Does finance benefit society? A language embedding approach

Manish Jha, Hongyi Liu, and Asaf Manela

Discussion by Alejandro Lopez-Lira

Summary

- Measure sentiment towards finance using hundreds of years of books
- Eight languages: American English, British English, Simplified Chinese, French, German, Italian, Russian, and Spanish
- Large time-series variation and large level differences between languages
- Finance sentiment declines one year before financial crises
    - No effect afterwards?
- Positive shocks to finance sentiment lead to greater output and credit growth
- Really cool paper!

Traditional approach

- Count number of good sentiment vs. bad sentiment words and divide by number of words
  - Advantage: Extremely simple, good dictionaries in finance proxying for good vs. bad
  - Disadvantage, context is lost, and sentences like ‘there is no way finance is not good for the poor’ will not give anything meaningful

How?

- Fix one sentence:
  - financial services benefit society
  - Project it into a 768-dimension vector space (BERT)
    - The vector space has nice properties
    - It’s built to capture context, meaning and order
    - “In a crisis, we could bank on financing from the government” != “Government’s financing for the bank is in crisis”
    - BERT is open source!

How?

- Compare each sentence (in vector space) against a positive sentiment sentence
- Cosine similarity (in 768 dimension vector space)

$$\text{cosine similarity} = S_C(A, B) := \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

How?

- Cosine similarity (in 768 dimension vector space)

$$\text{cosine similarity} = S_C(A, B) := \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

  - king − man ~ queen - woman -> king − man + woman ~ queen

How?

- Compare ‘finance’ against ‘positive’ - ‘negative’ to proxy for sentiment
  - Makes intuitive sense, but why those specific words?
    - Positive could appear in other contexts (math)
    - Why not, ‘good’ - ‘bad’ or ‘approve’ - ‘disapprove’? Or an average of all?
      - Referee will probably ask about the specifics

Still, it seems to works well!

Positive sentences	Negative sentences
financial services benefit society	financial services damage society
finance is good for society	finance is bad for society
finance professionals are mostly good people	finance professionals are mostly corrupt people
finance positively impacts our world	finance negatively impacts our world
financial system helps the economy	financial system hurts the economy

- It would be nice to have a metric of how well it works relative to simple word counts
  - Perhaps with some supervised examples
  - I think it's obviously better, but again, referee will want the method justified

Data

- Five-word sentences (5-grams) containing the stem of the word “finance” across eight languages
- Between 1870 and 2009
- From the 2012 edition of the Google Books Ngram Corpus
- American English, British English, Simplified Chinese, French, German, Italian, Russian, and Spanish
- Why not newspapers? In what sense is this better?

