International Journal of Forecasting 39 (2023) 570–586

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Forecasting electricity prices with expert, linear, and nonlinear models

Anna Gloria Billé [a], Angelica Gianfreda [c,d,*], Filippo Del Grosso [b], Francesco Ravazzolo [b,e,f]

[a] Department of Statistical Sciences, University of Bologna, Italy
[b] Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy
[c] Department of Economics, University of Modena and Reggio Emilia, Italy
[d] Energy Markets Group, London Business School, London, UK
[e] BI Norwegian Business School, Norway
[f] RCEA, Rimini Center for Economic Analysis, Italy

ARTICLE INFO

Keywords:
Demand
Wind
Solar
Biomass
Waste
Fossil fuels (coal, natural gas, CO_2)
Weighted inflows
Commercial and public forecasts

ABSTRACT

This paper compares several models for forecasting regional hourly day-ahead electricity prices, while accounting for fundamental drivers. Forecasts of demand, in-feed from renewable energy sources, fossil fuel prices, and physical flows are all included in linear and nonlinear specifications, ranging in the class of ARFIMA-GARCH models—hence including parsimonious autoregressive specifications (known as *expert-type* models). The results support the adoption of a simple structure that is able to adapt to market conditions. Indeed, we include forecasted demand, wind and solar power, actual generation from hydro, biomass, and waste, weighted imports, and traditional fossil fuels. The inclusion of these exogenous regressors, in both the conditional mean and variance equations, outperforms in point and, especially, in density forecasting when the superior set of models is considered. Indeed, using the model confidence set and considering northern Italian prices, predictions indicate the strong predictive power of regressors, in particular in an expert model augmented for GARCH-type time-varying volatility. Finally, we find that using professional and more timely predictions of consumption and renewable energy sources improves the forecast accuracy of electricity prices more than using predictions publicly available to researchers.

1. Introduction

Forecasting day-ahead electricity prices has always attracted the attention of practitioners and scholars because trading decisions are based on strategic and stochastic components, like arbitrage speculations, the impossibility of storing electricity, and variability introduced into the system by the effects of new regulations and the imperfect predictability of fundamental drivers. This paper investigates both aspects.

Day-ahead electricity prices are determined for each hour of the following day by the intersection of the aggregated curves of demand and supply. Therefore, factors that influence both curves have been extensively investigated in price modeling. Fundamental variables such as forecasted demand and weather conditions have been taken into account for the demand curve, whereas the predicted intermittent generation by renewable energy sources (RES) has been recently considered a risk source in the supply curve, together with import and export flows and the international movement of fossil fuel prices used in traditional thermal plants; for extensive reviews see Weron (2014), Nowotarski and Weron (2018), and Hong et al. (2020).

* Corresponding author at: Department of Economics, University of Modena and Reggio Emilia, Italy.
E-mail address: angelica.gianfreda@unimore.it (A. Gianfreda).

https://doi.org/10.1016/j.ijforecast.2022.01.003
0169-2070/

All these variables must be considered in the formulation of ex ante expectations of day-ahead electricity prices. Furthermore, in recent years, the power generated by RES has increased substantially, due to incentives and the worldwide goal of reducing carbon emissions. Indeed, as a country in the European Union (EU), Italy is among the top six countries in the world for renewable power capacity (not including hydro), after Germany and together with the United Kingdom. Specifically, Italy is among the top EU countries for wind and solar photovoltaic (PV) capacity additions in 2017 (REN21, 2018).

The increasing RES generation dispatched on the day-ahead (and intraday) market has a twofold effect. According to the merit order, producing units that pollute less have the priority of dispatch and move the supply curve towards the right as soon as their generation increases. Consequently, equilibrium prices decrease, due to new RES generation. On one hand, all this has the effect of moving thermal conventional technologies out of the day-ahead market, and, on the other hand, it reduces the spreads between maximum and minimum prices which make water pumping units less profitable. They have no more time-arbitrage opportunities for buying electricity in off-peak hours and selling it during peak hours in the day-ahead market. Then, those units allowed to act in real-time sessions can try to recover their profits there. This occurred in Italy, attracting the attention of the energy regulator in 2016, when enormous costs were generated within the system as a consequence of the speculative trading of few thermal units. Gianfreda, Parisio, and Pelagatti (2018) studied the auction/bid data for the northern Italian zone, characterized by high solar PV and hydro penetration. Considering all market sessions, from the day ahead to real time and passing through intraday sessions, they provide empirical evidence that balancing costs increased between two samples associated with low (in years 2006—08) and high (in years 2013—15) RES levels. They studied the up- and down-regulation in the balancing market sessions, which differ across the Italian physical zones because of the location and characteristics of RES capacity. It is intuitive that a geographically balanced portfolio may compensate easily and promptly for any variations in demand or in generation (due to the forecast errors of RES output). However, the authors observed that the northern zone appears to be subject to a systematic overestimation of PV generation capacity sold in the day-ahead market, hence requiring up-regulation to restore the system equilibrium at a price which is generally more costly than the one for down-regulation. Considering that the seasonality of solar production reduces the residual demand covered by conventional technologies during hours of irradiation and that it requires a strong increase in programmable and flexible production at sunset, the evening ramp increased from 8250 MW in 2012 to 11,050 MW in 2014; and it was contemporarily paired with the dismissal of a number of old thermal units. They observed that some generators, allowed to act on the balancing market, were withholding capacity on the day-ahead market (or closing their net position to zero over day-ahead and intraday sessions) and selling energy in the real-time sessions, where the pay-as-bid pricing mechanism grants the (higher) price declared in accepted bids. These Italian sessions have a limited number of traders and are dominated by conventional (thermal, hydro, and water pumping) technologies with no competition from RES units (indeed they are not allowed to participate in the Italian balancing sessions) and so they can only reduce the day-ahead prices, as an effect of the merit order.

To overcome these critical issues, some EU countries, including Italy, have started to discuss the possibility of allowing RES units to act in the balancing markets. Meanwhile, however, the prediction of prices on the day-ahead market is becoming an increasingly important and essential step in the evaluation of trading strategies, since thermal conventional as well as water pumping units consider the price spreads among the various sequential sessions—together with the possibility to act over a long-term capacity market. Based on all these arguments and because of the raised issue in 2016, Italy is an excellent case study. Moreover, the zonal structure allows for the consideration of the operators' bidding behavior across different areas and according to the composition of their generation mix. Northern Italy is, therefore, an exceptionally good example for several reasons. First, the zone is well interconnected with foreign countries, from where electricity can be imported at lower prices. Second, a high share of solar PV generation has been observed in recent years. Third, most of the hydro generation is located in the Alps. Fourth, and more importantly, the demand for electricity in this zone represents almost half of the national demand; hence, variations in demand and supply can boost the strategic use of balancing sessions. Finally, all three thermal conventional, hydro, and water pumping technologies act in this zone across all different market sessions. Therefore, the prediction of day-ahead electricity prices observed in northern Italy can increase the understanding of the main drivers of these prices, and could contribute to monitoring (and hence controlling) the bidding strategies across market sessions, according to the price levels expected in the day-ahead market. Other studies based on different markets that consider bidding strategies and their associated economic value are presented in Bunn, Gianfreda, and Kermer (2018), Lisi and Edoli (2018), Abramova and Bunn (2020), and Kath et al. (2020).

Others attempts to capture the impacts of economic, technical, strategic, and risk factors on intraday prices are presented in Karakatsani and Bunn (2008). Oberndorfer (2009) focused on the relationship between energy market developments, external shocks, and the pricing of European utility stocks. Hickey, Loomis, and Mohammadi (2012) implemented ARMAX-GARCH models with trend, dummy variables for seasonality, and load for five MISO pricing hubs. Subsequently, Maciejowska and Weron (2016) focused on the increased granularity of data available on the British market (where prices have a half-hour frequency) to test a set of fundamental explanatory variables (i.e. natural gas, coal, and CO_2 emissions). de Marcos, Bello, and Reneses (2019) proposed an econometric and fundamental approach to forecast short-term prices in the Iberian market by pairing a neural network with a set of expected and actual fundamental variables. Gianfreda, Ravazzolo, and Rossini (2020) compared several univariate and multivariate models augmented with fundamental variables, including demand forecasts and forecasted production from renewable energy sources, to predict hourly day-ahead electricity prices in several European markets.

According to the literature, few papers have inspected the predictability of day-ahead prices in northern Italy. The most notable studies are Gianfreda and Grossi (2012), Shah and Lisi (2019), and Bernardi and Lisi (2020). The latter two papers adopt a generalized additive location–scale model with a non-parametric estimation of the conditional mean and variance and a nonparametric functional autoregressive model based on individual bids, whereas the former one considers the Italian zonal prices during the years 2006–2008, when RES had a limited (or no) role in the determination of prices. Indeed, in that contribution, wind, solar, and hydro were not considered.

Accounting for the arguments that strong electricity price autocorrelations and long memory may be induced by the mean-reverting nature of market fundamentals, by the highly repetitive nature of electricity auctions, or by increased market integration (Knittel & Roberts, 2005; Haldrup & Nielsen, 2006; Conejo, Contreras, Espinola, & Plazas, 2005; Koopman, Ooms, & Carnero, 2007 and Jeon & Taylor, 2016), we select AR(FI)MA-GARCH-type models and compare their forecasting ability with and without a set of regressors, while adopting a rolling window approach and an adaptive scheme. The former approach recalls the dynamic evolution of fundamentals over time, in line with the time-varying parameter regression model implemented in Karakatsani and Bunn (2008) to adapt price structures to market changes continuously. Furthermore, the latter scheme develops to the estimation strategy implemented in Weron and Misiorek (2008), Chen and Bunn (2014), and Maciejowska and Weron (2016), by extending the selection to both the autoregressive and moving average lag orders for each calibration window and each model specification, including the options to switch from one model to another one and to replace negative forecasted prices with null prices (since negative pricing is not allowed in the Italian market). Additionally, parsimonious autoregressive models extended for regressors and time-varying volatility have been included in the analysis, following Ziel (2016) and Ziel and Weron (2018). Therefore, in what follows with refer to these models as those built on some *expert knowledge*.

It is worth noting that we expand these models by including our set of fundamentals (that is, predicted values for wind, solar PV, and demand, together with actual values for biomass, hydro, waste, and weighted flows plus fossil fuel prices). Then, we explore a total of 58 linear and nonlinear specifications to provide empirical evidence of their forecasting performance, given the mixed results in the literature (see Hong, Pinson, & Fan, 2014 among others). Specifically, we test several AR(FI)MAX-GARCH and *expert-type* models, and we additionally investigate LASSO variants for the selection of exogenous regressors, dummy variables, and autoregressive terms when the lag ordering is set at high values. Recent literature on LASSO and its applications can be found in Ziel, Steinert, and Husmann (2015), Ziel and Weron (2018), Uniejewski, Marcjasz, and Weron (2019), and Messner and Pinson (2019) among others.

In addition, our contribution relies on applying both the Diebold–Mariano (DM) (Diebold & Mariano, 1995) and the model confidence set (MCS) (Hansen, Lunde, & Nason, 2011) testing procedures to account for the large model uncertainty in evaluation. A density forecast exercise is also provided to guide practitioners in choosing the best model according to different hours.

