Energy Economics 44 (2014) 135–142

Contents lists available at ScienceDirect

Energy Economics

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

Forecasting energy markets using support vector machines

Theophilos Papadimitriou, Periklis Gogas *, Efthimios Stathakis

Department of Economics, Democritus University of Thrace, Greece

ARTICLE INFO

Article history:
Received 11 January 2014
Received in revised form 13 March 2014
Accepted 14 March 2014
Available online 3 April 2014

JEL classification:
C45
C51
E44
G10
G17
Q43

Keywords:
Support vector machines
Autoregressive model
European Energy Exchange
Day-ahead market

ABSTRACT

In this paper we investigate the efficiency of a support vector machine (SVM)-based forecasting model for the next-day directional change of electricity prices. We first adjust the best autoregressive SVM model and then we enhance it with various related variables. The system is tested on the daily Phelix index of the German and Austrian control area of the European Energy Exchange (EEX) wholesale electricity market. The forecast accuracy we achieved is 76.12% over a 200 day period.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The transition from regulated regional electricity monopolies to deregulated interacting electric energy markets started in the early 90s once electricity was considered a commodity. In Germany, the liberalization and deregulation process of the electricity sector began with the enactment of the Energy Act 1998 which transposed the EU energy directive 96/92/EC into a national law. In 2001, the German wholesale markets LPX in Leipzig and EEX in Frankfurt merged to establish the EEX AG wholesale market. More recently, the control area of the Austrian power grid (APG) was integrated into the market area of EEX.

Competition in deregulated markets leads suppliers to hold electricity prices near marginal cost, while the pressure to keep prices low provides incentives to adopt cost minimization policies more quickly. In this new environment, energy exchange markets play a key role in increasing efficiency and offering a centralized trading system within which buyers and sellers place their bids on specific amounts of electricity. In these systems, an independent operator aggregates all bids in demand and supply functions and then adjusts the price to the implied equilibrium level.

The EEX provides a uniform price auction mechanism through which participants can place their bids for an individual hour or a specific block of day. The EEX operates both a spot and a derivatives market. The former operates intraday and day-ahead auctions for the regions of Germany/Austria, France and Switzerland. In the latter power contracts of various lengths (weeks, months, and years) are traded. In this paper we focus on the power derivatives market for the region of Germany/Austria as they are described by the Phelix day base and Phelix day peak power indices. The base load refers to the overall load on the power grid in a day and the day peak refers just to the peak hours of the power grid, estimated between 8 am and 8 pm.

The behavior of electricity markets is determined by the special features inherent in electricity consumption. Some of these features are: multiple seasonality, the calendar effect (weekends and holidays consumption behavior), high volatility, non-stationarity and non-linearity (Mugele et al., 2005, Nogales et al., 2002). These characteristics create a complex environment where forecasting is quite a demanding task.

The relevant literature provides a wide range of methods used for the empirical forecasting of electricity prices. Statistical models and artificial neural networks (ANNs) are the most common of them. Contreras et al. (2003) developed an ARIMA model to forecast the hourly electricity prices of day-ahead markets of Spain and California. The average error was 10% for the Spanish wholesale electricity market and 5% for the wholesale electricity market of California. Cuaresma et al. (2004) investigated the efficiency of univariate models in forecasting electricity prices for the German electricity market. In order to evaluate the forecast ability of univariate models they used two sets of autoregressive (AR) and autoregressive moving average (ARMA) models with different setups,

* Corresponding author.
E-mail address: pgkogkas@ierd.duth.gr (P. Gogas).

http://dx.doi.org/10.1016/j.eneco.2014.03.017
0140-9883/© 2014 Elsevier B.V. All rights reserved.

single timeseries models and separate hour-by-hour timeseries models. Using out-of-sample to compare the models' accuracy they concluded that the separate timeseries ARMA models which take into account the price spikes outperform the single timeseries models and AR models. Conejo et al. (2005) employed a wavelet filter to decrease the fluctuation of the prices and eliminate the outliers before using the data into an ARIMA model. Placing emphasis on the non-stationary characteristics of electric energy prices, Garcia et al. (2005) proposed a GARCH model for forecasting the prices of electricity in the markets of Spain and California. Erlwein et al. (2010) proposed a hidden Markov model to forecast the prices of the following day taking into account the bidding strategy adopted by buyers and sellers, unpredictable weather changes and issues related to the productive process.

ANNs are widely used for forecasting the prices of electric energy. The advantage of ANN models is that they can conform to the non-linear behavior of electricity price series. Pao (2006) proposed an ANN for the long-term forecasting of prices in the EEX market. He compared the out-of-sample forecasting performance of an ANN and an auto-regressive model for three different periods of time, and he determined that the ANN model outperformed the autoregressive one. A fuzzy method which enhances the learning capability of a feed forward ANN was proposed by Amjady (2006) for the forecasting of the hourly prices in the Spanish day-ahead electricity market. He compared the forecast accuracy using daily and weekly mean error for various months of year with other forecasting models such as transfer function, dynamic regression, maximum likelihood programming, ARIMA and wavelet-ARIMA. The evaluation revealed that the fuzzy ANN model can provide more accurate forecasts in some cases. Coelho and Santos (2011) used an RBF-ANN in conjunction with a GARCH model for multi-step-ahead point forecast and directional changes forecast for the Spanish electricity market. The accuracy for point forecast models ranged from 6.3 to 7.4 in terms of RMSE and for directional change models the accuracy ranged from 83% to 86%. Zareipour et al. (2011) proposed an SVM model to forecast the directional changes of electricity prices. The binary categorization of the price time series is made according to whether they are above or below a specific threshold. Zhang and Tan (2013) presented a hybrid model combining a chaotic least squares SVM model and an exponential GARCH model to forecast the prices in the market of Spain.

