Lecture 1: Dictionary Methods

Julian Ashwin

London Business School

Heidelberg, 2022

Plan for today

1. General introduction - why use text data for social science?
2. Biggest challenge - data collection
3. What are dictionary methods?
4. Application to nowcasting Euro Area GDP

Why use text?

As social scientists, why should we use text data in our research?

- We are not linguists, poets or machine learning engineers so we don't care about it for its own sake.
- Want to extract a signal from it that isn't otherwise accessible.
- We can apply a range of different tools to do this...
- Correct tool depends on the information you want to extract - often a simple tool will work best!

Some general terminology

- A *corpus* is a collection of *documents*
- A *document* is a collection of *tokens*
- A *token* is a word or phrase
- The *vocabulary* is the complete list of unique *tokens* used in a *corpus*
- The *documents* may be associated with numerical *metadata*. This could be the date or authorship of the document, but can be much broader (e.g. third lag of GDP growth in the quarter it was published)

Sources of text data

In these four sessions we will see several examples:

- Newspaper articles
- Central Bank communication
- Interviews with refugees
- Movie reviews

But also many other commonly used types

- Social media
- Patent filings
- 10K filings
- Political speeches
- Legal judgements

The first (and biggest) challenge

Collecting and cleaning text data is often the most time consuming step.

There are some proprietary platforms:

- Factiva
- Lexis Nexis

Others available for free, but with limits

- Twitter API (for now)

Others will need to be scraped

- Sometimes quite easy - e.g. central bank communication
- Sometimes not easy - irregular formatted web pages, block regular requests from same IP address...

Or ask someone who you know has it - I'll happily provide central bank communication data!

Worked example

FOMC minutes - have text documents scraped from webpage and we want to convert into a nice clean csv file.

Many options for webscrapers (e.g. selenium, rvest, scrapy, Beautiful Soup)

In most cases it will involve a bit of knowledge of html code and a lot of patience (to begin with)

Once you have the raw documents, you'll probably want to turn them into nice dataframe...

Clone, or download, the intro-text-analysis-econ repository and we'll go through an example

Dictionary methods

Very simple idea: count number of certain tokens (words or phrases) in each documents.

Advantages:

- Easy to compute
- Transparent
- Focused on a particular signal

Disadvantages:

- May not be an existing dictionary for your purpose
- Very labour intensive to create new ones
- Don't take context into account

Elephant in the room

But language models are so good know, look at GPT-3, can we really not do better than word lists?

Q: How old is Joe Biden?

A: Joe Biden is 78 years old.

Q: Why use dictionary methods for text analysis?

A: Dictionary methods are used for text analysis because they are a quick and easy way to analyze a text. They can also be used to find the most common words in a text.

What do OpenAI, Google, Microsoft et al have that we don't?

- Huge quantities of *relevant* data
- Different aims (e.g. predicting text, translation)
- Vast computational (and human) resources
- And a lot more engineering skill...

How to pick a tool

Two most important questions

1. What is the signal you're trying to extract?
2. What are the data that you have?

Start simple - if there's a dictionary for it, use it!

Worked example: open the 1_dictionary_methods.R script

Motivation

Monitoring the economy in real time is crucial for policy-making, particularly during crisis periods.

Macroeconomic fundamentals like GDP are typically measured at a quarterly frequency and released with a substantial delay.

Market participants and policy makers therefore rely heavily on survey-based indicators, but these are not available in real-time either.

This paper: a daily news sentiment can improve real-time nowcasting performance compared to realistic benchmarks

Literature

- Text data:
  - Measure latent variables: Baker et al. (2016), Thorsrud (2020)
  - Forecasting: Azqueta-Gavaldón (2017), Larsen and Thorsrud (2019), Kalamara et al. (2020), Shapiro et al. (2020), Aguilar et al. (2021) and Barbaglia et al. (2022)
- Nowcasting models:
  - Traditional econometric models: factor-based models (Bańbura et al., 2011), BVARs (Cimadomo et al., 2021), bridge equations (Baffigi et al., 2004), MIDAS models Foroni and Marcellino (2014)
  - Non-linear machine learning methods Kim and Swanson (2018)

Contributions

News sentiment adds value to nowcasts of Euro Area GDP.

