

Forecasting Macroeconomic Data, Machine Learning and News

Aleksandr Shneider

October 19, 2026

Introduction

- Machine Learning (ML) methods are becoming increasingly popular among economists (Athey and Imbens, 2019). Economists are using texts as data (Gentzkow et al, 2019) and apply text algorithms in economic questions (Ash, Hansen, 2023)
- News about the economy contains a lot of information (economic context): uncertainty, sentiment, etc. Using ML techniques, this information can be extracted
- Economic context can be used for macroeconomic forecasting (Baker et al, 2016)
- Yet systematically integrating text-based economic context measures into real-time macroeconomic forecasting practice remains an open, active question

Motivation

- The volume of news is large: 300 to 500 thousand articles a year in seven US newspapers. It has to be processed automatically.
 - ▶ News volume
- News can both contain noise and relevant information about the current state of the economy
 - Distinguished empirically: BERT-based filtering/sentiment scores articles, and out-of-sample forecast performance decides what counts as relevant
- Forecast horse race: ML-extracted text information used *in addition* to forecasting models (both econ and ML)
- Previous literature does not focus on the comparative power of predictions over time or on filtering economic articles

Contribution

- A broad newspaper corpus instead of a single news wire: seven US newspapers, 11.76 million articles, 1990–2020 (Ellingsen et al, 2022, JoAE)
- Several text measures side by side instead of one: topics, policy uncertainty, dictionary and BERT-based sentiment (Baker et al, 2016, QJE)
- Automatic selection of the economic articles with a BERT-family model instead of human-based selection (Shapiro et al, 2015, JoE)
- Evaluation by period, with tests that allow for multi-step forecast errors: *when* news helps, not only whether

Two baseline papers

Preview of results

- **Does news beat a simple autoregression?** Only in the Great Recession (2008–2009): by 12–14% for GDP, 22–35% for consumption, 8–23% for investment. Without the four worst quarters it is 3–6% worse for GDP, level for consumption and 22–32% worse for investment. Hard data do no better.
- **Does news add to the hard data?** A little, for consumption: about 3% lower RMSE, significant in 10 of 48 comparisons and never worse; half of these hold without the recession quarters. Not for GDP outside the recession. Investment gets worse.
- **Does it matter which articles are used?** Yes. On economic articles the same model does significantly better in 27 of 144 comparisons and worse in 4, in normal times as well.
- **Why the full sample misleads.** Four quarters, 2008Q3–2009Q2, make up 41–51% of the benchmark's squared error for GDP, 31–37% for consumption and 51–60% for investment.

Data

- Macroeconomic data: FRED-MD monthly series in real-time vintages; outcomes from ALFRED
- Outcomes: quarterly growth of real GDP, consumption and investment; forecasts are evaluated through 2020Q1, so the COVID collapse of 2020Q2 is outside the sample
- Textual Data: seven US print newspapers, NYT and WSJ from 1990, seven papers from 1997 (sample through 2020). Focus only on the sources that cover the US economy in the English language. Extendable to other languages [▶ Coverage](#)
- News filter: economic articles selected with a BERT-family sentence-embedding model — 24.7% of 11.76M articles
- Do not consider FED statements (Handlan, 2022); Twitter/X posts and comments are also not considered (Gorodnichenko, Pham, Talavera, 2025, JoE)

The forecasting exercise

- **What is forecast.** Quarterly growth of real GDP, consumption and investment, y_{q+h} : the current quarter ($h = 0$, the nowcast) and one to three quarters ahead ($h = 1, 2, 3$).
- **When.** In each month $m = 1, 2, 3$ of quarter q , with the data available in that month.
- **With what.** Three sets of monthly predictors:
 - *hard data* (H): 101 FRED-MD series, real-time vintages;
 - *news* (N): 80 topic shares and their dispersion across newspapers (160 series);
 - *hard data + news* ($H+N$): both together.
- **How.** One model per outcome, horizon, month and set, re-estimated every quarter on an expanding window from 1997Q1; forecasts are scored out of sample over 2003Q1–2020Q1 (69 quarters).
- **Two comparisons.** (1) News against hard data. (2) Hard data against hard data + news.

The main specification adds 46 further text series to N (policy uncertainty, sentiment): “all text measures”.

The forecasting regression

For each outcome, horizon h , month m and predictor set $S \in \{H, N, H+N\}$ (Ellingsen et al. 2022):

$$y_{q+h} = \alpha + \sum_{j \in S} \sum_{k=0}^5 \beta_{j,k} x_{j,t_q-k}^{(m)} + \varepsilon_{q+h}$$

- $x_{j,t}$: monthly series j ; t_q : last month of quarter q . **Six monthly lags** of each series enter as separate regressors (unrestricted MIDAS): the current and the previous quarter.
- The superscript (m) : the forecast is made in month m of the quarter, so the later months of that quarter are missing. Each missing month is replaced by the average of the series up to then. News for month m is already available; hard data arrive one month late.
- Regressors. News: 80 topics, each as its average share across the papers and as its dispersion between them = 160 series, $\times 6$ lags = 960. Hard data: the 101 FRED-MD series available in every vintage, $\times 6 = 606$. Both: 1,566. No lag of y .
- Estimated with four methods that all use these regressors (next slides); trained on first-release data.

Lags and missing months

Two comparisons, one statistic

Forecast error of predictor set S over the T evaluation quarters:

$$\text{RMSE}_{h,m}^S = \sqrt{\frac{1}{T} \sum_q \left(y_{q+h} - \hat{y}_{q+h|q,m}^S \right)^2}$$

- ① **News against hard data:** $\text{RMSE}^N / \text{RMSE}^H$
Can text stand in for the hard data?
- ② **Hard data against hard data + news:** $\text{RMSE}^{H+N} / \text{RMSE}^H$
Does text add anything to the hard data, holding the model fixed?

A ratio below one means "yes". Comparison benchmarks (denominator): an autoregression (AR), constant growth, and the Survey of Professional Forecasters (SPF).

Why RMSE ratios

Four forecasting methods, three benchmarks

The same regression is estimated with four methods. Each gives its own forecast.

- **Elastic Net**: a linear regression with a penalty that shrinks the coefficients and sets most of them to zero.
- **PCA-OLS**: the predictors are first summarised by five common factors; the outcome is regressed on the factors.
- **Random forest**: the average of 300 regression trees, each grown on a random part of the data.
- **XGBoost**: regression trees added one after another, each fitted to the errors the previous ones left.

Tables report the most accurate of the four for each outcome, horizon and month (chosen after the fact, as in Ellingsen et al.), or each method separately.

Three benchmarks.

- **AR**: the outcome regressed on its own last 1–4 quarters (lag length by BIC). It captures persistence, not turning points.
- **Constant growth**: the historical average growth rate.
- **SPF**: the mean forecast of the Survey of Professional Forecasters, as published at the time.

How to read the tables

- **One exercise.** Every table in the main deck is a forecast of quarterly growth from monthly predictors: the U-MIDAS regression of the previous slides.
- **H0, H1, H2, H3:** the quarter being forecast — the current quarter (the nowcast) and one, two and three quarters ahead.
- **M1, M2, M3:** the month of the quarter in which the forecast is made. Main-deck tables average over the three months; the appendix shows them separately.
- **Each number is a ratio of forecast errors (RMSE).** Below 1: the first-named forecast is more accurate. Example: news only against AR, GDP, H1 = 0.95 means that one quarter ahead the best news-only model's RMSE is 5% below the autoregression's.
- **Stars:** the difference in accuracy is statistically significant at 10 / 5 / 1% (Diebold–Mariano test, adjusted for multi-step forecasts).

How the best model is chosen

Tables on the 564k-article sample (appendix) use an earlier version of the test; their frames say so.

Baseline, all articles: the two comparisons

Outcome	Text only vs. hard only				Hard + text vs. hard only			
	H0	H1	H2	H3	H0	H1	H2	H3
GDP	1.07	1.00	0.98	0.99	1.02	1.00	0.98	0.98
Consumption	0.99	1.02	0.97	0.93	0.99	1.01	0.98	0.94
Investment	1.29	1.17	1.13	1.08	1.02	1.06	1.05	1.02

- **How to read:** RMSE of the best text-only method (left block) or of the best hard + text method (right block), divided by the RMSE of the best hard-data method. Each number is the average over the three months in which the forecast is made. Below 1 = better than hard data alone.
- Text alone is behind the hard data for the GDP nowcast and for investment: significantly worse in 8 of 36 cells, better in 3.
- Adding news to the hard data changes little: the best hard + news model is significantly better in 3 of 36 cells and worse in 2.

Every method, every month

Against the benchmarks

Seven papers, all articles, 80 topics; the methods and text measures of the main specification. Cells: 3 outcomes $\times$ 4 horizons $\times$ 3 months = 36.

Baseline, all articles: adding news, model held fixed

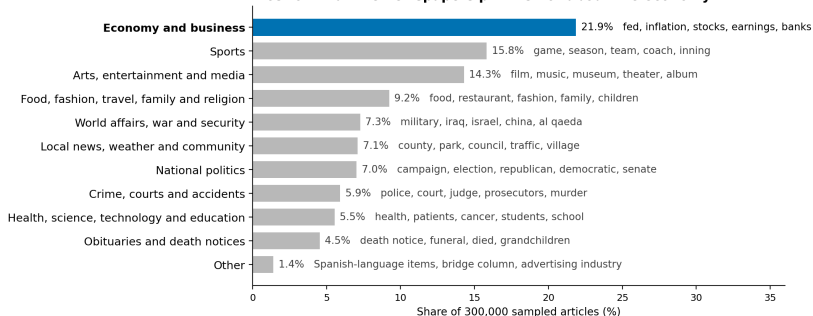
Model	H0	H1	H2	H3	Sig. better	Sig. worse
Elastic Net	1.03	1.02	1.04	0.96	0 of 36	4 of 36
PCA-OLS	1.05	1.02	0.99	0.96	4 of 36	1 of 36
Random forest	1.00	1.02	0.99	0.99	5 of 36	2 of 36
XGBoost	0.99	1.00	0.93	0.95	5 of 36	0 of 36
All models	1.02	1.01	0.99	0.97	14 of 144	7 of 144
GDP	1.03	1.00	0.98	0.96	1 of 48	4 of 48
Consumption	0.98	0.99	0.99	0.96	12 of 48	0 of 48
Investment	1.05	1.05	1.00	0.97	1 of 48	3 of 48

- **How to read:** each row compares the same method with and without news; below 1 = news helps. "Sig.": significant cases among the row's comparisons — 36 per method (3 outcomes $\times$ 4 horizons $\times$ 3 months), 48 per outcome, 144 in all.
- Adding news leaves the average ratio near one: 14 significant improvements and 7 deteriorations in 144 comparisons.
- Consumption gains (12 improvements, none worse); GDP and investment lose (2 improvements, 7 deteriorations).

All 144 comparisons

Most of what a newspaper prints is not about the economy

Most of what the newspapers print is not about the economy



- An embedding classifier keeps 24.7% of the 11.76 million articles (2.9 million) as economic. Three quarters are about something else.
- By cluster label, economy and business are 21.9% of 300,000 sampled articles; sports are 15.8%, arts and entertainment 14.3%.

250 clusters of sentence embeddings on 300,000 sampled articles; themes assigned by hand. Shares by cluster label.

Economic share over time

Themes and method

Text pipeline

- Text methods, first stage — select the economic articles with a BERT-family sentence-embedding model; implemented on the full corpus
- An earlier keyword rule (6–10% of articles) is kept as a comparison
- Second stage — turn the selected articles into monthly series: topic shares, policy-uncertainty counts, dictionary and FinBERT sentiment
- Third stage — forecast macro data with these series

The economic-article filter

- Each article (title and lead) is embedded with a BERT-family sentence encoder (all-MiniLM-L6-v2).
- 250 clusters are fitted on 300,000 sampled articles; 57 are labelled economic automatically, from their most characteristic words or their closeness to economic anchor sentences.
- An article is economic if its nearest cluster is economic, or if it is closer to economic than to non-economic anchor sentences by a margin of 0.10.
- Kept: 24.7% of articles (2.9 million of 11.76 million).
- The 80-topic model is re-fitted on the kept articles and all monthly series are rebuilt from them.

Keywords against the filter

How the filter is built

Cluster map

BERT explainer

Main specification

- **News:** seven print papers; economic articles only (24.7% of the corpus); 80 topics re-fitted on them; all text measures.
- **Hard data:** FRED-MD, the standard monthly macro database (McCracken and Ng). 101 series: those present in every real-time vintage since 1999. Outliers screened.
- **Sample:** training starts in 1997Q1, when five of the seven papers enter the database. Forecasts are evaluated from 2003Q1, after 24 quarters of training data, to 2020Q1: 69 quarters.
- **Methods:** Elastic Net, PCA-OLS, random forest, XGBoost; the best of the four is reported.
- The specification of Ellingsen et al. (LASSO, PCA-OLS, random forest; topics only) is in the appendix.

The filter works: main specification against the baseline

Outcome	News only						Hard data + news					
	H0	H1	H2	H3	Sig. < 1	Sig. > 1	H0	H1	H2	H3	Sig. < 1	Sig. > 1
GDP	0.93	0.97	0.99	0.99	8 of 48	2 of 48	0.98	0.97	0.97	0.99	9 of 48	0 of 48
Consumption	1.00	1.00	0.96	1.03	4 of 48	2 of 48	1.00	0.99	0.95	1.02	5 of 48	4 of 48
Investment	0.87	0.86	0.90	1.01	15 of 48	0 of 48	1.00	0.93	0.98	1.04	10 of 48	1 of 48
All outcomes	0.94	0.95	0.95	1.01	27 of 144	4 of 144	0.99	0.96	0.97	1.02	24 of 144	5 of 144

- **What is compared:** the main specification (economic articles) with the baseline (all articles). Same method, same hard data, same window; only the articles change.
- News only: 27 significant improvements and 4 deteriorations in 144 comparisons; about 7 of each would be expected by chance. Investment RMSE falls by 10–14% at H0–H2, the GDP nowcast by 6–7%; consumption is unaffected.
- Not a recession effect: without the four worst quarters, 30 improvements and 7 deteriorations.

RMSE on economic articles divided by RMSE on all articles, same model; mean over the four methods and M1–M3. Sig.: Diebold–Mariano test at 10% (Newey–West variance, lag h , small-sample correction). With the specification of Ellingsen et al.: 11 improvements, 1 deterioration in 81.

What the filter changes against the benchmark

Outcome	Full sample		2008–2009	
	All articles	Economic articles	All articles	Economic articles
GDP	1.01 / 0.99 / 0.98	0.94 / 0.95 / 0.95	1.00 / 0.94 / 0.93	0.88 / 0.88 / 0.86
Consumption	0.88 / 0.95 / 0.90	0.87 / 0.93 / 0.89	0.71 / 0.88 / 0.86	0.65 / 0.75 / 0.78
Investment	1.23 / 1.15 / 1.14	1.04 / 1.02 / 1.04	1.07 / 0.92 / 0.98	0.77 / 0.80 / 0.92

- On all articles the best news-only model is significantly worse than AR in 11 of 36 cells; on economic articles in none. The filter brings news from behind the benchmark to level with it.
- Against the hard data: on all articles text is significantly worse in 8 of 36 cells and better in 3; on economic articles no cell differs.
- Outside the four recession quarters the filter changes little for GDP: 1.05 / 1.05 / 1.06 on all articles, 1.03 / 1.03 / 1.06 on economic articles.

Other noise reductions All specifications

Best news-only model against the AR benchmark, mean over M1–M3, H0 / H1 / H2. Main specification; best model chosen ex post on the full sample.

Economic articles: the two comparisons

Outcome	Text only vs. hard only				Hard + text vs. hard only			
	H0	H1	H2	H3	H0	H1	H2	H3
GDP	0.99	0.96	0.95	0.97	0.99	0.97	0.96	0.97
Consumption	0.98	1.00	0.95	1.00	0.96	0.97	0.96	1.00
Investment	1.08	1.03	1.04	1.01	1.03	1.02	1.02	0.99

- No cell of 36 differs significantly from one in either comparison.
- Text is ahead of the hard data for GDP beyond the nowcast (0.96 / 0.95 / 0.97) and behind for investment.

Best model on each side chosen ex post, per outcome, horizon and month; mean over M1–M3.

Economic articles: adding news with the model held fixed

Model	H0	H1	H2	H3	Sig. better	Sig. worse
Elastic Net	1.02	0.97	1.02	1.05	1 of 36	1 of 36
PCA-OLS	1.00	0.93	0.92	0.93	6 of 36	0 of 36
Random forest	1.00	1.00	0.98	0.99	3 of 36	0 of 36
XGBoost	1.02	0.99	0.90	0.96	5 of 36	1 of 36
All models	1.01	0.97	0.96	0.98	15 of 144	2 of 144
GDP	1.01	0.97	0.95	0.95	4 of 48	2 of 48
Consumption	0.98	0.97	0.94	0.98	10 of 48	0 of 48
Investment	1.04	0.98	0.98	1.00	1 of 48	0 of 48

- Holding the model fixed, adding text to the hard data lowers RMSE in two thirds of 144 comparisons (median ratio 0.98): 15 significant improvements, 2 deteriorations. About 7 of each would be expected by chance.
- Ten of the 15 are consumption; eleven are at H2, where the average ratio is 0.96.

All 144 comparisons By period

Same model with and without text; ratio below one = text helps. Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Full sample: news only against the AR benchmark

Outcome	Ellingsen et al. (2022) Table 1	This paper, main specification	This paper, with the methods of Ellingsen et al.
GDP	0.94 / 0.95 / 0.96	0.94 / 0.95 / 0.95	1.01 / 0.98 / 0.98
Consumption	0.89 / 0.89 / 0.92	0.87 / 0.93 / 0.89	0.91 / 0.93 / 0.89
Investment	0.91 / 0.94 / 1.02	1.04 / 1.02 / 1.04	1.13 / 1.09 / 1.13

- On the full evaluation sample the best news-only model is below the AR benchmark for GDP in all 12 cells (0.91–0.98) and for consumption in all 12 (0.84–0.99). Ellingsen et al. (2022) report 0.91–0.97 and 0.87–0.96 at H0–H2.
- With a test that allows for serially correlated multi-step errors, the best news-only model is significantly better than AR in 2 of 36 cells and the best hard + news model in 4; all six are consumption at H0 or H1.
- The Survey of Professional Forecasters is ahead of the best news-only model in all 36 cells, significantly in 30.

Significance table

Specification of Ellingsen et al.

Relative RMSE, mean over M1–M3, H0 / H1 / H2. Best model chosen ex post (of three in the paper; of four / three here).

Paper: evaluation 1996Q1–2020Q1; here 2003Q1–2020Q1. Paper column: Ellingsen et al. (2022), Table 1. Stars:

Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Four quarters hold half of the forecast error

Outcome	H0	H1	H2	H3
GDP	41%	48%	50%	51%
Consumption	32%	31%	36%	37%
Investment	51%	59%	59%	60%

- Four quarters, 2008Q3–2009Q2, hold 41–51% of the AR benchmark's total squared error for GDP, 31–37% for consumption and 51–60% for investment. They are 4 of 69 evaluation quarters.
- Every full-sample RMSE ratio is therefore largely a statement about those four quarters.

How the periods are defined

Share of the AR benchmark's squared forecast errors over 2003Q1–2020Q1 that falls in 2008Q3–2009Q2.

