News effect in the European Emission Trading System

Áron Dénes Hartvig, Péter Pálos, and Áron Pap

June 2022

- We calculate EU ETS news index based on GDELT data with bag-of-words/TF-IDF method
- There are certain periods of 'hype' around climate change
- We identify the keywords with highest impact
- News indices have significant effect on ETS prices

1 Introduction

Carbon pricing is a effective method to incentivize emissions reduction. It captures the external costs of greenhouse gas (GHG) emissions. The first international carbon market, the European Trading System (ETS), was implemented in 2005 covering power generators and energy-intensive industries. The ETS sets a cap on GHGs emitted by these sectors by limiting the number of emission allowances that the companies receive or buy. In Phase 1 of ETS almost all allowances were given for free, but as the trading system evolved the number of free allowances decreased pushing its price higher. The ETS allowance prices exploded in 2021 reaching €88.9 and continues to increase.

However, volatile price changes makes investment decisions complex. Most companies covered by the ETS are risk averse in terms of trading with allowances. They are purchasing allowances as they are emitting GHGs and only speculate with a small fraction of the expected number of required allowances. Thus, a substantial increase in the ETS prices leads to higher production costs. Forecasting carbon prices can enhance the purchasing strategy and reduce allowance costs.

In this paper we propose alternative explanatory variables to forecast ETS allowance prices more accurately. We apply bag-of-words and TF-IDF feature extraction to GDELT news database to predict next day's ETS allowance price. Our results suggest that the occurrance of climate change related keywords in the most reliable news sites has significant impact on carbon prices.

We conclude that news improves forecast accuracy of carbon prices as ETS is a policy-driven system. EU climate goals and policies are constantly becoming more ambitious impacting the ETS sectors as well. New policy package proposals are widely discussed in the news especially around the vote. Furthermore, the empirical models confirm that the

European carbon market is an effective market integrating publicly available information. Including news data in our forecasting model improved forecast accuracy. Consequently, ETS allowance traders take news into consideration in their trade-decisions.

Several quantitative methods have already been developed in the literature to forecast carbon prices. Zhao et al. (2018) categorizes these papers into two groups: forecasting based on time-series data using carbon price-only, and the ones involving economic and energy data for forecasting and monitoring carbon prices.

The carbon price-only methods mostly include ARIMA models; however, they can only capture linear relationships (Zhu and Chevallier, 2017). Therefore, more advanced frameworks, like different varieties of generalized autoregressive conditional heteroscedasticity (GARCH) model (Arouri et al., 2012; Benschopa and López Cabreraa, 2014; Byun and Cho, 2013), and vector autoregressive (VAR) model (Arouri et al., 2012) has been applied to carbon prices.

Nevertheless, carbon price-only methods does not incorporate all available information in the market. Various articles that aim to forecast carbon prices use economic and energy related variables proxying the demand for the CO2 allowances (Gubrandsdóttir and Haraldsson, 2011; Zhao et al., 2018).

Recently, alternative predictors, e.g., news data through natural language processing (NLP), have been widely used to forecast market data. Furthermore, NLP has already been applied to ETS price prediction. Ye and Xue (2021) created a carbon tone index reflecting sentiment in news articles and showed that it has a strong predictive power on carbon prices.

To shed more light on the impact of news on ETS prices we create features by applying TF-IDF method to GDELT news dataset. TF-IDF (TF - term frequency, IDF - inverse document frequency) is a popular method to determine the importance of a term in news. It has been widely used to improve forecast accuracy of stock prices (Coyne et al., 2017; Lubis et al., 2021; Mittermayer, 2004; Nikfarjam et al., 2010).

The remainder of this paper is organized as follows. Section 2 provides a short description of our dataset and Section 3 outlines the methods used in to analyse it. This leads into Section 4 where we discuss forecasting performance of our models. Finally, Section 5 summarises our conclusions and ideas for work moving forward.

2 Data

We use both financial time-series and news data to forecast the ETS prices.

GDELT is a free open platform covering global news from numerous countries in over 100 languages and identifies the people, locations, organizations, themes, sources, emotions, counts, quotes, images and events driving the society (GDELT, 2022).

- What is GDELT
- Financial data sources, why did we choose these (control) variables
- Time span

3 Methods

- GDELT filtering, RAKE keywords, keyword filtering
- Bag-of-words method / TF-IDF method

4 Results

4.1 Indices from bag-of-keywords/TF-IDF

- Visualization of keyword 'indices'
- Highlight most important events
- Qualitative analysis

4.2 Regression

- Present results of stepwise regression
- Present ElasticNet results
- Present PCA results

5 Conclusions and further research opportunities

- EU ETS market is efficient, news effect appear in the prices with a few day lags.
- Keyword indices signal important events
- News indices enhance forecasting

5.1 Further research

- More complicated methods for NLP
- Extend news sources

References

Arouri, M. E. H., Jawadi, F., and Nguyen, D. K. (2012). Nonlinearities in carbon spot-futures price relationships during phase ii of the eu ets. *Economic Modelling*, 29(3):884–892.

Benschopa, T. and López Cabreraa, B. (2014). Volatility modelling of co2 emission allowance spot prices with regime-switching garch models. Technical report, SFB 649 Discussion Paper.

Byun, S. J. and Cho, H. (2013). Forecasting carbon futures volatility using garch models with energy volatilities. *Energy Economics*, 40:207–221.

Coyne, S., Madiraju, P., and Coelho, J. (2017). Forecasting stock prices using social media analysis. In *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pages 1031–1038. IEEE.

GDELT (2022).

Gubrandsdóttir, H. N. and Haraldsson, H. Ó. (2011). Predicting the price of eu ets carbon credits. *Systems Engineering Procedia*, 1:481–489.

Lubis, A. R., Nasution, M. K., Sitompul, O. S., and Zamzami, E. M. (2021). The effect of the tf-idf algorithm in times series in forecasting word on social media. *Indones. J. Electr. Eng. Comput. Sci.*, 22(2):976.

Mittermayer, M.-A. (2004). Forecasting intraday stock price trends with text mining techniques. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, pages 10–pp. IEEE.

Nikfarjam, A., Emadzadeh, E., and Muthaiyah, S. (2010). Text mining approaches for stock market prediction. In *2010 The 2nd international conference on computer and automation engineering (ICCAE)*, volume 4, pages 256–260. IEEE.

Ye, J. and Xue, M. (2021). Influences of sentiment from news articles on eu carbon prices. *Energy Economics*, 101:105393.

Zhao, X., Han, M., Ding, L., and Kang, W. (2018). Usefulness of economic and energy data at different frequencies for carbon price forecasting in the eu ets. *Applied Energy*, 216:132–141.

Zhu, B. and Chevallier, J. (2017). Carbon price forecasting with a hybrid arima and least squares support vector machines methodology. In *Pricing and forecasting carbon markets*, pages 87–107. Springer.