More importantly, given the issue of data availability, market transparency, and the economic relevance of accurate predictions, as discussed in Kezunovic et al. (2020) and Gonçalves, Pinson, and Bessa (2021), we include an interesting analysis in which we compare the forecasting performance when professional and more timely forecasts are used in place of public and freely available forecasts. Maciejowska, Nitka, and Weron (2021) show that these freely available forecasts of fundamental variables are biased and could be improved. We confirm that including fundamental factors improves the forecasting ability. In particular, our expert EX_4X model augmented with fundamental drivers gives more accurate point forecasts: none of the other 57 models is statistically superior to it at any hour, despite the large number of specifications found in the model confidence set.

The evidence is different when the loss function is generalized to density forecasting: all models with GARCH time-varying volatility give the most accurate density forecasts and they are statistically superior to models that exclude it.

When professional forecasts are used, the forecast power further increases, in particular for the early-morning and peak hours.

In detail, we find that the inclusion of exogenous regressors reduces both the RMSEs and the CRPSs, especially during peak hours. More specifically, an expert model (our EX_4X) and its GARCH specifications drastically outperform all other models in point forecasts. From a practical point of view, this expert model and its GARCH variants are the only ones retained for all hours in the MCS, and especially when hour 19 is considered. In addition, the results on CRPS and DM show that there are substantial improvements when all models are enlarged to include GARCH time-varying volatility.

In a context characterized by a limited number of regressors with respect to the amount of statistical information available, we find that there are no substantial improvements when LASSO models are considered.

In addition, for the first time, to our best knowledge, we provide empirical evidence that using commercial forecasts improves price forecasts substantially, especially during hours 1–7 and peak hours 8–20. Then, as soon as the forecasting horizon increases, as after hour 21, the benefits of these more timely forecasts disappear. And we emphasize that this evidence is driven by the usage of professional forecasts and not by considering simple or complex models (in both cases, we observe improvements).

Finally, we also assess the coefficients of the exogenous regressors in our best model to investigate their degree of significance through the considered sample. We provide evidence that a model accounting for the dependence of prices over their demeaned prices of the previous 8 days, and including forecasted load, wind, and solar, as well as actual hydro and natural gas prices, seems 'expert' enough to explain well and forecast even better the northern Italian day-ahead prices.

The remainder of the paper is structured as follows. Section 2 presents a brief description of the Italian market with a focus on the northern zone. Section 3 provides a detailed description of the data employed and the methodological strategy used to predict hourly electricity prices. Section 3.3 describes the estimation, and Section 4 presents the results. Finally, Section 5 concludes.

2. Italian market structure and the northern zone

The Italian electricity market is structured into three main segments: the day-ahead, intraday, and ancillary services markets. The latter is paired by the balancing market operated in real time on the day of delivery. Day-ahead and intraday segments are open to a variety of national and international operators (producers, consumers, traders), for a total of 258 different market participants in 2017.[1] Market participation is voluntary in both the day-ahead and intraday markets, whereas it is compulsory in the ancillary services market sessions where only balancing units with the required degree of flexibility are allowed to act. We focus on the day-ahead market, which opens nine days before the day of delivery and closes at noon on the day before delivery.

The Italian electricity market is structured into geographical and foreign virtual zones. The geographical zones represent a portion of the national grid delimited by bottlenecks in transmission capacity, and these are northern Italy, central-northern Italy, central-southern Italy, southern Italy, Sicily, and Sardinia. The foreign virtual zones are points of interconnection with neighboring countries. In this paper we consider northern Italy; thus, the foreign virtual zones in this analysis are France, Switzerland, Austria, and Slovenia.

Each geographical and virtual zone yields an hourly (clearing) price, obtained from an implicit bidding mechanism in which pairs of quantities (in MWh) and prices (in €/MWh) are considered by accounting for market splitting in case of congestion. Therefore, in the same hour, zonal prices in contiguous market zones can differ depending on transmission bottlenecks. The zonal prices concur to generate the single national price (or *prezzo unico nazionale*, PUN), that is, the average of zonal day-ahead prices weighted for total purchases, net of purchases for pumped-storage units, and purchases by neighboring zones. Additional details on the Italian market structure and the process of the creation of a system marginal price are found in Gianfreda, Parisio, and Pelagatti (2019), Gianfreda, Parisio, Pelagatti, et al. (2016), and Shah and Lisi (2019).

These researchers have emphasized the differences in the generation mix across regions and how the industrial activities are mainly concentrated in the northern area of the country, which is by far the most relevant in terms of consumption, due to the high concentration of population and industries. Northern consumption is 175,396 GWh over 303,443 GWh at the national level. Energy intensity is consistently higher, with an average of 6326 kWh per inhabitant versus a national average of 5024 kWh (Terna, 2018). In 2017, production in the northern zone was 149,204 GWh over a total of 289,708 GWh, or roughly 51%.

The northern area is also characterized by a varied, flexible generation mix, with 26% hydropower, and other renewables such as solar (6%) and biomass (8%), with conventional thermal generation covering the remaining proportion. Yearly details on the evolution of the portfolio generation are reported in Table B.7 in Appendix B for all zones and across the years 2015–2019. At first sight, given the low share of wind, a reader could argue about the choice of selecting northern Italy to understand the contribution of main drivers to the forecasts of future prices. However, we would like to emphasize that this zone has the highest hydro generation and demand; and more importantly, all three types of thermal, hydro, and water pumping units act in all market sessions. This zone is also connected with four foreign countries, whereas the others have only national connections or limited numbers of foreign connections.

Italy has arranged market-coupling agreements with Slovenia since 2011, and with France and Austria since 2015, which represent completion steps to the creation of a single internal electricity market in Europe. Market coupling allows for the simultaneous calculation of electricity prices and cross-border flows across coupled regions, and the main benefits are both an optimized and more efficient utilization of cross-border capacity and a better price alignment among different countries. Because of the relevant interconnection capacity between foreign countries and northern Italy, it is possible to import electricity at a lower price. For instance, in 2018, Italy imported 47,170 GWh of electricity (approximately equivalent to 15% of total consumption) from French, Swiss, and Slovenian borders. Table B.8 in Appendix B summarizes the information related to the local mix, and it reports the technology shares over total installed capacity in neighboring countries. Furthermore, the inspection of import/export flows presented in Table 1 shows the relevance of imports.[2] Hence, cross-border flows are included in this analysis through the construction of an artificial variable to account for prices determined in interconnected countries and in central-northern Italy, where the local mixes differ substantially. In this way, we account for their generating portfolio when using prices weighted by the quantity imported.

[1] The spot market is complemented by the forward market (a platform for different types of contracts) and by the bilateral contract platform (where all OTC energy transactions that require flows through the power grid are registered).

[2] Details on the dynamics of imports from neighboring countries over months and across hours are omitted but are available on request.

Table 1
Italian imports from and exports to other neighboring countries (in GW).
Data: ENTSO-E.

Year	France		Austria		Switzerland		Slovenia		Malta		Greece	
	Imports	Exports	Imports	Exports	Imports	Exports	Imports	Exports	Imports	Exports	Imports	Exports
2015	13,335	85	1526	33	25,263	47	6179	16	0	926	588	1657
2016	11,056	286	1420	55	19,846	315	6371	16	0	1522	302	1999
2017	10,860	280	1313	108	20,490	272	5784	23	33	887	325	1614
2018	13,102	79	1391	20	21,406	122	6707	11	8	606	1053	621
2019	15,134	98	1215	1	21,231	121	5140	170	18	654	55	3028

3. Data and methodology

This section provides a detailed overview of the available data and then explains the methodological strategy to predict hourly electricity prices. In particular, Section 3.1 describes both the endogenous and the exogenous variables used in our model specifications, while Section 3.2 shows all model specifications and the forecast procedure.

3.1. Data

To perform our analysis, we use day-ahead electricity prices determined hourly in the northern zone of Italy, and hourly forecasted load, wind, and solar generation, as well as actual biomass, waste, and hydropower generated in northern Italy, together with weighted imports and prices for fossil fuels (coal and natural gas) and CO_2 emissions.

Northern Italian zonal prices (in €/MWh) were collected directly from the website of the Italian system operator (Gestore dei Mercati Energetici, GME).[3] Forecasted load, wind, and solar were collected from the European Network of Transmission System Operators for Electricity (ENTSO-E) and from Refinitiv Thomson Reuters (RTR), and re-scaled from MW to GW.

Then, we use both public and private forecasts to compare the forecasting performances of our models. Specifically, we use public ENTSO-E forecast data[4] from 2015 to 2019 and professional RTR forecasts from 2018 to 2019, since hourly forecasted load, wind, and solar were fully available for the northern Italian region only from 2018. In the latter case, we consider the forecasts produced by two weather providers: the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Global Forecast System of the American weather service of the National Centers for Environmental Prediction (GFS). Both providers use two types of weather models—a deterministic one with no involved randomness and high resolution (the *operational* model), and a probabilistic one with lower resolution but with variations of weather conditions (the *ensemble* model)—and different runs (one run for the op and between 21 or 51 runs for the ens) at specific hours (namely at midnight, at 6 a.m., at 12 a.m., and at 6 p.m.). Then, according to their ending time of updates and publication, we use two different series of forecasts:[5] one for forecasting models running quickly (*fast*, F), and so including more recent information released at 7.40 a.m.; and one for models running less quickly (*less fast*, LF), then including the information released at 6.55 a.m. These contain the latest information available to market operators to run their forecasting models and formulate their day-ahead bidding strategy of 24 forecasted hourly prices to be submitted (by noon) on the day-ahead market. Hence, in this paper we compare public ENTSO-E with private RTR forecasts to inspect the different forecasting performances.

The relevant information for actual biomass, waste, and hydro (generated for all 24 h) and physical flows are collected from ENTSO-E. However, this information is not available in a timely manner for their inclusion in the forecasting models of all the 24 price series, because the quantities usually displayed before noon refer up to hour 11.[6] Therefore, we consider the lagged actual biomass, waste, and hydro generation together with flows for hours 1–10, and their realized values observed at hour 10 for hours 11–24.

To construct the weighted imports, we use ENTSO-E data for imports and prices of foreign countries. In particular, to consider the effect of imports from foreign countries and from the contiguous zone (central-northern Italy), we account for the different prices observed in neighboring foreign markets and we construct a series of average hourly prices (expressed in €/MWh) weighted for the quantity of electricity imported. Specifically, this is calculated as the average of day-ahead hourly prices determined in Austria, France, Switzerland, Slovenia, and central-northern Italy, weighted for the actual hourly electricity physical flows, to capture the effects of electricity transits across bordering markets and the neighboring national zone.

Finally, to account for the marginal costs of conventional thermal generation, we use the Dutch TTF natural gas prices (for delivery over the next month), the ICE API2 Rotterdam Future prices for coal, and the EEX-EU CO2

[3] http://www.mercatoelettrico.org.

[4] This information is published per time unit at the latest two hours before the gate closure time of the day-ahead market or at 12.00 (local time) at the latest, when the gate closure time does not apply. This represents the publication deadline for ENTSOE and actually refers to data available to market operators at (the latest) 10 a.m.