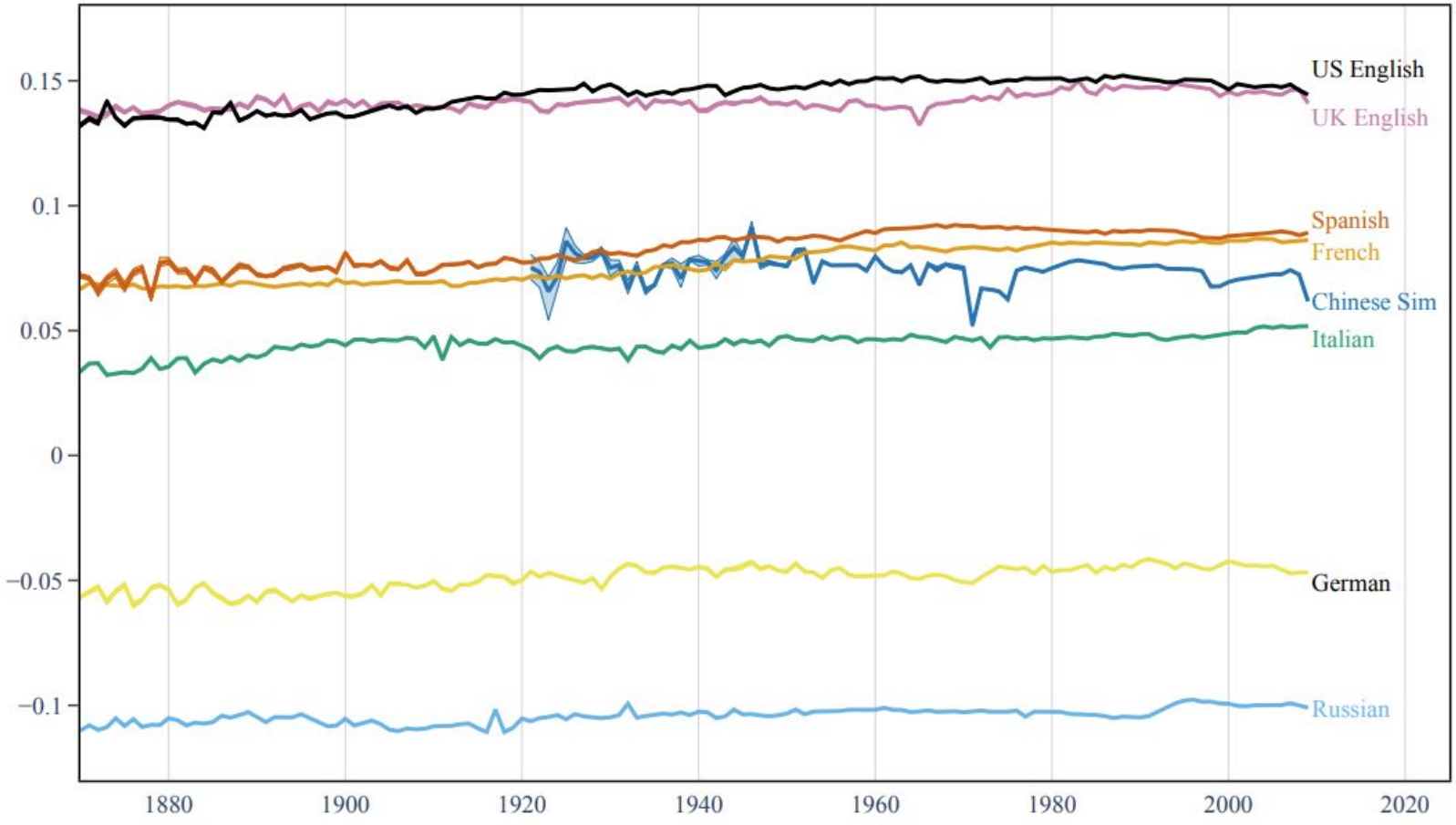
Sentiment over time

- Average of all sentences' sentiment for each year

$$f_{it} = \sum_j a_{ji} \times \frac{c_{jit}}{\sum_k c_{kit}}.$$

- Completely sensible
- Does not take into account the number of times books speak about finance though
  - Maybe speaking more about it reveals sentiment
  - Easy to control for

