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 an economically efficient instrument in the policy toolkit to signal the social cost of emissions and to stimulate investments in low-carbon technological innovations. Emissions trading systems are market-based mechanisms where a cap is set on the emissions of certain sectors and the entities that are covered are allowed to trade emissions permits. Trading ensures that emissions reduction takes place where abatement costs are the lowest.

The first international carbon market, the European Trading System (ETS), was implemented in 2005 covering power generators and energy-intensive industries. In Phase 1 of ETS almost all emissions permits, or allowances, were given for free, but as the trading system evolved the number of free allowances decreased pushing its price higher. ETS prices stayed low even in the early stage of Phase III (2013-2020) and finally started to rise in 2017. Unexpected, The ETS allowance prices exploded in 2021 reaching almost €100 in February 2022. Nevertheless, prices dropped below €60 in early-March due to the disruptions caused by the Russian invasion in Ukraine.

The aim of the emissions trading system is to incentivise emissions abatement investments. For a long time ETS prices were too low to become decisive in investment decisions, but the current price range and the peak price of €100 have caught the eyes of investors. However, volatile changes in the allowance prices and policy uncertainty around the future of the ETS increase risks associated with low-carbon technology investments. Furthermore, most companies covered by the ETS are risk averse in terms of trading with allowances (Haita-Falah, 2016) and volatile carbon prices increase uncertainty of their production costs since they purchase allowances as they emit.

*XYZ

Our contribution to the literature of carbon pricing is twofold. First, we propose a new climate change news index that tracks the ongoing discussion about climate change in the media. We apply TF-IDF feature extraction to GDELT news database to measure the frequency of climate change related keywords in the news associated with the EU. The TF-IDF features help to incorporate policy context to the analysis of carbon prices. Second, we apply the news index to predict next day's ETS allowance price returns. We compare the forecast accuracy of models incorporating the climate change news index against a set of control variables taken from the literature. Our results suggest that the occurrance of climate change related keywords in the most reliable news sites has significant impact on the ETS prices.

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 future work.

2 Data

We use both economic 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). The process leading to the extraction of the TF-IDF scores from the article URL collection provided by GDELT is detailed in the methodology section.

In order to measure the net impact of the news articles, we tried to filter out the distorting effects of the main economic variables. Based on (Ye and Xue, 2021), we included the following control variables in our analysis:

- Gas: Natural Gas Futures
- Electricity: Electricity Yearly Futures (ELCBASYc1)
- Coal: Rotterdam Coal Futures (ATWMc1)
- Stocks: Europe 600 Index (STOXX)
- Oil: Brent Oil Futures (LCOU2)

These control variables are not only advantageous because of the filtering effect but also suitable for creating a baseline model for later comparison. Using this, we can demonstrate the benefit of extra information from the news articles in our forecasting model.

The data was collected from 2017 until December 2021. The starting date was given by the availability of control variables, and the end was fixed in this way to avoid the possible distorting effect of the Russian-Ukrainian conflict.

2.1 Exploratory Data Analysis

On Figure 1 we show the evolution of the price and the volatility of the EU ETS carbon permits over the sample period. We can observe interesting dynamics, as the time series of the prices shows an upward trajectory, whereas the rolling volatility graph follows a downward trend which might suggest that the carbon market is becoming more efficient. However, we can also observe huge spikes in the chart, which shows that extreme market events (for example the crisis around the COVID-19 pandemic) can still move the carbon price substantially. It is also interesting to note that the price increase has significantly accelerated since the beginning of 2021 and especially after COP26.

On Figure 2 we show a correlation heatmap of the TF-IDF group scores, where we assigned each individual keyword to one particular group and aggregated the scores for each category afterwards. We can observe relatively high pairwise correlations for all groups, which indicates that the different groups are usually appearing together or at least the same time in the news. But there are also specific pairs with an even higher correlation, such as *gas-fossil fuel* and *policy-emissions*, which seem intuitive given that they are good proxies for related aspects of the carbon market.

On Figure 3 we exhibit the dynamics of the aggregated carbon market news index, which is based on the aggregation of the TF-IDF scores for all keywords that we considered. This provides a first assessment of the relevance of climate change in general, and the EU ETS carbon market in particular, over the global media. It is evident from this chart as well, that significant events like the COVID-19 pandemic and the COP26 manifest through the news data as well and our index is able to capture this *signal*.