[5] The first series for *fast* models uses forecasts obtained considering first the model ECens00 (which ends its updates at 7.40 a.m.). Any missing forecasts are then replaced by the ECop00 (since this ends at 6.55 a.m.). If necessary, we use the same replacement scheme using GFSen00 and GFSop00, ECen18 and ECop18, and GFSen18 and GFSop18. By contrast, the second series for *less fast* models simply starts with ECop00. Please note that the runs at 18 were available only from 2018.

[6] The hourly aggregated output is generally published no later than one hour after the operational period, as described by ENTSO-E.

emissions E/EUA prices in euros, all collected from RTR Datastream. These prices are settlement prices, released at the end of the day at approximately 7 p.m., and hence included with a time lag $t-1$.

Our final database comprises 35,064 hourly observations for each variable, from January 2015 to December 2019, apart from models using RTR forecasted regional data, which cover only 2018 and 2019.

Following Bunn (2000), Cuaresma, Hlouskova, Kossmeier, and Obersteiner (2004), and subsequent references, we adopt a variable segmentation approach. The modeling and forecasting process considers hourly time series per time. That is, we model and forecast each of the hourly prices individually. Moreover, the model specification strategy replaces missing or incomplete actual hourly data (when they are unavailable because they have not yet been published) with the corresponding information observed for the same hour on the day before.

Unlike Weron (2007) and Afanasyev and Fedorova (2019), we maintain the outliers in all the variable series, and we do not decompose the effects of seasonality. We claim that outliers represent peculiar characteristics of the Italian market, since they incorporate notable market information in terms of sample variance and arbitrage opportunity from a day-ahead trading perspective. In addition and in contrast to Conejo et al. (2005), Garcia, Contreras, van Akkeren, and Garcia (2005), Weron and Misiorek (2008), and Bordignon, Bunn, Lisi, and Nan (2013) among others, we do not apply logarithms to prices to improve normality and stabilize variance, since this transformation could mask the statistical price properties and volatility dynamics that we want to capture and model; see Karakatsani and Bunn (2010) and Paraschiv, Erni, and Pietsch (2014) for a similar choice.

The descriptive statistics of the selected variables are reported in Table 2, and their dynamics are depicted in Figs. 1 and 2. Prices show a degree of skewness and high kurtosis (as for solar, wind, and weighted imports). Even if the hourly electricity prices range between 5 and 206.12 €/MWh, Italian power prices have a floor of 0 €/MWh and a cap of 3000 €/MWh. Notably, even if wind generation in northern Italy exhibits low values (a range between 0 and 20 MW), we include this variable for the sake of generality, completeness, and consistency with the local generation mix, as suggested by Ziel et al. (2015); for the same reason, we include biomass and waste. This general approach can be applied to other zones or markets, since it is reasonable to include all fundamental drivers and to expect limited significance of those with lower generation shares. Moreover, it allows for possible changes in the local generation induced by changes in policy regulation or weather conditions.

Consumption and electricity prices present weekly and calendar seasonality, with consumption levels higher on working days and lower on weekends. These features are more evident in Fig. 2, where time series are presented for a sample of hours within peak and off-peak periods (i.e. hours 3, 9, 13, 15, 21, and 24). Consistently, monthly seasonality is characterized by a consumption peak in winter months (January and February) and a peak in summer months, because of the widespread use of cooling systems and heat pumps. Wind and solar PV generation fluctuate according to weather conditions, and solar PV generation also fluctuates according to the hours of solar radiation. Electricity inflows from the bordering central-northern Italian zone and foreign markets (Austria, France, Switzerland, and Slovenia) also exhibit strong seasonality, especially at the beginning of our sample. To help in understanding the effects of these regressors on prices, their intradaily dynamics are shown in Figure C.3 in Appendix C.

We consider the Jarque–Bera (JB) test to check for the normality of error terms (Jarque & Bera, 1987), and both the augmented Dickey–Fuller (ADF) (Dickey & Fuller, 1979; Said & Dickey, 1984) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests for the stationarity (Kwiatkowski, Phillips, Schmidt, & Shin, 1992). We observed non-normality according to the JB test, stationarity according to the ADF test, and both level and trend non-stationarity according to the KPSS test. These results for all hours are omitted but available on request.

3.2. Model specifications

We use several expert- and AR(FI)MA-GARCH-type models.

The first expert specification (EX_1) simply considers past prices observed on one, two, and seven days before with weekdays dummies. Formally, the hourly price y_t (for simplicity we omit the subscript h) is modeled as

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-7} + \sum_{k=1}^{6} \gamma_k D_t^k + \varepsilon_t \qquad (1)$$

where D_t^1 is equal to one for Mondays, D_t^2 for Tuesdays, and so on up to D_t^6 for Saturdays. We extend this model with a set of exogenous regressors, having the EX_1X defined as

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-7} + \sum_{k=1}^{6} \gamma_k D_t^k + \lambda' \mathbf{x}_t + \kappa' \mathbf{z}_{t-1} + \varepsilon_t \qquad (2)$$

where $\mathbf{x}_t$ is the vector at time t of exogenous regressors, which include forecasted load, wind, and solar PV generation, whereas $\mathbf{z}_{t-1}$ is a vector for exogenous regressors at time $t-1$ since we use actual hydro, biomass, and waste generation, together with weighted imports, natural gas, and CO_2 prices.

The second expert model (EX_2) builds upon the EX_1 model by including the lowest and the highest hourly prices observed on the previous day. Formally,

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-7} + \beta_4 y_{min,t-1} + \beta_5 y_{max,t-1} + \sum_{k=1}^{6} \gamma_k D_t^k + \varepsilon_t \qquad (3)$$

As before, the EX_2X includes the exogenous regressors $\mathbf{x}_t$ and $\mathbf{z}_{t-1}$.

The third expert model EX_3 expands the EX_2 by including the price at hour 24 of the previous day (this is

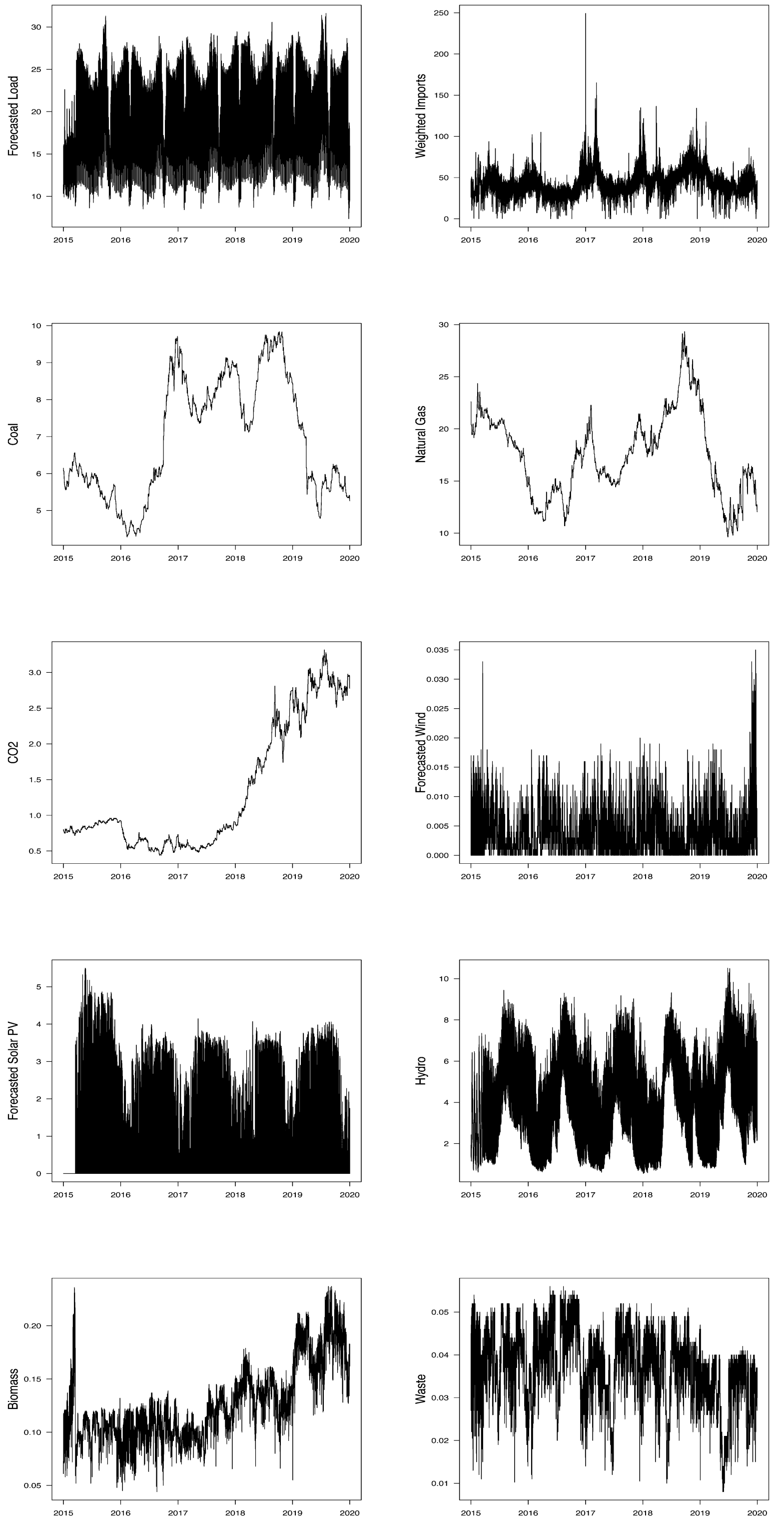

Fig. 1. Time series of all used exogenous variables.

Table 2
Descriptive statistics of fundamental variables computed over the full sample.

	Min	Mean	Max	Std. Dev.	Skewness	Kurtosis
Price	1.000	52.345	206.120	16.364	1.107	3.426
Forecasted load	7.344	18.624	31.617	4.858	0.164	−1.107
Weighted import	0.000	43.075	249.340	15.551	0.911	2.883
Coal	4.280	6.961	9.840	1.598	0.143	−1.350
Natural gas	9.630	17.575	29.330	3.986	0.296	−0.367
CO_2	0.440	1.330	3.316	0.877	0.829	−0.937
Forecasted solar	0.000	0.765	5.499	1.153	1.417	0.832
Forecasted wind	0.000	0.004	0.035	0.004	1.509	3.623
Hydro	0.550	3.910	10.510	2.029	0.348	−0.772
Biomass	0.044	0.128	0.237	0.036	0.818	−0.013
Waste	0.008	0.037	0.056	0.009	−0.532	0.058

Note that *Std. Dev.* means *standard deviation*.

