

Trading Macro Narratives: Evidence from LLM-Based Signals for Industry and Factor Portfolios

Gabin Taibi^{1,2}, Axel Gross-Klussmann², Joerg Osterrieder¹

¹University of Twente ²Quoniam Asset Management

April 8, 2026

Abstract

We study whether directional macro-narrative scores extracted from news can be used to form cross-sectional signals for U.S. equity and factor portfolios. Using 65 evergreen narratives over 2004–2025, we estimate expanding-window narrative betas and combine them with weekly changes in narrative scores to construct two related strategies: a characteristics-weighted portfolio that ranks assets by exposure to current narrative shifts, and a narrative momentum portfolio that ranks narratives by their recent score trends. We compare a prompted large language model (LLM) sentiment score with a simpler bag-of-words (BoW) attention measure and with a non-text benchmark based on principal components from the FRED-MD macro panel. Across both asset universes considered, the LLM-based signals generate positive returns in both portfolio constructions, while the BoW baseline is generally weaker and the macro benchmark remains competitive. The evidence is most naturally interpreted as showing that directional narrative measures extracted from text can complement traditional macro signals, rather than as a stand-alone replacement. Overall, the results are encouraging but still provisional: economic magnitudes are moderate, performance is not uniform across specifications, and questions of turnover, trading costs, and broader implementability remain open.

Keywords: Narrative economics, macro themes, cross-sectional allocation, narrative betas, large language models, characteristics-weighted portfolios

1 Introduction

Recent work in narrative economics argues that stories help propagate economic fluctuations and shape beliefs about spending, investing, and policy (?). In parallel, the text-as-data literature has shown that economic and financial text can be turned into quantitative state variables that capture sentiment, attention, and risk perceptions (???). News-based measures have been used to track business-cycle conditions, policy uncertainty, and broader macroeconomic narratives (???).

Within asset pricing, recent work has begun to map such narratives into tradable signals. ? show that narrative betas—sensitivities of returns to changes in narrative scores—can be linked to market- and firm-level returns, while ? document momentum in narrative attention. This literature suggests that narrative measures may contain information that is not fully summarized by standard factors and macro datasets, but it also leaves open a practical question: how should narratives be quantified for portfolio construction?

That question matters because different text models measure different objects. Count-based or similarity-based methods are often well suited to capturing attention or topic intensity, but they can struggle with direction and context. More flexible language models, by contrast, can be used to assign directional scores to the same underlying themes. In this paper, we do not treat these approaches as strict substitutes. Instead, we ask whether LLM-based directional narrative scores contain useful cross-sectional information beyond a simpler bag-of-words (BoW) attention measure and beyond a traditional macro benchmark based on FRED-MD principal components (?).

To answer this question, we construct weekly signals for 65 evergreen macro narratives using news headlines from January 2004 to July 2025. For each narrative and asset, we estimate expanding-window narrative betas from weekly returns on weekly changes in narrative scores. We then study two related portfolio constructions. The first is a characteristics-weighted (CW) strategy that ranks assets by the product of their narrative beta and the recent change in the corresponding narrative score. The second is a narrative momentum strategy in the spirit of ?, which forms narrative-mimicking portfolios and then goes long narratives with improving scores and short narratives with deteriorating scores. We apply both approaches to the 49 Fama–French industry portfolios and to Fama–French factor-spread portfolios augmented with momentum (FF5+Mom).

This design yields four main results. First, LLM-based directional scores produce positive CW returns and cleaner signal monotonicity than the BoW baseline in both asset universes, although the magnitudes are moderate rather than decisive. Second, narrative momentum offers additional evidence that the LLM scores contain information not captured well by the BoW measure, but the strength of this result depends on the exact implementation and is best interpreted cautiously. Third, a benchmark built from

McCracken-style macro data remains competitive, especially in slower-moving episodes, while the LLM-based signals appear most useful in fast-changing post-2020 environments. Taken together, the results point more toward complementarity with traditional macro signals than toward outright replacement. Fourth, the CW and narrative momentum strategies load on different dimensions of the same underlying data and therefore diversify each other, even though they rely on the same narrative scores and estimated betas.

Our contribution is therefore narrower than a claim that LLMs “solve” narrative trading. The paper provides new evidence that directional narrative scores extracted from news can be used to form portfolio signals that differ meaningfully from both count-based textual measures and standard macro factors. At the same time, the results leave several important questions open, including turnover, trading costs, concentration in the mimicking portfolios, and the stability of the signals across specifications. We view the evidence as a proof of concept and as a starting point for a broader comparison between textual narrative measures and conventional macroeconomic state variables.

Relatedly, our CW and narrative-momentum exercises are meant to illuminate two complementary uses of the same narrative panel. The CW strategy asks which assets should respond most to current narrative shifts; narrative momentum asks which narratives are themselves trending. Their joint performance therefore matters less as a pursuit of a single headline Sharpe ratio than as evidence that narrative information can be organized along multiple economically distinct dimensions.

The remainder of the paper proceeds as follows. Section ?? describes the narrative, return, and macro datasets. Section ?? outlines the beta estimation and portfolio constructions. Section ?? presents the empirical results, and the appendix reports additional crisis and sensitivity analyses.

2 Data

2.1 Narrative Scores

We use 65 evergreen narratives drawn from the taxonomy of ?, excluding three event-specific narratives (COVID-19, Trade War, Brexit) that are absent over significant portions of the sample. Each narrative belongs to one of seven thematic “reservoirs”: Corporate, US Macroeconomic, Global Macro & Geopolitics, Financial Markets & Instruments, Investment Strategies, Sectors & Industries, and Society.

For the BoW baseline, the daily narrative score $\theta_{n,t}^{\text{BoW}}$ is the average cosine similarity between all headlines on day t and the centroid embedding of narrative n , filtered for negative sentiment. This is an *attention* measure: it captures how prominently a narrative features in the news flow, without distinguishing the direction of sentiment. BoW scores are non-negative and unbounded, surging during crises as article counts rise.

For the LLM model, the daily score $\theta_{n,t}^{\text{LLM}}$ is a *sentiment* measure obtained from structured LLM calls. Each day’s headlines are grouped by overarching theme and passed to a prompted LLM that assigns one of five ordinal sentiment labels per narrative: *negative*, *mostly negative*, *can’t tell*, *mostly positive*, or *positive*, mapped to $\{-2, -1, 0, +1, +2\}$. The prompt instructs the model to assess whether the tone towards each narrative implies improving or deteriorating conditions (e.g., for “Recession,” *negative* indicates worsening recession risk, not negative sentiment *about* recessions). This design means the LLM score is bounded, ordinal, and inherently directional—properties that fundamentally shape its behaviour in both the CW portfolio and narrative momentum frameworks. Table ?? summarises score statistics by reservoir.

A notable feature is that cross-model change correlations (ρ_{Δ}) are negative across all reservoirs, ranging from -0.03 (US Macroeconomic) to -0.17 (Global Macro & Geopolitics). This suggests the two models respond to different aspects of narrative dynamics. Full narrative-level descriptives are omitted for brevity.

2.2 Industry Returns

We use value-weighted excess returns for the 49 Fama–French industry portfolios, obtained from Kenneth French’s data library. Daily returns are aggregated to weekly frequency by compounding. We also construct factor spread portfolios from the five Fama–French factors (MKT, SMB, HML, RMW, CMA) plus momentum (MOM). The sample runs from January 2004 through July 2025 (approximately 1,125 weekly observations).

2.3 McCracken PCA Macro-Factor Benchmark

As a non-textual benchmark, we use ten principal components extracted from the FRED-MD macroeconomic panel of ?. FRED-MD comprises over 100 monthly U.S. macroeconomic time series—covering employment, industrial production, interest rates, prices, housing, and financial conditions—transformed to stationarity and reduced via PCA to ten factors (F_1, \dots, F_{10}). We work with monthly first differences of these factors, yielding a panel of ten monthly macro-innovation series available from 1960 onward.

The McCracken benchmark applies the same signal-construction pipeline as the narrative models. Monthly expanding-window betas are estimated by regressing each asset’s monthly return on each factor change, using all data up to month t with a minimum window of 12 months. Because the FRED-MD panel begins decades before the narrative sample, the expanding window incorporates approximately 34 years of history by the start of our evaluation period in 2004, providing substantially more estimation data than the narrative betas (which begin in 2003). Monthly signals ($\beta_{i,f,t} \times \Delta F_{f,t}$) are forward-filled to weekly dates and then enter the same lagging, rank-scaling, and portfolio-construction steps described in Section ??.

Table 1: Narrative Score Descriptives by Reservoir (Weekly)

Reservoir	N	BoW \bar{I}	BoW σ	BoW AC(1)	LLM \bar{I}	LLM σ	LLM AC(1)	ρ_ℓ	ρ_Δ
Corporate	10	1.10	0.84	0.51	0.04	0.48	0.26	-0.09	-0.07
Financial Markets & Instruments	9	10.14	5.50	0.63	-0.13	0.55	0.20	-0.07	-0.13
Global Macro & Geopolitics	7	1.47	1.12	0.52	-0.20	0.47	0.42	-0.37	-0.17
Investment Strategies	11	1.03	0.89	0.69	0.14	0.31	0.14	0.01	-0.05
Sectors & Industries	5	0.89	0.79	0.41	-0.05	0.48	0.37	-0.29	-0.14
Society	8	0.73	0.89	0.43	-0.08	0.20	0.26	-0.27	-0.16
US Macroeconomic	15	12.14	7.49	0.45	-0.40	0.69	0.39	-0.15	-0.03

Notes: N is the number of narratives in each reservoir. $\bar{\theta}$ denotes the time-series mean of the weekly score. $AC(1)$ is the first-order autocorrelation. ρ_ℓ and ρ_Δ are the cross-model correlations in levels and first differences, respectively.

This benchmark tests whether the cross-sectional information in narrative signals could be replicated by standard macroeconomic data. If the narrative signal merely proxies for macro-factor exposures, the McCracken PCA strategy should perform at least as well.

3 Methodology

The signal construction is inspired by the narrative beta framework of ?, adapted for industry-level cross-sectional analysis. We first describe the shared building blocks (Sections ??–??), then the two portfolio strategies that use them: the characteristics-weighted (CW) portfolio (Section ??) and the narrative momentum strategy (Section ??).

3.1 Weekly Aggregation and Beta Estimation

Daily narrative scores are resampled to weekly frequency by averaging within each Monday–Friday window:

$$\bar{\theta}_{n,w} = \frac{1}{5} \sum_{d \in w} \theta_{n,d},$$

where w indexes the calendar week. We then define weekly score changes as

$$\Delta\theta_{n,w} = \bar{\theta}_{n,w} - \bar{\theta}_{n,w-1}.$$

Industry excess returns are compounded over the same window: $R_{i,w} = \prod_{d \in w} (1 + r_{i,d}) - 1$. Weekend observations are excluded from all series.

For each narrative n and industry i , we estimate an expanding-window univariate beta using all data up to week t (with a minimum of 52 weeks):

$$R_{i,\tau} = \alpha_{i,n} + \beta_{i,n,t} \Delta\theta_{n,\tau} + \varepsilon_{i,\tau}, \quad \tau \in [1, t]. \quad (1)$$

The beta $\beta_{i,n,t}$ measures industry i 's contemporaneous sensitivity to weekly changes in narrative n as estimated at week t , using an expanding window that incorporates all available history. Untabulated tests using a fixed 52-week rolling window lead to similar qualitative results.

3.2 Signal Construction

We define a general building block that underpins both the CW portfolio and the narrative momentum strategy. The smoothed score change for narrative n at horizon h weeks is:

$$\overline{\Delta\theta}_{n,t}^{(h)} = \frac{1}{h} \sum_{s=0}^{h-1} \Delta\theta_{n,t-s}, \quad \Delta\theta_{n,t} \equiv \bar{\theta}_{n,t} - \bar{\theta}_{n,t-1}. \quad (2)$$

Multiple lookbacks $\mathcal{H} = \{h_1, h_2, h_3\}$ can be combined into a composite:

$$\overline{\Delta\theta}_{n,t}^{\text{comp}} = \frac{1}{|\mathcal{H}|} \sum_{h \in \mathcal{H}} \frac{\overline{\Delta\theta}_{n,t}^{(h)}}{\sigma_t^{(h)}}, \quad \sigma_t^{(h)} = \text{std}_n(\overline{\Delta\theta}_{n,t}^{(h)}), \quad (3)$$

where $\sigma_t^{(h)}$ is the cross-sectional standard deviation of the rolling-mean score change across all narratives at horizon h . Each horizon’s rolling mean is divided by this dispersion before averaging, but *not* demeaned. This equalises the dispersion across horizons—shorter lookbacks have roughly twice the cross-sectional spread of longer ones—while preserving the cross-sectional mean, which carries economic information when the signal enters multiplicatively with β . The single-week change $\Delta\theta_{n,t}$ is the degenerate case $\mathcal{H} = \{1\}$. Short lookbacks (1–12 weeks) capture high-frequency news flow and are natural for cross-sectional asset allocation (“which industries are moving now?”), while longer lookbacks (12–52 weeks) identify persistent narrative trends suited for cross-narrative selection (“which stories have sustained momentum?”). We specify the horizons used by each strategy in their respective sections below.

