

---

# Text analysis in a capital market context

---

# Mari Paananen



- Associate professor of Business Administration at the School of Business, Economics & Law since 2017
- PhD University of North Texas
- Worked in the US and the UK for +10 years.
- Research focuses on capital market effects on corporate disclosure and firm behaviour
- Manager Business IT Lab (BitLab)



---

# Text sources in business research

- Financial reporting (and sustainability reports)
- Press releases
- Earnings Conference Calls (ECC)
- News items (Factiva)

---

# Earnings Conference Calls (ECC)

- A conference called by companies
- Some analysts are invited, other can listen in
- The audio of these calls is publicly available
- The calls comprise
  - An introduction
  - A Q&A session



---

# Text analysis using NLP

- Topic modelling using Latent Dirichlet Allocation (LDA)
- Bag of word approach to locate sentiment in a financial setting (Loughran & McDonald 2011)
- Cosine similarities
- XGBoost
- Named entity recognition (NER)
- LLM



---

# DO ANALYSTS' QUESTIONS CONFUSE INVESTORS?





---

# Background

- Prior research suggests that investors have limited information processing power
- Do diverging financial analysts' questions cause increase disagreement among investors?

---

# Sample

- US firms from 2007 to 2019 covering the following industries:
  - Capital Goods,
  - Energy, Software & IT,
  - Materials, and
  - Health Care Equipment & Services.

	<b>Total</b>
<b>Initial sample of ECCs</b>	<b>48,879</b>
<b>Less ECCs with format issues</b>	- <b>3,978</b>
<b>Less firms-quarters missing in IBES/Compustat/CRISP</b>	- <b>9,053</b>
<b>Final sample</b>	<b>35,848</b>



---

# Proxies for divergence

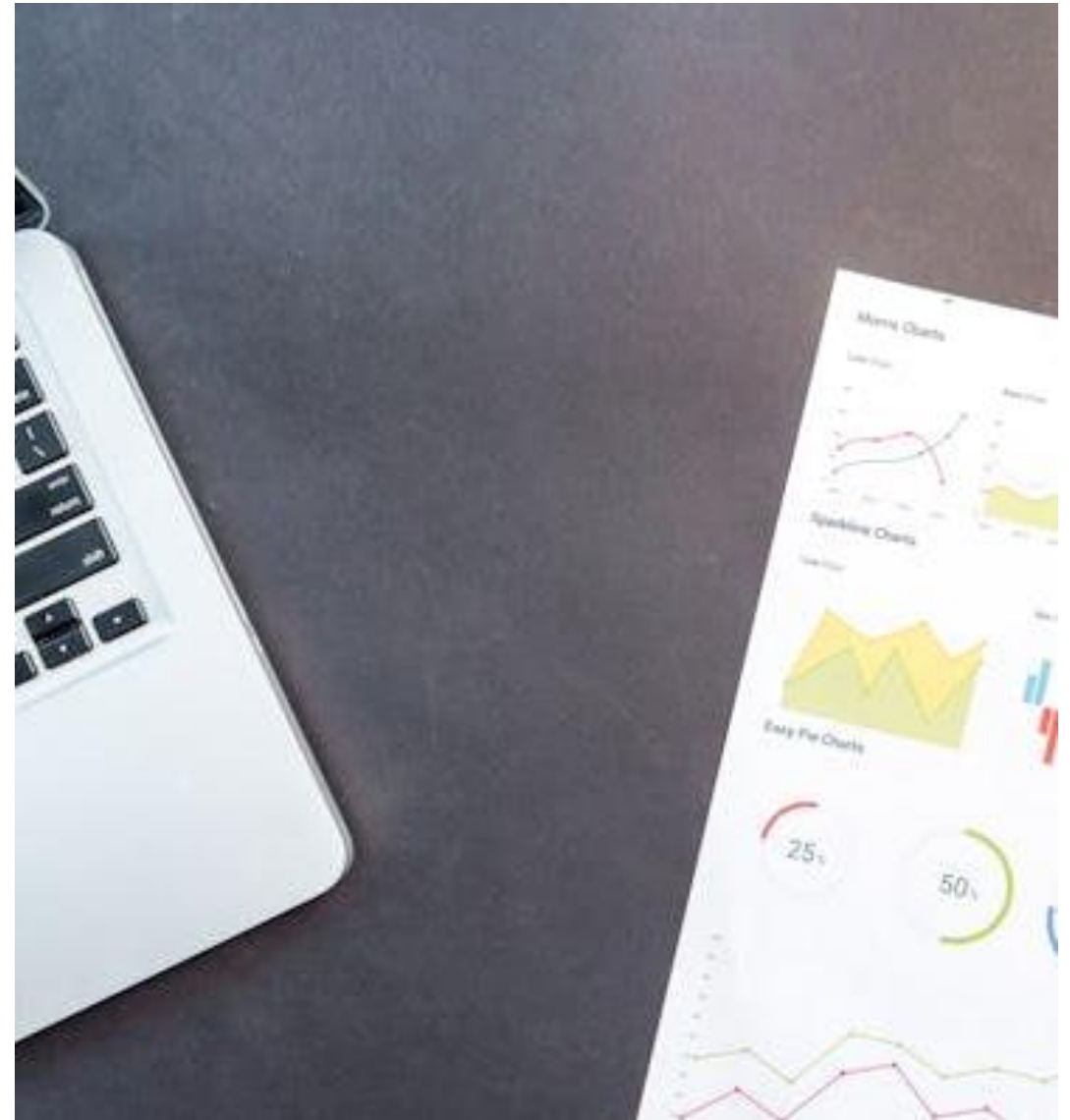
- Number of questions
- Standard deviation of analyst questions:
  - Novelty (divergence from introduction)
  - Themes divergence (among analysts' questions)
  - Tone (pos/neg) (L&M 2011)