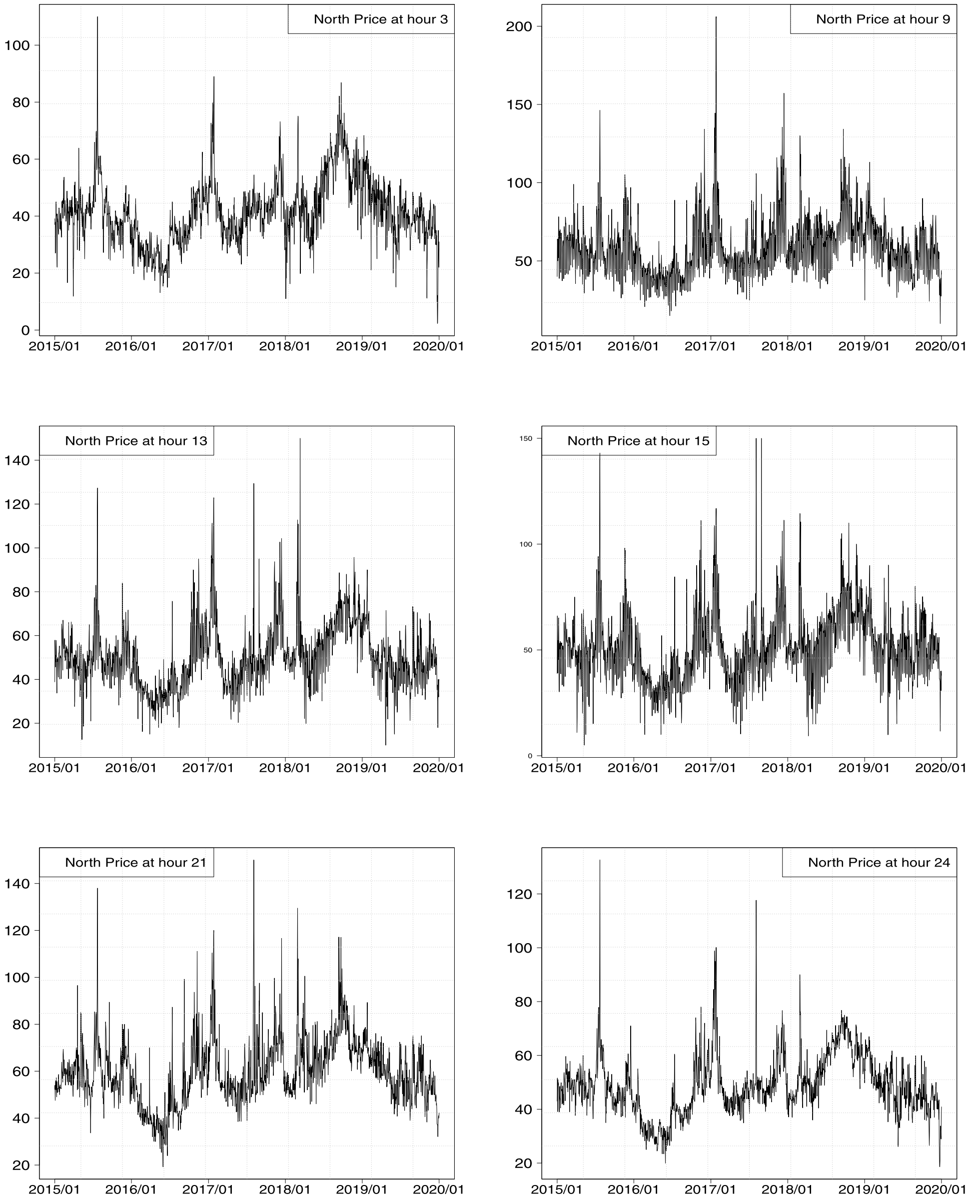

Fig. 2. Day-ahead electricity prices in northern Italy at hours 3, 9, 13, 15, 21, and 24.

omitted when the price at hour 24 is considered):

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-7} + \beta_4 y_{min,t-1} + \beta_5 y_{max,t-1} + \beta_6 y_{24,t-1} + \sum_{k=1}^{6} \gamma_k D_t^k + \varepsilon_t \quad (4)$$

and, similarly, we have model EX_3X augmented for regressors.

The last expert model EX_4 takes into account demeaned prices. Formally,

$$y_t = \alpha_0 + \alpha_1 \bar{y}_t^w + \sum_{k=1}^{8} \beta_k \left(y_{t-k} - \bar{y}_t^w\right) + \varepsilon_t \quad (5)$$

where $\bar{y}_t^w$ is the mean value of the (hourly) price over the week, and a possible dependency over the $k = 8$ past days is considered, as in Ziel and Weron (2018). Its augmented variant EX_4X is expanded by including daily dummies D_t^k (with $k = 1, 2, \ldots, 6$ for Mondays, Tuesdays, and so on to Saturdays) and all exogenous regressors.

Moving to the AR(FI)MA models, the first specification is an AR(7), a simple autoregressive process with seven lags given by the frequency of our data, and its variant AR(p) with lag length p estimated over a maximum length size of seven. Formally, our AR(p) models are defined as

$$y_t = \alpha + \sum_{k=1}^{4} \beta_s D_t^k + \sum_{j=1}^{11} \gamma_j M_t^j + \sum_{r=1}^{p} \phi_r y_{t-r} + \varepsilon_t \quad (6)$$

with D_t^k, differently from before, being dummies with $k = 1$ for Mondays, $k = 2$ for Saturdays, $k = 3$ for Sundays, and $k = 4$ for holidays not occurring on Saturdays or Sundays; M_t^j are dummies for months, with $j = 1, 2, \ldots, 11$ for January, February, ..., until November, excluding December. Monthly dummy variables are used to model calendar seasonality, whereas the $Monday_t$ dummy captures the impact of a change in consumption among working days and the first day after the weekend. ϕ_r with $r = 1, \ldots, p$ denotes the coefficients for the autoregressive terms, with p varying from 1 to 7. If p is fixed to 7, then we have the AR(7) process; by contrast, if p is estimated from the data, then we have the AR(p) process (details on the estimations are reported in the following section). We also consider their variants augmented for regressors, that is, ARX(7) and ARX(p).

Then, the autoregressive process is generalized to include moving average components and we consider the general ARMA models with p and q orders, fixed or again estimated from the data. The general formulation for an ARMA(p,q) is

$$y_t = \alpha + \sum_{k=1}^{4} \beta_s D_t^k + \sum_{j=1}^{11} \gamma_j M_t^j + \sum_{r=1}^{p} \phi_r y_{t-r} + \sum_{s=1}^{q} \theta_s \varepsilon_{t-s} + \varepsilon_t \quad (7)$$

with D_t^k dummies, with $k = 1$ for Mondays, $k = 2$ for Saturdays, $k = 3$ for Sundays, and $k = 4$ for holidays; θ_s with $s = 1, \ldots, q$ denotes the coefficients for the moving average terms, with q varying from 1 to 7; and again both are estimated, within a maximum range of 7, that is, $p_{max} = 7$ and $q_{max} = 7$. For comparisons, we have also considered several specifications of this general process with fixed values: the ARMA(7,7) with $p = q = 7$, the ARMA(7,1) with $p = 7$ and $q = 1$, and the ARMA(1,7) with $p = 1$ and $q = 7$. As for the other models, we include in our analysis all ARMAs augmented with exogenous regressors, then testing ARMAX(p,q), ARMAX(7,7), ARMAX(7,1), and ARMAX(7,1).

To account for long memory, we finally consider the autoregressive fractionally integrated moving-average, or ARFIMA(p, d, q) models, defined as

$$\Phi(L)(1-L)^d (y_t - \mu_t) = \Theta(L)\varepsilon_t \quad \text{with } \varepsilon_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}\left(0, \sigma^2\right) \quad (8)$$

where the normal distribution of the errors has a constant variance, σ^2 $\forall t$. d is the fractional integration parameter (with $0 < d < 0.5$), and μ_t is defined as

$$\mu_t = \mu + \sum_{k=1}^{4} \beta_s D_t^k + \sum_{j=1}^{11} \gamma_j M_t^j \quad (9)$$

with D_t^k dummies, with $k = 1$ for Mondays, $k = 2$ for Saturdays, $k = 3$ for Sundays, and $k = 4$ for holidays; and monthly dummies, M_t^j. As in the ARMA models, we set the p, d, and q to be estimated within a range of $p_{max} = 7$, $d_{max} = 2$, and $q_{max} = 7$. We extend this model with our sets of exogenous regressors, obtaining the ARFIMAX(p, d, q) model, and we compare it with ARFIMAX variants with fixed values for p and q, leaving instead d free to change between 0, 1, and 2. Specifically, we include in our analysis the ARFIMAX(7,d,7) and the ARFIMAX(7,d,0).

Moreover, to account for possible time-varying volatility patterns, asymmetries, and shocks induced by fundamental drivers, we expand our models by including GARCH-type specifications. A similar approach has been used by, for example, Koopman et al. (2007), Huurman, Ravazzolo, and Zhou (2012), Paraschiv et al. (2014), Ketterer (2014), Jeon and Taylor (2016), and Laporta, Merlo, and Petrella (2018). Therefore, we follow consolidated and well-established modeling approaches.

In particular, when the Italian market is considered, Bosco, Parisio, and Pelagatti (2007) used an ARMA-GARCH model, whereas Gianfreda and Grossi (2012) used ARFIMAX-GARCHX models.

Hence, we compare the performances of several AR(FI)MA models with their variants, including GARCH-type specifications, while allowing for an automatic selection of the length of autoregressive and moving average processes and the switching among models, when necessary. To this aim, the considered GARCH specifications are: the standard GARCH (SGARCH), the exponential GARCH (EGARCH), the threshold GARCH (TGARCH), and finally the GARCH-in-mean (GARCH-M), all with normal distribution.

These models differ according to the type of GARCH adopted. Thus, the second set of models extends the previous one with time-varying volatility expressed without loss of generality on day t as $\sigma_t^2 = \mathbb{V}(\varepsilon_t \mid \mathcal{F}_{t-1}) = Var(\varepsilon_t \mid \mathcal{F}_{t-1})$. These models are detailed in Appendix A.

In particular, the EGARCH(1,1) allows the conditional variance process to respond asymmetrically to rises and falls in electricity prices (Nelson, 1991). To account for asymmetries in volatility, making it a function of positive and negative values of the innovations, we also consider the TGARCH(1,1) process (Zakoian, 1994). Moreover, to consider the possibility that price levels may be influenced by their past price variability and by the fact that volatility in electricity prices is generally stronger when prices are high, we include the standard deviation, as obtained from the conditional variance equation, in the conditional mean equation, adopting the GARCH(1,1)-in-mean (or, simply, GARCH-M), as in Kyritsis, Andersson, and Serletis (2017) and Gianfreda and Scandolo (2018). These GARCH specifications are expanded to include the exogenous regressors $\mathbf{x}_t$ and $\mathbf{z}_{t-1}$, following the evidence in Huurman et al. (2012) that fundamental drivers improve accuracy when the volatility equation is also included.

Finally, we estimate the LASSO of further autoregressive models with 28 lags to account for changing market conditions in the last four weeks; their augmented specifications for regressors, that is, $AR(28)_{LASSO}$ and $ARX(28)_{LASSO}$; and also the formulation including the time-varying volatility, that is, the ARX(28)-GARCHX(1,1)-M_{LASSO}.