3.3 Characteristics-Weighted Portfolio

The CW signal for industry i , narrative n , week t is:

$$S_{i,n,t} = \beta_{i,n,t} \times \overline{\Delta\theta}_{n,t}^{\text{comp}}. \quad (4)$$

This captures both the direction and magnitude of the industry’s narrative exposure, scaled by recent narrative shifts. Because the signal enters multiplicatively with β , the magnitude of $\overline{\Delta\theta}$ carries economic information—a narrative undergoing a large shift should contribute more than one that is barely moving. We use a short-horizon composite $\mathcal{H} = \{4, 8, 12\}$ weeks ($\approx 1/2/3$ months), which smooths over high-frequency noise while preserving the economic magnitude that the beta multiplier requires.¹

We apply a two-week lag to the raw signal ($S_{i,n,t-2}$) to ensure implementability and to accommodate the negative autocorrelation induced by differencing. The lagged signals are then rank-scaled cross-sectionally within each week:

$$\tilde{S}_{i,n,t} = \text{rank-scale}(S_{i,n,t-2}) \in [-2, +2], \quad (5)$$

where the rank-scaling assigns the most negative signal a weight of -2 and the most positive a weight of $+2$, with linear interpolation in between. The ranking is performed across all (i, n) pairs on each date, ensuring cross-sectional comparability across the $N_{\text{industries}} \times 65$

¹The degenerate single-week case $\mathcal{H} = \{1\}$ produces virtually identical LLM-sentiment performance (Appendix ??).

signal entries.

Industry excess returns are volatility-adjusted by dividing by rolling industry-level standard deviation (annualized):

$$R_{i,t}^{\text{va}} = \frac{R_{i,t}}{\hat{\sigma}_i \sqrt{52}}. \quad (6)$$

The portfolio return on week t is the cross-sectional average of signal-weighted, volatility-adjusted returns:

$$R_t^{\text{cw}} = \frac{1}{N} \sum_{i,n} \tilde{S}_{i,n,t} \times R_{i,t}^{\text{va}}, \quad (7)$$

where N is the total number of industry-narrative combinations. This is a characteristics-weighted portfolio in the sense of ??: the weight on each asset is a linear function of its narrative signal characteristic $\tilde{S}_{i,n,t}$. By construction, the portfolio overweights industries with strong positive narrative signals and underweights those with negative signals.

3.4 Narrative Momentum Strategy

The CW framework exploits cross-*asset* variation in narrative betas: for a given narrative shift, which assets benefit most? A complementary question is which *narratives* are currently gaining or losing momentum, regardless of the specific assets they load on. Following ??, we exploit *momentum in narrative attention* directly. Rather than weighting existing industry portfolios by narrative signals, this approach constructs *narrative-mimicking portfolios*—long-short equity portfolios that isolate the return associated with each narrative—and then goes long narratives with rising scores and short those with declining scores.

Mimicking portfolios. For each narrative n and week t , we rank assets by their lagged expanding beta $\beta_{i,n,t-1}$ and go long the asset with the highest beta and short the asset with the lowest beta. Each mimicking portfolio return is vol-scaled to 1% annualised volatility so that all narratives contribute equally to the spread portfolio. With 65 ever-green narratives and $K = 1$ assets per leg, this yields 65 mimicking portfolio return series per model and universe.

Narrative score momentum. We measure narrative momentum using the same rolling-mean operator $\overline{\Delta\theta}_{n,t}^{(h)}$ defined in Equation (??), evaluated at lookback M weeks:

$$\text{Mom}_{n,t}^{(M)} = \overline{\Delta\theta}_{n,t}^{(M)} = \frac{1}{M} \sum_{s=0}^{M-1} \Delta\theta_{n,t-s}. \quad (8)$$

At each week t , narratives are ranked by $\text{Mom}_{n,t-1}^{(M)}$ (lagged one week for implementability). The top J narratives are classified as *rising* and the bottom J as *declining*.

Spread portfolio. The spread portfolio return on week t is:

$$r_t^{\text{spread}} = \frac{1}{J} \sum_{n \in \text{Rising}} r_{n,t}^{\text{mim}} - \frac{1}{J} \sum_{n \in \text{Declining}} r_{n,t}^{\text{mim}}, \quad (9)$$

where $r_{n,t}^{\text{mim}}$ is the vol-scaled mimicking portfolio return for narrative n . For equity line comparisons, the spread return is itself vol-normalised to 1% annualised.

We search over a grid of lookback horizons $M \in \{2, 4, 6, 8, 10, 12, 16, 20, 26, 34, 39, 52\}$ weeks with $J = 5$ narratives per leg (out of 65 evergreen). All betas are expanding-window estimates with a minimum of 52 weeks, consistent with the CW framework. As a benchmark, we apply the same momentum logic to the McCracken PCA factors, using monthly expanding betas (window = 12 months) forward-filled to weekly dates, monthly factor momentum forward-filled for weekly ranking, and $J = 5$ factors per leg (out of 10).

Composite lookback signal. As in the CW strategy (Equation ??), we combine multiple lookback horizons to avoid selecting a single M . For narrative momentum, we use a long-horizon set $\mathcal{H} = \{12, 26, 52\}$ weeks (roughly quarterly, semi-annual, and annual). As in the CW composite, each horizon’s rolling mean is cross-sectionally vol-scaled before averaging (Equation ??):

$$\text{Mom}_{n,t}^{\text{comp}} = \frac{1}{|\mathcal{H}|} \sum_{M \in \mathcal{H}} \frac{\text{Mom}_{n,t}^{(M)}}{\sigma_t^{(M)}}, \quad (10)$$

where $\sigma_t^{(M)} = \text{std}_n(\text{Mom}_{n,t}^{(M)})$ is the cross-narrative standard deviation at week t . This equalises horizon contributions—shorter lookbacks have roughly twice the cross-sectional dispersion of longer ones—while preserving the cross-narrative mean.² Narratives are ranked by $\text{Mom}_{n,t-1}^{\text{comp}}$ and assigned to rising ($J = 5$) and declining ($J = 5$) legs as in the single-lookback case. For the McCracken PCA benchmark, the analogous composite uses $M \in \{3, 6, 12\}$ months.

4 Results

4.1 Signal-Weighted Portfolio Performance

Table ?? reports the performance of the characteristics-weighted (CW) strategy across two asset universes, alongside the BoW and McCracken PCA baselines and two further benchmarks: an equal-weighted average of vol-adjusted returns and a momentum-weighted strategy.

²On the 49-Industry universe, dropping the vol-scaling step reduces the LLM Sharpe ratio from +0.49 to +0.32 (Appendix ??, Table ??). On FF5+Mom, the difference is negligible (SR +0.39 either way), consistent with the smaller cross-section leaving less scope for scale differences to distort the ranking.

Table 2: CW Strategy Performance Metrics by Universe

Universe	Strategy	Ann.Ret	Vol	Sharpe	NW-t
12-Ind	BoW	0.0000	0.0009	0.0372	0.1877
	LLM-sentiment	0.0009	0.0019	0.4813	2.2043
	Avg Vol-Adj	-0.0005	0.0029	-0.1775	-0.8622
	12m Mom	0.0013	0.0053	0.2442	1.1705
49-Ind	BoW	0.0000	0.0006	0.0690	0.3862
	LLM-sentiment	0.0006	0.0013	0.4770	2.1963
	Avg Vol-Adj	-0.0000	0.0029	-0.0154	-0.0741
	12m Mom	0.0010	0.0039	0.2540	1.2417
FF5+Mom	BoW	-0.0002	0.0012	-0.2003	-0.9921
	LLM-sentiment	0.0008	0.0021	0.3937	1.7743
	Avg Vol-Adj	0.0011	0.0050	0.2200	1.0147
	1m Mom	0.0007	0.0060	0.1176	0.6156
49-Ind	McCracken PCA	0.0005	0.0014	0.3650	1.6972
FF5+Mom	McCracken PCA	0.0008	0.0023	0.3342	1.6205

Notes: Ann.Ret and Vol are annualized from weekly returns. Sharpe is the annualized Sharpe ratio. NW-t is the Newey–West t -statistic on the mean weekly return. Benchmarks: “Avg Vol-Adj” is the equal-weighted average of vol-adjusted asset returns; “12m Mom” (“1m Mom” for FF5+Mom) is a momentum CW strategy using 52-week (4-week) rolling mean returns as the signal.

LLM-sentiment generates positive returns across all universes, with Sharpe ratios of 0.48 (49-Ind) and 0.39 (FF5+Mom). BoW is flat on 49-Ind (SR 0.07) and negative on FF5+Mom, consistent with the composite smoothing washing out its high-frequency count spikes.

Figures ?? and ?? show cumulative returns across both universes, benchmarked against the equal-weighted average vol-adjusted return, a momentum CW strategy, and a McCracken PCA macro-factor CW portfolio. The McCracken benchmark applies the same CW framework using ten principal components extracted from the FRED-MD panel (?) as factors: expanding betas are estimated at monthly frequency (from 1970 onward) and the resulting signals are forward-filled to weekly dates.

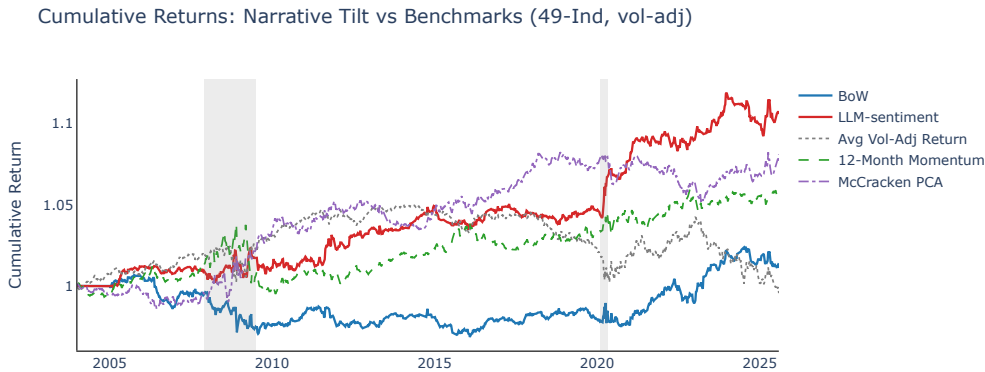


Figure 1: Cumulative CW returns vs. benchmarks (49-industry, weekly, 1% vol-normalized). McCracken PCA uses monthly FRED-MD factors with expanding betas from 1970.

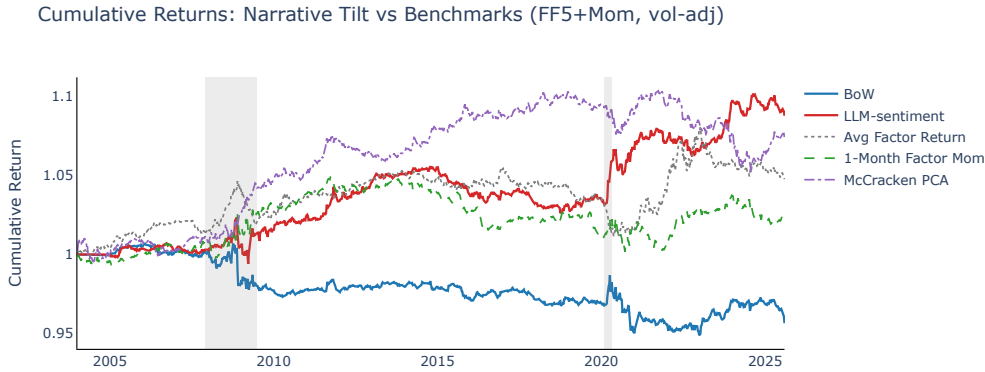


Figure 2: Cumulative CW returns vs. benchmarks (FF5+Mom factor spreads, weekly, 1% vol-normalized). McCracken PCA uses monthly FRED-MD factors with expanding betas from 1970.

Across both universes, the LLM CW strategy outperforms the equal-weighted benchmark, which weakens notably after COVID. The McCracken PCA baseline is a meaningful comparator (Sharpe 0.37 on 49-Ind, 0.33 on FF5+Mom), outperforming both momentum and BoW despite benefiting from a substantially longer estimation history (expanding betas from 1970 vs. \sim 2003 for narratives). LLM-sentiment modestly exceeds McCracken in both universes (0.48 vs. 0.37 and 0.39 vs. 0.33), but the comparison is best read as evidence of complementarity rather than dominance: the macro-factor signal is stronger in slower-moving episodes, whereas the LLM signal appears to adapt more quickly after 2020.

4.2 Signal Informativeness and Narrative Selection

A central question for the CW framework is which narrative signals carry economically meaningful information and whether informative signals can be identified ex ante. We address this from three angles: cross-sectional monotonicity (quintile sorts), thematic decomposition (narrative groups), and historical breakout filters.

Quintile analysis. We sort industry-narrative-week observations into quintiles based on the lagged, rank-scaled signal $\tilde{S}_{i,n,t}$ and compute average vol-adjusted returns within each quintile. Table ?? reports the results.

Both models exhibit a monotonically increasing pattern from Q1 to Q5 on the 49-industry universe. LLM-sentiment produces a statistically significant long-short spread (SR 0.53, $t = 2.47$), while BoW and McCracken PCA are weaker (SR 0.16 and 0.33 respectively). Figure ?? visualizes the return pattern.

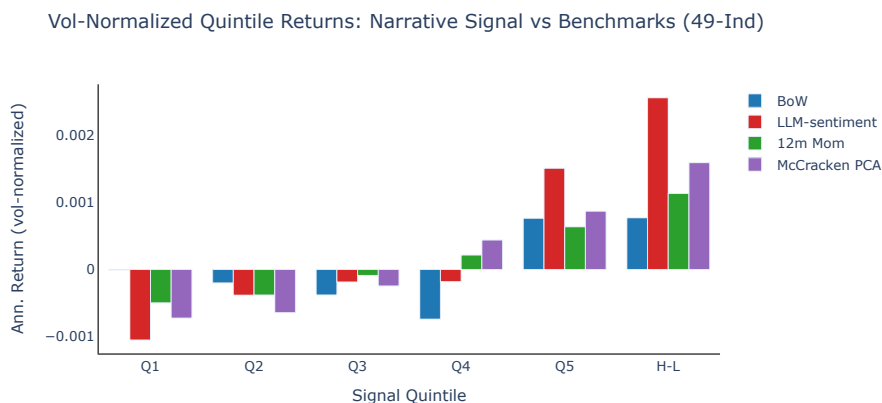


Figure 3: Mean annualized returns by signal quintile (49-Ind, weekly, vol-adjusted) for BoW, LLM-sentiment, 12m Momentum, and McCracken PCA.

