

Forecasting The Italian Day-Ahead Electricity Price Using Bootstrap Aggregation Method

Murad Harasheh

University of Pavia, Department of Economics and Management
Research Partner: The Italian Authority for Energy Regulations

doi: 10.19044/esj.2016.v12n28p51 [URL:http://dx.doi.org/10.19044/esj.2016.v12n28p51](http://dx.doi.org/10.19044/esj.2016.v12n28p51)

Abstract

Electricity price forecasting has become a crucial element for both private and public decision-making. This importance has been growing since the wave of deregulation and liberalization of the energy sector on a global scale since the late 1990s. Given these facts, this paper is an attempt to establish and demonstrate a precision based applicable forecasting model for wholesale electricity prices with respect to the Italian power market on an hourly basis. Artificial intelligence models such as neural networks and bagged regression trees are utilized, although they are rarely used to forecast electricity prices. After model calibration, bagged regression trees with exogenous variables comprised the final model. The selected model outperformed neural network and bagged regression with a single price used in this paper, it also outperformed other statistical and non-statistical models used in other studies. We also confirm certain theoretical specifications of the model. As a policy tool, this model could be used by energy traders, transmission system operators and energy regulators for an enhanced decision-making process.

Keywords: PUN, artificial intelligence models, regression tree, bootstrap aggregation, forecasting error

Introduction

The course of deregulation and the emergence of competitive electricity markets, witnessed towards the end of the 20th century, have been restructuring the power sectors which were traditionally controlled by governments as monopolists. In countries where power sectors were more liberal, market conditions assumed the role of the central government policies for trading various electricity products. Furthermore, energy research and practitioners continue to debate the categorization of electricity as a good or a service. Nevertheless, we propose that electricity, as a product,

is distinguished from other products in the sense that it is not physically storable, and thus the real-time balancing between generation and consumption is needed to keep the system in equilibrium. There are factors driving electricity demand and perhaps weather conditions such as, temperature and wind, as well as daily activities during the business hours and weekends are important factors that determine the composition of daily demand. Daily demand can be further categorized as base demand, peak and off-peak. As a consequence, these special characteristics of electricity demand lead to some phenomena, or rather, anomalies, in the behavior of electricity prices not witnessed in other markets, such as temporal seasonality and sudden spikes are among the major behaviors observed stemming from electricity pricing. In this manner, these facts have been encouraging energy research to intensify the endeavor for developing accurate price forecasting models, (Weron 2014; Kaminski, 2013; Shahidehpour, Yamin, & Li, 2002).

Furthermore, electricity price forecasting has become a crucial element for business decision-making as well as for public policy-making since energy security is a vital issue in all societies, (Weron 2014; Bunn, 2004; Eydeland & Wolyniec, 2003). Additionally, experience shows that energy is one of the most sensitive sectors in the economy since cost pass-through to final customers has many social and political implications as shown by California crisis of 2000-2001, (Joskow, 2001). Electric companies are, in this regard, very vulnerable to bankruptcies as trading any extra/shortage of energy as needed might result in significant economic losses. In electricity markets, hedging against volumes and price movements was mainly motivated by the extreme electricity price variations, especially during peak hours. Nevertheless, hedging positions in electricity markets can't be achieved without the temporal and on-going price forecast for different time frequencies ranging from half an hour to several months or years ahead. This forecast serves as the core interest of market participants on both demand and supply sides, as well as for investment managers in power sectors to continuously modify their actual production/consumption or for a temporal adjustment of financial portfolios. Along those lines, and based on the aforementioned attributes, the motives for this research reside under three main pillars as follows:

- 1- The emergence of a new set of forecasting models using machine algorithm.
- 2- These models are not widely tested on the Italian market, thus verifying their accuracy is needed.
- 3- The important changes in the wholesale electricity markets in Europe in general and Italy with its surrounding countries, such as regulatory

changes, market coupling, the growing impact of renewables and the distributed generation.

Our study contributes to the existing literature by investigating and verifying on of the increasingly growing group of “artificial intelligence” models, in particular, bootstrap aggregation of regression trees. This model is not widely explored in energy studies despite demonstrating enhanced forecasting. Furthermore, few forecasting papers are conducted on the Italian energy market and this is the first time that this particular model is being applied with respect to such a market. Finally, we provide detailed breakdowns of forecast errors by hour, hour type (peak and off-peak), day and by month which are rarely provided in other studies.

Before proceeding, the remaining sections of this paper are comprised in the following manner; part two is the review of the related literature; part three concerns the model and the bagging technique; part four is dedicated for data, model calibration and model specifications; part five for the forecasting results; and finally part six concerns conclusions and policy implications.

Literature Review

The inspiration of the liberalization of power sectors worldwide came as a consequence of liberalization in other sectors, during which privatization showed growing efficiencies and economic values of nations’ wealth. However, as indicated earlier, electricity is not like any other commodity, it is immediately produced as a response for demand, making real-time balancing and price forecasts important elements in electricity markets. In this context, wholesale energy markets emerged as a consequence of the deregulation and the liberalization of power sectors. Wholesale exchanges can be broadly categorized as a power pool in which the participation of market agents is compulsory as in Nordic EU countries, and power exchange with a voluntary involvement as in continental EU countries, with the latter being a more common type of power market. Trading in power exchanges is mostly concentrated on day-ahead spot products. Spot power markets work on a full efficiency to meet the full-time functioning, constituting hourly auctions for 24 hours a day, seven days a week and 365 days a year. Moreover, the dynamics and the special characteristics of this type of market makes price forecasting, from few hours ahead to several days, of high importance for market participants, (Girish et al., 2013; Girish & Vijayalakshmi, 2013).

Regarding the outcome of the power market, a market clearing price is finalized as a result of the intersection between the market aggregate demand and supply curves corresponding to each trading hour and each trading zone, and since the supply curve seems to be flat for base demand

stake, the electricity price, therefore, is less sensitive to changes in demand in that stake of the curve, (Girish & Vijayalakshmi, 2013).

Furthermore, the shape of the supply curve is simply the ranking of all power plants based on their marginal cost and their flexibility to be run as needed. As demand for electricity starts to increase, more expensive power plants have to be switched on to meet the extra demand driving the supply curve to be steeper. Normally, in this case, generation plants such as natural gas and refined oil plants with higher variable costs, are called upon to generate the extra electricity requested driving the marginal cost of production to be higher. Therefore, a small change in electricity demand in this area of the curve leads to a significant change in electricity price since the price is set as the marginal cost of the plant that met the required demand. Additionally, these plants have to charge higher prices to compensate themselves for the expensive fuel used and for the few hours of work. This situation occurs normally during the peak hours, and when demand starts to revert back to its base shape, the expensive generation plants are turned off. Furthermore, other factors might have significant effects on electricity price such as sudden outages, maintenance, network congestion and the exercise of market power by power market participants, (Girish et al., 2013; Weron, 2006; Misiorek et al., 2006; Bunn and Martoccia, 2005; Stoft, 2002; Mielczarski, Michalik & Widjaja, 1999).

In the following part, we concentrate on the classifications of electricity price forecasting models (EPF) as indicated in different studies. Depending on the purpose of the forecasting, different modeling and pricing techniques might be utilized. Generally, long-term forecasting serves investments spanning broader time horizons and strategic planning, medium-term forecasting is appropriate for general financial positions, derivatives and risk management, whereas short-term forecasting applies for daily/hourly auctioning and trading in power markets, (Misiorek et al., 2006; Girish et al., 2013).

