PREDICTIVE PRICING OF GERMAN POWER FUTURES

Capstone Project Summary

BY
TAMAS OLAH

Budapest
CENTRAL EUROPEAN UNIVERSITY
June 2020

Introduction

Wholesale electricity prices can be predicted based on weather forecasts. This is the working assumption of this capstone project, which aims to develop a price prediction framework for electricity trading on a leading European power exchange. The fundamental idea behind the project is that by training machine learning algorithms on historical market data and weather forecast data, one can predict electricity forward prices, if renewables represent a large share in the energy mix. This document will summarize the background and motivation of the project, chronologically explain the work done and its outcomes, and describe the main analytical steps yet to be taken.

The project is live and ongoing as of the submission of this document. There are two main reasons for the slower than expected project execution. One is inconsistent data availability. Second is a change in the target of analysis. The analyst shifted his focus on a new marketplace abandoning previous work due to the client's request. This change will be explained in detail below. The analyst remains in employment by the client and has the responsibility to oversee the project until it is deployed.

Background and motivation

Electricity is a unique commodity. It cannot be stored in large quantities. Moreover, frequency on the electricity grid must be stable. Therefore, electricity production and consumption must equal at all times. This means that the grid must always be balanced between producers and consumers. On coupled markets, meaning on markets where electricity grids are interconnected across borders, this vast system requires careful planning and balancing. This is why power exchanges exist.

Power exchanges are meant to operate markets and facilitate balancing by organizing trade. They offer a platform for players on the power market – producers, consumers, and traders – to connect and put in orders for buying or selling power. These orders represent supply and demand for a specific market area, in a specific point in time. The power exchange calculates a market clearing price based on the supply and demand curves. This way, wholesale electricity prices are a result of a transparent, open, and competitive process.

The proliferation of renewable technologies, e.g. wind and solar power generation, have naturally increased the variability in the system. Renewable generation is erratic at best. While non-renewable power plants can be turned on and off as needed, production from renewables is subject to weather. Therefore, the chance to forecast electricity prices based on weather forecasts have dramatically increased in recent years.

Market description

The **spot** electricity market is operated through the day-ahead and intraday markets. On the day-ahead market participants put in buy or sell orders on an auction. On the intraday market, trading runs continuously. This provides margins to profit from. While on the futures market one trades with financial derivatives of electricity, on the spot market one trades with actual, physical electricity.

Data selection

Three main data sources are used in this project. These are capacity data, weather data, and market data. This section will describe each source in detail.

Capacity data are obtained from the *Marktstammdatenregister* (MaStR), or the Core Energy Market Data Register, an open platform maintained by the German Federal Network Agency. The MaStR registers and lists all operating electricity generating units in the country. Knowing the location and capacity of e.g. each wind turbine in Germany will prove useful later in the analysis, serving as weight for weather forecasts.

Weather data is obtained from a private weather data vendor. It is important to note that this project uses archived weather forecast data instead of actual weather data. That is, we are not interested in knowing the actual wind speed on e.g. January 1, 2018, instead we are interested in the forecasted wind speed from the day before. This is important, because once the model is deployed, trivially, only weather forecast data will be available.

To better understand weather forecasts, the analyst experimented with the model output of the Global Forecast System (GFS) by the National Centers for Environmental Prediction, an agency of the U.S. government. The GFS is one for the main global Numerical Weather Prediction forecast systems. Its raw output comes in binary format and is unreadable to the untrained eye. Therefore, it required significant time to extract e.g. wind speed predictions from the dataset. Fortunately, the analyst could enlist the help and conduct interviews with meteorologists.

Market data was initially obtained from the **futures** marketplace following a lengthy process involving the marketplace and third-party data vendors. The data was not available at the outset of the project—it required the analyst to negotiate with both the marketplace, and third-party companies to make the data available. The global pandemic did not help these efforts but eventually, the marketplace published the data for purchase. After solving significant data quality issues, and performing cleaning and exploratory

analysis, the analyst and the client concluded that the data showed that the marketplace did not contain enough liquidity to commence trading. Therefore, the initial target of the analysis, the futures marketplace, was abandoned.

Change in focus

Therefore, the project shifted its focus to the **spot** marketplace. This is a significant change; the spot marketplace is structurally different from the futures marketplace, albeit both are operated by the same parent company. The working of the sport market is explained above. As of the submission of this document, spot market data has been purchased and is ongoing initial data cleaning and exploratory steps.

Steps to be taken

For reasons established above, the modeling and validation phase of this project is still ahead. This section will briefly describe the main analytical steps yet to be taken.

The spot market data will be partitioned into 60/20/20 training/validation/test splits. As the data is a time series, the partitioning will not be random; rather the ordered list of observations will be split into arbitrary chunks, such as 2019 as test, 2018 as validation, and 2017–2015 as training sets. Seasonality will be accounted for. The training set will be used to fit a Logistic Regression, a Random Forest, an XGBoost and a GBM model. The validation set will be used to evaluate the performance of the algorithms trained on the training set, and the best performing algorithm will be selected. Finally, the chosen algorithm will be applied to the test data.

For cross validation the walk forward optimization method will be used. This means that a small portion (one month) of the validation set following the training set is used for validation and the results are recorded. The in-sample time window is shifted forward by the period covered by the out of sample test, and the process repeated. This means that the first validation step (January 2018) will validate the January 2015–December 2017 period. The second validation step (February 2018) will validate the January 2015–January 2018 training period. The third validation step (March 2018) will validate the January 2015–February 2018 training period, and so forth. Finally, all of the recorded results will be aggregated to assess model performances.

The Area Under the ROC Curve (AUC) will be used to measure the performance of the models across all classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example.

Lessons learned (so far)

Without reciting the obvious, the most important lessons so far are that data vendors may be unreliable, the burden of cleaning and structuring supposedly clean data always falls on the analyst, and that even in 2020, some participants in the industry use outdated technologies.

This page is intentionally left blank.

This page is intentionally left blank.

This page is intentionally left blank.

This page is intentionally left blank.

This page is intentionally left blank.