Prediction of Electricity Prices

Sedlacek P.[1,*], Lancosova E.[1]

[1,2] Masaryk university, Brno

*corresponding author:e-mail: 369947@mail.muni.cz

Abstract

The aim of this paper is analyzing time series that might have an impact on prices of energy commodities. Analysis and econometric modelling reveal which time series are corelated and how. This is input for further modelling of forecasting tool for electricity prices on Czech energy market trading commodity futures.

Keywords: electricity, forecasting, prices

1. Introduction

For companies trading electricity as well as for electricity producers, the most convenient product is future year baseload (BL cal+1 – next year future). There are professionals deciding purchase strategy on their opinion (based on fundamental and technical analysis), but there is no exact model. Even though there is a tool for predicting future prices of spot electricity prices (predictions for Day ahead market) on Reuters platform Eikon, there is no prediction tool for the market prices of futures in near future. This model was created for Czech commodity market in this paper based on time series from from January 2015 to December 2018. Selected time series are electricity prices of BL cal+1 on electricity commodity markets in Germany, Poland, France, Slovakia, Czech Republic, Italy, Hungary and contracted quantities per day in Czech Republic. Those were chosen to figure out the level of correlation of the markets and eventual delay of price trends of the commodity markets. Then we have chosen time series of prices of gas (NCG cal+1 and cal+2), LGO (light gas oil), oil, coal and uranium, as energy "substitutes". We have also selected prices of emission allowances and information about daily electricity production by source and prices on spot electricity market of that day (representing the point of view of electricity producers). We have also selected exchange rates of CZK/EUR and EUR/USD, weather data (temperature, sunshine and wind), day of week, stock exchange indexes (PX and DAX), representing the trends in economies, as well as stocks of ČEZ and EON (Czech and German electricity trading and distributing companies).

2. Methodology

As stated in chapter 1, our dataset consists of high number of time series predictors. In such settings as is often the case in financial markets, several drawbacks like multicollinearity, autocorrelation, missing values, necessity to detect high number of irrelevant variables and the most problematic of all, debatable stationarity, are to be expected. Therefore, methods for feature pre-selection were chosen in such a way that each method would leverage different characteristic of our data. In order to satisfy stationarity assumption, one day differences of all variables were used thorough whole analysis.

Lasso and ridge were both chosen for maintaining linear structure which they share with ARIMAX models chosen for prediction model specification. Both methods fare better when applied to standardized input which is problematic in many real-world time series settings where stationarity is difficult to reach. Therefore, last method considered for feature preselection are Random Forests (Pedregosa, 2011) because of for their endurance against different scales, multicollinearity and autocorrelation thanks to random sampling from data common to all bagging algorithms. Also, as a CART based method, RF are able to deal with missing observations by surrogate splits.

Since multicollinearity is to be expected in financial markets setting, Principal Components Analysis (PCA) was employed to orthogonalize some of the predictors exhibiting high correlation and to engineer new predictors with potentially higher prediction power (Hastie, 2003). Having significantly reduced our feature space through use of before mentioned methods, ARIMAX was employed to assess variable relevance more finely (Hyndman, 2019). It was chosen for its ability to take full advantage of non-trivial link between past and present values of observed variables as well as for its interpretability and transparency common to all linear models. Finally, prediction on strictly independent test sample was created, accurately assessing model's prediction abilities to be expected in future years.

3. Results And Discussion

Movements in prices of electricity in European commodity energy markets are strongly positively correlated. Also, movement in production of different kinds of power plants is correlated but not as strongly as prices. Both these relationships are later exploited by PCA.

A strong positive correlation between sunshine and photovoltaic plant production, and between wind and wind power plant production is to be expected as well as well-known correlation between prices of electricity, gas, coal and emission allowance.

On the other hand, one could be surprised by lack of correlation between electricity volume and price and also minor correlation of national stock market indexes and energy prices (can be explained as economy is flourishing, expected consumption of energies is increasing causing higher prices).

Table 1 contains specification of model used for prediction as a result of feature preselection, PCA, manual feature selection and optimal lag determination.
We can see that electricity prices are strongly determined by past prices up to 1 day, the latency is quite short which is probably due to high volatility of the market.