Against the benchmark, by period

Outcome	Best model	All 69 quarters	Without the four worst quarters	2008–2009 only
GDP	News only	0.95	1.04	0.88
	Hard data only	0.98	1.02	0.96
	Hard data + news	0.95	1.05	0.90
Consumption	News only	0.89	0.98	0.73
	Hard data only	0.91	0.99	0.83
	Hard data + news	0.88	0.97	0.74
Investment	News only	1.03	1.28	0.83
	Hard data only	0.98	1.05	0.96
	Hard data + news	1.01	1.16	0.93

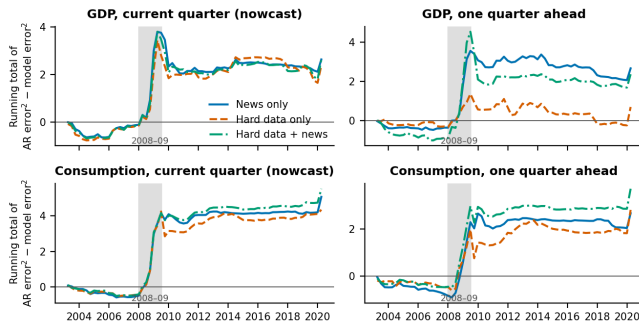
- **How to read:** RMSE of the best model divided by the RMSE of the AR benchmark, averaged over the nowcast and one and two quarters ahead. Below 1 = better than AR.
- Without the four worst quarters no model beats the AR benchmark for GDP: 1.04, 1.02, 1.05.
- In 2008–2009 news alone is 12% better than AR for GDP and 27% better for consumption; hard data alone 4% and 17%. Eight quarters: no test.

By horizon

Consumption and investment

Main specification. Best model chosen ex post on the full sample and not re-chosen. Four worst quarters: 2008Q3–2009Q2. Periods are defined on the quarter being forecast.

When the gains over the benchmark are made



- **How to read:** a running total of the AR benchmark's squared error minus the model's. The line rises in a quarter when the model was more accurate than AR and falls when AR was.
- Until 2007 the lines stay at or below zero: no gain. In 2008–09 they jump; after 2010 they are roughly flat.
- Current quarter: the three lines jump together. One quarter ahead: the lines with news jump, hard data alone less — there the gain comes from text.

The step at the right edge is 2020Q1, the first COVID quarter.

Two quarters ahead

Two recessions: 2001 and 2008–2009

New York Times and Wall Street Journal only, the two papers with articles in every year: training from 1990, evaluation 1996Q1–2020Q1 (97 quarters), main specification.

Best model	2001 recession		2008–2009	
	GDP	Consumption	GDP	Consumption
News only	1.10	1.23	0.85	0.74
Hard data only	1.06	1.19	0.88	0.76
Hard data + news	1.10	1.26	0.86	0.75

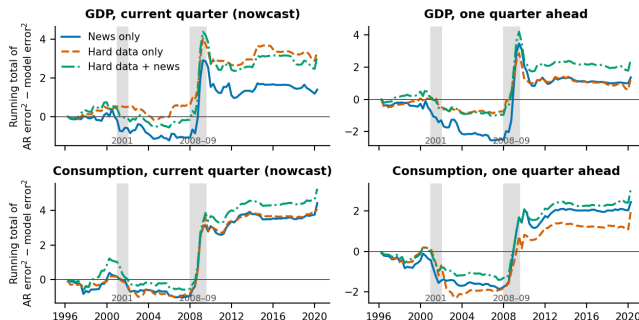
- **How to read:** RMSE of the best model divided by the RMSE of the AR benchmark inside each recession, averaged over the nowcast and one and two quarters ahead. Below 1 = better than AR.
- In the 2001 recession nothing beat the AR benchmark: news 1.10 for GDP and 1.23 for consumption, hard data 1.06 and 1.19.
- In 2008–2009 everything did: news 0.85 and 0.74, hard data 0.88 and 0.76. With two papers, text and hard data are level.

By horizon

All periods

Best model chosen ex post on the full sample and not re-chosen. 2001 recession = 2001Q1–2001Q4; 2008–2009 = 2008Q1–2009Q4.

Two recessions: when the gains are made



- Same reading: a rising line means the model was more accurate than the AR benchmark in that quarter. Two papers, 1996–2020, main specification.
- Around 2001 the lines fall or stay flat: no model gained on AR in that recession.
- In 2008–09 they all jump, text and hard data alike.

Two quarters ahead

Robustness: two further exercises

- **Outlet lean and party control** (564k-article sample). Is news from Democrat-leaning outlets more useful for forecasting when Democrats hold the presidency, the Senate or the House?
- **Result:** no. Pooled over GDP, consumption and investment the gap does not move with party control (placebo p-values 0.25–0.88). **Party control**
- **Look-ahead bias** (Sarkar and Vafa, 2024). The sentence encoder behind the filter and FinBERT are pretrained on text that includes years after the dates being forecast.
- **Result so far:** sentiment scored with a language model trained only on text up to 1989 (StoriesLM) forecasts no differently from FinBERT (39,492 US articles, earlier targets). Without it the best news-only model is at 1.00 / 0.94 / 0.93 of AR for GDP in 2008–2009, against 0.88 / 0.88 / 0.86 with it. **Look-ahead bias**

StoriesLM: 1989 checkpoint with a sentiment head trained on Financial PhraseBank, run of 2026-09-10. Filter: main specification on all articles against economic articles; H0 / H1 / H2, mean over M1–M3.

Conclusions

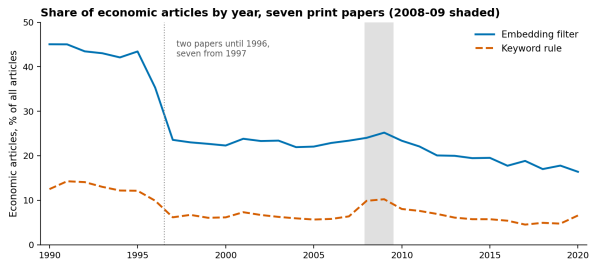
- Text was a turning-point signal in the Great Recession (2008–2009): the best news-only model beat the AR benchmark by 12–14% for GDP and 22–35% for consumption.
- In normal times neither text nor hard data beat a simple autoregression.
- Added to the hard data, text gives a small, significant gain for consumption (about 3%), inside and outside the recession. For GDP it gives none outside the recession.
- Investment is the exception: outside the recession news is far behind the benchmark and makes the hard-data forecast worse.
- The filter matters: the same model does significantly better on economic articles in 27 of 144 comparisons and worse in 4, in normal times as well.
- It is one episode: the 2001 recession shows no such gain.

Two baseline papers

- Baseline Paper 1: News media versus FRED-MD for macroeconomic forecasting by Ellingsen et al (2022) [▶ Base Paper](#)
 - Ex-post best-model selection
 - Proprietary single corpus
- Baseline Paper 2: Measuring Economic Policy Uncertainty (Baker et al, 2016) [▶ Paper](#)
 - No OOS forecasting
 - Dictionary-based measurement
 - Attention $\neq$ content

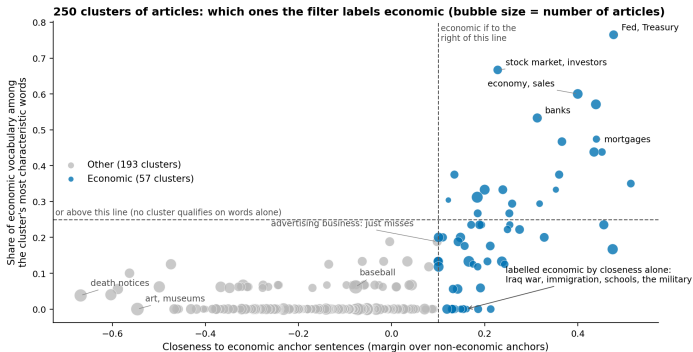
[◀ Back](#)

The economic share is small and moves over time



- A keyword rule flags 7.9% of all articles; the embedding filter keeps 24.7%.
- The share is not stable: by the embedding filter 42–45% in 1990–95 (two papers), 22–25% in 1997–2009, 16–19% in 2016–20. By the keyword rule it falls from 8% (2010) to 4.7% (2019).
- Papers write more about the economy in recessions: by the keyword rule the share rises from 6% to 10% in 2008–09 (2001: 6% to 7%); by the embedding filter from 23% to 25%.
- Topic shares computed on all articles also move with the other three quarters of the paper. For a macro forecast that movement is noise.

Which clusters the filter labels economic



- Each bubble is one of the 250 clusters. A cluster is labelled economic when it is close to economic anchor sentences (right of the vertical line) or when its characteristic words are economic (above the horizontal line).
- All 57 economic clusters qualify by closeness; 21 of them also by their words. The 36 that qualify by closeness alone include clusters on the Iraq war, immigration, schools and the military.

Keywords against the embedding filter: how much each catches

Flagged as economic by	Articles	Share
Both rules	802,973	6.8%
Embedding filter only	2,105,994	17.9%
Keywords only	120,390	1.0%
Neither	8,730,997	74.2%
Keyword rule, total	923,363	7.9%
Embedding filter, total	2,908,967	24.7%

- Of the articles the keywords flag, 87% are also kept by the embedding filter.
- The embedding filter adds 2.1 million articles that the keywords miss: 72% of what it keeps.
- Only the keywords flag 120 thousand articles, 1% of the corpus.

Keyword rule: two of 20 term groups in the title and lead, or one in the title. Embedding filter: economic cluster, or anchor margin of 0.10.

Keywords against the embedding filter: examples

Keywords only	Embedding filter only	Both rules
Raiders Close to Hiring Payton (Los Angeles Times, 2004)	How debt-ceiling talks came apart (Chicago Tribune, 2011)	Strong I.B.M. Report Helps Propel Dow to Another High (New York Times, 1999)
Italy lending Bernini art to the Getty (Los Angeles Times, 2008)	Investors say no to bid from Genzyme (Boston Globe, 2007)	Bank to Buy Credit Card Business for \$6.45 Billion (New York Times, 2005)
Starbucks dips toe into beer and wine sales (USA Today, 2009)	Web Ads Hit Rock Bottom: Some Are Free (Wall Street Journal, 2002)	Bush Plan Would Hold Increase In 2004 Spending to Tight 4% (Wall Street Journal, 2003)
Trump May Choose Critic of Fannie and Freddie to Lead Agency Overseeing Them (New York Times, 2018)	A spin around the old vinyl (Boston Globe, 2014)	Dell, EMC Complete Merger (Wall Street Journal, 2016)

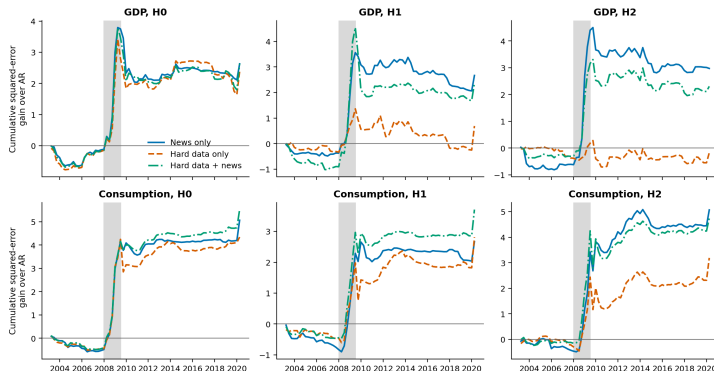
- Keywords alone misfire on words out of context: “hiring” a coach, “lending” a sculpture.
- The embedding filter picks up company, deal and budget news without those words. It has false alarms of its own.

All 45 headlines

Four of 15 randomly sampled headlines per group: seven print papers, 1997–2020, fixed seed. Headlines only.

When the gains are made: nowcast to two quarters ahead

Gains over the AR benchmark are made in the Great Recession (shaded); beyond the nowcast they come from text

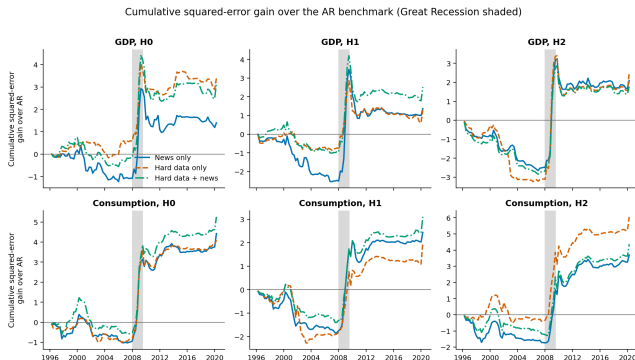


Beyond the nowcast the Great Recession gain over AR comes from text; hard data alone gain little at H1 and H2.

Cumulative sum of the AR benchmark's squared error minus the model's, best model per set, mean over M1–M3; shaded: 2007Q4–2009Q2.

◀ Back

Two recessions: nowcast to two quarters ahead

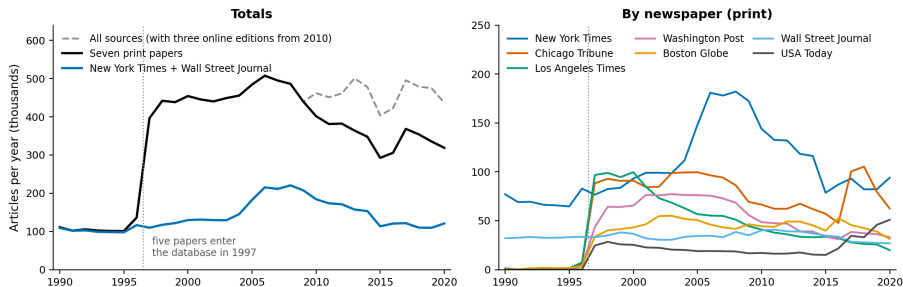


Cumulative squared-error gain over the AR benchmark, two papers, main specification, 1996–2020. Before 2008 every model set loses ground to AR, including around the 2001 recession; the gains are made in 2008–09, by text and by hard data alike.

Cumulative sum of the AR benchmark's squared error minus the model's, best model per set, mean over M1–M3; shaded: 2007Q4–2009Q2.

Data — Newspaper Coverage

Articles per year in the corpus: a step in 1997, a peak in 2006, no growth after



- Seven print papers: about 100 thousand articles a year until 1995 (two papers), 397 thousand in 1997, a peak of 508 thousand in 2006, 319 thousand in 2020.
- With the online editions of three papers, from 2010, the total stays between 400 and 500 thousand a year.

Dotted line: five of the seven papers enter the database in 1997.

◀ Motivation ▶ Data

Data — Newspaper Coverage (cont.)

Articles per newspaper, selected years

Newspaper	1992	1994	1996	1997	1998	2005	2015	2020
New York Times	69,308	65,698	82,832	76,605	82,329	147,960	78,596	93,927
Wall Street Journal	33,344	32,573	33,595	33,058	35,024	34,394	34,764	26,956
Boston Globe	1,342	1,542	3,513	33,649	40,106	50,606	39,878	31,460
Los Angeles Times	1,339	1,361	7,352	96,738	98,952	56,712	33,479	19,942
Chicago Tribune	19	5	5,782	88,249	92,891	99,626	57,088	62,520
Washington Post	0	0	3,152	43,612	64,262	76,136	33,762	33,159
USA Today	0	0	4	24,617	28,426	18,921	15,001	50,921

Five of the seven papers have essentially no articles before 1997; the cross-section is NYT+WSJ until 1996 and seven papers after.

◀ Back

Sampled headlines: keywords only

Newspaper	Year	Headline (cut at 54 characters)
Los Angeles Times	2006	Red Cross Chapter Chief Resigns Amid Complaints; Cont...
Chicago Tribune	2017	American Dental Association Foundation Announces Hiri...
Los Angeles Times	2020	Artists are essential too; Stop assuming that content...
Los Angeles Times	2004	AROUND THE NFL; Raiders Close to Hiring Payton
Wall Street Journal	2000	The Homestretch
New York Times	2018	Trump May Choose Critic of Fannie and Freddie to Lead...
Boston Globe	2018	New England colleges have one big worry: 2025
USA Today	2009	Starbucks dips toe into beer and wine sales
Los Angeles Times	2008	Italy lending Bernini art to the Getty; Major exhibit...
New York Times	2014	In Detroit, Judge Takes Offbeat Tack in Hiring Help
Wall Street Journal	2012	Next: Planning & Living The New Retirement (A Special...
Los Angeles Times	1998	Don't Rush to Raise Teacher Pay
Los Angeles Times	2008	HARD TIMES: CULTURE AT A CROSSROADS; Panic and anger...
Boston Globe	2009	Gerald Scully; applied economics to professional base...
Chicago Tribune	2014	King's abdication could be end of royal line in Spain

15 headlines drawn at random from the 93,665 articles of the seven print papers, 1997–2020, that only the keyword rule flags.

◀ Back

Sampled headlines: embedding filter only

Newspaper	Year	Headline (cut at 54 characters)
Chicago Tribune	2011	How debt-ceiling talks came apart
Washington Post	2006	WORKING
Boston Globe	2007	Investors say no to bid from Genzyme; After rare defe...
Wall Street Journal	2014	Overheard
Boston Globe	2003	HYDROGEN HOT AIR
Wall Street Journal	2002	Web Ads Hit Rock Bottom: Some Are Free
Wall Street Journal	1997	Party Line: The Battle for MCI Takes Another Twist: N...
Boston Globe	2014	A spin around the old vinyl
New York Times	1999	Newport News Joins in a New Venture
Boston Globe	2003	OUR DREAM OF A GREAT SOCIETY
New York Times	2010	An Unlikely Path to a Great Workplace
Los Angeles Times	2008	SMALL BUSINESS; Shoemaker in step with clients; J&A S...
Chicago Tribune	2002	'Are you serious? That's him? ; Former Enron executiv...
Los Angeles Times	2011	Amy Baer out as head of CBS Films; The executive left...
Chicago Tribune	2001	A few simple steps key to short storage

15 headlines drawn at random from the 1,576,849 articles of the seven print papers, 1997–2020, that only the embedding filter keeps.

◀ Back

Sampled headlines: both rules

Newspaper	Year	Headline (cut at 54 characters)
New York Times	1999	THE MARKETS: STOCKS; Strong I.B.M. Report Helps Prope...
New York Times	2005	Bank to Buy Credit Card Business for \$6.45 Billion
Los Angeles Times	2010	MONDAY BUSINESS; SMALL BUSINESS / Q & A; Getting more...
Wall Street Journal	2003	Bush Plan Would Hold Increase In 2004 Spending to Tig...
New York Times	2020	Kodak C.E.O. Got Stock Options Day Before News of Loa...
USA Today	2002	Enron workers get \$29M 'beacon of hope' ; 4,200 forme...
Los Angeles Times	1999	China Budgets Red Ink to Meet 7% Growth Plan
Wall Street Journal	2003	Beiersdorf Sharpens Its Brand Focus
Washington Post	1999	Save Our Steel
Los Angeles Times	2000	WALL STREET, CALIFORNIA; MONEY MAKE-OVER / Southern C...
Wall Street Journal	2016	Dell, EMC Complete Merger
Wall Street Journal	2018	Universal Income, Incentives and Well-Being
Wall Street Journal	2006	Currency Games
Wall Street Journal	2019	World News: China Counts the Costs Of Its Big Bet on...
Chicago Tribune	1997	PRICIEST OF ALL POSSIBLE WORLDS INFLATION MUST BE A T...

15 headlines drawn at random from the 547,721 articles of the seven print papers, 1997–2020, that both rules flag.

◀ Back

Filter effect: all specifications

Specification	News only vs. AR, H0 / H1 / H2						Hard + news vs. hard, same model			News vs. hard,
	Consumption		GDP	Investment		Median	Below one	Sig. < 1 / > 1	best vs. best	
Three models, all articles	0.90 / 0.94 / 0.89	1.03 / 1.00 / 0.99	1.21 / 1.12 / 1.17	1.00	52%	5 / 8	1.10 / 1.05 / 1.03			
Three models, economic articles	0.91 / 0.93 / 0.89	1.01 / 0.98 / 0.98	1.13 / 1.09 / 1.13	0.99	59%	9 / 8	1.08 / 1.03 / 1.01			
Four models, all articles	0.88 / 0.95 / 0.90	1.01 / 0.99 / 0.98	1.23 / 1.15 / 1.13	1.00	49%	5 / 7	1.12 / 1.06 / 1.03			
Four models, economic articles	0.87 / 0.93 / 0.89	0.94 / 0.95 / 0.95	1.04 / 1.02 / 1.04	0.98	64%	11 / 2	1.02 / 1.00 / 0.98			

Same hard data (full FRED-MD, outliers screened), same window. Three models: LASSO, PCA-OLS (one factor), random forest, topics only. Four models: Elastic Net, PCA-OLS (five factors), random forest, XGBoost, all text measures. H0-H2; corrected test.