Figure 1: EU ETS carbon price dynamics in the sample period

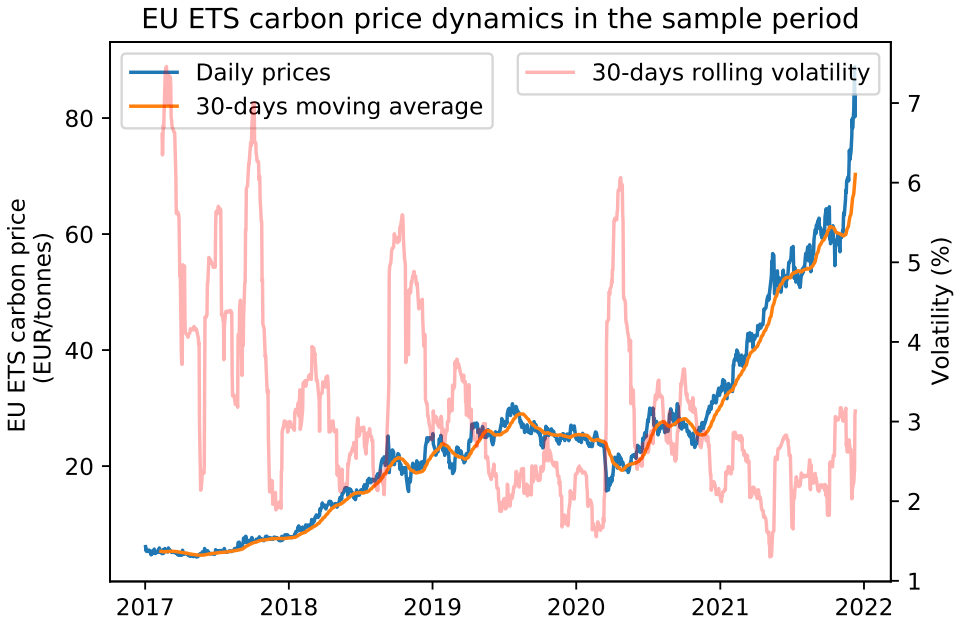


Figure 2: Correlation of TF-IDF group scores

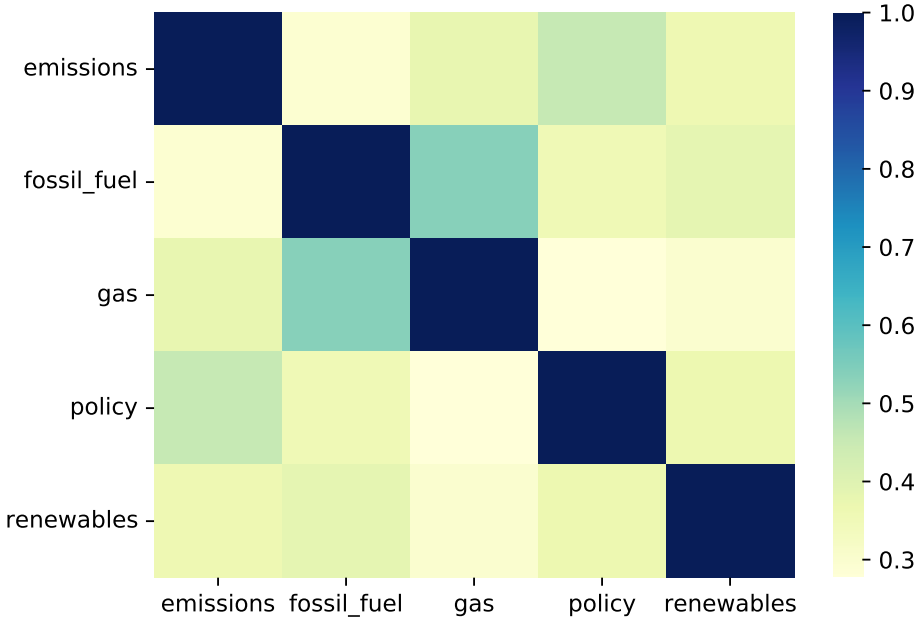
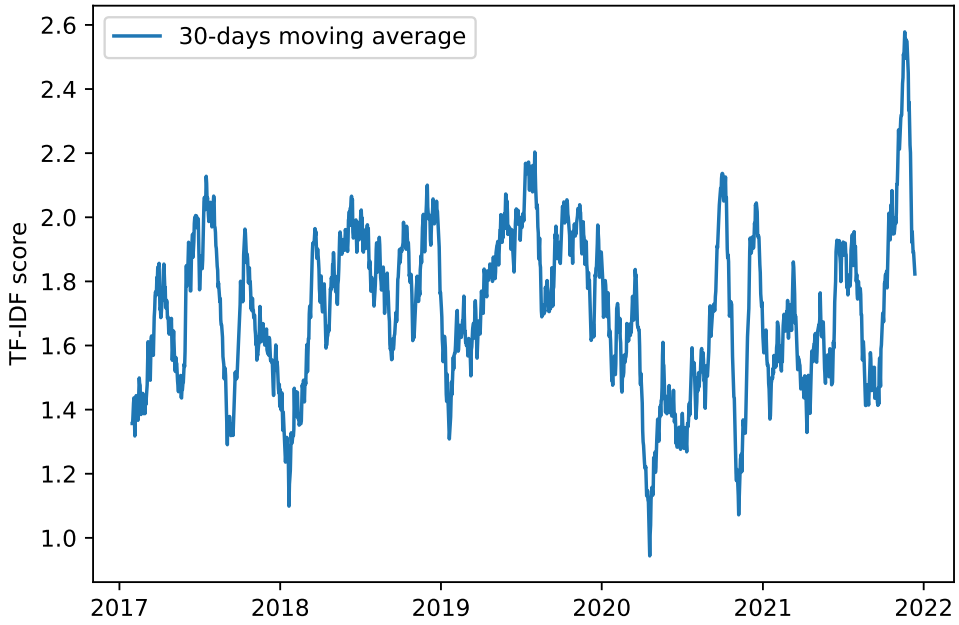