Weron, 2014 classifies forecasting models into five main groups, agents-based, fundamental, reduced-form, statistical, and artificial intelligence models. Generally, these are the main forecasting groups, but can also be classified according to different criteria such as dynamic/static, linear/nonlinear, parametric/non-parametric and deterministic/stochastic. Additional classifications such as optimization models, equilibrium models, agent-based models, and artificial intelligence models or game theory models, simulation models and the time-series models as in (Aggarwal et al., 2009) can also be used. Table (1) summarizes the EPF models as in (Weron, 2014), however, we add the last type in computational intelligence models (bootstrap aggregation of regression trees).

Table 1: Electricity price forecasting models

Multi-agent	Electricity price models			
	Fundamental	Reduced-form	Statistical	Computational intelligence
Nash-Cournot framework	Parametric fundamental	Jump-diffusion	Similar-day exponential smoothing	Feed-forward neural network
Supply function equilibrium	Parsimonious structural	Markov regime-switching	Regression models	Recurrent neural network
Strategic production cost			(AR)(MA)(X)	Fuzzy neural network
Agent-based			Threshold AR	Support vector machine
			(G)(AR)(CH)	Bagged regression tree

Source: Self reproduction (Weron, 2014)

The groups of models are summarized as follows as in (Weron, 2014):

- Multi-agent models are based on simulating the behaviors of heterogeneous agents in the power market, and then construct a pricing process by matching the demand and supply in the market.
- Fundamental methods try to look for important fundamental variables that might explain electricity price dynamics.
- Reduced-form models look at the statistical features of temporal electricity prices for the purpose of derivatives valuation and risk management.
- Statistical approaches utilize various econometric models ranging from time series with single price to various types of regression models with exogenous variables.
- Artificial intelligence techniques combine elements of learning, evolution and fuzziness to create approaches that are capable of adapting to complex dynamic systems.

 (Girish et al., 2014; Hong, 2014; Aggarwal, Mohan & Kumar, 2009; Karakastani & Bunn, 2008), among others show different classification of energy forecasting models, they summarize many studies from different countries using various models with varying forecasting accuracy measured in different ways, nevertheless they all agree that until now there is no clear evidence that a certain model is superior to others. In fact there are pros and cons for each model or set of models, however, what is clear is that some models are more appropriate in certain cases i.e., some of them are more appropriate for price forecast rather than for volume forecast, some are good

to forecast short-term, while others are better in forecasting medium or long-term values.

In this regard and depending on the type of model and the target forecast (load or price); various forecasting errors might be utilized to assess how good the forecasting model is. The most common measures are MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), MASE (Mean Absolute Scaled Error), Theil's U index, and RMSE (Root Mean Square Error).

With respect to the Italian market, very few studies use different models for electricity pricing at the wholesale level. Gianfreda & Grossi, 2013, use a combined statistical model of ARFIRMA-GARCH with exogenous variables to forecast zonal electricity prices and then the PUN simulation, they measure predictive goodness by RMSE, MAPE, and Theil's U-index. Guerci, Rastegar, & Cincotti, 2010, propose an agent-based computational model. Particularly, they investigate the relationship between the strategic behavior of the thermal power plants and the level of the price at a zonal and national levels. Their model replicates exactly the market clearing process in the day-ahead market and the Italian high-voltage transmission network with its zonal subdivision. Moreover, Beccali et al., 2007, utilize a specific artificial intelligence model, Neural Networks, to forecast the short-term household electricity consumption related to a suburban area in Sicily.

Various methods and models have been applied to electricity price forecasting (EPF), with varying degrees of success depending on the country, economic cycle, time evolution, and other micro/macro-structure variations. In this study, the focus is on the last type of computational intelligence models, bootstrap aggregation or bagged regression trees. Our study contributes to the existing literature by investigating and verifying on of the rapidly growing set of models "artificial intelligence models", specifically, bootstrap aggregation of regression trees which is not widely explored in energy aspects though it shows a relatively better forecasting power. Furthermore, few forecasting papers are conducted on the Italian energy market and this model is the first time to be tested on such a market.

Our objective is trying to come up with a precise and a flexible forecasting model for the hourly wholesale electricity price in Italy which might be utilized by the regulatory authorities and utility providers for a better decision-making process.

The Italian Context

Since the end of 1962, the whole electricity market was under the monopoly of a single vertically integrated and state-owned company Enel, providing generation, transmission, and distribution, with the exception of

few cities where the local municipalities serve as the electricity provider. Moreover, the liberalization process, started early 1990s by Bersani's decree, which brought about the replacement of old administrative procedures with market mechanisms more suited to the new framework for power trade, and consequently, the Italian Power Exchange was created (AEEG, 2008).

The Italian Power Exchange (IPEX) is the main instrument through which the execution of the requirements provided by law no. 79/1999 for the implementation of the free market. It is an organized system that is able to promote both competition relating to the production and sale of electricity, and the protection of end customers through a maximum transparency ensured through the coordination of the unified electricity market entrusted to the Electricity Market Operator (GME)². The sale of electricity is carried out daily by resorting to bargaining on an hourly basis and for each Geographic Area (North, Central North, Central, South, Sicily, and Sardinia) to the Single National Price (PUN).

The Italian Power Exchange was established in Italy as of 1 April 2004, as the platform whereby the offers of producers meet consumers' demand and electricity wholesalers. The exchange commenced with a traded volume of 232 TWh in 2004 reaching 282 TWh in 2014 with a peak volume of 337 TWh achieved in 2008. Regarding the structure of the market, IPEX consists of three different platforms, the Spot Electricity Market (MPE) is a platform for physical delivery; the Forward Electricity Market (MTE) was launched on November 2008; and the platform for physical delivery of financial derivatives (CDE) negotiated on the financial exchange (IDEX).

Regarding the market mechanisms, the MPE is divided into Day-Ahead Market (MGP), Intra-Day Market (MI), and Services Market (MSD). The Day-Ahead Market (MGP) is the venue for the majority of transactions for the sale of electricity and in which power is exchanged for the next day. It opens at 8:00 AM for nine days before the delivery day until 9:00 AM of the day before the day of delivery when the MGP closes. Here operators shall submit their offers in which they define the amount and the maximum and minimum price at which they are willing to buy and sell electricity.

The Intra-Day Market (MI) allows operators to modify the schedules defined in the Day-Ahead Market through further purchase or sale. MI takes place in four sessions (MI1, MI2, MI3 and MI4) with different opening and closing times.

The Services Market (MSD) is the instrument through which Terna S.p.a (the transmission system operator) procures the resources necessary for the management and the control of the system (congestion resolution intra-

² GME (*Gestore Mercati Energetici*) is a company wholly owned by the Ministry of Economy and Finance and officially the only body responsible for managing physical energy trading platforms in Italy.

zonal, creation of energy reserve, real-time balancing). In the MSD, Terna acts as a central counterparty and offers accepted are remunerated at the price presented (pay-as-bid), (GME, 2015; Geanfreda, 2012).