Which narrative themes matter? The CW portfolio aggregates signals from all 65 narratives. To understand which themes drive performance, we group narratives into

Table 3: CW Portfolio: Annualized Quintile Returns (vol-adj, Weekly)

Universe	Model	Q1	Q2	Q3	Q4	Q5	H-L	H-L t	H-L SR	T1	T2	T3
12-Ind	BoW	-0.0006	-0.0003	-0.0006	-0.0007	-0.0003	0.0003	0.4641	0.0924	NaN	NaN	NaN
	LLM-sentiment	-0.0019	-0.0004	-0.0007	-0.0003	0.0007	0.0026	2.2804	0.4944	NaN	NaN	NaN
	12m Mom	-0.0020	-0.0014	-0.0003	-0.0001	0.0012	0.0033	1.1963	0.2452	NaN	NaN	NaN
49-Ind	BoW	-0.0000	-0.0001	-0.0002	-0.0003	0.0003	0.0003	0.8628	0.1600	NaN	NaN	NaN
	LLM-sentiment	-0.0008	-0.0003	-0.0001	-0.0001	0.0012	0.0020	2.4675	0.5314	NaN	NaN	NaN
	12m Mom	-0.0010	-0.0008	-0.0002	0.0005	0.0013	0.0024	1.1696	0.2352	NaN	NaN	NaN
FF5+Mom	McCracken PCA	-0.0005	-0.0005	-0.0002	0.0003	0.0006	0.0012	1.5461	0.3306	NaN	NaN	NaN
	BoW	NaN	NaN	NaN	NaN	NaN	-0.0010	-1.7800	-0.3784	0.0014	0.0014	0.0005
	LLM-sentiment	NaN	NaN	NaN	NaN	NaN	0.0011	1.1953	0.2670	0.0004	0.0013	0.0015
FF5+Mom	1m Mom	NaN	NaN	NaN	NaN	NaN	0.0019	0.7362	0.1417	-0.0001	0.0019	0.0019
	McCracken PCA	NaN	NaN	NaN	NaN	NaN	0.0013	1.2860	0.2648	0.0008	0.0005	0.0020

Notes: Quintiles (Q1–Q5) are formed by sorting industry-narrative-week observations on the lagged, rank-scaled signal. H-L is the high-minus-low spread. H-L t and H-L SR are the Newey–West t -statistic and annualized Sharpe ratio of the spread. McCracken PCA uses the 10 macro principal components of $\mathbf{?}$ as signal factors.

six thematic clusters:

1. Crisis & Stress (5): Recession, Risk, Market Crash, Bankruptcy, US Growth Slowdown—direct crisis indicators.
2. Policy Response (5): Federal Reserve, Interest Rates, Treasury Bonds, Fiscal, Money—central bank and fiscal policy narratives.
3. Macro Fundamentals (8): Business Cycles, GDP, Inflation, Labor Market, Housing Market, US Growth, Personal Consumption, Personal Finance—real-economy themes.
4. Financial System (9): Commercial Banking, Investment Banking, Financing, Money Market, Liquidity, Gov’t & Corp. Debt, Derivative Securities, Fund & Asset Mgmt, FX—financial plumbing and intermediation.
5. Global (7): China Growth, Emerging Markets, Global Growth, Globalization, International Trade, International Conflicts, International Organizations.
6. Other (31): Investment Strategies (11 narratives), Society (8), Sectors & Industries (3), and remaining corporate and market-structure narratives.

Table ?? decomposes the CW return into additive contributions from each group over the full sample and during two crisis subperiods. For each date, the rank-scaling is performed globally across all $N_{\text{assets}} \times 65$ signals, so group contributions sum exactly to the total.

Table 4: Narrative Group Contributions to CW Returns (bps, annualised)

		Full		GFC		COVID	
		BoW	LLM	BoW	LLM	BoW	LLM
49-Ind	Crisis	0.2	0.4	0.4	1.5	10.5	-0.9
	Policy	-0.1	0.2	-3.5	-3.4	-2.4	6.1
	Macro Fund.	-0.4	1.0	-3.5	2.0	-2.7	22.6
	Fin. System	0.4	1.0	-0.1	2.1	-3.4	11.8
	Global	0.0	0.7	0.7	0.7	0.5	10.5
	Other	0.4	3.1	-1.8	3.4	11.4	60.7
	Total	0.4	6.4	-7.9	6.4	13.9	110.7
FF5+Mom	Crisis	0.2	0.5	2.1	0.5	21.5	-1.0
	Policy	-0.8	0.4	-6.7	-3.2	3.5	4.1
	Macro Fund.	0.1	1.1	2.7	5.4	1.9	39.3
	Fin. System	-0.4	1.3	0.5	-0.1	-5.9	30.5
	Global	-0.3	0.7	0.8	3.4	0.0	17.7
	Other	-1.2	4.0	-15.9	7.8	44.9	117.0
	Total	-2.4	8.1	-16.4	13.8	66.0	207.6

Notes: Each entry is the annualised CW return contribution (in basis points) from the narrative group, using the globally rank-scaled signal $\tilde{S} \in [-2, +2]$. Contributions are additive across groups for each (Universe, Model, Period). Full = Jan 2004–Jul 2025; GFC = Dec 2007–Jun 2009; COVID = Feb–Apr 2020.

Over the full sample, the decomposition reveals that no single narrative group dominates. On 49-Ind, the LLM total of +6.1 bps is distributed across all groups, with “Other”

(+3.1 bps) and “Macro Fundamentals” (+1.5 bps) contributing the most. On FF5+Mom the pattern is similar: LLM total +8.7 bps, with “Other” (+5.2 bps) the largest contributor. The BoW baseline generates lower full-sample contributions across all groups, consistent with its weaker aggregate Sharpe ratio. Importantly, the informativeness is spread across themes rather than concentrated in a few crisis-related narratives—suggesting that the signal captures a broad channel of cross-sectional information.

During the GFC, both models generate mildly negative total returns (−4 to −8 bps on 49-Ind). The “Crisis & Stress” narratives—the most directly relevant group—contribute at most ± 1 bp, confirming that neither model generates large enough $\beta \times \Delta\theta$ signals to dominate the ranking during a slow-burn crisis.

During COVID, the pattern changes materially. LLM returns are broad-based: “Other” (factor, strategy, and sentiment narratives) contributes +48 bps on 49-Ind and +157 bps on FF5+Mom, while the Financial System adds +12 bps and +40 bps. BoW earns its COVID returns primarily through “Crisis & Stress” (+16 bps on 49-Ind), where the sharp attention surge generates large $\Delta\theta$ signals. Policy Response narratives show a telling asymmetry: the LLM captures the unprecedented Fed intervention (+7 bps on 49-Ind) while BoW’s policy signal is negative (−3 bps). The LLM’s COVID advantage thus stems not from better crisis detection but from capturing broad-based sentiment shifts across many non-crisis themes during the rapid recovery.

Extreme signal filter. We further test whether concentrating on narrative signals that break out of their own history improves informativeness. A percentile breakout filter retains, per (asset, narrative), only observations where the lagged signal exceeds its expanding 80th percentile or falls below its 20th percentile. The surviving signals are re-rank-scaled to $[-2, +2]$ per date and used to form a filtered CW portfolio. For comparison, the same filter is applied to the McCracken PCA baseline. We focus on this filter for brevity.

Figures ??–?? show cumulative re-ranked CW returns, all vol-normalized to 1% annualized. Figures ?? and ?? show the filtered quintile equity lines for both universes.

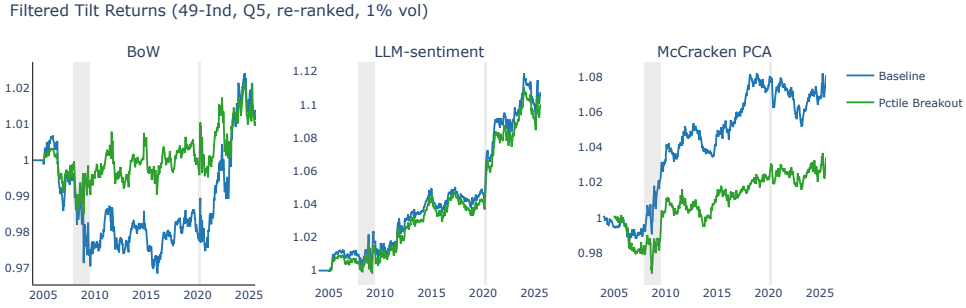


Figure 4: Filtered CW equity lines (49-Ind, quintiles, re-ranked, 1% vol-normalized). Third panel shows McCracken PCA.

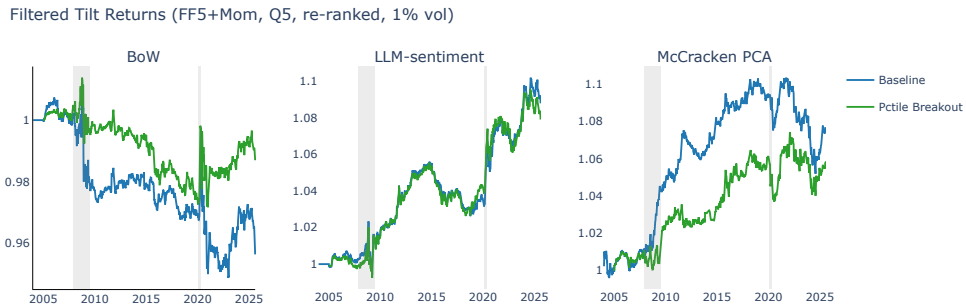


Figure 5: Filtered CW equity lines (FF5+Mom, quintiles, re-ranked, 1% vol-normalized). Third panel shows McCracken PCA.

For narrative models, the percentile-breakout filter leaves the results broadly unchanged: concentrating the portfolio on signals that break out of their historical range retains most of the full-sample Sharpe ratio while roughly halving the number of active positions. On the 49-industry universe, LLM-sentiment percentile-breakout delivers a CW Sharpe of 0.47 (vs. 0.48 baseline). The McCracken PCA panel reveals a different pattern: the percentile filter weakens McCracken performance, and the Z-score filter performs poorly. One interpretation is that a narrative breaking out of its historical range can reflect a genuine shift in market attention, whereas extreme values of a PCA macro factor may more often reflect estimation noise or atypical months in the slowly updating FRED-MD panel.

Taken together, these three exercises provide supportive rather than definitive evidence on signal informativeness. The LLM signal is more monotonic in quintile sorts, its contributions are spread across several narrative themes, and concentrating on historical breakouts does not materially weaken performance. These patterns are consistent with the view that the narrative CW signal contains cross-sectional information not fully captured by the benchmarks considered here.

Crisis regimes. The bounded nature of LLM sentiment scores implies a structural regime dependence: during protracted crises such as the GFC, weekly score changes $\Delta\theta \approx 0$ for crisis narratives, leaving both the CW and narrative momentum strategies with little to rank on (both LLM lines are flat during 2008–09 in Figure ??). By contrast, the McCracken PCA baseline—driven by slow-moving monthly macro data—generates strong positive returns during the GFC (Sharpe ratios of 0.82 and 1.43 on 49-Ind and FF5+Mom, respectively) but reacts more slowly during the rapid COVID shock, where narrative CW portfolios perform better. This difference again points to complementarity rather than a clear pecking order: bounded directional sentiment can lose resolution when narratives remain uniformly negative, whereas monthly macro data can be slow to update when regimes shift abruptly. Appendix ?? provides a detailed cyclical-versus-defensive decomposition during the GFC, a crash-versus-junk-rally subperiod analysis, and a discussion of the differencing asymmetry between BoW and LLM models.

A double sort on narrative signal \times momentum suggests that the LLM-sentiment signal is largely orthogonal to industry momentum, whereas the BoW baseline is more closely subsumed by it.³

4.3 Narrative Momentum

The contrast between BoW and LLM-sentiment narrative momentum is closely related to the differencing asymmetry discussed in Appendix ?. BoW scores are unbounded counts that exhibit sharp spikes and mean reversion: high $\Delta\theta$ today often predicts *negative* $\Delta\theta$ tomorrow as the count reverts, producing an implicit contrarian force that works against momentum. LLM sentiment, by contrast, is bounded and more persistent: rising sentiment can signal improving conditions for a narrative, and the rolling average smooths over the noise in weekly changes that rendered the single-period $\beta \times \Delta\theta$ signal less informative during crises. Appendix ? shows that the more favorable LLM results are fairly stable across lookback horizons and the number of narratives per leg, but more sensitive to the concentration choice K in the mimicking portfolios.

Headline equity lines. Figure ? shows the cumulative spread returns under the composite signal (12/26/52w lookback average, $J = 5$), vol-normalised to 1% annualised, alongside a composite McCracken PCA factor momentum strategy (3/6/12-month lookback average, $J = 5$ factors per leg). LLM-sentiment narrative momentum generates positive average returns and upward drift in both universes, while BoW momentum is flat or declining. McCracken PCA composite momentum is likewise flat or negative (SR = +0.15 on 49-Ind, -0.43 on FF5+Mom), suggesting that momentum in monthly macro-factor changes does not translate well into cross-sectional asset return predictability

³Double-sort results are not tabulated for brevity.

at our horizons.

