Energy Policy 123 (2018) 390–403

Contents lists available at ScienceDirect

Energy Policy

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

Managing electricity price modeling risk via ensemble forecasting: The case of Turkey☆

Ezgi Avci[a,c,*], Wolfgang Ketter[b], Eric van Heck[a]

[a] Rotterdam School of Management, Erasmus University, the Netherlands
[b] Turkish Standards Institute, Ankara, Turkey
[c] Institute of Energy Economics, University of Cologne, Germany

ARTICLE INFO

Keywords:
Electricity markets
Day ahead auctions
Risk of price modeling
Ensemble forecasting
Multi-seasonality
Exogenous variable selection
Semi-transparent market

ABSTRACT

There are two ways of managing market price risk in electricity day ahead markets, forecasting and hedging. In emerging markets, since hedging possibilities are limited, forecasting becomes the foremost important tool to manage spot price risk. Despite the existence of great diversity of spot price forecasting methods, due to the unique characteristics of electricity as a commodity, there are still three key forecasting challenges that a market participant has to take into account: risk of selection of an inadequate forecasting method and transparency level of the market (availability level of public data) and country-specific multi-seasonality factors. We address these challenges by using a detailed market-level data from the Turkish electricity day-ahead auctions, which is an interesting research setting in that it presents a number of challenges for forecasting. We reveal the key distinguishing features of this market quantitatively which then allow us to propose individual and ensemble forecasting models that are particularly well suited to it. This forecasting study is pioneering for Turkey as it is the very first to focus specifically on electricity spot prices since the country's day-ahead market was established in 2012. We also suggested applicable policy and managerial implications for both regulatory bodies, market makers and participants.

1. Introduction

Electricity day-ahead auctions play a central role for the sustainability of electricity markets since they reveal the *reference price* for all market participants. In emerging markets, since the number and variety of hedging tools are limited, for market participants accurate forecasting becomes the most essential tool for managing spot price risk. On the other hand designing a market with a proper transparency level is one of main responsibilities of the policy makers to let market participants generate reasonable forecasts using public information.

Especially after the well-known California crisis in the 2000s, the number of studies on day-ahead price forecasting increased substantially, as the need for such studies became apparent (e.g., Borenstein, 2001). Since the storability of electricity is limited, electricity prices reveal characteristics that differ from other commodities and present specific forecasting challenges. Studies focus first on the basic characteristics of electricity, namely non-storability and inelasticity of supply/demand (Geman and Roncoroni, 2006; Lucia and Schwartz, 2002), and then examine spikes, nonstationarity and mean reversion (Haugom and Ullrich, 2012; Knittel and Roberts, 2005). However there are still three challenges that needs to be dealt with in electricity price forecasting (Weron, 2014).

The first one is a methodological issue, the *risk of selection of an inappropriate forecasting model*. Although various individual forecasting methods are suggested in the literature, none of them has been proven to be superior (Chen and Bunn, 2010; Weron, 2014), and the

☆ The corresponding author greatly acknowledge financial support from Ministry of Science, Industry of Technology-The Scientific and Technological Research Council of Turkey (TUBITAK) under the project 114K601 and Jean Monnet Scholarship (2015–2016) in the area of energy funded by EU through Ministry for EU Affairs. The authors are grateful to Doganbey AKGUL, Ph.D. (Head of Strategy Development Department at TR. Ministry of Energy and Natural Resources (MENR), Zafer DEMIRCAN, Ph.D. (General Directorate of Energy Affairs General Directorate at TR.MENR), Abdulkadir ONGUN (Vice President at Electricity Transmission System Operator of Turkey (TEIAS), Fatih BAYTUGAN (Head of Information systems at TEIAS) and Engin ILSEVEN for providing the data set and their valuable comments. The authors also greatly acknowledge the feedback from participants of 5th International Symposium on Environment and Energy Finance Issues (ISEFI-2017), Paris, France and INFORMS-2017, Houston, Texas, USA. It should be noted that the views in this article are entirely those of the authors and do not represent those of the institutions with which they are affiliated with.

* Corresponding author.
E-mail address: eavci@tse.org.tr (E. Avci).

https://doi.org/10.1016/j.enpol.2018.08.053
Received 23 January 2018; Received in revised form 24 July 2018; Accepted 22 August 2018
0301-4215/

performance of individual models depends on the periods being considered and the characteristics of market (Aggarwal et al., 2009). To compensate for the weaknesses of each individual method, the combined forecasting approach was developed (Crane and Crotty, 1967; Bates and Granger, 1969). The main advantage of combining forecasts is not that the ex-post performance of best ensembles are better than that of best individuals, but that it is less risky to ensemble forecasts than to select ex-ante one individual forecasting method (Hibon and Evgeniou, 2005). Although ensemble approach has been well-studied in other contexts (Stock and Watson, 2004; Timmermann, 2006), the number of combined forecasting studies relating to electricity markets is limited (Weron, 2014). The existing studies are mostly on electricity load forecasting (e.g. Bunn, 1975, 1977; Taylor, 2010). In the context of day-ahead price forecasting, leading point forecasting studies are, Bordignon et al. (2013) and Nowotarski et al. (2014), both of which are from mature electricity markets (the UK and NordPool). The recent few studies (e.g. Maciejowska et al., 2016; Gaillard et al., 2016, Maciejowska and Nowotarski, 2016) focus on probabilistic forecasting and are from the results of the Global Energy Forecasting (GEF) Competition-2014 in which US zonal prices are used. One of the main findings of these papers is that analysis of different market conditions can provide important insights in terms of comparing individual and ensemble models.

Second challenge is a more market-specific one which is the *appropriate selection of exogenous variables* (e.g., Keles et al., 2016), since the quality and availability of public data (transparency level of the market) and their influence on price may differ depending on the market studied (Aggarwal et al., 2009; Von der Fehr, 2013). Including exogenous variables (e.g., demand) generally increases the accuracy of price forecasts. However, when the number of exogenous variables is increased, the probability of data quality issues and access problems can increase, and may lead to even worse price forecasts. Therefore, in devising a sound process for selecting variables the features of the market need to be taken into account.

The last challenge is catching market-specific *multi-seasonality* (e.g., Janczura et al., 2013) characteristics of the spot price. Multi-seasonality is defined as having simultaneously daily, weekly and annual components. The *annual seasonality* is more difficult to detect, as it is masked by more irregular patterns, and it is often ignored in studies as it is generally believed to add complexity to already parsimonious models (Weron, 2014; Nowotarski and Weron, 2016). The problem of *daily seasonality* is solved by either taking daily averages or modeling each hour separately (Misiorek et al., 2006; Karakatsani and Bunn, 2008, 2010). Since predictability level (and thus market efficiency) of electricity prices can change over the course of the day, choosing forecasting models by considering fractal properties is very important (Avci-Surucu et al., 2016). The *dual-calendar effect* is another factor that influences prices in countries which follow both the Hijri and Gregorian calendars for holidays (De Livera et al., 2011).

In this paper our aim is to examine the performance of carefully selected individual and ensemble models in an emerging EDA market, Turkey, which is an interesting research setting in that it presents several challenges for forecasters. First, since it is an emerging market, hedging possibilities are limited, forecasting becomes the foremost important tool to manage spot price risk for power agents. As stated in Hong (2015), even one percent decrease in short term price forecasting error could result in a hundred thousands of profit per year for a medium sized utility. Even worse, as in most of the emerging markets, there is no information transparency platform from which market participants can obtain relevant data easily and use them in order to decrease their forecast errors. This makes the Turkey day ahead market (DAM) a *semi-transparent* one in terms of information dissemination.[1] Secondly, there is no nuclear power plant which eases the horizontal shift of the base load in the hourly merit order curve (MOC). Third, there is no gas forward market as gas prices are regulated by the government, therefore it is more difficult to catch the trend seasonality by calculating the marginal costs of bidders. Fourth, the observance of multiple religious holidays in Turkey means that dual-calendar seasonal effects can be found and this situation complicates modeling seasonality component. Lastly, bids by the state-owned hydro power plants (PP) have a crucial impact on the shape of the MOC, therefore appropriate selection of the exogenous variables relating to supply stack characteristics becomes vital. Thus, our main research question is '*How to manage price modeling risk via ensemble forecasting in the Turkish electricity day ahead auctions'.* In sum, we aim to develop well-performing ensemble forecast models for an emerging market and compare their performance with alternative individual models. This will enable us to address some of the shortcomings of existing ensemble modeling studies, specifically the inadequate handling of trend seasonality component, the inappropriate selection of exogenous variables when one considers the relationships that exist between price and system constraints, and the methodological similarity of the individual models they have considered. We contribute to the energy economics literature on both theoretical and practical dimensions. Firstly, regarding the variety and type of individual models, ours is one of the first studies to include a long-term seasonality component for ensemble forecasts and to provide evidence of their performance. Secondly, since all of the previous ensemble forecasting studies of electricity price are from well-developed markets (such as the UK and Nord Pool) and ours is the first to look at a developing market with a semi-transparent structure, our findings may be informative for market participants and policymakers in other developing markets with similar transparency features. Lastly, this forecasting study is also pioneering in terms of the Turkish electricity market since it is the very first to examine electricity day-ahead auction prices after the establishment of the DAM in 2012.

The remainder of the paper is organized as follows. In Section 2, we briefly explain the theoretical framework of the individual and ensemble models that we adopt for this study. Section 3 illustrates the conceptual background related to general processes and fundamental drivers of the clearing price in day ahead auctions. Section 4 describes the data and its temporal properties, Section 5 explains the theoretical framework and presents the results, and Section 6 concludes the paper with a summary of the findings and suggestions for future research.

2. Theoretical framework

We consider three classes of individual models; econometric time series, artificial neural networks and seasonality models; which are chosen by virtue of their suitability for catching the specific features of the price dynamics of electricity day ahead auctions and also for minimizing the methodological similarities between them guided by the previous literature. We aim to understand how their forecasting performance changes according to the particular electricity market. We derive our ensemble models by using *equally weighted means* which has been highly advocated (Makridakis et al., 1998; Stock and Watson, 2004). All the models (both individual and ensemble) are executed with a set of explanatory variables which are carefully selected from the literature due to their fundamental relation with auction mechanism of the day ahead markets and their correlation with the clearing price (Karakatsani and Bunn, 2008; Nan, 2009).

2.1. Exogenous variables affecting electricity day ahead auction prices

The selection of exogenous variables is a crucial step for developing forecasting models. Market characteristics, nonstrategic uncertainties, other stochastic uncertainties, behavioral indices, and temporal effects are main classes of input variables effecting electricity prices (Karakatsani and Bunn, 2008). *Historical electricity prices* (e.g., price lags of 1–7, 14, 21, 28 or 364 days) are the most extensively used variable.

[1] Details about the *transparency level* is given in Section 3.

