Trading Macro Narratives

A Plain-English Walkthrough

Gabin Taibi[1,2] Axel Gross-Klussmann[2] Joerg Osterrieder[1]

April 2026

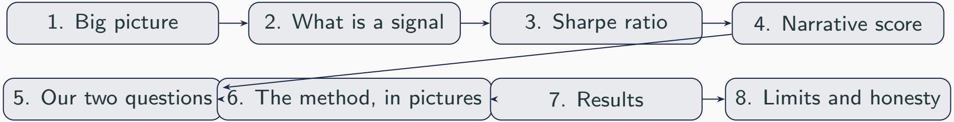
[1]University of Twente [2]Quoniam Asset Management

The One-Sentence Takeaway

Computer-read news sentiment about recurring economic stories may carry useful information about which industries will outperform others next week.

Hedged on purpose: in-sample evidence over 21 years, not a recommendation.

Where This Deck Goes



Eight short arcs. One concept per slide. Every term gets defined before it is used.

A Signal, and What "Cross-Sectional" Means

A signal is any number a researcher computes today, hoping it relates to tomorrow's return.

- **Analogy:** a weather forecast temperature is a signal for tomorrow's outdoor plans.
- Our signal will be: a computer-read score of the tone in news about a recurring economic story.

Cross-sectional question

Which assets will go up *more than others* this week? *(this paper)*

Time-series question

Will the market *as a whole* go up or down? *(not this paper)*

We only try to rank assets against each other. We do not predict the market level.

Why Cross-Sectional Is Easier

- **Analogy:** picking the best runner in a race is easier than predicting the race's finishing time.
- Many things move the whole market (interest rates, growth shocks). They mostly cancel out when we ask *which industry beats the average*.
- In our paper: we sort 49 US industry portfolios each week and ask which ones our signal says will outperform the others.

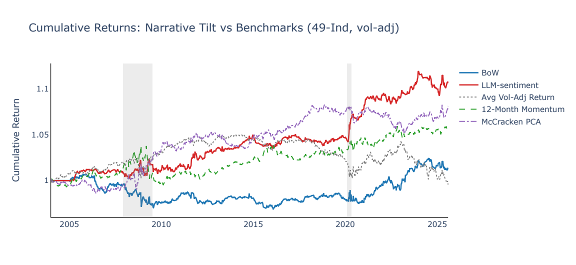
No level forecast required: only the *ordering* has to be useful.

What Is a Sharpe Ratio?

Return per unit of risk, annualised.

- **Analogy:** a gig's per-minute pay relative to its risk. Higher is better.
- Rough reference bands (long-short strategies):
  - ≈ 0.3 weak
  - ≈ 0.5 decent
  - ≥ 1.0 strong
- Our best combined number in this paper is about 0.65 in-sample (paper line 566).

How to Read a Sharpe Ratio on a Chart



Reading guide:

- The line shows *cumulative* return over time.
- A Sharpe of 0.48 (paper L296) means: drifts up steadily, with week-to-week bumps.
- The flat line is BoW: roughly zero return per unit of risk.
- The drifting line is LLM: positive but bumpy.

What Is a Narrative?

A narrative is a recurring economic story that journalists write about repeatedly.

- Examples: "recession fears", "Fed tightening", "oil supply shock".
- Recurring, not one-off. Evergreen, not topical. (One-off topics like COVID itself are excluded.)
- Taxonomy comes from Bhargava et al. (2023). 65 evergreen narratives in total (paper line 98).

65 Narratives, Grouped into 7 Reservoirs

The 65 narratives are bucketed into 7 thematic reservoirs (paper line 99):

1. **Corporate** (company stories: earnings, governance, M&A)
2. **US Macroeconomic** (domestic growth, employment, inflation)
3. **Global Macro and Geopolitics** (trade, conflict, foreign growth)
4. **Financial Markets and Instruments** (volatility, credit, fixed income)
5. **Investment Strategies** (value, momentum, ESG)
6. **Sectors and Industries** (tech, energy, banking)
7. **Society** (demographics, education, social trends)

Plain-English glosses in parentheses; the bold labels are the paper's verbatim names.

Two Ways a Computer Can Read a Narrative

BoW (bag-of-words): counting

Count how often keywords for a narrative appear.

Captures: attention, intensity of coverage.
Misses: direction (panic vs. relief).

LLM sentiment: reading the tone

A language model reads each headline and gives a directional score:
$\{-2, -1, 0, +1, +2\}$
(very negative to very positive).

Captures: direction of belief.

Counting vs. understanding. Two different views of the same headline stream.

Same Week, Two Scores: A Tiny Example

Suppose this week's news contains many articles like "Fed signals pause".