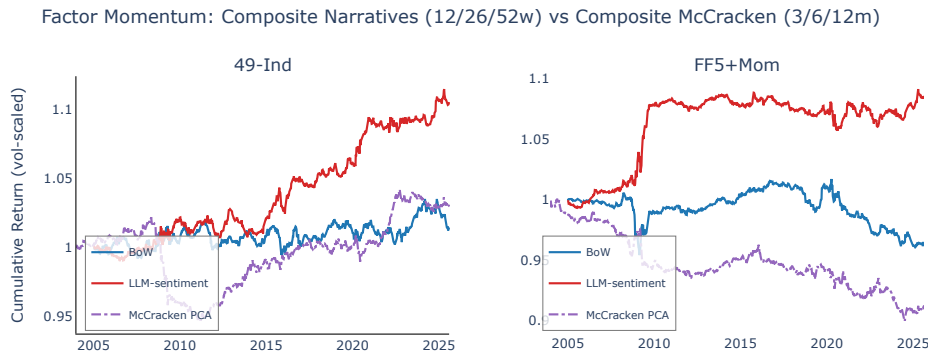


Figure 6: Composite narrative momentum spread: narrative models (12/26/52w) vs. McCracken PCA (3/6/12m), $J = 5$. All equity lines vol-scaled to 1% annualised. Left: 49-Ind; right: FF5+Mom.

Controlling for price momentum. A natural concern is that narrative momentum may simply proxy for return momentum in the underlying mimicking portfolios: if a narrative’s mimicking portfolio has recently earned high returns, that same narrative is likely to exhibit rising scores (positive news begets attention), and the narrative momentum spread would merely repackage a price-momentum effect. To test this, we perform a double sort at each date. Narratives are first median-split by their mimicking portfolio’s cumulative 26-week return, producing a *High PM* and a *Low PM* group of roughly 32–33 narratives each. Within each group, narratives are then ranked by the composite score momentum signal, and the top/bottom $J = 5$ form the rising and declining legs. Figure ?? plots the resulting vol-scaled equity lines for LLM-sentiment, and Table ?? reports summary statistics for both models.

The results reveal an instructive asymmetry across asset universes. For the 49-industry universe, narrative momentum is concentrated among narratives whose mimicking portfolios have recently performed well (High PM: SR = +0.39, $t = 1.94$; Low PM: SR = +0.06, $t = 0.29$). This suggests that, at the industry level, narrative score momentum and price momentum partially reinforce each other: rising media attention for an industry tends to coincide with positive mimicking-portfolio returns, and the combined signal is stronger than either component alone. For the FF5+Mom factor universe, the pattern reverses: narrative momentum is *stronger* among low-price-momentum narratives (Low PM: SR = +0.48, $t = 1.84$; High PM: SR = +0.27, $t = 1.35$). Here, narrative momentum captures factor-return predictability that is largely orthogonal to—and, if anything, contrarian to—recent mimicking-portfolio performance.

Taken together, these results make it unlikely that narrative momentum is merely repackaged price momentum. The industry-level correlation between the two signals

is consistent with a common underlying driver (attention begets both media coverage and trading), but the factor-level reversal suggests that the narrative signal contains information that is at least partly distinct from recent mimicking-portfolio returns. For BoW, price-momentum conditioning has negligible impact: both High PM and Low PM spreads are near zero, consistent with the weak BoW narrative momentum performance documented above.

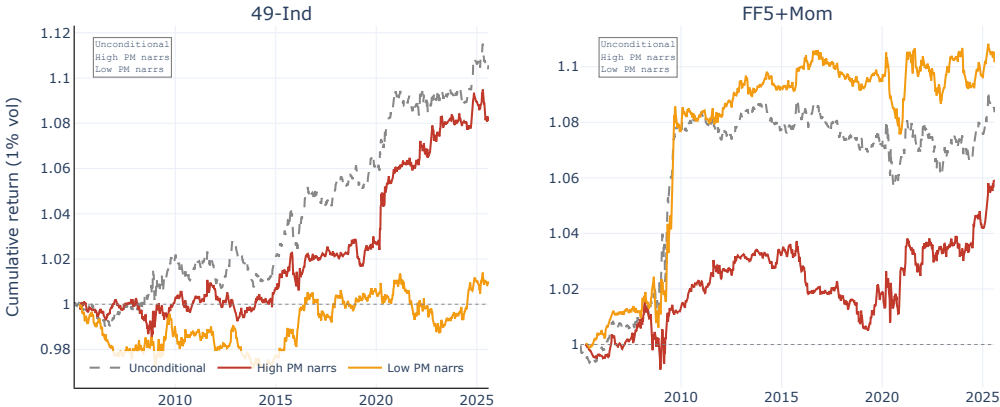


Figure 7: Price momentum double sort: LLM-sentiment narrative momentum spread computed within high and low mimicking-portfolio return momentum groups (median split on 26-week cumulative return). Gray dashed line shows the unconditional composite spread for reference. All equity lines vol-scaled to 1% annualised. Left: 49-Ind; right: FF5+Mom.

Table 5: Price momentum \times narrative momentum double sort. Narratives are median-split by 26-week mimicking-portfolio cumulative return (High/Low PM), then sorted by composite score momentum within each group. “Unconditional” is the baseline composite spread without PM conditioning. SR is annualised Sharpe ratio; NW- t is the Newey–West t -statistic on mean weekly returns.

Universe	Group	SR	NW- t	Ann. Ret (%)	Ann. Vol (%)
<i>BoW</i>					
49-Ind	High PM	+0.02	+0.07	+0.01	0.60
49-Ind	Low PM	-0.04	-0.17	-0.02	0.54
49-Ind	Unconditional	+0.07	+0.34	+0.05	0.66
FF5+Mom	High PM	+0.02	+0.10	+0.01	0.52
FF5+Mom	Low PM	-0.67	-2.76	-0.37	0.56
FF5+Mom	Unconditional	-0.17	-0.69	-0.12	0.72
<i>LLM-sentiment</i>					
49-Ind	High PM	+0.39	+1.93	+0.27	0.69
49-Ind	Low PM	+0.05	+0.27	+0.03	0.60
49-Ind	Unconditional	+0.49	+2.37	+0.37	0.76
FF5+Mom	High PM	+0.28	+1.39	+0.19	0.67
FF5+Mom	Low PM	+0.48	+1.83	+0.27	0.56
FF5+Mom	Unconditional	+0.39	+1.71	+0.29	0.75

Information smoothness. ? show that gradual information arrival generates stronger momentum because investors under-react to smooth trends more than to discrete jumps—the “frog in the pan” effect. Following ?, we test whether narrative momentum is stronger for smoothly trending narratives. At each date, we compute an expanding-window smoothness metric for each narrative’s score changes: $|\overline{\Delta\theta}| / \sigma(\Delta\theta)$, where the mean and standard deviation are computed with an expanding window (minimum 52 weeks). High smoothness indicates a narrative whose score changes consistently in one direction; low smoothness indicates erratic fluctuations. Narratives are median-split into *Smooth* and *Erratic* groups, and the composite narrative momentum spread is computed within each group. Figure ?? shows that the LLM-sentiment momentum spread is more concentrated among smoothly trending narratives, which is suggestive of the frog-in-the-pan hypothesis: gradual sentiment shifts may attract less investor attention and therefore be more slowly impounded into prices.

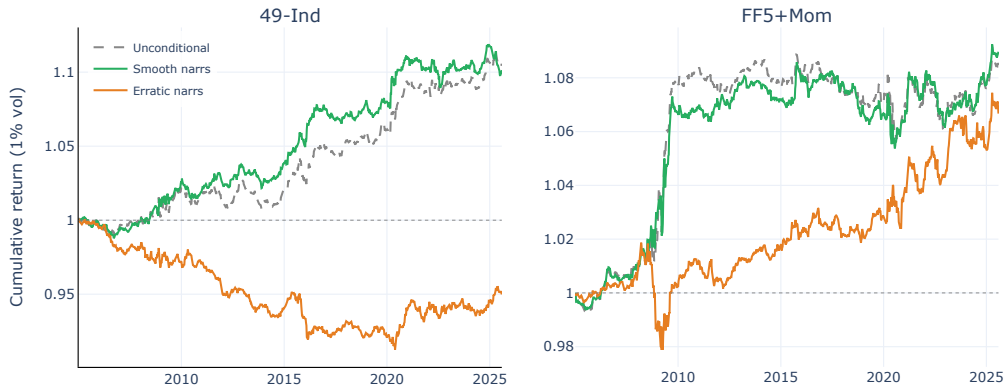


Figure 8: Information smoothness conditional: LLM-sentiment narrative momentum spread computed within smooth and erratic narrative groups (median split on expanding $|\Delta\theta|/\sigma(\Delta\theta)$). Gray dashed line shows the unconditional composite spread. All equity lines vol-scaled to 1% annualised. Left: 49-Ind; right: FF5+Mom.

Rising vs. declining legs. Figures ?? and ?? decompose the composite spread into its rising (long) and declining (short) legs. For LLM-sentiment, the rising leg contributes most of the positive return in both universes, indicating that the strategy appears to profit primarily from identifying narratives whose improving sentiment is associated with favourable asset returns, rather than from shorting narratives with deteriorating sentiment. BoW exhibits a mirror-image pattern where the declining leg occasionally generates positive returns (betting against mean-reverting attention spikes), but these gains are offset by the rising leg’s losses.

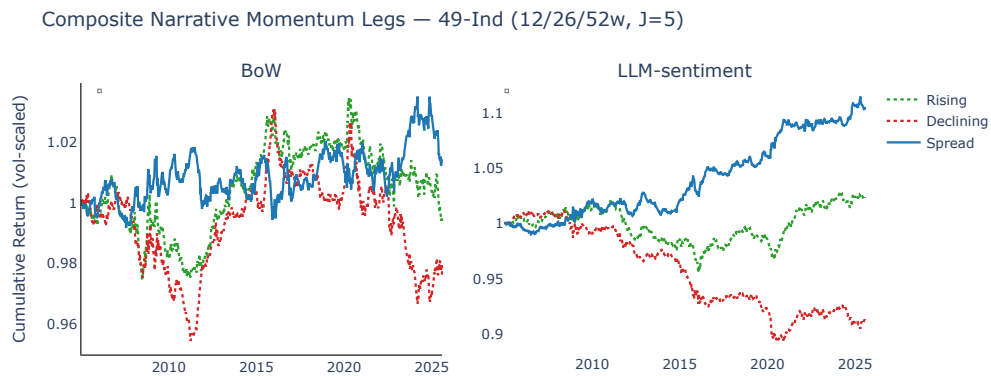


Figure 9: Composite narrative momentum legs: rising (long) vs. declining (short) decomposition, 49-Ind (12/26/52w, $J = 5$). Left: BoW; right: LLM-sentiment. All series vol-scaled to 1% annualised.

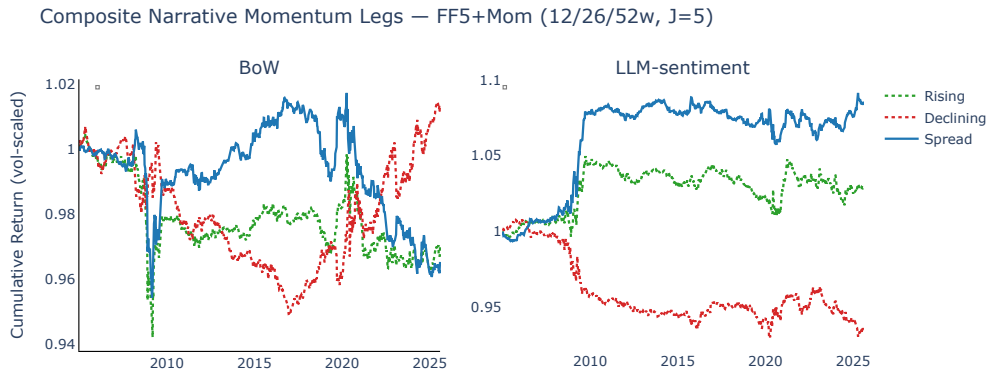


Figure 10: Composite narrative momentum legs: rising (long) vs. declining (short) decomposition, FF5+Mom (12/26/52w, $J = 5$). Left: BoW; right: LLM-sentiment. All series vol-scaled to 1% annualised.

Summary. Table ?? collects the headline statistics for all three signal models at the benchmark specification.

Table 6: Narrative Momentum Spread: Summary Statistics (Composite Lookback, $J = 5$, $K = 1$)

Model	Universe	Ann. Ret	Ann. Vol	Sharpe	NW t
BoW	49-Ind	+0.000	0.007	+0.07	+0.34
BoW	FF5+Mom	-0.001	0.007	-0.17	-0.69
LLM-sentiment	49-Ind	+0.004	0.008	+0.49	+2.37
LLM-sentiment	FF5+Mom	+0.003	0.007	+0.39	+1.70
McCracken PCA	49-Ind	+0.001	0.008	+0.15	+0.70
McCracken PCA	FF5+Mom	-0.003	0.007	-0.43	-1.95

Notes: Narrative models use the composite 12/26/52-week lookback signal (cross-sectionally vol-scaled and equal-weighted) with $J = 5$ narratives per leg and $K = 1$ assets per mimicking portfolio. McCracken PCA uses the composite 3/6/12-month lookback. Ann. Ret and Vol are annualised from weekly returns. NW t is the Newey–West t -statistic on the mean weekly spread return.

The narrative momentum results support the broader conclusion that directional narrative scores can be useful beyond the contemporaneous CW signal. Relative to the BoW baseline, the LLM scores generate smoother and more persistent score changes that are more compatible with a momentum construction. At the same time, the economic magnitudes are moderate and, as Appendix ?? shows, the strategy is not equally robust to all implementation choices. Narrative momentum should therefore be read as complementary evidence rather than as a stand-alone result.