Due to its strong correlation with price, *demand* is the next most often used input variable. There are different approaches to using demand in the models, namely demand forecasts made by Independent System Operator (ISO) (e.g. Nan, 2009; Bordignon et al., 2013), in-house demand forecasts (e.g. Georgilakis, 2006, Mandal et al., 2006) and historical demand data (e.g. Weron, 2006). Since storage of electricity is limited, variables that indicate stochastic uncertainties (e.g., system constraints) can be very useful for modeling the spikes or extreme conditions. *Reserve margin* (the difference between available capacity and demand) (Eydeland and Wolyniec, 2003; Harris, 2006) and *reserve margin ratio* (the ratio of demand divided by available capacity, and is also an indicator of safe functioning of power system) (Anderson and Davison, 2008; Cartea et al., 2009; Davison et al., 2002; Maryniak, 2013; Maryniak and Weron, 2014) are two of the most commonly used system capacity constraint variables. Another possible explanatory variable is *temperature* as it is strongly correlated to demand (Mandal et al., 2006; Guo ve Luh, 2003; 2004). However, if demand can be forecasted accurately, the explanatory power of temperature can be ignored considering the data quality and accessibility issues. The marginal costs of market participants affect their bidding strategies and shape (i.e., steepness) of the MOC. Thus *natural gas and oil prices* have also been used to explain prices (Guo and Luh, 2003, 2004; Gao et al., 2000; Zhang and Luh, 2005; Bordignon et al., 2013). Importance level of these exogeneous variables for electricity prices may change according to the market design and characteristics, and their effect may be limited (Weron, 2014). For the UK market, for example, whilst Bordignon et al. (2013) find demand and gas price to have significant effects on the UK spot electricity prices, Maciejowska (2014) finds these fundamental variables to have only minor effects. Therefore optimal selection of exogeneous variables depends on the characteristics of the market, the type of model used, data access,[2] heuristics, and experience of the modeller.[3]

In this paper, to select the appropriate exogenous variables we follow the approach of Weron (2006, 2014), Karakatsani and Bunn (2008) and Bordignon et al. (2013), and use only publicly available explanatory data even though we also have access to insider information. We have three reasons for this; first, we are looking at how these methods are used by energy experts in the market and policy makers; second, it allows us to compare our forecast accuracy results with the previous studies; and third, although we have access to insider information (the level of all dams with reservoirs, forecasted geospatial temperatures) we prefer not use these data, due to the data being of poor quality and not available until relatively late (usually not before the gate's closure in the day-ahead market). From these variables we chose the ones which have been most widely used in the literature and which are most relevant for the Turkish electricity market. We run each model with and without exogenous variables to see the effect of adding explanatory variables on performance. The exogenous variables (X) used in the models are summarized in Table 1.

2.2. Individual forecasting methods

In the following subsections, we briefly explain the parametric and nonparametric methods that are adopted in this paper for generating individual forecast models.

2.2.1. Econometric time series models

Econometric time series methods forecast the current price by using a mathematical combination of historical prices and/or historical or current values of exogenous variables. Although these methods are not good for spiky periods, they perform well in terms of capturing the general patterns in electricity price data (Weron, 2014).

Table 1
Exogenous variables used in the models.

Variables	Description	Notation and lags
Forecasted demand (d)	Day-ahead demand forecast published by the system operator	d_{t-j} (j = 0,1, 2, 3, 7,14,21, 28)
Margin (m)	Available capacity (generation, surplus) – demand forecast	m_{t-i} (i = 0,1, 2)
Day-specific dummy	Three dummy variables for Saturday, Sunday and Monday, separately	D_{sat}, D_{sun} , D_{mon}
Holiday dummy	Dummy variable for official and religious holidays	D_{hol}

Seasonal autoregressive moving average (ARMA) models (SARMA)

If the dependence on the past prices is likely to occur at certain seasonal lag s, it is appropriate to introduce autoregressive and moving average polynomials that identify with the seasonal lags. The resulting seasonal autoregressive moving average model is ARMA $(P;Q)_s$, which can be written as $\Phi_P(B^s)x_t = \Theta_Q(B^s)\omega_t$, where $\Phi_P(B^s) = 1-\Phi_1 B^s - \Phi_2 B^{2s} - ... - \Phi_p B^{Ps}$, and $\Phi_Q(B^s) = 1+\Phi_1 B^s + \Phi_2 B^{2s} + ... + \Phi_Q B^{Qs}$ are the seasonal autoregressive operator and the seasonal moving average operator of orders P and Q, respectively, with seasonal period s. SARIMA models are most often used as a benchmark in electricity price literature. The following AR formula gives a general expression of the model structure adopted in this study:

$$\begin{aligned} p_t = \; & \varnothing_1 p_{t-1} + \varnothing_2 p_{t-2} + \varnothing_3 p_{t-3} + \varnothing_4 p_{t-7} + \varnothing_5 p_{t-14} \\ & + \varnothing_6 p_{t-21} + \varnothing_7 p_{t-28} + \psi_1 d_t + \psi_2 d_{t-1} + \psi_3 d_{t-2} \\ & + \psi_4 d_{t-3} + \psi_5 d_{t-7} + \psi_6 d_{t-14} + \psi_7 d_{t-21} + \psi_8 d_{t-28} \\ & + \delta_1 m_t + \delta_2 m_{t-1} + \delta_3 m_{t-2} + \gamma_1 D_{sat} + \gamma_2 D_{sun} \\ & + \gamma_3 D_{mon} + \alpha_1 D_{hol} \end{aligned} \tag{1}$$

where p_{t-i} denotes the ith lag of the clearing price which represents the autoregressive effects of the previous days, d_{t-i} is for the ith lag of the demand forecast where i = 0 represents the demand forecast for the next day. m_{t-i} is for the ith lag of margin variable where i = 0 represents the planned margin for the next day. D_{sat}, D_{sun} , D_{mon} are dummy variables for days Saturday, Sunday and Monday to consider the weekly seasonality. The dummy variable D_{hol} refers to the official and religious holidays in the whole dataset. The orders of our models are identified through Akaike Information Criterion (AIC) and investigation of the residual diagnostics.

2.2.2. Artificial intelligence (AI) models

AI methods model price processes using non-parametric approaches. Due to their flexibility and their ability to handle complexity and nonlinearity they have been preferred by many authors (e.g., Catalao et al., 2007; Pino et al., 2008; Vahidinasab et al., 2008) for short-term price forecasting. Despite their popularity for individual modeling, there is still an open question for their performance in ensemble models (Keles et al., 2016) with and without exogenous variables. We use the *nonlinear autoregressive exogenous (NARX)* AI method which have been proposed by Lin et al. (1996). These recurrent networks also have very good learning capabilities and generalization performance (Weron, 2014) and have shown that they can greatly improve performance on long-term dependency problems. They relate the current value of a time series both to past values of the same series and to current and past values of the exogenous variables. NARX can be written implicitly as:

$$y_t = F(y_{t-1}, y_{t-1}, y_{t-3}, ..., u_t, u_{t-1}, u_{t-2}, u_{t-3}, \cdots) + \varepsilon_t \tag{2}$$

where u are the exogenous variables and ε is the error term. The function F is a neural network. A NARX $(p,P,k)_m$ model has inputs $(y_{t-1}, y_{t-2}, ..., y_{t-p}, y_{t-m}, y_{t-2m}, y_{t-pm})$ and k neurons in the hidden layer. For this paper we use a *feed-forward single layer algorithm* NARX.

[2] For a detailed discussion, see Weron (2014) and, Aggarwal et al. (2009).
[3] For a detailed discussion, see Amjady and Hemmati (2006).

2.2.3. Seasonality models

Adequate treatment of seasonality in electricity prices is of utmost importance for developing sound forecasting models (Janczura et al., 2013; Lisi and Nan, 2014) There are different approaches to modeling seasonality such as piecewise constant/sinusoidal functions and wavelets. A failure to acknowledge seasonality or to handle it appropriately can significantly reduce the accuracy of the model (Ketter et al., 2009, 2012; De Livera et al., 2011). For short term electricity price forecasting; although daily and weekly components have been taken into account, long-term seasonal component (LTSC) has been neglected due to the general belief that it adds unnecessary complexity to already parsimonious models (Weron, 2014). However, recently Nowotarski and Weron (2016) have shown that taking the LTSC component into account in short term electricity price forecasting can significantly improve the model performance and create a natural opportunity to include them in ensemble forecasts. This can be done in a number of ways; polynomial and sinusoidal regression, polynomial kernel regression, linear regression and smoothing splines, filters, state space models etc. (for details see Hyndman et al., 2002; De Livera et al., 2011, Lisi and Nan, 2014). After a thorough review of this extensive literature we adopt two of the recently developed, state space based models, namely *Exponential smoothing state space (ETS)* and *Trigonometric Box-Cox Transformed ARMA errorred Seasonal (TBATS)* models, by virtue of their ability for handling complex seasonal patterns existing in electricity prices.

Hyndman et al. (2002) have expanded earlier work by Ord et al. (1997) on special types of innovation state-space models which underlie exponential smoothing models and have developed a new and more general set of methods which bring exponential smoothing into the same class as ARMA models. These are known as ETS models. For this paper we utilize the algorithm of Hyndman et al. (2002) to estimate the parameters of our ETS models.

Commonly used seasonality models, *single seasonal exponential smoothing* (Makridakis et al., 1982; Makridakis and Hibon, 2000; Snyder, Koehler, and Ord, 2002) and *second seasonal Holt–Winters* (Taylor, 2003) methods cannot catch complex seasonal patterns such as non-integer seasonality and calendar effects, or time series with non-nested seasonal patterns. The *nonlinear versions of the state space models* can be unstable, as they have infinite forecast variances beyond a certain forecasting horizon, and analytical results for the prediction distributions are not available (Akram et al., 2009). Thus we use the TBATS algorithm, introduced by De Livera et al. (2011), which is developed as an alternative estimation method to allow for non-integer seasonality and calendar effects, or time series with non-nested seasonal patterns; and handle a wider variety of seasonal patterns.

2.3. Ensemble models

The idea of ensemble forecasts was developed by Reid (1968, 1969) and Bates and Granger (1969), and has been extensively studied (e.g., Bunn, 1975, 1977; De Menezes et al., 2000; Timmermann, 2006; Altavilla and De Grauwe, 2010; Clark and McCracken, 2009). The idea behind ensembling forecasting techniques is straightforward: individual models have their own weaknesses and none of them is superior to the others. Ensemble forecasting enables us to compensate for the weaknesses of individual models. In general, an ensemble forecast including a set of K competing spot price predictors ($\hat{P}_t^{(1)},....,\hat{P}_t^{(K)}$) can be written as:

$$\hat{P}_t^{(C)} = f(\hat{P}_t^{(1)},....,\hat{P}_t^{(K)};\omega) \quad (3)$$

where f is a generic function and ω is a parameter vector. Using linear functions, this expression can be written as $\hat{P}_t^C = \sum_{k=1}^K \omega_k \hat{P}_t^{(k)}$. In general, weights ω_k can be constant or time-varying, ω_{t-k}. Several studies have shown that, due to the effect of finite-sample error in estimating the combining weights, an equally weighted mean is often the best choice (Clemen, 1989; Makridakis and Winkler, 1983; Smith and Wallis, 2009; Stock and Watson, 2004). Although ensemble forecasts have generally been shown in theoretical studies to outperform the individual methods (e.g., Chen and Yang, 2007) they have not stood out in the context of electricity markets. The first ensemble forecasting studies in electricity markets were done by Bunn (1985) and Bunn and Farmer (1985), looking at load forecasting provided a theoretical discussion of the merits of ensemble modeling. Smith (1989) provides empirical evidence to show the superiority of forecast combinations of ARMA models over the individual ARMA models. Taylor and Majithia (2000) find that the performance of ensemble models differs with respect to the time of the day.