---

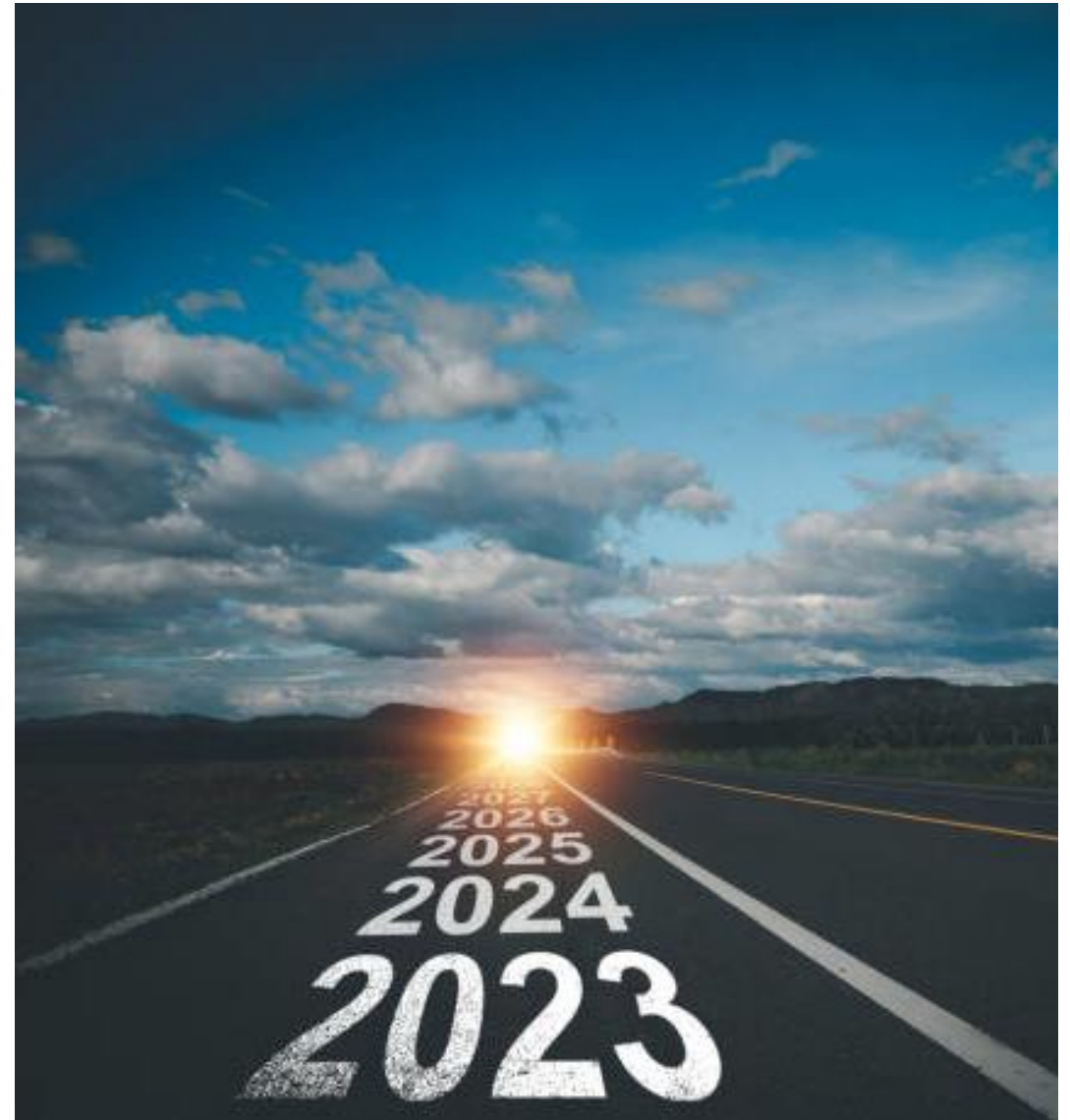
# Results

- All proxies for divergence are positively associated with an increased trading volume around the ECC
- Inference: Variability in analyst reaction/behaviour influence investors' reaction to ECCs (disagreement about firm value)



---

# DO ECC DISCUSSIONS AFFECT POST-EARNINGS- ANNOUNCEMENT DRIFT (PEAD)





- PEAD is the tendency for a stock's cumulative abnormal returns to drift for several weeks (even several months) following the positive earnings announcement.