[◀ Back](#)

Baseline, all articles: news only against the benchmarks

Outcome	Benchmark	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.02	1.02	1.00	0.99	0.98	1.00	0.99	0.97	0.99	0.99	0.99	0.99
	RW	0.93	0.93	0.91	0.96	0.96	0.97	0.99	0.97	0.98	0.99	0.99	0.99
	SPF	1.30**	1.29**	1.27**	1.15*	1.15*	1.16**	1.11**	1.08**	1.10**	1.05	1.05*	1.05
Consumption	AR	0.88	0.91	0.85*	0.95	0.95	0.94	0.90	0.91	0.90*	0.91**	0.91**	0.92*
	RW	0.85*	0.87	0.82**	0.92	0.92	0.91*	0.93	0.93	0.92	0.91**	0.91**	0.92*
	SPF	1.20**	1.23***	1.16**	1.17**	1.17**	1.16**	1.12	1.13*	1.11*	1.06	1.05	1.06
Investment	AR	1.25***	1.24***	1.20***	1.17**	1.14*	1.14	1.13*	1.12*	1.15*	1.17***	1.16**	1.11**
	RW	1.08	1.07	1.04	1.13	1.11	1.10	1.14*	1.14*	1.17*	1.18***	1.18***	1.13**
	SPF	1.64***	1.63***	1.58**	1.45***	1.42***	1.41***	1.32***	1.31***	1.35***	1.30***	1.29***	1.24**

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

◀ Back

Baseline, all articles: news against hard data

Outcome	Model	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	Elastic Net	1.15***	1.11***	1.17***	1.05	1.04	0.96	0.95	0.93	1.00	0.82	0.86	0.97
	PCA-OLS	1.13**	1.11*	1.13*	1.00	1.03	1.00	0.95	0.96	0.98	0.86*	0.87	0.91
	Random forest	1.01	1.01	1.00	0.99	0.98	1.01	0.98	0.96	0.99	0.97	0.98	1.01
	XGBoost	1.09*	1.06	1.07	1.00	1.06	1.10	0.95	0.99	1.04	0.94	1.00	0.98
	<i>Best vs. best</i>	1.07	1.05	1.09	0.99	1.01	1.01	0.98	0.96	0.99	0.97	0.98	1.00
Consumption	Elastic Net	1.06	1.05	0.93	0.88	0.97	0.93	0.92	1.01	1.12	0.94	0.90	0.86
	PCA-OLS	1.09	1.04	1.04	1.05	1.06	1.05	0.95	0.99	0.99	0.87**	0.85	0.88
	Random forest	0.98	0.97	0.93*	1.01	0.98	0.97	0.94*	0.94	0.95	0.94**	0.93**	0.93*
	XGBoost	0.97	0.98	0.92	1.01	1.04	0.99	0.88	0.95	0.99	1.01	0.95	0.96
	<i>Best vs. best</i>	0.99	1.00	0.99	1.01	1.03	1.02	0.94	0.98	0.99	0.94**	0.93**	0.93*
Investment	Elastic Net	1.27**	1.18	1.30*	0.99	1.09	1.24*	1.13	1.07	1.26*	1.06	0.85	0.92
	PCA-OLS	1.37**	1.42**	1.37**	1.18	1.14	1.11	1.11	1.03	1.07	0.99	0.90	1.04
	Random forest	1.27***	1.29***	1.31***	1.17**	1.16***	1.18*	1.07	1.11*	1.19*	1.09	1.10*	1.11
	XGBoost	1.25	1.35**	1.43***	1.02	1.42	1.42*	0.91	1.00	1.17	1.07	1.09	1.22*
	<i>Best vs. best</i>	1.27***	1.29***	1.30*	1.17**	1.16***	1.18*	1.08	1.11*	1.19*	1.09	1.09	1.06

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.[◀ Back](#)

Baseline, all articles: adding news, model by model

Outcome	Model	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	Elastic Net	1.12**	1.04*	1.13*	1.01	1.04	0.99	0.97	1.01	1.05*	0.89	0.97	1.01
	PCA-OLS	1.00	1.02	1.04	0.97	1.01	0.97	0.93	1.00	1.00	0.90**	0.95	0.96
	Random forest	1.00	1.00	0.99	0.99	0.99	1.00	0.98	0.97	0.99	0.97	0.98	1.00
	XGBoost	1.05	0.98	1.00	1.02	1.03	1.00	0.96	0.94	0.98	0.96	1.00	0.96
Consumption	Elastic Net	0.97	0.99	0.93	0.93	0.92	0.95	1.03	1.07	1.17	1.08	1.06	1.03
	PCA-OLS	1.02	0.98	1.00	1.00	1.06	1.06	0.93**	1.00	1.01	0.91**	0.92*	0.93
	Random forest	0.99	0.99	0.96	1.01	0.99	0.98	0.95*	0.95	0.96*	0.95***	0.95***	0.94**
	XGBoost	0.98	0.98	0.99	0.99	0.98	0.97	0.92*	0.95	0.98	0.95*	0.92***	0.94**
Investment	Elastic Net	1.06	0.96	1.07	1.07	1.15	1.09	1.03	1.00	1.02	0.99	0.82	0.84
	PCA-OLS	1.10	1.19*	1.15	1.04	1.06	0.99	0.97	1.07	1.02	0.94	1.04	1.09
	Random forest	1.04	1.04	1.03	1.07*	1.08**	1.03	1.04	1.04	1.07	1.03	1.04	1.06
	XGBoost	0.96	1.01	0.94	0.95	1.03	1.00	0.88*	0.80	1.00	0.95	0.90	0.96

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.[◀ Back](#)

Baseline, all articles: against the benchmark, by period

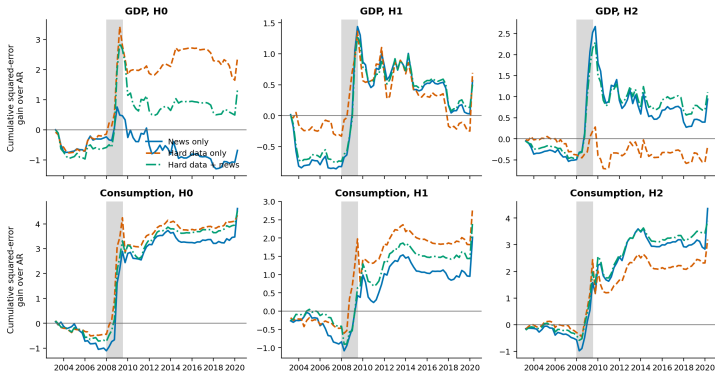
Outcome	Best model	Full sample (69 q.)	Without 2008Q3–2009Q2 (65 q.)	2008–2009 (8 q.)
GDP	News only	1.01 / 0.99 / 0.98 / 0.99	1.05 / 1.05 / 1.06 / 1.03	1.00 / 0.94 / 0.93 / 0.95
	Hard data only	0.95 / 0.99 / 1.00 / 1.00	1.01 / 1.02 / 1.03 / 1.03	0.90 / 0.97 / 1.00 / 0.98
	Hard data + news	0.97 / 0.99 / 0.98 / 0.99	1.07 / 1.05 / 1.05 / 1.03	0.92 / 0.95 / 0.94 / 0.94
Consumption	News only	0.88 / 0.95 / 0.90 / 0.91	0.96 / 0.98 / 0.93 / 0.91	0.71 / 0.88 / 0.86 / 0.91
	Hard data only	0.89 / 0.93 / 0.93 / 0.98	0.99 / 0.99 / 0.99 / 1.00	0.74 / 0.88 / 0.89 / 0.93
	Hard data + news	0.88 / 0.94 / 0.91 / 0.92	0.97 / 0.97 / 0.95 / 0.93	0.72 / 0.89 / 0.85 / 0.90
Investment	News only	1.23 / 1.15 / 1.14 / 1.15	1.41 / 1.45 / 1.36 / 1.31	1.07 / 0.92 / 0.98 / 1.04
	Hard data only	0.96 / 0.98 / 1.01 / 1.06	1.00 / 1.09 / 1.04 / 1.05	0.96 / 0.93 / 1.01 / 1.07
	Hard data + news	0.98 / 1.04 / 1.05 / 1.08	1.09 / 1.24 / 1.17 / 1.14	0.92 / 0.91 / 0.99 / 1.04

Relative RMSE against the AR benchmark, mean over M1–M3, H0 / H1 / H2 / H3. Best model chosen ex post on the full sample and not re-chosen.

[◀ Back](#)

Baseline, all articles: where the gains are made

All articles, four models: cumulative squared-error gain over the AR benchmark (Great Recession shaded)



Cumulative sum of the AR benchmark's squared error minus the model's, best model per set, mean over M1–M3; shaded: 2007Q4–2009Q2.

[◀ Back](#)

Against the benchmark, by period: every horizon

Outcome	Best model	Full sample (69 q.)	Without 2008Q3–2009Q2 (65 q.)	2008–2009 (8 q.)
GDP	News only	0.94 / 0.95 / 0.95 / 0.97	1.03 / 1.03 / 1.06 / 1.01	0.88 / 0.88 / 0.86 / 0.94
	Hard data only	0.95 / 0.99 / 1.00 / 1.00	1.01 / 1.02 / 1.03 / 1.03	0.90 / 0.97 / 1.00 / 0.98
	Hard data + news	0.94 / 0.96 / 0.96 / 0.97	1.02 / 1.08 / 1.03 / 1.01	0.89 / 0.90 / 0.91 / 0.95
Consumption	News only	0.87 / 0.93 / 0.89 / 0.97	0.96 / 1.01 / 0.96 / 0.98	0.65 / 0.75 / 0.78 / 0.97
	Hard data only	0.89 / 0.93 / 0.93 / 0.98	0.99 / 0.99 / 0.99 / 1.00	0.74 / 0.88 / 0.89 / 0.93
	Hard data + news	0.85 / 0.90 / 0.89 / 0.98	0.94 / 0.98 / 0.98 / 0.97	0.67 / 0.76 / 0.79 / 0.98
Investment	News only	1.04 / 1.02 / 1.04 / 1.07	1.30 / 1.32 / 1.22 / 1.20	0.77 / 0.80 / 0.92 / 0.98
	Hard data only	0.96 / 0.98 / 1.01 / 1.06	1.00 / 1.09 / 1.04 / 1.05	0.96 / 0.93 / 1.01 / 1.07
	Hard data + news	0.99 / 1.00 / 1.03 / 1.05	1.08 / 1.23 / 1.18 / 1.13	0.92 / 0.90 / 0.96 / 1.00

- Without them no model beats the AR benchmark for GDP: news only 1.01–1.06, hard data 1.01–1.03, both together 1.01–1.08.
- In 2008–2009 the best news-only model is 12–14% below AR for GDP and 22–35% below for consumption at H0–H2; the hard-data model gains less beyond the nowcast. Eight quarters: no test.

Relative RMSE against the AR benchmark, mean over M1–M3, H0 / H1 / H2 / H3. Best model chosen ex post on the full sample and not re-chosen. Samples are defined on the quarter being forecast.

What the newspapers print: themes

What the articles are about	Share	Among their most characteristic words
Economy and business	21.9%	fed, inflation, stocks, earnings, banks, oil prices, mortgage, jobs
Sports	15.8%	game, season, team, coach, inning, quarterback
Arts, entertainment and media	14.3%	film, music, museum, theater, album, book
Food, fashion, travel, family and religion	9.2%	food, restaurant, fashion, family, children, church
World affairs, war and security	7.3%	military, iraq, israel, china, al qaeda, minister
Local news, weather and community	7.1%	county, park, council, traffic, village, hurricane
National politics	7.0%	campaign, election, republican, democratic, senate, white house
Crime, courts and accidents	5.9%	police, court, judge, prosecutors, murder, crash
Health, science, technology and education	5.5%	health, patients, cancer, students, school, scientists
Obituaries and death notices	4.5%	death notice, funeral, died, grandchildren
Other	1.4%	Spanish-language items, bridge column, advertising industry
All other themes together	78.1%	

250 clusters of sentence embeddings fitted on 300,000 sampled articles (1990–2020). Economy and business = the 57 clusters the filter labels economic; the other 193 are grouped into themes by hand from their most characteristic words. Shares by cluster label; the full rule (cluster or zero-shot) keeps 24.7% of all articles.

[◀ Back](#)

Real-time design

- Outcomes: real GDP, real consumption, real non-residential investment; quarterly log growth $\times 100$. Models are trained on first releases and scored against final-vintage actuals.
- Hard data: FRED-MD monthly vintages (300, 1999-08 to 2024-07); at each forecast origin the vintage of that month. Hard series are known through the previous month, news through the current month.
- U-MIDAS: every monthly predictor enters with six monthly lags as separate regressors; the same columns at M1, M2, M3 — only the information set changes.
- H0 = nowcast; H1–H3 = direct forecasts 1–3 quarters ahead. One model per outcome, horizon, month and predictor set.
- Expanding window; a training target is used only once it has been released; at least 20 training quarters.
- Main specification: training from 1997Q1, evaluation 2003Q1–2020Q1 (69 quarters).

Real-time design

Lags and missing months

Does text add to the hard data? By period

Evaluation sample	Quarters	All	GDP	Consumption	Investment
Full sample	69	0.98	0.97	0.97	1.00
Without 2008Q3–2009Q2	65	1.00	1.00	0.97	1.04
Through 2007	20	1.00	0.98	0.98	1.03
2008–2009	8	0.95	0.95	0.96	0.94
2010–2019	40	1.02	0.99	0.97	1.11

- By period the ratio is 0.95 in 2008–2009, 1.00 in 2003–2007 and 1.02 in 2010–2019.

How the periods are defined

Same model with and without text; ratio below one = text helps. Mean over models, horizons and months. Samples are defined on the quarter being forecast.

◀ Back

Against the benchmark, by period (cont.)

- Outside the recession consumption forecasts stay within a few percent of AR: news 0.96–1.01, hard data 0.99–1.00, both 0.94–0.98.
- Investment: the best news-only model is far above AR outside the recession (1.20–1.32).

Table on the previous frame. Relative RMSE against the AR benchmark, mean over M1–M3, H0 / H1 / H2 / H3. Best model chosen ex post on the full sample and not re-chosen. Samples are defined on the quarter being forecast.

◀ Back

Can the news measures be made less noisy?

	News only vs. AR, H0 / H1 / H2			Hard + news vs. hard, same model		
News measures	GDP	Consumption	Investment	Median	Below one	Sig. < 1 / > 1
Baseline: 160 series, six free lags	1.01 / 0.98 / 0.97	0.92 / 0.93 / 0.89	1.13 / 1.09 / 1.13	0.99	59%	9 / 8
Detrended topic series	1.02 / 0.99 / 0.96	0.99 / 1.02 / 0.99	1.03 / 1.02 / 1.00	0.99	53%	12 / 13
Quarterly averages	1.02 / 0.98 / 0.97	0.92 / 0.93 / 0.89	1.12 / 1.08 / 1.10	0.99	65%	9 / 5
Consensus only (80 series)	1.01 / 0.97 / 0.96	0.91 / 0.94 / 0.89	1.13 / 1.11 / 1.11	0.99	64%	11 / 4

	GDP			Consumption			Investment		
News measures	Full	2008–09	2010–19	Full	2008–09	2010–19	Full	2008–09	2010–19
Baseline: 160 series, six free lags	0.99	0.97	1.03	0.91	0.87	0.97	1.12	1.04	1.29
Detrended topic series	0.99	1.00	1.00	1.00	0.99	0.98	1.02	1.04	1.03
Quarterly averages	0.99	0.98	1.03	0.91	0.88	0.97	1.10	1.01	1.27
Consensus only (80 series)	0.98	0.97	1.00	0.91	0.86	0.96	1.12	1.03	1.29

- Quarterly averages instead of six free monthly lags, or dropping the 80 cross-publisher dispersion series, move no ratio in the table by more than 0.03.
- Detrending the topic series removes the investment losses of 2010–2019 ($1.29 \rightarrow 1.03$ against AR) and also the 2008–2009 gain for consumption ($0.87 \rightarrow 0.99$).
- Drift in topic shares accounts for the investment losses; no variant improves on the baseline overall.

Variants

Paper's three models, main specification. Second table: news only against AR, mean over H0–H2 and M1–M3, by evaluation sample.

Two recessions: two papers, 1996–2020

Outcome	Best model	Full sample (97 q.)	1996–2007 (48 q.)	2008–2009 (8 q.)	2010–2019 (40 q.)
GDP	News only	0.98 / 0.98 / 0.97 / 1.01	1.05 / 1.10 / 1.10 / 1.12	0.87 / 0.84 / 0.85 / 0.91	1.03 / 1.09 / 1.05 / 1.08
	Hard data only	0.95 / 0.98 / 0.97 / 1.00	0.97 / 1.03 / 1.12 / 1.10	0.89 / 0.93 / 0.83 / 0.89	0.99 / 1.05 / 1.03 / 1.13
	Hard data + news	0.95 / 0.97 / 0.98 / 1.00	1.01 / 1.04 / 1.11 / 1.10	0.86 / 0.89 / 0.85 / 0.92	1.02 / 1.05 / 1.04 / 1.05
Consumption	News only	0.91 / 0.95 / 0.94 / 1.00	1.05 / 1.10 / 1.09 / 1.18	0.70 / 0.72 / 0.78 / 0.93	0.91 / 1.00 / 0.90 / 1.05
	Hard data only	0.92 / 0.96 / 0.90 / 0.99	1.04 / 1.11 / 1.02 / 1.14	0.72 / 0.82 / 0.76 / 0.87	0.89 / 0.96 / 0.88 / 0.98
	Hard data + news	0.90 / 0.94 / 0.93 / 1.01	1.03 / 1.08 / 1.06 / 1.18	0.69 / 0.76 / 0.79 / 0.89	0.87 / 0.95 / 0.90 / 1.00
Investment	News only	1.03 / 0.96 / 1.00 / 1.05	0.96 / 0.90 / 0.96 / 0.97	0.88 / 0.87 / 0.95 / 0.95	1.33 / 1.29 / 1.20 / 1.41
	Hard data only	0.88 / 0.92 / 0.97 / 0.97	0.88 / 0.87 / 0.95 / 0.97	0.87 / 0.92 / 1.00 / 0.97	0.87 / 1.07 / 0.97 / 0.98
	Hard data + news	0.91 / 0.93 / 0.99 / 1.03	0.94 / 0.88 / 0.96 / 0.97	0.84 / 0.87 / 0.93 / 0.91	0.96 / 1.17 / 1.17 / 1.38

- New York Times and Wall Street Journal, four models, 97 quarters: news only against AR is 0.97–0.98 for GDP and 0.91–0.95 for consumption at H0–H2; one cell of 36 is significant.
- Before 2008 it is above one (GDP 1.05–1.12, consumption 1.05–1.18); in 2008–2009 it is 0.84–0.87 and 0.70–0.78 at H0–H2.

Relative RMSE against the AR benchmark, mean over M1–M3, H0 / H1 / H2 / H3. Best model chosen ex post on the full sample and not re-chosen. Samples are defined on the quarter being forecast. 2001 recession = 2001Q1–2001Q4.

Main specification: significant cells by model set and benchmark

Best model	Cells significantly below / above one, against		
	AR	Constant growth	SPF
News only	2 / 0	4 / 2	0 / 30
Hard data only	0 / 0	4 / 1	0 / 31
Hard data + news	4 / 1	6 / 2	0 / 33

Best model per set chosen ex post; 36 cells (3 outcomes $\times$ H0–H3 $\times$ M1–M3). Counted with the corrected test at 10%.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

◀ Back

How much of it is significant?