1. Translation: machine translated text with English-language metrics performs best.
2. Novel approach to construct truly real-time daily indicator that still corresponds to quarterly frequency.
3. Choice of sentiment metric matters for nowcasting: general purpose dictionaries are more robust.
4. Nowcasting gains depend on the model in which metrics are used.

Text data

- Major print newspapers from Big 4 EA economies from Factiva.
- Restrict to articles tagged as economic, corporate or financial news.
- Wide range of political leanings.

Table: Total number of articles per country

	France	Germany	Italy	Spain	All
Sources	Les Échos Le Figaro Le Monde	Die Welt Süddeutsche Zeitung Der Tagesspiegel German Collection	Corriere della Sera La Repubblica Il Sole 24 Ore La Stampa	Expansión El Mundo El País La Vanguardia	
Total articles	1,255,472	833,914	1,497,909	1,407,534	4,994,829

Sentiment metrics

- Best performance - English language metrics on translated articles.
- Test a range on sentiment metrics, varying in
    1. General purpose vs econ specific
    2. word or sentence level
    3. integer or continuous scale

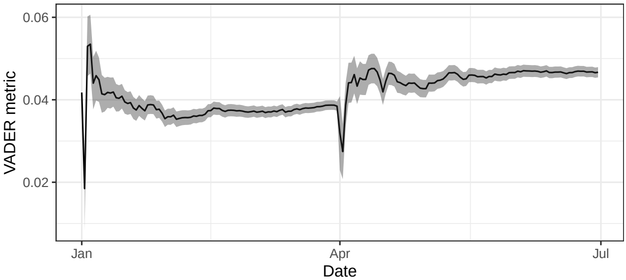
Table: English language sentiment dictionaries used

Initials	Source	Units	Application
AFINN	(Nielsen, 2011)	words (+5 to -5)	general purpose.
CGLM	(Correa et al., 2017)	words (+1 or -1)	financial stability.
HIV	(Tetlock, 2007)	words (+1 or -1)	general purpose.
HL	(Hu and Liu, 2004)	words (+1 or -1)	opinion in reviews
LM	(Loughran and McDonald, 2013)	words (+1 or -1)	economics/finance.
NKTGOS	(Nyman et al., 2018)	words (+1 or -1)	finance
VADER	(Hutto and Gilbert, 2014)	sentences (-1 to 1)	general purpose

A Daily Indicator

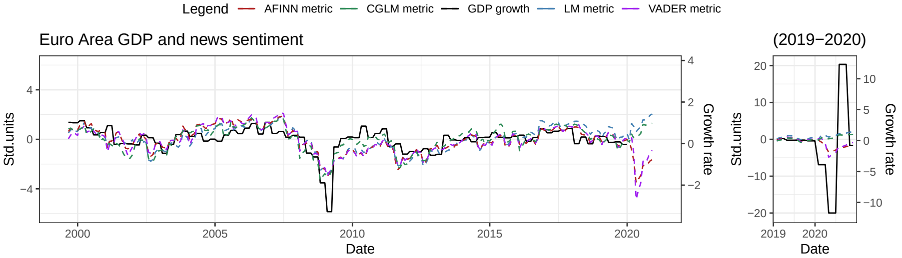
- A given day has $N_{q,d}$ words/sentences, where (q, d) denoting the dth day of quarter q. that have a sentiment score $sent_{q,d,n}$.
- Use all available articles, weighting each word/sentence equally.

$$sent_{q,d} = \frac{\sum_{t \leq d} \sum_{n=1}^{N_{q,t}} (sent_{q,t,n})}{\sum_{t \leq d} N_{q,t}}$$



EA metric

Compute each country separately and then aggregate using real-time Eurostat GDP weights.



