Company-Specific News and High-Frequency Equity Prices: Green impact and greenwashing

Massimiliano Caporin[a] Yufeng Mao[a] Sandra Paterlini[b]

[a]University of Padova

[b]University of Trento

2 December 2024 University Ca' Foscari Venice

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

GRINS FOUNDATION

Motivation

- Companies face growing scrutiny from regulators, investors, and the public regarding their environmental practices, making it vital to assess how such pressures influence market valuations.
- Advances in data accessibility enable the collection of high-frequency trading and news data, providing a unique opportunity to study immediate market reactions.
- The release of macro-related and company-specific news impacts on equity prices, even leading to sudden movements (jumps), and the impact might also be modulated by the sentiment we could extract from news
- Given the role of the environmental exposure of companies, the adoption of ESG-related practices, and the growing interest for these aspects both from the point of view of investor and regulators, do environmentally-related news (green news) impact on equity prices? Is their impact comparable to news related to, say, earnings announcements or macro-related news?

Motivation

- Greenwashing refers to misleading claims about the environmental soundness of firms' practices.

  In April 2023, the Wall Street Journal reported that "nearly three-quarters of corporate leaders say most organizations in their industry would be caught greenwashing if they were investigated thoroughly".

- Can we develop a company-specific greenwashing index derived from news data? What would be the impact of changes in this index on (high-frequency) equity prices? If we are capable of detecting greenwashing-related news, is their impact equal to other green news?

Related Literature

Recent studies have sought to construct greenwashing indices using diverse data sources.

- Gourier and Mathurin (2024) construct a news-implied index of greenwashing which measures the fraction of climate-related news articles mentioning firms' greenwashing.

  Gourier and Mathurin (2024) apply an advanced Natural Language Processing (NLP) algorithm to the history of paper-based Wall Street Journal articles. The algorithm is trained to identify the articles that allude to greenwashing.

  They show that greenwashing impacts investors' behavior. Unexpected increases in the greenwashing index are followed by decreases of flows into funds advertised as sustainable, both for retail and institutional investors.

Related Literature (Cont.)

- Lagasio (2024) constructs the ESG-washing Severity Index (ESGSI) for quantitatively assessing discrepancies between portrayed and actual sustainability practices in corporate disclosures.

  Lagasio (2024) uses advanced NLP techniques and analyze sustainability reports from 749 listed companies, integrating sentiment analysis with the frequency of sustainability terms to calculate the index.

  The findings reveal significant variation in ESG-washing practices across industries and geographical regions, highlighting the need for stricter sustainability reporting standards and more effective regulatory frameworks against ESG-washing.

Related Literature (Cont.)

Several recent papers have tried to identify greenwashing firms.

- Bingler, Kraus, Leippold, and Webersinke (2022) use a BERT model trained on climate resources to assess whether companies' TCFD disclosures are meaningful or "cheap talk."
- Bingler, Kraus, Leippold, and Webersinke (2024) analyze companies' annual reports to investigate greenwashing practices.
- Kacperczyk and Peydró (2022) find evidence of greenwashing by banks.
- Parise and Rubin (2023) find that mutual funds manipulate sustainability ratings by strategically timing purchases and sales of sustainable assets.
- Dumitrescu, Gil-Bazo, and Zhou (2023) propose a definition of greenwashing for mutual funds and find that a third of self-labeled ESG funds qualify as greenwashing under this definition.

Contribution

- First step: verify if green news do have a differential impact on high-frequency equity prices when compared to other news related, for instance, to release of company indicators (EPS) or macro-related news
- Use an event-study framework extended in a high-frequency framework and panel form to derive proper inferential procedures; possibly extend the model by using pervasive common factors (beyond the market - latent factors)
- Second step: introduce in the evaluation greenwashing news and greenwashing index and determine how these event impact on equity prices; similar approach as in the previous case (which represents the baseline)

Data

- High frequency equity prices data
- Focus on European data
- Database: Eurofidai dataset including equities belonging to the Stoxx600 index, traded in Euronext, Xetra and LSE
- Coverage: from 2010 to 2023 at the 1-second frequency
- Content: 378 equities (79 Germany, 104 France, 176 United Kingdom, 19 other); data for France do present some missing periods - not a big concern as we will focus on events and their impact on prices in a (short) event window

Data (Cont.)