- Traditional research analyse PEAD based on traditional measures of earnings surprises.
- Do analysts' limited processing power make it harder for them to process text, causing longer/higher PEAD?

---

# Proxies for text content that might affect PEAD

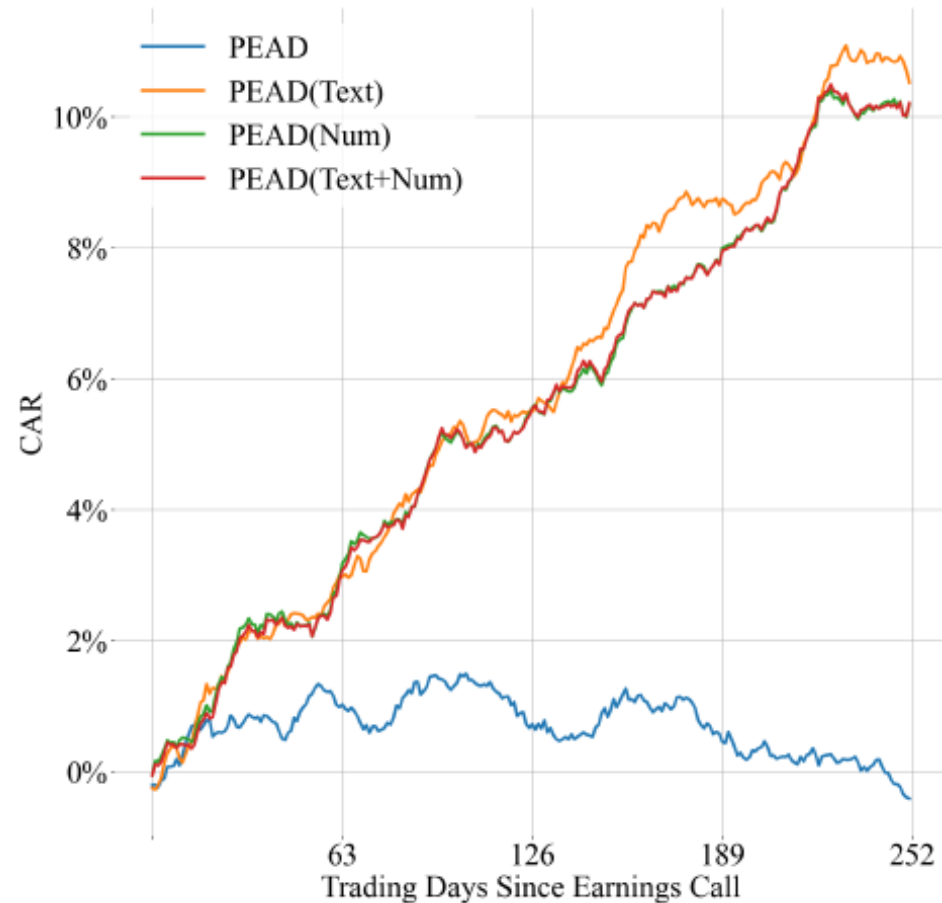
- Themes
- Novelty
- Specificity
- L&M (2011) characteristics:
  - Tone (pos/neg)
  - Complexity
  - Uncertainty