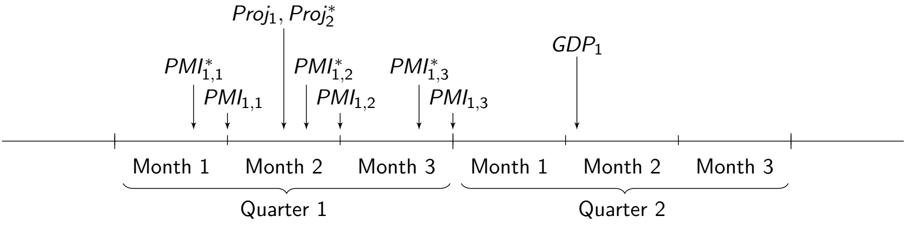
Key points:

1. Econ-focused dictionaries miss 2020 recession.
2. Relationship non-linear in crisis periods.

Macro indicators

- Target: EA real GDP growth
- Benchmark: Purchasing Managers' composite index (PMI)
- Benchmark: historical ECB macroeconomic projection

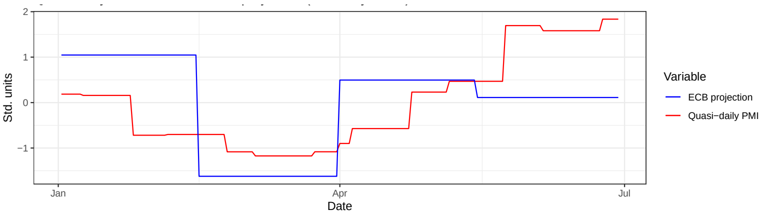
Figure: Example data release cycle



Real time information only

- Create quasi-daily PMI indicator including all available months for the quarter
- ECB projection revised in middle of quarter.

Figure: Real time daily PMI metric and ECB projection



Translation

A key challenge for multi-lingual Euro Area, so try three approaches:

1. **Translate articles**
    - Previous work suggests that in other contexts, machine translation preserves sentiment (Shalunts et al., 2016; De Vries et al., 2018)
2. **Translate dictionaries**
3. **Language-specific dictionaries**
    - Available for Germany: Bannier et al. (2019)
    - Far less extensively tried and tested
    - Questions around aggregation across countries.

Translation and GDP correlation

Translation	Metric		France	Germany	Italy	Spain	Euro area
Article	CGLM	Economic	0.615	0.527	0.542	0.71	0.698
	LM	Economic	0.593	0.461	0.412	0.619	0.636
	NKTGOS	Economic	0.38	0.373	0.364	0.618	0.538
	AFINN	General	0.583	0.371	0.41	0.731	0.637
	HIV	General	0.58	0.317	0.254	0.524	0.575
	HL	General	0.513	0.403	0.33	0.664	0.597
	VADER	General	0.583	0.369	0.292	0.701	0.611
Dict	CGLM	Economic	0.407	0.376	0.487	0.624	0.593
	LM	Economic	0.329	0.214	0.445	0.571	0.528
	NKTGOS	Economic	0.157	0.145	0.389	0.478	0.365
	AFINN	General	0.334	0.268	0.472	0.628	0.536
	HIV	General	0.288	0.24	0.187	0.38	0.446
	HL	General	0.246	0.304	0.253	0.553	0.478
Own lang				0.355			

Framework

Variety of nowcasting models with general form

$$\hat{y}_{q,d} = g\left(x_{q,d}, \theta, \eta\right)$$

In general we include the sentiment metric as well as the PMI metric in $x_{q,d}$.

Two classes of real-time benchmark:

1. Same $g()$ but only the PMI metric and not the text metric.
2. Latest available ECB projection - represent synthesis of all available data and expert judgement.

Models re-trained each day, using only data available in real time.

Assessing nowcasts

Compare the out-of-sample mean squared error (MSE) for each day of the quarter.