In our case Gas NCG cal2 was statistically significant but this was not in accordance with financial markets behaviour, therefore it was replaced by GAS NCG cal1 leading to more stable and powerful model.

The interpretation is intuitive, if GAS NCG cal1 increases by 1 (EUR/MWh) unit today, we can expect increase in electricity price of 0.015 units (EUR/MWh) tomorrow. Interpretation of the rest of variables is similar, increase in EU emission allowance, coal price index and alternative power plant production all lead to increase in price tomorrow while increase in LGO would lead to drop in price following day.

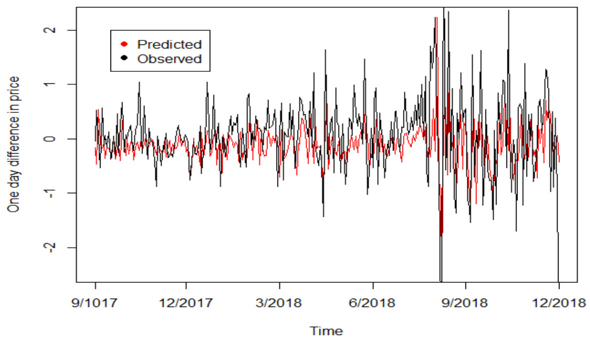
The high impact of alternative power plant production on tomorrow electricity prices can be interpreted as lack of energy. Those power plants are being ready to produce energy in case of lack in the grid, there is regular subsidy for them for not producing and only staing in standby mode (usualy from gas or other source with immidiate production and low costs for activation and deactivation).

Table 1. Prediction model specification

Predictor	Coefficient	P-value
AR 1	-0.471	0.003
MA 1	0.266	0.125
Gas NCG cal1	0.015	0.864
LGO	-0.006	0.017
EU emission allowance	0.570	0
Coal price index	0.110	0
Alternative power plant	0.007	0.009

Figure 2 shows development of electricity prices from sepentember 2017 untill December 2018 and how our model predicted these values

Figure 1. Comparison of predicted values and reality



We can see that while in many cases we are able to assess general trend, we often underestimate its size. This is caused by change in time series behaviour around August 2018 where we have witnessed strong increase in volatility. One of the main reason of volatility increase is uncertainty about Brexit, where enormous amounts of speculative money were invested in EU emmision allowance market. This change forced our model to extrapolate data causing its deteriorated but still relevant prediction power. If we reduce our model to binary output of decrease or increase in price and compare its prediction to reality, we are able to correctly predict 59.2% of data.

Table 2. Confusion matrix

		Observed	
Predicted		Decrease	Increase
	Decrease	95	87
	Increase	25	68

It is important to note that figures in table 2 were obtained on out of sample test data that were not in any way part of training process. This ensures applicability of our model on future data, that it has never seen before, and makes it likely that similar results would be reached by any investor using our model.It is easy to see in table 2 that our model predicts poorly decrease in price but is quite confident in price increase with correct prediction rate of 73.1 %. Therefore the recommendation would be to proceed as usual when model predicts decrease and to buy when it predict increase of electrity price.

4. Conclusion

With rising market volatility is this model with 73 % certainty of prediction of following day price increase very useful and implementing this knowledge into purchase strategy can cause extra profits that might be invested for example to new “green” powerplants.

The support of the Masaryk University internal grant No. 2156 is gratefully acknowledged.

References

Greene W. H. (2000), Econometric Analysis, Prentice Hall Inc, 1026, 123-131

Hastie, T.R., Tibshirani R. and Friedman J.H. (2003), The elements of statistical learning: data mining, inference, and prediction, Corrected ed. New York: Springer, 699, 43-132 and 219-379

Hyndman R., Athanasopoulos G., Bergmeir C., Caceres G., Chhay L., O'Hara-Wild M., Petropoulos F., Razbash S., Wang E., Yasmeen F. (2019), forecast: Forecasting functions for time series and linear models

Pedregosa et al (2011), Scikit-learn: Machine Learning in Python, JMLR 12, 2825-2