

Discussion:

(Almost) 200 Years of News-Based Economic Sentiment

J. H. van Binsbergen, S. Bryzgalova, M. Mukhopadhyay, V. Sharma

Yinan Su

Johns Hopkins University Carey Business School

AFA paper session “Asset Pricing: Big Data and AI”

Jan. 4, 2025, San Francisco

Summary

Text-based measures of economic sentiment for long time series and geographical cross-sections

Overall, very positive sentiment about this paper :)

Summary

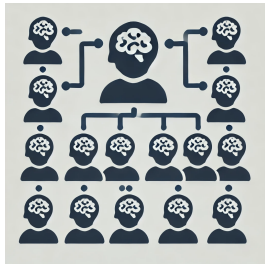
Text-based measures of economic sentiment for long time series and geographical cross-sections

Overall, very positive sentiment about this paper :)

Novelty/contributions:

- ▶ Data
- ▶ Economic relevance
- ▶ Method

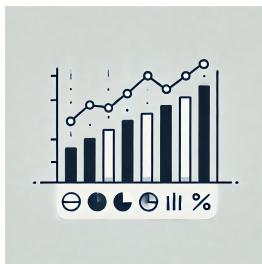
News for economics



Economic agents



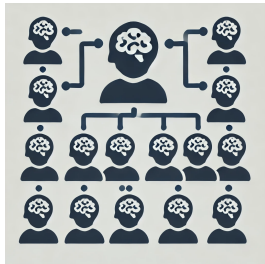
"Think"



Economic outcomes

- ▶ Economic models: left to right
 - "think" based on information, "rational expectations," choose (invest, consume, trade...) based on utility maximization...
 - or "animal spirits", social networks...

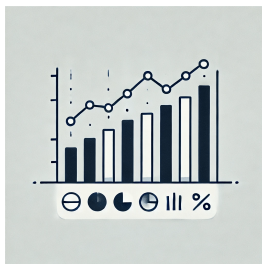
News for economics



Economic agents



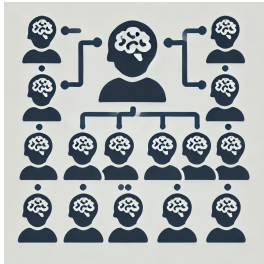
“Think”



Economic outcomes

- ▶ Economic models: left to right
 - “think” based on information, “rational expectations,” choose (invest, consume, trade...) based on utility maximization...
 - or “animal spirits”, social networks...
- ▶ Empirical research: right to left
 - observe historical economic outcomes
 - fit models
 - infer how agents “think”

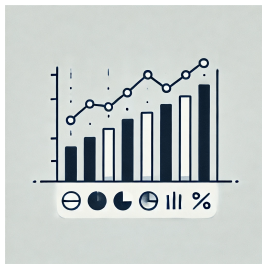
News for economics



Economic agents



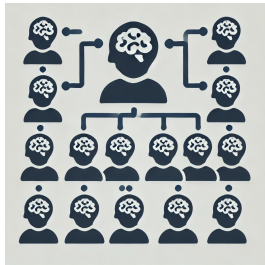
"Think"



Economic outcomes

- ▶ Economic outcomes as data:
Economic activities' historical records

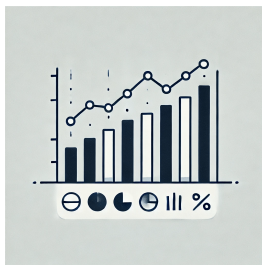
News for economics



Economic agents



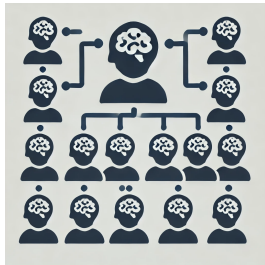
"Think"



Economic outcomes

- ▶ Economic outcomes as data:
Economic activities' historical records
- ▶ Historical records of "think"?
ideal data for being closer to the core of economic model
(no neural link data yet...)