- **BoW (counting):** the count of "Fed" headlines surges. The signal goes up.
- **LLM (tone):** the tone shifts from negative ("Fed will hike") to neutral ("Fed will pause"). The signal moves toward zero or slightly positive.
- **Both numbers move, but in different ways.** BoW measures *how much* we talk about the Fed; LLM measures *how worried* we are about the Fed.

Attention and Tone Often Disagree

The week-to-week *changes* in BoW (attention) and LLM (tone) are negatively correlated across all 7 reservoirs:

Reservoir	Correlation of changes
US Macroeconomic	−0.03
Global Macro and Geopolitics	−0.17
other reservoirs	between these two values

Range −0.03 to −0.17 (paper line 121).

The two signals are not redundant. They argue with each other.

Our Two Questions, in Plain English

Question 1 (CW):

Which *industry* reacts most to the current tone shift in a given narrative?

Rank assets.

Question 2 (NarrMom):

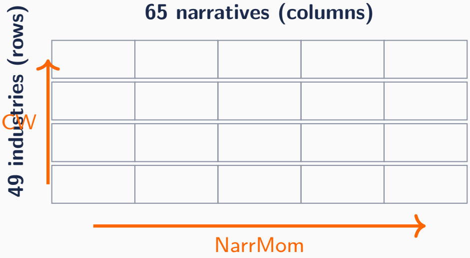
Which *narrative* is itself on an upswing in tone right now?

Rank narratives.

Both are ranking questions, not level-forecast questions.

Both apply weekly to a fixed cross-section.

Two Dimensions of the Same Table



CW ranks along the rows (industries). **NarrMom** ranks along the columns (narratives).

Same data, two different axes.

What Is a “Beta”? (Our Version)

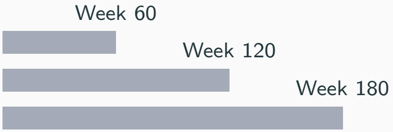
Our beta = how strongly one industry’s return moves on weeks when one narrative’s tone changes.

- **Analogy:** a kitchen thermometer’s responsiveness to opening the oven door.
- **Not** the CAPM market beta. Different concept; same word.
- One beta per (industry, narrative, week). About $49 \times 65 = 3{,}185$ betas per week.
- Some industries are very sensitive to “Fed tightening” (high beta), others barely respond (low beta).

What Is an "Expanding Window"?

We estimate today's beta using all data from the start of the sample up to last week.

- As time passes, the window *grows*.
- Contrast: a *rolling* window has a fixed length and drops old data.
- Contrast: *in-sample* uses future data, which we forbid.
- Minimum window: 52 weeks before any beta is used.



What Is “Rank-Scaling”?

Replace raw signal values with their cross-sectional ranks, then map to $[-2, +2]$.

- **Analogy:** grading on a curve. We do not care about the raw score, only the ordering.
- Top-ranked asset gets weight $+2$, bottom-ranked gets -2, middle assets get values in between.
- Weights sum to zero each week (long-short).
- Outliers are neutralised: the top asset cannot dominate the portfolio just because its raw signal is huge.

What Is “Vol-Adjusted” and Why

Divide each asset’s returns by that asset’s own volatility before forming the portfolio.

- **Analogy:** a sound engineer adjusting volumes of different instruments so none drowns out the others.
- One naturally noisy industry (e.g. tech) would otherwise dominate; vol-adjusting puts every industry on a 1% annualised volatility scale.
- Source: Brandt, Santa-Clara, and Valkanov (2009), the standard reference for parametric portfolio policies.

The Two-Week Lag, and Why

We compute the signal at time t but only use it to trade in week $t+2$.
Never t. Never $t+1$.

- **Analogy:** even a real trader needs time to act on a report once it lands on their desk.
- Purpose: no look-ahead bias. The signal must be fully observable before the trade is placed.
- Two weeks is conservative; a one-week lag is more common in academic papers. We chose two to be safe.

What Is the McCracken PCA Benchmark?

Our non-text comparison signal.

- Start with 100+ US monthly macro time series (the FRED-MD dataset).
- Extract 10 principal components: summary indices that capture the biggest common swings across the series.
- Treat these 10 indices as if they were narrative scores. Run them through the same CW pipeline.
- Source: McCracken and Ng (2016).

Why have a benchmark? So we can ask: is text signal really adding something beyond classical macro?

Strategy 1: Characteristics-Weighted (CW)

Each week, for each (industry, narrative) pair:

1. Compute *beta* (sensitivity) $\times$ *current tone change*. Call this the raw signal.
2. Add it up per industry (across all 65 narratives) to get each industry's score.
3. Rank-scale the scores to $[-2, +2]$ (frame 18).
4. Vol-adjust the returns (frame 19) and apply the two-week lag (frame 20).
5. Long top-scored industries, short bottom-scored.

Lineage: Brandt–Santa-Clara–Valkanov (2009) for the weighting; Bhargava et al. (2023) for the narrative betas.