Recently, European energy markets and the Italian market in particular, are undergoing a period of transformation (transitional regulatory period) which will reshape the structure of energy markets along the lines of integrity, transparency and security for the energy sector in Europe. Perhaps the most noticeable structural change is the market coupling among different EU neighboring nations. Market coupling was derived from the European policy in establishing a European internal electricity market for the purpose of achieving a better capacity allocation, more efficient congestion management, security of supply, transparency, liquidity, and finally the convergence of electricity prices in the coupled zones. Italy and Slovenia – among others- established a coupling agreement in January 2011 and ultimately evolving into a tangible and official market coupling in February 2015. Since then the two markets operate in an integrated way and utilize a common matching algorithm for quantity allocation and price determination.

The other noticeable structural change in the European energy sector is the financialization of energy sector activities under the new European financial reforms under which a considerable part of energy derivative contracts will be regulated under EU financial regulations.

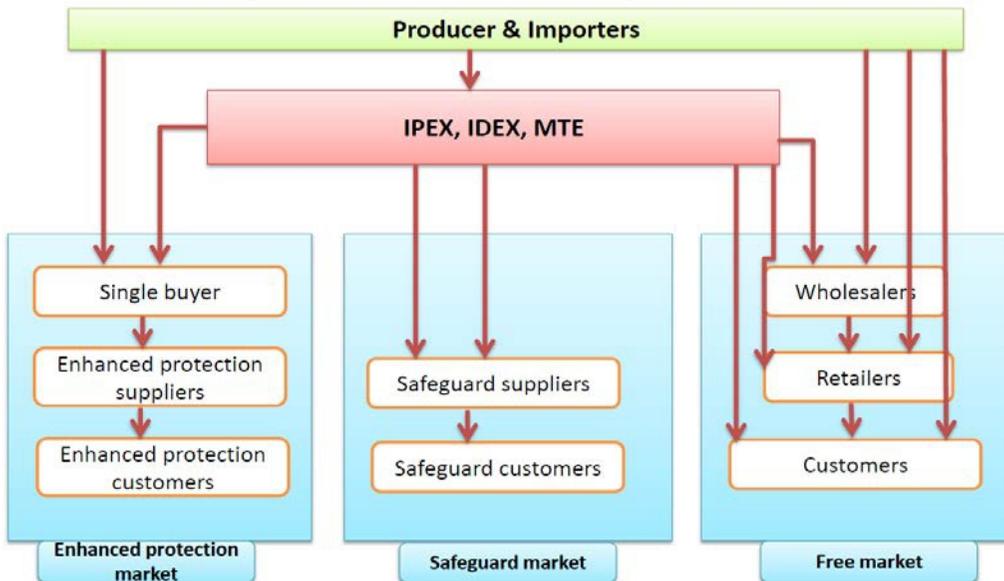


Figure 1: The structure of the Italian Power Market, (Source: Acquirente Unico, 2008)

On the wholesale side, bilateral contracts and power exchange capture the vast majority of wholesale transactions. Whereas on the retail side, customers are classified as free market customers or eligible customers

and the protected segment (*standard offer market*³) that will be phased out by 2018.

The model

In the model description we start from the broad group of intelligence models, then to regression trees and finally to bootstrap aggregation (bagging) of regression trees.

According to Weron, 2014; Aggarwal et al., 2009, in general, there is no common consensus of what Artificial Intelligence (AI) means, and it may not mean the same thing for all people. We can broadly say that the failure of other forecasting models, statistical in particular, to adapt for complex systems was the main drivers for the spread of AI models. AI models have the added value over other models to adapt to complex dynamic settings and to handle non-linear relations by creating approaches using learning elements, fuzzy logic, and evolution. And with no doubt this added value makes them excellent for short-term price modeling and forecasting as indicated by different authors. Nonetheless artificial neural networks (ANN) have received a great attention in empirical studies for load and price forecasting, other dynamic techniques are also used in hybrid structures such as fuzzy logic, genetic algorithms, evolutionary programming, swarm intelligence and bagging.

In Breiman et al., 1984, regression trees (RTs) or decision trees are designed to predict the value of a numeric variable. RTs use learning elements for building prediction models from a sample data. The data space is recursively split and a simple fitting model is created for each partition, then the final model is obtained by aggregating all partitions together. Regression trees are represented in hierarchical structures (*as shown in figure 2*), which contains nodes that represent the tests on the input elements, branches that correspond to the results of the test, and the prediction for the value of the target attribute is stored in a leaf. Regression tree leaves contain constant values as predictions for the target variable.

In the same regard, Bootstrap Aggregation, or rather Bagging, was introduced by Breiman as a technique for improving the prediction of regression trees. Since then it has been gaining momentum in response to the increased interest in bootstrapping and other processes in statistical modeling. Bagging works on creating multiple similar datasets, re-running the tree analysis, and then collecting the aggregate of the results and re-calculating the tree and associated statistics based on this aggregate (Breiman, 1998;1996). Notwithstanding Bagging is mainly applied to

³ Standard offer market is the option that guarantees consumers the energy supply with reference to a standard price set by the regulatory authority and is revised each semester, this market will be phased out by 2018 under the Italian energy regulations.

regression tree as an accuracy enhancement technique of predictions, it can also be applied to other methods such as neural networks. Furthermore, it is worth noting that Bagging is mainly very useful for unstable regression being bagged, means that, when small changes in the learning sample can simply result in significant variations in the predictions obtained using a specified method, bagging can result in an appreciable reduction in average prediction error (Sutton, 2005).

Tree bagging algorithm

The training algorithm applies the general technique of bootstrap aggregating to tree learners to improve the predictive performance of tree models. Bagging works as follows, (ETH Zurich, 2008):

1- Create a bootstrap sample

$$(X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*) \tag{1}$$

Then compute the bootstrapped estimator $\hat{g}^*(\cdot)$.

2- Repeat step 1 B times, yielding

$$\hat{g}^{*1}(\cdot), \dots, \hat{g}^{*B}(\cdot). \tag{2}$$

3- Aggregate the bootstrap estimates

$$\hat{g}_{Bag}(\cdot) = B^{-1} \sum_{i=1}^B \hat{g}^{*i}(\cdot). \tag{3}$$

The term $\hat{g}_{Bag}(\cdot)$ in the model is our electricity price estimate PUN.

Bagging algorithm can also be considered an approximation

$$\hat{g}_{Bag}(\cdot) \approx \mathbb{E}^* [\hat{g}^*(\cdot)] \tag{4}$$

Which can be improved by increasing B . the original point is that we should use $\mathbb{E}^* [\hat{g}^*(\cdot)]$ as a new estimator.

A trivial identity hints at some properties of bagging: write (the theoretical version of bagging with $B = \infty$)

$$\begin{aligned} \hat{g}_{Bag}(\cdot) &= \hat{g}(\cdot) + (\mathbb{E}^* [\hat{g}^*(\cdot)] - \hat{g}(\cdot)) \\ &= \hat{g}(\cdot) + \text{bootstrap bias estimate} \end{aligned} \tag{5}$$

Instead of subtracting the bootstrap bias estimate, we are adding it! What we can hope for is a variance reduction at the price of a higher bias. This turns out to be true if $\hat{g}(\cdot)$ is a tree-based estimator.