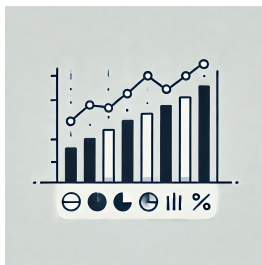
News for economics



Economic agents



"Think"



Economic outcomes

- ▶ Economic outcomes as data:
Economic activities' historical records
- ▶ Historical records of "think"?
ideal data for being closer to the core of economic model
(no neural link data yet...)
- ▶ Proxies:
 - survey data, social media
 - **news textual data**
- ▶ Advantages of news: important, timely, forward-looking

News data are like **tree rings** in climate science



News data are like **tree rings** in climate science



This paper's data contributions

Against this backdrop, this paper:

- ▶ **Expansive** data:
 - 13,000 local newspapers from the United States
 - 200m pages
 - 1b articles
 - almost 200 years

This paper's data contributions

Against this backdrop, this paper:

- ▶ **Expansive** data:
 - 13,000 local newspapers from the United States
 - 200m pages
 - 1b articles
 - almost 200 years

This paper's data contributions

Against this backdrop, this paper:

- ▶ **Expansive** data:
 - 13,000 local newspapers from the United States
 - 200m pages
 - 1b articles
 - almost 200 years
- ▶ This paper sets the standard and provides the critical infrastructure for future research in this area

Comment on data

- ▶ So many local newspapers
~ 20 newspapers in Maryland per year

Comment on data

- ▶ So many local newspapers
~ 20 newspapers in Maryland per year
- ▶ Who are they? examples with names? example articles?
...just to have a concrete sense of the data

100 Years Ago Today: 12/26/1924 (60 issues)



[The midland journal.](#) (8pp.)
Rising Sun, Md.



[Putnam patriot.](#) [volume] (8pp.)
Putnam, Conn.



[Coulee City dispatch.](#) [volume] (4pp.)
Coulee City, Wash.

source: <https://chroniclingamerica.loc.gov/>

lay, December 26th 2024

e Sun eNewspaper

THE BALTIMORE SUN

eNewspaper

Latest Headlines

Sports

News Obituaries

Death Notices

Things To Do

Opinion

Weather

TRENDING:

 Holiday travel poll

 Luigi Mangione case coverage

 Baltimore homicide map

 Md. COVID-19 data

 Top Workple

Bill creating Chesapeake National Recreation Area dies in Congress

After passing the U.S. Senate, a bill establishing a National Recreation Area for the Chesapeake fell short this year.



Barstool Sports founder Dave Portnoy to give \$60,000 to keep Federal Hill pizzeria open



5 things we learned from the Ravens' 31-2 win over the Houston Texans

LATEST HEADLINES

Ravens, MVP hopeful Lamar Jackson crush Texans, 31-2, to take AFC North lead

15 hours ago

One dead in Harford County house fire on Christmas morning

1 day ago

Ravens' Lamar Jackson sets NFL QB rushing record, eclipsing Michael Vick

14 hours ago

Baltimore mayor, boards disagree whether 'immediate action' needed on police oversight powers

2 days ago

Maryland tribe's first female chief fights for identity, equality and preservation

1 day ago

Instant analysis from Ravens' 31-2 win over

END-OF-YEAR SALE! Start 2025 with digital access: 6 months for 99¢.

ACT NOW

[Bay Area](#)

[Election](#)

[S.F.](#)

[Weather](#)

[Puzzles](#)

[Real Estate](#)

[Sports](#)

[Food](#)

[College Guide](#)

[Datebook](#)

[Opinion](#)

WINE

'We only want NDA wines': This secretive practice could get you a \$300 wine for \$30

Here's why some wineries require nondisclosure agreements to sell their wines.

- This Napa Valley Cabernet cost only \$35. Now, it's back after a yearlong hiatus
- This type of Napa wine is notoriously expensive. Here are 6 bottles with excellent value
- The stories that defined California wine in 2024



BAY AREA & CALIFORNIA

The Santa Cruz Wharf needed repairs. A court battle followed instead

- Bay Area could see 30-foot-tall waves again amid high surf advisory

POLITICS

California lawmakers are resurrecting these bills for another shot at becoming law

TOTAL SF

Free wedding gowns and loaner prom dresses? This S.F. group swaps formal wear for free

- This S.F. fire station planned its holiday decor for months — and claims 1st prize in contest
- Did this Netflix show just film the single best episode of San Francisco television?