To summarize, our model set contains 58 models divided in five groups: (i) four *expert* models (EX_1, EX_2, EX_3, and EX_4), their extensions with fundamental drivers (EX_1X, EX_2X, EX_3X, and EX_4X), and the EX_4X extended with the time-varying volatility (that is, EX_4X-SGARCHX, EX_4X-EGARCHX, EX_4X-TGARCHX, and EX_4X-GARCHX-M); (ii) autoregressive AR and ARX models with the order p estimated for the AR(p) and ARX(p), or a priori fixed for the AR(7) with the ARX(p) and ARX(7) extended with time-varying volatility; (iii) ARMA and ARMAX models, where AR and MA lags are estimated or fixed (ARMA(p,q), ARMA(7,7), ARMA(1,7), and ARMA(7,1)), their extensions for regressors (ARMAX(p,q), ARMAX(7,7), ARMAX(1,7), and ARMAX(7,1)), and their ARMA(p,q) and ARMA(7,7) with time-varying volatility; (iv) ARFIMA and ARFIMAX models, where the AR and MA lags and the fractional integration order are estimated or fixed (ARFIMA(p,d,q), ARFIMAX(p,d,q), ARFIMAX(7,d,7), and ARFIMAX(7,d,0)), and their extensions with the time-varying volatility; (v) the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996) for the AR and ARX models with up to 28 lags and their extension with GARCH-in-mean volatility.

3.3. Estimation methods and iterative optimization procedures

The iterative procedure adopted for the selection of model ordering allows us to adapt the price structure to the changing market conditions, as the increasing RES shares in generation, or changes in import/export flows due to additional interconnections, or more generally to agent learning, regulatory and market structural changes. However, to account for possible bias in day-ahead predictions induced by the iterative ordering selection, we compare the iterative models with the ones with ex ante and pre-determined orders. Let us now describe the iterative model selection process while defining the estimation and optimization procedures.

Iterative model selection is essentially a two-step estimation procedure. In the first step the autoregressive and moving average orders p and q, and (eventually) the fractionally integrated parameter d, are estimated through a grid search process by finding the best model according to the corrected AIC value (AICc), a modification of the original AIC for small sample sizes. The maximum values of the orders are set to seven in order to consider the seven-day-per-week frequency of our data, so $p_{max} = 7$ and $q_{max} = 7$, respectively, whereas the maximum value of the fractional integration parameter is $d_{max} = 2$.

The second step is then used for the ARFIMAX(p, d, q) and the ARMAX(p, q)-GARCH models, both with exogenous regressors. In the former cases, the estimated orders $\left(\hat{p}, \hat{q}\right)$ enter in the ARFIMAX process, and then the fractional integration parameter d is estimated simultaneously with the other parameters of interest. In the latter case, the estimated orders $\left(\hat{p}, \hat{q}\right)$ enter in the ARMAX(p, q)-GARCH processes, and the GARCH orders are estimated simultaneously with the other parameters. For both, we found the *nloptr* nonlinear optimization R package (Johnson, 2021) to be suitable for estimating these types of models. Due to convergence problems in some specific cases, and in order to ensure the invertibility of the processes, we changed the numeric tolerance of the solver, or, alternatively, tried a combination of other solvers. Finally, the model parameters are all estimated by maximum likelihood.

Models with fixed orders are instead estimated without any adaptive scheme, due to the ex ante pre-determined specified orders. In these situations, the estimation procedure is based on the conditional sum of squares to find the starting values, and then on the maximum likelihood, with the use of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm for optimization; see Broyden (1970), Fletcher (1970), Goldfarb (1970), and Shanno (1970). As before, if convergence problems occur, the procedure allows us to fit the model via maximum likelihood and the optimization via a modification of the simulated annealing (SANN) of Bélisle (1992), which always guarantees convergence, even with non-differentiable functions, although it can be relatively slow.

As far as the LASSO is concerned, we proceed in a way that the relevant exogenous regressors $\mathbf{x}_t$ and $\mathbf{z}_{t-1}$, combined with the autoregressive terms up to 28 earlier periods, i.e. $\mathbf{y}_{t-h}$ with $h = 1, \ldots, 28$, are selected by considering a simple linear model. In this way, we are able to properly define a potential subgroup of regressors and autoregressive terms selected at each iteration and for each hour. The criterion used for the statistical bias–variance tradeoff, which determines the tuning/penalty parameter, is the standard cross-validation (cv) that minimizes the average error.

All computations were executed using the software R and using an AMD EPYC 7542 32-core 2.90 GHz processor.

3.4. Assessment of the forecasting performance

We compare different model specifications for modeling and forecasting the electricity zonal prices observed over individual hours: each hour is modeled separately by following a daily frequency for prices and drivers. Because all information is available or reconstructed at approximately 11 a.m. (i.e. before the market closure when traders must submit their offers), we are able to model all 24 h and forecast them for the next day by a simple prediction process that produces a set of 24 price forecasts for the 24 h of the following day.

We use the first 730 days of our dataset (i.e. from 1/1/2015 to 31/12/2016) for the in-sample estimation, and then the first out-of-sample prediction is obtained for 1/1/2017. Thereafter, the window is rolled one step ahead with further estimation and forecasts obtained for 2/1/2017, and so forth, until the last observation in the sample. Therefore, we produce forecasts over three years from 1/1/2017 to 31/12/2019 using the ENTSO-E forecasted data. We recall that the modeling and forecasting process is undertaken on day t to provide a set of 24 hourly prices forecasted for the next day $t + 1$. These forecasts must be submitted before the closure of the market, i.e. before noon on day t (thus, we assume that these models must complete their runs before noon). To predict the day-ahead hourly price on day $t + 1$, we use the information referred to that specific hour as follows: we assume that market operators submit their bids by noon on day t, based on predicted prices for day $t + 1$, obtained by considering fuel prices determined on day $t - 1$; the forecasted values for RES and zonal load available on day t; the hydro generation, weighted imports, biomass, and waste observed on day $t - 1$ for hours 1–10; and their realized values observed at hour 10 on the day t for modeling and forecasting the electricity prices of hours 11–24 on day $t + 1$. Indeed, Maciejowska and Nowotarski (2016) and Ziel (2016) note and suggest that prices for early morning hours depend more on the latest information than on information contained at the same hour but on the previous day.

To assess the forecasting performance of implemented models, we use both point and density metrics, as the root mean square error (RMSE) and the continuous ranked probability score (CRPS); see for example Gneiting and Ranjan (2011) and Groen, Paap, and Ravazzolo (2013) for early applications in economics, and Gianfreda et al. (2020) for an application to Italian electricity prices. In addition, we implement the one-sided Diebold–Mariano (DM) test to judge the superiority between two competing models (see Diebold & Mariano, 1995 and West, 1996), and the Hansen–Luden–Nason procedure of the model confidence set (MCS) to verify the statistical significance in terms of differences in forecasting performances among the selected models (Hansen et al., 2011). The DM test compares the forecast residuals of only two competing models, and the MCS procedure is a sequence of statistical tests in which the null hypothesis is built on the equal predictive ability (EPA) of several model specifications. Given that the EPA statistical tests can be calculated for different loss functions (depending on the aim of the comparison), we consider a loss function for level forecasts because of our interest in a comparison of the predictability power in the mean between our models. We also consider a comparison in terms of the full density forecasts by applying the DM and MCS tests to the CRPS metrics.

Finally, we compare the forecasting ability of the best performing model when the RTR professional forecasts for consumption, wind, and solar replace the ones provided by ENTSO-E. In this latter analysis, the in-sample estimation considers only 365 observations for 2018 and produces forecasts for the whole of 2019, because of the reduced size of the RTR Italian regional forecasts.

4. Results

To judge the quality of the forecasted prices, RMSEs and CRPSs are computed and presented in Tables 3 and 4, respectively. We include the superior set of models and the DM tests. These results refer to hours 3, 9, 11, 13, 19, and 21, to the average metrics computed over 24 h (Avg_{1-24}), and over the peak hours 8–20 (Avg_{8-20}). Results for other hours are omitted but are available on request.

Firstly, we observe that the inclusion of exogenous regressors reduces both the RMSEs and the CRPSs, especially during peak hours. Therefore, we extend the empirical evidence in Gianfreda et al. (2020) on the predictive power of a large set of exogenous regressors to forecast, this time, regional prices, whereas single national prices were forecasted in the cited reference.

Considering the whole set of 58 models, it can be easily observed that the expert EX_4X model drastically outperforms all other models in point forecasts, with the lowest average errors of around 7 and 6 €/MWh over peak and base periods, respectively. These results clearly declare the EX_4X model as the superior specification in point forecasting. Interestingly, including time-varying variance does not significantly increase the point forecasting accuracy of these Italian zonal prices; however, in general, some potential improvements can be expected, as discussed in Ziel et al. (2015). The other AR(FI)MA models provide higher and similar errors, even if they differ in their structures.

Notably, the forecasting precision drastically decreases during the ramp-up and ramp-down phases (hours 9 and 19), when conventional thermal generation is necessary to restore the balance between demand and supply. Across peak hours, the non-programmable renewables (especially solar and wind) bid at 0 €/MWh and have priority of dispatch of the produced energy. Therefore, their intermittent, erratic in-feed increases the variability of prices and consequently affects the forecasting errors, especially when demand is at higher levels (at hours 9 and 19).

Furthermore, the predictability power of fundamental variables decreases during the evening hours because the forecast horizons are longer than those for the morning hours. This argument is particularly notable for RES because the accuracy of weather predictions decreases substantially with the length of forecasting horizons.

Table 3
RMSEs of all selected models and for a section of hours.