For example, for first day of quarter:

$$MSE_1 = \frac{1}{Q} \sum_{q=1}^{Q} (y_q - \hat{y}_{q,1})^2$$

We can thus assess nowcast performance on a daily basis throughout the quarter

Also use Diebold and Mariano (1995) tests for statistically significance of difference in performance.

Nowcasting models

Vary regression models

- OLS regression
- Ridge regression: L2 regularization on coefficients
- Random Forest: non-parametric regression based on the average across several decision trees. Max depth cross-validated.
- Neural Network: flexible non-linear function approximator. The number of nodes and layers are cross-validated.
- Also analyse average prediction across all models.

Vary potential inputs

- Levels and first differences of sentiment metrics, PMI and the ECB projections.
- In all cases, compare the models with and without text.

Linear example (pre-2020)

Exclude 2020 from example as growth rates are so extreme.

Our linear text model is

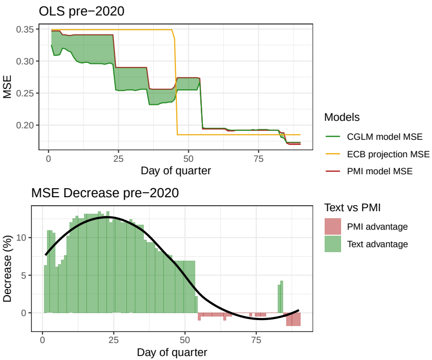
$$g_{text}(x_d, \theta, \eta) = \theta_0 + \theta_1 PMI_{q,d} + \theta_2 sent_{q,d},$$

where $sent_{q,d}$ is the sentiment metric and $PMI_{q,d}$ is the quasi-daily PMI metric. The PMI model used as a benchmark is therefore

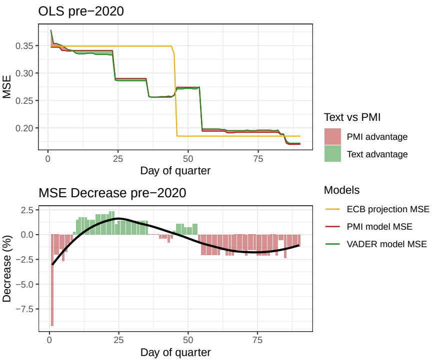
$$g_{pmi}(x_d, \theta, \eta) = \theta_0 + \theta_1 PMI_{q,d}$$

Linear example (pre-2020)

(a) CGLM metric

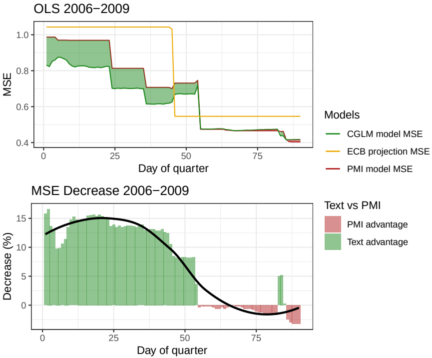


(b) VADER metric

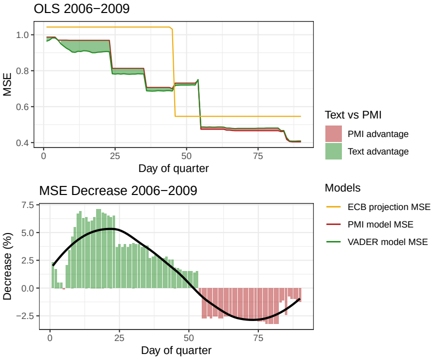


Great Recession (2006-2009)

