Can We Predict Business Cycles With Natural Language Processing?

Exploring the Meaning of Boom and Bust

Albert Flak
albert.flak@student.unisg.ch

University of St. Gallen
Swiss Institute for International Economics and Applied Economic Research (SIAW)

Presentation for the Annual Congress of the Swiss Society of Economics and Statistics (SSES/SGVS)

November 11, 2021
Table of Contents

1. Research Problem and Motivation

2. Method(s)

3. Results

4. Conclusions and Perspectives
Countries in recession

Each mark represents a country

<table>
<thead>
<tr>
<th>IMF forecast*</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In October 2008 the IMF forecast that just seven countries would be in recession in 2009. The real number was 91, including most major economies.

* Made in October of previous year

Source: IMF
© FT
Research Problem and Motivation

Substantial efforts dedicated to predicting key macroeconomic variables, yet ...

- It is difficult to predict crises: Brigden (2019) has found that IMF macroeconomic projections correctly predicted only 4 out of 469 recessions in 194 countries.
- Common business cycle indicators do not use text analysis despite its potential (new insights into agents’ expectations, high-frequency of text data).

Current literature establishing connection between text corpora and macroeconomic variables


These recent attempts at using text analysis to nowcast or forecast macroeconomic variables use bag-of-words techniques and thus entirely disregard context and word ordering.
The Purpose

Emphasise the value and importance of context and meaning for textual analysis in macroeconomics.

My contribution:

• Create a real-time text-based index of *macroeconomic sentiments* exhibiting (at least some of these) useful properties
  
  i) stable relationship to macroeconomic variables,
  
  ii) predictive power towards GDP / general macroeconomic conditions,
  
  iii) predictive power towards business cycle turning points

Final product: The Relative Sentiment Index

An index tracking *macroeconomic sentiments* based on computationally inferred meaning of key business cycle terms.
Research Problem and Motivation

Context and Meaning in Computational Linguistics

The Distributional Hypothesis

*You shall know a word by the company it keeps.*

— Firth (1957, p. 179), *A Synopsis of Linguistic Theory*

**Implication 1:** A word’s meaning can be described by a probability distribution over its context words.

**Implication 2:** If a pair of words has a high *relative* co-occurrence probability, these words should have similar meanings.

In order for computer to understand meaning it needs to learn about co-occurrence probabilities of words.

Examples:

- *severe* and *crisis* vs. *severe* and *prosperity*
- *sustainable* and *recovery* vs. *sustainable* and *depression*
Method(s)

Text Corpora

**Corpus I. (Business Cycle News):**
Purpose: Insights into how agents think about business cycles

*Search Query:* Economy/Economic + "Business Cycle Words" (in heading/first paragraph, close to each other)
*Time Span:* 01/1990 – 04/2020
*Number of Articles (Unique Words):* 61,675 (278,470)

**Corpus II. (Economic Expectations News):**
Purpose: Proxy beliefs about future economic outcomes

*Search Query:* Economy/Economic + "Forward-Looking Words" (in heading/1st paragraph, close to each other)
*Newspapers:* NY Times, Washington Post, The WSJ, USA Today, Reuters
*Number of Articles (Unique Words):* 31,259 (146,639)

Search Query Words Examples

- Business cycle words (14): expansion, contraction, crisis, depression, prosperity...
- Forward-looking words (67): expectation, prediction, forecast, ...
Methodology
The Relative Sentiment Index

I. Collect News Articles and Represent in Matrix

- Word-Word Co-Occurrence Matrix

II. Infer Meaning

- GloVe: Global Vectors for Word Representation (Pennington, Socher, & Manning, 2014)
- Other algorithms out there: BERT, Word2Vec, Sent2Vec, etc.
- Dimensionality reduction and stochastic gradient descent filter out "patterns" in language

\[ F(w_i, w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}}; \quad P_{ik} = \frac{X_{ik}}{X_i} \]

III. Examine Euclidian Space

- Locate Contractionary and Expansionary Positions
- Word vectors are "meaningful": Use their properties
- Examples:
  - suprime + lending \(\approx\) mortgage
  - yellen – fed + ecb \(\approx\) trichet
Method(s)
The Relative Sentiment Index

IV. Find Similar Words: Expansionary and Contractionary Vocabulary

- Once the expansionary and contractionary vectors have been located: Find most similar word vectors
- Use proximity metric: Cosine Distance

\[
\cos(\theta) = \frac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\sum_{i=1}^{n} u_i^2} \sqrt{\sum_{i=1}^{n} v_i^2}}
\]

V. Count: Create Index

- Calculate proportion of words in articles from the respective vocabulary per time period (month)
- Create Expansionary and Contractionary Sentiment Time Series
- Difference: Expansionary - Contractionary = Relative Sentiment Index

VI. Analyse Properties

- Time Series Patterns
- Simple Linear Regressions
- Granger-Causality Tests
- Structural Breaks

A. Flak  Can We Predict Business Cycles With Natural Language Processing?  November 11, 2021
Results

Expansionary and Contractionary Words

Size by cosine distance from expansionary and contractionary positions.
Results: Relative Sentiment Index

Interesting Patterns

Figure: Business Cycle Sentiment: Expansionary and Contractionary
Results: Relative Sentiment Index

Interesting Patterns

Figure: Relative Cycle Sentiment ($R_{St}$) Index
Results and Evaluation: Relative Sentiment Index
Does the Relative Sentiment Index lead or lag macroeconomic variables?

