

Deriving Sustainability Data from News Feeds

Thomas Dangl
Vienna University of Technology

Michael Halling
University of Luxembourg

Stefan Salbrechter
Vienna University of Technology

This project represents work-in-progress.



FACULTY OF
ECONOMICS
AND FINANCE

Motivation

- **Central questions in Sustainable Finance:**
 - Are ESG-related risks, in particular carbon risk, systematic (i.e., affect expected rates of returns)?
 - Put differently, do firms with high exposure to carbon risk earn higher expected rates of returns?
- To answer this question, we need to identify firms with large exposure to carbon risk.
- Ideally, such a measure:
 - ...is **comprehensive** and does not just rely on observable emissions.
 - ...is available at a **reasonably high frequency** (e.g., monthly).
 - ...**varies over time**, capturing changes at the firm-level (e.g., when firms improve their carbon footprint).
- One source of information to exploit: **news articles**.

What we do

- Measure the **exposure to carbon risk at the firm-level** and at a high(er) frequency using news.
 - Advantage of news: (potentially) forward-looking.
 - Disadvantage of news: many firms (especially, smaller ones, not listed, etc.) will not be covered.
- Develop an approach – using state-of-the-art **Natural Language Processing techniques** – that is flexible in terms of (a) topic (e.g., can be extended beyond carbon risk) and (b) scope (e.g., can be extended to differentiate between positive and negative news).
- Once firm-level carbon risk measures are in place, study how capital markets respond to this information in terms of **expected returns and traditional risk measures**.



Main Results (Overview)

- Market responds **very negatively** in the month in which firms are associated with carbon risk (**strong contemporaneous effect**).
- This negative effect, however, vanishes quickly during the subsequent 1-6 months.
- Implications (tentatively):
 - Markets efficiently price in (negative) news about carbon risk exposure at the firm-level.
 - Expected returns (moving forward) do not seem to change noticeably.
 - The effect seems to be idiosyncratic rather than systematic.

Related Literature

- Most closely related paper is Engle et al. (2020). They use news to create an aggregate carbon risk index.
 - They count words or use a 3rd-party provider for a more sophisticated approach.
 - They do not use news at the individual firm level. Instead, they conjecture that firms with high ESG-scores (in particular, high E-scores) seem to provide a hedge against innovations in the aggregate carbon risk index.

Data

- **Sample period:** 1996 to the end of 2019.
- **Thomson Reuters News Dataset:** around 13 million news articles
 - Use all news in English. Perform some rough pre-selection based on pre-tagged topic codes.
 - Feed the news through a data cleaning pipeline to remove special characters, identify word stems, etc.
 - Use the news to identify (a) topic maps and (b) the exposure of individual firms to these topic maps.
- **Firm universe:**
 - US-listed firms with a market capitalization above the median (i.e., large firm sample) because small firms are only covered very irregularly in the news.
 - Use monthly return data from CRSP.



Topic Modeling

- We create word embeddings using Word2Vec (shallow neural network with one hidden layer).
- Main idea:
 - Word embeddings represent vector representations (we use 64 dimensions) of individual words and bi-grams.
 - Similar words will be embedded in vectors that are (mathematically) similar (e.g., using cosine similarity).
- Topic clustering:
 - The process needs seed words: here we use “carbon pollution” and “carbon tax”.
 - Iterative clustering algorithm then forms word clusters (topics) based on the distance of words to the plane spanned by the original seed words.