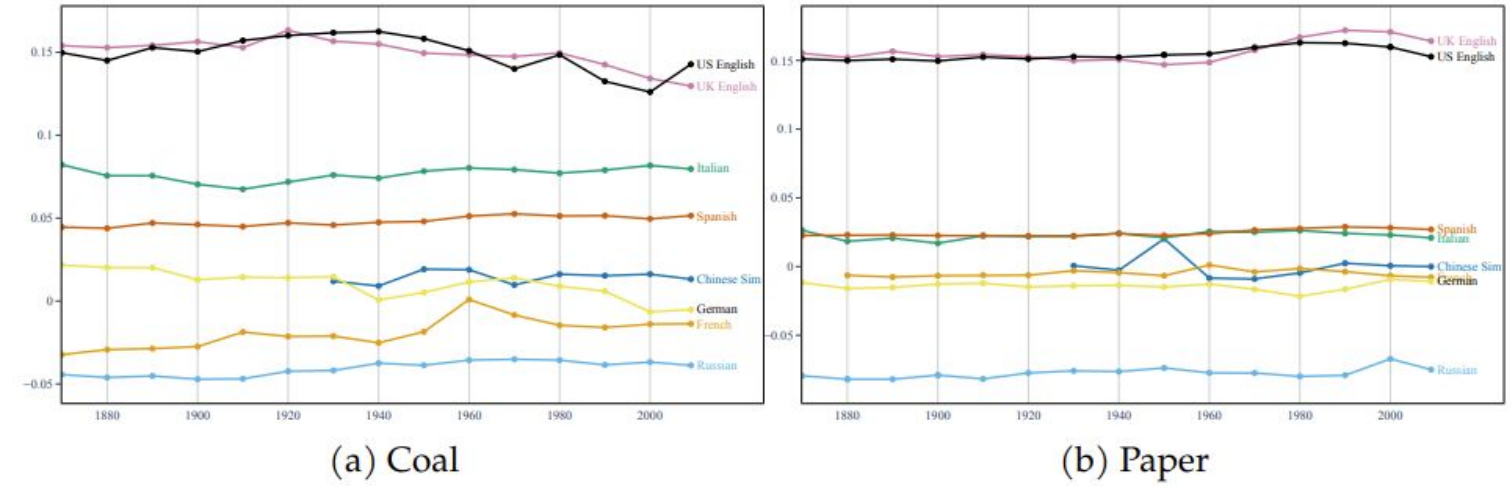
Figure 2: Sentiment toward finance



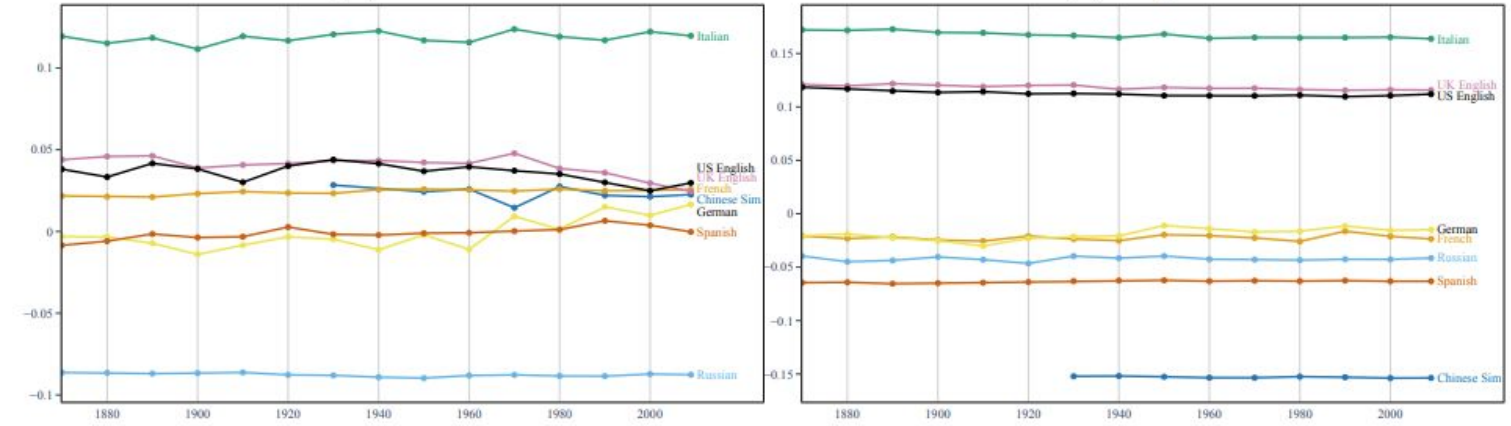
Sentiment across countries

- There seem to be persistent differences across languages
- But, the level is hard to interpret
  - Maybe standardizing is better
- ‘positive’ - ‘negative’ may not have a similar (rotated) vector in every language
- So comparing ‘finance’ vs ‘positive’ - ‘negative’ may not lead to the same level results
- It would be good to see the graph with ‘good’ - ‘bad’ or ‘approve’ - ‘disapprove’
- It will be tricky to do the comparison across countries, and the paper is really nice even without that part