(a) CGLM metric



(b) VADER metric

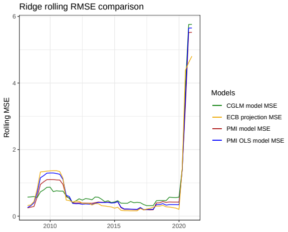


Notes: The results shown here are for April 2006 to December 2009

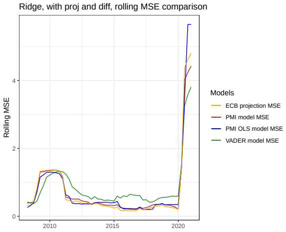
Tale of Two Crises

Figure: MSE using an 8-quarter rolling window

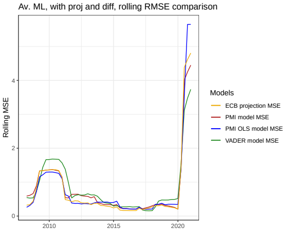
(a) CGLM Ridge



(b) VADER Ridge



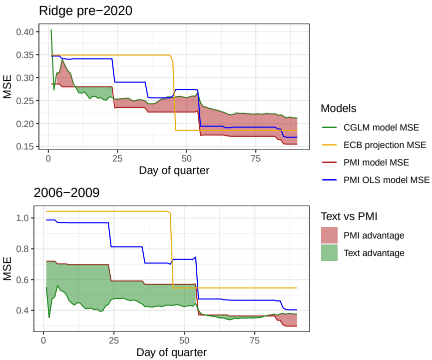
(c) VADER AvML metric



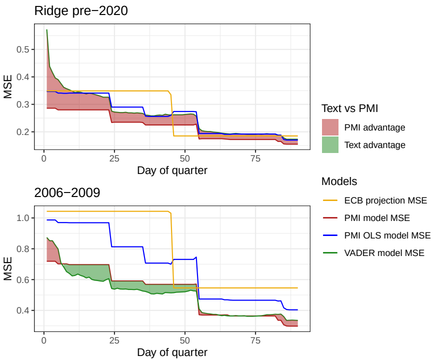
Notes: Green lines denote the text-based model (CGLM or VADER) while red lines are for the PMI-only model. Yellow lines denote the ECB projections errors and blue lines are the linear PMI-only benchmark.

Ridge regression results

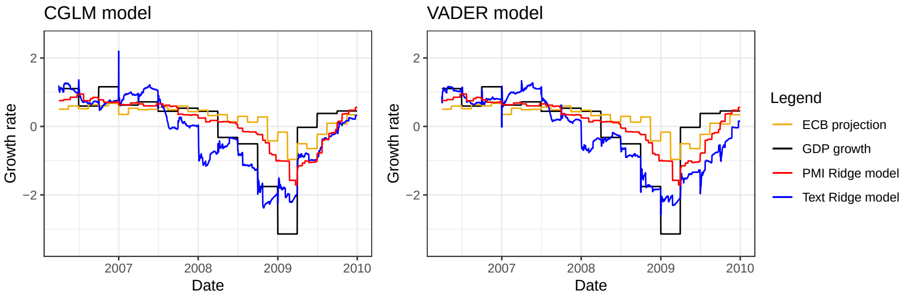
(a) CGLM metric



(b) VADER metric



Nowcasts in Great Recession

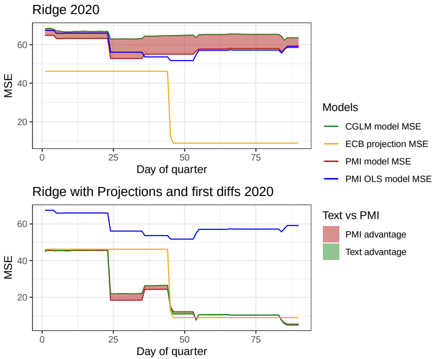


Notes: This Figure compares the real-time nowcasts of text-based PMI ridge regression models from April 2006 to December 2009.

