

(Almost) 200 Years of News-Based Economic Sentiment

Jules H. van Binsbergen, Svetlana Bryzgalova, Mayukh Mukhopadhyay, Varun Sharma

解读：王梦涵

2025 年 1 月 15 日

Motivation

- Economic sentiment is important to understand the business cycle
 - Sentiment shocks can affect real economic activity (Shapiro et al., 2022).
- Existing economic sentiment measures are limited by short time spans or low granularity.
 - Surveys
 - Constructed from the text data of economic news
- Use a historical collection of 170 years of digitized newspapers to construct economic sentiment
 - 1 billion articles, 200 million pages from 13,000 local newspapers

Challenges

- Large size text corpus and focus on historical data
 - BERT and ChatGPT are not feasible
- Traditional dictionary methods rely on manually created dictionary and word-count techniques
 - Hard to create dictionary mutually for such a long period
 - May omit complex patterns of language by ignoring context, sequence of words
- Thus, this paper uses *Word2vec + Sentprop + Smooth Inverse – Frequency*
 - Automatically constructed dictionary
 - Measure sentiment through a word vector approach

Question

- Q1: Does the new sentiment measure effective?
 - Yes
- Q2: Can economic sentiment predict GDP growth rates?
 - Yes, for nation-level and state-level
 - Its predictability surpasses the survey of professional forecasters.
 - It operates mainly through the labor channel rather than the capital one
- Q3: How does economic sentiment interact with the monetary policy?
 - Sentiment has a large influence on the key policy rate

Contribution

- Literature on economic sentiment measure derived from newspaper articles
 - Prior: National, short time series
 - Ext: Expands existing measures in both the time series and the cross-section
- Literature on Textual analysis in the economics and finance
 - Prior: Rely on manually created sentiment dictionary and simple word counts
 - Ext: Relies on automatically constructed dictionaries and word vector approach
- Literature on expectational distortions of fundamentals and/or financials and their importance for real outcomes
 - Prior: Focus on predicting stock returns and behavioral biases
 - Ext: Predict future fundamentals and posits sentiment can be rational

Data

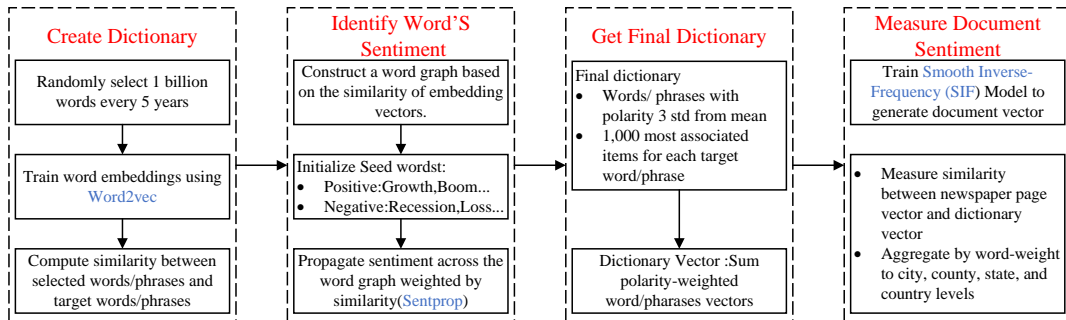
- Newspaper:1850-2020,47/50 states,city-level
 - 1 billion articles, 200 million pages from 13,000 local newspapers
- University of Michigan Consumer Sentiment survey()
- Macroeconomic data from FRED/Maddison Project Database(Quarter)
 - Real **GDP** per capita; **CPI** for all urban consumers;total non-farm payroll **employment**;gross private domestic **investment**; **industrial production**;real personal **consumption** expenditures
 - Forecasts of macroeconomic variables
- Federal funds rate (FFR)

Economic sentiment

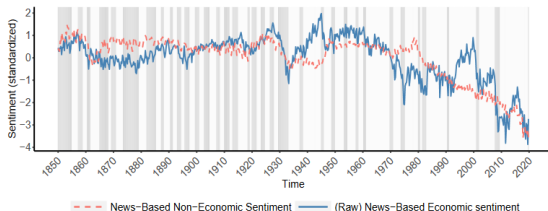
- *Word2vec*
 - Generate word embeddings by capturing semantic relationships from context
 - Window: 10 words; Vector length:300
- *Sentprop*: a label propagation algorithm that
 - Classifies all the words in the dictionary, beginning with a few initial seeds
 - Assigns continuous polarity scores to each word
- *Smooth Inverse – Frequency*(SIF): method for generating document vector
 - Computing a weighted average of word embeddings ($w_i = \frac{a}{a+p(w_i)}$)
 - Remove the first principal component to reduce noise('a','the'..)