Outcome	Benchmark	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	0.97	0.95	0.91	0.97	0.96	0.92	0.95	0.94	0.95	0.97	0.98	0.96
	RW	0.88	0.86	0.83*	0.94	0.94	0.90	0.95	0.94	0.95	0.97	0.98	0.96
	SPF	1.23**	1.20**	1.16*	1.13*	1.12*	1.07	1.06*	1.05	1.06*	1.03*	1.04	1.01
Consumption	AR	0.89	0.87*	0.84**	0.95	0.94	0.91	0.90	0.89	0.87	0.99	0.98	0.95
	RW	0.86*	0.84**	0.81**	0.91	0.90	0.88	0.93	0.91	0.89	0.99	0.98	0.95
	SPF	1.21***	1.19***	1.14**	1.16***	1.15***	1.11***	1.12**	1.10**	1.08**	1.15	1.14	1.10*
Investment	AR	1.05	1.03	1.03	1.05	1.03	0.97	1.05	1.04	1.04	1.06	1.05	1.10
	RW	0.91	0.89	0.89	1.02	0.99	0.94	1.06	1.05	1.05	1.07	1.07*	1.12*
	SPF	1.37***	1.35***	1.35***	1.31**	1.27**	1.21**	1.23***	1.22***	1.22*	1.17***	1.17***	1.23**

- With a test that allows for serially correlated multi-step errors, the best news-only model is significantly better than AR in 2 of 36 cells and the best hard + news model in 4; all six are consumption at H0 or H1.
- The Survey of Professional Forecasters is ahead of the best news-only model in all 36 cells, significantly in 30.

Benchmarks and test

All tables

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Main specification: by evaluation sample, five samples

Outcome	Best model	Full sample (69 q.)				Without 2008Q3–2009Q2 (65 q.)				2003–2007 (20 q.)				2008–2009 (8 q.)				2010–2019 (40 q.)			
GDP	News only	0.94	0.95	0.95	0.97	1.03	1.03	1.06	1.01	1.02	1.06	1.11	0.99	0.88	0.88	0.86	0.94	1.02	1.08	1.05	1.05
	Hard data only	0.95	0.99	1.00	1.00	1.01	1.02	1.03	1.03	1.02	1.06	1.07	1.08	0.90	0.97	1.00	0.98	1.01	1.07	1.01	1.07
	Hard data + news	0.94	0.96	0.96	0.97	1.02	1.08	1.03	1.01	1.02	1.15	1.06	1.01	0.89	0.90	0.91	0.95	1.02	1.03	1.03	1.03
Consumption	News only	0.87	0.93	0.89	0.97	0.96	1.01	0.96	0.98	1.10	1.17	1.10	1.08	0.65	0.75	0.78	0.97	0.99	1.08	0.93	0.92
	Hard data only	0.89	0.93	0.93	0.98	0.99	0.99	0.99	1.00	1.09	1.10	1.07	1.10	0.74	0.88	0.89	0.93	0.87	0.95	0.96	1.01
	Hard data + news	0.85	0.90	0.89	0.98	0.94	0.98	0.98	0.97	1.08	1.09	1.04	1.13	0.67	0.76	0.79	0.98	0.90	1.01	0.97	1.08
Investment	News only	1.04	1.02	1.04	1.07	1.30	1.32	1.22	1.20	1.20	1.25	1.25	0.96	0.77	0.80	0.92	0.98	1.32	1.37	1.21	1.31
	Hard data only	0.96	0.98	1.01	1.06	1.00	1.09	1.04	1.05	1.09	1.05	1.00	0.88	0.96	0.93	1.01	1.07	0.88	1.10	1.01	1.10
	Hard data + news	0.99	1.00	1.03	1.05	1.08	1.23	1.18	1.13	1.11	1.30	1.06	0.93	0.92	0.90	0.96	1.00	0.85	1.12	1.17	1.22

Relative RMSE against the AR benchmark, mean over M1–M3, H0 / H1 / H2 / H3. Best model chosen ex post on the full sample and not re-chosen. Samples are defined on the quarter being forecast.

[◀ Back](#)

Main specification: best hard-only model vs. benchmarks

Outcome	Benchmark	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	0.95	0.97	0.92	1.00	0.98	0.99	1.01	1.01	0.99	1.01	1.01	0.99
	RW	0.87*	0.89	0.84	0.97	0.95	0.97	1.01	1.00	0.99	1.01	1.01	0.99
	SPF	1.21	1.24**	1.17**	1.16*	1.14	1.15*	1.12	1.12	1.10*	1.07**	1.07**	1.05*
Consumption	AR	0.89	0.91	0.86	0.94	0.92	0.92	0.96	0.92	0.91	0.97	0.98	0.98
	RW	0.86	0.87	0.83*	0.91	0.89	0.89*	0.99	0.95	0.93	0.97	0.98	0.98
	SPF	1.22**	1.24***	1.17**	1.15**	1.13*	1.13*	1.19*	1.14*	1.13*	1.13**	1.13**	1.14**
Investment	AR	0.99	0.96	0.92	1.00	0.99	0.96	1.04	1.01	0.97	1.07	1.07	1.05
	RW	0.85	0.83*	0.80	0.97	0.95	0.93	1.06*	1.02	0.98	1.08	1.08	1.06
	SPF	1.29**	1.26*	1.21	1.24**	1.22*	1.20*	1.22*	1.18*	1.13**	1.19*	1.18**	1.16*

Best hard-only model (101 FRED-MD series, outliers screened) / benchmark RMSE.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

◀ Back

Main specification: best hard + news model vs. benchmarks

Outcome	Benchmark	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	0.96	0.95	0.93	0.96	0.96	0.95	0.97	0.96	0.95	0.97	0.99	0.96
	RW	0.87	0.86	0.85	0.94	0.94	0.92	0.97	0.96	0.95	0.97	0.99	0.96
	SPF	1.21**	1.20**	1.18**	1.12*	1.12*	1.10	1.08*	1.07*	1.06*	1.03*	1.05*	1.02
Consumption	AR	0.87**	0.86**	0.83**	0.92	0.91	0.89*	0.91	0.89	0.88	0.98	0.97	0.97
	RW	0.84**	0.83**	0.80**	0.88	0.87*	0.86*	0.93	0.92	0.91	0.98	0.97	0.97
	SPF	1.19***	1.17***	1.13***	1.12***	1.11***	1.09**	1.12**	1.11**	1.10**	1.14*	1.13*	1.13*
Investment	AR	1.02	1.00	0.95	1.01	1.01	0.98	1.03	1.05	1.01	1.04	1.04	1.08*
	RW	0.88	0.86	0.82*	0.98	0.98	0.95	1.04	1.06	1.03	1.06	1.05*	1.09**
	SPF	1.33**	1.31**	1.25**	1.25**	1.25**	1.22**	1.20	1.23**	1.19***	1.16***	1.15***	1.20***

Best hard + news model / benchmark RMSE.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

[◀ Back](#)

Main specification: news only vs. hard only, same model

Outcome	Model	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	Elastic Net	1.05	0.99	1.14*	1.13*	1.07	1.00	1.10*	1.09*	1.08	0.86	0.88	1.05
	PCA-OLS	0.99	0.97	0.99	0.92	0.94	0.90	0.92	0.93	0.92	0.89**	0.90	0.92
	Random forest	0.99	0.99	1.00	0.97	0.96	0.98	0.96	0.94	0.96	0.96	0.97	0.96
	XGBoost	1.05	0.99	1.02	0.96	0.99	0.99	0.83*	0.85	0.95	0.90	0.90	0.89
	Best vs. best	1.02	0.97	0.99	0.97	0.99	0.93	0.94	0.93	0.96	0.96	0.97	0.96
Consumption	Elastic Net	1.09	0.97	0.92	1.01	1.02	0.94	0.86	0.88	1.07	0.88	0.92	0.88
	PCA-OLS	1.00	0.96	0.97	1.00	1.02	0.98	0.95	0.99	0.96	0.92	0.91	0.90
	Random forest	1.01	1.00	0.96	1.02	0.99	0.98	0.95	0.92*	0.94*	1.02	1.00	0.99
	XGBoost	1.04	1.04	0.98	1.02	1.00	0.97	0.88**	0.86**	0.92**	1.02	0.95	0.95
	Best vs. best	1.00	0.96	0.97	1.01	1.02	0.98	0.94	0.96	0.96	1.02	1.00	0.97
Investment	Elastic Net	1.22	1.12	1.30*	0.90	0.94	1.02	1.14	1.07	1.14	1.46	1.06	1.14
	PCA-OLS	1.05	1.07	1.08	0.91	0.89	0.86	0.95	0.90	0.89	0.97	0.92	0.96
	Random forest	1.14**	1.16**	1.18**	1.11	1.09	1.11	1.03	1.08*	1.13*	0.99	0.99	1.05
	XGBoost	1.03	1.10	1.30**	0.93	1.21	1.24*	0.76	0.74	1.03	0.93	0.92	1.10
	Best vs. best	1.06	1.08	1.11	1.05	1.04	1.01	1.01	1.03	1.08	0.99	0.99	1.05

News-only / hard-only RMSE for the same model; “best vs. best” = ex post best on each side.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Main specification: hard + news vs. hard only, same model

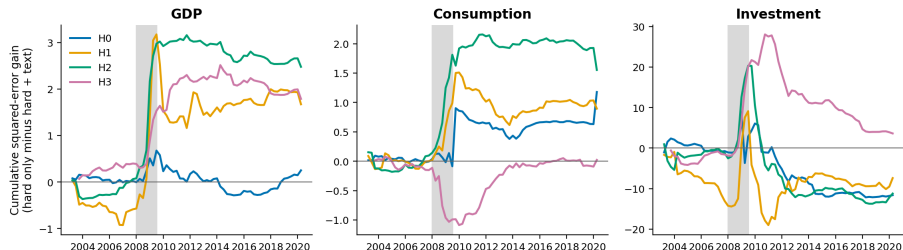
Outcome	Model	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	Elastic Net	1.08**	0.98	1.03	1.04	1.04	0.96	1.01	1.02	1.04	0.90	1.03	1.03
	PCA–OLS	0.97	0.97	1.01	0.91	0.94	0.93	0.90*	0.92	0.95	0.90***	0.93	0.96
	Random forest	0.99	0.99	0.99	0.97	0.97	0.98	0.96	0.95	0.96	0.96	0.98	0.97
	XGBoost	1.06*	1.00	1.03	0.96	1.00	1.00	0.90*	0.88*	0.94	0.90	0.96	0.91
Consumption	Elastic Net	0.99	0.96	0.89	0.91	1.02	0.90	0.87**	0.90	0.96	1.05	1.00	0.98
	PCA–OLS	0.98	0.94	0.97	0.97	0.98	0.97	0.93**	0.97	0.97	0.92**	0.92*	0.93
	Random forest	1.01	1.00	0.97	1.00	0.99	0.98	0.95*	0.94**	0.96*	1.01	0.99	0.99
	XGBoost	1.04	1.02	0.99	1.01	1.00	0.96	0.93*	0.93*	0.94*	1.04	0.98	0.97
Investment	Elastic Net	1.12	0.98	1.16	0.98	0.89	1.02	1.29	1.01	1.10	1.39	0.92	1.10
	PCA–OLS	1.02	1.08	1.07	0.87	0.88	0.90	0.86	0.87	0.91	0.91**	0.91	0.98
	Random forest	1.03	1.04	1.02	1.06	1.07	1.02	1.01	1.04	1.05	0.98	0.97	1.03
	XGBoost	0.94	1.06	0.99	0.95	1.05	1.02	0.82	0.82	0.95	0.96	0.90	1.00

Hard + news / hard-only RMSE for the same model; ratio below one = text helps.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Main specification: hard + text against hard only, over time

Adding text to the hard data pays off around the Great Recession (shaded), not after it

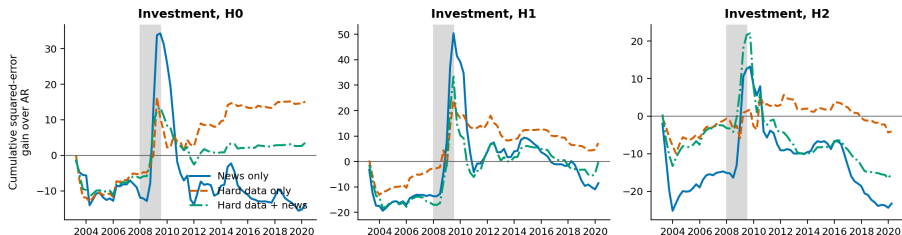


Cumulative sum of the hard-only model's squared error minus the hard + text model's.

◀ Back

Main specification: investment against AR, over time

Investment: cumulative squared-error gain over the AR benchmark



Investment: the best news-only model is far above AR outside the recession (1.20–1.32).

Cumulative sum of the AR benchmark's squared error minus the model's, best model per set, mean over M1–M3; shaded: 2007Q4–2009Q2.

◀ Back

Paper's three models: best news-only model vs. benchmarks

Outcome	Benchmark	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.01	1.01	1.01	0.99	0.99	0.97	1.00	0.98	0.95
	RW	0.92	0.92	0.92	0.96	0.96	0.95	0.99	0.97	0.95
	SPF	1.28**	1.28*	1.29*	1.15*	1.15*	1.13*	1.11*	1.09	1.06
Consumption	AR	0.91	0.92	0.91*	0.93	0.94	0.93	0.89	0.89*	0.89*
	RW	0.88*	0.89**	0.88**	0.90*	0.90*	0.90*	0.92	0.91	0.92
	SPF	1.24***	1.26**	1.24**	1.15**	1.15*	1.14*	1.11*	1.10*	1.11
Investment	AR	1.14***	1.15***	1.11*	1.08	1.12*	1.08*	1.12	1.15**	1.11
	RW	0.98	0.99	0.95	1.05	1.08	1.05*	1.13	1.17*	1.12
	SPF	1.49***	1.50***	1.45**	1.34*	1.39**	1.34*	1.31*	1.35***	1.30**

LASSO, PCA-OLS (one factor), random forest; topics only; main specification. With the paper's three models and topics only: consumption 0.89–0.94, GDP 0.95–1.01, investment 1.08–1.15. Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Paper's three models: significant cells

Best model	Cells significantly below / above one, against		
	AR	Constant growth	SPF
News only	3 / 6	6 / 2	0 / 24
Hard data only	5 / 0	8 / 1	0 / 21
Hard data + news	5 / 1	7 / 2	0 / 25

Best model per set chosen ex post. Counted with the corrected test at 10%.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Paper's three models: against the benchmark, by period

Outcome	Best model	Full sample (69 q.)	Without 2008Q3–2009Q2 (65 q.)	2008–2009 (8 q.)
GDP	News only	1.01 / 0.98 / 0.97	1.03 / 1.05 / 1.05	1.05 / 0.95 / 0.93
	Hard data only	0.99 / 0.99 / 1.01	0.99 / 1.00 / 1.04	1.02 / 0.99 / 1.00
	Hard data + news	1.00 / 0.98 / 0.98	1.00 / 1.03 / 1.05	1.03 / 0.96 / 0.95
Consumption	News only	0.92 / 0.93 / 0.89	0.95 / 0.94 / 0.91	0.85 / 0.91 / 0.83
	Hard data only	0.91 / 0.94 / 0.95	0.95 / 0.95 / 0.95	0.87 / 0.92 / 0.93
	Hard data + news	0.91 / 0.93 / 0.91	0.96 / 0.95 / 0.97	0.85 / 0.90 / 0.82
Investment	News only	1.13 / 1.09 / 1.13	1.19 / 1.17 / 1.29	1.09 / 1.03 / 1.01
	Hard data only	0.94 / 0.99 / 1.01	0.98 / 1.08 / 1.07	0.94 / 0.96 / 1.01
	Hard data + news	1.00 / 1.04 / 1.09	1.06 / 1.18 / 1.20	0.98 / 0.96 / 1.03

Relative RMSE against the AR benchmark, mean over M1–M3, H0 / H1 / H2 / H3. Best model chosen ex post on the full sample and not re-chosen. Samples are defined on the quarter being forecast.

Paper's three models: does text add to the hard data?

Model	H0	H1	H2	Sig. better	Sig. worse
LASSO	1.07	0.96	0.98	3 of 27	3 of 27
PCA-OLS	0.98	0.96	0.98	5 of 27	0 of 27
Random forest	1.02	1.01	1.00	1 of 27	5 of 27
All models	1.03	0.98	0.98	9 of 81	8 of 81
GDP	1.00	0.99	0.99	0 of 27	0 of 27
Consumption	1.02	0.96	0.94	7 of 27	2 of 27
Investment	1.05	0.98	1.02	2 of 27	6 of 27

Evaluation sample	Quarters	All	GDP	Consumption	Investment
Full sample	69	1.00	1.00	0.97	1.02
Without 2008Q3–2009Q2	65	1.02	1.03	0.99	1.03
Through 2007	20	1.00	0.99	0.98	1.03
2008–2009	8	0.97	0.97	0.95	0.99
2010–2019	40	1.05	1.04	1.03	1.07

Same model with and without text; ratio below one = text helps. Second table: mean over models, horizons and months, by evaluation sample (defined on the quarter being forecast). Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

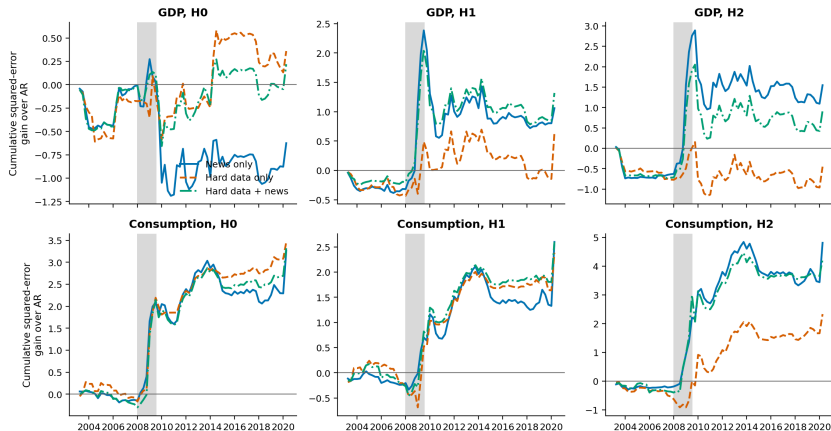
Paper's three models: text against hard data

Outcome	Text only vs. hard only			Hard + text vs. hard only		
	H0	H1	H2	H0	H1	H2
GDP	1.02	0.99	0.97	1.00	0.99	0.98
Consumption	1.01	0.99	0.94	1.00	0.99	0.95
Investment	1.20	1.11	1.11	1.06	1.05	1.07

Best model on each side chosen ex post, per outcome, horizon and month; mean over M1–M3.

Paper's three models: where the gains are made

Cumulative squared-error gain over the AR benchmark (Great Recession shaded)



Cumulative sum of the AR benchmark's squared error minus the model's, best model per set, mean over M1–M3; shaded: 2007Q4–2009Q2.

Without the outlier screen

Hard data	Hard only vs. AR				Hard + news vs. hard, same model			
	GDP		Consumption		Comparisons	Median	Below one	Sig. < 1 / > 1
Not screened	0.94 / 0.99 / 1.00 / 1.01	0.89 / 0.89 / 0.95 / 0.96	141	0.98	66%	12 / 3		
Outliers screened	0.95 / 0.99 / 1.00 / 1.00	0.89 / 0.93 / 0.93 / 0.98	144	0.98	66%	15 / 2		

- Screening the hard data for outliers changes no conclusion: 15 significant improvements and 2 deteriorations with the screen, 12 and 3 without.

Hard only vs. AR: best model, mean over M1–M3, H0 / H1 / H2 / H3. Same-model comparisons counted with the corrected test at 10%; without the screen the outlier-driven hard-only cells (2008Q2) are excluded, hence 141 comparisons.