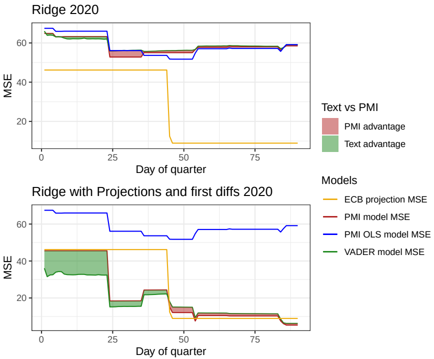
Results 2020

Figure: Ridge with and without projections: 2020

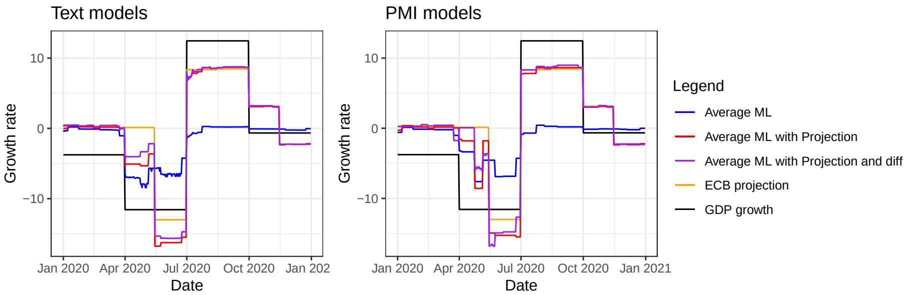
(a) CGLM metric



(b) VADER metric



Nowcasts in 2020



Notes: This Figure compares the real-time nowcasts of various text models and PMI models throughout 2020.

Conclusions and Next Steps

Conclusions:

- Machine translated newspaper text can provide information about current economic outlook for the EA.
- Improve over competitive benchmarks, especially during crisis periods.
- Choice of sentiment metric matters - general purpose more robust

Dictionaries and context

Words change based on their context (more on this tomorrow).

Vader takes steps towards this, but context matters differently in different contexts!

For example,

- Sentiment metrics generally classify words like “increasing” and “rising” as positive.
- Is “rising inflation” in a central bankers’ speech the same as “increasing unemployment”?
- One suggests rate hikes, the other rate cuts...

Modifiers and Keywords

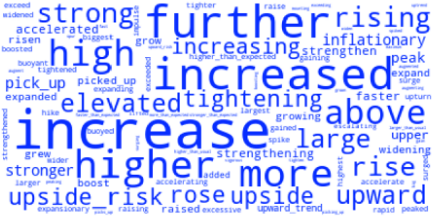
Based on Apel and Grimaldi (2012); Gonzalez et al. (2021), we can create a hawkishness index.

Identify modifiers and assign them to a keyword.

- Modifiers come in four categories:
  1. high (e.g., increase, upward)
  2. low (e.g., decrease, downward)
  3. positive (e.g., improve, good)
  4. negative (e.g., worsen, bad)
- Keywords have dual definition
  1. One of six topics: policy, prices, growth, finance, global and other
  2. Either hawkish or dovish: is "high" hawkish or dovish?

Modifiers

High modifiers



Negative modifiers

Low modifiers

Positive modifiers

Keywords

Hawkish Growth

Dovish Growth

Examples

"For their part, some indicators of **domestic demand** continue to show a **favorable** performance, although in some cases a slight **slowdown** has been observed."

- "favorable" + "domestic demand": positive and hawkish growth → +1
- "slowdown" + "domestic demand": low and hawkish growth → -1

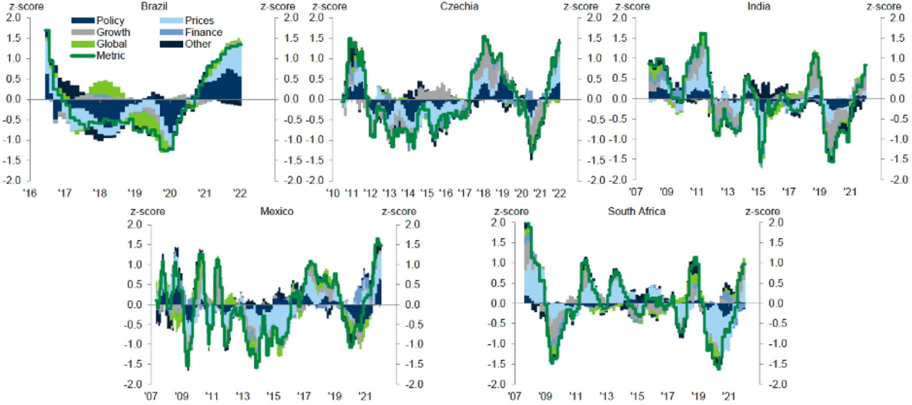
"In April, however, supply disruptions took a toll and **reversed the softening** of food **inflation**, which **surged** to 8.6 per cent from 7.8 per cent in March."