OPEN FORUM

Here's why I am opting out of TSA facial recognition — and you should

Economic implications

Very intriguing predictive results:

- ▶ Predicts GPD growth (national and state level) controlling for lagged GPD and SPF consensus forecast
- ▶ Sources of predictability
 - in articles about a) the economy b) views of the future
 - via labor market outcomes and consumption
 - not investments or return stock returns
 - ...

Economic implications

Very intriguing predictive results:

- ▶ Predicts GDP growth (national and state level) controlling for lagged GDP and SPF consensus forecast
- ▶ Sources of predictability
 - in articles about a) the economy b) views of the future
 - via labor market outcomes and consumption
 - not investments or return stock returns
 - ...
- ▶ *recent textual analysis in finance and economics is the **tip of the iceberg**...an exciting research agenda [in the future], in which economists gradually expand sourced text corpora and **increasingly refine their ability** to elicit information from that text*
(Chen, Kelly, Xiu, 2023)

About the method

Translating textual data into numerical data

- 1. embedding + 2. dimension reduction

1. (Article-level) embedding:

$$\text{text_in_article}_i \rightarrow \text{emb_vector}_i$$

2. dimension reduction weights for the topic of interest
(economy, sentiment, future vs. current)

$$\text{weights}^{\text{sentiment}}$$

About the method

Translating textual data into numerical data

- ▶ 1. embedding + 2. dimension reduction

1. (Article-level) embedding:

$$\text{text_in_article}_i \rightarrow \text{emb_vector}_i$$

2. dimension reduction weights for the topic of interest
(economy, sentiment, future vs. current)

$$\text{weights}^{\text{sentiment}}$$

- ▶ 1. + 2. = a one-dimensional “score” for each article i

$$\text{weights}^{\text{sentiment}} \top \text{emb_vector}_i$$

About the method

This paper:

- ▶ 1. article representation
 - I think this is a primary component of the methodology and should be more elaborated. I think this part is still bag-of-word rather than contextualized representation?
 - one paragraph: “*final step ... use Smooth Inverse-Frequency (SIF) ...*”

About the method

This paper:

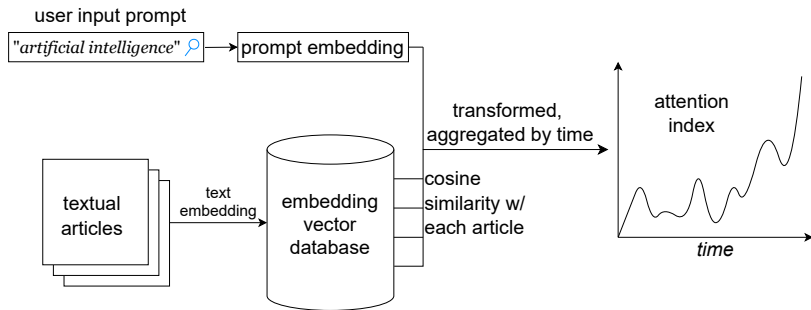
- ▶ 1. article representation
 - I think this is a primary component of the methodology and should be more elaborated. I think this part is still bag-of-word rather than contextualized representation?
 - one paragraph: “*final step ... use Smooth Inverse-Frequency (SIF) ...*”
- ▶ 2. dimension reduction weights
 - a dictionary of sentiment words, weighted by sentiment strength

About the method

This paper:

- ▶ 1. article representation
 - I think this is a primary component of the methodology and should be more elaborated. I think this part is still bag-of-word rather than contextualized representation?
 - one paragraph: “*final step ... use Smooth Inverse-Frequency (SIF) ...*”
- ▶ 2. dimension reduction weights
 - a dictionary of sentiment words, weighted by sentiment strength
- ▶ “word vector approach” vs. word count approach
 - I think the final cosine similarity step is still on the representations produced with a dictionary and word counts. Is this true?
 - “word vector” (i.e. Word2Vec) are for a.) automatic sentiment dictionary generation and b.) weighted word count by sentiment strength