- Factset StreetAccount News data:

  StreetAccount News provides concise, real-time, and actionable financial market updates, covering corporate events, macroeconomic data, and market trends.

  The news is labelled by industries, company identifiers, and subjects.

- We aim to download historical StreetAccount News, ideally for the same period covered by Eurofidai data.

  However, the data provider's application and API are optimized for downloading news data in real-time (!) or from a recent past period (!!), rather than for accessing extensive historical news data.

  Currently discussing with the FactSet team to determine the most suitable solution for downloading the historical news data covering a wide range of companies for a long time period.

Data (Cont.)

- Example of Factset StreetAccount News data:

FACTSET | streetaccount

Apple to settle EU antitrust investigation into its mobile payments system -- FT ($216.67, 0.00)
Tuesday, June 18, 2024 11:47:55 AM (GMT)

- Citing three people familiar with the situation, the article notes AAPL plans to end an investigation by the EC into its mobile payments system which started in 2022, and avoid a large fine, through a series of concessions that will allow competitors more access to its contactless NFC technology.
- *Editor's Note: This comment was revised to include distribution for Regulatory Antitrust, Articles/Reports, Conjecture/Published Reports and ESG (L&G) (18-Jun-2024, 07:59 ET)*

Reference Links:
- Financial Times

Industries: Computer Hardware
Primary Identifiers: AAPL-US
Related Identifiers: AAPL-US
Subjects: Antitrust (DOJ, FTC, EC, etc.), Articles, Reports, Conjecture, ESG, Leadership & Governance, Media Summaries, Published Reports, Regulatory
Related Stories:
- Apple offers commitments to settle EU'S Apple pay probe ($188.63, 0.00)
- EC sends Apple Statement of Objections regarding Apple Pay ($157.65, 0.00)

Detection of Greenwashing in News Data

We propose to use ChatGPT to detect news with potential indications of greenwashing (alternative: DeepInfra).

There are several applications in the literature demonstrating the use of ChatGPT for data generation:

- Eloundou et. al. (2024) propose a framework to generate exposures of occupations to large language models using ChatGPT.
- Eisfeldt et. al. (2023) use ChatGPT to generate measure of firms' workforce exposures to Generative AI.

Detection of Greenwashing in News Data (Cont.)

To detect greenwashing using ChatGPT, based on the framework outlined in Eloundou et al. (2024), the following key elements need to be determined:

- **Greenwashing rubric**: A structured framework that provides detailed definitions and classifications for different levels of greenwashing.

  This rubric categorizes greenwashing based on the severity, intent, and impact of misleading claims, offering clear distinctions to evaluate and rank instances of greenwashing.
- **Greenwashing rubric prompt**: The carefully designed input prompt provided to ChatGPT to guide it in generating results for identifying potential greenwashing instances.
- **Robustness checking scheme**: Given the inherent variability and potential instability of large language models, conducting robustness checks on classification results is crucial to ensure their reliability and replicability.

Econometric Framework

- To evaluate the impact of the greenwashing activities on equity prices in a high-frequency setting, we propose extending the traditional event study framework in finance to accommodate high-frequency data.

  This approach will allow for a more precise evaluation of the immediate and dynamic effects of greenwashing-related news on stock prices.

- The current literature usually apply the basic event study framework even in the high frequency context, without adapting it to account for the distinct characteristics of high-frequency data.

  We seek to address these limitations, providing a more robust framework for understanding the influence of greenwashing activities on financial markets.

