EXAMINING OPPORTUNITIES OF THE PARTICIPATION IN SEQUENTIAL MARKETS AND FINDING THE OPTIMAL BALANCE FOR IT

Written by: Johannes Weyers, Julian Cornea, Kim Ferres, Mert Erdogan and Rufin Korbmacher

University of Cologne
Advanced Analytics and Algorithms

1. Business Goal:

1.1 Who is the stakeholder or client?

An energy company

1.2 A description of the business goal. What are the business benefits of implementing this idea?

Maximizing profits from energy sales

1.3 What opportunity is it creating? What shortcomings does it address?

Splitting available energy between the day-ahead and intra-day auctions. It would increase effectiveness of dealing with the energy market's price volatility and decrease current opportunity costs

1.4 What would be considered a success?

Identifying whether participation in the intra-day auction is viable and, if it is, deriving a recommendation for pricing auction offers

2. Analytics and Data Mining Goal:

2.1 A description of the analytics objective:

Accordingly, to the issue description of the analytics objective it is necessary to identify parameters which are responsible for the existing price difference between the day-ahead and intraday auctions. Based on these parameters it would be needful to develop a regression model to emphasize this issue in a graphical way.

2.2 Is this a supervised or unsupervised task? Is it predictive or descriptive? Is it retrospective or forward-looking?

Furthermore, this task is supervised. The reason in order to that is the available price data. The first part of the regression is obviously descriptive, but the second part cannot exactly be estimated. It is predictive and things with a predictive character are never exact. Both parts are also retrospective. But the regression model can additionally be used for prospective predictions, required the utilized parameters data is available at the point of prediction.

2.3 What is the main outcome variable of interest?

In term of the main outcome variables it is important to consider several parameters which have a causal relationship with the prices at both auctions and the kind of their relationship.

3. Data:

3.1 Brief description of available data:

The dataset to be examined shall include volume (in MWh) and price (in €/MWh) of the energy sold at each auction sale. For the day-ahead auction the data is provided in hourly-, for the intraday auction in 15-minute resolution. As a whole the dataset encompasses 731 days, so two years, which results in 17544 data points for the day-ahead data and 70176 data points for the intraday auction data of the year 2016 and 2017.

3.2 Some guidance on the data subset that will be used and the re-processing or preparation that might be needed based on your past experience:

The first step in order to create the dataset to be used for advanced analysis will be the scraping of relevant information from the respective websites. After the successful collection and persistence of day-ahead and intra-day auction price and volume information the hourly data needs to be converted into 15-minute increments and attached to the data file of the intra-day auction data. Since there is no missing data, no respective preparations are required. As a unique identifier of each row in the merged data file, a timestamp will be used. Additionally, the difference between the prices of the two power markets are mapped to a new column, which is called price premium throughout our analysis.

3.3 Sample that will be used:

In order to get an overview of the features and the nature of the data, a copy of ten rows of the dataset is provided.

Timestamp	15-minute increment	MWh (day-ahead)	€/MWh (day-ahead)	MWh (intraday)	€/MWh (intraday)
2017-01-01 00:00:00	00:00	69,20	2736,55	46,56	616,5
2017-01-01 00:15:00	00:15	69,20	2736,55	39,96	282,6
2017-01-01 00:30:00	00:30	69,20	2736,55	34,44	257,1
2017-01-01 00:45:00	00:45	69,20	2736,55	27,18	426,2
2017-01-01 01:00:00	01:00	61,07	2642,975	38,68	456,9
2017-01-01 01:15:00	01:15	61,07	2642,975	34,1	326,7
2017-01-01 01:30:00	01:30	61,07	2642,975	40,52	396,2
2017-01-01 01:45:00	01:45	61,07	2642,975	36,41	404,5
2017-01-01 02:00:00	02:00	55,47	2651,8	39,83	348,2
2017-01-01 02:15:00	02:15	55,47	2651,8	37,76	231,4
2017-01-01 02:30:00	02:30	55,47	2651,8	35,24	277,4
2017-01-01 02:45:00	02:45	55,47	2651,8	32,79	288

Figure 1: A sample overview over the first few lines of the data

4. Methods

4.1 What are some data mining methods to consider?

To find underlying patterns in the given dataset based on the price premium, a supervised data mining task needs to be solved. This results out of the fact that price premium is our target variable which should be explained through a good fitting descriptive model. The values of price premiums are already known prior to the model step in our analysis. Due to the fact that the target variable is continuous regression models will be considered. For the following step, a forecast on whether the price premiums will rise or fall in the near future, a predictive model is necessary which can be the previous build regression model, but not necessarily has to. In order to get discrete values from the pre-trained regression model the output needs to be categorized in rise or fall through a threshold. Apart from that model any classification model can be used for this predictive task as well. For

example, Neural Networks, Decision Trees or Naïve Bayes. The final model selection process will be determined through performance measurements. Hence the choice is objectified.

4.2 Which performance measures are appropriate?

The crucial point in any data mining task is evaluation of models and therefore selecting the right performance measures, not only to select the best model, but also to infer the real performance without the interference of random data constellations. Each task implies specific Evaluation approaches and therefore specific measures. The descriptive task needs a measure to evaluate the goodness of fit and the fitting of the regression. The predictive task requires a good reflection of the performance on new data. A common method to simulate the performance on new data is to split the original dataset into training and holdout data. From that, advanced measures are derived from the comparison of actual prices of a holdout dataset and the predictions of an applied predictive model. Furthermore, cross-validation as an industry standard is used to validate the model and test the generalization performance.

4.3 How do they map to the business goal?

The goodness of fit shows if the derived parameters are responsible for the price difference between the auctions. The prediction accuracy shows how well the causal relationship between the parameters and the price has been estimated.

5. Implementation and Production:

5.1 Operational requirements and constraints:

The predictive model will not be continuously run in real-time, however it certainly can be used for that. This would however require automated data collection, an automated data preparation, a thorough monitoring system, and a backup system in case the prediction performance deteriorates too much.