Model	3	9	11	13	19	21	Avg_{8-20}	Avg_{1-24}
EX_1	6.468	12.501	10.822	9.834	11.836	8.911	10.973	9.150
EX_2	6.437	12.162	10.533	9.636	11.748	8.809	10.760	9.016
EX_3	5.984	11.960	10.309	9.411	11.804	8.756	10.616	8.824
EX_4	5.204	9.469	8.284	7.998	8.365	6.643	8.516	7.074
EX_1X	6.151	11.691	10.196	9.144	11.169	8.614	10.303	8.644
EX_2X	6.113	11.563	10.062	9.062	11.175	8.636	10.237	8.602
EX_3X	5.814	11.378	9.864	8.895	11.237	8.756	10.131	8.499
EX_4X	4.860	8.691	7.708	7.318	8.106	6.466	7.871	6.615
EX_4X-SGARCHX	4.861	8.834	7.712	7.274	8.284	6.449	7.919	6.628
EX_4X-EGARCHX	4.912	8.790	7.805	7.230	8.207	6.473	7.907	6.626
EX_4X-TGARCHX	4.894	8.835	7.791	7.286	8.401	6.595	7.992	6.689
EX_4X-GARCHX-M	4.925	8.846	7.788	7.313	8.264	6.420	7.976	6.662
AR(7)	5.577	11.500	9.867	9.228	10.272	8.122	10.130	8.327
AR(p)	5.529	11.977	10.170	9.545	10.521	8.223	10.582	8.596
ARX(7)	5.489	10.706	9.162	8.521	9.932	7.837	9.412	7.847
ARX(p)	5.449	11.000	9.428	8.739	10.135	7.871	9.703	8.024
ARX(7)-SGARCHX	5.462	10.722	9.177	8.337	9.739	7.689	9.367	7.795
ARX(7)-EGARCHX	5.462	10.873	9.225	8.455	9.712	7.786	9.395	7.826
ARX(7)-TGARCHX	5.506	10.596	9.208	8.489	10.048	7.745	9.472	7.871
ARX(7)-GARCHX-M	5.470	10.692	9.184	8.348	9.804	7.689	9.405	7.843
ARX(p)-SGARCHX	5.436	10.924	9.404	8.591	10.088	7.777	9.736	8.034
ARX(p)-EGARCHX	5.446	11.074	9.536	8.618	10.021	7.899	9.720	8.030
ARX(p)-TGARCHX	5.498	11.001	9.484	8.979	10.309	7.837	9.932	8.164
ARX(p)-GARCHX-M	5.468	11.025	9.571	8.603	10.091	7.813	9.824	8.114
ARMA(7,7)	5.717	13.542	10.430	9.402	10.629	9.974	10.615	8.934
ARMA(1,7)	5.589	11.808	9.942	9.362	10.449	8.192	10.309	8.447
ARMA(7,1)	5.584	11.483	9.821	9.191	10.310	8.105	10.091	8.300
ARMA(p,q)	5.561	11.805	10.066	9.429	10.520	8.160	10.450	8.516
ARMAX(7,7)	5.533	10.418	9.197	8.493	9.950	7.685	9.339	7.824
ARMAX(1,7)	5.518	10.770	9.176	8.571	10.003	7.819	9.444	7.873
ARMAX(7,1)	5.494	10.710	9.136	8.498	9.968	7.832	9.394	7.836
ARMAX(p,q)	5.467	10.959	9.353	8.660	10.129	7.827	9.630	7.975
ARMAX(7,7)-SGARCHX	5.693	10.562	10.574	8.461	9.831	7.893	9.708	8.044
ARMAX(7,7)-EGARCHX	5.578	12.429	9.732	8.662	9.827	8.050	9.702	8.045
ARMAX(7,7)-TGARCHX	5.624	10.768	9.261	8.762	10.003	7.752	9.495	7.910
ARMAX(7,7)-GARCHX-M	5.689	10.746	9.228	8.530	9.959	7.719	9.519	8.135
ARMAX(p,q)-SGARCHX	5.453	10.867	9.314	8.649	10.060	7.737	9.625	7.959
ARMAX(p,q)-EGARCHX	5.456	10.875	9.376	8.522	10.138	7.779	9.626	8.009
ARMAX(p,q)-TGARCHX	5.507	10.959	9.248	8.638	10.303	7.822	9.781	8.060
ARMAX(p,q)-GARCHX-M	5.501	11.157	9.526	8.766	10.112	7.756	9.991	8.195
ARFIMA(p,d,q)	5.572	11.261	9.827	9.267	10.054	8.223	10.063	8.289
ARFIMAX(p,d,q)	5.467	10.959	9.353	8.661	10.121	7.827	9.630	7.975
ARFIMAX(p,d,q)-SGARCHX	5.459	10.847	9.303	8.620	10.044	7.762	9.617	7.947
ARFIMAX(7,d,7)-SGARCHX	5.604	10.782	9.134	8.462	10.064	7.799	9.513	7.923
ARFIMAX(7,d,0)-SGARCHX	5.455	10.806	9.143	8.317	9.895	7.698	9.385	7.804
ARFIMAX(p,d,q)-EGARCHX	5.468	10.807	9.482	8.678	10.132	7.745	9.653	7.975
ARFIMAX(7,d,7)-EGARCHX	5.763	10.655	9.828	10.534	10.007	8.823	9.703	8.110
ARFIMAX(7,d,0)-EGARCHX	5.458	10.727	9.064	8.473	9.775	7.749	9.354	7.796
ARFIMAX(p,d,q)-TGARCHX	5.480	11.010	9.247	8.543	10.213	7.782	9.702	8.007
ARFIMAX(7,d,7)-TGARCHX	5.588	10.447	9.217	8.664	9.895	7.780	9.424	7.884
ARFIMAX(7,d,0)-TGARCHX	5.506	10.671	9.110	8.383	10.009	7.689	9.400	7.833
ARFIMAX(p,d,q)-GARCHX-M	5.480	10.999	9.424	8.671	10.178	7.839	10.068	8.262
ARFIMAX(7,d,7)-GARCHX-M	6.586	10.883	9.334	8.595	10.209	7.962	9.913	8.402
ARFIMAX(7,d,0)-GARCHX-M	5.480	10.782	9.084	8.289	10.021	7.714	9.436	7.852
$AR(28)_{LASSO}$	6.415	12.395	10.802	9.960	11.462	9.007	10.914	9.117
$ARX(28)_{LASSO}$	6.197	11.736	10.094	9.210	11.051	8.649	10.259	8.636
AR(28)-GARCH-M_{LASSO}	6.394	12.614	10.845	10.045	11.859	9.171	11.143	9.275
ARX(28)-GARCHX-M_{LASSO}	6.295	11.577	10.258	9.195	11.060	8.678	10.361	8.721

The average over 24 h and the average over peak hours 8–20 are also included. Grey cells refer to specifications excluded from the superior set of models selected according to the Hansen–Luden–Nason MCS procedure at $\alpha = 0.15$. ***, **, and * indicate that a model is more accurate than the EX_4X benchmark model at the 0.1%, 1%, and 5% significance levels, respectively, according to the one-sided DM test. The absence of stars indicates that none of the alternative specifications provides more accurate forecasts.

There are no substantial improvements when LASSO models are considered. Therefore, based on this evidence and on previous explorations,[7] we conclude that the

[7] In previous analyses on LASSO specifications, we note that on average LASSOs perform better when considering the simultaneous selection of the autoregressive terms with the exogenous regressors, revealing that not all the lagged terms are useful at each iteration. Moreover, including exogenous regressors in both the conditional mean and conditional variance does not improve on average the power predictability of the same model.

Table 4
CRPSs of all selected models and for a section of hours.

Model	3	9	11	13	19	21	Avg_{8-20}	Avg_{1-24}
EX_1	0.255	0.332	0.317	0.298	0.323	0.313	0.317	0.298
EX_2	0.254	0.332	0.318	0.299	0.323	0.313	0.317	0.298
EX_3	0.253	0.333	0.318	0.300	0.324	0.313	0.318	0.298
EX_4	0.252	0.323	0.309	0.291	0.314	0.309	0.309	0.292
EX_1X	0.254	0.327	0.312	0.294	0.319	0.314	0.312	0.295
EX_2X	0.253	0.327	0.312	0.294	0.319	0.314	0.312	0.295
EX_3X	0.253	0.327	0.312	0.294	0.320	0.313	0.313	0.295
EX_4X	0.251	0.320	0.307	0.289	0.313	0.308	0.307	0.290
EX_4X-SGARCHX	0.125***	0.199***	0.174***	0.156***	0.178***	0.166***	0.178	0.158
EX_4X-EGARCHX	0.125***	0.198***	0.173***	0.155***	0.180***	0.164***	0.178	0.158
EX_4X-TGARCHX	0.124***	0.199***	0.174***	0.155***	0.178***	0.167***	0.179	0.158
EX_4X-GARCHX-M	0.125***	0.201***	0.173***	0.155***	0.178***	0.164***	0.179	0.158
AR(7)	0.254	0.326	0.312	0.294	0.317	0.312	0.312	0.295
AR(p)	0.252	0.324	0.310	0.292	0.314	0.309	0.310	0.292
ARX(7)	0.252	0.323	0.309	0.292	0.315	0.310	0.309	0.292
ARX(p)	0.252	0.324	0.310	0.292	0.314	0.310	0.310	0.293
ARX(7)-SGARCHX	0.124***	0.201***	0.176***	0.158***	0.177***	0.163***	0.179	0.158
ARX(7)-EGARCHX	0.124***	0.205***	0.179***	0.159***	0.177***	0.166***	0.181	0.159
ARX(7)-TGARCHX	0.124***	0.200***	0.175***	0.154***	0.179***	0.162***	0.178	0.158
ARX(7)-GARCHX-M	0.124***	0.201***	0.175***	0.158***	0.177***	0.163***	0.179	0.158
ARX(p)-SGARCHX	0.124***	0.200***	0.174***	0.153***	0.176***	0.163***	0.177	0.157
ARX(p)-EGARCHX	0.124***	0.208***	0.179***	0.156***	0.177***	0.166***	0.181	0.160
ARX(p)-TGARCHX	0.124***	0.201***	0.173***	0.149***	0.177***	0.161***	0.176	0.156
ARX(p)-GARCHX-M	0.124***	0.201***	0.175***	0.154***	0.176***	0.163***	0.178	0.158
ARMA(7,7)	0.264	0.341	0.326	0.308	0.328	0.324	0.326	0.307
ARMA(1,7)	0.264	0.338	0.324	0.307	0.328	0.322	0.324	0.306
ARMA(7,1)	0.264	0.338	0.324	0.307	0.328	0.322	0.324	0.306
ARMA(p,q)	0.262	0.337	0.323	0.305	0.327	0.320	0.323	0.304
ARMAX(7,7)	0.252	0.323	0.310	0.292	0.315	0.310	0.310	0.292
ARMAX(1,7)	0.261	0.335	0.321	0.303	0.326	0.319	0.321	0.303
ARMAX(7,1)	0.252	0.323	0.309	0.292	0.315	0.310	0.309	0.292
ARMAX(p,q)	0.261	0.335	0.321	0.303	0.326	0.319	0.321	0.303
ARMAX(7,7)-SGARCHX	0.125***	0.202***	0.179***	0.157***	0.180***	0.165***	0.180	0.159
ARMAX(7,7)-EGARCHX	0.126***	0.203***	0.183***	0.159***	0.179***	0.166***	0.182	0.160
ARMAX(7,7)-TGARCHX	0.125***	0.201***	0.178***	0.156***	0.180***	0.163***	0.180	0.159
ARMAX(7,7)-GARCHX-M	0.125***	0.202***	0.177***	0.158***	0.178***	0.164***	0.181	0.159
ARMAX(p,q)-SGARCHX	0.124***	0.202***	0.176***	0.157***	0.178***	0.163***	0.179	0.158
ARMAX(p,q)-EGARCHX	0.124***	0.206***	0.179***	0.157***	0.177***	0.166***	0.181	0.159
ARMAX(p,q)-TGARCHX	0.124***	0.201***	0.175***	0.152***	0.179***	0.161***	0.178	0.157
ARMAX(p,q)-GARCHX-M	0.124***	0.202***	0.176***	0.158***	0.177***	0.163***	0.179	0.158
ARFIMA(p,d,q)	0.263	0.339	0.325	0.307	0.328	0.321	0.325	0.306
ARFIMAX(p,d,q)	0.261	0.335	0.321	0.303	0.326	0.319	0.321	0.303
ARFIMAX(p,d,q)-SGARCHX	0.124***	0.201***	0.175***	0.157***	0.178***	0.164***	0.178	0.158
ARFIMAX(7,d,7)-SGARCHX	0.125***	0.202***	0.178***	0.158***	0.179***	0.164***	0.181	0.160
ARFIMAX(7,d,0)-SGARCHX	0.125***	0.201***	0.176***	0.158***	0.177***	0.163***	0.179	0.158
ARFIMAX(p,d,q)-EGARCHX	0.125***	0.206***	0.179***	0.158***	0.177***	0.166***	0.181	0.160
ARFIMAX(7,d,7)-EGARCHX	0.126***	0.204***	0.182***	0.161***	0.179***	0.167***	0.183	0.161
ARFIMAX(7,d,0)-EGARCHX	0.125***	0.205***	0.179***	0.159***	0.178***	0.167***	0.181	0.160
ARFIMAX(p,d,q)-TGARCHX	0.124***	0.200***	0.176***	0.152***	0.179***	0.162***	0.178	0.157
ARFIMAX(7,d,7)-TGARCHX	0.125***	0.200***	0.178***	0.157***	0.179***	0.164***	0.180	0.159
ARFIMAX(7,d,0)-TGARCHX	0.124***	0.200***	0.176***	0.154***	0.179***	0.162***	0.178	0.158
ARFIMAX(p,d,q)-GARCHX-M	0.124***	0.202***	0.175***	0.157***	0.177***	0.163***	0.178	0.158
ARFIMAX(7,d,7)-GARCHX-M	0.126***	0.203***	0.177***	0.158***	0.178***	0.165***	0.180	0.160
ARFIMAX(7,d,0)-GARCHX-M	0.124***	0.201***	0.176***	0.158***	0.177***	0.164***	0.179	0.158
$AR(28)_{LASSO}$	0.254	0.327	0.313	0.294	0.317	0.312	0.312	0.295
$ARX(28)_{LASSO}$	0.256	0.327	0.311	0.294	0.318	0.313	0.312	0.295
AR(28)-GARCH-M_{LASSO}	0.124***	0.201***	0.176***	0.158***	0.181***	0.163***	0.179	0.158
ARX(28)-GARCHX-M_{LASSO}	0.125***	0.200***	0.173***	0.153***	0.179***	0.166***	0.178	0.158