Figure 3: Time series of the aggregated TF-IDF scores



3 Methodology

One of the main added values of our research is that we paid particular attention to reliability and to the filtering of distorting content at every point of data extraction. The exact details of the process are described in this chapter.

3.1 Article collection

Unfortunately, the GDELT package is not suitable for extracting raw texts, so we had to do this manually. To get a more focused scope we filtered to the actor: EU, and extracted the domains from the links of each item.

In GDELT's article collection process, the diversity of news sources is more emphasized than their reliability, so we guaranteed the factuality with a further filtering process. For this purpose we used the https://igniparoustempest.github.io/mediabiasfactcheck-bias/ repo, which summarizes on of the most citated media bias resource, https://igniparoustempest.github.io/me bias/ . MBFC calculates bias on 4 levels, namely: Biased Wording/Headlines- Does the source use loaded words to convey emotion to sway the reader. Do headlines match the story? Factual/Sourcing- Does the source report factually and back up claims with well-sourced evidence. Story Choices: Does the source report news from both sides, or do they only publish one side. Political Affiliation: How strongly does the source endorse a particular political ideology? Who do the owners support or donate to? For our research the Factual/Sourcing level is the most relevant, in which the derived database distinguishes 3 categories, high, mixed and questionable. We accepted the first two of these and excluded the questinable ones. After the filtering from the overall 9497 news sources only 719 remained, to which 84,280 articles belong. The list of filtered news sources is available in the attachment. Unfortunately many articles were protected or deleted, thus after running our custom scraper script 27777 articles were collected, the loss was 67

4

- Bag-of-words method / TF-IDF method

4.1 Feature generation workflow

$$log\left(1 + f_{t,d}\right) \cdot log\left(1 + \frac{N}{n_t}\right) \tag{1}$$

TEST: As can be seen in equation 1...

5 Results

In this section we present the TF-IDF feature produced by our method described in **??**. First, we interpret the TF-IDF feature as an index for the presence of global warming in the news. Relying on visualization of the results we

Then, we use ARIMAX and ElasticNet models to test the forecasting ability of the index.

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

Global warming is a widely discussed topic in the media. There are multiple reasons why climate change can appear in the news. Wildfires, policy changes, energy prices or technological breakthroughs all draw attention to the topic. Therefore, the TF-IDF feature that incorporates the most important keywords of climate change can track the cycles in the discussion.

Since policy uncertainty is substantial around the ETS system it is essential to somehow measure the intensity of the discussion around it. One of the key drivers of the ETS prices are the EU's emissions targets which set a cap on the number of ETS allowances. Nevertheless, various other policy measures impact carbon prices as sectoral policies, like green energy mandates.