# Results

- Text based portfolios (long position in high text variables and short other text variables), appear to be a better investment strategy.

Figure 1: Comparison of classic PEAD and PEAD derived from text or numbers



---

# IS THERE A DEMAND FOR DISCLOSURE ON CORPORATE ENVIRONMENTAL IMPACT?



---

# Double materiality

- **Financial materiality** - focuses on investors, creditors, and other stakeholders' decision-making
- **Environmental/Sustainability Materiality** - the environmental and social impacts of a company's operations, as well as its broader contribution to sustainable development.







---

# Sustainability reporting standards and double materiality

- ESRS, TNFD implicitly applies the concept of double materiality
- ISSB more centered on the entity's decision-making of specific stakeholders (financial materiality)
- **What disclosure do capital market participants want, and how would that affect investment decisions?**

---

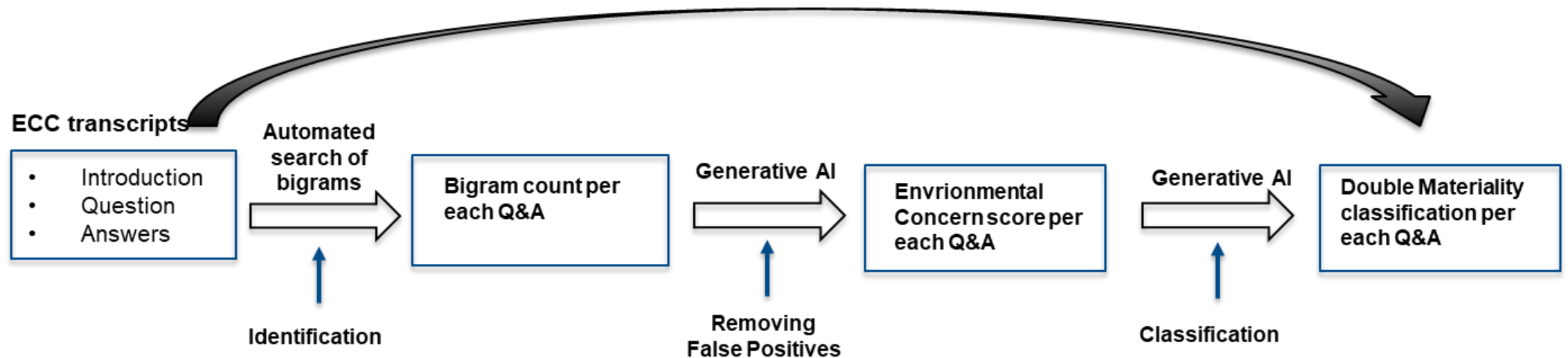
# Operationalisation

- Using earnings conference calls (ECC) to measure what type of information financial analysts demand (financial and/or sustainable materiality)
- Analyse the impact of the disclosure of the information on cost of capital



---

# How to extract the data from ECCs



---

# Example output

## Analyst question

On your B.C. operations what percent would be beetle kill going through mills right now?