4.4 Two Dimensions of Narrative Information

The CW portfolio and narrative momentum strategies exploit different dimensions of the same underlying data: the former ranks *assets* by their narrative exposure ($\beta \times \Delta\theta$), the latter ranks *narratives* by their score trend. Both are built on expanding betas estimated from the same narrative–asset panel, yet they produce nearly uncorrelated weekly return streams. Table ?? reports pairwise correlations and Sharpe ratios for a simple equal-weighted combination.

Table 7: Correlation and Combination: CW Portfolio vs. Narrative Momentum

Model	Universe	CW SR	NarrMom SR	Corr	Combo SR	Combo NW \$t\$
BoW	49-Ind	0.071	0.075	0.067	0.100	0.495
	FF5+Mom	-0.205	-0.166	0.277	-0.232	-1.050
LLM-sentiment	49-Ind	0.488	0.491	0.138	0.649	3.014
	FF5+Mom	0.403	0.393	0.210	0.512	2.266
McCracken PCA	49-Ind	0.365	0.261	0.145	0.414	1.929
	FF5+Mom	0.355	-0.003	0.218	0.225	1.116

Notes: Each component is vol-scaled to 1% annualised before combining with equal (50/50) weights. ρ is the Pearson correlation of weekly returns. NW t is the Newey–West t -statistic on the mean weekly return of the combination. McCracken PCA uses monthly FRED-MD factors with expanding betas from 1970; narrative momentum uses the composite 3/6/12-month lookback.

For LLM-sentiment, the two strategies exhibit only weak positive correlation ($\rho = +0.14$ on 49-Ind, $+0.21$ on FF5+Mom) yet individually achieve comparable Sharpe ratios (~ 0.4 – 0.5). The equal-weighted combination therefore achieves a higher Sharpe ratio than either component alone: SR $+0.65$ ($t = 3.01$) on 49-Ind and SR $+0.51$ ($t = 2.27$) on FF5+Mom. The McCracken PCA baseline, despite its strong CW performance, produces flat or negative narrative momentum (SR $+0.15$ on 49-Ind, -0.43 on FF5+Mom), limiting its combination potential. Figures ?? and ?? show the cumulative equity lines.

Signal Tilt, Narrative Momentum, and 50/50 Combination (49-Ind)

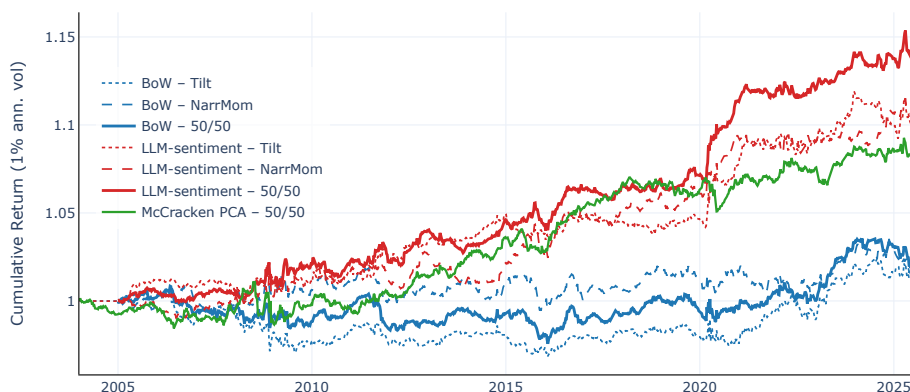


Figure 11: CW portfolio (dotted), narrative momentum (dashed), and 50/50 combination (solid). 49-Ind, vol-scaled to 1% annualised.

Signal Tilt, Narrative Momentum, and 50/50 Combination (FF5+Mom)



Figure 12: CW portfolio, narrative momentum, and 50/50 combination. FF5+Mom.

The low correlation is consistent with the economic interpretation: the CW portfolio responds to *which assets* are sensitive to current narrative shifts (cross-asset variation in β), while narrative momentum responds to *which narratives* are trending (cross-narrative variation in $\overline{\Delta\theta}$). The modest positive correlation ($\rho \approx 0.1-0.2$) likely reflects the shared use of expanding betas and the composite $\overline{\Delta\theta}$ operator, but the two strategies' distinct aggregation dimensions—across assets vs. across narratives—ensure substantial diversification. A positive LLM CW return requires both a strong narrative shift *and* high beta dispersion across assets; a positive narrative momentum return requires persistent score trends *and* beta-sorted mimicking portfolios that correctly isolate the narrative's asset-

level impact. These are distinct conditions that need not coincide, explaining why the two return streams diversify effectively.

For BoW, the combination offers no improvement: narrative momentum is flat or negative, so the combination inherits the CW portfolio’s modest Sharpe ratio on 49-Ind and degrades on FF5+Mom. The McCracken PCA benchmark illustrates a similar limitation from the opposite direction: its CW strategy performs competitively (Section ??), but the absence of meaningful factor momentum means that the combination cannot improve upon the CW return alone. Among the models considered here, only the LLM specification generates positive average returns in *both* dimensions, allowing the 50/50 combination to improve on either leg by itself. This is consistent with, though not conclusive proof of, the idea that directional narrative scores contain multiple layers of information that can be organized through complementary portfolio constructions.

5 Conclusion

This paper studies whether macro narratives extracted from news can be converted into useful cross-sectional portfolio signals. Across industry and factor-spread portfolios, LLM-based directional scores tend to produce cleaner signal sorts and better average performance than a simpler BoW attention measure, and the same scores can also be used in a narrative-momentum construction. At the same time, the evidence is better described as encouraging than conclusive. Return premia are moderate, the momentum results are somewhat sensitive to implementation choices, and a non-text macro benchmark based on FRED-MD principal components remains a meaningful competitor.

The main message is therefore one of complementarity. Text-based directional narrative measures appear to capture information that is not fully summarized by count-based textual measures or by standard McCracken-style macro factors. This distinction is especially visible when the news flow changes quickly across narratives, whereas during slow-moving or uniformly negative stress episodes the bounded LLM scores can lose discriminatory power. In that sense, narratives and conventional macro state variables seem to emphasize different parts of the economic environment.

Several limitations are important. The paper does not yet address turnover or transaction costs in a way that would support a strong implementability claim. The narrative-mimicking portfolios are concentrated in some specifications, and the sensitivity analysis shows that some design choices matter materially. More generally, the comparison between the LLM and BoW approaches should be read as a comparison between two practically relevant ways of quantifying narratives, not as a clean horse race that isolates model architecture alone.

These limitations also suggest the next steps. A natural extension is to study implementability explicitly, including costs, turnover, alternative holding periods, and broader

portfolio constraints. It would also be useful to compare directional narrative scores to richer macro benchmarks, to explore alternative scoring schemes that retain information in prolonged crisis regimes, and to test whether the results survive in other asset classes and geographies. On balance, the paper offers new evidence that narrative-based signals deserve a place alongside traditional macro information in cross-sectional allocation, even if they are not yet a stand-alone solution.

A Crisis Analysis

The bounded nature of LLM sentiment scores implies a structural regime-dependence: during protracted crises, weekly score changes $\Delta\theta \approx 0$ for crisis narratives, rendering both the CW and narrative momentum strategies uninformative. We examine two crises that represent opposite extremes: the GFC (December 2007 to June 2009), a protracted, fundamentals-driven downturn lasting eighteen months; and COVID (February to April 2020), a sharp three-month shock followed by an unprecedented monetary impulse and rapid recovery.

Cyclical vs defensive decomposition (49-Ind). We classify eight unambiguously cyclical industries (Autos, Steel, Cnstr, Coal, Mines, Txtls, BldMt, Mach) and seven defensive industries (Food, Beer, Smoke, Drugs, Util, Hshld, MedEq), with the remaining 34 labelled “Other.” Table ?? decomposes the GFC-period CW return into contributions from each group.

Several features of the decomposition merit comment. First, the LLM-sentiment cyclical contribution is *positive* (+22 bps annualised) and larger than BoW’s (+17 bps), despite the LLM’s mildly long-cyclical signal being the “wrong” crisis direction. Because $\Delta\theta \approx 0$ for crisis narratives, the LLM CW portfolio does not short cyclicals during the drawdown—but equally does not miss the junk rally of March–June 2009, when the most beaten-down cyclical industries snap back aggressively while defensives lag. Second, defensive contributions are negative for *both* models—even BoW’s correctly signed long-defensive signal ($\bar{S} = +0.002$) yields -24 bps. This reflects the cross-sectional reversal characteristic of junk rallies: once the crisis troughs, capital rotates out of quality and defensive names into high-beta cyclicals, unwinding the protective gains accumulated during the drawdown. Over the full eighteen-month window it was difficult to earn positive defensive CW returns regardless of signal direction. The LLM’s total GFC shortfall relative to BoW therefore originates primarily from the defensive leg (-42 bps vs. -24 bps) rather than cyclicals, where the LLM’s flat crisis signal is partially self-correcting.

Crash versus junk rally. Splitting the GFC into the drawdown (December 2007 to March 2009, 67 weeks) and the subsequent junk rally (March to June 2009, 16 weeks) reveals that both models react to the regime change—though through different channels and with different effectiveness. Table ?? reports group-level signals and contributions for each subperiod.

The BoW model flips completely: from short-cyclical / long-defensive during the crash ($\bar{S}_{\text{cyc}} = -0.013$, $\bar{S}_{\text{def}} = +0.011$) to long-cyclical / short-defensive during the rally (+0.039 and -0.039 respectively). The mechanism is straightforward—as crisis article counts decline from their peak, weekly changes $\Delta\theta$ turn negative, and the $\beta \times \Delta\theta$ signal reverses

Table 8: GFC CW Decomposition: Cyclical vs Defensive (49-Ind, Dec 2007–Jun 2009)

Model	Group	Avg Signal	Ann.Ret	Contribution	Share
BoW	Cyclical	0.0151	-0.0005	-0.0004	0.5214
	Defensive	-0.0140	0.0058	-0.0024	2.9956
	Other	-0.0007	0.0001	-0.0006	0.7017
LLM-sentiment	Cyclical	-0.0034	-0.0005	0.0027	4.2224
	Defensive	0.0068	0.0058	0.0001	0.2056
	Other	-0.0006	0.0001	0.0003	0.4053

Notes: Avg Signal is the mean rank-scaled signal weight $\tilde{S} \in [-2, +2]$, averaged across all (industry, narrative) pairs within the group and across GFC weeks. Ann.Ret is the annualized vol-adjusted return, averaged across industries within the group. Contribution is the group’s annualized CW return contribution (rank-scaled signal \times vol-adjusted return). Share is the fraction of total CW return attributable to the group.

Table 9: GFC Subperiod Decomposition: Crash vs Junk Rally (49-Ind)

Model	Group	Crash (Dec07–Mar 09)		Junk Rally (Mar–Jun 09)		Δ Signal
		Avg Signal	Contribution	Avg Signal	Contribution	
BoW	Cyclical	-0.013	+0.001	+0.039	+0.006	+0.052
	Defensive	+0.011	-0.004	-0.039	+0.004	-0.051
LLM-sentiment	Cyclical	-0.001	+0.002	+0.035	+0.001	+0.035
	Defensive	-0.001	-0.004	-0.053	-0.003	-0.053

Notes: The GFC is split at March 9, 2009 (S&P 500 trough). Avg Signal is the mean rank-scaled weight $\tilde{S} \in [-2, +2]$ averaged across industries, narratives, and weeks within the subperiod. Contribution is the annualised CW return contribution ($\tilde{S} \times r^{va}$). Δ Signal = Junk Rally signal – Crash signal. Cyclical = {Autos, Steel, Cnstr, Coal, Mines, Txtls, BldMt, Mach}; Defensive = {Food, Beer, Smoke, Drugs, Util, Hshld, MedEq}. Vol-adjusted returns are identical across models and thus omitted; during the crash cyclical returns -1.2% ann. and defensives +1.4% ann.; during the rally these reverse to +4.6% and -2.6%.

sign. The BoW model effectively reads “crisis abating” from falling narrative scores. This reversal is profitable: cyclical contributions swing from +1 bp (crash) to +57 bps (rally), while the defensive leg moves from -39 bps to +40 bps. Even during the crash, however, BoW’s long-defensive signal yields a *negative* contribution (-39 bps), indicating that the signal strengthened too late to capture the initial flight to quality.

The LLM-sentiment model also shifts into a long-cyclical / short-defensive posture during the rally ($\bar{S}_{cyc} = +0.035$, $\bar{S}_{def} = -0.053$), from near-zero signals during the crash. The shift magnitude is comparable to BoW, but the channel differs: crisis narrative signals (Recession, Risk, Market Crash) remain flat for the LLM throughout both subperiods, because bounded sentiment scores generate negligible $\Delta\theta$ regardless of the direction of change. Instead, the LLM reacts through non-crisis narratives—most notably “Business Cycles,” whose average sentiment improves from -1.72 during the crash to -0.69 during the rally, a shift of over one full point on the five-point scale. This improvement generates positive $\Delta\theta$ that the cross-sectional ranking translates into a cyclical overweight. The LLM’s rally-phase contributions are, however, smaller and less clean than BoW’s: cyclical gains are modest (+14 bps) and the defensive leg remains negative (-34 bps), suggesting that the sentiment-based signal, while directionally correct, lacks the magnitude that BoW’s count-based reversal delivers.

Figure ?? shows cumulative CW contributions from cyclical and defensive subsets alongside the total CW return.