We present the evaluation of our climate change news index between January 1, 2020 and November 30, 2022 on Figure 4. The index is highly volatile, but several cycles are outlined in the 30-day moving average. In January and February 2020 the index reached a relatively high and stable level due to the recent presentation of the EU Green Deal. Then, in March, 2020 the index started to steadily decrease as the COVID-19 pandemic overtook the news headlines. However, soon the concept of 'green recovery' emerged and climate change keywords again became trending. In October, 2020 the COVID-19 cases soared pushing down the index. The European Council endorsed a binding EU target of a net domestic reduction of greenhouse gas emissions by at least 55% by 2030, compared to 1990 levels in December leading to a local peak at the end of the year. The proceeding period was less volatile, the index remained in the range of 1.32 and 1.85. Then, the index broke out of the range in June, 2021 as the European Council endorsed the new renovation wave strategy and in July the 'Fit for 55' package was presented. Finally, the climate change news index peaked in November, 2021. The 26th Conference of the Parties (COP26) was held between the October 31 and November 13. The COP26 summit brought parties together to accelerate action towards the goals of the Paris Agreement.

5.2 Regression

Natural language processing is an adequate tool to incorporate the various uncertainties around ETS prices in our forecasts.

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

TEST in Table 1...

Figure 4: Climate change news index between January 1, 2020 and November 30, 2021

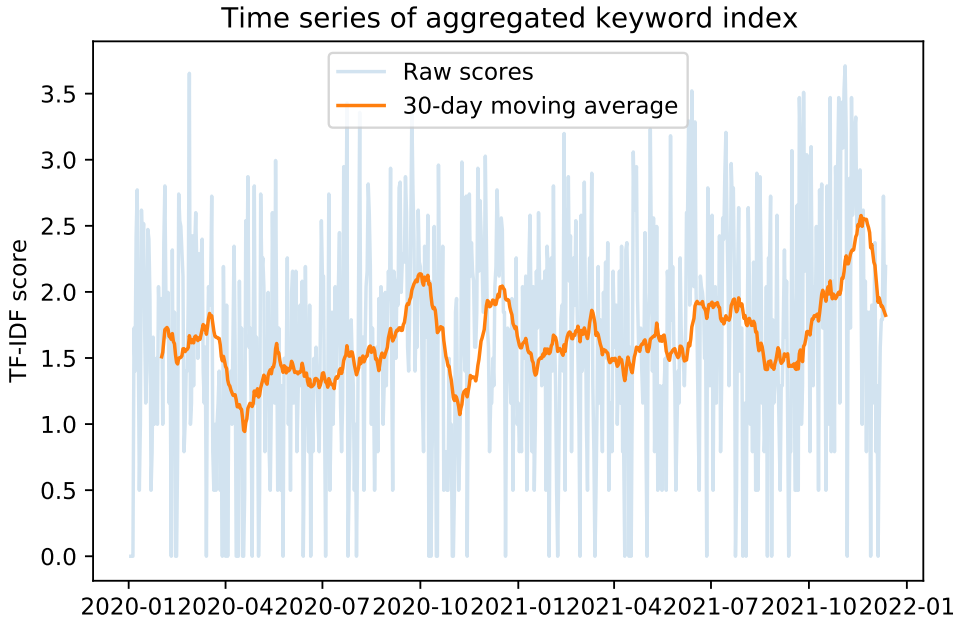


Table 1: Test table

Window	Forecast measure	Model	TF-IDF	Control
50	MAE	ARIMAX	0.019402	0.019420
		ElasticNet	0.020456	0.020484
		Min	**0.019402**	0.019420
	RMSE	ARIMAX	0.026125	0.026126
		ElasticNet	0.027924	0.027878
		Min	**0.026125**	0.026126
75	MAE	ARIMAX	0.017722	0.017763
		ElasticNet	0.018352	0.018392
		Min	**0.017722**	0.017763
	RMSE	ARIMAX	0.024594	0.024610
		ElasticNet	0.025671	0.025643
		Min	**0.024594**	0.024610
100	MAE	ARIMAX	0.016905	0.016971
		ElasticNet	0.017484	0.017543
		Min	**0.016905**	0.016971
	RMSE	ARIMAX	0.023265	0.023629
		ElasticNet	0.024167	0.024137
		Min	**0.023265**	0.023629

6 Conclusions and further research opportunities

We conclude that incorporating unstructured data from news articles improves forecast accuracy for carbon price volatility 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.

- 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

6.1 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.