Data:Economic sentiment

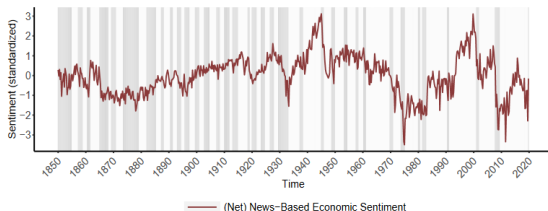
• How to measure economic sentiment?



Data:Economic sentiment



(b) Economic and Non-economic Sentiment

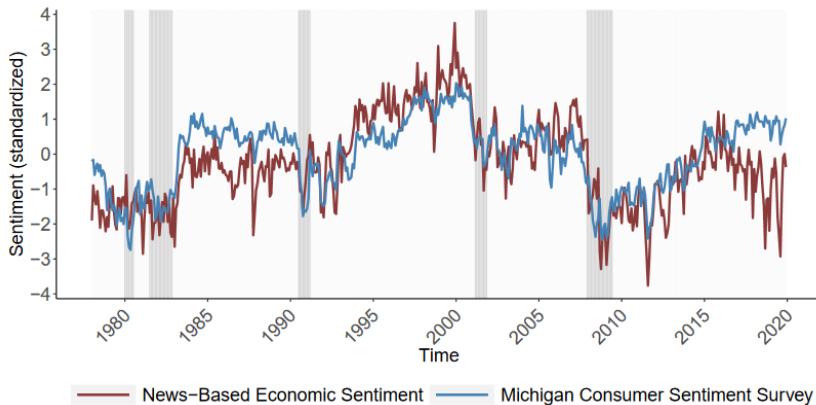


(c) Net Economic Sentiment

- Text-based sentiment has overall increasing negativity bias
- Original economic sentiment can be debiased by the non-economic sentiment

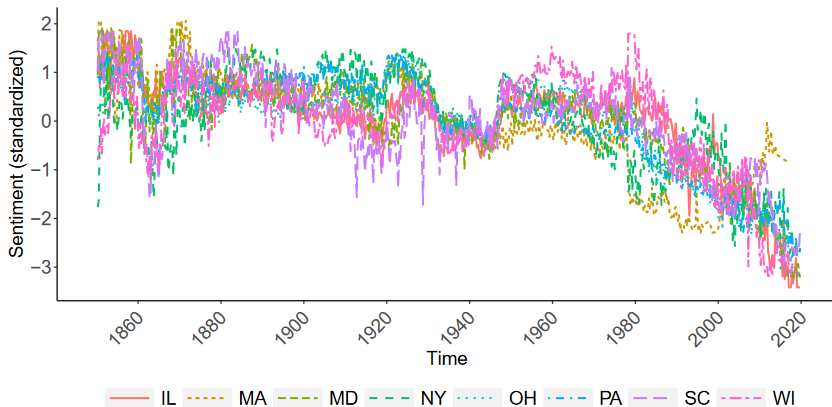
Q1: Does the new sentiment measure effective?

- News-based sentiment VS survey sentiment



Q1: Does the new sentiment measure effective?

- State-level economic sentiment



- SC: decline during the dot-com crash; NY: decline during the GFC

Q2:Economic sentiment and economy– predict GDP

$$\bullet \Delta \ln(GDP)_t = \sum_{n=1,2,\dots,6} \theta_n \Delta Sent_{t-n} + \gamma \Delta \ln(GDP)_{t-1} + \beta TermSpread_{t-1} + \epsilon_t$$

	$\Delta \ln(GDP_t)$								
	I	II	III	IV	V	VI	VII	VIII	IX
$\Delta \ln(GDP_{t-1})$	0.355*** (5.91)	0.356*** (5.79)		0.333*** (5.57)	0.335*** (5.48)			0.327*** (5.17)	0.330*** (5.15)
$Term\ Spread_{t-1}$		0.072* (1.97)			0.063* (1.69)			0.060 (1.59)	0.060 (1.58)
$\Delta Sent_{t-1}$			0.333*** (3.88)	0.255*** (2.78)	0.242*** (2.60)	0.421*** (4.05)	0.390*** (3.53)	0.288*** (2.73)	0.250*** (2.23)
$\Delta Sent_{t-2}$						0.167* (1.67)	0.176* (1.73)	-0.015 (-0.16)	-0.006 (-0.06)
$\Delta Sent_{t-3}$						0.317*** (2.60)	0.316** (2.58)	0.225** (2.04)	0.223** (2.01)
$\Delta Sent_{t-4}$						0.267** (2.23)	0.267** (2.20)	0.152 (1.37)	0.151 (1.35)
$\Delta Sent_{t-5}$						0.074 (0.74)	0.055 (0.54)	0.007 (0.08)	-0.017 (-0.18)
Q4 Dummy							-0.107 (-0.90)		-0.131 (-1.05)
$\Delta Sent_{t-1} \times Q4Dummy$							0.051 (0.25)		0.057 (0.30)
F statistic ($\Delta Sent$ (t-2 to t-n))						2.235	2.242	2.499	2.435
P-value (all $\Delta Sent$ (t-2 to t-n))						0.065	0.065	0.043	0.048
Observations	290	290	290	290	290	290	290	290	290
Adjusted R-squared	0.12	0.13	0.04	0.14	0.15	0.06	0.06	0.16	0.16

