Data Science in Finance and Economics

https://www.aimspress.com/journal/dsfe

DSFE, 6(4): 567–594.
DOI:10.3934/DSFE.2026020
Received: 27 December 2025
Revised: 15 September 2026
Accepted: 22 September 2026
Published: 27 October 2026

*Research article*

# Narratives as risk transmitters

**José Gabriel Astaíza-Gómez***

Area of Macroeconomics and Financial Systems, Universidad EAFIT, Carrera 49 N° 7 Sur - 50, Medellín, Colombia

* **Correspondence:** Email: jastaiza@eafit.edu.co.

**Abstract:** This work documented that investor narratives are net originators of information that propagates into systematic risk factors, reframing Fama-French factors as endogenous downstream constructs rather than exogenous risk primitives. Using daily data from 2015 to 2024 for 503 S&P 500 constituents, variance-weighted indices were constructed from analyst recommendations, social media, and stock prices, combined with Fama-French factors and Global Database of Events, Language and Tone (GDELT) media-attention measures. Quantile vector autoregression (VAR) network analysis showed that the market, size, and value factors are net information receivers, whereas growth and value narratives are net transmitters. Granger causality tests indicated that analyst recommendations predict factor movements, while social-media sentiment follows rather than leads prices. Structural impulse responses identified a transmission mechanism: growth narrative shocks significantly reduce the size premium, consistent with narrative-driven capital reallocation toward large-cap stocks.

**Keywords:** narrative finance; asset pricing; Fama-French factors; vector autoregression; information transmission

**JEL Codes:** G11, G12, G14, G41

## 1. Introduction

This work investigates whether investor narratives serve as fundamental transmitters of financial risk rather than mere reflections of market conditions. While traditional asset-pricing models treat risk factors as exogenous forces driven by economic fundamentals, I examine the hypothesis that narratives, i.e., coherent stories investors tell about market themes like value and growth, actively originate and propagate information flows that influence systematic risk factors. This paper's primary contribution is to establish that investor narratives are net originators of information that propagates into systematic risk factors, reframing Fama-French factors as endogenous downstream constructs rather than exogenous

risk primitives. The remaining findings of the paper on the analyst-versus-social-media hierarchy and on style-narrative reflexivity are not co-equal results but the constituent links that make this single causal claim empirically identifiable. Using the net pairwise directional connectedness (NPDC) framework applied to a comprehensive dataset of narrative intensity measures and Fama-French factors, I provide empirical evidence on the directional relationship between narrative dynamics and risk transmission. My findings have direct implications for understanding whether narratives are endogenous drivers of risk transmission or simply respond to underlying economic shocks, a distinction critical for asset-pricing theory (Bybee et al., 2023), risk management (Goetzmann et al., 2022), and the behavioral foundations of market efficiency (Johnson and Tuckett, 2022).

This study emphasizes signal transmission rather than the forecasting power of financial data or statistical models. Abstracting from noise and the limitations that opaque assets impose on signal transmission, I argue that variance serves as a fundamental measure of signal clarity in financial markets. Specifically, I extend insights from the analyst forecast consistency literature to construct information-weighted market indicators that better capture the quality of transmitted signals. My framework for deriving optimal weights shares conceptual foundations with Hilary and Hsu's influential work on forecast reliability, though I apply these principles to aggregate multiple information sources into composite indices. Hilary and Hsu (2013) operationalized consistency as the standard deviation of forecast errors, demonstrating that analysts with lower error variance provide more valuable information to investors, even when their forecasts exhibit systematic bias. Their central insight, that predictability enables decision-makers to adjust for systematic patterns, motivates my approach to index construction using optimal weights that minimize variance in the underlying signals. I construct indices for stock prices, Twitter counts, Twitter sentiment, numbers of buy and sell recommendations, and analyst price targets from 2015 onward, where each index aggregates constituent signals weighted by their consistency properties. The parallel to Hilary and Hsu is precise: just as consistently biased forecasts prove more useful than erratic predictions, consistently weighted information sources produce more reliable aggregate measures that facilitate systematic analysis of market dynamics.

Recent work in financial data science has advanced along three complementary strands that frame the contribution of this paper. Afzal et al. (2025) integrated long short-term memory (LSTM) networks with the Fama-French six-factor model to forecast Shenzhen portfolio returns, treating factors as exogenous inputs to a non-linear predictor. Dabara and Tahir (2025) developed voting and stacking ensembles of gradient boosting, neural networks, and Bayesian ridge regression across six global indices, demonstrating that hybrid architectures minimize forecasting error but remain agnostic to the economic origin of predictive signals. Li et al. (2026) modeled China's commodity futures through an Ornstein-Uhlenbeck jump process coupled to a high-order hidden markov chain, inferring latent information-flow regimes that lead official macroeconomic indicators. Collectively, these studies treat factors, prices, or latent states as the objects of prediction or inference. This paper takes a different, causally oriented route: rather than forecasting returns or extracting hidden regimes from prices alone, it uses quantile VAR with net pairwise directional connectedness on explicitly measured narrative, analyst, and factor channels to show that Fama-French factors are net receivers in an information hierarchy originated by professional analyst narratives and media-based investment-style discourse.