The literature on ensemble modeling for electricity price forecasting is very rare. As far as we know, there have been only three studies; Nan (2009), Bordignon et al. (2013) and Nowotarski et al. (2014). Nan (2009) is the pioneer study in this area. She examines the UK market prices and uses 19 individual models which are derived from four basic models; namely linear regression, ARMAX, time-varying regression and Markov regime switching models. The author then selects subsets of individual models through using model confidence set and encompassing approaches. Her ensemble models outpetform the individual ones in most cases. In a follow up study, with the same data set, Bordignon et al. (2013) examine the performance of ensemble models against five individual models; linear regression, ARMAX, time-varying regression and two Markov regime switching models and conclude that most ensemble models perform better than individual ones. Nowotarski et al. (2014) as an extension of these two studies, increased the number of considered markets (to Nordpool, EEX and PJM), time periods and individual models (to AR- and mean reversion based) and similarly find supporting evidence for the outperformance of ensemble models. However they show that the performance of combined forecasts may differ with respect to the market considered and periods.

In this paper we aim to develop well-performing ensemble forecast models for an emerging market and compare their performances with competing individual models addressing some short comings of existing ensemble modeling studies. First, we increase the variety of considered individual models to minimize the methodological similarity between them. We extend the set of individual models to seasonality methods; which have not been considered by the prior literature; in virtue of their potential to describe some specific characteristics of electricity prices. We adopt ETS and TBATS seasonality models to take into account inadequate handling of trend-seasonality component and to deal with the complex multi-seasonality in electricity prices. Second we conduct the analysis following the literature which has found that price volatility, fractality, market efficiency, predictability of prices, costs and operational constraints differ for each hour/hour block during the course of a day (Shahidehpour et al., 2002; Huisman et al., 2007; Avci-Surucu et al., 2016). Focus of the previous studies are on testing the performance of ensemble models for each hour separately (e.g. Huisman et al., 2007) or taking daily averages (e.g. Bunn, 2004; Bunn and Karakatsani, 2003) . We rather stayed focus on the market efficiency and predictability of prices in a day and provide important insights on the performance of our models considering three different tariff timezones – namely T1 (day): 06:00–17:00, T2 (peak): 17:00–22:00, and T3 (night): 22:00–06:00, leading to three sets of parameters for each day. Third, most of the previous studies are from well-developed markets; namely UK, Nordpool and PJM. We extend these studies to an emerging market which has different price characteristics in terms of seasonality, spikes and fractality level. Thus we believe that our findings can be informative for the market participants/policy makers in other developing markets with similar characteristics (e.g. share of renewables, natural gas etc.). Lastly, while testing the additional gains from ensemble modeling approach we also test the additional benefit from including exogeneous variables considering the fundamental relations between price and system constraints through using a complete and big data set.

3. Conceptual background

3.1. Information transparency level of the market

The general principles and processes of the DAM in Turkey is similar to most European electricity markets. Between 2012 and 2015 DAM was operated by Electricity Market Financial Reconciliation Center (PMUM), and is currently operated by the Istanbul Energy Exchange (EXIST). Prices and volumes are determined each hour on a daily basis. Until September 2015 information on the market-specific variables was made available by PMUM, then EXIST took over this responsibility and now disseminates both forecasted (ex-ante), real-time and realized (ex-post) data through its Market Transparency Platform in a timely manner. DAM-related information revealed through the transparency platform and its details are illustrated in Tables A.1 and A.2. As can be observed from these tables, until the beginning of 2016 information transparency level of the market is low which can be defined as *semi-transparent*. DAM prices in our sample (Jan.2012-Dec.2015) were comprised of aggregation of expectations of market participants making decisions in such a transparency level. Thus to imitate the real decision making setting of power agents, we utilize from only public data released during that time period.

3.2. Fundamental drivers of EDA clearing price in Turkey

Changes in the supply and demand characteristics, market architecture and/or ownership structure can affect the dynamics of the clearing price in uniform-price auctions (Petrella and Sapio, 2009; Glachant and Saguan, 2007) and lead to structural breaks in the series. In this section we investigate these characteristics in the Turkish DAM to understand whether there have been any structural breaks in the price series between 2012 and 2015. The capacity mix in Turkey is not particularly well balanced and is mostly based on fossil fuels, which constituted 67.7% of the total generation in 2015. Natural gas is the major fuel in electricity generation (38.6%), followed by coal (28.3%), and some fuel oil (0.8%). The capacity mix in Turkey can change year on year, owing to the seasonality of hydro supply and availability of old lignite plants. Currently, generation activities are carried out by three parties: Electricity Generation Company (EUAS), private generators and autoproducers. In 2012, the share held by EUAS and its affiliates was 42% of the total installed capacity. In 2015 this state-owned share decreased to 28% with a portfolio of hydro and lignite power plants. Independent power producers (IPPs), build operate transfer (BOT), build operate own (BOO) and transfer of operating rights (TOOR) utilities are responsible for most of the power generation, and hold 70% of the total capacity. The important issue here is that the privatized EUAS power plants were those that had not generally been marginal generators in the MOC. Therefore we can confidently state that there has not been any major change in industry structure that has led to a structural break in the price series.

3.3. Electricity price forecasting studies in Turkey

There is an absence of substantial research on electricity price forecasting in Turkey and those studies that do exist are in the form of PhD theses or conference proceedings and represent the period before the day ahead market was established. Hayfavi and Talasli (2014) model logarithmic daily average spot prices for electricity as the sum of a deterministic function and a multi-factor stochastic process. They do not use any exogenous variables. Their data interval runs from December 2009 to July 2011, which was the *Day Ahead Planning* period and does thus not reflect the market regulations and conditions of the day ahead market currently operating in Turkey. Yıldırım et al. (2012) use dynamic regression, conic multivariate adaptive regression splines (CMARS) and robust CMARS (RCMARS) to forecast the next day's electricity prices. They do not use any exogenous variables in their models. For their training data they use only one month of data because of the limitations in their RCMARS method, thus the validity of their results is questionable in the context of electricity markets. Kölmek and Navruz (2013) use ANN and AR methods to forecast the electricity prices with a training data interval running from December 2009 to November 2010. For the ANN model they use historical day ahead prices, demand forecast, bilateral contact, and available capacity as exogenous variables, however, they do not use any exogenous variables for the AR model. The authors compared these two methods with respect to mean absolute percentage error (MAPE) and conclude that performance of the ANN method is higher than that of the AR. This study is valuable in the sense that it is the first published paper which uses ANN for electricity price forecasting in Turkey. However, it has some deficiencies. First, their data is from the day ahead planning period and does not represent the current electricity market conditions. Second, they compare these two methods, one with exogenous variables and one without. It is known that models with exogenous variables tend to perform better than the ones without them, thus the validity of their performance comparison is also questionable.

Unlu (2012) has studied the linear relationship between temperature and day ahead electricity prices using a data set of 35 days. He investigates in particular the seasonality and predictability issue in temperature. In the final stage, in which linear models (AR and ARX) and 30 days training data are used, he attempts to predict the electricity price for the next five days. The drawback of this study is that it takes no account of the well-known nonlinear relationship between temperature and electricity prices, and thus concludes that there is no relationship between these two variables. From the studies undertaken up to now, it is very difficult to understand the general characteristics of electricity prices in Turkey and the driving factors behind them since all the studies use datasets from the day head planning period, the training sets are not large enough for the results to be generalizable, and all have some methodological deficiencies.

Ozyildirim and Beyazit (2014) forecast the day ahead electricity prices using linear regression and radial basis function with a dataset from the period 2010–2013. They found that the out-of-sample performance of RBF is slightly better than that of linear regression. They use hourly market clearing price (MCP) without logarithmic transformation and integration. Their exogeneous variables are MCP lags, temperature, square of temperature (to deal with the nonlinearity issue) and hourly, daily, monthly, and holiday dummies. By using graphs and descriptive statistics, they showed that hourly MCP have distinct clusters with respect to time zones.

As far as we know the only study that reflects current market conditions is Taysi et al. (2015). They forecast MCP using SARIMA and ANN methods with exogeneous variables: historical prices and calendar dummies. They showed that the performance of these two methods is very similar. The crucial deficiency of this study is that it uses 1,2,3,4 lagged MCP as exogenous variables. This issue has been well-studied in the energy economics literature and has been found to run counter to the electricity market mechanism. Thus the models in this study cannot be implemented in real life since bidders do not know the prices up to 24 h ahead. Our examination of the literature on the Turkish electricity market shows that there are no electricity price forecasting studies that reflect current market conditions and mechanisms.

4. The data

The data set used in this study consists of hourly day ahead auction clearing price time series from the Turkey Electricity Market. The sample starts on 1 January 2012 and ends on 27 December 2015 providing a total of 34.944 hourly observations. This period covers the time after the DAM in Turkey was established and does not have any significant structural changes in the market mechanism and auction design, until the launch of the Market Transparency Platform at the EXIST. The time series data set were obtained from the General

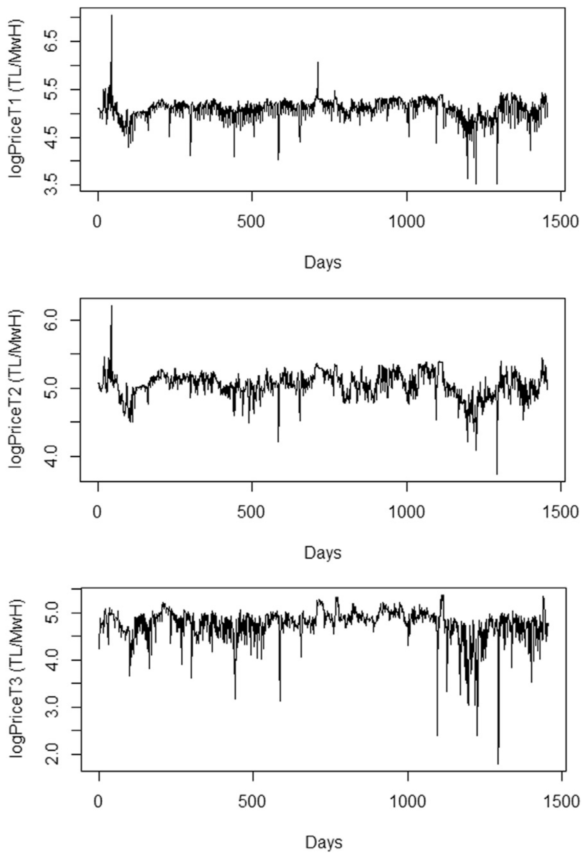

Fig. 1. Time series plot of market clearing price with respect to three tariff time zones.

Directorate of Energy Affairs at the Ministry of Energy and Natural Resources without any missing values or doubled values. Although in the literature some studies (e.g. Weron, 2006) advocate the substitution of outliers with arithmetic averages of some neighbour values or the general mean, we intentionally did not preprocess the data set in terms of outliers (at 10.02.2012, 11.02.2012 and 13.02.2012) to see the real performance of the artificial intelligence and seasonality models. We used R 3.4.3 for the conduction of our analyses.

