

Economic Narratives and Market Outcomes: A Semi-Supervised Topic Modeling Approach



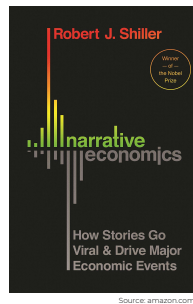
Dat Mai (datmai@mail.missouri.edu)

Kuntara Pukthuanthong (pukthuanthongk@missouri.edu)

Introduction

Inspiration

- Popular stories in daily conversation can affect individual and collective behaviors.
- Studying these stories helps predict major economic events.



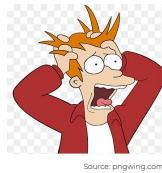
Research question

- Extract 10 narratives from 7 million New York Times (NYT) articles over 150 years
- Examine whether these stories can predict the market returns

Measure Panic?

A model-free method combining

- text data
- computing power
- machine learning.



Main findings

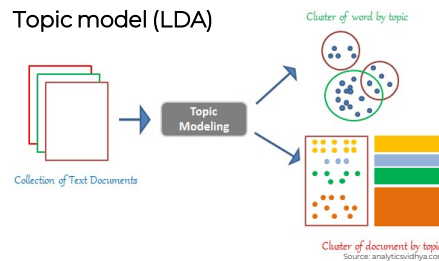
- Panic is the most important topic encompassing various anxiety-related themes.
- Panic is a strong positive predictor of market return (both in- and out-of-sample).
- Panic proxies for time-varying risk aversion consistent with the ICAPM.
- Panic has predicting power beyond the actual events.

Methodology

General idea

Extract the attention of news articles allocated to each topic (narrative) in any period

Topic model (LDA)



Seeded LDA

Provide seed words per topic to guide the clustering of words towards the predefined themes.

Seed words for Panic

bank failure, bank panic, crisis, depression, epidemic, fear, financial panic, pandemic, panic, recession, tension, war, etc.

Figure 1. Rolling estimation scheme

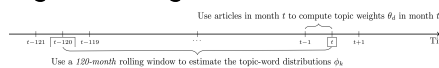
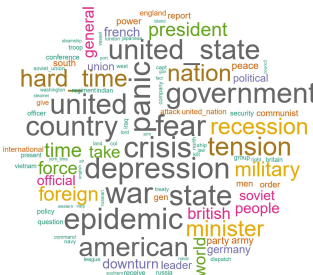


Figure 3. Panic content



Main results

Figure 4. Time series of Panic weight

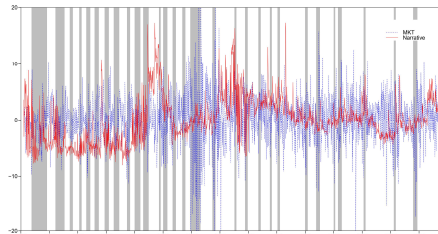


Figure 5. Important articles for Panic

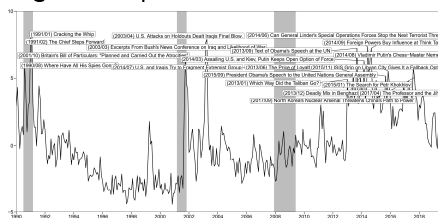


Table 4. Predicting market returns

$$R_{t+1-t+h}^e = \alpha + \beta x_t + \epsilon_{t+1-t+h}$$

	Panic (%)	t-stat	R ² (%)	PLS (%)	t-stat	R ² (%)	N
Panel A: 1871-2010							
h = 1	3.44 ***	(3.18)	0.31	5.60 ***	(4.36)	0.91	1784
h = 3	2.62 ***	(2.71)	0.47	3.47 ***	(3.08)	0.86	1784
h = 6	2.64 ***	(2.62)	1.02	3.00 ***	(2.80)	1.33	1784
h = 12	2.49 **	(2.81)	1.53	2.71 **	(2.17)	1.83	1784
h = 24	1.85 *	(1.84)	1.45	2.36 **	(2.55)	2.45	1776
h = 36	1.92 **	(1.91)	2.16	2.73 ***	(2.88)	4.43	1761
Panel B: 2000-2019							
h = 1	10.85 ***	(3.74)	4.22	11.75 ***	(3.13)	5.02	238
h = 3	8.01 ***	(4.10)	6.72	6.25 **	(2.44)	3.92	238
h = 6	4.02 ***	(3.15)	6.62	3.92 **	(1.85)	2.57	238
h = 12	5.80 **	(2.51)	11.14	2.18	(1.17)	1.21	238
h = 24	5.03 **	(2.24)	14.25	1.57	(0.98)	1.00	230
h = 36	4.89 ***	(2.72)	20.16	1.72	(1.41)	2.09	218

Table 10. Out of sample R²

	1891-2019	1891-1949	1950-1999	2000-2019
Panel A: OLS				
Dividend-price ratio (DP)	-0.39	-0.48	-0.38	0.05
Dividend yield (DY)	-0.34	-0.15	-0.93	0.04
Earnings-price ratio (EP)	-0.05	0.07	-0.23	-0.35
Dividend payout ratio (DE)	-0.65	-0.84	-0.02	-1.06
Stock variance (SVAR)	-1.67	-2.19	-0.76	-0.86
Treasury bill rate (TBL)	0.07 **	-0.04	0.18 **	0.45
Panic:	0.28 ***	0.05	0.37 **	1.41 ***
Panel B: PLS				
Economic	-0.62	-0.76	-0.30	-0.55
Narrative	0.24 **	-0.26	0.81 **	1.71 *

Additional results

- Prediction results hold at the portfolio and daily level.
- Topics from seeded LDA outperform those from frequency count of seed words.
- Results with 2 millions WSJ articles:
 - The narrative index is still a positive market predictor.
 - Stock market bubble is the most important topic and is a negative predictor.
 - Media with different target audiences have diverse impacts on the market.

Conclusion

- This paper extracts Shiller's 10 narratives from 7 million NYT and 2 million WSJ articles.
- Panic positively predicts market, proxies for time-varying risk aversion, and has power beyond actual stressful events.
- The estimation scheme can be extended by
 - daily estimation
 - applying to other languages

Acknowledgements

This paper benefits from discussions and suggestions by seminar participants at Mizzou, University of Missouri-St. Louis, Missouri State University, UT Austin PhD Symposium, EFA DT, CQA Academic Competition, NFA, FMA, SFA, and LongTail Alpha LLC. I acknowledge computing support from the RCSS at Mizzou.