From all articles to the main specification

Specification	News only vs. AR, H0 / H1 / H2			Hard + news vs. hard, same model			News vs. hard, best vs. best
	Consumption	GDP	Investment	Median	Below one	Sig. < 1 / > 1	
7 papers, all articles, 17 hard series	0.90 / 0.94 / 0.89	1.03 / 1.00 / 0.99	1.21 / 1.12 / 1.17	0.99	54%	1 / 13	1.12 / 1.04 / 1.00
+ economic articles	0.94 / 0.96 / 0.90	1.04 / 0.99 / 0.99	1.17 / 1.07 / 1.05	0.99	57%	5 / 5	1.12 / 1.03 / 0.97
+ topics re-fitted on them	0.91 / 0.93 / 0.89	1.01 / 0.98 / 0.98	1.13 / 1.09 / 1.13	0.99	53%	4 / 7	1.09 / 1.02 / 0.98
All articles, full FRED-MD	0.90 / 0.94 / 0.89	1.03 / 1.00 / 0.99	1.21 / 1.12 / 1.17	1.00	50%	6 / 10	1.10 / 1.05 / 1.03
Main spec, paper's three models	0.91 / 0.93 / 0.89	1.01 / 0.98 / 0.98	1.13 / 1.09 / 1.13	0.99	59%	9 / 8	1.08 / 1.03 / 1.01
Main spec, four models, all text	0.87 / 0.93 / 0.89	0.94 / 0.95 / 0.95	1.04 / 1.02 / 1.04	0.98	64%	11 / 2	1.02 / 1.00 / 0.98

Rows 1–5: paper's three models, topics; row 6: four models, all text measures. All rows: seven papers, training from 1997Q1, evaluation 2003Q1–2020Q1. “News vs. hard, best vs. best”: H0 / H1 / H2, mean over outcomes and M1–M3.
H0–H2; corrected test.

Walk, step 1: all articles, 17 hard-data series

Outcome	Benchmark	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.03	1.02	1.03	1.00	1.00	1.00	0.99	0.99	1.00
	RW	0.94	0.93	0.94	0.98	0.98	0.98	0.99	0.99	0.99
	SPF	1.31**	1.30**	1.31**	1.17*	1.17*	1.17*	1.10*	1.10	1.11*
Consumption	AR	0.89*	0.90*	0.92*	0.94	0.94	0.93	0.89	0.89	0.89
	RW	0.86**	0.87**	0.88**	0.91*	0.90*	0.90*	0.92	0.91	0.92
	SPF	1.22***	1.23**	1.25**	1.16*	1.15*	1.14*	1.11*	1.10*	1.11*
Investment	AR	1.20***	1.21***	1.22***	1.08	1.13*	1.14**	1.21***	1.17**	1.12*
	RW	1.03	1.05	1.05	1.04	1.10*	1.10	1.22***	1.19**	1.13*
	SPF	1.57**	1.59***	1.60***	1.33*	1.40**	1.41***	1.42***	1.37***	1.31***

Best news-only model / benchmark RMSE.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction: 10 / 5 / 1%.

◀ Back

Walk, step 1: all articles, 17 hard-data series (cont.)

Outcome	Model	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	LASSO	1.12**	1.14**	1.08	0.89	1.00	1.08**	0.96	0.96	0.97
	PCA-OLS	0.97	0.98	1.10	0.96	0.96	1.01	0.99	0.98	0.98
	Random forest	1.01	0.99	1.02	0.99	1.01	0.99	1.00	0.99	0.98
Consumption	LASSO	0.92	0.87	0.88	0.97	0.98	0.93	0.96	0.96	0.89
	PCA-OLS	0.98	0.99	0.98	0.96*	0.95	0.96	0.95	0.93	0.90
	Random forest	1.01	1.02	1.00	1.01	1.02	1.01	0.97	0.97	0.96
Investment	LASSO	1.14**	1.06	1.13*	0.80	1.02	1.14	1.35**	0.98	1.22**
	PCA-OLS	0.95	0.95	1.26*	0.98	0.94	1.09	1.07	1.03	1.06
	Random forest	1.16***	1.19***	1.19***	1.08	1.11**	1.15**	1.03	1.00	1.05

Hard + news / hard-only RMSE for the same model; ratio below one = text helps.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Walk, step 2: economic articles

Outcome	Benchmark	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.04	1.04	1.03	0.99	0.99	0.98	1.00	0.99	0.97
	RW	0.94	0.94	0.94	0.96	0.96	0.95	0.99	0.98	0.97
	SPF	1.32 [*]	1.31 [*]	1.31 [*]	1.15 [*]	1.15 [*]	1.14 [*]	1.11 [*]	1.10 [*]	1.08
Consumption	AR	0.94	0.94	0.95	0.95	0.95	0.97	0.90	0.90	0.91
	RW	0.90 [*]	0.90 [*]	0.91 ^{**}	0.91 [*]	0.92 [*]	0.93 [*]	0.92	0.92	0.94
	SPF	1.27 ^{**}	1.28 ^{**}	1.29 ^{***}	1.16 [*]	1.17 [*]	1.18 [*]	1.11 [*]	1.11	1.13
Investment	AR	1.17 [*]	1.17	1.18 ^{***}	1.08 [*]	1.08 [*]	1.06	1.08 [*]	1.04	1.02
	RW	1.01	1.01	1.02	1.05	1.04	1.02	1.09	1.05	1.03
	SPF	1.53 ^{**}	1.53 [*]	1.54 ^{**}	1.34 ^{**}	1.34 ^{**}	1.31 ^{**}	1.26 ^{**}	1.22 ^{**}	1.19 ^{***}

Best news-only model / benchmark RMSE.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

◀ Back

Walk, step 2: economic articles (cont.)

Outcome	Model	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	LASSO	1.05	1.06	1.05	0.89	0.99	1.02	0.94	0.97	1.00
	PCA-OLS	0.98	0.99	1.12	0.97	0.97	1.02	0.99	0.98	0.98
	Random forest	1.01	1.00	1.02	0.95*	0.98	0.98	0.96	0.95**	0.95
Consumption	LASSO	1.06	0.99	0.95	0.99	0.94	0.98	0.93	0.92	0.89
	PCA-OLS	0.99	1.01	1.01	0.97	0.97	0.99	0.96	0.94	0.92
	Random forest	1.05**	1.05*	1.04	1.01	1.04	1.03	0.98	0.96*	0.95**
Investment	LASSO	1.08	1.03	0.97	0.70	0.96	1.10	1.08	0.79	1.12
	PCA-OLS	0.95*	0.94	1.24	0.98	0.93	1.07	1.06	1.01	1.02
	Random forest	1.14**	1.17***	1.14**	1.01	1.03	1.08	0.94	0.93	0.97

Hard + news / hard-only RMSE for the same model; ratio below one = text helps.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

◀ Back

Walk, step 3: topics re-fitted on the economic articles

Outcome	Benchmark	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.01	1.01	1.01	0.99	0.99	0.97	1.00	0.98	0.95
	RW	0.92	0.92	0.92	0.96	0.96	0.95	0.99	0.97	0.95
	SPF	1.28**	1.28*	1.29*	1.15*	1.15*	1.13*	1.11*	1.09	1.06
Consumption	AR	0.91	0.92	0.91*	0.93	0.94	0.93	0.89	0.89*	0.89*
	RW	0.88*	0.89**	0.88**	0.90*	0.90*	0.90*	0.92	0.91	0.92
	SPF	1.24***	1.26**	1.24**	1.15**	1.15*	1.14*	1.11*	1.10*	1.11
Investment	AR	1.14***	1.15***	1.11*	1.08	1.12*	1.08*	1.12	1.15**	1.11
	RW	0.98	0.99	0.95	1.05	1.08	1.05*	1.13	1.17*	1.12
	SPF	1.49***	1.50***	1.45**	1.34*	1.39**	1.34*	1.31*	1.35***	1.30**

Best news-only model / benchmark RMSE.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

[◀ Back](#)

Walk, step 3: topics re-fitted on the economic articles (cont.)

Outcome	Model	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	LASSO	1.03	1.04	1.07	0.96	1.02	1.02	1.05	1.07	1.07
	PCA-OLS	0.98	0.98	1.10	0.97	0.97	1.02	0.99	0.98	0.98
	Random forest	0.99	0.99	1.00	0.94	0.97	0.97	0.97	0.95	0.93*
Consumption	LASSO	1.04	0.98	1.01	1.05	1.00	0.92	0.90	0.86**	0.82*
	PCA-OLS	0.99	0.99	0.97	0.96	0.95	0.96	0.96	0.93	0.91
	Random forest	1.02	1.03	1.03	1.01	1.03	1.03	0.98	0.94	0.94
Investment	LASSO	1.25*	1.35	1.20**	0.66	0.91	0.99	1.01	0.82	1.12
	PCA-OLS	0.95*	0.94	1.24*	0.98	0.94	1.08	1.06	1.01	1.05
	Random forest	1.12**	1.16***	1.15***	1.06	1.08	1.11*	1.03	1.01	1.05

Hard + news / hard-only RMSE for the same model; ratio below one = text helps.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Walk, step 4: all articles, full FRED-MD (no outlier screen)

Outcome	Benchmark	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.03	1.02	1.03	1.00	1.00	1.00	0.99	0.99	1.00
	RW	0.94	0.93	0.94	0.98	0.98	0.98	0.99	0.99	0.99
	SPF	1.31**	1.30**	1.31**	1.17*	1.17*	1.17*	1.10*	1.10	1.11*
Consumption	AR	0.89*	0.90*	0.92*	0.94	0.94	0.93	0.89	0.89	0.89
	RW	0.86**	0.87**	0.88**	0.91*	0.90*	0.90*	0.92	0.91	0.92
	SPF	1.22***	1.23**	1.25**	1.16*	1.15*	1.14*	1.11*	1.10*	1.11*
Investment	AR	1.20***	1.21***	1.22***	1.08	1.13*	1.14**	1.21***	1.17**	1.12*
	RW	1.03	1.05	1.05	1.04	1.10*	1.10	1.22***	1.19**	1.13*
	SPF	1.57**	1.59***	1.60***	1.33*	1.40**	1.41***	1.42***	1.37***	1.31***

Best news-only model / benchmark RMSE.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

[◀ Back](#)

Walk, step 4: all articles, full FRED-MD (no outlier screen) (cont.)

Outcome	Model	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	LASSO	1.09*	1.03	0.97	1.02	1.01	1.01	1.04	1.02	1.04
	PCA-OLS	0.98	0.98	1.02	0.98	0.97	0.95	0.99	0.98	0.96
	Random forest	1.01	1.00	1.02	1.02	1.03	1.02	1.03	1.00	1.01
Consumption	LASSO	0.80†	0.25†	0.80†	0.91*	0.97	0.98	0.96	1.04	0.99
	PCA-OLS	0.99	0.98	0.96	0.97*	0.96*	0.92*	0.96	0.95	0.90
	Random forest	1.00	1.00	0.98	1.02	1.00	0.98	0.97	0.96	0.98
Investment	LASSO	1.04	0.94	0.99	0.97	1.23*	1.02	1.05	1.18	1.02
	PCA-OLS	0.97*	0.95*	1.02	0.98	0.96	0.90	1.02	1.01	0.97
	Random forest	1.07***	1.07***	1.08***	1.07**	1.07**	1.05**	1.05	1.08**	1.06*

Hard + news / hard-only RMSE for the same model; ratio below one = text helps.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%. † hard-only forecast for 2008Q2 is outlier-driven: excluded from summaries.

Walk, step 4: against the benchmark, by period

Outcome	Best model	Full sample (69 q.)	Without 2008Q3–2009Q2 (65 q.)	2008–2009 (8 q.)
GDP	News only	1.03 / 1.00 / 0.99	1.08 / 1.03 / 1.03	1.04 / 0.96 / 0.96
	Hard data only	0.99 / 0.99 / 1.00	0.99 / 1.01 / 1.02	1.02 / 0.98 / 0.99
	Hard data + news	1.00 / 1.01 / 1.00	1.02 / 1.05 / 1.06	1.03 / 0.98 / 0.97
Consumption	News only	0.90 / 0.94 / 0.89	0.95 / 0.93 / 0.92	0.83 / 0.94 / 0.81
	Hard data only	0.92 / 0.94 / 0.95	0.95 / 0.95 / 0.98	0.87 / 0.92 / 0.90
	Hard data + news	0.91 / 0.94 / 0.92	0.95 / 0.96 / 0.94	0.85 / 0.89 / 0.86
Investment	News only	1.21 / 1.12 / 1.17	1.35 / 1.23 / 1.39	1.12 / 1.02 / 1.02
	Hard data only	0.94 / 0.98 / 1.01	0.97 / 1.07 / 1.04	0.94 / 0.96 / 1.02
	Hard data + news	0.99 / 1.05 / 1.07	1.04 / 1.19 / 1.19	0.97 / 0.97 / 1.01

Relative RMSE against the AR benchmark, mean over M1–M3, H0 / H1 / H2. Best model chosen ex post on the full sample and not re-chosen. Samples are defined on the quarter being forecast.

◀ Back

Two papers, four models: best news-only model vs. benchmarks

Outcome	Benchmark	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.00	0.98	0.95	0.99	0.98	0.97	0.98	0.97	0.97	1.01	1.01	1.02
	RW	0.92	0.90	0.87	0.97	0.96	0.95	0.98	0.97	0.96	1.01	1.01	1.02
	SPF	1.18**	1.16**	1.11*	1.07*	1.06	1.05	1.03	1.02	1.01	1.01	1.01	1.02
Consumption	AR	0.94	0.91	0.89*	0.96	0.94	0.96	0.95	0.94	0.93	0.96	1.01	1.02
	RW	0.91*	0.88*	0.86**	0.92	0.90	0.91	0.98	0.96	0.96	0.96	1.01	1.02
	SPF	1.14**	1.09*	1.07	1.08*	1.05	1.07	1.08*	1.06	1.06	1.02	1.07	1.08**
Investment	AR	1.04	1.03	1.01	0.98	0.96	0.95	1.02	1.02	0.96	1.06	1.04	1.06
	RW	0.88*	0.87*	0.85	0.95	0.92	0.92	1.03	1.03	0.96	1.08	1.06	1.09
	SPF	1.36***	1.35***	1.32***	1.21***	1.18**	1.18**	1.15**	1.15**	1.08	1.13	1.11	1.14*

New York Times and Wall Street Journal; training from 1990Q1, evaluation 1996Q1–2020Q1 (97 quarters).

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Abstract

Complete this form to determine: Complete and submit on the 1st of each month.

[Back](#)

Two papers, paper's three models: two recessions

Best model	2001 recession (4 q.)		2008–2009 (8 q.)	
	GDP	Consumption	GDP	Consumption
News only	1.23 / 1.04 / 1.10	1.17 / 1.12 / 1.26	0.99 / 0.92 / 0.90	0.76 / 0.83 / 0.82
Hard data only	0.86 / 1.00 / 1.06	1.05 / 0.98 / 1.14	0.90 / 0.94 / 0.87	0.76 / 0.91 / 0.83
Hard data + news	1.11 / 1.01 / 1.08	1.08 / 1.05 / 1.21	0.95 / 0.92 / 0.91	0.71 / 0.85 / 0.82

Relative RMSE against the AR benchmark, mean over M1–M3, H0 / H1 / H2. Best model chosen ex post on the full sample and not re-chosen. Samples are defined on the quarter being forecast. 2001 recession = 2001Q1–2001Q4.

◀ Back

Earlier window: evaluation 1996–2020, 17 hard-data series

Five of the seven papers have no articles before 1997; the main specification starts in 1997.

Intermediate Results

Full-scale real-time U-MIDAS horse race, 1990–2020, seven print newspapers:

- Hard-data nowcasts improve from M1 to M3 as monthly releases arrive (best hard model vs. AR: 1.01 $\rightarrow$ 0.91 for GDP); news-based nowcasts barely move — the paper's mechanism, reproduced.
- In a real-time optimal combination, news carries about 60% of the weight for GDP nowcasts at M1, falling as hard data arrives; the RMSE gain from admitting news is at most 1%.
- Restricting to economic articles (6–10% of the corpus) removes most of the 2010s deterioration and the LASSO overfitting but does not change the level: hard data still wins at H0.
- Outlet lean $\times$ party control (564k-article sample): pooled over outcomes no effect (placebo $p \geq 0.25$); a consumption-only effect under Democratic presidents ($p = 0.02$) comes from the Clinton years and has the opposite sign for the House. [► Results](#)
- An economics-heavy sample and keyword + topic filters leave the picture unchanged. [► Sample](#)
- Cross-publisher disagreement measures (raw and weighted std. dev.) tested against consensus topic intensity — disagreement $\approx$ consensus, no cell where disagreement clearly wins [► Explainer](#)
- BERT sentiment scoring completed across 4 pretrained models (FinBERT, FinBERT-tone, RoBERTa-sentiment, DistilBERT-SST2) [► Appendix](#)

Replication on the newspaper corpus: news-only models vs. benchmarks

Outcome	Benchmark	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.04	1.03	1.05	1.03	1.03	1.02	1.02	1.01	1.00
	RW	0.96	0.95	0.96	1.01	1.01	1.00	1.02	1.01	1.00
	SPF	1.22**	1.21**	1.23**	1.12	1.11	1.11	1.06	1.05	1.05
Consumption	AR	0.95	0.95	0.96	1.01	1.01	1.00	1.00	0.99	0.97
	RW	0.92	0.92	0.93	0.96	0.97	0.95	1.03	1.02	1.00
	SPF	1.14*	1.15*	1.16**	1.13*	1.14*	1.12*	1.13*	1.12*	1.10
Investment	AR	1.18***	1.16***	1.10*	1.10	1.10	1.08	1.08	1.06	1.04
	RW	1.00	0.98	0.93	1.07	1.06	1.05	1.09	1.07	1.05
	SPF	1.54***	1.51***	1.44***	1.37***	1.36**	1.34**	1.22**	1.20**	1.18**

Best news-only model / benchmark RMSE; values < 1: lower RMSE than the benchmark. Ellingsen spec (LASSO, RF, PCA-1; topic consensus + disagreement). Real-time training on first releases, final-vintage actuals, 1996Q1–2020Q1, seven print papers. H0–H2: 0–2 quarters ahead; M1–M3: month of the forecast. DM stars (NW) */**/** = 10/5/1%. [▶ Hard-only / hard+news](#)

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

My approach, same layout (H0–H3)

Outcome	Benchmark	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	0.99	1.01	0.97	1.01	1.01	1.01	1.02	1.03	1.05	1.04	1.04	1.05
	RW	0.91	0.93	0.89	0.99	0.99	0.99	1.02	1.02	1.05	1.04	1.04	1.05
	SPF	1.17**	1.19**	1.14*	1.10*	1.09*	1.10**	1.07	1.07	1.10*	1.04	1.04	1.06
Consumption	AR	0.94	0.94	0.91	1.01	1.01	1.00	0.98	0.98	0.99	1.04	1.04	1.04
	RW	0.91	0.91*	0.88*	0.97	0.96	0.95	1.01	1.01	1.02	1.04	1.04	1.03
	SPF	1.13*	1.14*	1.09	1.14**	1.13**	1.12*	1.11	1.11*	1.12*	1.10	1.10	1.09
Investment	AR	1.08	1.13**	1.14**	1.10	1.08	1.08	1.06	1.08	1.10	1.13	1.10	1.08
	RW	0.91	0.95	0.97	1.06	1.04	1.05	1.07	1.09	1.11	1.16**	1.13*	1.11
	SPF	1.41***	1.47***	1.49***	1.36***	1.34***	1.34***	1.20**	1.22**	1.24*	1.22***	1.18**	1.16**

Best news-only model / benchmark RMSE; values < 1: lower RMSE than the benchmark.

Models: Elastic Net, RF, PCA-5, XGBoost; text families: topics + EPU + Loughran–McDonald + FinBERT. Real-time training on first releases, final-vintage actuals, 1996Q1–2020Q1, seven print papers; H3 scored on 95 origins. DM stars (NW) */**/** = 10/5/1%. Benchmarks: AR = autoregressive, RW = random walk, SPF = Survey of Professional Forecasters.