4.1. Summary of the descriptives

Based on the findings of Avci-Surucu et al. (2016), we conduct our analyses considering three tariff time zones T1 (day): 06:00–17:00, T2 (peak): 17:00–22:00, and T3 (night): 22:00–06:00 where each time series has 1456 observations. Fig. 1 illustrates plots of each series and Table 2 gives summary statistics for each time zone. T3 series has the highest standard deviation and skewness, but the lowest mean. T1 and T2 series have similar statistical characteristics.

From an empirical point of view, electricity spot prices has different forms of nonlinear dynamics, with the crucial one being the strong dependence of the variability of the series on its own past. 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 (Weron, 2014). In literature it has been established that it is often advantageous to discard the models with the worst performance to eliminate too much variability which is introduced in the combinations because of models that perform very poorly during particular seasons and/or for particular hours. In the earlier versions of the paper, we used GARCH-type models and their ensembles however since their performance are very poor, we did not consider them as a benchmark.

4.2. Stationarity, long term correlation and predictability

We investigate stationarity and predictability level of price series through their fractal dynamics since accurate measurement of fractality is crucial for correct statistical inference and forecast uncertainty (Lildholdt, 2000). There are three unit root tests commonly used to test the stationarity of a process: 1) the AugmentedDickey-Fuller (ADF) test, 2) the Phillips– Peron (PP) test and 3) the Kwiatkowski–-Phillips–Schmidt–Shin (KPSS) test. However if we wish to test stationarity as a null and have strong priors in its favour, employing the ADF test may not be useful (Baillie et al., 1996). An empirical series with d close to 0.5 will probably be misclassified as nonstationary. Therefore, we use a combination of the PP and KPSS tests allowing us to determine the four possible outcomes of the series (Baillie et al., 1996): 1) if the PP is significant and the KPSS is not, then the data are probably stationary with $d \in (0;0.5)$—strong evidence of a covariance stationary process; 2) if the PP is insignificant and the KPSS is significant, then the data may indicate having brown noise—a strong indicator of a unit root, i.e., an I(0) process; 3) if neither the PP nor the KPSS is significant, then the data are insufficiently informative regarding the long memory of the process; and 4) if both the PP and the KPSS are significant, then the data are not well described as either an I(1) or an I(0) process—$d \in (0; 1)$.

Table 3 presents the unit root tests for logarithm of prices. As can be observed from the PP and KPSS values, price series for each time zone are not well described as either an I(1) or an I(0) process which means the differencing parameter (d) is not an integer but between 0 and 1. We estimate this differencing parameter through *Geweke–Porter-Hudak (GPH)* and *Local Whittle estimator (Whittle)*. We adopt these methods to benefit from their different statistical properties; namely GPH's common usage and comparability with the literature, Whittle's parametric efficiency and consistency. As can be seen from Table 3, each price series has different fractal dynamics. T1 tariff zone has the lowest differencing average value (0.372) which indicates that it has the highest long-term correlation and prices in this tariff zone are the most predictable. T3 price series has the highest fractal value demonstrating that predictability level of prices in this tariff zone are the lowest. This confirms the proposition that at peak load (here T1 and T2), marginal generators give hyperbolically decaying weights to information by considering the prices of a day/week before (Sapio, 2004). At off-peak load, if marginal generators bid at their marginal costs, then there is no fractal noise, as we observed in the night tariff time zone T3.

4.3. Multi-seasonality

Multi-seasonality and complexity of the long-term correlation structure in electricity prices are two of the most important challenges for forecasters. Multi-seasonality is mainly shaped by the variation in

Table 2
Descriptive statistics for three tariff time zones.

Variables	Min.	Max.	Range	Median	Mean	Variance	Std.dev	Skewn.	Kurtosis
logPriceT1	3.53	7.06	3.53	5.1358	5.09	0.05	0.22	− 0.98	15
logPriceT2	3.74	6.21	2.47	5.0626	5.05	0.04	0.19	− 0.66	6.8
logPriceT3	1.79	5.38	3.58	4.8283	4.75	0.11	0.33	− 2.7	16

Table 3
Unit Root results for logarithm of prices in three tariff time zones.

Variables	PP	KPSS	GPH	Whittle	Avr. Fract. value
logPriceT1	− 1259.6**	0.302*	0.450	0.295	0.372
logPriceT2	− 440.81**	0.476*	0.428	0.492	0.460
logPriceT3	− 345.2**	0.857**	0.477	0.567	0.522

Notes: PP null hypothesis: non-stationary, KPSS null hypothesis: stationarity. *Avr. Fract.value* is the average of the PP, KPSS, GPH and Whittle estimates for the corresponding row.
*** indicates rejection of the null hypothesis at the 1% significance level. ** indicates rejection of the null hypothesis at the 5% significance level. * indicates rejection of the null hypothesis at the 10% significance level.

Table 4
Correlation table for the exogenous variables.

	MCP	Forecasted demand	Margin	Demand
MCP	1	0.604**	− .151**	0.602**
Forecasted Demand	0.604**	1	− 0.515**	0.987**
Margin	− 0.151**	− .515**	1	− 0.476**
Demand	0.602**	0.987**	− .476**	1

** Correlation is significant at the 0.01 level (two-tailed).

demand. The daily cycle refers to variations between day and night and during the different points in the day. This daily variation in prices reflects the working habits of the population. To remove *intra-day seasonality*, it is possible to model each hour (or half-hour) of the day separately, as different commodities (Ramanathan et al., 1997; Guthrie and Videbeck, 2002) or by taking daily averages. However it has been shown that daily average prices do not capture the microstructure of the day ahead auctions since the level of mean reversion, volatility structure and predictability (fractality) level in clearing prices are not constant throughout the day (Huisman et al., 2007; Avci-Surucu et al., 2016). Following the approaches of the aforementioned studies, for our modeling purposes we use the average price for each tariff time zone since first, we do not want to lose information about the microstructure of day-ahead prices, as would be the case were we to use daily averaging. Second, previous studies that have considered each hour separately conclude that there is a block-structured correlation between specific hours resembling the time zones. Third, taking average with respect to each time zone is more intuitive in the sense that electricity market participants have different incentives and bidding strategies for each time zone.

The *intra-week variability* is also non-negligible. The load profile for Saturday and Sunday is generally lower than for the weekdays; this feature is called the *weekend effect*. Similar to the findings of Weron (2006), our results indicate that price characteristics for Monday are different from those of other weekdays. There are two approaches in the literature to eliminate the intra-week seasonality; omitting the weekends and developing models only for weekdays (e.g. Bordignon et al., 2013) or introducing dummies for the weekends (and some of the days which have different characteristics from those of the rest of weekdays) (e.g. Weron and Misiorek, 2008). We follow the approach of the latter studies and introduce daily dummy variables for Saturday, Sunday and Monday.

Electricity prices also contain strong *seasonal fluctuations (annual cycle)*, reflecting the use of lighting and heating in winter and the growing use of air conditioning in summer. To consider the effects of different profiles on official and religious holidays (*calendar effects*), we include holiday dummies for each of these days. This approach is commonly used in the electricity price forecasting literature as an alternative to removing these days.

4.4. Selection and description of exogeneous variables

Historical market clearing prices are the lagged values of MCP which are determined by means of the autocorrelation function (ACF) and partial ACF (PACF) graphs of electricity prices with respect to time zones. *Forecasted demand* is the hourly demand forecasts published each day by Turkish Electricity Transmission Corporation (TEIAS) for the next physical day. The prices strongly reflect the level of demand with very high peaks in winter and summer, especially during high-demand load periods. In the literature, for the demand variable, there are different approaches. Some researchers (e.g., Bordignon et al., 2013) use forecasted demand published by the transmission system operator (TSO) and its lagged values, while others (Weron, 2006) use the forecasted demand by the TSO for the next day's demand and take the realized load values for the lagged values. There are three types of information published on the webpage of the National Load Dispatch Center: historical load data, forecasted load schedule and real-time load consumption. Since market participants are making their price forecasts based on forecasted demand and the merit order curves are aggregations of individual price expectations, we would expect there to be a higher correlation between price and forecasted demand. When we analyze the training data, as presented in Table 4, the correlation between price and forecasted demand is 0.604, which is slightly higher than the correlation between price and realized demand. Thus we choose to use forecasted demand data and its lagged values as one of the exogeneous variables. *Reserve Margin* is the difference between the demand forecast and the *Daily Production Program* (The generation values with respect to the settlement delivery point which a settlement aggregation entity (SAE) expects to realize in the following day and notifies the market operator at the beginning of day ahead balancing stage, according to the obligations of the balance responsible party to which the SAE is attached). It can be seen from Table 4 that there is a significant relationship between margin and MCP, with a correlation level of 0.151. *Natural gas prices* have been used as an exogenous variable especially for markets in which there are natural gas exchanges and the share of natural gas in electricity production is high. Some of the previous studies utilize the daily forward gas as the forecasted gas price. However, when we analyze the training data, the price of natural gas has increased very slightly. Also there is no natural gas market in Turkey, and the price of natural gas is regulated by the government rather than randomly determined by a liberalized market. Thus we do not use the price of natural gas as an exogenous variable.

5. Results

5.1. Data Partition

For model validation purposes, the whole dataset (Jan 1st, 2012–Dec. 31th, 2015) is divided into two parts. The first part, *training (in-sample) set*, covering the period Jan 1st, 2012–December 31th, 2014, is used only for regressor selection and model building. The remaining period, *test set*, (January 1st, 2015– December 30th, 2015) is used for *out-of-sample* forecast evaluation. Following the previous literature (e.g. Weron and Misiorek, 2008) all the series (clearing prices and exogenous variables) are considered using a logarithmic scale to obtain a more stable variance and have the opportunity to make our results comparable with the previous studies.

5.2. Experimental framework for comparing forecast accuracies

5.2.1. Accuracy measure

We adopt the most commonly used accuracy measure (Weron, 2014) in the electricity pricing literature; the mean absolute percentage error (MAPE);

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{p_t - \hat{p}_t}{p_t} \right| \tag{4}$$

to compare the results of our models since we do not have any negative prices in our dataset. It is important to note that we also got the results with Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error(MAE), however since results are similar, we omit them to save from space.

5.2.2. Criteria for comparing model performance

We compare performance of the models according to *three criteria*. In the first criteria we adopt an *ex-post* approach in which we use out-of-sample MAPE values to compare the models. In this criteria, we take ARMA as the benchmark model and compare the performance of the individual and ensemble models within themselves. In the *second criteria*, we compare the *ex-post* performance of the *best individual (BI)* model with that of the *best ensemble (BE)* based on their out-of-sample MAPE values. Although this criteria has been commonly used in the literature, it has also been criticized for not being a realistic forecasting setting. In the *third criteria*, we adopt a more realistic approach and decide the BI and BE based on in-sample MAPE values (*ex-ante*), then compare their performances using their out-of-sample MAPEs.

5.2.3. Test for statistical significance

To evaluate competing forecasting models, we use the *Diebold-Mariano* (DM) (Diebold and Mariano, 1995) test which verifies the existence of statistical significant difference between forecasting accuracy of two models. The test statistic is based on the loss differential d_t which we take in this paper as week-ahead (zonely) forecast errors. The two-sided DM test evaluates the accuracy of the competing forecasting models by testing the null hypothesis, (H_0:$d_t = 0$), indicating there is no statistical significant difference between the compared two forecasting models. The alternative hypothesis is that the two forecasts have different levels of accuracy.