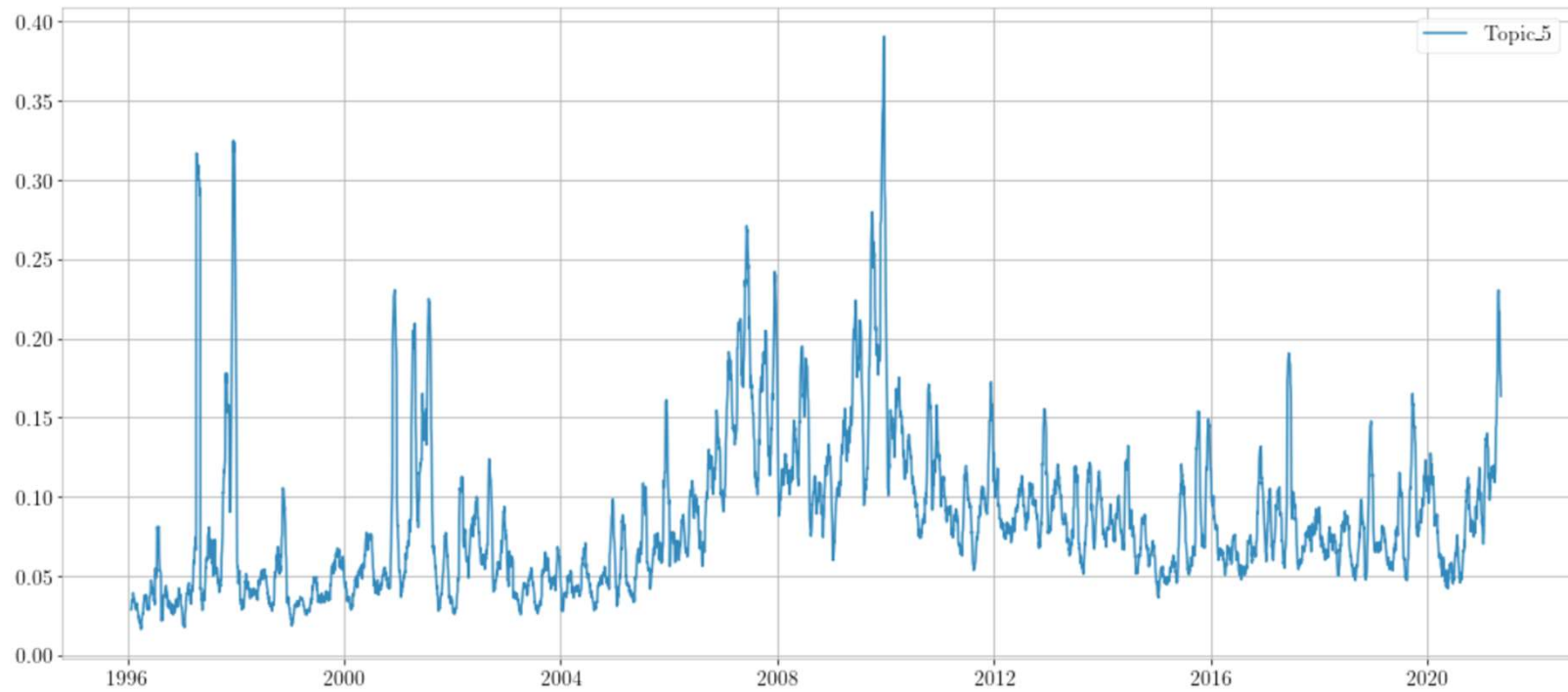


Construction of Carbon Risk News Indices

- A carbon risk news index shows the exposure of an individual firm or the aggregate market to the carbon-risk-topic in the Thomson Reuters news stories from 1996 to 2019.
- Construction of the aggregate index:
 - Weighted count of the topic words for each news article.
 - Counts are calculated at the daily frequency (arithmetic mean) and then aggregated to the monthly frequency.
- Construction of company-specific indices:
 - Only news that are tagged to be related to a given company are used to construct the index.