	$\Delta \ln(GSP_{s,t})$					
	I	II	III	IV	V	VI
$\Delta Sentiment_{s,t-1}$	0.263 (0.268)	0.648** (0.273)	0.723** (0.287)	-0.153 (0.320)	0.182 (0.342)	0.247 (0.357)
$\Delta Sentiment_{s,t-2}$		1.244*** (0.246)	1.362*** (0.273)		0.777*** (0.274)	0.901*** (0.302)
$\Delta Sentiment_{s,t-3}$		1.156*** (0.289)	1.321*** (0.329)		1.017*** (0.289)	1.173*** (0.330)
$\Delta Sentiment_{s,t-4}$		1.301*** (0.283)	1.455*** (0.288)		1.281*** (0.289)	1.424*** (0.293)
$\Delta Sentiment_{s,t-5}$			0.334 (0.257)			0.353 (0.266)
$\Delta Sentiment_{s,t-6}$			0.740** (0.319)			0.611* (0.321)
$\Delta \ln(GSP_{s,t-1})$				0.028 (0.022)	0.013 (0.024)	0.010 (0.025)
$\Delta \ln(GDP_{t-1})$				0.488*** (0.055)	0.421*** (0.071)	0.404*** (0.075)
$\Delta NationalSentiment_{t-1}$				0.229 (0.186)	0.276 (0.192)	0.305 (0.193)
Controls	No	No	No	Yes	Yes	Yes
Observations	1,172	1,172	1,172	1,172	1,172	1,172
R-squared	0.00	0.03	0.03	0.03	0.04	0.05

- Local and national sentiment changes predict GDP growth

Q2:Economic sentiment and economy– predict GDP

- *GDPForecast*: The Survey of Professional Forecasters(1968-)

	<i>GDP Forecast_{t,t+1}</i>						
	I	II	III	IV	V	VI	VII
<i>GDP Forecast_{t-1}</i>	0.786*** (13.15)	0.727*** (12.52)		0.770*** (11.90)	0.716*** (11.82)		0.726*** (12.43)
<i>Term Spread_{t-1}</i>		0.057*** (3.85)			0.054*** (3.68)		0.054*** (3.73)
$\Delta Sent_{t-1}$			0.154*** (3.35)	0.094*** (3.10)	0.084*** (3.05)	0.205*** (3.98)	0.082*** (2.89)
$\Delta Sent_{t-2}$						0.151*** (3.39)	-0.038 (-1.06)
$\Delta Sent_{t-3}$						0.163*** (3.30)	0.020 (0.47)
$\Delta Sent_{t-4}$						0.118** (2.31)	0.004 (0.12)
$\Delta Sent_{t-5}$						0.058 (1.28)	-0.001 (-0.04)
F statistic ($\Delta Sent$ (t-2 to t-n))						3.708	0.971
P-value ($\Delta Sent$ (t-2 to t-n))						0.006	0.425
Observations	204	204	204	204	204	204	204
Adjusted R-squared	0.62	0.65	0.05	0.63	0.67	0.11	0.67

- Economic sentiment leads the survey of professional forecasters.