Haita-Falah, C. (2016). Uncertainty and speculators in an auction for emissions permits. *Journal of Regulatory Economics*, 49(3):315–343.

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.

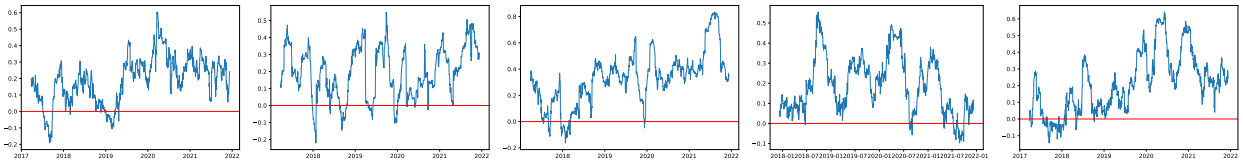
Ye, J. and Xue, M. (2021). Influences of sentiment from news articles on eu carbon prices. *Energy Economics*, 101:105393.

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.

A Data

Figure 5: TEST FIGURE

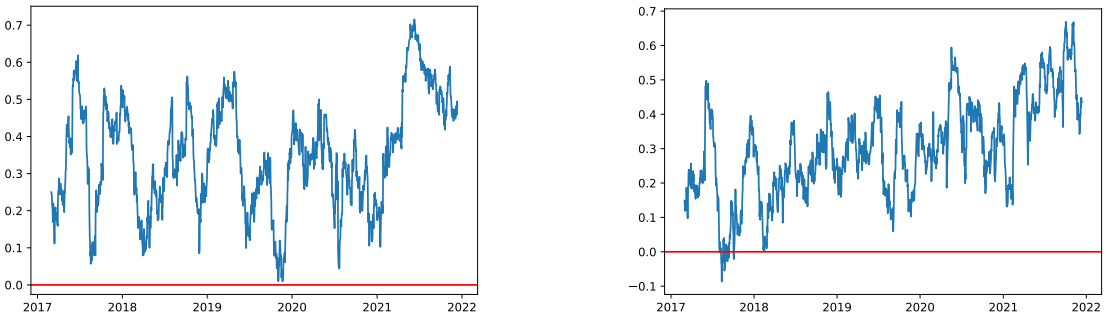


Oil Coal Natural Gas Electricity STOXX Europe 600

TEST FIGURE

TEST: As can be seen on Figure 5...

Figure 6: Rolling 60-days correlation of the TF-IDF scores of the keyword groups



Policy and renewables

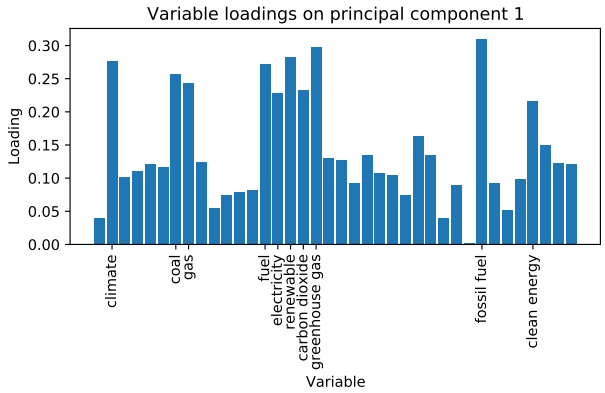
Gas and renewables

TEST FIGURE

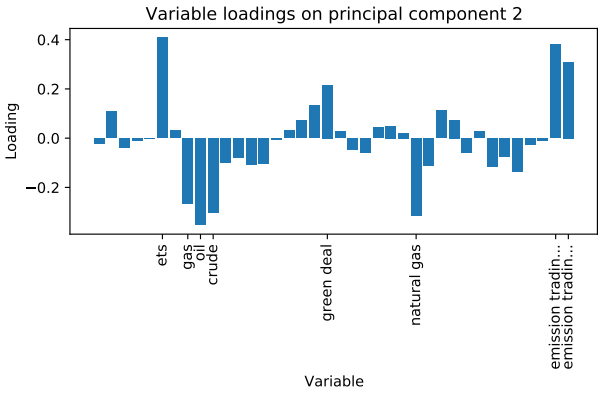
TEST: As can be seen on Figure 6...

TEST: As can be seen on Figure 7...

Figure 7: TEST FIGURE



First principal component



Second principal component

TEST FIGURE