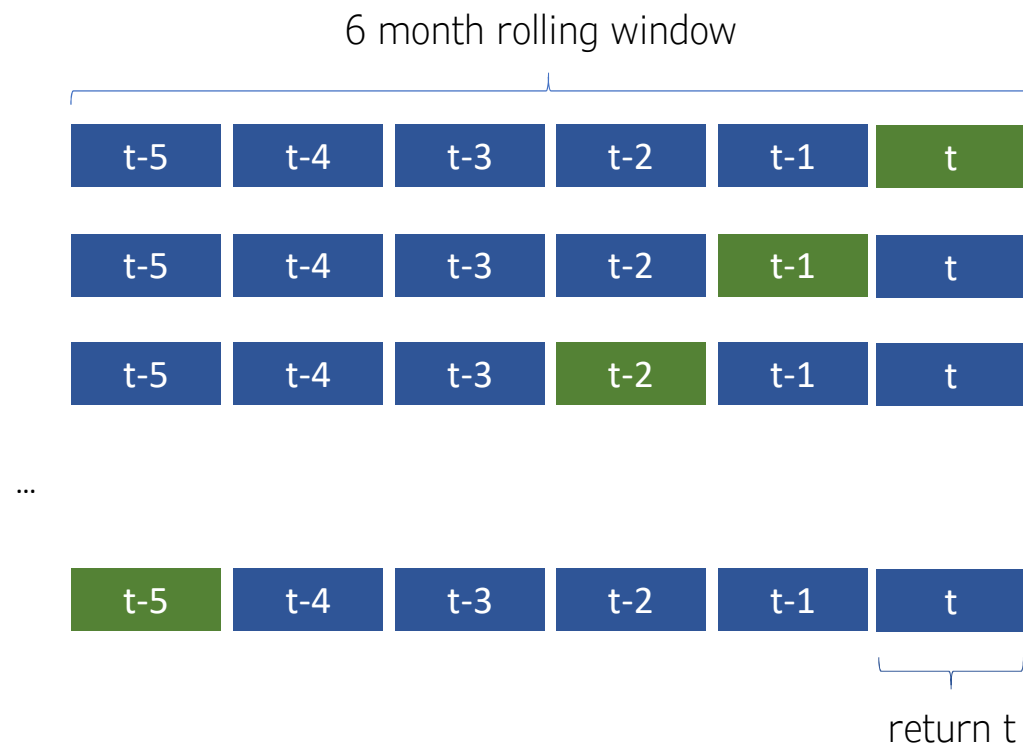


Market-Wide Carbon Risk News Index



Market Response Over Different Horizons

- Invest in a portfolio of stocks that are covered in the carbon-risk-news at $t=0$ (or $t=1, t=2$, etc.)
- The first portfolio is not investable because it looks at the contemporaneous effect.



Market Response Over Different Horizons

- Port1 = portfolio based on news at $t=0$ (blue line):
 - Shows strong underperformance relative to the S&P 500 (orange line).
 - Consistent with the idea that market prices drop when firms are mentioned in news articles on carbon risk.
- But: if the news coverage happened 5 (port 5) or 6 (port 6) months ago, performance differences are small.



Market Response Over Different Horizons

- The bad performance of port1 also translates into a statistically sign. and economically large alpha of -75 bps per month (9% p.a).
- Once we introduce lags, alpha estimates drop and become mostly insignificant.
- Results do not change much if we control for the market-wide news index or for SMB and HML in these regressions.

	<i>Dependent variable: Lagged carbon risk portfolio</i>					
	lag(0)	lag(1)	lag(2)	lag(3)	lag(4)	lag(5)
const	-0.0075*** (-2.7478)	-0.0008 (-0.3899)	-0.0034 (-1.4740)	-0.0036* (-1.7558)	0.0012 (0.5475)	0.0003 (0.1014)
beta _{Mkt-RF}	1.3065*** (21.3467)	1.1219*** (24.0440)	1.1156*** (21.6304)	1.2272*** (27.0140)	1.1825*** (25.1345)	1.2391*** (21.9737)
Observations	300	300	300	300	300	300
R ²	0.4201	0.6926	0.6353	0.7271	0.6439	0.6250
Adjusted R ²	0.4181	0.6916	0.6341	0.7262	0.6427	0.6237
Residual Std. Error	0.0731	0.0349	0.0395	0.0352	0.0401	0.0455
F Statistic	215.8428***	671.5516***	519.0711***	793.8465***	538.7262***	496.5877***

Note:

*p<0.1; **p<0.05; ***p<0.01



Conclusion & Further Research

- NLP can be used to identify firms that are exposed to carbon risk based on news (at least, for large, listed firms).
- When firms are covered by the news, the market views this negatively in the same month. Over 6 months, however, firms usually recover.
- Further research:
 - Perform standard cross-sectional asset pricing tests – including the definition of a carbon risk factor – to assess whether carbon risk is a systematic source of risk or not.
 - Further improve the NLP approach to explicitly disentangle positive carbon risk news from negative carbon risk news.