It can be shown that for tree-based estimators $\hat{g}(\cdot)$,

$$Var(\hat{g}_{Bag}(x)) \stackrel{asympt}{<} Var(\hat{g}(x)), \tag{6}$$

for very many x . Thus, bagging is a variance reduction technique. The reason for this is that a bagged tree turns out to be a product of probit functions $\Phi(d - \cdot)$ instead of indicator functions $\mathbf{1}_{[\cdot \leq d]}$. This causes a variance reduction at the price of some bias. For example,

$$Var(\mathbf{1}_{[x \leq d]}) = \mathbb{P}[X \leq d](1 - \mathbb{P}[X \leq d]). \tag{7}$$

If $X \sim N(0,1)$ and $d = 0$, the latter quantity equals $\frac{1}{4}$. On the other hand,

$$Var (\Phi(-X)) = Var (U) = \frac{1}{12}, \quad U \sim Unif. ([0,1]), \quad (8)$$

Which reduced the variance by the factor 3!. We should use large trees for bagging, because the variance reduction due to bagging asks for a large tree to balance the bias-variance trade-off.

The key point in bagging to reduce the variance without affecting the bias is that the estimates of a single tree are very sensitive to noise, whereas the average aggregated value of many trees is not a noise-sensitive given that the trees are not correlated. Bootstrap sampling is a way of de-correlating the trees by showing them different training sets as a non-deterministic approach. Furthermore, cross-validation might be used to determine the optimal number of trees B even it is a free parameter. Normally, a number of trees ranging from few hundreds to thousands can be used depending on the size of the training set. B can also be determined by observing the out-of-bag error. The training and test error tend to level off after some number of trees have been fit (Hastie et al., 2008; Breiman, 2001).

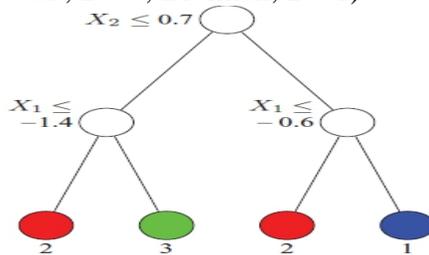


Figure 2: Regression tree. (Source, Wei-Yen Loh, 2011)

- Leaf is the lowermost node on the tree also called a terminal node.
- Node is a variable that significantly contributes to model explanatory power.
- Recursive Partitioning is the method by which RT operates – subsequently dividing the data into smaller portions to isolate most important variables.
- Root is the top most node on the tree – the most important explanatory variable.
- Splitting (Purity) Criterion: In RT some criteria need to be used to determine the “best” primary way to split the data. While there are different criteria available, the Gini⁴ impurity criterion is most widely used. In RT, this process of splitting makes the tree more “pure.”

⁴ Used by the CART (classification and regression tree) algorithm, Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset. Gini impurity can be computed by summing the probability f_i of each item being chosen times the probability $1 - f_i$ of a mistake in categorizing that item. It reaches its minimum (zero) when all cases in the node fall into a single target category.

Figure (3) demonstrates the accepted range of leaves in the RT.

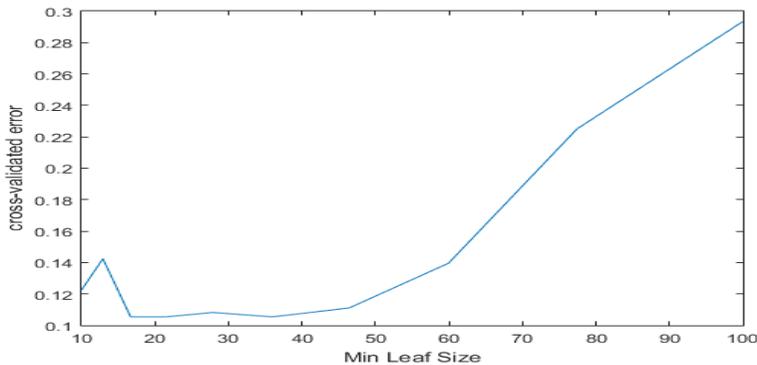


Figure 3: Minimum leaf size. (Source, [ww.mathworth.com](http://www.mathworth.com))

As shown in figure (3), the best leaf size is between 20 and 50 observations per leaf.

Data and model calibration

We gather hourly data for the wholesale energy market for load, electricity price and natural gas price. Load and gas price are traditionally used as strong determinants of electricity price. Our data sources are the Italian authority of energy and the GME (Gestore dei Mercati Energetici).

Moreover, the data is gathered on hourly basis for 2012, 2013, 2014, and the first half of 2015. Each year consists 8,760 hours (except 2012 was a leap year which consisted one additional day making 8,784 hours) resulting in 30,408 observations for the given period. For the data sets, no missing values are registered, however, one zero-value for the day-ahead price was registered in 2013, and given the large number of observations this value would not affect the forecasting performance.

For the purpose of performing the model, we divide the data into a training set which consists 26,304 observations from 2012 to 2014, and a test data set which consists 4,104 observation representing the first half of 2015 until 20/06/2015. The training set is used to build the model by applying the appropriate algorithm of the model then fitting current observations to future ones (estimating its parameters). Then when the system sets the model with defined parameters. The test set is used only for forecasting to test the performance of the model on out-of-sample data.

In this regard, we utilize various artificial intelligence models to fit the same data to be tested on the first half of 2015, then we select the model that minimizes errors of the forecast to generate future predictions of the second half of 2015 starting from 21/06/2015 to 21/12/2015 consisting 4,656 hours. All model calibration and analysis were performed using MATLAB 15.a. Table (2) demonstrates the division of model samples.

Table 2: Model samples

Sample	Observations
Fitting and training 2012 to 2014	26,304
Testing 1 st half of 2015	4,104
Predict 2 nd half of 2015	4,656

Before we move to the calibration, it is worth providing the various forecast errors used in this study:

(1) Mean Absolute Error

$$MAE = \frac{1}{24} \sum_{n=1}^{24} |a_t - f_t| \quad (9)$$

(2) Mean Absolute Percentage Error

$$MAPE = \frac{1}{24} \sum_{n=1}^{24} \frac{|a_t - f_t|}{a_t} \quad (10)$$

(3) Theil' U index: Measured in terms of Theil's U2. The U2 statistic takes the value 1 under the naïve forecasting method. Values less than 1 indicate greater forecasting accuracy than the naïve forecasts, values greater than 1 indicate the opposite.

$$U2 = \sqrt{\frac{\sum_{t=1}^{n-1} \left(\frac{f_{t+1} - a_{t+1}}{a_t} \right)^2}{\sum_{t=1}^{n-1} \left(\frac{a_{t+1} - a_t}{a_t} \right)^2}} = \frac{RMSE \text{ Forecast}}{RMSE \text{ Naive}} \quad (11)$$

(4) Peak MAPE

$$\text{Peak MAPE} = \text{MAPE for price at peak} \quad (12)$$

Where in equations (9), (10) and (11) a is the actual price for a certain hour, and f is the corresponding forecasted price for the same hour.

In this regard, the calibration of the forecasting models is conducted on three phases as follows:

Phase one: we select different artificial intelligence models as shown in table (3), in this regard, regression models of neural network and bagged regression are built to forecast the day-ahead price given the training data, the models are then used on the test data to validate its accuracy.

For the neural network, we initially started with a default one of two layers with 20 neurons, and after calibrations for different layers and neurons, the defaults networks performed better using the default Levenburg-Marquardt algorithm.

For the tree bagger, a set of regression trees each with different set of rules for performing the non-linear regression. We initially build an aggregate of 20 such trees, with a minimum leaf size of 20. The larger is the leaf size the smaller the tree is. This provides a control for over fitting and performance, and Breiman's random forest algorithm is applied as shown in the algorithm section.