The Market Model

We define the following market model for each firm i

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it}, \quad t = 1, \ldots, T_i, \tag{1}$$

where R_{mt} is market return and ε_{it} is the idiosyncratic error. Then we write the market model in matrix form

$$R_i = X_i \theta_i + \varepsilon_i, \tag{2}$$

where

$$R_i = \begin{pmatrix} R_{i1} \\ \vdots \\ R_{iT_i} \end{pmatrix}; \quad X_i = \begin{pmatrix} 1 & R_{m1} \\ \vdots & \vdots \\ 1 & R_{mT_i} \end{pmatrix}; \quad \varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}; \quad \theta_i = \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix}. \tag{3}$$

Measures of Abnormal Returns

Define the event window ξ_i consists of observations $t = t_{0i} - k_i, \ldots, t_{0i} + k_i$, and let $\tau_i = 2k_i + 1$ be the length of the event window. Define the estimation window ζ_i consists of observations $t = t_{0i} - h_i, t_{0i} - h_i - 1, \ldots, t_{0i} - h_i - T_i + 1$, where $h_i > k_i$. The event window errors (abnormal returns) are defined as

$$\varepsilon_i^* = R_i^* - X_i^* \theta_i, \tag{4}$$

where

$$R_i^* = \begin{pmatrix} R_{i,t_{0i}-k_i} \\ \vdots \\ R_{i,t_{0i}+k_i} \end{pmatrix}; \quad X_i^* = \begin{pmatrix} 1 & R_{m,t_{0i}-k_i} \\ \vdots & \vdots \\ 1 & R_{m,t_{0i}+k_i} \end{pmatrix}; \quad \varepsilon_i^* = \begin{pmatrix} \varepsilon^*_{i,t_{0i}-k_i} \\ \vdots \\ \varepsilon^*_{i,t_{0i}+k_i} \end{pmatrix}. \tag{5}$$

The estimated event window errors (abnormal returns) are calculated as

$$\widehat{\varepsilon}_i^* = R_i^* - X_i^* \widehat{\theta}_i, \tag{6}$$

where $\widehat{\theta}_i$ is the estimated parameters obtained from the market model (2) using the estimation window data.

Aggregating Abnormal Returns across the Cross-Section

For simplicity, we change the time index t to the time relative to the origin t_{0i} (event time). Then we can write $\widehat{\varepsilon}^*_{i;e}$, $e = -k_i, \ldots, k_i$. Define the estimated cumulative abnormal return

$$\widehat{CAR}_i(c_i) = \sum_{e=-k_i}^{k_i} c_{i;e}\widehat{\varepsilon}^*_{i;e}, \tag{7}$$

where $c_i = (c_{i,-k_i}, \ldots, c_{i,k_i})'$ are weights. If we take $c_i \in \mathbb{R}^{\tau_i} = (1, 1, \ldots, 1, 0, \ldots, 0)'$ with the first zero in position $k_i + l_i + 2$, then $\widehat{CAR}_i(c)$ is the sum of abnormal returns up to the period $l_i \leq k_i$. Note that

$$\widehat{\varepsilon}^*_i = R^*_i - X^*_i\theta_i - X^*_i(\widehat{\theta}_i - \theta_i) = \varepsilon^*_i - X^*_i(\widehat{\theta}_i - \theta_i) = \varepsilon^*_i + o_P(1). \tag{8}$$

It follows that the distribution of $\widehat{CAR}_i$ is unknown if the error distribution is not specified. Also, since the event window is usually small in practice, we cannot rely on a CLT to obtain the asymptotic distribution of the cumulative abnormal returns. Thus, valid inference about the (cumulative) abnormal returns cannot be made in this case.

Aggregating Abnormal Returns across the Cross-Section

One solution is to aggregate the abnormal returns across the cross-section with similar types of events and rely on the large cross-section to impose a CLT.

We suppose that the event times may be different across firms and are ordered so that $t_{01} \leq t_{02} \leq \cdots \leq t_{0n}$.