## Management response

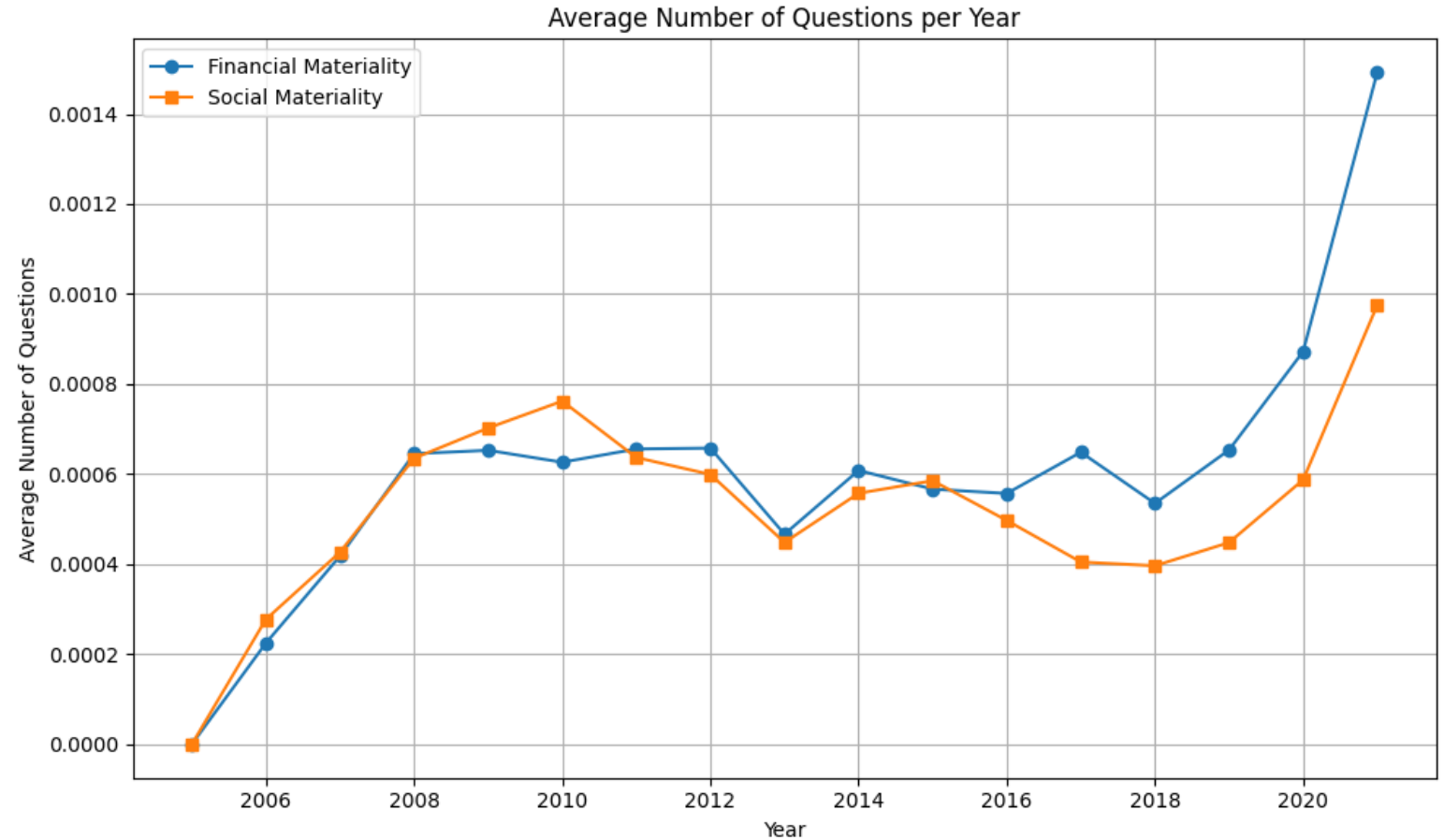
We're approximately 80% of the volume going through our mills today is from mountain pine beetle impacted stands. Not all of those trees obviously are all attacked by mountain pine beetle.