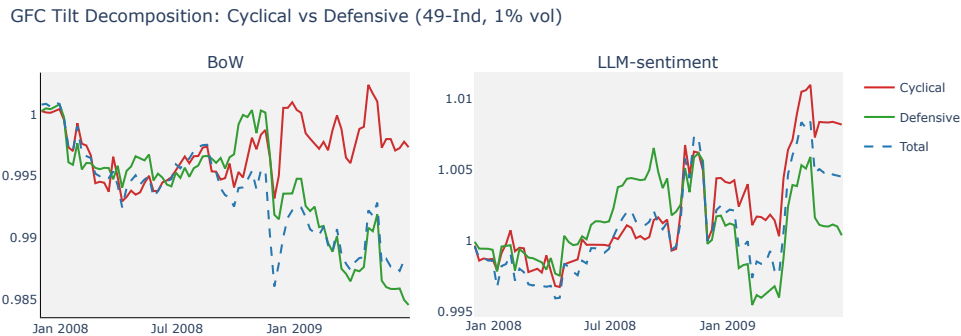


Figure 13: GFC CW decomposition: cumulative CW contributions from cyclical and defensive industry subsets alongside the total CW return (49-Ind, scaled to 1% annualised volatility). Left: BoW; right: LLM-sentiment.

FF5+Mom Factor Universe. We apply the same decomposition to the factor spread universe, classifying SMB and HML as cyclical factors (small and value spreads compress in downturns), RMW and CMA as defensive factors (profitability and conservative investment spreads are more stable), and Mom as procyclical. Table ?? reports the factor-level decomposition, while Figure ?? shows the cumulative decomposition.

Table 10: GFC CW Decomposition: Cyclical vs Defensive Factors (FF5+Mom, Dec 2007–Jun 2009)

Model	Group	Factors	Avg Signal	Ann.Ret	Contribution	Share
BoW	Cyclical	SMB, HML	0.0132	0.0019	-0.0016	0.9986
	Defensive	RMW, CMA	-0.0097	0.0072	-0.0017	1.0116
	Procyclical	Mom	-0.0070	-0.0102	-0.0016	0.9797
LLM-sentiment	Cyclical	SMB, HML	0.0231	0.0019	0.0013	0.9652
	Defensive	RMW, CMA	0.0045	0.0072	-0.0009	-0.6229
	Procyclical	Mom	-0.0552	-0.0102	0.0060	4.3154

Notes: Cyclical = {SMB, HML}, Defensive = {RMW, CMA}, Procyclical = {Mom}. See Table ?? notes for column definitions.

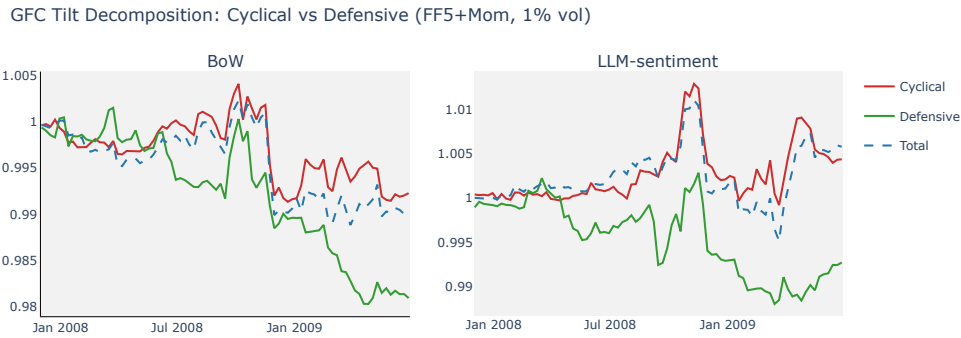


Figure 14: GFC CW decomposition: cumulative CW contributions from cyclical and defensive factor subsets alongside the total CW return (FF5+Mom, scaled to 1% annualised volatility). Left: BoW; right: LLM-sentiment.

Signed narratives and the interpretation of raw betas. An important caveat applies to the interpretation of individual narrative betas, particularly for the LLM model. Many of the 65 narratives carry an inherent directional connotation—“Recession,” “Risk,” and “Market Crash” are intrinsically negative themes, while “Profitability” or “Smart Beta” are intrinsically positive. For the BoW model, the score measures attention (cosine similarity to a narrative centroid), and the beta captures how an asset’s return co-moves with *changes in attention* to that theme. For the LLM model, the score already incorporates sentiment direction: a persistently negative narrative like “Recession” will typically receive a negative LLM score regardless of whether conditions are improving or deteriorating. In this case, the LLM score acts primarily as an attention measure whose sign is predetermined by the narrative label, and the estimated beta captures exposure to *variation in a pre-signed quantity* rather than to a directionally neutral factor. This means that comparing the sign of BoW and LLM betas for individual narratives—or interpreting cyclical–defensive beta patterns for the LLM—is not straightforward: a positive LLM beta to “Recession” does not imply that the asset benefits from recessions but rather that it co-moves with fluctuations in how negatively the LLM scores recession-related headlines. We therefore focus the GFC analysis on the aggregate signal weights and CW contributions reported in Tables ??–??. which reflect the full pipeline ($\beta \times \Delta\theta$, rank-scaled cross-sectionally) and do not require interpreting the sign of individual betas.

Differencing and the LLM crisis signal. An important asymmetry between the two models emerges during the GFC. The signal construction uses weekly changes $\Delta\theta_{n,t}$ as narrative factors (Section ??), which is natural for the BoW model: BoW scores are non-negative attention measures that surge during crises—“Recession” rises from a pre-GFC weekly average of 0.3 to 5.3 during the downturn, an eighteen-fold increase—so differencing captures the onset and escalation of crisis narratives. The LLM model,

by contrast, scores each narrative on a bounded ordinal scale $\{-2, -1, 0, 1, 2\}$ that is stationary by construction. During the GFC, LLM “Recession” scores average -1.2 , only modestly below their pre-crisis level of -0.9 : the score is *persistently* negative rather than *newly* negative, so its weekly change $\Delta\theta \approx 0$ conveys little information. Because $\Delta\theta$ is small for crisis narratives, the resulting $\beta \times \Delta\theta$ signals lack magnitude relative to non-crisis narratives whose scores fluctuate more freely. In the cross-sectional ranking that determines portfolio weights, these non-crisis narratives dominate, and the aggregate signal does not shift toward the defensive posture that the LLM *levels* would imply. Using levels directly as factors restores the correct crisis-period ordering, but at the cost of full-sample performance: levels are too persistent to generate meaningful weekly variation in the signal ranking. This tension between crisis-regime identification and high-frequency predictability is inherent to any $\Delta\theta$ -based framework applied to bounded scores—it affects both the CW portfolio (which ranks assets by $\beta \times \Delta\theta$) and narrative momentum (which ranks narratives by rolling-mean $\Delta\theta$). The multi-horizon composite $\bar{\Delta}I$ smooths over the shortest zero-signal windows, but does not resolve the fundamental issue during protracted crises where sentiment is persistently flat for months.

The narrative group decomposition in Table ?? confirms this at the thematic level: during the GFC, no narrative group contributes meaningfully in either direction, with “Crisis & Stress” narratives adding at most ± 1 bp. During COVID, however, the LLM captures broad-based sentiment shifts: its returns are driven by non-crisis themes (factor, strategy, and policy narratives) rather than by “Crisis & Stress” signals, which remain flat for the LLM but provide BoW’s largest COVID contribution through sharp attention surges.

McCracken PCA vs Narratives: GFC and COVID. Table ?? compares McCracken PCA CW returns with narrative CW returns during the two major crises.

During the GFC, the McCracken PCA CW dominates, with Sharpe ratios of 0.82 (49-Ind) and 1.43 (FF5+Mom), while LLM-sentiment is essentially flat ($SR \approx 0$) and BoW mildly negative. The slow-moving monthly macro factors—which capture employment, industrial production, and interest rate dynamics—are well-suited to a protracted, fundamentals-driven downturn where bounded narrative sentiment generates negligible $\Delta\theta$. During COVID, the pattern reverses completely: narrative CW portfolios generate Sharpe ratios of 1.8–5.6, while McCracken is flat (0.48 on 49-Ind) or negative (-0.58 on FF5+Mom). The three-month COVID window is too short for monthly macro data to update meaningfully, whereas daily news narratives adjust within days. Figure ?? illustrates this contrast.

Table 11: McCracken PCA vs Narrative CW: GFC and COVID Subperiods

Universe	Strategy	Ann.Ret (COVID)	Ann.Ret (GFC)	Sharpe (COVID)	Sharpe (GFC)	Vol (COVID)	Vol (GFC)
49-Ind	BoW	0.001	-0.001	0.532	-0.741	0.003	0.001
	LLM-sentiment	0.011	0.001	3.010	0.285	0.004	0.002
	McCracken PCA	0.001	0.002	0.482	0.820	0.002	0.003
FF5+Mom	BoW	0.007	-0.002	1.722	-0.635	0.004	0.003
	LLM-sentiment	0.021	0.001	4.085	0.367	0.005	0.004
	McCracken PCA	-0.002	0.004	-0.581	1.427	0.004	0.003

Notes: GFC = Dec 2007–Jun 2009; COVID = Feb–Apr 2020. Ann.Ret and Vol are annualized from weekly returns. McCracken PCA uses ten FRED-MD principal components (?) with expanding betas estimated at monthly frequency from 1970, forward-filled to weekly dates.

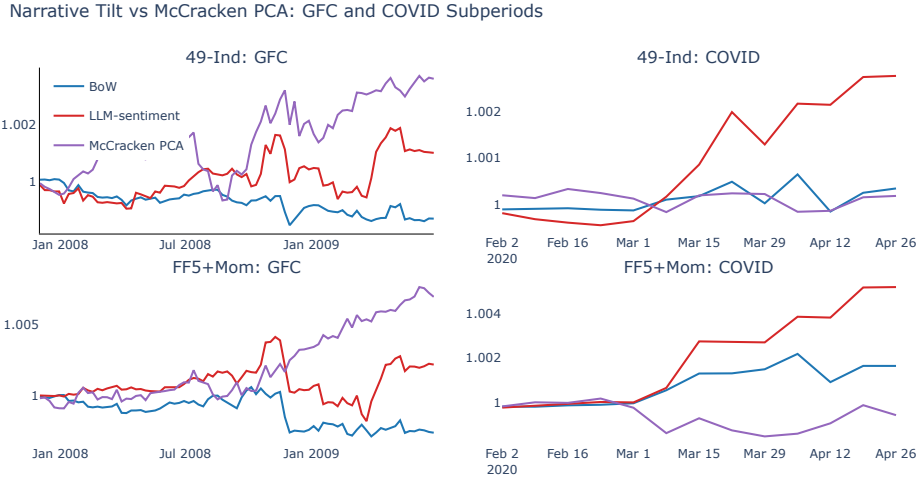


Figure 15: Narrative CW vs McCracken PCA during the GFC (left) and COVID (right) for 49-Ind (top) and FF5+Mom (bottom). McCracken dominates during the slow-burn GFC; narratives dominate during the rapid COVID shock.

B Narrative Momentum Sensitivity

The headline narrative momentum results in Section ?? use a composite lookback (12/26/52w), $J = 5$ narratives per leg, and $K = 1$ assets per mimicking portfolio. This appendix examines sensitivity to each of these parameters in turn.

B.1 Lookback Horizon (M)

Figure ?? plots the annualised Sharpe ratio of the narrative momentum spread as a function of the lookback horizon M , with $J = 5$ and $K = 1$. BoW narrative momentum produces Sharpe ratios near zero or negative across nearly all lookback windows and both universes, while LLM-sentiment exhibits positive Sharpe ratios over most of the grid, with a peak at $M = 26$ weeks (≈ 6 months) for FF5+Mom (SR = 0.69, $t = 3.06$) and competitive performance at $M = 52$ weeks for 49-Ind (SR = 0.34, $t = 1.60$). The six-month lookback—the headline specification in ?—delivers SR = 0.41 ($t = 1.99$) on 49-Ind.

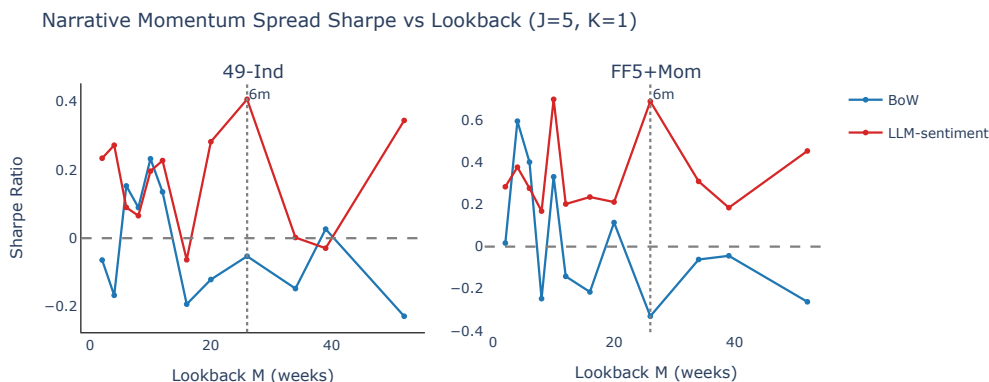


Figure 16: Narrative momentum spread: annualised Sharpe ratio vs. lookback horizon M (weeks), $J = 5$ narratives per leg, $K = 1$ assets per mimicking portfolio. Left: 49-Ind; right: FF5+Mom. Vertical dotted line marks $M = 26$ weeks (≈ 6 months).