Figure 4: General sentiment



(a) Coal



(b) Paper

(c) Tobacco

(d) January

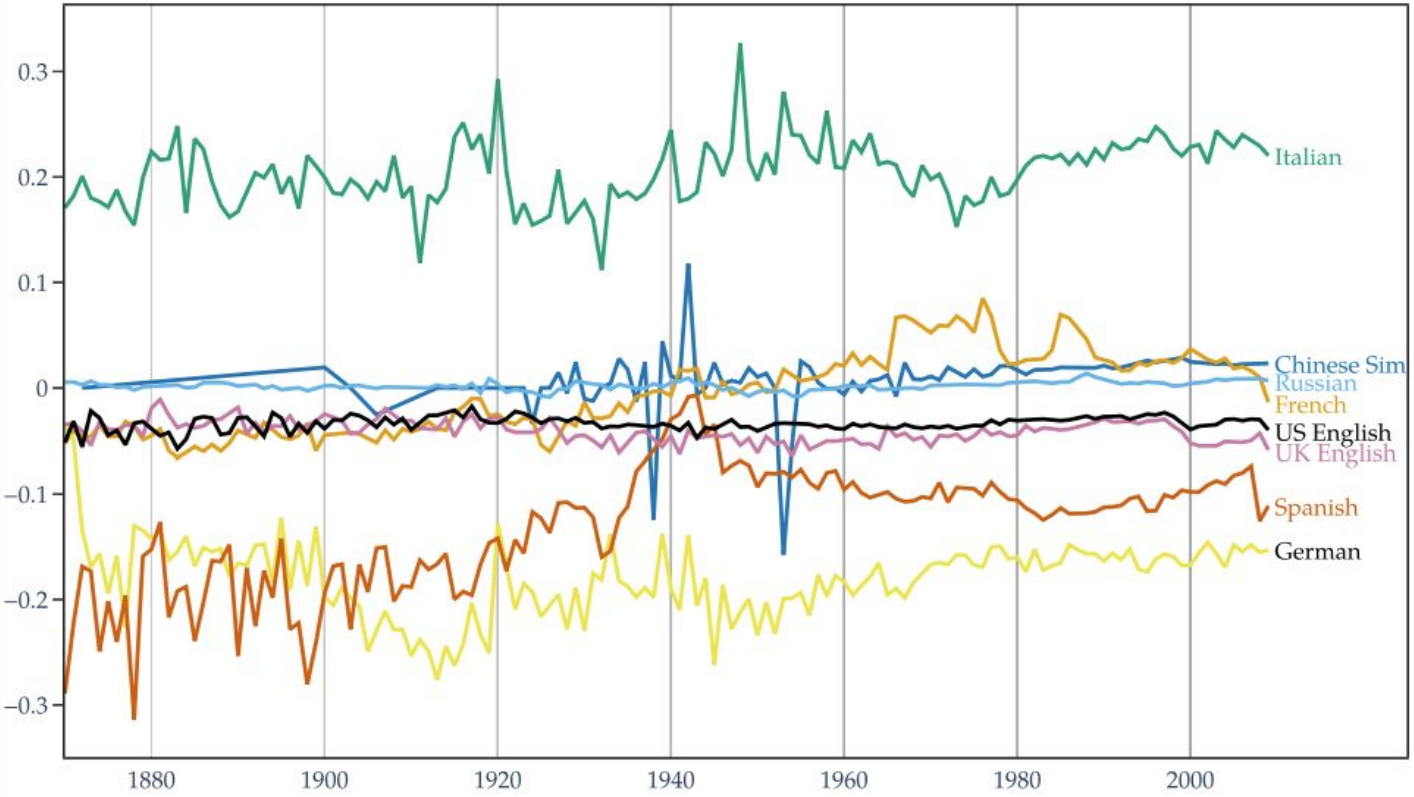
- English always positive
- Russian always negative
- Maybe levels are not comparable

Finance sentiment growth

$$\Delta f_{it} = \frac{f_{it} - f_{it-1}}{|f_{it-1}|} \times 100.$$

- But, sentiment is already stationary, and has meaningful units, why not just first difference? I'd imagine it gives the same results anyways
- At least it takes care of the level issue

Figure 5: Sentiment toward finance using an alternative dictionary-based approach