Q2:Economic sentiment and economy– predict GDP

- Source of the predictability: Capital and labor

	$\Delta \ln(Investment_t)$					$\Delta \ln(IndustrialProduction_t)$				
	I	II	III	IV	V	VI	VII	VIII	IX	X
$\Delta \ln(Investment_{t-1})$	0.738*** (13.70)			0.738*** (13.79)	0.743*** (13.83)					
$Term\ Spread_{t-1}$	-0.046** (-2.08)			-0.044** (-1.99)	-0.045** (-2.08)	0.225* (1.88)			0.226* (1.89)	0.215* (1.78)
$\Delta Sent_{t-1}$		-0.073 (-0.90)	-0.073 (-0.66)	-0.038 (-0.72)	-0.019 (-0.34)	0.472 (0.99)	0.826* (1.74)	-0.039 (-0.09)	0.133 (0.33)	
$\Delta Sent_{t-2}$			0.016 (0.12)		0.108 (1.54)		0.739* (1.97)		0.213 (0.58)	
$\Delta Sent_{t-3}$			-0.043 (-0.32)		-0.024 (-0.38)		1.152*** (3.27)		0.702* (1.81)	
$\Delta Sent_{t-4}$			-0.018 (-0.14)		0.032 (0.44)		0.450 (1.03)		-0.060 (-0.15)	
$\Delta Sent_{t-5}$			-0.030 (-0.24)		-0.006 (-0.10)		0.092 (0.25)		0.061 (0.20)	
$\Delta \ln(IndustrialProduction_{t-1})$						0.409*** (3.92)			0.410*** (4.13)	0.402*** (4.04)
F statistic ($\Delta Sent$ (t-2 to t-n))			0.175		1.062		3.144		0.873	
P-value ($\Delta Sent$ (t-2 to t-n))			0.951		0.376		0.015		0.480	
Observations	290	290	290	290	290	399	399	399	399	399
Adjusted R-squared	0.61	-0.00	-0.01	0.61	0.61	0.17	0.00	0.01	0.17	0.17

	$\Delta \ln(Employment_t)$					$\Delta \ln(Consumption_t)$				
	I	II	III	IV	V	VI	VII	VIII	IX	X
$\Delta \ln(Employment_{t-1})$	0.757*** (14.73)			0.748*** (14.65)	0.748*** (16.61)					
$Term\ Spread_{t-1}$	0.042** (2.29)			0.036* (1.94)	0.036* (1.85)	0.068 (1.50)			0.064 (1.40)	0.059 (1.27)
$\Delta Sent_{t-1}$		0.249*** (3.11)	0.339*** (3.67)	0.154*** (3.06)	0.159*** (2.92)		0.135** (2.04)	0.202** (2.55)	0.106 (1.34)	0.173* (1.74)
$\Delta Sent_{t-2}$			0.315*** (2.71)		0.005 (0.06)			0.185** (2.13)		0.164* (1.76)
$\Delta Sent_{t-3}$			0.258* (1.79)		-0.014 (-0.16)			0.234** (2.21)		0.216* (1.75)
$\Delta Sent_{t-4}$			0.241* (1.76)		0.049 (0.70)			0.126 (1.11)		0.115 (1.18)
$\Delta Sent_{t-5}$			0.107 (0.97)		-0.029 (-0.52)			0.097 (1.19)		0.100 (1.14)
$\Delta \ln(Consumption_{t-1})$						0.083 (0.52)			0.071 (0.44)	0.039 (0.25)
F statistic ($\Delta Sent$ (t-2 to t-n))			2.234		0.736			1.581		1.075
P-value ($\Delta Sent$ (t-2 to t-n))			0.065		0.568			0.179		0.369
Observations	322	322	322	322	322	290	290	290	290	290
Adjusted R-squared	0.56	0.02	0.06	0.57	0.57	0.01	0.01	0.02	0.02	0.02

- It predicts employment and consumption only.

Q3: Interaction between economic sentiment and monetary policy

- Taylor Rule: $i_t = r^* + \pi_t + \alpha(\pi_t - \pi^*) + \beta(y_t - y^*)$
- Sentiment could induce the Fed to deviate from the Taylor rule

	ΔFFR_t			
	I	II	III	IV
$\Delta Sent_{t-1,t-7}$		0.063** (2.13)	0.037* (1.67)	0.032 (1.41)
$\Delta Sent_{t-1,t-7} \times Recession_{t-1}$			0.195* (1.95)	0.198** (1.98)
<i>Predicted GDP Growth</i>	0.197*** (2.68)	0.119 (1.34)	0.113 (1.29)	0.072 (0.86)
<i>Predicted GDP Growth (Two Quarters Ahead)</i>				0.073 (1.10)
Specification	R&R	R&R	R&R	R&R
Observations	346	346	346	346
Adjusted R-squared	0.24	0.25	0.27	0.27

- Sentiment remains predictive beyond its effect on GDP growth.

New ideas

- Explore the role of sector-specific sentiment in predicting economic outcomes
- Cross-regional sentiment dispersion in different country;eg: China VS US
- **New approach to use dictionary method in textual analysis**
 - Social media text
 - Other large text data...