

Using text analysis to study media and sentiment

Alistair Macaulay

University of Surrey

SUERF | ECB | BAFFI Bocconi workshop

April 2025

Overview

Key statistics from text (newspaper, broadcast transcript, social media, etc.):

1. Sentiment

- Dictionary methods: count positive/negative words e.g. Tetlock 2007
- ML methods: BERT, LLMs, + many more e.g. Fraiberger et al 2021, Lopez-Lira & Tang 2023

Overview

Key statistics from text (newspaper, broadcast transcript, social media, etc.):

1. Sentiment

- Dictionary methods: count positive/negative words e.g. Tetlock 2007
- ML methods: BERT, LLMs, + many more e.g. Fraiberger et al 2021, Lopez-Lira & Tang 2023

2. Others, including

- Topics: researcher-specified or data-led e.g. Larsen & Thorsrud 2018, Hirshleifer et al 2025
- Narratives: hand-labelling, semantic similarity, RELATIO Andre et al 2025, Goetzmann et al 2022, Ash et al 2024

Example: Macaulay & Song (2025)

Question: U.S. Yield Curve inverts in September 2019. Did **media narratives** affect the sentiment response?

Example: Macaulay & Song (2025)

Question: U.S. Yield Curve inverts in September 2019. Did **media narratives** affect the sentiment response?

- Separate newspaper articles covering the event according to their principle narrative (LDA).

Example: Macaulay & Song (2025)

Question: U.S. Yield Curve inverts in September 2019. Did **media narratives** affect the sentiment response?

- Separate newspaper articles covering the event according to their principle narrative (LDA).
- Link to the Twitter (now X) activity of those who engage with each article.

Example: Macaulay & Song (2025)

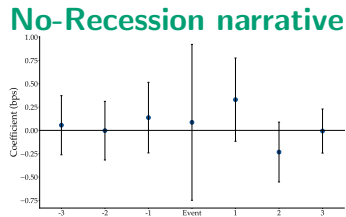
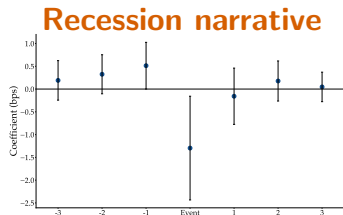
Question: U.S. Yield Curve inverts in September 2019. Did **media narratives** affect the sentiment response?

- Separate newspaper articles covering the event according to their principle narrative (LDA).
- Link to the Twitter (now X) activity of those who engage with each article.
- Measure Twitter sentiment before and after exposure.

Example: Macaulay & Song (2025)

Question: U.S. Yield Curve inverts in September 2019. Did **media narratives** affect the sentiment response?

- Separate newspaper articles covering the event according to their principle narrative (LDA).
- Link to the Twitter (now X) activity of those who engage with each article.
- Measure Twitter sentiment before and after exposure.

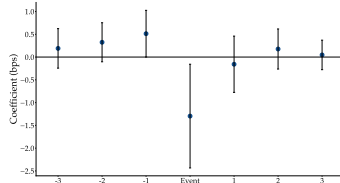


Example: Macaulay & Song (2025)

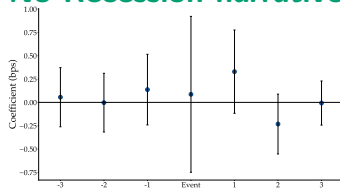
Question: U.S. Yield Curve inverts in September 2019. Did **media narratives** affect the sentiment response?

- Separate newspaper articles covering the event according to their principle narrative (LDA).
- Link to the Twitter (now X) activity of those who engage with each article.
- Measure Twitter sentiment before and after exposure.