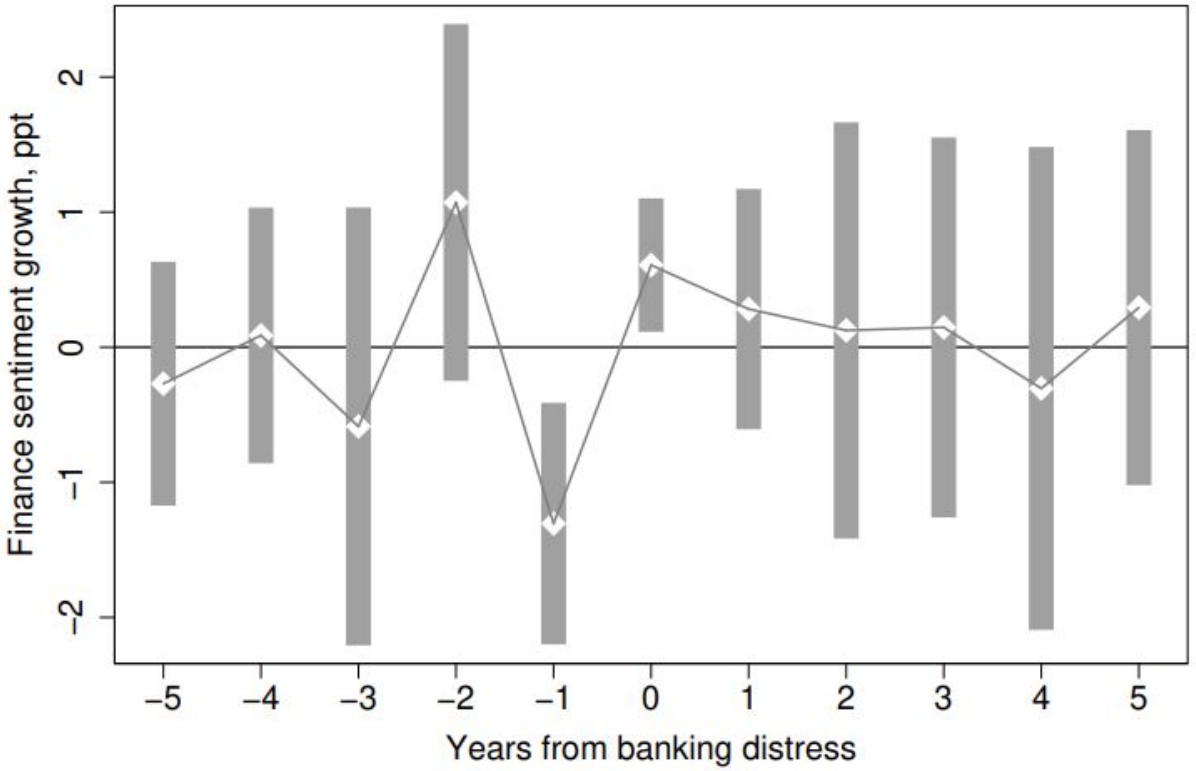
- Why is this a worse measure?
- Level issue still arises

Table 6: Alternative approaches

	Bag-of-words	Kozlowski-Taddy-Evans			Our approach
Languages	LM Dictionary	fastText	GloVe	Word2Vec	BERT
Chinese	-0.06	0.02	0.03	-0.03	0.03
	(0.10)	(0.26)	(0.14)	(0.09)	(0.06)
French	-0.01	-0.25	-0.14	-0.02	0.08
	(0.04)	(0.04)	(0.09)	(0.05)	(0.01)
German	-0.18	-0.10	-0.02	-0.02	-0.05
	(0.04)	(0.07)	(0.09)	(0.07)	(0.00)
Italian	0.20	-0.09	-0.10	0.10	0.04
	(0.03)	(0.06)	(0.15)	(0.07)	(0.00)
Russian	0.00	-0.05	0.04	0.03	-0.10
	(0.00)	(0.10)	(0.11)	(0.07)	(0.00)
Spanish	-0.13	-0.16	0.05	0.06	0.08
	(0.06)	(0.06)	(0.11)	(0.09)	(0.01)
UK English	-0.04	0.16	0.03	0.08	0.14
	(0.01)	(0.09)	(0.12)	(0.05)	(0.00)
US English	-0.03	0.17	0.01	0.14	0.14
	(0.01)	(0.06)	(0.07)	(0.08)	(0.01)
Total	-0.02	-0.03	-0.02	0.04	0.05
	(0.11)	(0.18)	(0.13)	(0.10)	(0.09)
Average	-0.03	-0.04	-0.01	0.04	0.04
	(0.04)	(0.09)	(0.11)	(0.07)	(0.01)

- Why is this a worse measure?
- What if sentiment is really volatile?
- A supervised test would give objective metrics

Figure 6: Finance sentiment around financial crises



Conclusion

- Really cool paper!
- Nice datasets and techniques
- Mostly missing an objective measure of why BERT is better
  - Holds in theory
  - But we still need the results
- Cross-language comparisons seem hard (and unnecesary) to defend
- Sentiment in the sample is not zero, but it could be under robustness checks, percentage growth seems unnecesary since series is already stationary