Figure: Cross-Correlation Functions: Relative Sentiment with U.S. GDP Growth, VIX, UoM Consumer Sentiment Survey, Economic Policy Uncertainty
Results and Evaluation: Relative Sentiment Index

Do other indices lead or lag macroeconomic variables?

Figure: Cross-Correlation Functions: Other Indices with U.S. GDP Growth, VIX, UoM Consumer Sentiment Survey, Economic Policy Uncertainty
Results and Evaluation: Relative Sentiment Index

Is the Relative Sentiment Index useful in predicting macroeconomic variables?

Table 8: Granger Causality: Constructed Series → Comparative Series

<table>
<thead>
<tr>
<th>Series</th>
<th>GDP, nom, QoQ, growth</th>
<th>GDP, real, QoQ, growth</th>
<th>VIX</th>
<th>UMCSSENT</th>
<th>EPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Sentiment</td>
<td>19.8*** (8.5 * 10^-6)</td>
<td>15.4*** (8.9 * 10^-5)</td>
<td>13.3*  (0.065)</td>
<td>30.9*** (2.6e - 05)</td>
<td>32.4*** (0.0002)</td>
</tr>
<tr>
<td>Entropy</td>
<td>0.41 (0.52)</td>
<td>0.2 (0.65)</td>
<td>0.39 (0.94)</td>
<td>0.23 (0.89)</td>
<td>8.3* (0.08)</td>
</tr>
</tbody>
</table>

Table 9: Granger Causality: Comparative Series → Constructed Series

<table>
<thead>
<tr>
<th>Time Series</th>
<th>Relative Sentiment</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP, nominal, QoQ Growth (U.S.), quarterly</td>
<td>1.2 (0.27)</td>
<td>0.12 (0.73)</td>
</tr>
<tr>
<td>GDP, real, QoQ Growth (U.S.), quarterly</td>
<td>3.2* (0.075)</td>
<td>0.003 (0.95)</td>
</tr>
<tr>
<td>VIX (CBOE Volatility Index), monthly average</td>
<td>21.5*** (0.0003)</td>
<td>2.6 (0.46)</td>
</tr>
<tr>
<td>University of Michigan Consumer Sentiment (UMCSSENT)</td>
<td>8.4 (0.21)</td>
<td>0.14 (0.93)</td>
</tr>
<tr>
<td>Economic Policy Uncertainty (EPU), U.S., monthly</td>
<td>30.5*** (0.0004)</td>
<td>1.7 (0.8)</td>
</tr>
</tbody>
</table>

Figure: Granger-Causality Tests: Method from Toda and Yamamoto (1995)
Results and Evaluation: Relative Sentiment Index

Does the Relative Sentiment Index exhibit structural breaks pre-business-cycle turning points?

Figure: Structural Breaks for $RS_t$ Index. According to Bai and Perron (2003).
Conclusions and Perspectives

Conclusions

Relative Sentiment Index

- Stable Relationship to Macroeconomic Data
- Predicting GDP / General Macroeconomic Conditions
- Predicting Business Cycle Turning Points

Evidence of all three

Limitations:

- More robustness checks necessary.
- For a credibly predictive index (out-of-sample testing), I would either need time-dependent lexicons or only use an early subset of articles to estimate the word embeddings.
- Need more news articles.
Perspectives: Further Research

There are many ways my work could be extended and/or improved

- Use more/different insights from the trained vectors.
- Make the (word) vectors supervised.
- Make the vectors time-variable.
- Use other word/sentence/document-vectorisation techniques.
- More news articles.

There are many opportunities for further research with word embeddings and vectorisation techniques.
Conclusions and Perspectives

Text Embeddings: A Step Towards Economic Narratives?

We need to capture meaning to find economic narratives. A narrative can be defined as:

An account of a series of events, facts, etc., given in order and with the establishing of connections between them; a narration, a story, an account.

A story or representation used to give an explanatory or justificatory account of a society, period, etc.

— Oxford English Dictionary (2003), Definition 1 and 2

Implication 1: Need to account for order and connection between words to find (economic) narratives.

Implication 2: Need to account for context and meaning behind words to find (economic) narratives.
Thank You for Your Attention!


