



Epex forecasting with XGBoost

AIHACK 2022

Kee Kwan Liu and Jacky Zhang | 13-3-2022

1: Data exploration

From Epex, spot price, and system price database, 47858, 51840, and 47186 entries and a total of 7 features are found. All data frames are converted to time series. In the systems data frame, unlike the other two databases, "Settlement Date" along with "Settlement period" is used instead of timestamp (This will be converted later on). The mean of "apx_da_hourly" is found to be 65.31, with values ranging from -38.8 to 1860 and a standard deviation of 71.6. In the spots data frame, 516 invalid entries are found.

2: Data visualization

The Epex price soared when approaching the year 2021 relative to that in 2019 and 2020. The hourly change of Epex price for most of the time ranges from 60% to 160%. In general, the Epex curve has an upward trend and seasonality pattern occurring every week, with prices relatively high at the start of the week and descending at the end of the week. There are in total of 72 extreme days defined by a residual value greater than 500. They occur after 2021 in January and between September and December.

In the Augmented Dickey-Fuller test, the ADF statistic (-4.760525) is less than the critical value at 1% (-3.431). Thus, we can confidently reject the null hypothesis at a significance level of 1%. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, The KPSS statistic (7.957) is higher than the critical value at 1% (0.739), thus, the null hypothesis is not rejected. In both tests, it is evident that the time series is stationary, which is important in the model we use later on.

In the partial autocorrelation plot, the correlation is only strong within a time lag of 4 indices. It is beneficial to know in order to appropriately extract features for prediction.

3: Feature engineering

All the features in epex, spot price, and systems price data frame are joined. In addition, hour, month, and weekday detail are generated. It is found that the spot price is highly correlated with the epex price and System sell price and buying price having a moderately low correlation.

All the data are fitted with MinMaxScaler and PCA. The training set is the first 95% of the data with testing set the remaining 5%

4: Forecasting

An XGBoost model is used to forecast the Epex graph. It is proven to be rather accurate and is capable of capturing the trends of the curve with RMSE of hour-ahead electricity price XGBoost forecast equals 47.762