- "reversed the" + "softening" + "inflation": negator with low and hawkish price → +1
- "surged" + "inflation": high and hawkish price → +1

"On the one hand, (i) possible propagation through inertial mechanisms of **low inflation** levels in the past and the **high** level of **economic slack** may lead to a **lower-than-expected** prospective **inflation** trajectory."

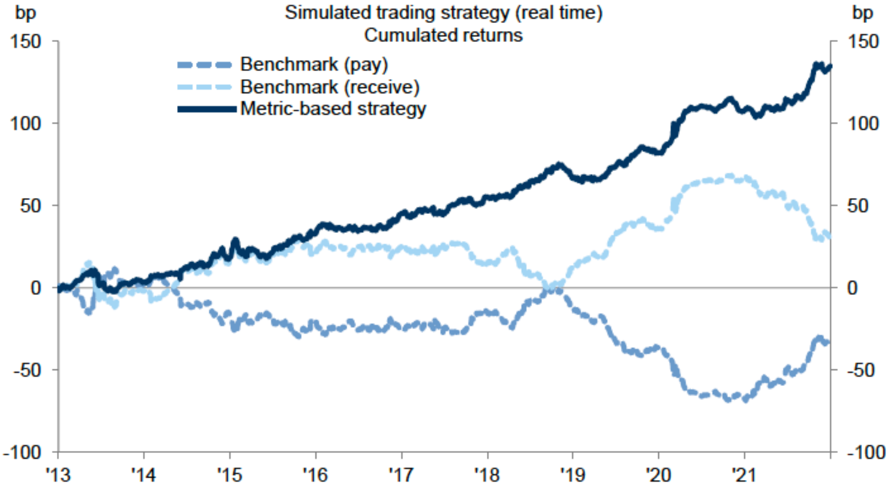
- "low" + "inflation": low and hawkish price → -1
- "high" + "economic slack": high and dovish growth → of -1
- "lower than expected" + "inflation": low and hawkish price → -1

Metrics



Brazil
Czechia
India
Mexico
South Africa
z-score
Policy
Growth
Global
Metric
Prices
Finance
Other

Trading strategy



Summary

- Dictionary methods are transparent and easy to implement
- Can be extended/adapted to different contexts
- Extract signals that are of economic relevance
- But will not be suited to all questions - so tomorrow we will move on to unsupervised methods...

References I

Aguilar, P., Ghirelli, C., Pacce, M., and Urtasun, A. (2021). Can news help measure economic sentiment? An application in COVID-19 times. *Economics Letters*, 199:109730.

Apel, M. and Grimaldi, M. (2012). The information content of central bank minutes. *Riksbank Research Paper Series No. 92*.

Azqueta-Gavaldón, A. (2017). Developing news-based economic policy uncertainty index with unsupervised machine learning. *Economics Letters*, 158:47–50.

Baffigi, A., Golinelli, R., and Parigi, G. (2004). Bridge models to forecast the euro area GDP. *International Journal of forecasting*, 20(3):447–460.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Bańbura, M., Giannone, D., and Reichlin, L. (2011). Nowcasting with daily data. *2012 Meeting Papers 555, Society for Economic Dynamics*.

Bannier, C., Pauls, T., and Walter, A. (2019). Content analysis of business communication: introducing a German dictionary. *Journal of Business Economics*, 89(1):79–123.

Barbaglia, L., Consoli, S., and Manzan, S. (2022). Forecasting with economic news. *Journal of Business & Economic Statistics*, pages 1–12.