The sensitivity of the LLM-sentiment Sharpe ratio to the choice of lookback—especially on 49-Ind, where it ranges from -0.06 to $+0.41$ —motivates the composite approach described in Section ???. Figure ??? compares the single-lookback Sharpe ratios (dotted lines) with the composite signal (horizontal solid lines). For LLM-sentiment, the composite delivers $SR = +0.49$ ($t = 2.37$) on 49-Ind—above the best single lookback—and $SR = +0.39$ ($t = 1.70$) on FF5+Mom. BoW momentum remains flat or negative under the composite, confirming that the BoW–LLM contrast is robust to lookback aggregation.

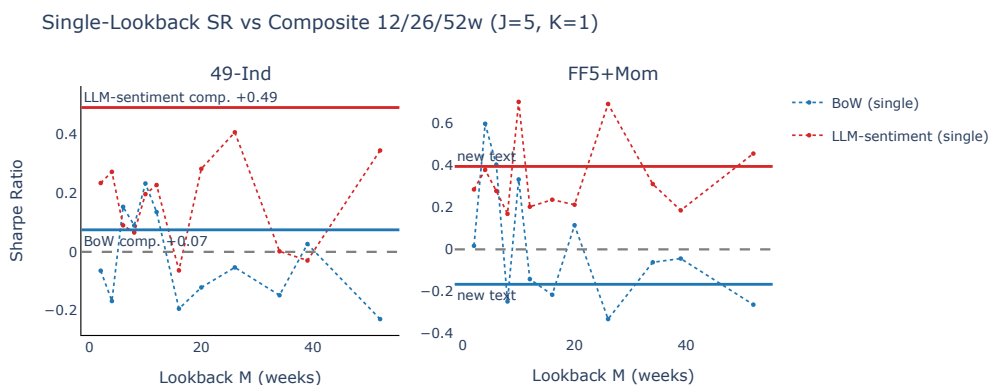


Figure 17: Narrative momentum Sharpe ratio: single-lookback sensitivity (dotted) vs. composite 12/26/52w signal (solid horizontal lines). The composite eliminates lookback selection while preserving the LLM-sentiment advantage.

B.2 Narratives per Leg (J)

Figure ??? varies the number of narratives per leg (J) from 3 to 30 while holding $K = 1$ and using the composite lookback signal (12/26/52w). LLM-sentiment Sharpe ratios are

positive across the entire range of J on both universes; BoW momentum remains flat or negative throughout. The baseline $J = 5$ sits near the peak for 49-Ind, while FF5+Mom is largely insensitive to J , reflecting the limited cross-sectional dispersion among its five assets.

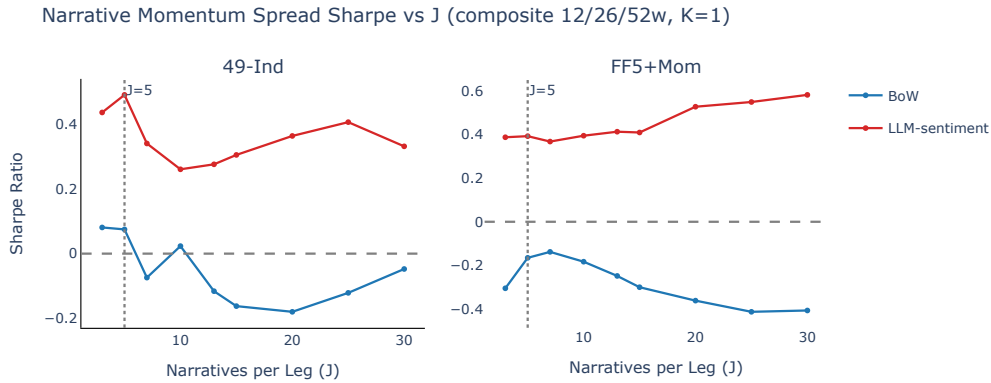


Figure 18: Narrative momentum spread: annualised Sharpe ratio vs. number of narratives per leg J (composite 12/26/52w lookback, $K = 1$). Vertical dotted line marks the baseline $J = 5$.

B.3 Assets per Mimicking Portfolio (K)

Figure ?? varies the number of assets per mimicking portfolio (K) while holding $J = 5$ and using the composite lookback. $K = 1$ is clearly optimal: increasing K dilutes narrative-specific exposure in the mimicking portfolios, and the LLM-sentiment Sharpe ratio on 49-Ind declines monotonically from its peak at $K = 1$. FF5+Mom constrains $K \leq 2$ due to its five-asset universe; even moving from $K = 1$ to $K = 2$ reduces the spread Sharpe ratio.

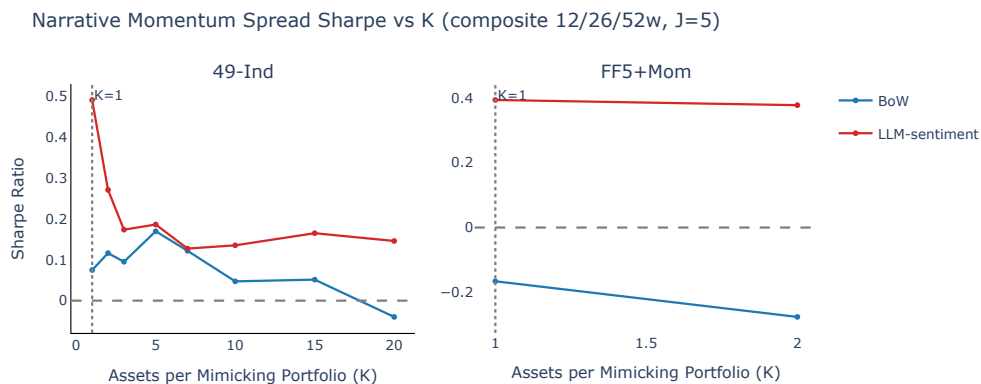


Figure 19: Narrative momentum spread: annualised Sharpe ratio vs. assets per mimicking portfolio K (composite 12/26/52w lookback, $J = 5$). FF5+Mom is limited to $K \leq 2$. Vertical dotted line marks the baseline $K = 1$.

Figures ??-?? further illustrate the effect of wider mimicking portfolios using single-lookback ($M = 26w$) results. Increasing K from 1 to 10 ($J = 5$) eliminates the LLM-sentiment advantage entirely (SR falls from +0.41 to +0.01), while doubling J to 10 with $K = 5$ yields only a modest LLM Sharpe of +0.12 ($t = 0.59$). These results confirm that $K = 1$ is the critical design choice: each mimicking portfolio must isolate the single most exposed and least exposed asset to preserve the narrative-specific signal.

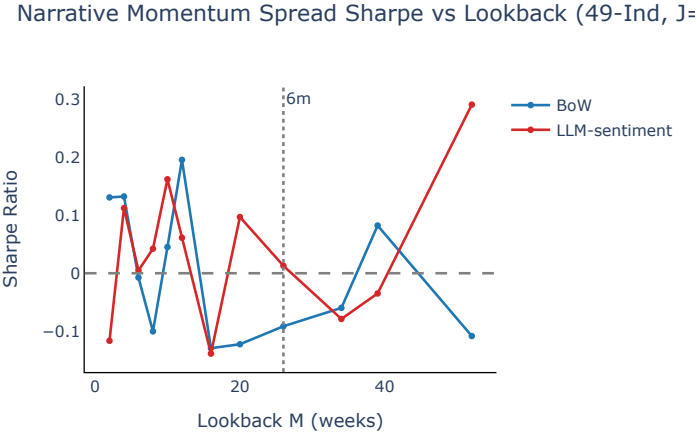


Figure 20: Narrative momentum spread Sharpe ratio vs. lookback (49-Ind, $J = 5$, $K = 10$).

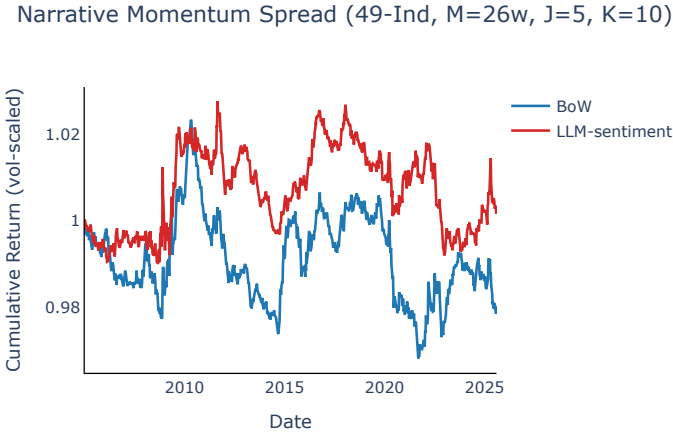


Figure 21: Narrative momentum spread equity lines (49-Ind, $M = 26w$, $J = 5$, $K = 10$, vol-scaled to 1% ann.).

Narrative Momentum Spread Sharpe vs Lookback (49-Ind, J=

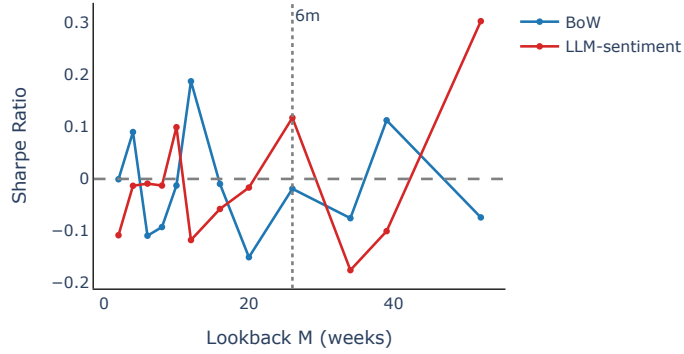


Figure 22: Narrative momentum spread Sharpe ratio vs. lookback (49-Ind, $J = 10$, $K = 5$).

Narrative Momentum Spread (49-Ind, M=26w, J=10, K=5)

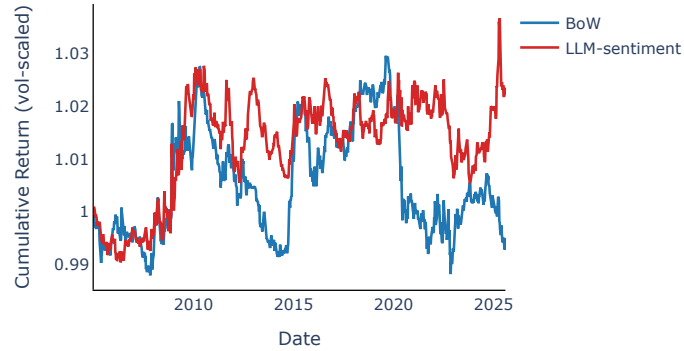


Figure 23: Narrative momentum spread equity lines (49-Ind, $M = 26w$, $J = 10$, $K = 5$, vol-scaled to 1% ann.).

C Multi-Horizon Composite Signal

The main text uses a short-horizon composite $\bar{\Delta}I$ with $\mathcal{H} = \{4, 8, 12\}$ weeks for the CW portfolio and a long composite $\mathcal{H} = \{12, 26, 52\}$ weeks for narrative momentum. Here we test sensitivity to the composite composition by varying the three lookback horizons while holding the number of lookbacks fixed at three. We consider five specifications:

1. Very short $\{4, 8, 12\}w \approx 1/2/3$ months (CW baseline);
2. Short-tilted $\{4, 12, 26\}w$ — shifts weight toward recent dynamics;
3. Baseline $\{12, 26, 52\}w \approx 3/6/12$ months (narrative momentum baseline);
4. Long-tilted $\{26, 39, 52\}w$ — persistent trends only;

5. Wide spread $\{4, 26, 52\}w$ — maximises horizon diversity.

CW portfolio. Figures ?? and ?? overlay CW equity lines across all five horizon sets. LLM-sentiment is remarkably robust: Sharpe ratios range from +0.29 to +0.48 on 49-Ind and +0.36 to +0.48 on FF5+Mom, with the very short baseline (4/8/12w) performing best on 49-Ind and the long composite (12/26/52w) on FF5+Mom. BoW degrades monotonically with longer horizons, reflecting its reliance on high-frequency count spikes.

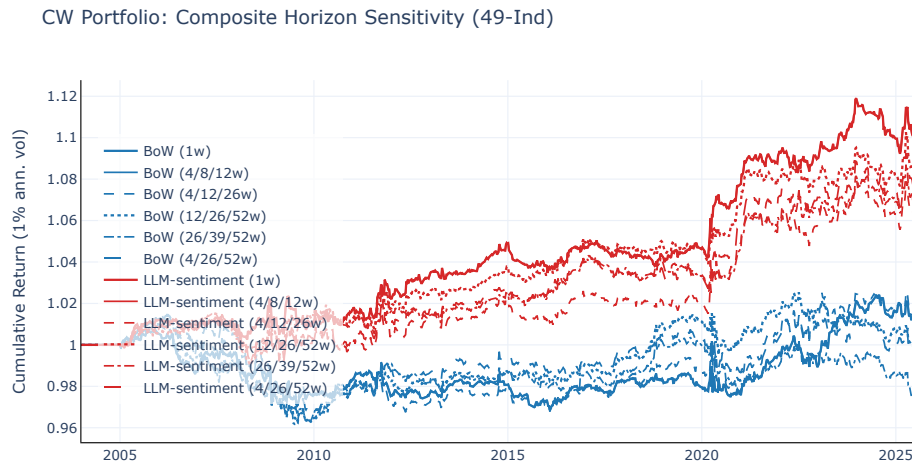


Figure 24: CW equity lines across five composite horizon sets. 49-Ind, vol-scaled to 1% ann.