► [Hard-only](#) / [hard+news](#)

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Data vs. news: same model, news-only vs. hard-only

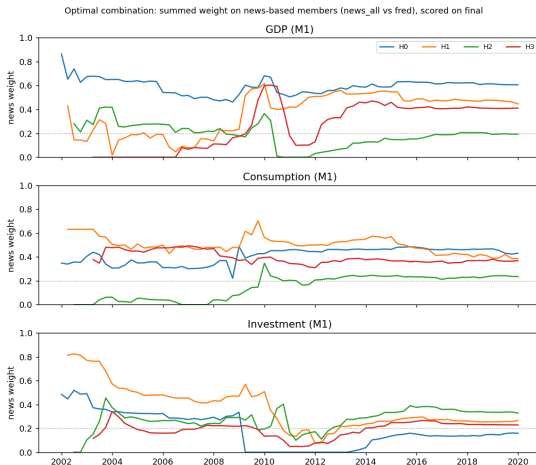
Outcome	Model	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	LASSO	1.05	1.16**	1.18**	1.55	1.98	1.59	1.56	2.10	1.71
	PCA-OLS	0.99	0.99	1.13	1.00	1.00	1.06	1.00	1.01	1.02
	Random forest	1.03	1.04	1.09**	1.05	1.04	1.04	1.06	1.04	1.05**
	<i>Best vs. best</i>	1.03	1.14**	1.15**	1.02	1.03	1.06	1.03	1.01	1.02
Consumption	LASSO	1.07*	1.15**	1.22**	1.01	1.07*	1.02	1.02	1.01	1.03
	PCA-OLS	1.03	1.06*	1.13**	1.03	1.05*	1.13*	1.05	1.03	1.05
	Random forest	1.02	1.07*	1.08*	1.02	1.04	1.02	1.04	1.04	1.02
	<i>Best vs. best</i>	1.02	1.07*	1.12	1.02	1.04	1.06	1.04	1.04	1.02
Investment	LASSO	1.26***	1.12	1.12	1.51	1.29	1.43	1.65	1.38	2.25
	PCA-OLS	0.92**	0.90**	1.23	0.97	0.92**	1.10	1.01	1.00	1.05
	Random forest	1.31***	1.31***	1.34***	1.13**	1.17***	1.24***	1.06	1.06	1.12*
	<i>Best vs. best</i>	1.31***	1.31***	1.28***	1.13**	1.15**	1.24**	1.05	1.01	1.06

News-only / hard-only RMSE for the same model (Ellingsen spec, cf. paper's Fig. 4). Values < 1: lower RMSE with text than with the 17 FRED series. "Best vs. best" = ex post best news model vs. ex post best hard model. Final-vintage actuals, 1996Q1–2020Q1. DM stars (NW) */**/** =

10/5/1%. [► My approach](#)

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%. 

Forecast combination: weight on news (ELT Sec. 4.2)



Recursive optimal combination, My approach, M1, final-vintage actuals. Dotted line: 0.2.

► Ellingsen spec

Forecast combination: news weight and RMSE gain

Specification	Outcome	News weight at M1 (M3)				RMSE: all members / hard only			
		H0	H1	H2	H3	H0	H1	H2	H3
Ellingsen spec (6 members)	GDP	0.69 (0.10)	0.17 (0.26)	0.06 (0.35)	–	1.00	1.01	1.01	–
	Consumption	0.40 (0.11)	0.70 (0.35)	0.29 (0.48)	–	1.01	1.01	1.03	–
	Investment	0.00 (0.12)	0.37 (0.00)	0.61 (0.14)	–	1.01	1.01	1.02	–
My approach (8 members)	GDP	0.60 (0.31)	0.37 (0.55)	0.19 (0.04)	0.26 (0.10)	1.00	0.99	1.01	1.04
	Consumption	0.41 (0.07)	0.50 (0.26)	0.15 (0.02)	0.39 (0.39)	0.99	0.99	1.00	0.99
	Investment	0.18 (0.21)	0.37 (0.30)	0.27 (0.34)	0.19 (0.39)	0.99	1.05	1.00	1.05

Left: summed weight on news-based members at M1 (M3 in parentheses). Right: RMSE of the combination of all members / combination of hard-only members; values < 1: admitting news helps.

Is it the corpus? Economic articles only

Model	Full corpus			Economic articles only		
	H0	H1	H2	H0	H1	H2
LASSO	1.15	1.38	1.52	1.20	1.02	1.01
PCA-OLS	1.04	1.03	1.02	1.06	1.04	1.04
Random forest	1.14	1.08	1.05	1.12	1.06	1.01

News-only / hard-only RMSE by model and horizon, full corpus vs. economic articles only (keyword rule, 6–10% of articles). Values < 1 : lower RMSE with news than with hard data.

Is it the corpus? Economic articles only (cont.)

Outcome	H	Full corpus				Economic articles only			
		1996–2007	2008–09	2010–19	2020	1996–2007	2008–09	2010–19	2020
GDP	H0	0.99	1.08	1.08	0.98	1.01	1.12	1.08	0.96
	H1	1.02	1.01	1.25	0.98	1.01	1.01	1.03	1.01
Consumption	H0	1.08	1.07	1.12	0.95	1.08	1.15	1.09	0.97
	H1	1.00	1.05	1.13	1.02	1.00	1.05	1.03	1.03
Investment	H0	1.09	1.22	1.34	0.96	1.09	1.19	1.15	0.75
	H1	1.05	1.05	1.36	0.86	1.07	1.07	1.15	1.00

News / hard RMSE ratio (random forest, H0/H1) by sub-period, full vs. filtered corpus; first-release actuals.

Best hard-only model vs. benchmarks — Ellingsen spec

Outcome	Benchmark	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.01	0.91	0.91 [*]	1.01	1.00	0.97	0.99	0.99	0.98
	RW	0.93	0.84 [*]	0.84 ^{**}	0.99	0.98	0.95	0.99	0.99	0.98
	SPF	1.19 [*]	1.07	1.07	1.10	1.08	1.05	1.04	1.04	1.03
Consumption	AR	0.93 [*]	0.89 ^{**}	0.86 [*]	0.99	0.98	0.95	0.96	0.95	0.95
	RW	0.90 ^{**}	0.86 ^{***}	0.83 ^{**}	0.94	0.93	0.90 ^{**}	0.99	0.98	0.98
	SPF	1.12	1.07	1.03	1.11 [*]	1.09 [*]	1.06	1.09	1.08	1.08
Investment	AR	0.90 ^{**}	0.89 ^{**}	0.86 ^{**}	0.97	0.95	0.88 ^{**}	1.03	1.05	0.98
	RW	0.76 ^{***}	0.75 ^{***}	0.73 ^{***}	0.94	0.92 [*]	0.85 ^{**}	1.03	1.06	0.99
	SPF	1.18 ^{**}	1.16 ^{**}	1.13 [*]	1.21 [*]	1.18 [*]	1.09	1.16 [*]	1.19	1.11

Best hard-only model (17 FRED series) / benchmark RMSE (paper's Table SD1 analogue).

Final-vintage actuals, 1996Q1–2020Q1. DM stars (NW) */**/** = 10/5/1%.

[◀ Back](#)

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Best hard-only model vs. benchmarks — My approach

Outcome	Benchmark	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.00	0.93	0.91*	1.01	0.98	0.99	0.99	1.01	0.99	1.01	1.01	1.00
	RW	0.92	0.85*	0.83**	0.99	0.96	0.97	0.99	1.00	0.99	1.01	1.01	1.00
	SPF	1.17*	1.09	1.06	1.09	1.06	1.07	1.04	1.05	1.04	1.01	1.01	1.00
Consumption	AR	0.94	0.88*	0.86*	1.01	0.95	0.98	0.97	0.96	0.93	1.01	0.99	1.02
	RW	0.91**	0.85**	0.83**	0.96	0.91*	0.94	1.00	0.99	0.96	1.01	0.99	1.02
	SPF	1.13*	1.06	1.04	1.13**	1.07	1.10*	1.10	1.09	1.06	1.06	1.05	1.07
Investment	AR	0.86	0.82	0.86	0.93	0.93	0.92	0.99	0.97	0.96	1.01	1.00	0.99
	RW	0.73**	0.69**	0.73*	0.90	0.90*	0.89	1.00	0.98	0.96	1.04	1.03	1.02
	SPF	1.13	1.07	1.13	1.15**	1.15*	1.14**	1.12	1.10	1.08	1.09	1.08	1.07

Best hard-only model (17 FRED series) / benchmark RMSE. Final-vintage actuals, 1996Q1–2020Q1; H3 scored on 95 origins. DM stars (NW) */**/** = 10/5/1%. Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

[◀ Back](#)

Best hard+news model vs. benchmarks — Ellingsen spec

Outcome	Benchmark	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.02	1.02	0.96	1.03	1.03	1.02	1.02	1.01	1.00
	RW	0.94	0.94	0.89*	1.01	1.01	1.00	1.02	1.00	1.00
	SPF	1.20**	1.20**	1.13	1.12	1.11	1.11	1.06	1.05	1.05
Consumption	AR	0.93	0.93	0.93	1.00	1.01	1.00	0.97	0.97	0.97
	RW	0.90*	0.90*	0.90**	0.96	0.96	0.95	1.00	1.00	1.00
	SPF	1.12	1.12	1.12	1.12*	1.13*	1.12*	1.10	1.10	1.10
Investment	AR	1.01	1.00	0.98	1.08	1.07	0.99	1.08	1.06	1.04
	RW	0.85**	0.85**	0.83***	1.05	1.03	0.96	1.09	1.07	1.05
	SPF	1.32**	1.31***	1.28**	1.34***	1.32***	1.23**	1.22**	1.20**	1.18**

Best hard+news (“grand model”) / benchmark RMSE. Final-vintage actuals, 1996Q1–2020Q1.

DM stars (NW) */**/** = 10/5/1%. [◀ Back](#)

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Best hard+news model vs. benchmarks — My approach

Outcome	Benchmark	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	0.99	0.99	0.91*	1.01	1.00	1.01	1.02	1.03	1.05	1.04	1.04	1.01
	RW	0.91	0.91	0.84**	0.99	0.98	0.98	1.02	1.02	1.04	1.04	1.04	1.01
	SPF	1.17**	1.17**	1.07	1.09*	1.09*	1.09*	1.07	1.07	1.09*	1.04	1.04	1.01
Consumption	AR	0.93	0.93	0.89*	1.01	1.00	0.99	0.96	0.97	0.98	1.02	1.04	1.01
	RW	0.90*	0.90*	0.86**	0.96	0.96	0.95	0.99	0.99	1.01	1.02	1.03	1.01
	SPF	1.13*	1.13*	1.08	1.13**	1.13**	1.11*	1.09	1.10	1.11	1.08	1.09	1.07
Investment	AR	1.04	0.97	0.96	1.06	1.05	0.99	1.05	1.06	1.08	1.12	1.08	1.06
	RW	0.88	0.82**	0.81*	1.03	1.02	0.96	1.05	1.06	1.08	1.15*	1.11	1.09
	SPF	1.36***	1.27**	1.25***	1.32***	1.30***	1.23***	1.19**	1.20**	1.22*	1.21**	1.16**	1.14**

Best hard+news (“grand model”) / benchmark RMSE. Final-vintage actuals, 1996Q1–2020Q1;

H3 scored on 95 origins. DM stars (NW) */**/** = 10/5/1%. [◀ Back](#)

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Data vs. news — My approach

Outcome	Model	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	Elastic Net	1.02	1.13*	1.12*	1.16	1.87	1.38	1.85	2.11	1.48	1.04	1.07	1.13
	PCA-OLS	1.05	1.08	1.06	1.04	1.07	0.93	1.01	1.00	1.02	1.02	0.98	1.05
	Random forest	0.99	1.01	1.01	1.01	1.01	1.02	1.01	1.01	1.05*	1.03	1.02	1.04
	XGBoost	0.99	1.03	1.03	1.01	1.01	0.99	0.99	0.99	1.05	1.06	1.02	0.99
	<i>Best vs. best</i>	0.99	1.09	1.07	1.01	1.02	1.02	1.03	1.02	1.06	1.03	1.03	1.05
Consumption	Elastic Net	1.11	1.19**	1.21**	1.00	1.08*	1.06	1.06*	1.02	0.99	1.05	1.10	1.07
	PCA-OLS	0.98	1.02	1.01	1.00	1.13*	1.01	0.97	0.98	1.07	1.00	1.03	1.03
	Random forest	1.00	1.02	1.03	1.00	1.01	1.01	1.01	1.02	1.04	1.01	1.01	1.02
	XGBoost	0.98	1.09	1.02	1.03	1.05	1.02	0.95	0.99	1.01	1.04	1.05	0.96
	<i>Best vs. best</i>	0.99	1.07	1.05	1.00	1.06	1.02	1.01	1.02	1.06	1.03	1.05	1.02
Investment	Elastic Net	1.14*	1.07	1.24***	1.40	1.19	1.23	1.84	1.29	1.97	1.18	1.14	1.27
	PCA-OLS	1.44***	1.50***	1.45***	1.19**	1.19*	1.18	1.14	0.92	1.03	1.15**	1.01	1.13**
	Random forest	1.19***	1.25***	1.28***	1.17**	1.17**	1.18*	1.07	1.11*	1.15	1.12*	1.10**	1.09
	XGBoost	1.25**	1.18	1.30**	1.18	1.04	1.08	0.99	0.98	0.96	1.14*	0.94	1.08
	<i>Best vs. best</i>	1.25**	1.38***	1.33**	1.18*	1.17**	1.18*	1.07	1.11*	1.15	1.12*	1.10**	1.09

News-only / hard-only RMSE for the same model; values < 1: lower RMSE with text than with the 17 FRED series. “Best vs. best” = ex post best news model vs. ex post best hard model.

DM stars (NW) */**/** = 10/5/1%.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Data vs. data + news — Ellingsen spec, full table

Outcome	Model	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	LASSO	1.08*	1.13*	1.06	1.53	1.24	1.16	1.28	1.95	1.91
	PCA-OLS	0.99	0.99	1.13	1.00	1.00	1.06	1.00	1.01	1.02
	Random forest	1.01	1.02	1.05*	1.03	1.03	1.03	1.06	1.03	1.04**
Consumption	LASSO	1.08**	1.13*	1.14*	1.01	1.08**	1.02	0.98	0.99	1.03
	PCA-OLS	1.03	1.06*	1.13*	1.03	1.05*	1.13*	1.05	1.03	1.05
	Random forest	1.00	1.05	1.05	1.02	1.04	1.02	1.03	1.03	1.02
Investment	LASSO	1.06	0.97	1.07	1.59	1.63	1.29	2.01	1.29	2.30
	PCA-OLS	0.93**	0.90**	1.23	0.97	0.93**	1.10	1.01	1.00	1.05
	Random forest	1.13***	1.13***	1.13***	1.11**	1.12***	1.13***	1.05	1.04	1.08

Hard+news / hard-only RMSE for the same model; values < 1: adding news helps. DM stars (NW) */**/** = 10/5/1%.

Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

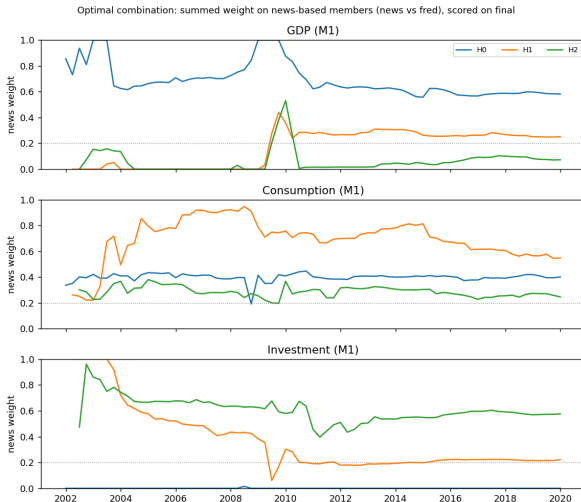
Data vs. data + news — My approach, full table

Outcome	Model	H0			H1			H2			H3		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	Elastic Net	1.03	1.09	1.06	1.26	1.74	1.30	1.49	1.66	1.77	1.04	1.06	1.01
	PCA-OLS	1.04	1.06	1.04	1.03	1.05	0.92	0.99	0.98	1.01	0.99	0.96	1.04
	Random forest	0.99	1.01	0.99	1.00	1.00	1.02	1.02	1.01	1.05*	1.03	1.02	1.03
	XGBoost	0.99	1.01	0.97	0.97	0.96	0.97	1.01	0.97	1.01	1.06	1.02	0.96
Consumption	Elastic Net	1.11	1.12	1.10	1.00	1.08**	1.04	1.06	0.99	1.01	1.06	1.13*	1.09
	PCA-OLS	0.98	1.01	1.00	0.99	1.12*	1.01	0.95	0.96	1.06	0.99	1.01	1.03
	Random forest	0.99	1.01	1.01	1.00	1.00	1.00	1.01	1.02	1.04	1.01	1.01	1.01
	XGBoost	0.98	1.07	1.01	1.02	1.03	1.02	0.95	0.96	0.99	1.03	1.06*	0.94
Investment	Elastic Net	1.20	0.92	1.60	1.11	1.08	1.04	1.89	1.22	2.07	1.17	1.10*	1.19
	PCA-OLS	1.42***	1.43***	1.37***	1.16*	1.14	1.13	1.10	0.90	1.00	1.11	0.97	1.11*
	Random forest	1.07**	1.08**	1.09**	1.14*	1.13**	1.08**	1.05	1.09*	1.12	1.11*	1.08**	1.07
	XGBoost	1.09	1.01	0.99	1.05	1.07	1.00	1.02	0.96	1.00	1.04	0.98	0.98

Hard+news / hard-only RMSE for the same model; values < 1: adding news helps. DM stars (NW) */**/** = 10/5/1%.

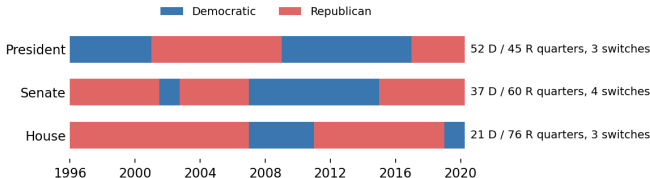
Stars: Diebold–Mariano test, Newey–West variance (lag h), small-sample correction; 10 / 5 / 1%.

Forecast combination: weight on news — Ellingsen spec



Recursive optimal combination, Ellingsen spec, M1, final-vintage actuals.

Extension: does outlet lean interact with party control?



- Is news from Democrat-leaning outlets more useful for forecasting when Democrats hold the presidency, the Senate or the House?
- Estimand: forecast loss with left-outlet news minus forecast loss with other-outlet news, Democratic minus Republican control. Permanent outlet differences and quarter effects cancel.
- Left: Boston Globe, LA Times, New York Times, Washington Post. Other: Chicago Tribune, USA Today, Wall Street Journal.
- 1996Q1–2020Q1 contains 3 presidential, 4 Senate and 3 House switches, so inference uses a placebo test: the control series shifted in time.

► Lean split Political exercise

Mac, 564k-article sample, run `us_only`. Political exercise; own inference (placebo, Driscoll–Kraay).