Cimadomo, J., Giannone, D., Lenza, M., Monti, F., and Sokol, A. (2021). Nowcasting with large Bayesian vector autoregressions. *CEPR Discussion Paper No. DP15854*.

Correa, R., Garud, K., Londono, J. M., Mislang, N., et al. (2017). Constructing a dictionary for financial stability. *IFDP notes. Board of Governors of the Federal Reserve System, Washington, DC*.

De Vries, E., Schoonvelde, M., and Schumacher, G. (2018). No longer lost in translation: Evidence that google translate works for comparative bag-of-words text applications. *Political Analysis*, 26(4):417–430.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144.

References II

Foroni, C. and Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates. *International Journal of Forecasting*, 30(3):554–568.

Gonzalez, M., Tadle, R. C., et al. (2021). Monetary policy press releases: An international comparison. Technical report, Central Bank of Chile.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM.

Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8.

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., and Kapadia, S. (2020). Making text count: economic forecasting using newspaper text. *Bank of England Staff Working Paper*, 865.

Kim, H. H. and Swanson, N. R. (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*, 34(2):339–354.

Larsen, V. H. and Thorsrud, L. A. (2019). The value of news for economic developments. *Journal of Econometrics*, 210(1):203–218.

Loughran, T. and McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, 109(2):307–326.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Nyman, R., Kapadia, S., Tuckett, D., Gregory, D., Ormerod, P., and Smith, R. (2018). News and narratives in financial systems: exploiting big data for systemic risk assessment. *Bank of England Staff Working Papers*, 704.

Shalunts, G., Backfried, G., and Commeignes, N. (2016). The impact of machine translation on sentiment analysis. *Data Analytics*, 63:51–56.

References III

Shapiro, A. H., Sudhof, M., and Wilson, D. J. (2020). Measuring news sentiment. *Journal of Econometrics*.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business and Economic Statistics*, 38(2):1–17.

Newspapers

Country	Source	Total articles	Daily Circulation	Political leaning
France	Les Échos[1]	691,287	120,546	economic liberal
	Le Figaro[1]	341,925	313,541	centre-right
	Le Monde[1]	222,260	302,624	centre-left
Germany	Süddeutsche Zeitung[2]	514,284	361,507	centre-left
	Die Welt[2]	180,694	165,686	centre-right
	Der Tagesspiegel[2]	71,095	113,716	liberal
	German Collection[3]	67,841	-	-
Italy	Corriere della Sera[4]	412,944	258,991	liberal
	La Repubblica[4]	263,339	176,010	progressive
	Il Sole 24 Ore[4]	605,480	145,685	liberal
	La Stampa[4]	216,146	115,870	social liberal
Spain	Expansión[5]	634,659	50,180	liberal conservative
	El Mundo[5]	174,651	248,463	liberal conservative
	El País[5]	354,613	359,809	centre-left
	La Vanguardia[5]	243,611	180,939	liberal

1: Circulation from `https://en.wikipedia.org/wiki/List_of_newspapers_in_France`, .

2: Circulation from `https://en.wikipedia.org/wiki/List_of_newspapers_in_Germany`, .

3: Collection including *Boersen-Zeitung*, *Handelsblatt*, *Süddeutsche Zeitung* and *Frankfurter Allgemeine Zeitung*.

4: Circulation from `https://en.wikipedia.org/wiki/List_of_newspapers_in_Italy`, .

5: Circulation from `https://en.wikipedia.org/wiki/List_of_newspapers_in_Spain`, .

VADER

Valence Aware Dictionary and sEntiment Reasoner (VADER) Hutto and Gilbert (2014).

- Valence-based, not polarity-based: takes intensity into account.
- Based on a lexicon of over 7,500 text features, including commonly used abbreviations and emojis.
- These features are then rated by workers on Amazon Mechanical Turk on a scale from -4 to +4.
- Applies general heuristics to account for context of word:
  - Modifiers such as "extremely" increase the intensity of sentiment.
  - Negation switches polarity