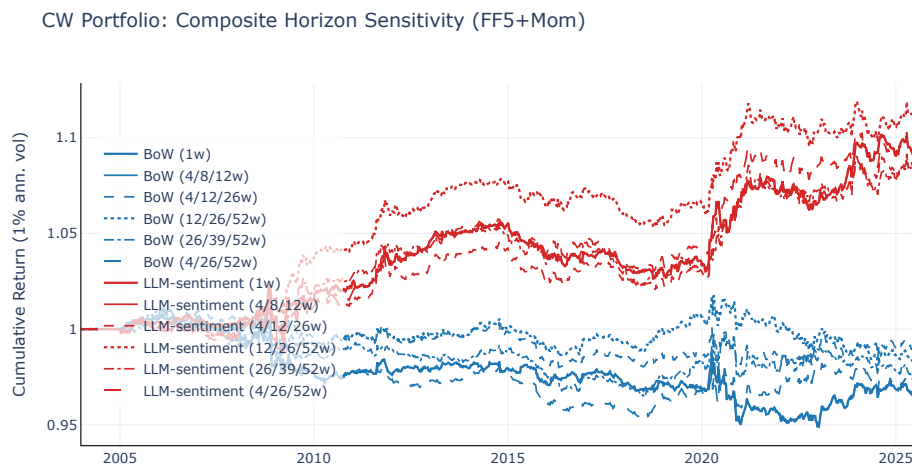


Figure 25: CW equity lines across five composite horizon sets. FF5+Mom.

Narrative momentum. Figures ?? and ?? compare narrative momentum spreads across the same five horizon sets. The horizon preference is universe-dependent. On

49-Ind, the baseline (12/26/52w) dominates with LLM SR +0.49 ($t = 2.37$), while the very short composite (4/8/12w) drops to SR +0.15, indicating that industry-level narrative momentum requires longer lookbacks. The short-tilted (4/12/26w) recovers much of the performance (SR +0.40, $t = 2.08$), suggesting that the 26-week horizon carries most of the signal. On FF5+Mom, the pattern reverses: the long-tilted composite (26/39/52w) achieves the highest LLM SR (+0.51, $t = 2.14$), but all five specifications deliver SR $\geq +0.39$, indicating that factor-level narrative momentum is robust to horizon choice.

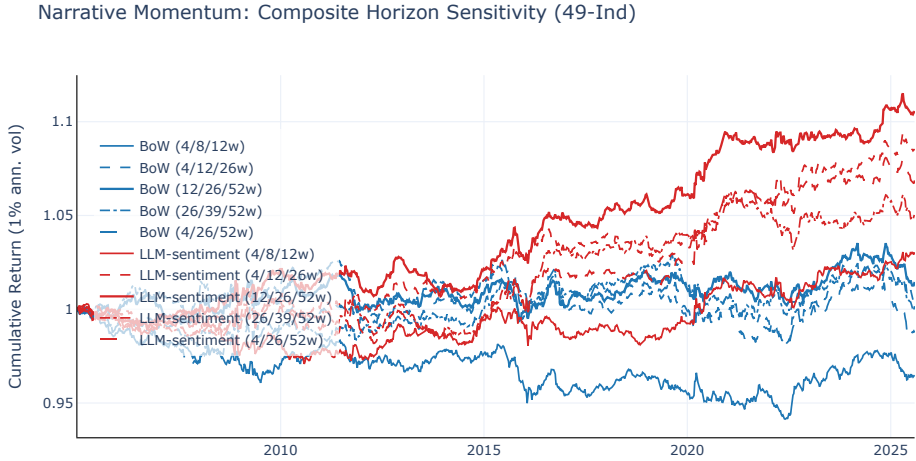


Figure 26: Narrative momentum spread across five composite horizon sets. 49-Ind, vol-scaled to 1% ann.

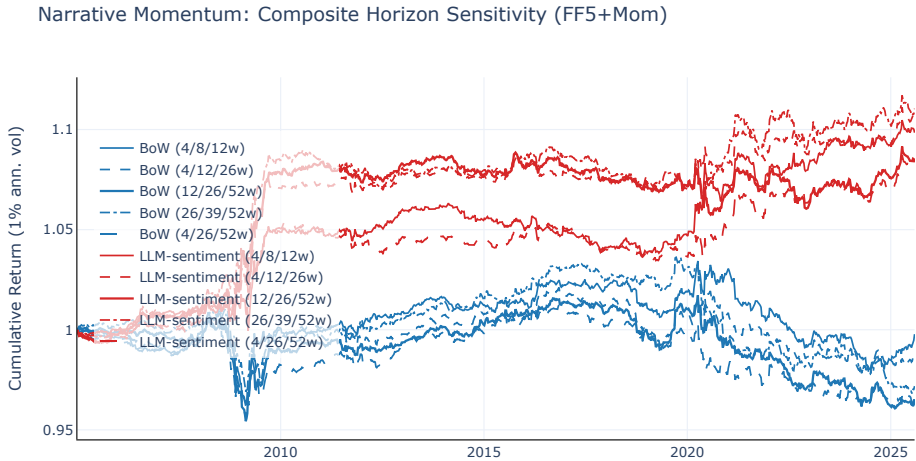


Figure 27: Narrative momentum spread across five composite horizon sets. FF5+Mom.

Narrative momentum without vol-scaling. The baseline composite cross-sectionally vol-scales each horizon's rolling-mean $\Delta\theta$ before averaging, ensuring all lookbacks contribute equally. Figures ?? and ?? compare the vol-scaled composite with a raw average of rolling means for the baseline horizons (12/26/52w). On 49-Ind, the raw composite is materially weaker (LLM SR +0.32, $t = 1.60$ vs. +0.49, $t = 2.37$), indicating that the vol-scaling step is important for equalising dispersion across horizons when ranking 65 narratives. On FF5+Mom, the two variants are virtually identical (SR +0.39 in both cases), consistent with the smaller cross-section leaving less room for scale differences to matter.

Narrative Momentum: Vol-Scaled vs Raw Composite (49-Ind)

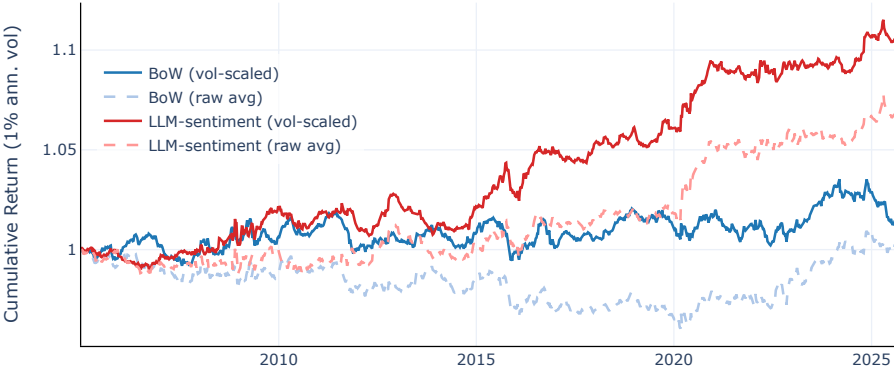


Figure 28: Narrative momentum spread: vol-scaled composite (solid) vs. raw average composite (dashed). 49-Ind, baseline horizons (12/26/52w).

Narrative Momentum: Vol-Scaled vs Raw Composite (FF5+Mom)

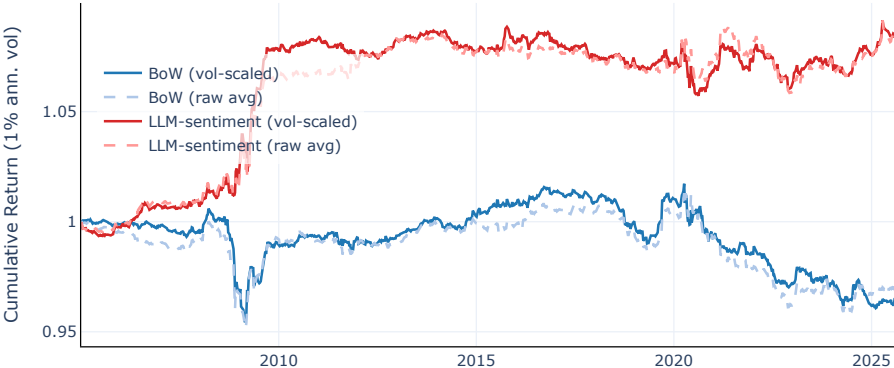


Figure 29: Narrative momentum spread: vol-scaled vs. raw composite. FF5+Mom.

Summary. Table ?? collects results across both strategies, all five horizon sets, and the raw-average variant. Four findings stand out. First, the LLM-sentiment CW portfolio is insensitive to the composite horizon—SR ranges from +0.29 to +0.48 on 49-Ind and +0.36 to +0.48 on FF5+Mom—while BoW degrades monotonically with smoothing. Second, the two strategies have opposite horizon preferences: the CW portfolio favours short smoothing (preserving high-frequency $\beta \times \Delta\theta$ variation), while narrative momentum on 49-Ind requires longer lookbacks to identify persistent cross-narrative trends. Third, the short-tilted composite (4/12/26w) offers a reasonable compromise for narrative momentum, recovering $\sim 80\%$ of the baseline SR on 49-Ind while slightly improving on FF5+Mom. Fourth, the vol-scaling step matters for narrative momentum on 49-Ind (SR drops from +0.49 to +0.32 without it) but is inconsequential on FF5+Mom, suggesting that normalisation is most valuable when the cross-section is large enough for scale differences across horizons to distort the ranking.

49-Ind	Composite (4/12/26w)	0.000	0.001	0.081	0.416
	Composite (12/26/52w)	-0.000	0.001	-0.005	-0.029
	Composite (26/39/52w)	-0.000	0.001	-0.127	-0.670
	Composite (4/26/52w)	-0.000	0.001	-0.004	-0.020
BoW	Baseline (4/8/12w)	-0.000	0.001	-0.200	-0.992
	Composite (4/8/12w)	-0.000	0.001	-0.200	-0.992
	Composite (4/12/26w)	-0.000	0.001	-0.080	-0.422
	Composite (12/26/52w)	-0.000	0.001	-0.098	-0.558
	Composite (26/39/52w)	-0.000	0.001	-0.117	-0.659
	Composite (4/26/52w)	-0.000	0.001	-0.172	-0.931
	Baseline (4/8/12w)	0.001	0.001	0.477	2.196
	Composite (4/8/12w)	0.001	0.001	0.477	2.196
FF5+Mom	Composite (4/12/26w)	0.000	0.001	0.299	1.360
	Composite (12/26/52w)	0.001	0.001	0.378	1.772
	Composite (26/39/52w)	0.000	0.001	0.294	1.364
	Composite (4/26/52w)	0.000	0.001	0.341	1.592
	Baseline (4/8/12w)	0.001	0.002	0.394	1.774
	Composite (4/8/12w)	0.001	0.002	0.394	1.774
	Composite (4/12/26w)	0.001	0.002	0.358	1.629
	Composite (12/26/52w)	0.001	0.002	0.477	2.274
FF5+Mom	Composite (26/39/52w)	0.001	0.002	0.368	1.756
	Composite (4/26/52w)	0.001	0.002	0.407	1.848
	4/8/12w	-0.001	0.007	-0.168	-0.769
	4/12/26w	0.000	0.007	0.033	0.150
	12/26/52w	0.000	0.007	0.075	0.337
	26/39/52w	0.000	0.007	0.014	0.062
	4/26/52w	-0.000	0.007	-0.056	-0.252
	12/26/52w raw	0.000	0.007	0.017	0.084
BoW	4/8/12w	-0.000	0.007	-0.006	-0.024
	4/12/26w	-0.000	0.007	-0.043	-0.179
	12/26/52w	-0.001	0.007	-0.166	-0.689
	26/39/52w	-0.001	0.006	-0.133	-0.563
	4/26/52w	-0.001	0.006	-0.158	-0.693
	12/26/52w raw	-0.001	0.008	-0.133	-0.567
	4/8/12w	0.001	0.008	0.150	0.788
	4/12/26w	0.003	0.008	0.402	2.083
49-Ind	12/26/52w	0.004	0.008	0.491	2.369
	26/39/52w	0.002	0.007	0.244	1.132
	4/26/52w	0.002	0.008	0.326	1.576
	12/26/52w raw	0.003	0.008	0.323	1.605
	4/8/12w	0.004	0.008	0.459	2.158
	4/12/26w	0.004	0.008	0.482	2.151
	12/26/52w	0.003	0.007	0.393	1.707
	26/39/52w	0.004	0.007	0.512	2.140
FF5+Mom	4/8/12w	0.004	0.008	0.459	2.158
	4/12/26w	0.004	0.008	0.482	2.151
	12/26/52w	0.003	0.007	0.393	1.707
	26/39/52w	0.004	0.007	0.512	2.140
	4/8/12w	0.004	0.008	0.459	2.158
	4/12/26w	0.004	0.008	0.482	2.151
	12/26/52w	0.003	0.007	0.393	1.707
	26/39/52w	0.004	0.007	0.512	2.140
LLM-sentiment	4/8/12w	0.004	0.008	0.459	2.158
	4/12/26w	0.004	0.008	0.482	2.151
	12/26/52w	0.003	0.007	0.393	1.707
	26/39/52w	0.004	0.007	0.512	2.140
	4/8/12w	0.004	0.008	0.459	2.158
	4/12/26w	0.004	0.008	0.482	2.151
	12/26/52w	0.003	0.007	0.393	1.707
	26/39/52w	0.004	0.007	0.512	2.140
Narr. Momentum	4/8/12w	0.004	0.008	0.459	2.158
	4/12/26w	0.004	0.008	0.482	2.151
	12/26/52w	0.003	0.007	0.393	1.707
	26/39/52w	0.004	0.007	0.512	2.140
	4/8/12w	0.004	0.008	0.459	2.158
	4/12/26w	0.004	0.008	0.482	2.151
	12/26/52w	0.003	0.007	0.393	1.707
	26/39/52w	0.004	0.007	0.512	2.140