Table 3: Trained models

Model	MAE	MAPE (%)	Theil's U	Peak MAPE (%)
Regression trees with exogenous variables	1,83	3,87	0,360	2,29
Regression trees without exogenous variables	2,0	4,13	0,389	2,59
Neural networks with exogenous variables	2,44	5,1	0,475	3,16

The performance shows that the first model “bagged regression trees” with exogenous variables produces the least error forecast as shown in bold. Additionally, it is worth to notice that neural networks model did not perform well in our case with respect to bagging technique. Equations (12) and (13) show the input variables for our models. Equation (12) is applied to the first model of bagged regression and to the neural network, in both models we utilize the same set of exogenous variables (multivariate equation).

Elec. Price =

$$f(\text{date, hour, weekday, workday, weather data, prev. load, prev. elec. price, gas price}) \tag{12}$$

Meanwhile, the second model of bagged regression without exogenous variables utilizes only the historical series of electricity price with dummy variables, and equation (13) represents the model. The frequency of previous prices (previous week, days, and hour) and the dummies are used to capture the seasonality effect on prices. In addition to load, gas and dummies affecting the energy price, numerous studies have illustrated the influence of weather variables on electricity price. The approaches so far employed for price load forecasting include statistical models or econometric models, engineering models, structural models, fuzzy logic and expert systems (Beccali, 2007).

It is worth to mention here that the values of variables including prices are taken without transformation because machine learning, random forest for classification in particular, are used for high-dimensional data, and we are interested in the predictive distribution of the response variable, rather than just the conditional mean, moreover, machine learning concerns classifications which does not require variable transformation, (Kuhn & Johnson, 2016)

$$Elec. Price = f(\text{date, hour, weekday, workday, prev. elec. price}) \tag{13}$$

Input data for the multivariate equation:

- 1- *Date*
- 2- *Hours of the day*
- 3- *Is week day or not*
- 4- *Is working day or not*
- 5- *Dry bulb temperature*
- 6- *Dew point*

- 7- *Load: measured in MWh*
 - Current Load
 - Previous Week Same Hour Load
 - Previous Day Same Hour Load
 - Previous 24 hours Average Load
- 8- *Electricity price: measured in €/MWh*
 - Previous Week Same Hour Price
 - Previous Day Same Hour Price
 - Previous 24Hr Average Price
- 9- *Gas Price: measured in €/MWh*
 - Previous Day Gas Price
 - Previous Week Average Gas Price

Phase two: the second phase of the model calibration process is to select the first bagged regression trees with exogenous variables model and simulate it for a different number of leaves of the tree. It is worth mentioning that the number of leaves has an influence on the forecasting accuracy, and as mentioned earlier that the optimal number of leafs range from 20 to50. Table (4) show the results of the model simulation:

Table 4: Bagged regression tree with diff. leaves

Model	MAE	MAPE (%)	Theil's U	Peak MAPE (%)
Regression trees with 20 leafs	1,92	3,99	0,385	2,39
Regression trees with 25 leafs	1,83	3,87	0,360	2,29
Regression trees with 30 leafs	1,91	3,96	0,379	2,41

Table (4) clearly shows that forecasting errors change by changing the number of tree leaves. A tree with 25 leafs produces the least forecasting errors with respect to other trees.

Phase three: in this phase we select the bagged regression tree with 25 leaves and then we simulate for different forecasting terms (horizon of the forecast). We use week ahead forecast, month ahead forecast, quarter ahead forecast, and semi-annum ahead forecast, and in all cases the frequency of the forecast is the hourly PUN. Table (5) summarizes the results of the horizon forecast simulation.

Table 5: Bagged regression with diff term forecast

Term forecast	MAE	MAPE (%)	Theil's U	Peak MAPE (%)
Week-ahead forecast	2.22	4.21	0,425	8.23
Month-ahead forecast	2.07	4.02	0,412	4.26
Quarter-ahead forecast	1.93	3.98	0,401	3.67
6-month-ahead forecast	1.83	3.87	0,360	2.29

Table (5) shows that as we move to longer horizons, forecasting performance improves, and the 6-month ahead forecast produces better results in terms of forecasting errors compared to other term forecasts.

Empirical Findings

Descriptive Statistics

Table (6) demonstrates the basic descriptive statistics for the main variables used in this paper. We can clearly notice the declining trend of load and gas price which leads to a decline in the PUN price, this might be due to different facts such as the economics crises dated back to 2011, less severe winters in the last couple of years, or increased regulatory, consumption and management efficiencies of energy, the growth of renewables, and finally the sharp decline in oil prices internationally. Additionally, the standard deviation is decreasing as well supporting the premise of increasing efficiencies in regulation, consumption, and management.

Table 6: Descriptive statistics

All sample (2012-2015)					
	Mean	Max	Min	STD	N
PUN (€/MWh)	61,7	324,2	0,0	20,6	30.408
Load (MWh)	32.523	50.394	17.961	6.741	30.408
Gas (€/MWh)	26,2	65,0	17,4	3,5	30.408
2012					
PUN	75,5	324,2	12,1	22,2	8.784
Load	33.534	50.394	19.387	6.928	8.784
Gas	28,7	65,0	25,9	2,7	8.784
2013					
PUN	63,0	151,9	0,0	18,2	8.760
Load	32.557	47.164	18.692	6.720	8.760
Gas	28,0	34,0	25,6	1,4	8.760
2014					
PUN	52,1	149,4	2,2	15,6	8.760
Load	31.783	46.106	18.065	6.441	8.760
Gas	23,3	29,3	17,4	3,1	8.760
2015 (until 20/06/2015)					
PUN	50,1	100,8	5,6	10,7	4.104
Load	31.864	46.075	17.961	6.726	4.104
Gas	23,5	30,2	21,8	1,6	4.104

PUN and load data are directly obtained from power market platforms while gas price is referred to PSV Italian gas hub price (Punto di Scambio Virtuale).

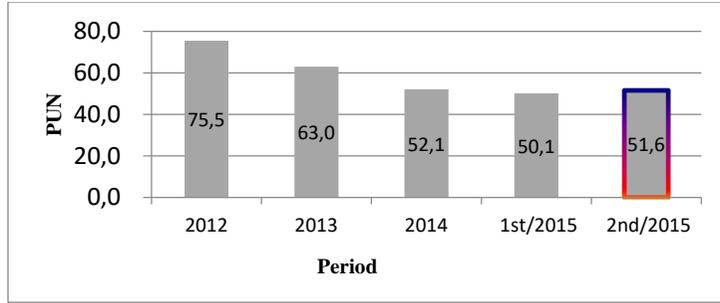


Figure 4: PUN trend

As shown in figure (4), the average PUN exhibits a negative trend in the last 3 years. The average of the first half of 2015 is €50.1 and based on our forecast, the average of second half PUN is €51.6 producing an annual average of €50.85 being 2.4% less than 2014 PUN.

Model Results

The output of our model is the hourly PUN forecast for the other 194 days of the years (second half of 2015) starting from June 21, 2015, classified as daily, weekly, and monthly, for each 24 hours a day. The following graph summarizes the price profile of the model forecast.