## ChatGPT rationale for classification as DM

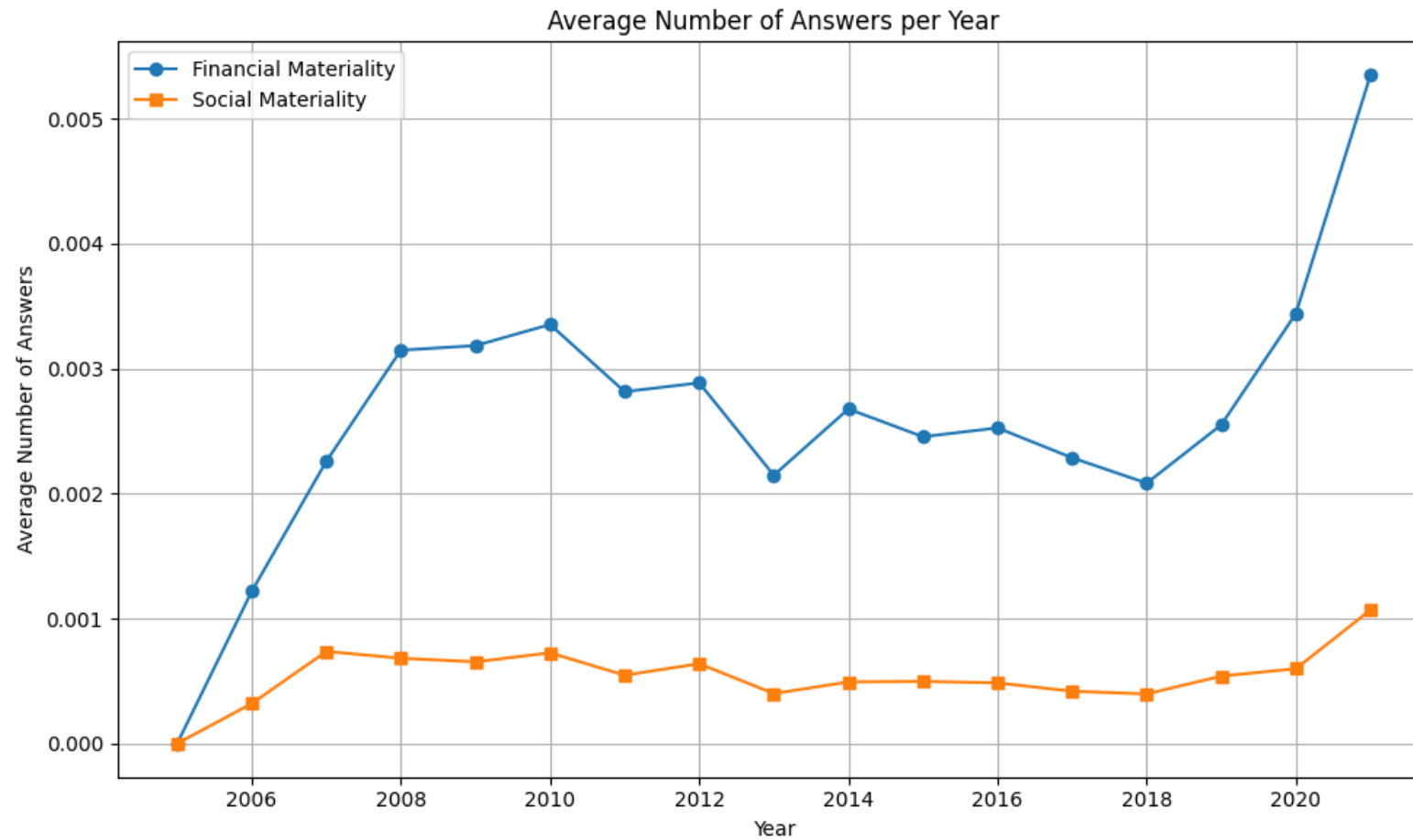
The primary focus of the text is on the environmental impact of the mountain pine beetle on the entity's operations. The mention of "mountain pine beetle impacted stands" and "trees attacked by mountain pine beetle" indicates a concern for the ecological sustainability of the entity's activities. This aligns with the concept of social and environmental materiality, which encompasses the entity's impact on the environment..

---

# ANALYSTS' DEMAND FOR DOUBLE MATERIALITY DISCLOSURE



# Management responses





---

# Results

- Analysts are asking for information related to environmental issues that have to do with *both* financial performance and social/sustainable matters, although more so about financial performance.
- Information related to both financial performance and social/sustainable matters as the disclosure is associated with lower cost of capital

---

# WHEN DO FIRMS RESORT TO HALF- TRUTHS?





---

# Half-truths

Statements that contain some truth but give a false impression because only part of the whole truth is included.





---

# Ideas/Questions

- Is it possible to use:
  - LLM to identify half-truths?
  - Dialogue semantics
- When (and why) do firms resort to half-truths?
- Are investors eventually picking up half-truths?
  - If so, when and why?
  - What are the capital market consequences?



Thank you for listening!