The high volatility of electricity prices increases risk for both producers and consumers. A hedging strategy to reduce this risk against price volatility can be established through a spot and future contract mix (Redl et al., 2009; Wolak, 2007). In this case, the forecasting of the exact value of electricity prices may not be as useful as the directional forecasting. For example, in order to decide whether to buy or sell a derivative with a daily delivery period, the information for the directional change of the daily average electricity prices is often more useful than an exact numerical forecast, especially considering that the directional forecasting is usually more accurate than the exact value forecasting. A classic way to consider the directional forecasting is as a binary classification problem. In this case, the model attempts to correctly classify the future direction of a time series, using an input set with past and present values.

The SVM technique is considered as the state-of-the-art machine learning technique for binary classification. The advantages of SVM in comparison to other machine learning techniques, such ANNs, are the generalization performance and the fewer free parameters. SVM uses convex quadratic programming which always converges to a global solution to determine the classification boundaries.

The majority of machine learning techniques and methods need large data sets for model training. This is essential in schemes involving ANNs or deep learning (DL) architectures and this is the main reason we can find several attempts to use ANN and DL methods within a financial context, though the same is not true for applications where the sampling frequency prohibits the creation of long time series. However, this is not a prerequisite for SVM based systems.

Electricity price time series are characterized by non-linear and non-stationary processes; SVM is a technique that can deal with such characteristics. The basic idea behind SVMs is to project the datasets into a higher dimensional space where the transformed phenomenon can be described by a linear function. The "kernel trick" caries out the projection concept idea while keeping the computational cost low by using just dot products within the original space through the "kernel functions", instead of explicitly computing the mapping of each data point.

In this paper an SVM model is employed for the short term forecasting of the electricity price directional change for the area of Germany and Austria in EEX wholesale market. The lags of daily-average spot price (the dependent variable) and a large set of other relevant explanatory variables, such as the prices of liquid and solid fuel and the total volume-of-trade of electric energy are used as input variables. The best forecasting model was selected through extensive search for the best set of input variables and parameter values.

The paper is organized as follows: In Section 2 there is a brief discussion of the SVM classifiers. In Section 3 the proposed forecasting model and empirical results are presented. Finally, in Section 4 there is an evaluation of the findings.

2. Support vector machines

The support vector machine technique is a supervised machine learning method used for binary data classification. Roughly speaking, the basic idea of an SVM is to select a small number of data points from the dataset, called support vectors (SVs) that can define a hyperplane separating the two classes of observations. When the problem is not linearly-separable, then the SVM is coupled with a non-linear Kernel mapping procedure, projecting the data points to a higher dimensional space, called feature space, where the classes are linearly separable.

The learning procedure has two steps: the training step and the testing step. In the training step, the largest part of the dataset is used for the estimation of the separating hyperplane; in the testing step, the generalization ability of the model is evaluated by investigating the model's performance in the small subset that was left aside in the first step. Typically, 80%–95% of the dataset is used for the training step and the rest is reserved for testing.

In the following Section we briefly describe the mathematical derivations of the SVM technique.

2.1. Linearly separable case

We consider a dataset (vectors) $\mathbf{x}_i \in \mathbb{R}^n$ (i = 1,2,…,N) belonging to two classes (output targets) $y_i \in \{-1, +1\}$. If the two classes are linearly separable, then we define a separator

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} - b = 0. \tag{1}$$

In the linear separable case all data satisfy the above constrains

$$\begin{aligned} &\mathbf{x}_i\mathbf{w} - b > 0 \text{ for } y_i = +1 \\ &\mathbf{x}_i\mathbf{w} - b < 0 \text{ for } y_i = -1 \end{aligned} \tag{2}$$

in such that $y_i f(\mathbf{x}_i) > 0\ \forall i$. Where $\mathbf{w}$ is the weight vector and b is the bias.

The optimal hyper plane is selected as the decision boundary that classifies each data vector to the correct class and has the maximum distance from both classes. This distance is often called "main". In Fig. 1, the SVs are represented with the pronounced contour, the margin lines (defining the distance of the hyperplane with each class) are represented with the continuous lines and the hyper plane is represented with the dashed line.

The solution to the problem of finding the hyperplane can be dealt with through the Lagrange relaxation procedure in the following equation:

$$\min_{\boldsymbol{w},b} \max_{\boldsymbol{a}} \left(\frac{1}{2} \|\boldsymbol{w}\|^2 - \sum_{i=1}^{N} a_i \left[y_i \left(\boldsymbol{w}^T \boldsymbol{x}_i - b \right) - 1 \right] \right) \tag{3}$$

Download English Version:

https://daneshyari.com/en/article/5064629

Download Persian Version:

https://daneshyari.com/article/5064629

Daneshyari.com