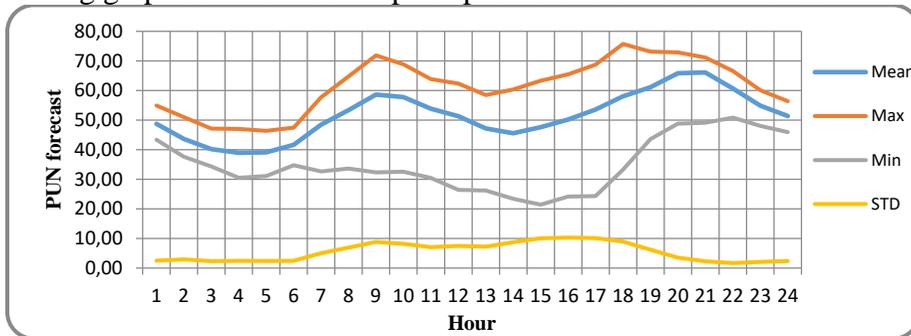


Figure 5: Forecasting results

Figure (5) demonstrates the basic statistics for each of 24 hours averages for the 194 days of forecast. First of all, the price profiles resemble peak and off-peak price movement. Moreover, the average price forecast is higher in the second half of 2015, the price ranges between €21.4 and €75.7. Additionally, peak hours experience higher standard deviation as well as higher prices compared to other types of hours as we will see in next part of the analysis. The penetration of renewables can partially explain the price variance and min-max spread given that if the demand is met by renewables during off-peak hours, this drags the price down and stretching the difference between peak-off peak prices.

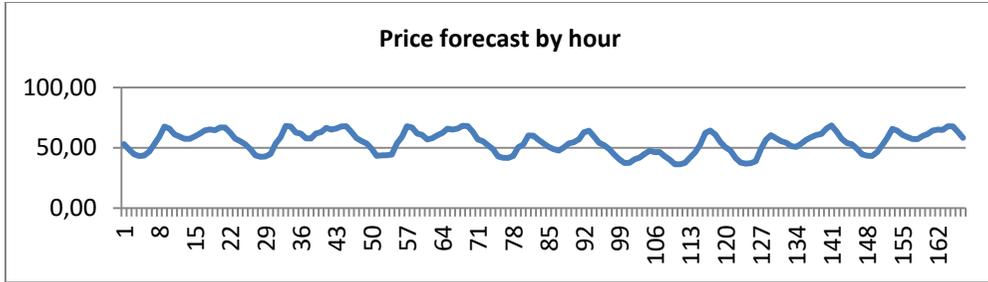


Figure 6: Price forecast for the first week of July 2015

Figure (6) shows the s selected sample of the day-ahead price forecast for the first week of July 2015 which corresponds to 168 hours.

Forecasting accuracy

The forecasting accuracy is measured by MAE, MAPE, and MAPE at peak. Figure (8) and the upper part of figure (7) together show how the forecasting error is normally distributed around zero, and the mean value of the error is very close to zero, which makes the model appropriate for forecasting purposes. Additionally, the lower parts of figure (7) demonstrate the distributions of both MAE and MAPE on the tails of error distribution. We can, moreover, notice that MAE is less than 2 and MAPE is less than 5 which are clarified numerically in table (3).

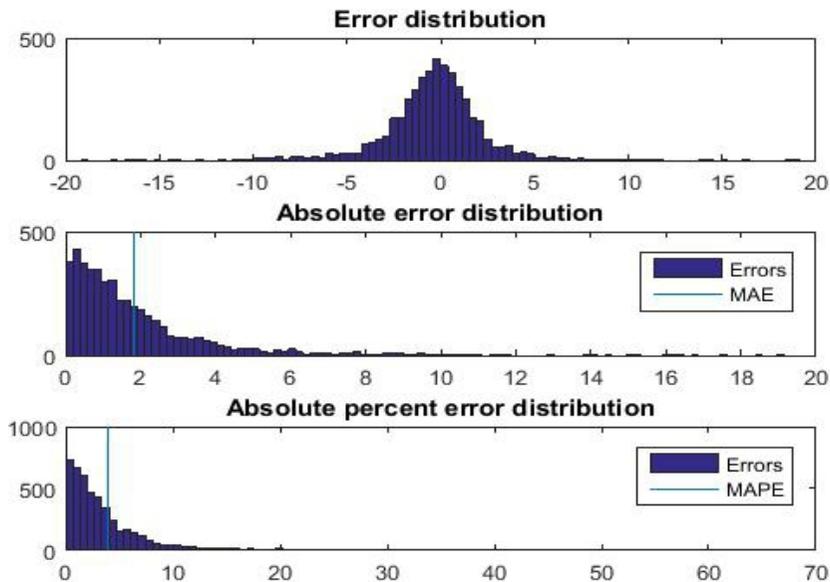


Figure 7: Error distribution

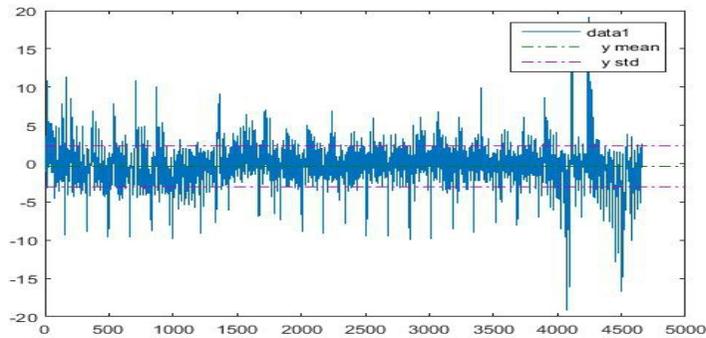


Figure 8: Error statistics

Breakdown of the forecasting error

In this part, we show how the forecasting error changes with respect to the frequency of the forecast. Figure (9) shows the breakdown of the forecasting error for each hour of the day. We can clearly notice that peak hours (from 9:00 to 19:00) have a higher forecasting error. This result confirms the official reports published by the Italian Power Exchange and the Italian Energy Regulator, however, it was difficult to compare our result with other studies since our paper is –to our knowledge- the only to provide this type of detailed analysis. As we mentioned earlier, growth in renewables might partly explain forecasting errors at peaks hours, we say partly given the fact that the renewables share in the Italian generation mix is not as high compared to Germany or to the cheap nuclear generation in France.

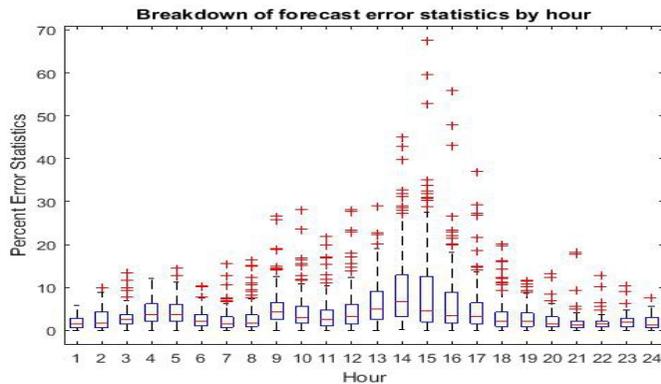


Figure 9: Forecast error by hour

In table (7) we classify the forecasting error by each type of hour (peak, base, and off-peak) for each month of the forecast. Additionally, we graph the MAE per each type of hour per month in figure (10). The graph demonstrates the results provided in table (7) but in a more observable manner, we can see that hour 1 type which represent the peak hours experience higher error forecast with respect to other hour types especially in

the month of December in which we will see later that December has the highest forecast error.