5.3. Forecast accuracy comparison according to Criteria 1

5.3.1. Accuracy comparison of individual models

Our findings are summarized in Tables 5 and 6. Table 5 presents the summary of individual models with the lowest forecast errors according to *in-sample MAPE* values. NARX outperforms other competing individual models in most of the cases (96%). T3 and T1 tariff time zones have the highest and lowest MAPE values, respectively. This is an expected result and can be explained by T3 having the lowest long term correlation and T1 having the highest.

Table 6 illustrates summary of individual models with the lowest forecast errors according to *out-of-sample MAPE* values and their related DM test statistics according to Criteria-1. As can be observed from the table, especially for T2 and T3 series, dominant outperformance of NARX is replaced by ARMA-based and seasonality models. Similar to the results of Table 4, in most of the cases T3 has the highest MAPE values. Performance of ARMA is the one of the worst among the other competing individual models. Its MAPE value is the lowest in only 5 out of 48 cases (11%). NARX model has the lowest MAPE value in 14 out of 48 cases (30%) (mostly in T1 time zone), however accuracy difference is statistically significant in only 2% of them. ARMAX is the second best performer and outperformed the other individual models in 26% of the cases with a 5% ratio of significant cases. ETS, NAR, ARMA and TBATS are the following outperforming models with 15%, 12%, 11% and 6% outperforming cases respectively. In sum, according to DM test, ARMA can be outperformed significantly in 15% of the cases. Models with exogenous variables generally perform better than models without exogenous variables. This finding is inline with the previous literature and arises because of the high cross-correlation between demand, margin and price. Further this result also shows the appropriate selection of our input variables.

5.3.2. Accuracy comparison of ensemble models

Table 7 presents the summary of ensemble models with the lowest forecast errors according to *in-sample MAPE* values. ARMAXNARX outperforms other competing ensemble models in all of the cases (100%). T3 and T1 tariff time zones have the highest and lowest MAPE values, respectively. This is an expected result and can be explained by T3 having the lowest long term correlation and T1 having the highest.

Table 8 illustrates summary of ensemble models with the lowest forecast errors according to *out-of-sample MAPE* values and their related DM test statistics according to Criteria-1. As can be observed from the table, especially for T2 and T3 series, dominant outperformance of ARMAXNARX is replaced by ensemble seasonality models such as ARMAXETS and ARMAXTBATS. Similar to the results of Table 7, in most of the cases T3 has the highest MAPE values. According to DM test results, these accuracy differences are statistically significant in only in 9 out of 48 cases (19%). 45% of these cases are in the T3 zone indicating that for T3 price series, usage of ensemble seasonality models could be useful to increase forecast accuracy.

ARMAXNARX model has the lowest MAPE value in 12 out of 48 cases (25%) (mostly in T1 time zone), however accuracy difference is statistically significant in only 16% of them. ARMAXETS is the second

Table 5
Summary of individual models with best performance according to *in-sample MAPE*.

WEEK NO	T1 training	MAPE (in)	T2 training	MAPE (in)	T3 training	MAPE (in)	MAX MAPE (in)	MIN MAPE (in)
Winter1	NARX	0. 145	NARX	0.250	NARX	0.449	T3	T1
Winter2	NARX	0.149	NARX	0.257	NARX	0.467	T3	T1
Winter3	NARX	0.138	NARX	0.248	NARX	0.451	T3	T1
Winter4	NARX	0.133	NARX	0.237	NARX	0.445	T3	T1
Spring1	NARX	0.139	NARX	0.219	NARX	0.481	T3	T1
Spring2	NARX	0.135	NARX	0.240	NARX	0.453	T3	T1
Spring3	NARX	0.135	NARX	0.220	NARX	0.477	T3	T1
Spring4	NARX	0.130	NARX	0.223	NARX	0.476	T3	T1
Summer1	NARX	0.172	NARX	0.265	NARX	0.496	T3	T1
Summer2	NARX	0.178	NARX	0.275	NARX	0.593	T3	T1
Summer3	NARX	0.180	NARX	0.285	NARX	0.309	T3	T1
Summer4	NARX	0.180	NARX	0.279	NARX	0.310	T3	T1
Autumn1	NAR	0.074	NARX	0.325	NARX	0.636	T3	T1
Autumn2	NARX	0.192	NARX	0.312	NARX	0.660	T3	T1
Autumn3	NAR	0.123	NARX	0.315	NAR	0.182	T2	T1
Autumn4	NARX	0.180	NARX	0.285	NARX	0.676	T3	T1

Table 6
Summary of individual models with best model performance (*out-of-sample MAPE*) and Diebold-Marino statistics and p values (second row) according to Criteria-1.

WEEK NO	T1 Test	MAPE (OUT)	DMSTAT	T2 Test	MAPE (OUT)	DMSTAT	T3 Test	MAPE (OUT)	DMSTAT	MAX MAPE (out)	MIN MAPE (out)
Winter1	NARX	0.856	1.3283	NAR	0.728	0.900	**ARMA**	4.364	–	T3	T2
			0.232			0.402					
Winter2	NARX	0.966	1.039	NARX	1.107	0.252	**ARMA**	3.321	–	T3	T1
			0.338			0.809					
Winter3	ARMAX	0.632	**3.263****	ETS	0.994	0.612	ARMAX	1.307	1.927	T3	T1
			0.017			0.562					
Winter4	NAR	1.182	0.944	NARX	0.696	**2.246***	ARMA	1.277	− 0.102	T3	T2
			0.424			0.065					
Spring1	NAR	1.179	1.122	NARX	1.669	1.174	NARX	4.563	**2.727****	T3	T1
			0.304			0.284			0.034		
Spring2	NAR	1.272	0.789	NARX	1.189	0.627	NAR	7.832	0.824	T3	T2
			0.459			0.553			0.441		
Spring3	NARX	1.447	1.329	ETS	1.792	**3.079****	TBATS	4.789	− 0.949	T3	T1
			0.232			0.021			0.379		
Spring4	NARX	1.348	1.027	TBATS	1.463	1.810	NARX	8.898	0.150	T3	T1
			0.343			0.120			0.885		
Summer1	NARX	1.685	0.802	**ARMA**	0.897	–	ETS	4.971	1.672	T3	T2
			0.453						0.145		
Summer2	ARMAX	1.340	0.891	ARMAX	1.344	0.956	**ARMA**	4.851	–	T3	T1
			0.407			0.375					
Summer3	ARIMAX	1.413	1.149	NARX	1.136	− 1.221	ARMAX	4.355	**2.414***	T3	T2
			0.294			0.267			0.052		
Summer4	ETS	1.559	1.122	ETS	0.985	0.539	ETS	1.028	**1.943***	T1	T2
			0.304			0.609			0.099		
Autumn1	ARMAX	0.745	**2.360***	ARMAX	1.917	1.748	NAR	3.227	− 0.441	T3	T1
			0.056			0.130			0.674		
Autumn2	ARMAX	0.833	1.546	**ARMA**	3.109	–	ARMAX	3.819	0.901	T3	T1
			0.173						0.402		
Autumn3	ARMAX	1.181	1.232	ARMAX	2.203	0.156	NARX	3.313	0.095	T3	T1
			0.263			0.880			0.927		
Autumn4	NARX	3.203	1.131	TBATS	2.170	0.727	ETS	8.807	0.749	T3	T2
			0.300			0.494			0.481		

Note: *** indicates rejection of the null hypothesis at the 1% significance level. ** indicates rejection of the null hypothesis at the 5% significance level. * indicates rejection of the null hypothesis at the 10% significance level.

best performer and outperformed the other individual models in 21% of the cases with a 2% ratio of significant cases. ARMAXTBATS, ARMANAR, TBATSNAR and ETSNAR are the following outperforming models with 15%, 11%, 11% and 8% outperforming cases respectively. In sum, according to DM test with 5% significance level, ARMA can be outperformed significantly in 19% of the cases, mostly for the T3 time zone. Models with exogenous variables generally perform better than models without exogenous variables.

5.4. Forecast accuracy comparison according to Criteria 2

Our findings according to Criteria-2 are summarized in Table 9. In order to interpret the table, for example in T1-Winter1, '4–12' means comparing the BI model 4 with BE model 12. As can be observed from the table, the difference between the BI and BE is statistically significant in 9 out of 48 cases (19%). 70% of these significant cases are in the T2 zone. However BE model outperforms the BI in only 3 out of 9

Table 7
Summary of ensemble models with best model performance (according to *in-sample MAPE*).

WEEK NO	T1-training	MAPE (in)	T2-training	MAPE (in)	T3-training	MAPE (in)	MAX MAPE (in)	MIN MAPE (in)
Winter1	ARMAXNARX	0.605	ARMAXETS	0.049	ARMAXNARX	1.326	T3	T2
Winter2	ARMAXNARX	0.603	ARMAXNARX	0.715	ARMAXNARX	1.344	T3	T1
Winter3	ARMAXNARX	0.591	ARMAXNARX	0.710	ARMAXNARX	1.342	T3	T1
Winter4	ARMAXNARX	0.578	ARMAXNARX	0.704	ARMAXNARX	1.332	T3	T1
Spring1	ARMAXNARX	0.578	ARMAXNARX	0.698	ARMAXNARX	1.324	T3	T1
Spring2	ARMAXNARX	0.574	ARMAXNARX	0.698	ARMAXNARX	1.322	T3	T1
Spring3	ARMAXNARX	0.575	ARMAXNARX	0.700	ARMAXNARX	1.343	T3	T1
Spring4	ARMAXNARX	0.578	ARMAXNARX	0.704	ARMAXNARX	1.350	T3	T1
Summer1	ARMAXNARX	0.674	ARMAXNARX	0.780	ARMAXNARX	1.566	T3	T1
Summer2	ARMAXNARX	0.672	ARMAXNARX	0.783	ARMAXNARX	1.624	T3	T1
Summer3	ARMAXNARX	0.685	ARMAXNARX	0.791	ARMAXNARX	1.499	T3	T1
Summer4	ARMAXNARX	0.687	ARMAXNARX	0.791	ARMAXNARX	1.493	T3	T1
Autumn1	ARMAXNARX	0.696	ARMAXNARX	0.835	ARMAXNARX	1.745	T3	T1
Autumn2	ARMAXNARX	0.692	ARMAXNARX	0.839	ARMAXNARX	1.751	T3	T1
Autumn3	ARMAXNARX	0.700	ARMAXNARX	0.847	ARMAXNARX	1.762	T3	T1
Autumn4	ARMAXNARX	0.699	ARMAXNARX	0.834	ARMAXNARX	1.749	T3	T1

Table 8
Summary of ensemble models with best model performance (*out-of-sample MAPE*) and Diebold Marino statistic and p values (second row) according to criteria1.