The average over 24 h and the average over peak hours 8–20 are also included. Grey cells refer to specifications excluded from the superior set of models selected according to the Hansen–Luden–Nason MCS procedure at α = 0.15. ***, **, and * indicate that a model is more accurate than the EX_4X benchmark model at the 0.1%, 1%, and 5% significance levels, respectively, according to the one-sided DM test. The absence of stars indicates that the alternative specification does not provide more accurate forecasts.

Table 5
RMSEs of the best performing model with ENTSO-E forecasts (EX_4X and EX_4X-SGARCHX-norm) and with ETR forecasts for *fast* (EX_4X-F and EX_4X-SGARCHX-norm-F) and *less fast* (EX_4X-LF and EX_4X-SGARCHX-norm-LF) models over 365 forecasts computed for the whole of 2019 for a selection of hours.

	3	9	11	13	19	21	Avg_{1-24}	Avg_{8-20}
EX_4X	5.082	9.782	7.515	6.888	6.719	5.105	6.605	7.814
EX_4X-F	4.629*	6.562***	6.325***	5.681***	5.354***	4.668**	5.264	5.938
EX_4X-LF	4.624*	6.553***	6.329***	5.694***	5.357***	4.673**	5.266	5.941
EX_4X-SGARCHX	5.090	9.939	7.663	7.074	6.739	5.269	6.745	8.031
EX_4X-SGARCHX-F	4.592**	6.573***	6.273***	5.525***	5.400***	4.645**	5.302	5.932
EX_4X-SGARCHX-LF	4.580**	6.593***	6.307***	5.597***	5.362***	4.647**	5.303	5.937

The average over 24 h and the average over peak hours 8–20 are also included. Grey cells refer to specifications excluded from the superior set of models selected according to the Hansen–Luden–Nason MCS procedure at α = 0.15. ***, **, and * indicate that a model is more accurate than the EX_4X benchmark model at the 0.1%, 1%, and 5% significance levels, respectively, according to the one-sided DM test. The absence of stars indicates that the alternative specification does not provide more accurate forecasts.

Table 6
CRPSs of the best performing model with ENTSO-E forecasts (EX_4X and EX_4X-SGARCHX-norm) and with ETR forecasts for *fast* (EX_4X-F and EX_4X-SGARCHX-norm-F) and *less fast* (EX_4X-LF and EX_4X-SGARCHX-norm-LF) models, over 365 forecasts computed for the whole 2019 for a selection of hours.

	3	9	11	13	19	21	Avg_{1-24}	Avg_{8-20}
EX_4X	0.239	0.310	0.301	0.280	0.293	0.295	0.278	0.296
EX_4X-F	0.269	0.297***	0.295***	0.307***	0.313***	0.301***	0.285	0.299
EX_4X-LF	0.269	0.297***	0.295***	0.307***	0.313***	0.301***	0.285	0.299
EX_4X-SGARCHX	0.103***	0.179***	0.164***	0.140***	0.144***	0.134***	0.137	0.159
EX_4X-SGARCHX-F	0.098***	0.128***	0.113***	0.117***	0.141***	0.121***	0.112	0.125
EX_4X-SGARCHX-LF	0.099***	0.126***	0.112***	0.117***	0.144***	0.122***	0.112	0.125

The average over the 24 h and the average over peak hours 8-20 are also included. Grey cells refer to specifications excluded from the Superior Set of Models selected according to the Hansen-Luden-Nason MCS procedure at α = 0.15. ***, ** and * indicate that a model is more accurate than the EX_4X benchmark model at the 0.1%, 1%, 5% significance levels according to the one-sided DM test. Absence of stars indicates that the alternative specification does not provide more accurate forecasts.

LASSO is not necessary to improve accuracy in our context, characterized by a limited number of regressors with respect to the amount of statistical information available.

Indeed, when all models are simultaneously compared, the computations of the superior set of models,[8] in terms of minimum loss function for level forecasts, show that the LASSO models are always discarded. Moreover, none of the models provides more accurate forecasts than those of the EX_4X, which is considered as the benchmark in the one-sided DM tests (under the alternative hypothesis that any other model is more accurate than the EX_4X). This model is always retained in the superior set of models and the DM tests confirm its out-performance in pairwise comparisons. More importantly from a practical point of view, this expert model and its GARCH variants are the only ones retained for all hours in the MCS, and especially when hour 19 is considered. Hence, market operators willing to adopt a single model to forecast all hourly prices should consider this relevant and clearly assessed fact.

For completeness, we extend the analysis to density forecasting and investigate whether a more general loss function provides different evidence. Looking at Table 4, the results on CRPS show that there are substantial improvements when all models are enlarged to include GARCH time-varying volatility. Indeed, the average of CRPS over the 24 h of all the models with time-varying volatility is in the range of 0.156–0.160, whereas the same average for models without time-varying volatility is around 0.3. Specifications for only the conditional mean are always excluded from the MCS, apart from the $AR(28)_{LASSO}$, which, however, provides forecasts that are not statistically more accurate than the benchmark EX_4X in the DM tests. The expert models augmented with time-varying volatility are the only (class of) models never excluded from the MCS. Moreover, the DM tests show that, in general, all GARCH specifications are statistically superior to the benchmark, confirming the importance of including time-varying volatility. Therefore, when the loss function is generalized to the full distribution, sophisticated specifications that allow for time-varying volatility are essential to improve the forecast accuracy.

Given the focus on forecasting, we compare the forecasting performance of the EX_4X and EX_4X-SGARCHX models when professional and more timely forecasts are used in place of public and freely available forecasts. The RMSEs and CRPSs for a selection of hours and averages over base and peak hours are presented in Tables 5 and 6. However, to show the full performance of these models, we decided to report the results for all 24 h in Tables B.9 and B.10 in Appendix B.

We compare the forecasting performances of EX_4X when ENTSO-E forecasts are considered, with those obtained by the same model when RTR forecasts are used instead. As anticipated, these professional forecasts are released more often and represent the best updated information available at 6.55 a.m. and at 7.40 a.m. when market operators can start running their forecasting models to formulate their day-ahead bidding strategy. Then,

[8] We implement the MCS procedure with the $T_{max,\mathcal{M}}$ test (Hansen et al., 2011, p. 465) at the $\alpha = 0.15$ significance level by using the R function `MCSprocedure` within the package `MCS` written by Bernardi and Catania (2018).

we distinguish between models which can run quickly (and their forecasts are labeled with F for *fast*) from those running less quickly (hence labeled with LF for *less fast*). In our case, we compare the forecasting performances of EX_4X-F and EX_4X-LF with EX_4X and the ones from EX_4X-SGARCHX-F and EX_4X-SGARCHX-LF with EX_4X-SGARCHX. Also for this exercise, we studied other model specifications with professional forecasts, and the results were qualitative similar. We emphasize that the evidence is not driven by considering simple or complex models, but by the usage of professional forecasts.

The results clearly show that using professional forecasts substantially improves price forecasts, especially during hours 1–7 and peak hours 8–20. Then, as soon as the forecasting horizon increases, as after hour 21, the benefits of using professional forecasts disappear. Moreover, in the very short horizon up to hour 18, there is no difference between the two forecasting models running with the latest information: indeed *fast* and *less fast* models perform equivalently. They diverge when the fast model shows better (but small) gains at hour 19, before losing any forecasting power as soon as the forecasting horizon further increases to hours 21–24. Therefore, these results emphasize the importance of implementing forecasting models with accurate and professionally computed forecasts; and, if possible, traders should wait for the latest published forecasts to take longer benefits of the forecasting gain. Even in this case, GARCH specifications do not substantially improve the point forecasts, whereas the opposite occurs for the density forecasts. The take-home message is that traders and market operators are encouraged to use models accounting for higher moments of the distributions, as suggested in Gianfreda and Bunn (2018), while considering professional forecasts.

Finally, in what follows, we discuss the estimated coefficients (with confidence intervals at 90%) of the EX_4X model. The results generally refer to hours 3, 9, 15, and 21 in the out-of-sample period. However, some additional hours are considered with respect to the intradaily profiles of drivers, and the results for the remaining hours are omitted but are available upon request.

Consistently with the literature, forecasted load is statistically significant with a positive effect on day-ahead price, meaning that prices do respond to load, as shown in Figure C.4 in Appendix C. However, for hour 3, we document an increasing influence in 2018, then decreasing in 2019. Hours 9 and 15 show different dynamics with an almost constant influence until the end of 2018 but a substantial lower and progressively decreasing influencing power over the whole of 2019, which may reflect the *negative demand* effect played by solar PV generation according to its generation and new additions. By contrast, hour 21 exhibits a decreasing influence already from July 2017, probably for more conventional power available to cover demand (and so being less at the margin).

The estimated coefficients for solar PV forecasts are depicted in Figure C.5, and it shows that it is statistically significant at hours 13 and 15 with a negative sign, implying the reduction of the mean level of zonal prices. At hour 9 or 11 when sun starts to shine, it turns from significant to non-significant from the last part of 2017 or the middle of 2018 and through all the other years. Notwithstanding its limited generation in northern Italy, forecasted wind is found to have a significant negative effect at hour 3, whereas its effect turns from significant to non-significant from roughly the beginning of 2018 at hours 9 and 21. Instead, it is almost never significant at hour 15; see Figure C.6.

Looking at hydro and its intradaily profile, we were expecting significant (negative) effects at hours 9 and 21 when its generation is at its maximum. However, the estimated coefficients for actual hydropower generation were not found to be statistically significant at these hours. Figure C.7 shows that it is statistically significant only at hour 3, when it does not suffer from the competition of solar PV (and wind to a lesser extent).

