(1), 99–105.
- Zhao, J. H., Dong, Z. Y., Xu, Z., & Wong, K. P. (2008). A statistical approach for interval forecasting of the electricity price. *IEEE Transactions on Power Systems*, 23(2), 267–276.
- Zięba, M. M., Tomczak, J. M., Lubicz, M., & Świątek, J. (2014). Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients. *Applied Soft Computing*, 14, 99–108.
- Zou, H., & Yang, Y. (2004). Combining time series models for forecasting. *International Journal of Forecasting*, 20, 69–84.