[◀ Back](#)

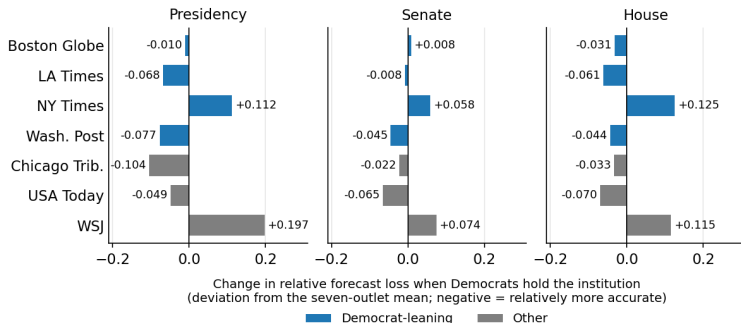
Lean $\times$ party control: pooled estimates (564k-article sample)

Design	Outcomes	Horizons	Democrats hold the ...		
			Presidency	Senate	House
Lean groups: left vs. other	All three	H0	−0.012 (0.72)	0.032 (0.42)	−0.007 (0.88)
	All three	H0–H2	−0.010 (0.78)	0.010 (0.73)	−0.022 (0.84)
	Consumption	H0	−0.102 (0.24)	0.104 (0.15)	0.122 (0.04)
	Consumption	H0–H2	−0.052 (0.09)	0.019 (0.66)	0.067 (0.05)
Outlet panel: seven outlets	All three	H0	−0.022 (0.26)	0.047 (0.25)	0.042 (0.57)
	All three	H0–H2	−0.025 (0.33)	0.008 (0.84)	−0.006 (0.87)
	Consumption	H0	−0.106 (0.02)	0.096 (0.28)	0.169 (0.04)
	Consumption	H0–H2	−0.074 (0.04)	0.044 (0.51)	0.117 (0.17)

- Pooled over GDP, consumption and investment, the gap does not move with party control: estimates between -0.03 and $+0.05$ of mean loss, placebo p-values 0.25–0.88.
- Consumption alone: left outlets do relatively better under Democratic presidents (seven-outlet panel -0.11 , $p = 0.02$) and relatively worse under a Democratic House ($+0.17$, $p = 0.04$).

β = change in the left-minus-other relative squared-error loss when Democrats hold the institution, as a share of mean loss; negative = Democrat-leaning outlets relatively more accurate; placebo p-values in parentheses (control series shifted in time). Equal-weight ensemble of LASSO, RF, PCA(1); pooled over M1–M3; final-vintage actuals, 1996Q1–2020Q1. Mac. 564k-article sample. run us only. Political exercise; own inference (placebo, Driscoll–Kraay).

Not a lean pattern: the estimate by outlet (564k-article sample)



- The change under a Democratic president is unfavourable for the New York Times (+0.11) and the Wall Street Journal (+0.20) — one outlet from each group — and favourable for the other five (−0.01 to −0.10).
- The all-outcome presidency estimate runs from −0.09 to +0.07 as single outlets are dropped, and from −0.18 (Ad Fontes score, $p = 0.01$) to +0.02 (endorsement record, $p = 0.42$) across lean measures.

► Robustness ► By president

Mac, 564k-article sample, run `us_only`. Political exercise; own inference (placebo, Driscoll–Kraay).

Political exercise: the lean split (sample)

Group	Outlet	AllSides (news pages)	Ad Fontes bias	Endorsements 1992–2016	G–M ADA score
Democrat-leaning	Boston Globe	Lean Left	−7.80	DDDDDDDD	–
	Los Angeles Times	Lean Left	−6.21	---- DDD	70.0
	New York Times	Lean Left	−7.87	DDDDDDDD	73.7
	Washington Post	Lean Left	−7.01	DDDDDDDD	66.6
Other	Chicago Tribune	Center	−4.38	RRRRDDL	–
	USA Today	Center	−3.94	-----	63.4
	Wall Street Journal	Center	4.32	-----	85.1

AllSides = rating of the news pages that held during 1990–2020; Ad Fontes values re-read on 2026-10-01, the estimates use an earlier snapshot that differs by at most 0.41; endorsements are the seven presidential elections 1992–2016 (D, R, L = Libertarian, - = none); ADA = Groseclose–Milyo (2005) score of the news pages, not available for two papers. Mac, 564k-article sample, run `us_only`. Political exercise; own inference.

Political exercise: who held what (sample)

Institution	Quarters under		Democratic spells	Switches
	Democrats	Republicans		
President	52	45	1996Q1–2000Q4, 2009Q1–2016Q4	3
Senate	37	60	2001Q3–2002Q3, 2007Q1–2014Q4	4
House	21	76	2007Q1–2010Q4, 2019Q1–2020Q1	3

Evaluation window 1996Q1–2020Q1 (97 quarters). A quarter is assigned to the party in control for most of its days; the regime is dated at the forecast-origin quarter.
Mac, 564k-article sample, run `us_only`. Political exercise; own inference.

Lean × party control by outcome: lean groups (sample)

Outcomes	Horizons	Presidency			Senate			House		
		β	p_{DK}	p_{plac}	β	p_{DK}	p_{plac}	β	p_{DK}	p_{plac}
All three	H0	-0.012	0.76	0.72	0.032	0.51	0.42	-0.007	0.93	0.88
All three	H0-H2	-0.010	0.75	0.78	0.010	0.80	0.73	-0.022	0.69	0.84
Consumption	H0	-0.102	0.33	0.24	0.104	0.26	0.15	0.122	0.16	0.04
Consumption	H0-H2	-0.052	0.30	0.09	0.019	0.67	0.66	0.067	0.13	0.05
GDP	H0	0.056	0.19	0.15	-0.084	0.12	0.21	-0.111	0.11	0.12
Investment	H0	0.009	0.91	0.90	0.076	0.49	0.32	-0.032	0.86	0.88

p-DK = Driscoll-Kraay; p-plac = placebo. 564k-article sample.
 Mac, 564k-article sample, run us_only. Political exercise; own inference.

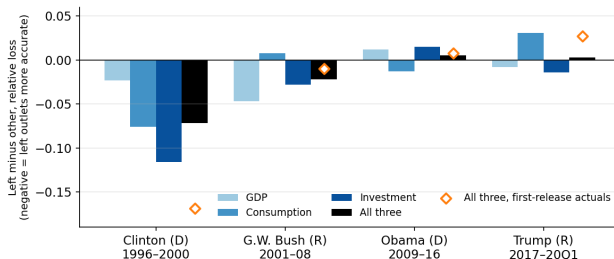
Robustness: lean measure and leave-one-outlet-out (sample)

Lean measure	Presidency	Senate	House
Binary split (main)	−0.025 (0.33)	0.008 (0.84)	−0.006 (0.87)
Ad Fontes bias score	−0.184 (0.01)	−0.055 (0.62)	−0.091 (0.67)
Net Democratic endorsements, 1992–2016	0.017 (0.42)	0.017 (0.60)	0.022 (0.67)

Democrats hold the	All seven outlets	Lowest (outlet dropped)	Highest (outlet dropped)
Presidency	−0.025	−0.085 (Chicago Tribune)	0.067 (Wall Street Journal)
Senate	0.008	−0.022 (USA Today)	0.048 (Wall Street Journal)
House	−0.006	−0.047 (New York Times)	0.048 (Wall Street Journal)

Outlet panel, all outcomes, H0–H2, equal-weight ensemble; placebo p-values in parentheses. 564k-article sample. Mac, 564k-article sample, run `us_only`. Political exercise; own inference.

By president (sample)



President	Final-vintage actuals				First release
	GDP	Consumption	Investment	All three	All three
Clinton (D), 1996-2000	-0.023	-0.076	-0.116	-0.072	-0.169
G.W. Bush (R), 2001-08	-0.047	0.008	-0.028	-0.022	-0.010
Obama (D), 2009-16	0.012	-0.013	0.015	0.005	0.008
Trump (R), 2017-20Q1	-0.008	0.031	-0.014	0.003	0.027

The left-minus-other gap is -0.07 under Clinton and between -0.02 and $+0.01$ under Bush, Obama and Trump; in the sample, Washington Post, USA Today and Chicago Tribune contribute almost nothing before 1997.

Mac, 564k-article sample, run us_only. Political exercise; own inference.

Presidency, Senate and House in one regression (sample)

Design	Outcomes	Horizons	Presidency	Senate	House
Lean groups	All three	H0	−0.026 (0.58)	0.049 (0.09)	−0.036 (0.76)
Lean groups	All three	H0–H2	−0.019 (0.76)	0.028 (0.35)	−0.040 (0.67)
Outlet panel, binary split	All three	H0	−0.029 (0.30)	0.048 (0.15)	0.013 (0.90)
Outlet panel, binary split	All three	H0–H2	−0.032 (0.32)	0.021 (0.52)	−0.023 (0.73)
Outlet panel, binary split	Consumption	H0	−0.104 (0.07)	0.076 (0.28)	0.111 (0.04)
Outlet panel, Ad Fontes score	All three	H0–H2	−0.207 (0.04)	0.032 (0.61)	−0.146 (0.55)
Outlet panel, endorsements	All three	H0–H2	0.019 (0.51)	0.005 (0.85)	0.023 (0.67)

The three control series overlap, so the separate estimates do not hold the other institutions fixed; this one does. Placebo

p-values in parentheses. 564k-article sample.

Mac, 564k-article sample, run `us_only`. Political exercise; own inference.

Which actuals: final vintage vs. first release (sample)

Estimate	Final-vintage actuals	First-release actuals
Outlet panel, all outcomes, H0, presidency	-0.022 (0.26)	-0.074 (0.052)
Outlet panel, all outcomes, H0-H2, presidency	-0.025 (0.33)	-0.067 (0.061)
Outlet panel, consumption, H0, House	0.169 (0.04)	0.089 (0.39)
Outlet panel (Ad Fontes), all outcomes, H0, presidency	-0.119 (0.10)	-0.196 (0.01)

The four estimates whose reading changes between final-vintage and first-release actuals; placebo p-values in parentheses.

564k-article sample.

Mac, 564k-article sample, run `us_only`. Political exercise; own inference.

Coverage: do topic shares move with lean $\times$ control? (sample)

Democrats hold the	Outlet + month effects	+ outlet-specific trends
Presidency	12	6
Senate	20	11
House	6	9
Expected by chance	about 8 of 80	

Of 80 topics, number with placebo $p < 0.10$; no sentiment or uncertainty measures exist for the sample. Mac, 564k-article sample, run `us_only`. Political exercise; own inference.

Not a lean pattern: the estimate by outlet, table (sample)

Group	Outlet	Presidency	Senate	House
Democrat-leaning	Boston Globe	-0.010	0.008	-0.031
	Los Angeles Times	-0.068	-0.008	-0.061
	New York Times	0.112	0.058	0.125
	Washington Post	-0.077	-0.045	-0.044
Other	Chicago Tribune	-0.104	-0.022	-0.033
	USA Today	-0.049	-0.065	-0.070
	Wall Street Journal	0.197	0.074	0.115

Per-outlet change in relative loss under Democratic control (deviation from the cross-outlet mean); negative = that outlet's news relatively better. Table behind the outlet figure. 564k-article sample.
Mac, 564k-article sample, run `us_only`. Political exercise; own inference.

Is it the corpus? An economics-heavy sample gives the same Table 1

- The same chain on a 564k-article sample of the seven papers, with its own topic model, reproduces the full-corpus Table 1: consumption nowcasts 0.93–0.95 vs. AR (full corpus 0.95–0.96), GDP 1.04–1.05 (1.03–1.05), investment 1.10–1.13 (1.10–1.18).
- The sample is economics-heavy: 39% of its articles pass the economic keyword rule (27–50% across the seven papers), against 6–10% in the full corpus. More economic content, same result.

Outcome	Benchmark	Corpus	H0			H1			H2		
			M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	Full corpus	1.04	1.03	1.05	1.03	1.03	1.02	1.02	1.01	1.00
		Sample	1.04	1.04	1.05**	1.01	1.03	1.02	1.04	1.03	1.02
	RW	Full corpus	0.96	0.95	0.96	1.01	1.01	1.00	1.02	1.01	1.00
		Sample	0.96	0.96	0.97	0.99	1.00	1.00	1.03	1.03	1.02
Consumption	AR	Full corpus	0.95	0.95	0.96	1.01	1.01	1.00	1.00	0.99	0.97
		Sample	0.93	0.94	0.95	0.96	0.97	0.98	0.95	0.94	0.95
	RW	Full corpus	0.92	0.92	0.93	0.96	0.97	0.95	1.03	1.02	1.00
		Sample	0.90*	0.91**	0.92**	0.91**	0.93*	0.93*	0.98	0.97	0.98
Investment	AR	Full corpus	1.18***	1.16***	1.10*	1.10*	1.10*	1.08	1.08	1.06	1.04
		Sample	1.12**	1.10*	1.13**	1.05	1.07	1.07	1.05	1.05	1.06
	RW	Full corpus	1.00	0.98	0.93	1.07	1.06	1.05	1.09*	1.07	1.05
		Sample	0.95	0.93	0.96	1.02	1.03	1.03	1.05	1.06	1.07

Full corpus = 11.76M articles; sample = 564k articles of the same seven papers, own 80-topic model. Best news-only model, relative RMSE. Real-time training on first releases, final-vintage actuals, 1996Q1–2020Q1. DM stars */**/** = 10/5/1%.

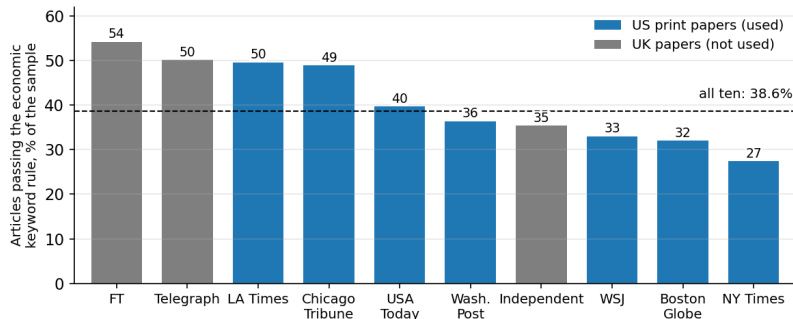
Significance here uses the earlier test, which overstates it beyond the nowcast. Mac, 564k-article sample, run us_only. Full-corpus rows:

The 564k-article sample

	Newspaper	Articles	Pass the economic keyword rule
Seven US print papers	New York Times	110,177	27.4%
	Wall Street Journal	98,640	33.0%
	Washington Post	61,791	36.4%
	Boston Globe	40,637	32.0%
	Los Angeles Times	32,467	49.6%
	Chicago Tribune	26,207	48.9%
	USA Today	25,469	39.7%
UK papers (not used)	Financial Times	98,085	54.2%
	The Independent	66,281	35.4%
	Daily Telegraph	18,909	50.1%
All ten		578,663	38.6%

Sample = the Mac corpus of the same newspapers; the seven US print papers are used (564k articles with topic scores from the sample's own 80-topic model). Keyword rule: ≥ 2 economic term groups in the lead or ≥ 1 in the headline. Mac, 564k-article sample.

The sample: share passing the economic keyword rule



564k-article sample.
Mac, 564k-article sample.

Sample: Table 1 (news-only vs. benchmarks)

Outcome	Benchmark	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	1.04	1.04	1.05**	1.01	1.03	1.02	1.04	1.03	1.02
	RW	0.96	0.96	0.97	0.99	1.00	1.00	1.03	1.03	1.02
	SPF	1.22**	1.22**	1.24**	1.10*	1.11**	1.10*	1.08	1.08	1.07**
Consumption	AR	0.93	0.94	0.95	0.96	0.97	0.98	0.95	0.94	0.95
	RW	0.90*	0.91**	0.92**	0.91**	0.93*	0.93*	0.98	0.97	0.98
	SPF	1.12*	1.13*	1.15*	1.07	1.09*	1.10*	1.08	1.07	1.08
Investment	AR	1.12**	1.10*	1.13**	1.05	1.07	1.07	1.05	1.05	1.06
	RW	0.95	0.93	0.96	1.02	1.03	1.03	1.05	1.06	1.07
	SPF	1.47***	1.43***	1.48***	1.31***	1.32***	1.32***	1.18**	1.19***	1.20***

Best news-only model / benchmark RMSE, 564k-article sample, same layout as the full-corpus Table 1. Final-vintage actuals,

1996Q1–2020Q1. DM stars */**/** = 10/5/1%.

Significance here uses the earlier test, which overstates it beyond the nowcast. Mac, 564k-article sample, run us_only.

Sample: hard data and hard + news vs. AR

Outcome	Predictors	H0			H1			H2		
		M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	Hard data only	1.01	0.91	0.90*	1.01	1.00	0.97	0.99	0.99	0.98
	Hard data + news	1.03	0.98	0.90	1.01	1.01	1.01	1.03	1.02	1.02
Consumption	Hard data only	0.93*	0.89**	0.86*	0.99	0.97	0.95*	0.96	0.95	0.95
	Hard data + news	0.93	0.91	0.92*	0.95	0.97	0.97	0.95	0.94*	0.95
Investment	Hard data only	0.90**	0.89**	0.86***	0.97	0.95	0.88**	1.03	1.05	0.98
	Hard data + news	0.99	0.98	0.95	1.03	1.04	0.95	1.08	1.07	1.05

Best hard-only and best hard+news model / AR RMSE, 564k-article sample. Final-vintage actuals, 1996Q1–2020Q1. DM stars

*/**/** = 10/5/1%.

Significance here uses the earlier test, which overstates it beyond the nowcast. Mac, 564k-article sample, run us_only.

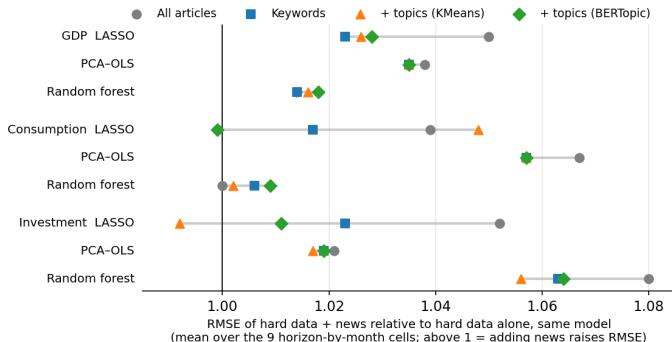
Economic filter on the sample: keywords, then a topic model

Articles used	Articles	Share of sample	Share of keyword-flagged	Thin outlet-months
All articles	578,663	100%	—	6.5%
Keywords	223,332	38.6%	100%	36.6%
Keywords + topic model (KMeans)	213,614	36.9%	95.6%	39.6%
Keywords + topic model (BERTopic)	218,500	37.8%	97.8%	37.3%

- A topic model after the keyword rule removes almost nothing: KMeans keeps 95.6% and BERTopic 97.8% of keyword-flagged articles, and the two agree on 94.5%.
- Filtered panels are thin on a sample this size: 37–40% of outlet-months in the evaluation window have fewer than 30 articles (6.5% unfiltered).

Thin outlet-months = share of outlet-months of the seven papers in 1996–2020Q1 with fewer than 30 kept articles; topic stage uses automatic cluster labels. 564k-article sample.
Mac, 564k-article sample.

Does filtering change the increment? (sample)



Significance here uses the earlier test, which overstates it beyond the nowcast. Mac, 564k-article sample, runs `us_only`, `us_only_keywords`, `us_only_topics`, `us_only_topics_bertopic`.

Filter on the sample: news-only Table 1 by article set

Outcome	Benchmark	Articles used	H0			H1			H2		
			M1	M2	M3	M1	M2	M3	M1	M2	M3
GDP	AR	All articles	1.04	1.04	1.05**	1.01	1.03	1.02	1.04	1.03	1.02
		Keywords	1.02	1.02	1.05*	1.03	1.00	1.03	0.98	1.02	0.99
		Keywords + topics (KMeans)	1.03	1.03	1.06**	1.04	1.03	1.04	1.01	1.02	1.01
		Keywords + topics (BERTopic)	1.03	1.03	1.05*	1.04	1.01	1.03	1.02	1.02	1.01
Consumption	AR	All articles	0.93	0.94	0.95	0.96	0.97	0.98	0.95	0.94	0.95
		Keywords	0.93	0.94	0.95	0.98	0.99	0.99	0.93	0.95	0.97
		Keywords + topics (KMeans)	0.92*	0.93	0.94*	0.97	0.98	0.98	0.94	0.93*	0.95*
		Keywords + topics (BERTopic)	0.90	0.94	0.90	0.98	0.97	0.98	0.94	0.95	0.97
Consumption	RW	All articles	0.90*	0.91**	0.92**	0.91**	0.93*	0.93*	0.98	0.97	0.98
		Keywords + topics (BERTopic)	0.87**	0.91**	0.87*	0.93*	0.93*	0.93	0.97	0.98	1.00
Investment	AR	All articles	1.12**	1.10*	1.13**	1.05	1.07	1.07	1.05	1.05	1.06
		Keywords	1.10**	1.10**	1.13**	1.06	1.03	1.04	1.06	1.06	1.02
		Keywords + topics (KMeans)	1.10*	1.11**	1.12**	1.03	1.06	1.07	1.03	1.03	0.99
		Keywords + topics (BERTopic)	1.12**	1.08*	1.12**	1.06	1.01	1.03	1.05	1.06	1.03

Best news-only model / AR RMSE (and consumption vs. RW) by article set, H×M. Final-vintage actuals, 1996Q1–2020Q1.