Table 7: Forecast errors by hour type and month

Month	Hour type ⁵	MAE	Min	Max
6	1	2,55	0,04	9,31
	2	2,41	0,04	9,28
	3	1,73	0,01	10,67
		2,23		
7	1	2,63	0,00	12,63
	2	2,25	0,01	10,04
	3	1,97	0,00	6,76
		2,28		
8	1	2,58	0,01	8,95
	2	2,16	0,02	10,78
	3	1,32	0,00	6,56
		2,02		
9	1	2,02	0,04	8,17
	2	1,86	0,01	7,29
	3	1,05	0,01	3,43
		1,64		
10	1	1,52	0,00	7,93
	2	2,2	0,03	9,36
	3	1,15	0,00	4,70
		1,62		
11	1	1,72	0,04	8,10
	2	1,78	0,00	6,58
	3	0,98	0,00	4,69
		1,49		
12	1	4,23	0,09	18,67
	2	1,97	0,01	11,23
	3	1,86	0,01	16,21
		2,69		

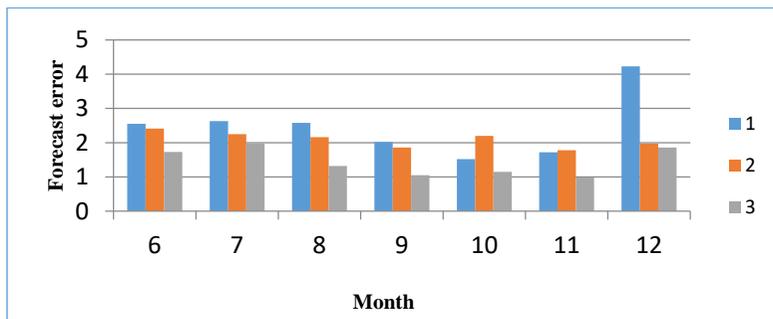


Figure 10: Forecast error by hour type and month

Hour group	Type	Hour
1	Peak	9 to 19
2	Base	8, and 20 to 23
3	Off-Peak	1 to 7, and 24

Regarding the breakdown of forecast error for days of the week; figure (11) shows that Saturdays and Mondays have relatively higher forecast error. It is worth mentioning here that Saturday hours are considered on-peak or normal hours in terms of electricity demand but not the forecasting error, and higher error does not necessarily mean higher electricity demand.

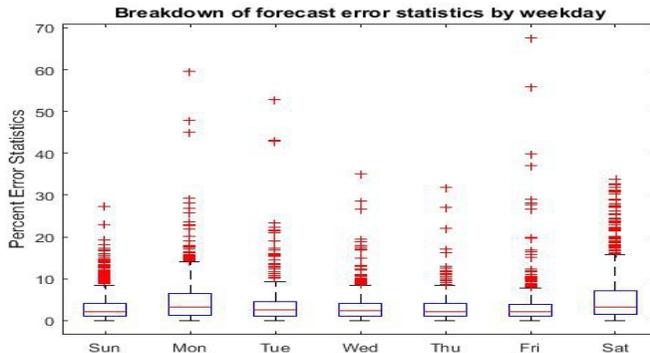


Figure 11: Forecast error by day

Our results here are very much comparable to Weron, 2006, table 1, in which he shows forecast error by days of the week for nine different time series models, we both show the same daily risk profile, but ours slightly outperformed. Notwithstanding Adel Aal, 2012, table 3, demonstrates the same daily risk profile, his results outperformed both of us.

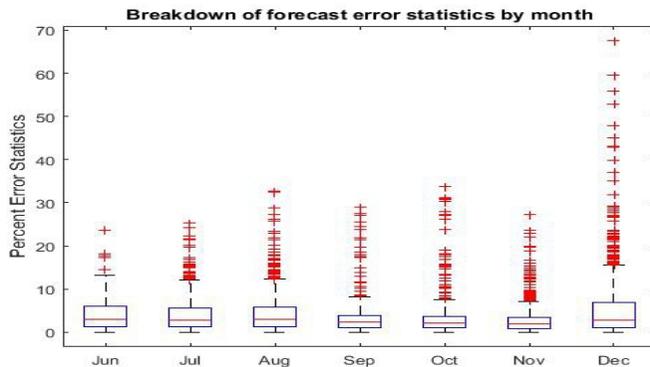


Figure 12: Forecast error by month

In this regard, the last breakdown (figure 12) shows the forecast error by month from June to December, and as indicated previously, the month of December exhibits the highest forecast error especially for the peak hours. Once again, the breakdown by month corresponds to official reports published but is not comparable to other research since it is an additional contribution to our paper.

The classification of the forecasting error by hour, day, and month has an important implication on the physical as well as the commercial relationships in the power market, it helps in setting the power reserves for various demand profiles, and bidding strategies.

Before concluding the analysis, it is worth comparing our results to other related papers. As indicated in (Weron, 2006), the use of errors varies a lot since there is no clear standard for how errors are defined in the literature, and what makes it even more complex is that the same error term is used differently. Accordingly, the comparability among different papers does not give a clear picture even though the same dataset is used. For example, some papers define the “Mean Weekly Error” as the weekly MAPE (literally as the average of the seven daily “average prediction errors”, i.e. daily MAPE values) while others normalize it for the hours of the week. Additional complexity for the comparison is that our study forecasts hourly price while others forecast daily or zonal price. However, we try to do our best to find common forecast measures to construct a comparison table.

Table 8: Comparison of forecast performance

Paper	Energy market	Model	Forecast performance				
			M AE	MA PE	MAPE peak	WM AE	Theil's U
Our paper	(IPEX): Italy	Bagged regression tree	1,83	3,87	2,29		0,360
Cuaresma et al (2004)	(LPX): Germany	ARMA intercepts & jumps	2,568				
Kristiansen (2012)	Nord pool	ARX		5 to 11			
Weron & Misioerek (2008)	California	P-ARX				12	
	Nord pool (98-99)	SNAR				4,04	
	Nord pool (03-04)	SNARX				3,2	
Gianfreda & Grossi (2012)	Italy -Zonal	ARFIMA-GARCH		11,96			0,685
Bowden & Payne (2008)	Five US hubs	Three time series models	11	26			
Ghosh et al (2014)	India	ARIMA-EGARCH		5,86			

Table (8) demonstrates the results of various forecasting measures used in different papers. We can clearly notice that our model “bagged regression tree with exogenous variables” outperformed other models in terms of MAE and MAPE, we additionally provide the peak MAPE which measures the forecast error at spikes. Furthermore, if we can consider

WMAE as a variant MAPE (as in Weron 2006), our model is superior to the models of Weron & Misiorek for both California and Nord pool.

Conclusion & implications

In the present paper, we forecast the Italian hourly electricity price “known as PUN” using a particular forecasting model. Artificial intelligence models are rarely applied on the Italian context. We focus on two main intelligence models, the neural networks, and the bagged regression trees. Furthermore, we calibrate the models using historical data of 2012, 2013, 2014, and the first half of 2015 while our objective is the hourly price forecast for the second half of 2015. During the calibration process we exclude neural network for their underperformance compared to bagged regression trees, then we exclude the bagged regression trees with single price for producing higher forecasting errors. By the end of model calibration, we end up using “bagged regression trees with exogenous variables”. This model is one of the least used forecasting models and additionally has never been applied to forecast the Italian wholesale electricity price. All models’ calibration and forecasting are conducted using MATLAB 2015a. Our final model outperformed neural network model tested in this study and it also showed superiority to other models used in other studies in Italy and in other markets, and above all, it is a flexible model that can be well-customized to fit particular needs (*White box modeling*). We provide a verification for the theoretical model specification in which the optimal leaf size lies between 20 and 50 (*in our case is 25*). Moreover, we also show that larger sample size with longer forecasting horizons may reduce the variance without affecting the bias.