Strategy 2 (Part 1): Mimicking Portfolios

For each narrative, build a tiny long-short portfolio that represents "how it feels when this narrative moves".

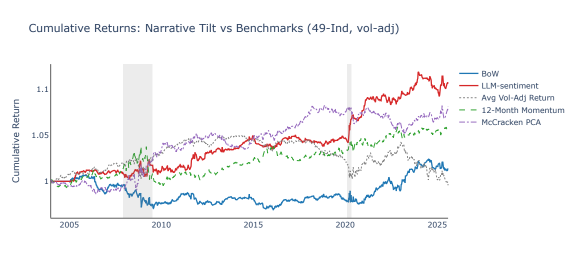
- **Analogy:** if the narrative is "inflation fears", go long the industry most exposed to inflation fears and short the industry least exposed.
- This single long-short pair is called the narrative's mimicking portfolio.
- One mimicking portfolio per narrative; 65 in total.
- Source: Lee et al. (2025) introduce narrative-mimicking for use in equity returns.

Strategy 2 (Part 2): Ride the Trending Narratives

Now rank the 65 narratives by whether their tone has been trending up or down.

- Long the rising narratives' mimicking portfolios.
- Short the falling narratives' mimicking portfolios.
- This is narrative momentum: ride the narratives whose tone has been improving lately.
- Specific parameter choices (how many to long, how many to short) live in the glossary frame.

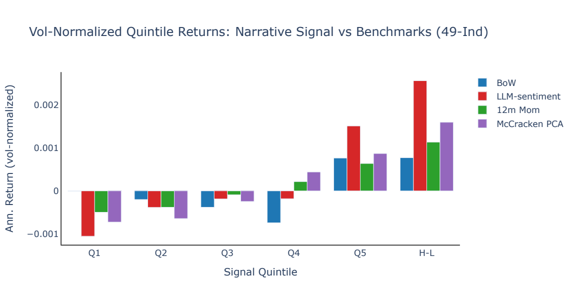
Result 1: LLM Reads News Better than Counting Words



Strategy 1 (CW) on 49 US industries:

- LLM: Sharpe 0.48 (paper L296)
- BoW: Sharpe near zero
- Same pipeline, same data; only the way the news is read changes.

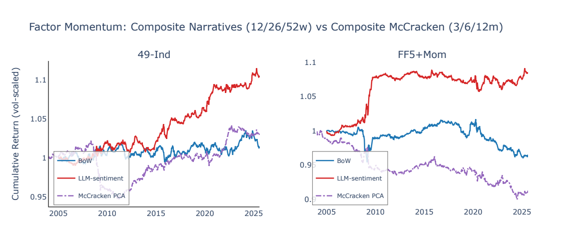
Result 2: The Signal Is Monotonic



Sort industry-weeks into 5 buckets by signal (Q1 lowest to Q5 highest):

- Returns increase *monotonically* from Q1 to Q5.
- High-minus-Low Sharpe $\approx$ 0.53, $t = 2.47$ (paper L341).
- The ordering reflects real information; no “U-shape”.

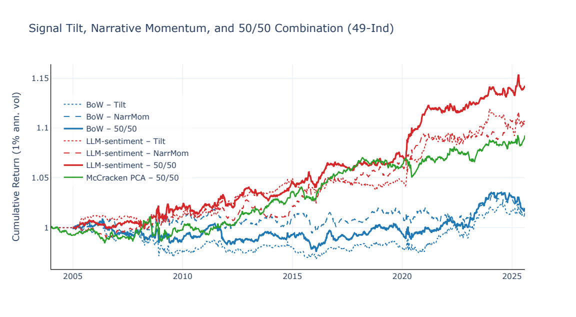
Result 3: Narrative Momentum Adds Something New



Strategy 2 (NarrMom) on 49 industries:

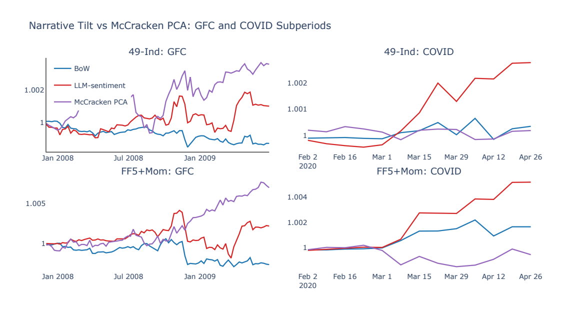
- LLM: Sharpe ≈ 0.49
- BoW NarrMom: roughly zero.
- Only the LLM-based version shows positive long-run drift.

Result 4: The Two Strategies Diversify Each Other



- Correlation between CW and NarrMom: $\rho \approx 0.14$ on 49-Ind.
- Combined (50/50): Sharpe 0.65, $t = 3.01$ (paper L566).
- Different questions $\Rightarrow$ low correlation $\Rightarrow$ diversification bonus.