Recession narrative



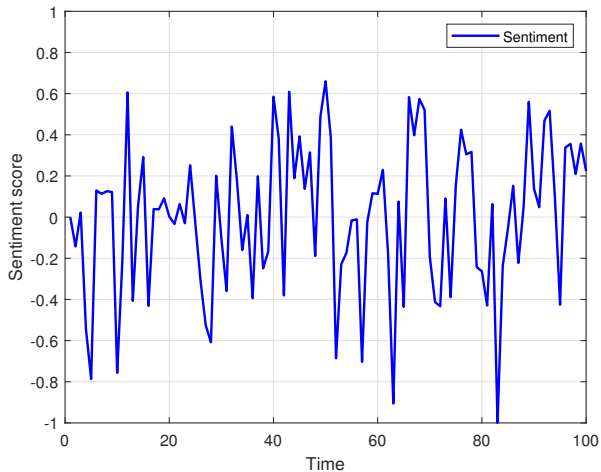
No-Recession narrative



Lesson: when the yield curve inverts again, look at prevailing media narratives to predict sentiment.

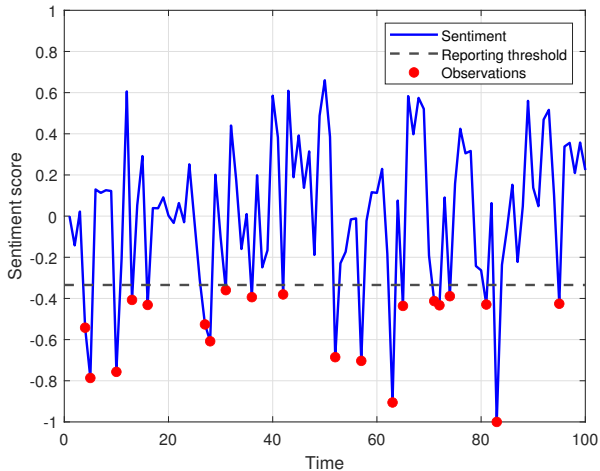
The challenge of editorial decision-making

Some simulated sentiment data: mean=0



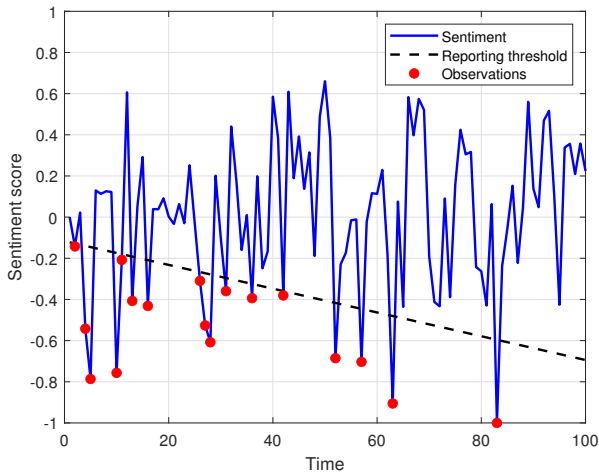
The challenge of editorial decision-making

“If it bleeds, it leads” $\Rightarrow$ biased measurement



The challenge of editorial decision-making

Changing reporting thresholds $\Rightarrow$ spurious sentiment shifts



Solutions

- Use reports with constant frequency.
- Compare reporting within events, across e.g. narratives (Macaulay & Song 2023; 2025), outlet type (Rees & Twedt 2022), language (Wabitsch 2024).
- Econometric fixes
 - Selection models (a la Heckman 1979; Lee 2009)
 - Heuristically: study frequency of reporting as well as content
- Use selection as another data point
 - How does topic selection predict sentiment (Larsen & Thorsrud 2018; 2021)?

Wrap up

- Text data is high-frequency, real time, + opens up questions you can't otherwise reach;
- Media is especially important – media narratives strongly affect sentiment;
- But be mindful of **selective reporting**.

Wrap up

- Text data is high-frequency, real time, + opens up questions you can't otherwise reach;
- Media is especially important – media narratives strongly affect sentiment;
- But be mindful of **selective reporting**.

Opportunities:

- Other forms of social media?
- Different media diets by demographics/regions/other characteristics?
- Systematic tracking of media narratives as a policy input?