Various policy implications might be drawn from this study. First: exogenous variables such as gas price and load still play a positive role in the accuracy of the models. Second: the average PUN forecast for the second half of 2015 can serve public policy makers such as the energy regulators for price reference for the standard offer market, electricity market operators for their short-term operating needs and hedging purposes, as well as for investment companies for their energy portfolio balancing. Third: for regulatory authorities and generators, higher reserves have to be dedicated for hours, days, and months with higher forecasting errors. And lastly: the model can be adopted by the analysis departments of the interested parties and can be customized, calibrated, and simulated according to their particular needs.

Finally, further research might be needed to validate the accuracy of the model for different time periods or extended sample series; it can also be verified by applying the model to different energy markets. In our study, we assume that load demand accounts for various exogenous variables, but the

research might also be extended to include more of these variables such as weather forecast measures, the impact of renewable on the electricity network, production technologies, the new trend of distributed generation of energy, and market coupling.

Acknowledgment

I would like to thank the participants in the “2016 commodity markets conference” held in Hannover University, Germany in June 2016, for their fruitful comments. I would also like to thank professor Carolina Castagnetti, from the University of Pavia-Italy, for her useful feedback to improve the this paper. I’m also grateful to the Italian Authority for Energy Regulations for the data provision and suggestions.

References:

1. Abdel-Aal, R. E. (2012). Modeling and Forecasting Electric Daily Peak Loads Using Abductive Networks. Unpublished paper, King Fahed University, Saudi Arabia.
2. Aggarwal, S. K., Saini, L. M., & Kumar, A. (2009). Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power and Energy Systems*, 31, 13–22.
3. Beccali, M., Cellura, M., Lo Brano, V., & Marvuglia, A. (2007). Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area. *Renewable & sustainable energy reviews*, 1-26.
4. Bowden, N., & Payne, J. E. (2008). Short term forecasting of electricity prices for MISO hubs: Evidence from ARIMA-EGARCH models. *Energy Economics*, 30(6), 3186-3197.
5. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth.
6. Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140.
7. Breiman, L. (1998). Arcing classifiers (with discussion). *Annals of Statistics*, 26, 801–849.
8. Breiman, L. (2001). Random Forests. *Machine Learning*, 45 (1), 5–32
9. Bunn, D. W. (Ed.) (2004). *Modelling prices in competitive electricity markets*. Chichester: Wiley.
10. Bunn, D. W., & Martoccia, M. (2005). Unilateral and collusive market power in the electricity pool of England and Wales. *Energy Economics*, 7, 305-315.

11. Cuaresma, J. C., Hlouskova, J., Kossmeier, S., & Obersteiner, M. (2004), Forecasting electricity spotprices using linear univariate time-series models. *Applied Energy*, 77(1), 87-106.
12. Duch, W. (2007). What is computational intelligence and where is it going? In W. Duch, & J. Mandziuk (Eds.), *Springer studies in computational intelligence: Vol. 63. Challenges for computational intelligence* (pp. 1–13).
13. Lecture notes in statistics, ETH Zurich, (2008). Available at www.stat.ethz.ch.
14. Eydeland, A., & Wolyniec, K. (2003). *Energy and power risk management*. Hoboken, NJ: Wiley.
15. Gianfreda, A., & Grossi, L. (2012). Forecasting Italian electricity zonal prices with exogenous variables. *Energy Economics*, 34(6), 2228–2239.
16. Girish, G. P., Vijayalakshmi, S. (2013). Determinants of Electricity Price in Competitive Power Market. *International Journal of Business and Management*, 8(21), 70-75.
17. Girish, G. P., Vijayalakshmi, S., Ajaya, K., & Badri, R. 2013. Forecasting Electricity Prices in Deregulated Wholesale Spot Electricity Market: A Review. *International Journal Of Energy Economics And Policy* 4(1): 32-42
18. Ghosh, S., & Kanjilal, K. (2014). Modeling and Forecasting day ahead electricity price in Indian Energy Exchange: Evidence from MSARIMA-EGARCH model. *Int. J. Indian Culture and Business Management*, 8 (3), 413-423.
19. Guerci, E., Rastegar, M. A., & Cincotti, S. (2010). Agent-based modeling and simulation of competitive wholesale electricity markets. In S. Rebennack, et al. (Eds.), *Handbook of power systems II—energy systems* (pp. 241–286). Springer.
20. Hastie, T; Tibshirani, R; & Friedman, J.(2008). *The Elements of Statistical Learning* (2nd ed.). Springer.
21. Hong, T., Pinson, P., & Fan, S. (2014). Global Energy Forecasting Competition 2012. *International Journal of Forecasting*, 30(2), 357–363.
22. Joskow, P. L. (2001). California’s electricity crisis. *Oxford Review of Economic Policy*, 17(3), 365–388.
23. Kaminski, V. (2013). *Energy markets*. Risk Books.
24. Karakatsani, N. V., & Bunn, D. W. (2008). Forecasting electricity prices: the impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, 24(4), 764–785.
25. Kristiansen, T. (2012), Forecasting NordPool day ahead prices with an autoregressive model. *Energy Policy*, 49, 328–332.

26. Kuhn, M & Johnson, K. (2016). Applied Predictive Modeling. Springer
27. Matlab user manual (2015).
28. Mielczarski, W., Michalik, G., & Widjaja, M. (1999). Bidding strategies in electricity markets. Proceedings of the 21st IEEE International Conference Power Industry Computer Applications, 71-76.
29. Misiorek, A., Trueck, S., Weron, R. (2006). Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models. Nonlinear Analysis of Electricity Prices, 10, 1-36.
30. Periodic publications and data bank of the AEEG (Italian authority of energy) available at www.autorita.energia.it.
31. Periodic publications and data bank of the GME (gestore dei mercati energetici) available at www.mercatoelettrico.org.
32. Shahidehpour, M., Yamin, H., & Li, Z. (2002). Market operations in electric power systems: forecasting, scheduling, and risk management. Wiley.
33. Stoft, S. (2002). Power System Economics: Designing Markets for Electricity. Wiley IEEE Press.
34. Sutton, C. D. (2005). Classification and Regression Trees, Bagging, and Boosting. Elsevier handbook of statistics, 24, 303-329.
35. Wei-Yen, L. (2011). Classification and regression trees, overview. Data mining knowl. Dicov. John wiley & sons, 1, 14-23.
36. Weron, R. (2006). Modeling and forecasting electricity loads and prices: a statistical approach. Chichester: Wiley.
37. Weron, R., Misiorek, A. (2008). Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. International journal of forecasting, 24, 744-762.
38. Weron, R. (2013). Electricity price forecasting: A review of the state-of-the-art with a look into the future. International journal of forecasting, 30, 1030-1081.
39. Weron, R., & Misiorek, A. (2006). Short-term electricity price forecasting with time series models: A review and evaluation. In W. Mielczarski (Ed.), Complex electricity markets (pp. 231–254). Łódź: IEPE& SEP.