WEEK NO	T1-Test	MAPE (out)	DMSTAT	T2_Test	MAPE (out)	DMSTAT	T3_Test	MAPE (out)	DMSTAT	MAX MAPE	MIN MAPE
Winter1	ARMAXTBATS	0.960	1.826 0.117	ARMANAR	0.876	-0.716 0.500	ARMAETS	4.900	-1.821 0.118	T3	T2
Winter2	ARMAXETS	0.418	1.876 0.109	ETSNAR	1.091	0.412 0.694	ARMANAR	3.573	-0.782 0.463	T3	T1
Winter3	ARMAXTBATS	0.753	**2.803**** 0.031	TBATSNAR	0.974	0.3602 0.731	ARMAXTBATS	1.567	**2.542**** 0.043	T3	T1
Winter4	ARMAXNARX	0.293	1.482 0.188	ARMAXNARX	0.650	**2.799**** 0.031	ARMANAR	0.914	1.165 0.287	T3	T1
Spring1	TBATSNAR	1.395	1.222 0.267	TBATSNAR	1.741	1.454 0.196	ETSNAR	6.038	**3.203**** 0.018	T3	T1
Spring2	ARMAXNARX	1.269	1.013 0.350	ARMANAR	1.165	0.847 0.429	ARMAXNARX	7.909	0.976 0.366	T3	T2
Spring3	ARMAXNARX	1.455	1.367 0.220	ARMAXETS	2.224	**2.090*** 0.081	ARMAXTBATS	4.795	0.373 0.721	T3	T1
Spring4	ETSNNAR	1.294	1.282 0.247	ARMAXETS	1.247	1.643 0.151	ETSNAR	8.345	0.721 0.497	T3	T2
Summer1	ARMAXNARX	1.192	0.991 0.359	ARMATBATS	1.139	-0.283 0.786	ARMAXETS	4.504	1.258 0.254	T3	T2
Summer2	TBATSNAR	1.121	1.155 0.291	ARMAXETS	1.818	0.392 0.708	ARMAXTBATS	8.806	**-7.670***** 0.0002	T3	T1
Summer3	ARMAXNARX	1.081	**2.169*** 0.073	ARMAXNARX	1.377	-0.212 0.838	ARMAXETS	6.334	1.161 0.289	T3	T1
Summer4	ARMAXETS	1.213	1.146 0.295	ARMAXETS	0.824	1.438 0.200	ETSNAR	1.453	**2.810**** 0.030	T3	T2
Autumn1	ARMAXTBATS	0.537	**2.589**** 0.041	ARMAXNARX	1.981	-0.447 0.670	TBATSNAR	3.699	-1.062 0.328	T3	T1
Autumn2	ARMAXTBATS	1.046	1.808 0.120	ARMATBATSETS	3.104	0.437 0.677	ARMAXETS	3.804	0.853 0.426	T3	T1
Autumn3	ARMAXNARX	1.202	1.236 0.262	ARMAXETS	2.181	0.501 0.634	ARMANAR	3.205	-0.406 0.698	T3	T1
Autumn4	ARMAXNARX	2.823	1.127 0.302	ARMATBATS	2.161	0.796 0.455	ARMAXNARX	8.289	0.634 0.549	T3	T2

Notes: *** indicates rejection of the null hypothesis at the 1% significance level. ** indicates rejection of the null hypothesis at the 5% significance level. * indicates rejection of the null hypothesis at the 10% significance level.

significant cases (30%). These results indicate that, looking from an ex-post performance evaluation, using ensemble models does not increase price forecast accuracy significantly. On the contrary for the peak time load time (T2), in most of the cases, BI models significantly outperforms BEs.

5.5. Forecast accuracy comparison according to Criteria 3

Our findings according to Criteria-3 are summarized in Table 10. As can be observed from the table, the difference between the BI and BE is statistically significant in 18 out of 48 cases (38%). 45% of these significant cases are in the T2 zone. BE model outperforms the BI in 14 out of 18 significant cases (75%). These results indicate that, looking from an ex-ante performance evaluation, using ensemble models increase price forecast accuracy significantly in 30% of the cases. In the remaining 70% of the cases, there is no statistical significant difference between forecast accuracies of BI and BEs. Thus for a power market participant, looking from a more realistic ex-ante decision making approach, using ensemble forecasting is useful to manage electricity price modeling risk.

6. Conclusion and suggestions for future work

In this paper we examine the relative forecasting performances of six promising individual models and ensemble models derived from them through equal weighting scheme in a semi-transparent electricity day ahead market. Although ensemble models are well studied in other contexts, the number of these studies are very limited for the electricity markets. Previous studies are from mature electricity markets (UK, NordPool and PJM) and variety of individual methods used for the ensemble models in them are limited. Following the main findings of these studies; namely performance of ensemble models are different for different market conditions, time periods and model combinations; we develop well-performing ensemble forecast models for an emerging market, Turkey, and compare their performance with alternative individual models according to three benchmark criteria. This enables us to address three forecasting challenges which have been encountered in the previous ensemble modeling studies, specifically the inadequate handling of trend seasonality component and varying market power over the course of a day, the inappropriate selection of exogenous variables when one considers the relationships that exist between price and system constraints, and the methodological similarity of the individual models they have considered.

We addressed the following research question: *How to manage price modeling risk via ensemble forecasting in the Turkish electricity day ahead auctions?*

We compare performance of the models according to three criteria. In the first criteria, we take ARMA as the benchmark model and compare the performance of the individual and ensemble models within themselves according to both in-sample and out-of-sample MAPE values. In the second criteria, we compare the ex-post performance of the best individual (BI) model with that of the best ensemble (BE) based on their out-of-sample MAPE values. In the third criteria, we adopt a more realistic approach and decide the BI and BE based on in-sample MAPE values (ex-ante), then compare their performances using their out-of-sample MAPEs.

According to Criteria-1, regarding the individual models, we find that although NARX outperforms other competing individual models in most of the cases (96%) with respect to in-sample MAPE; for out-of-sample MAPE, dominant outperformance of NARX is replaced by ARMA-based and seasonality models. However, according to DM test with 5% significance level, ARMA can be outperformed significantly in

Table 9
Summary of ensemble models with best model performance (*out-of-sample MAPE*) and Diebold Marino statistic and p values (second row) according to criteria2.

WEEK NO	T1	DMTEST	T2	DMTEST	T3	DMTEST
Winter1	4–12	-0.536	3–7	**-3.693*****	1–9	-1.821
		0.610		0.000		0.118
Winter2	4–10	1.881	4–13	-0.539	1–7	-0.782
		0.108		0.590		0.463
Winter3	2–12	-0.696	5–14	0.806	2–12	-0.665
		0.512		0.421		0.530
Winter4	3–8	**2.521****	4–8	0.015	1–7	1.165
		0.045		0.987		0.287
Spring1	3–14	-0.888	4–14	**-3.452*****	4–13	-1.926
		0.408		0.001		0.102
Spring2	3–8	0.839	4–7	1.6159	3–8	0.752
		0.433		0.110		0.480
Spring3	4–8	-0.909	5–10	**-5.712*****	6–12	0.692
		0.398		0.001		0.514
Spring4	4–13	-0.195	6–10	**4.451*****	4–13	1.220
		0.851		0.001		0.268
Summer1	4–8	0.011	1–11	-0.602	5–10	0.712
		0.991		0.552		0.502
Summer2	2–14	0.624	2–10	-1.229	1–12	**-7.670*****
		0.555		0.264		0.000
Summer3	2–8	1.007	4–8	**2.767****	2–10	**-2.718****
		0.352		0.0325		0.034
Summer4	5–10	1.113	5–10	0.668	5–13	-0.225
		0.308		0.528		0.829
Autumn1	2–12	0.734	2–8	**-2.192****	3–13	-1.436
		0.469		0.037		0.201
Autumn2	2–12	-0.616	1–15	0.437	2–10	-0.952
		0.56		0.677		0.377
Autumn3	2–8	-0.522	2–10	0.227	4–7	0.021
		0.619		0.827		0.984
Autumn4	4–8	0.417	6–11	1.314	5–8	0.205
		0.690		0.236		0.843

Note: Model no (1)ARMA (2)ARMAX (3)NAR (4)NARX (5) ETS (6)TBATS (7)ARMANAR (8)ARMAXNARX (9)ARMAETS (10)ARMAXETS (11)ARMATBATS (12)ARMAXTBATS (13)ETSNAR (14)TBATSNAR (15)ARMATBATSETS (16)ARMATBATSNAR (17)HYBRIDALL.
*** indicates rejection of the null hypothesis at the 1% significance level. ** indicates rejection of the null hypothesis at the 5% significance level. * indicates rejection of the null hypothesis at the 10% significance level.

Table 10
Summary of ensemble models with best model performance (*in-sample MAPE*) and Diebold Marino statistic and p values (second row) according to criteria3.

WEEK NO	T1	DMTEST	T2	DMTEST	T3	DMTEST
Winter1	4–8	-1.2854	4–10	**6.405*****	4–8	-1.247
		0.246		0.001		0.258
Winter2	4–8	1.2982	4–8	**-7.252*****	4–8	-0.129
		0.241		0.001		0.901
Winter3	4–8	**4.1369*****	4–8	**6.819*****	4–8	0.766
		0.006		0.001		0.472
Winter4	4–8	**2.1307***	4–8	0.015	4–8	**1.992***
		0.077		0.987		0.093
Spring1	4–8	-0.727	4–8	**-3.585*****	4–8	**-2.610****
		0.494		0.001		0.040
Spring2	4–8	0.272	4–8	**-5.305*****	4–8	-0.265
		0.794		0.001		0.799
Spring3	4–8	-0.90	4–8	-1.402	4–8	1.843
		0.398		0.165		0.115
Spring4	4–8	0.011	4–8	-0.116	4–8	-0.065
		0.991		0.908		0.950
Summer1	4–8	0.011	4–8	**2.373****	4–8	0.234
		0.991		0.025		0.822
Summer2	4–8	1.873	4–8	**2.767****	4–8	**4.004*****
		0.110		0.032		0.007
Summer3	4–8	1.458	4–8	**2.767****	4–8	**2.878****
		0.195		0.032		0.028
Summer4	4–8	1.796	4–8	1.272	4–8	**4.779*****
		0.122		0.250		0.003
Autumn1	3–8	**2.684****	4–8	1.220	4–8	**3.089****
		0.012		0.232		0.021
Autumn2	4–8	0.684	4–8	1.256	4–8	1.501
		0.519		0.255		0.183
Autumn3	3–8	**2.736****	4–8	0.762	3–8	0.387
		0.034		0.474		0.711
Autumn4	4–8	0.417	4–8	1.001	4–8	0.261
		0.690		0.355		0.802

Note:Model no (1)ARMA (2)ARMAX (3)NAR (4)NARX (5) ETS (6)TBATS (7)ARMANAR (8)ARMAXNARX (9)ARMAETS (10)ARMAXETS (11)ARMATBATS (12)ARMAXTBATS (13)ETSNAR (14)TBATSNAR (15)ARMATBATSETS (16)ARMATBATSNAR (17)HYBRIDALL.
*** indicates rejection of the null hypothesis at the 1% significance level. ** indicates rejection of the null hypothesis at the 5% significance level. * indicates rejection of the null hypothesis at the 10% significance level.

only 15% of the cases. Regarding the ensemble models, considering in-sample MAPE, ARMAXNARX outperforms other competing ensemble models in all of the cases (100%). For out-of-sample MAPE, dominant outperformance of ARMAXNARX is replaced by ensemble seasonality models such as ARMAXETS and ARMAXTBATS. According to DM test with 5% significance level, ARMA can be outperformed significantly in only 19% of the cases, mostly for the T3 time zone. According to Criteria-2, the difference between the BI and BE is found to be statistically significant for 19% of cases. 75% of these significant cases are in the T2 zone. These results demonstrate that, looking from an ex-post performance evaluation, using ensemble models does not increase price forecast accuracy significantly. According to Criteria-3, which is the most realistic decision making setting among the other two criteria, we find that the difference between the BI and BE is statistically significant for 30% of the cases. BE model outperforms the BI for 65% of these significant cases. These results indicate that, looking from an ex-ante performance evaluation, using ensemble models increase price forecast accuracy significantly in 30% of the cases. In the remaining 70% of the cases (after subtracting BI significant cases), there is no statistical significant difference between forecast accuracies of BI and BEs. Thus for a power agent, looking from a more realistic ex-ante decision making approach, using ensemble models can be very useful to manage price modeling risk. Thus our findings support the additional benefits of ensemble forecasts especially according to an ex-ante (more realistic) decision making setting and in line with the previous findings indicating ensemble modeling is less uncertain and more accurate than the ex-ante best individual model.