Result 5: Different Crises Reward Different Signals



- **GFC (slow burn):** McCracken PCA wins. SR 0.82 (paper L749).
- **COVID (fast shock):** narrative signal wins. SR strongly positive (paper L751).
- Macro data is fast in slow crises; news is fast in fast crises.

Numbers-Summary: Every Cited Number

Number	Where	Paper line
SR 0.48 (LLM CW, 49-Ind)	Frame 8, 24	296
SR 0.07 (BoW CW, 49-Ind)	Frame 24	296 to 297
SR 0.53, $t = 2.47$ (H-L quintile)	Frame 25	341
SR 0.49 (LLM NarrMom, 49-Ind)	Frame 26	532
$\rho = 0.14$ (CW vs NarrMom, 49-Ind)	Frame 27	565
SR 0.65, $t = 3.01$ (combined, 49-Ind)	Frame 27	566
Cross-model ρ_Δ in $[-0.17, -0.03]$	Frame 13	121
GFC McCracken SR 0.82 (49-Ind)	Frame 28	749
COVID narrative SR 1.8 to 5.6	Frame 28	751
65 narratives, 7 reservoirs	Frame 9, 10	98 to 99
Sample January 2004 to July 2025	Frame 24	131

FF5+Mom numbers (the factor-spread universe) are reported in the paper but kept off the storytelling frames here.

Honest Limitations

- **In-sample only.** All results are within the 2004 to 2025 window. No out-of-sample test.
- **Gross returns.** Trading costs and turnover are not modelled; net Sharpe could be materially lower.
- **Concentration risk.** Strategy 2 relies on one industry on each side of each mimicking portfolio; broader baskets degrade the signal.
- **Two crises.** The regime-asymmetry claim rests on GFC and COVID only.
- **No factor-model alpha.** We do not adjust the Sharpe for standard factor exposures.

The paper documents an association, not a causal claim.

Conclusion and Next Steps

LLM narrative scores and macro PCA carry different information; combining them may improve cross-sectional risk-adjusted returns in this sample.

Next steps:

- **Transaction costs.** Quantify net-of-cost Sharpe for both strategies.
- **Topic-model benchmark.** Bybee et al. (2024) provide a richer non-textual comparator.
- **Out-of-sample.** Hold out a window; replicate the headline results.

Glossary

- **Signal** number computed today, hoped to relate to tomorrow's return
- **Cross-sectional** ranking assets against each other in the same week
- **Time-series** forecasting the market level over time
- **Sharpe ratio** return per unit of risk, annualised
- **Narrative** recurring economic story (one of 65)
- **Reservoir** bucket of related narratives (7 of them)
- **BoW** bag-of-words; counts headlines
- **LLM sentiment** language model directional score $\{-2 \ldots +2\}$
- **Beta (this paper)** responsiveness of one industry to one narrative's tone
- **CAPM beta** not our beta; classic market beta
- **Expanding window** estimation window that grows over time
- **Rank-scaling** convert ranks to $[-2, +2]$
- **Two-week lag** signal at time t, trade at $t + 2$
- **McCracken PCA** 10 principal components from FRED-MD
- **FF5+Mom** Fama-French 5 factors plus momentum
- **49-industry** standard US industry portfolios
- **CW** Characteristics-Weighted portfolio (strategy 1)
- **NarrMom** Narrative Momentum (strategy 2)
- **Mimicking portfolio** long one industry, short another, to track a narrative
- **K=1** one industry on each side of a mimicking portfolio
- **J=5** top 5 rising narratives long, bottom 5 falling short
- **Quintile** fifth of the data (Q1 lowest, Q5 highest)
- **In-sample** uses data from the evaluation period
- **Out-of-sample** reserved data, not used in

Selected References

- Shiller, R. (2017). Narrative Economics. *American Economic Review*.
- Gentzkow, M., Kelly, B. T., and Taddy, M. (2019). Text as Data. *Journal of Economic Literature*.
- Bhargava, A., et al. (2023). Quantifying Narratives. Working paper.
- Lee, C., et al. (2025). Narrative Momentum. Working paper.
- McCracken, M. W., and Ng, S. (2016). FRED-MD. *Journal of Business and Economic Statistics*.
- Da, Z., Gurun, U., and Warachka, M. (2014). Frog in the Pan. *Review of Financial Studies*.
- Bybee, L., et al. (2024). Business News and Business Cycles. *Journal of Finance*.
- Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric Portfolio Policies. *Review of Financial Studies*.

Thank You

Questions?

Gabin Taibi	g.taibi@utwente.nl
Axel Gross-Klussmann	axel.gross-klussmann@quoniam.com
Joerg Osterrieder	j.osterrieder@utwente.nl

Paper available upon request