As far as weighted imports are considered, they are not found to be significant (as reported at hours 3, 9, 15 and 21); see Figure C.10. Therefore, foreign prices and imported quantities do not seem to affect northern Italian electricity prices via scheduled capacity on interconnectors. The same conclusion is drawn for biomass and waste. Coal is instead found to be significant only at hour 15 and up to the beginning of 2018, and then it turned out to be misplaced by the progressive penetration of RES. Figure C.12 shows that natural gas confirms its attitude to increase electricity prices at hour 3. This finding is surprising considering the relevant share of electricity generation covered by combined cycle gas turbine plants in northern Italy. Similarly, CO_2 emission prices almost never exhibit a significant effect; see Figure C.13.

However, it must be noted that these conclusions on the dynamics of coefficients for exogenous regressors are based on a model accounting for the dependence of prices over the previous eight demeaned prices. Thus, the model seems 'expert' enough with the inclusion of past essential information together with the contribution of load, wind, solar, hydro, and natural gas.

5. Conclusions

Forecasting day-ahead electricity prices has become extremely important for generation planning, given the imperfect predictability of weather conditions that affect both demand and RES generation, and for trading decisions influenced by the exploitation of possible arbitrage opportunities that can occur in subsequent market sessions. Hence, this paper provided a comparison of expert and AR(FI)MA models with GARCH specifications with fixed or estimated structures through a flexible model selection by an iterative and adaptive procedure. The results showed that the best performing model is an expert one augmented for exogenous regressors and time-varying volatility, especially if density forecasting has to be assessed. The importance of producing good and timely predictions of hourly day-ahead prices for northern Italy was also tested against the usage of commercial forecasts, since monitoring the bidding strategies for detecting strategic behaviors across market sessions is becoming crucial to avoid market speculations and the consequent increasing costs for final customers.

Using a set of drivers, including forecasted demand, forecasted wind and solar PV generation, fossil fuels, and actual hydro, biomass, and waste generation together with price-weighted flows, northern Italian electricity prices were forecasted through linear and nonlinear models, some of them with a flexible structure iteratively selected at both the autoregressive and moving average orders over each calibration window, including the possibility of switching from one model to another one. Our results clearly showed that if point forecasts are of concern, a simple expert model overcomes all other specifications, and that adopting a flexible structure that changes with time-varying market conditions and avoids over-parametrization in an ex ante ordering selection performs equally well, although this is not recommended for all hours.

We provided evidence that fundamental factors can drive zonal electricity prices differently within trading periods and that their simultaneous inclusion (fuels, imports, and RES as well) substantially improves the forecast accuracy. However, when studying the density forecasting, only nonlinear models that allow for time-varying volatility and second-moment dynamics provided more accurate results. Finally, we found that using professional and more timely consumption and RES predictions improves the forecast accuracy of electricity prices more than using predictions freely available to researchers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the editor, the associate editor, and two anonymous reviewers for useful comments and suggestions which led us to improve the paper. The authors also thank seminar and conference participants at the 39th International Symposium on Forecasting in Thessaloniki. Special acknowledgements go to Giorgio Battisti and Diego Ganz for useful discussions and suggestions. Alperia Energy S.p.A. is acknowledged for funding this research project. In addition, Angelica Gianfreda wishes to acknowledge RTDcall2017 support for the FoMoPM project on *Forecasting and Monitoring electricity Prices, volumes and market Mechanisms* and RTDcall2018 support for the ERMUn project on *Energy Risk Modelling Under uncertainties*, funded by the Free University of Bozen-Bolzano and received when she was there.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2022.01.003.

References

Abramova, E., & Bunn, D. (2020). Forecasting the intra-day spread densities of electricity prices. *Energies*, *13*(3).

Afanasyev, D. O., & Fedorova, E. A. (2019). On the impact of outlier filtering on the electricity price forecasting accuracy. *Applied Energy*, *236*, 196–210.

Bélisle, C. J. (1992). Convergence theorems for a class of simulated annealing algorithms on Rd. *Journal of Applied Probability*, 885–895.

Bernardi, M., & Catania, L. (2018). The model confidence set package for R. *International Journal of Computational Economics and Econometrics*, *8*(2), 144–158.

Bernardi, M., & Lisi, F. (2020). Point and interval forecasting of zonal electricity prices and demand using heteroscedastic models: The IPEX case. *Energies*, *13*(23).

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

Bosco, B. P., Parisio, L. P., & Pelagatti, M. M. (2007). Deregulated wholesale electricity prices in Italy: an empirical analysis. *International Advances in Economic Research*, *13*(4), 415–432.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms: 2. The new algorithm. *IMA Journal of Applied Mathematics*, *6*(3), 222–231.

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

Bunn, D. W., Gianfreda, A., & Kermer, S. (2018). A trading-based evaluation of density forecasts in a real-time electricity market. *Energies*, *11*(10).

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

Conejo, A. J., Contreras, J., Espinola, R., & Plazas, M. A. (2005). Forecasting Electricity Prices for a day-ahead pool-based electricity energy market. *International Journal of Forecasting*, *21*(3), 435–462.

Cuaresma, J. C., Hlouskova, J., Kossmeier, S., & Obersteiner, M. (2004). Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy*, *77*(1), 87–106.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366a), 427–431.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, *13*(3), 317–322.

Garcia, R. C., Contreras, J., van Akkeren, M., & Garcia, J. (2005). A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems*, *20*(2), 867–874.

Gianfreda, A., & Bunn, D. (2018). A stochastic latent moment model for electricity price formation. *Operations Research*, *66*, 1189–1456.

Gianfreda, A., & Grossi, L. (2012). Forecasting Italian electricity zonal prices with exogenous variables. *Energy Economics*, *34*(6), 2228–2239.

Gianfreda, A., Parisio, L., & Pelagatti, M. (2018). A review of balancing costs in Italy before and after RES introduction. *Renewable and Sustainable Energy Reviews*, *91*, 549–563.

Gianfreda, A., Parisio, L., & Pelagatti, M. (2019). The RES-induced switching effect across fossil fuels: An analysis of day-ahead and balancing prices. *Energy Journal*, *40*.

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

Gianfreda, A., Ravazzolo, F., & Rossini, L. (2020). Comparing the forecasting performances of linear models for electricity prices with high RES penetration. *International Journal of Forecasting*, *36*(3), 974–986.

Gianfreda, A., & Scandolo, G. (2018). Measuring model risk in the european energy exchange, chapter 5. In *Handbook of Recent Advances in Commodity and Financial Modeling*. Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-61320-8_5.

Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold and quantile weighted scoring rules. *Journal of Business & Economic Statistics*, *29*, 411–422.

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, *24*(109), 23–26.

Gonçalves, C., Pinson, P., & Bessa, R. J. (2021). Towards data markets in renewable energy forecasting. *IEEE Transactions on Sustainable Energy*, *12*(1), 533–542.

Groen, J. J. J., Paap, R., & Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Stastistics*, *31*, 29–44.

Haldrup, N., & Nielsen, M. Ø. (2006). A regime switching long memory model for electricity prices. *Journal of Econometrics*, *135*(1–2), 349–376.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Hickey, E., Loomis, D. G., & Mohammadi, H. (2012). Forecasting hourly electricity prices using ARMAX–GARCH models: An application to MISO hubs. *Energy Economics*, *34*(1), 307–315.

Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, *30*(2), 357–363.

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

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

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163–172.

Jeon, J., & Taylor, J. W. (2016). Short-term density forecasting of wave energy using ARMA–GARCH models and kernel density estimation. *International Journal of Forecasting*, *32*(3), 991–1004.

Johnson, S. G. (2021). *The NLopt nonlinear-optimization package*: *Technical report*, http://github.com/stevengj/nlopt.

Karakatsani, N., & Bunn, D. (2008). Forecasting electricity prices: the impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, *24*(4), 764–785.

Karakatsani, N., & Bunn, D. W. (2010). Fundamental and behavioural drivers of electricity price volatility. *Studies in Nonlinear Dynamics & Econometrics*, *14*(4), 1–42.

Kath, C., Nitka, W., Serafin, T., Weron, T., Zaleski, P., & Weron, R. (2020). Balancing generation from renewable energy sources: Profitability of an energy trader. *Energies*, *13*(1).

Ketterer, J. C. (2014). The impact of wind power generation on the electricity price in Germany. *Energy Economics*, *44*, 270–280.

Kezunovic, M., Pinson, P., Obradovic, Z., Grijalva, S., Hong, T., & Bessa, R. (2020). Big data analytics for future electricity grids. *Electric Power Systems Research*, *189*, Article 106788.

Knittel, C. R., & Roberts, M. R. (2005). An empirical examination of restructured electricity prices. *Energy Economics*, *27*(5), 791–817.

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.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1–3), 159–178.

Kyritsis, E., Andersson, J., & Serletis, A. (2017). Electricity prices, large-scale renewable integration, and policy implications. *Energy Policy*, *101*, 550–560.

Laporta, A. G., Merlo, L., & Petrella, L. (2018). Selection of Value at Risk models for energy commodities. *Energy Economics*, *74*, 628–643.

Lisi, F., & Edoli, E. (2018). Analyzing and forecasting zonal imbalance signs in the Italian electricity market. *The Energy Journal*, *39*(5).

Maciejowska, K., Nitka, W., & Weron, T. (2021). Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices. *Energy Economics*, *99*(C).

Maciejowska, K., & Nowotarski, J. (2016). A hybrid model for GEFCom2014 probabilistic electricity price forecasting. *International Journal of Forecasting*, *32*(3), 1051–1056.

Maciejowska, K., & Weron, R. (2016). Short-and mid-term forecasting of baseload electricity prices in the UK: The impact of intra-day price relationships and market fundamentals. *IEEE Transactions on Power Systems*, *31*(2), 994–1005.

de Marcos, R. A., Bello, A., & Reneses, J. (2019). Electricity price forecasting in the short term hybridising fundamental and econometric modelling. *Electric Power Systems Research*, *167*, 240–251.

Messner, J. W., & Pinson, P. (2019). Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *International Journal of Forecasting*, *35*(4), 1485–1498.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 347–370.

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

Oberndorfer, U. (2009). Energy prices, volatility, and the stock market: Evidence from the Eurozone. *Energy Policy*, *37*(12), 5787–5795.

Paraschiv, F., Erni, D., & Pietsch, R. (2014). The impact of renewable energies on EEX day-ahead electricity prices. *Energy Policy*, *73*, 196–210.

REN21 (2018). Renewables 2018 - global status report. ISBN: 978-3-9818107-6-9.

Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, *71*(3), 599–607.

Shah, I., & Lisi, F. (2019). Forecasting of electricity price through a functional prediction of sale and purchase curves. *Journal of Forecasting,*.

Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, *24*(111), 647–656.

Terna (2018). *Statistical data on electricity in Italy*: *Technical report*, https://www.terna.it/en/electric-system/statistical-data-forecast/statistical-publications.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*(1), 267–288.

Uniejewski, B., Marcjasz, G., & Weron, R. (2019). Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO. *International Journal of Forecasting*, *35*(4), 1533–1547.

Weron, R. (2007). *Modeling and forecasting electricity loads and prices: A statistical approach, vol. 403*. John Wiley & Sons.

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

Weron, R., & Misiorek, A. (2008). Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, *24*(4), 744–763.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 1067–1084.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics & Control*, *18*(5), 931–955.

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

Ziel, F., Steinert, R., & Husmann, S. (2015). Efficient modeling and forecasting of electricity spot prices. *Energy Economics*, *47*(C), 98–111.

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