6.1. Policy implications

For energy regulators and policy makers, using ensemble models can be useful to manage electricity price modelling risk for ex-ante policy impact assessment and lead to better policy decisions.

Electricity price forecasts are used by energy regulators as one of the main input variables for ex-ante policy impact assessment (Shahidehpour et al., 2002). In forecasting, uncertainity is reflected in the forecast error and the source of risk arises from the unobservability of full information set underlying the individual forecasts which could be differently affected by statistical properties of the related price series (Timmermann, 2006). Therefore using ensemble forecasting could mitigate this risk related to decision making of a policy maker (Bunn, 1985).

Considering fractal dynamics of price could improve decision making of policy makers.

We find that for both individual and ensemble models, in most of the cases T3 (night tariff time zone) has the highest MAPE values. This can be explained by the lower predictability level of this price series compared to T1 and T2, and the stronger impact of seasonality factors on demand during the night. Thus if fractal dynamics of price is prominently different for some time zones during the course of a day, ensemble modeling is less risky than individual models in terms of the risk of selection of an inappropriate individual forecasting model. Further, choosing ensemble models based upon the fractal dynamics of each time zone could improve policy maker's forecast accuracy.

Energy regulators could enhance predictability level of prices, especially for off-peak load periods by increasing transparency level of the market through disseminating data on primary resource based available installed capacity and planned generation schedules

We find that models with exogenous variables generally perform better than models without exogenous variables. This finding is in line with the previous literature and arises because of the high cross-correlation between demand, margin and price. Further this result also shows the appropriate selection of our input variables. On the other hand we find that T3 price series has lowest level of predictability and no long term correlation indicating that marginal bidders bid at their marginal costs (Sapio, 2004). This means if an energy regulator wants to enable power agents to forecast the prices accurately for off-peak hours, it needs to publish prior information on *primary resore based available installed capacity* (the active power capacity that a generation unit can provide to the system) and *final daily production program* (firm-level) in order to give a signal of the possible future supply stack for each hour and technology of the marginal generator.

6.2. Suggestions for future work

In emerging markets, from which most of them are semi-transparent, since market data is very limited for price forecasting, market participants tend to utilize from international data provider consultancy firms to get information which is not public. The type of data they seek for is usually on the planned generation schedules of large firms, dam level of some state-owned hydro power plants (for Turkey case), primary resorce based available installed capacity and final daily production program (firm-level) and matching quantities(ex-post) on the hourly merit order curve. We call this type of data private (not insider) since it does not include any firm-level *confidential* data. Since most of the market participants do not have this information, the ones that have it can make a more fundamental analysis of the market, improve their forecast accuracies and beat the market. Based on this point of view, it would be interesting to examine the worth of this information to market participants considering the improvement in forecast accuracy of models (both individual and ensemble).

Appendix

See Tables A.1 and A.2.

Table A.1
Data related to day ahead auctions in Turkey. Note: Settlement aggregation entity (UEVCB): Active electric energy generating or consuming entities which are defined by market participants so that the settlement calculations can be made for each market participant.

Class	Type	Name	Abbreviation	Definition
Generation	Planning	Daily production program	DPP	The generation values with respect to settlement delivery point which a settlement aggregation entity anticipates to realize in the following day and notifies the market operator at the beginning of day ahead balancing stage, according to the obligations of the balance responsible party to which it is attached.
		Final daily production program	FDPP	The generation or consumption values which a settlement aggregation entity expects to realize in the following day and notifies the system operator at the opening of the balancing power market according to the obligations of the balance responsible party to which it is attached and the result of day ahead balancing
		Available installed capacity	AIC	The active power capacity that a generation unit can provide to the system
	Real-time	Real-time generation	RGT	Hourly resource-based power generation
	Ex-post	Injection quantity (settlement volume)	IQ	Hourly injection of aggregate energy of the injection units (UEVCB)[1] through a settlement period (In terms of the settlement delivery point, the reading values of the meters in the settlement aggregation entity configuration will be used as a basis in settlement calculations)
Demand	Forecast	Demand forecast	DemFor	The hourly consumption forecasts published on a daily basis for the next physical day
	Real-time	Consumption		Hourly real-time consumption
	Ex-post	Withdrawal quantity		Hourly withdrawal of aggregate energy of withdrawal units through a settlement period
		Eligible customer withdrawal quantity (settlement withdrawl volume)		–
		Withdrawal quantity under supply liability		–
Price	Planning	Interim market clearing price	IMCP	Temporary hourly energy price which is determined within the objection period with respect to bids that are cleared according to total supply and demand
	Real-time	Market clearing price	MCP	Hourly energy price that is determined with respect to bids that are cleared according to total supply and demand
	Ex-post	Trade value		Hourly aggregate of cleared bids and offers
		Price independent offer		Hourly aggregate of offer quantity at 0 TL/MwH
		Price independent bid		Hourly aggregate of bid quantity at 2000 TL/MwH
		Supply/demand	MOC	Hourly merit order curve
		Bid quantity		Hourly and block bid(purchase) quantity at 0 TL/MwH
		Offer quantity		Hourly, block and flexible offer(sale) quantity at 2000 TL/MwH
		Block bids		Aggregate quantity of cleared block bid that covers a minimum of 4 to a maximum of 24 h
		Block offers		Aggregate quantity of cleared block bid that covers a minimum of 4 to a maximum of 24 h
		Matching quantity		Hourly aggregate quantity of cleared bids

Table A.2
Dissemination details for day ahead auction data.

Name	Responsible Institution	Period of Time	Publish Time	Detail	Public from
DPP	TEIAS	Hourly	D-1	Firm[a]-level	Aggregate-level: 2011 Firm-level: 2016
FDPP	TEIAS	Hourly	D-1	Firm-level	Aggregate-level: 2011 Firm-level: 2016
AIC	TEIAS	Hourly	D-1	Firm-level	Aggregate-level: 2011 Firm-level: 2016
RGT	TEIAS	Hourly	Real-time	Primary-resource-based	2016
IQ	TEIAS	Hourly	D + 1	Primary-resource-based	Aggregate-level: 2011 Primary-resource-based:2016
Demand Forecast	TEIAS	Hourly	D-1	Aggregate	2011
Consumption	TEIAS	Hourly	Real-time	Aggregate	2016
Withdrawal quantity	TEIAS	Hourly	D + S	Aggregate	2016
Interim Market Clearing Price	EXIST	Hourly	D-1	Aggregate	2016
Market Clearing Price	EXIST	Hourly	D-1	Aggregate	2011
Trade value	EXIST	Hourly	D-1	Aggregate	2011
Price independent offer	EXIST	Hourly	D-1	Aggregate	2016
Price independent bid	EXIST	Hourly	D-1	Aggregate	2016
Supply/Demand	EXIST	Hourly	D-1	Aggregate	2014
Bid quantity	EXIST	Hourly	D-1	Aggregate	2016
Offer quantity	EXIST	Hourly	D-1	Aggregate	2016
Block bids	EXIST	Hourly	D-1	Aggregate	2016
Block offers	EXIST	Hourly	D-1	Aggregate	2016
Matching Quantity	EXIST	Hourly	D-1	Aggregate	2014

[a] Firm represents the settlement aggregation entity in the day ahead auctions.