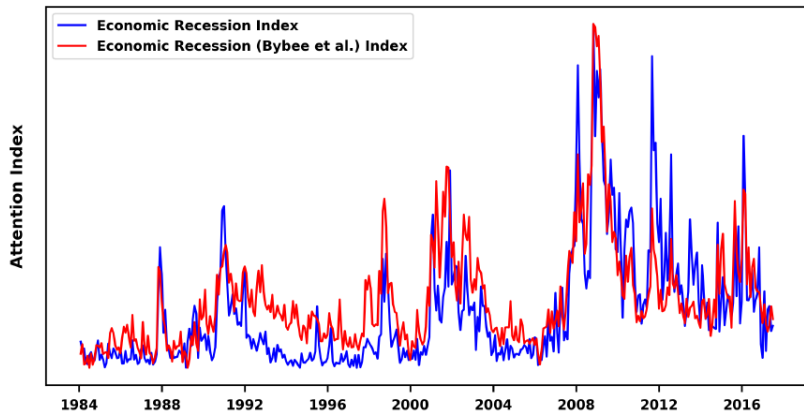
“Prompting Narratives from News Texts” (in progress)



- ▶ prompting **any** topic
- ▶ web-based service open to **all**

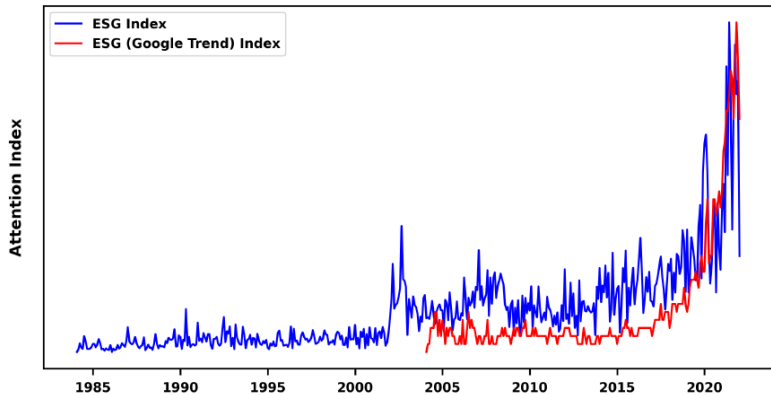
Example: recession

Figure 3: Replicating Economic Recession Index

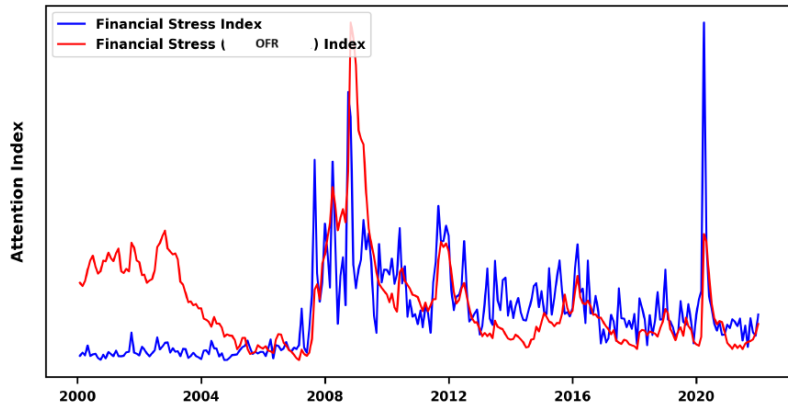


Example: ESG

Figure 5: ESG Index



Example: financial stress



Discussion:

(Almost) 200 Years of News-Based Economic Sentiment

J. H. van Binsbergen, S. Bryzgalova, M. Mukhopadhyay, V. Sharma

Yinan Su

Johns Hopkins University Carey Business School

AFA paper session “Asset Pricing: Big Data and AI”

Jan. 4, 2025, San Francisco