Denote $\varepsilon^* = (\varepsilon_1^{*\prime}, \ldots, \varepsilon_N^{*\prime})'$ and $\widehat{\varepsilon}^* = (\widehat{\varepsilon}_1^{*\prime}, \ldots, \widehat{\varepsilon}_N^{*\prime})'$. Then ε^* represents all the potential and actual information we have about the effect of the event. When the event windows for different firms do not overlap at all, we have $\Omega^* = E[\varepsilon^* \varepsilon^{*\prime} | \mathcal{X}]$ is block diagonal with diagonal elements $\Omega_{ii}^* = E[\varepsilon_i^* \varepsilon_i^{*\prime} | \mathcal{X}]$. Define

$$CAR(c) = \sum_{i=1}^{N} \sum_{e=-k_i}^{k_i} c_{i;e} \varepsilon_{i;e}^*, \tag{9}$$

and

$$\widehat{CAR}(c) = \sum_{i=1}^{N} \sum_{e=-k_i}^{k_i} c_{i;e} \widehat{\varepsilon}_{i;e}^*. \tag{10}$$

Aggregating Abnormal Returns across the Cross-Section

Further, we define the standardised cumulative abnormal returns

$$SCAR(c) = \frac{CAR(c)}{\sigma}, \tag{11}$$

where $\sigma = [Var(CAR(C))]^{1/2}$, and

$$\widehat{SCAR}(c) = \frac{\widehat{CAR(c)}}{\widehat{\sigma}}. \tag{12}$$

where $\widehat{\sigma} = \sigma + o_P(1)$. Given $\widehat{\theta}_i = \theta_i + o_P(1)$ as $T \to \infty$, it is easy to show that $\widehat{SCAR}(c) = SCAR(c) + o_P(1)$ as $T \to \infty$. We may check Lindeberg's condition and apply a CLT to $SCAR(c)$.

Then we have $SCAR(c) \to_D N(0,1)$ as $N \to \infty$ given Lindeberg's condition holds for SCAR(c). The specific form of σ and $\widehat{\sigma}$ can be easily figured out.

Next steps

- Finalize the construction of the news dataset
- Define the most appropriate prompt for separating green news from other news (which would also be classified according to the information privided by Factset)
- Define the prompt for greenwashing detection

References

Bingler, J., Kraus, M., Leippold, M., & Webersinke, N. (2022). Cheap talk in corporate climate commitments: The effectiveness of climate initiatives. *Swiss Finance Institute Research Paper*, (22-54).

Bingler, J. A., Kraus, M., Leippold, M., & Webersinke, N. (2024). How cheap talk in climate disclosures relates to climate initiatives, corporate emissions, and reputation risk. *Journal of Banking & Finance*, 164, 107191.

Dumitrescu, A., J. Gil-Bazo, and F. Zhou (2023). Defining greenwashing. *Working Paper*.

Eisfeldt, A. L., Schubert, G., Zhang, M. B., & Taska, B. (2023). The Labor Impact of Generative AI on Firm Values. *Available at SSRN 4436627*.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. Science, 384(6702), 1306-1308.

Gourier, E., & Mathurin, H. (2024). A greenwashing index. *Available at SSRN 4715053*.

References

Kacperczyk, M. T., & Peydró, J. L. (2022). Carbon emissions and the bank-lending channel. *Available at SSRN 3915486*.

Lagasio, V. (2024). ESG-washing detection in corporate sustainability reports. *International Review of Financial Analysis*, 103742.

Parise, G. and M. Rubin (2023). Green window dressing, *Working Paper*.

FactSet

- The GRINS network has access to FactSet
- We do have 150 licenses, currently 90 allocated, if you need a FactSet license for GRINS activities, email me!
- We also have: i) API access to FactSet (1 access to be transferred across active licenses); ii) access to FactSet streetaccount - can be activated on demand, access is through an FTP server, three licenses are allowed to include other licenses in the whitelist capable of accessing the FTP server; in case you are interested, email me!
- Please report to me problems in accessing the provider, and issues with data coverage that you believe are of potential interest for the entire network.