References

Aggarwal, S.K., Saini, L.M., Kumar, A., 2009. Electricity price forecasting in deregulated markets: a review and evaluation. Int. J. Electr. Power Energy Syst. 31, 13–22.
Akram, M., Hyndman, R., Ord, J., 2009. Exponential smoothing and non-negative data. Aust. N.Z. J. Stat. 51 (4), 415–432.
Altavilla, C., De Grauwe, P., 2010. Forecasting and combining competing models of exchange rate determination. Appl. Econ. 42 (27), 3455–3480.
Amjady, N., Hemmati, M., 2006. Energy price forecasting. IEEE Power Energy Mag. 20–29.
Anderson, C.L., Davison, M., 2008. A hybrid system-econometric model for electricity spot prices: considering spike sensitivity to forced outage distributions. IEEE Trans. Power Syst. 23 (3), 927–937.
Avci-Surucu, E., Aydogan, A.K., Akgul, D., 2016. Bidding structure, market efficiency and persistence in a multi-time tariff setting. Energy Econ. 54, 77–87.
Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. J. Econ. 74, 3–30.
Bates, J.M., Granger, C.W., 1969. The combination of forecasts. J. Oper. Res. Soc. 20 (4), 451–468.
Bordignon, S., Bunn, D.W., Lisi, F., Nan, F., 2013. Combining day-ahead forecasts for British electricity prices. Energy Econ. 35, 88–103.
Borenstein, S., 2001. The Trouble with Electricity Markets (and Some Solutions). University of California Energy Institute.
Bunn, D.W., 1975. A Bayesian approach to the linear combination of forecasts. Oper. Res. Q. (1970–1977) 26, 325–329.
Bunn, D.W., 1977. A comparative evaluation of the outperformance and minimum variance procedures for the linear synthesis of forecasts. Oper. Res. Q. (1970–1977) 28, 653–662.
Bunn, D.W., 1985. Statistical efficiency in the linear combination of forecasts. Int. J. Forecast. 1 (2), 151–163.
Bunn, D.W., Farmer, E.D. (Eds.), 1985. Comparative Models for Electrical Load Forecasting. Wiley.
Bunn, D.W., Karakatsani, N., 2003. Forecasting electricity prices. Working Paper, London Business School.
Bunn, D.W., 2004. Modelling prices in competitive electricity markets. Wiley.
Cartea, A., Figueroa, M., Geman, H., 2009. Modelling electricity prices with forward looking capacity constraints. Appl. Math. Financ. 16 (2), 103–122.
Catalao, 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. Electr. Power Syst. Res. 77, 1297–1304.
Chen, D., Bunn, D.W., 2010. Analysis of the nonlinear response of electricity prices to fundamental and strategic factors. IEEE Trans. Power Syst. 25, 595–606.
Chen, Z., Yang, Y., 2007. Time series models for forecasting: testing or combining? Stud. Nonlinear Dyn. Econ. 11 (1).
Clark, T.E., McCracken, M.W., 2009. Combining forecasts from nested models. Oxf. Bull. Econ. Stat. 71, 303–329.
Clemen, R.T., 1989. Combining forecasts: a review and annotated bibliography. Int. J. Forecast. 5, 559–583.
Crane, D.B., Crotty, J.R., 1967. A two-stage forecasting model: exponential smoothing and multiple regression. Manag. Sci. 13 (8) (B-501).
Davison, M., Anderson, C.L., Marcus, B., Anderson, K., 2002. Development of a hybrid model for electrical power spot prices. IEEE Trans. Power Syst. 17 (2), 257–264.
De Livera, A.M., Hyndman, R.J., Snyder, R.D., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. J. Am. Stat. Assoc. 106 (496), 1513–1527.
De Menezes, L.M., Bunn, D.W., Taylor, L.W., 2000. Review of guidelines for the use of combined forecasts. Eur. J. Oper. Res. 120, 190–204.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13, 253–263.
Eydeland, A., Wolyniec, K., 2003. Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging 206 John Wiley & Sons.
Gaillard, P., Goude, Y., Nedellec, R., 2016. Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. Int. J. Forecast. 32 (3), 1038–1050.
Gao, F., Cao, X., Papalexopoulous, A., 2000. Forecasting power market clearing price and quantity using a neural network method. IEEE Power Eng. Soc. Summer Meet. 4, 2183–2188.
Geman, H., Roncoroni, A., 2006. Understanding the fine structure of electricity prices. J. Bus. 79.
Georgilakis, P.S. (2006). Market clearing price forecasting in deregulated electricity markets using adaptively trained neural networks. Lecture notes in artificial intelligence LNAI 3955. Berlin, Heidelberg, Springer-Verlag, 2006, pp. 56–66.
Glachant, J.M., Saguan, M., 2007. An institutional frame to compare alternative market designs in EU electricity balancing.
Guo, J., Luh, P.B., 2003. Selecting input factors for clusters of Gaussian radial basis function networks to improve market clearing price prediction. IEEE Trans. Power Syst. 18 (2), 655–672.
Guo, J., Luh, P.B., 2004. Improving market clearing price prediction by using a committee machine of neural networks. IEEE Trans. Power Syst. 19 (4), 1867–1876.
Guthrie, G., Videbeck, S., 2002. Electricity spot price dynamics: beyond financial models. Energy Policy 35, 5614–5621.
Harris, C., 2006. Electricity Markets: Pricing, Structures and Economics. Wiley, Chichester.
Haugom, E., Ullrich, C.J., 2012. Forecasting spot price volatility using the short-term forward curve. Energy Econ. 34, 1826–1833.
Hayfavi, A., Talasli, I., 2014. Stochastic multifactor modeling of spot electricity prices. J. Comput. Appl. Math. 259, 434–442.
Hibon, M., Evgeniou, T., 2005. To combine or not to combine: selecting among forecasts and their combinations. Int. J. Forecast. 21 (1), 15–24.
Hong, T., 2015. Crystal Ball Lessons in Predictive Analytics. Spring, EnergyBiz, pp. 35–37.
Huisman, R., Huurman, C., Mahieu, R., 2007. Hourly electricity prices in day-ahead markets. Energy Econ. 29, 240–248.
Hyndman, R.J., Koehler, A.B., Snyder, R.D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. Int. J. Forecast. 18 (3), 439–454.
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 Econ. 38, 96–110.
Karakatsani, N.V., Bunn, D.W., 2008. Forecasting electricity prices: the impact of fundamentals and time-varying coefficients. Int. J. Forecast. 24 (4), 764–785.
Karakatsani, N.V., Bunn, D.W., 2010. Fundamental and behavioral drivers of electricity price volatility. Stud. Nonlinear Dyn. Econ. 14 (4).

Keles, D., Scelle, J., Paraschiv, F., Fichtner, W., 2016. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. Appl. Energy 162, 218–230.
Ketter, W., Collins, J., Gini, M., Gupta, A., Schrater, P., 2009. Detecting and forecasting economic regimes in multi-agent automated exchanges. Decis. Support Syst. 47 (4), 307–318.
Ketter, W., Collins, J., Gini, M., Gupta, A., Schrater, P., 2012. Real-time tactical and strategic sales management for intelligent agents guided by economic regimes. Inf. Syst. Res. 23 (4), 1263–1283.
Knittel, C.R., Roberts, M.R., 2005. An empirical examination of restructured electricity prices. Energy Econ. 27, 791–817.
Kölmek, M.A., Navruz, İ., 2013. Forecasting the day-ahead price in electricity balancing and settlement market of Turkey by using articial neural networks. Turk. J. Electr. Eng. Comp. Sci. 23, 841–852.
Lildholdt, P., 2000. Long Memory and ARFIMA Modelling of Inflation Rates (M.Sc. Dissertation). Department of Economics, University of Aarhus.
Lin, T., Horne, B.G., Tino, P., Giles, C.L., 1996. Learning long-term dependencies in NARX recurrent neural networks. IEEE Trans. Neural Netw. 7 (6), 1329–1338.
Lisi, F., Nan, F., 2014. Component estimation for electricity prices: procedures and comparisons. Energy Econ. 44, 143–159.
Lucia, J.J., Schwartz, E.S., 2002. Electricity prices and power derivatives: evidence from the Nordic power exchange. Rev. Deriv. Res. 5, 5–50.
Maciejowska, K., 2014. Fundamental and speculative shocks, what drives electricity prices? In: IEEE Conference Proceedings-EEM14.
Maciejowska, K., Nowotarski, J., 2016. A hybrid model for GEFCom2014 probabilistic electricity price forecasting. Int. J. Forecast. 32 (3), 1051–1056.
Maciejowska, K., Nowotarski, J., Weron, R., 2016. Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. Int. J. Forecast. 32 (3), 957–965.
Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R., 1982. The accuracy of extrapolation (Time Series) methods: results of a forecasting competition. Int. J. Forecast. 1, 111–153.
Makridakis, S., Winkler, R.L., 1983. Averages of forecasts: some empirical results. Manag. Sci. 29 (9), 987–996.
Makridakis, S., Wheelwright, S.C., Hyndman, R.J., 1998. Forecasting: methods and Applications. Wiley, New York.
Makridakis, S., Hibon, M., 2000. The M3-competition: results, conclusions and implications. Int. J. Forecast. 16, 451–476.
Mandal, P., Senjyu, T., Urasaki, N., Funabashi, T., 2006. A neural network based several hour-ahead electric load forecasting using similar days approach. Electr. Power Energy Syst. 28, 367–373.
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 (No. HSC/14/11). Hugo Steinhaus Center, Wroclaw University of Technology.
Misiorek, A., Trueck, S., Weron, R., 2006. Point and interval forecasting of spot electricity prices: linear vs. non-linear time series models. Stud. Nonlinear Dyn. Econ. 10 (3), 2.
Nan, F., 2009. Forecasting Next-day Electricity Prices: From Different Models to Combination (Ph.D. Thesis). Universita degli Studi di Padova.
Nowotarski, J., Raviv, E., Trück, S., Weron, R., 2014. An empirical comparison of alternative schemes for combining electricity spot price forecasts. Energy Econ. 46, 395–412.
Nowotarski, J., Weron, R., 2016. On the importance of the long-term seasonal component in day-ahead electricity price forecasting. Energy Econ. 57, 228–235.
Ord, J.K., Koehler, A.B., Snyder, R.D., 1997. Estimation and prediction for a class of dynamic nonlinear statistical models. J. Am. Stat. Assoc. 92, 1621–1629.
Ozyildirim, C., Beyazit, M.F., 2014. Forecasting and modeling of electricity prices by radial basis functions: Turkish electricity market experiment. Iktisat Islet. Finans 29 (344), 31–54 (2014).
Petrella, A. and Sapio, S. (2009). A time series analysis of day-aheadprices on the italian power exchange. In 6th International Conference on theEuropean Energy Market, 2009., pages 1–6. IEEE.
Pino, R., Parreno, J., Gomez, A., Priore, P., 2008. Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks. Eng. Appl. Artif. Intell. 21 (1), 53–62.
Ramanathan, R., Engle, R., Granger, C.W., Vahid-Araghi, F., Brace, C., 1997. Short-run forecasts of electricity loads and peaks. Int. J. Forecast. 13 (2), 161–174.
Reid, D.J., 1968. Combining three estimates of gross domestic product. Economica 35, 431–444.
Reid, D.J., 1969. A Comparative Study of Time Series Prediction Techniques on Economic Data (Ph.D. Thesis). University of Nottingham, Nottingham.
Sapio, S., 2004. Markets design, bidding rules, and long memory in electricity prices. Rev. deconomie Ind. Program. Natl. Perse. 107 (1), 151–170.
Shahidehpour, M., Yamin, H., Li, Z., 2002. Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management. Wiley.
Smith, D.G., 1989. Combination of forecasts in electricity demand prediction. J. Forecast. 8 (3), 349–356.
Smith, J., Wallis, K.F., 2009. A simple explanation of the forecast combination puzzle. Oxf. Bull. Econ. Stat. 71 (3), 331–355.
Snyder, R., Koehler, A., Ord, J., 2002. Forecasting for Inventory Control with Exponential Smoothing. Int. J. Forecast. 18 (1), 5–18.
Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. J. Forecast. 23 (6), 405–430.
Taylor, J.W., Majithia, S., 2000. Using combined forecasts with changing weights for electricity demand profiling. J. Oper. Res. Soc. 51, 72–82.
Taylor, J.W., 2003. Short-term electricity demand forecasting using double seasonal exponential smoothing. J. Oper. Res. Soc. 54, 799–805.
Taylor, J.W., 2010. Triple seasonal methods for short-term electricity demand forecasting. Eur. J. Oper. Res. 204, 139–152.
Taysi, Z.C., Biricik, G., Bozkurt, O.O., 2015. Evaluation of price forecast systems for Turkish Electric Market. In: Proceedings of the IEEE Signal Processing and Communications Applications Conference (SIU),pp. 620–623.
Timmermann, A.G., 2006. Forecast Combinations. Handbook of Economic Forecasting 1. Elsevier, Amsterdam, pp. 135–196.
Unlu, K.D., 2012. Temperature in Turkey and Turkish day Ahead Electricity Market Prices: Modeling and Forecasting (M.Sc. Thesis). Financial Mathematics, Institute of Applied Mathematics, Middle East Technical University.
Vahidinasab, V., Jadid, S., Kazemi, A., 2008. Day-ahead price forecasting in restructured power systems using artificial neural networks. Electr. Power Syst. Res. 78 (8), 1332–1342.
Von der Fehr, N.H.M., 2013. Transparency in electricity markets (No. 13/2013). Memorandum, Department of Economics, University of Oslo.
Weron, R., 2006. Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach. Wiley.
Weron, R., Misiorek, A., 2008. Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models. Int. J. Forecast. 24, 744–763.
Weron, R., 2014. Electricity price forecasting: a review of the state-of-the-art with a look into the future. Int. J. Forecast. 30 (4), 1030–1081.
Yıldırım, M.H., Özmen, A., Bayrak, Ö.T., Weber, G.W., 2012. Electricity Price Modelling for Turkey. Operations Research Proceedings 2011. Springer, pp. 39–44.
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 Trans. Power Syst. 20 (1), 59–66.