DM stars */**/** = 10/5/1%. 564k-article sample.

Significance here uses the earlier test, which overstates it beyond the nowcast. Mac, 564k-article sample, runs `us_only`, `us_only_keywords`, `us_only_topics`, `us_only_topics_bertopic`.

Elligsen Replication Status

- All figures and the main relative-RMSE table [▶ Details](#)
- Topic Interpretability. Topics aren't statistical noise — they're recognizable economic themes.
- Predictor sparsity: LASSO selects only a handful of predictors (out of $\simeq 230$ candidate predictors—news topics + hard data together) at any given estimation date, and selections are short-lived.
- Influential-topic patterns: factor importance splits meaningfully between news and hard data rather than news being ignored, and the influential news topics are economically sensible ones.
- The paper's SPF/AR comparison pattern (SPF beats news short-horizon, converges by H2) [◀ Back](#)

Mac sample, older quarterly design, earlier targets (pre-2026-09-29).

Plain Models vs Shock-Based Models

Why plain VAR, ARIMA and MIDAS?

- Current solution: use plain-vanilla model specification
- Using plain models allows for compatibility of the results
- Forecasting literature uses shock-specific models (Schorfheide, Song, 2021)
- Text analysis literature uses shock-specific equations (Bianchi, Ludvigson, Ma, 2024)

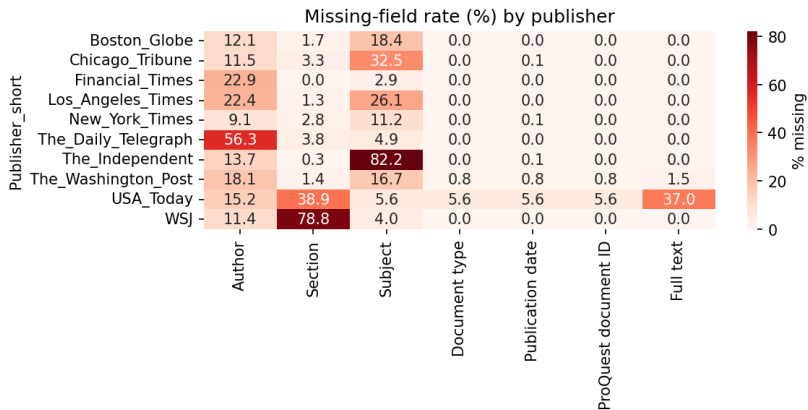
Methods Used — Explainers

- **MIDAS** (Mixed Data Sampling): stacks monthly predictors onto a quarterly target without frequency aggregation [▶ MIDAS](#)
- **Elastic Net**: penalized linear regression ($L_1 + L_2$) built for many correlated predictors [▶ Elastic Net](#)
- **Random Forest (RF)**: ensemble of regression trees, captures nonlinearities without a parametric form [▶ RF](#)
- **XGBoost**: gradient-boosted trees, builds an ensemble sequentially by correcting prior errors [▶ XGBoost](#)
- **PCA** (Principal Component Analysis): compresses many predictors into a handful of common factors before OLS [▶ PCA](#)
- **BERT**: bidirectional Transformer language model, used here for sentiment scoring [▶ BERT](#) [◀ Back](#)

Exploratory Data Analysis — News Corpus

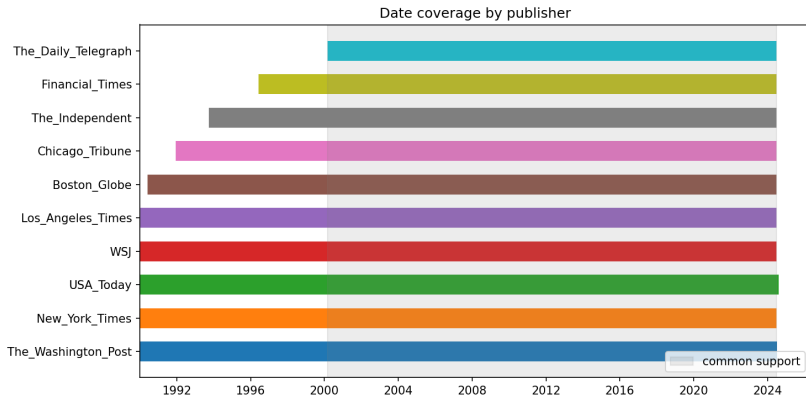
Coverage, text-length, and data-quality diagnostics across the 10-publisher,
1990–2024 corpus

EDA — Missing-Field Rates



% missing rate for key metadata fields, by publisher.

EDA — Date Coverage by Publisher



First/last date and coverage span per publisher (Gantt view).

Base Paper

- News media versus FRED-MD for macroeconomic forecasting by Ellingsen et al (2022). Forecast US GDP, consumption and investment growth using both econometrics and ML techniques
- Econometrics - OLS, MIDAS; ML techniques - LASSO, RF, PCA on the predictor set
- Baseline Idea - improve both Econometrics (add VAR) and ML (add BERT)
- Improving: add more macro variables and text sources

◀ Back

Uncertainty Paper

- Forecast Economic Activity with text analysis on both Micro and Macro levels
- Create an economic policy uncertainty (EPU) index (daily and monthly-based, can be policy specific) and look at the correlation with economic activity and try to predict it
- Data - baseline US, 10 newspapers for modern data, 6 newspapers for older observations. G10 economies as well.

◀ Back

MIDAS — Overview

- **Mixed Data Sampling:** relates a low-frequency target (quarterly GDP growth) to high-frequency predictors (monthly indicators) directly, without first averaging them to quarterly frequency
- Standard OLS forces every regressor onto the same frequency as y ; MIDAS keeps the within-quarter monthly information intact instead of discarding it
- Two flavors: *restricted* MIDAS imposes a distributed-lag polynomial (e.g. Almon) on the monthly lag weights; *unrestricted* MIDAS (U-MIDAS) simply includes each monthly lag as its own regressor
- This project uses **U-MIDAS**: $p = 6$ monthly lags per predictor, no functional-form restriction on how the lag weights decay

MIDAS — Model Setup

$$y_t^{(Q)} = \beta_0 + \sum_{i=1}^p \beta_i x_{3t-i+1}^{(M)} + \sum_{j=1}^q \gamma_j y_{t-j}^{(Q)} + \varepsilon_t$$

- $y_t^{(Q)}$: quarterly target (e.g. GDP growth) at quarter t
- $x_{3t-i+1}^{(M)}$: monthly predictor, stacked back $p = 6$ months from the quarter's last month
- Each monthly lag gets its own freely-estimated coefficient β_i — that's the “unrestricted” part
- Real-time vintage alignment matters: only monthly data actually available as of the forecast origin enters $x^{(M)}$

U-MIDAS vs. Restricted MIDAS — Trade-offs

- **U-MIDAS**: flexible, but $p = 6$ free parameters per predictor $\Rightarrow$ parameter proliferation once many monthly series are included
- **Restricted (Almon-lag) MIDAS**: imposes a smooth decay shape on lag weights, far fewer parameters, but risks misspecifying that shape
- Why U-MIDAS here: the predictor count is manageable, and it avoids committing to a polynomial shape the data may not support

MIDAS — Implementation Notes

- Implemented in the Stata forecasting pipeline (Forecasting Part/), vintage-aware FRED-MD alignment (ALFRED real-time scoring)
- $p = 6$ monthly lags stacked per predictor onto the quarterly target, as in the Ellingsen et al. replication
- Benchmarked via relative RMSE against AR, constant-growth, and SPF, with Diebold–Mariano and Clark–West tests

◀ Back

Elastic Net — Overview

- Linear regression with a combined L_1 (LASSO) + L_2 (ridge) penalty:

$$\hat{\beta} = \arg \min_{\beta} \|y - X\beta\|_2^2 + \lambda \left[\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right]$$

- L_1 shrinks some coefficients exactly to zero (feature selection); L_2 shrinks correlated coefficients together instead of arbitrarily picking one
- λ controls overall shrinkage strength, α (`l1_ratio`) controls the LASSO/ridge mix

Elastic Net — Why Here

- Topic-intensity predictors are highly collinear — by construction they sum to roughly 1 across topics each period
- Pure LASSO tends to pick one predictor from a correlated group and zero out the rest, arbitrarily; the ridge component in Elastic Net keeps correlated predictors together
- Useful as a genuine feature-selection model to contrast against OLS-factors/PCA, which never sets a predictor exactly to zero

Elastic Net — Tuning Choices

- `l1_ratio` grid restricted to $\{0.5, 1.0\}$, `n_alphas=20`, cross-validated with `TimeSeriesSplit(3)` — expanding-window CV, no look-ahead
- `max_iter=2000`, `tol=1e-3`: loosened from a slower default because the highly collinear topic features made coordinate descent slow to converge; the ranking of models is unaffected by this precision loss
- Features standardized before fitting (penalty is scale-sensitive)

Elastic Net — Diagnostic Value

- In the disagreement horse race, Elastic Net is the one model class that performs genuine feature selection rather than blind shrinkage
- When it is allowed to discard uninformative predictors, PCE $h=1$ relative-RMSE ratios rise *above* 1 (1.08–1.19, full sample) — the model finds nothing worth keeping and regularization noise hurts the forecast
- This is the tell that distinguishes “real signal” from “shrinkage toward the mean” in the other model classes

◀ Back

Random Forest — Overview

- An ensemble of regression trees: each tree is grown on a bootstrap resample of the data, splitting on a random subset of predictors at each node
- Final prediction = average of all trees' predictions — averaging cancels out each individual tree's overfitting (bagging)
- Captures nonlinearities and interactions among predictors automatically, with no parametric functional form to specify

Random Forest — Why Here

- Macro-text forecasting plausibly involves nonlinear/threshold effects (e.g. disagreement mattering only above some volatility level) that linear models can't capture
- Robust to irrelevant or noisy predictors — useful with ~ 80 topic-intensity features per predictor set
- Serves as the nonlinear counterpart to Elastic Net and OLS-factors in the horse race

Random Forest — Hyperparameters Used

- `n_estimators=300`: enough trees to stabilize the ranking across models; 500 was tried and found to add runtime without changing conclusions
- `max_depth=6`, `min_samples_leaf=5`: shallow trees and a minimum leaf size, both acting as regularization
- Fixed `random_state=42` for reproducibility across the horse race scripts

Random Forest — Caveats

- Training windows are short in quarters (expanding from ~ 30 –80 observations); deep, many-tree forests would overfit badly at this sample size
- Shallow depth (≤ 6) and a minimum leaf size of 5 are the guardrails against that overfitting
- Direct h -step forecasting (a separate model per horizon) rather than iterated forecasting, matching the other model classes in the horse race

◀ Back

- Principal Component Analysis compresses a large set of correlated predictors (here up to 1,566 regressors) into a handful of orthogonal factors capturing their common variation
- Each factor is a linear combination of all original predictors, ordered by how much variance it explains
- “OLS-factors” = standardize predictors $\rightarrow$ PCA $\rightarrow$ OLS of the target on the extracted factors. Main results: one factor (paper’s models) or five (four-model set), no lags of the target. **PCA-OLS**

PCA — Why Here

- With $T \approx 30\text{--}80$ quarterly training observations and up to 80 predictors, direct OLS is infeasible (more regressors than observations)
- Reducing to a small number of factors first restores identification and stabilizes estimation
- Standard approach in the dynamic-factor forecasting literature (and the one used by Ellingsen et al. for the “hard” and “news” blocks)

PCA — Choosing the Number of Factors

- $k = 5$ factors used by default; sandbox experimentation found $k \in \{3, 4, 5\}$ gave essentially the same out-of-sample performance
- Trade-off: too few factors discard predictive variation; too many re-introduce the overfitting PCA was meant to avoid

PCA — Interpretation Trade-off

- Cost: factors are not individually interpretable — a coefficient on “Factor 2” doesn’t map to any single topic or news feature
- Benefit: captures broad co-movement (e.g. a common “macro news tone” factor) that no single topic series would reveal alone
- Used throughout the horse race as the linear benchmark against which Elastic Net and Random Forest are compared

◀ Back

XGBoost — Overview

- Gradient-boosted trees: builds an ensemble **sequentially** — each new tree is fit to the residual (negative gradient) of the current ensemble's loss, rather than averaging independent trees as in Random Forest
- Prediction = sum of many small, shallow “weak learner” trees, each one nudging the ensemble a little closer to the target
- Regularized objective adds L_1/L_2 penalties on tree leaf weights on top of shrinkage (the learning rate), to keep any one tree from overfitting the residual it's chasing

XGBoost — Why Here

- Complements Random Forest: both are nonlinear tree ensembles, but boosting (sequential, error-correcting) trades off bias and variance differently than bagging (parallel, variance-reducing)
- Included as the fifth model class in the main horse race, alongside AR, OLS-factors, Elastic Net, and Random Forest
- Guarded by an `XGB_AVAILABLE` import check in the code — the horse race falls back to 4 models if `xgboost` isn't installed, rather than failing

XGBoost — Hyperparameters Used

- `n_estimators=300`, `max_depth=4`: shallower than RF's `max_depth=6` — boosting adds trees sequentially, so shallower trees are needed to avoid compounding overfit across rounds
- `learning_rate=0.05`: a small step size, so each tree corrects only a fraction of the remaining error
- `subsample=0.8`, `colsample_bytree=0.8`: each tree trained on 80% of rows/columns (stochastic gradient boosting) for extra regularization
- `reg_alpha=0.1` (L_1), `reg_lambda=1.0` (L_2): penalize leaf weights directly, on top of shrinkage

XGBoost — Caveats

- With ~ 80 quarterly training observations, deep or many-round boosting overfits fast — every knob above (shallow depth, low learning rate, subsampling, L_1/L_2) works together as a regularization stack, not a single safeguard
- Same one-time tuning discipline as the other models: hyperparameters tuned once on the 1996Q1–2003Q4 holdout rather than re-tuned at every forecast origin, to avoid overfitting the tuning itself to the OOS window

◀ Back

BERT Model

- **B**idirectional **E**ncoder **R**epresentations from **T**ransformers
- Designed to handle natural language processing tasks efficiently
- Key advantage: Uses a bidirectional context (looks at both left and right words)
- Proven effectiveness in tasks like sentiment analysis, question answering, and headline classification [◀ Back](#)

The full-corpus filter follows this pipeline with k-means (250 clusters) in place of HDBSCAN.

Core Concept

- BERT is built on the Transformer architecture, relying heavily on **self-attention** mechanisms
- Provides a **contextual embedding** of words, capturing nuances of the sentence
- *Masked Language Model (MLM)*: Randomly masks some tokens and predicts them
- *Next Sentence Prediction (NSP)*: Learns to predict whether two sentences follow each other

Pipeline Overview

1 Embedding $\rightarrow$ Projection

Convert each document d_i into a dense vector $\mathbf{e}_i \in \mathbb{R}^p$ (SBERT) and project to $\mathbf{z}_i \in \mathbb{R}^q$ (UMAP).

2 Density-based clustering

HDBSCAN finds K variable-sized clusters $C_k \subset \{\mathbf{z}_i\}$ and a noise set.

3 Topic representation

Aggregate the raw texts in each C_k and rank n-grams via c-TF-IDF.

4 Membership probabilities

HDBSCAN yields a confidence $P_{ik} = P(\mathbf{z}_i \in C_k)$ for the *assigned* cluster.

Step 1 — Embedding & Projection

$$d_i \xrightarrow{\text{SBERT}} \mathbf{e}_i \in \mathbb{R}^p \xrightarrow{\text{UMAP}} \mathbf{z}_i \in \mathbb{R}^q, \quad q \ll p$$

- SBERT captures semantic meaning; tunable by domain-specific fine-tuning.
- UMAP preserves *local* neighborhood structure critical for density-aware clustering.
- *Key hyper-parameters:* `n_neighbors`, `n_components`, `min_dist`.

Step 2 — Density-Based Clustering (HDBSCAN)

- HDBSCAN builds a condensed dendrogram of the mutual-reachability graph of $\{\mathbf{z}_i\}$.
- Clusters survive at high density thresholds; sparse points are labeled noise = -1.
- **Soft assignment:** for every member $\mathbf{z}_i \in C_k$

$$P_{ik} = \exp(-\text{core_dist}(i, k)) \in [0, 1]$$

where `core_dist` is the distance from $\mathbf{z}_i$ to its nearest “core” point along the cluster hierarchy.

Step 2 (cont.) — Key Hyper-Parameters

Practical knobs

<code>min_cluster_size</code>	smallest allowed cluster (granularity)
<code>min_samples</code>	outlier sensitivity (robustness)
<code>metric</code>	distance metric in UMAP space
<code>cluster_selection</code>	rule for cutting dendrogram — eom vs. leaf
<code>alpha</code>	balance between cluster density and separation
<code>approx_min_span_tree</code>	speed/accuracy trade-off for very large datasets

Step 3 — Topic Representation via c-TF-IDF

Class aggregation

For cluster k with documents $C_k = \{d_i\}$, form its *class document* $D_k = [\cup_{i \in C_k} d_i]$.

$$\text{c-tf-idf}(t, k) = \frac{\text{tf}(t, D_k)}{|D_k|} \ln\left(\frac{N}{\sum_{j=1}^K \mathbf{1}_{\{t \in D_j\}}}\right)$$

↑ importance of term t to cluster k

- Rank terms; top- n form the human-readable label.
- Optional diversity filter drops near-duplicates to widen semantic coverage.

Step 4 — Document–Topic Probabilities

- **Dominant topic:** each document *inherits* its cluster label $k^*(d_i) \in \{-1, 0, \dots, K-1\}$.
- **Membership confidence:** $P_{ik^*} = P(\mathbf{z}_i \in C_{k^*})$ from HDBSCAN's soft clustering.
- *Important:* BERTopic does not yield an LDA-style multinomial θ_{ik} distribution.

Take-away

“Probability” in BERTopic is a *confidence in cluster membership*, not the fractional mixture of topics within a document.

Example: Sentiment Analysis on Headlines

Input Headline:

"Fed Keeps Rates Steady and Flags Heightened Uncertainty About the Economy"

- Step 1: Tokenize (WordPiece)
- Step 2: BERT → contextual embeddings
- Step 3: Feed embeddings into a classification head (e.g., for sentiment: positive, negative, neutral)
- Step 4: **Output:** Probability distribution over sentiment classes

Look-ahead bias and time-stamped language models

- A pretrained language model has read text written after the dates it is applied to. Using it in a forecast can leak the future: lookahead bias (Sarkar and Vafa, 2024).
- In this project two components are pretrained on text that includes years after the dates being forecast: the sentence encoder behind the economic-article filter, and FinBERT sentiment.
- Topic shares, policy-uncertainty counts and dictionary sentiment use no pretrained model themselves; in the main specification they are computed on the articles the filter selects.
- Language models with time-stamped training data exist: StoriesLM (Sarkar, 2024); ChronoBERT and ChronoGPT (He, Lv, Manela and Wu, 2025).
- Done for sentiment: a 1989 StoriesLM checkpoint forecasts no differently from FinBERT (39,492 US articles, earlier targets). Next: the filter's encoder.