I incorporate two Fama-French factors and two GDELT publication time series capturing media attention to "growth investing" and "value investing" narratives. Network connections estimated from the QVAR reveal how information flows across these channels conditional on market states, while

impulse response functions show how shocks propagate through this information network. Granger causality tests provide computationally efficient baseline assessments of predictive relationships and inform the specification of structural vector autoregressions. Network analysis shifts the focus from predictive ability to net spillover dominance, identifying which variables drive the system versus respond to it. For studies investigating information transmission in financial markets, this approach offers several advantages. It distinguishes between predictive causality (Granger) and spillover dominance (NPDC), which, while related, capture distinct economic phenomena. The network representation provides intuitive visualization of interconnection patterns, facilitating communication with both academic and practitioner audiences. By identifying net transmitters at upper quantiles, the framework pinpoints variables that amplify rather than absorb extreme positive shocks.

The empirical support for this contribution rests on three interlocking components of one causal chain, rather than three independent findings. First, establishing the endpoint of the chain: net pairwise directional connectedness analysis identifies the market premium, small minus big (SMB), and high minus low (HML) as net receivers of information, and Granger causality tests confirm that analyst recommendations systematically predict subsequent factor movements, together demonstrating that Fama-French risk factors are endogenous to narrative dynamics rather than purely exogenous risk exposures. Second, identifying the origin of the chain: Information transmission exhibits a hierarchical structure in which professional analyst narratives originate pricing-relevant signals that cause, in the sense of Granger, both stock prices and factor returns, whereas social-media discourse reflects rather than leads market dynamics. The hierarchy justifies treating professional narratives, rather than retail sentiment, as the exogenous input in the structural identification. Third, closing the chain with an explicit mechanism: Investment style narratives exhibit reflexive relationships with systematic factors, as growth and value media attention emerge as net transmitters with mutual Granger causality, and structural impulse responses reveal that growth narrative shocks generate significant negative effects on the size premium, consistent with growth-oriented discourse directing capital toward large-capitalization stocks, thereby providing the concrete transmission channel through which narratives move factors.

This document is organized as follows. Section 2 presents the literature review. Section 3 describes the data and methodology, where I detail analyst coverage and social media metrics, Fama-French factors, and GDELT media attention measures for investment style narratives. I explain the construction of composite indices that maximize signal consistency and outline the econometric framework including vector autoregression, quantile VAR-based connectedness analysis, and structural identification. Section 4 presents empirical results, beginning with multivariate Granger causality tests, followed by interconnectedness analysis revealing the hierarchical network structure, and concluding with impulse response functions. Section 5 reconciles findings with existing finance literature, and addresses practical implications for portfolio construction and risk modeling.

## 2. Literature review

This paper investigates whether investor narratives actively originate and propagate information flows into systematic risk factors, rather than merely reflecting market conditions. Three central results emerge: (i) Fama-French factors (Mkt-RF, SMB, HML) are net receivers of information and are endogenous to narrative dynamics; (ii) professional analyst recommendations originate pricing-relevant signals, whereas Twitter sentiment follows prices rather than leads them; and (iii) growth and value narratives

are net transmitters, and growth shocks generate significant negative responses in SMB.

Afzal et al. (2025) integrated a long short-term memory network with the Fama-French six-factor model to predict portfolio returns on the Shenzhen Stock Exchange (1997—2022), complementing LSTM with a random-forest surrogate to capture non-linear threshold effects. They reported robust forecasting gains, particularly for high-downside-risk portfolios, and concluded that factor-return relationships are inherently non-linear. Their framework, however, treats the six Fama-French factors as exogenous inputs in a return-forecasting model. The present paper inverts this logic: rather than using factors to predict returns, it asks what drives the factors themselves and, applying NPDC at upper quantiles, finds that Mkt-RF, SMB, and HML act as net receivers of information flowing from analyst and narrative channels. Where Afzal et al. (2025) enriched factor models with non-linear learners for prediction, this paper challenges the premise that factors are primitive explanatory variables and shows they are downstream in the information hierarchy. The two approaches are methodologically and conceptually orthogonal: deep-learning augmentation of factor pricing versus causal-directional decomposition of factor origins.

Dabara and Tahir (2025) pursued a purely data-driven forecasting strategy, combining gradient boosting, decision trees, neural networks, and Bayesian ridge regression through voting and stacking ensembles across six global indices (S&P 500, NASDAQ, DAX, FTSE, Nikkei 225, Hang Seng). Their stacking ensemble achieved the lowest mean absolute error (MAE) (0.0332), and analysis of variance (ANOVA) with Tukey's honest significant difference (HSD) confirming the superiority of ensemble architectures over standalone models. Their design is agnostic to the economic origin of predictive signals and treats price series as self-referential. The present paper is complementary rather than competitive: it does not optimize forecast accuracy but identifies which upstream signals (specifically professional analyst recommendations and growth/value media narratives) carry causal information into return-generating processes. The variance-weighted narrative indices constructed here, along with the NPDC-identified transmitters, could plausibly serve as additional inputs to ensemble architectures such as Dabara and Tahir's, providing economically interpretable features that compensate for the black-box character of pure data-driven stacking.

Li et al. (2026) modeled China's commodity futures market (Nanhua indices, 2004—2023) through a three-layer fusion of an Ornstein-Uhlenbeck process with compound Poisson jumps and a high-order hidden markov chain, estimated recursively via an expectation-maximization (EM) recursive filtering procedure. They identified latent high- and low-volatility regimes, documented time-varying market efficiency consistent with the adaptive market hypothesis, found pronounced sectoral heterogeneity, and showed that hidden information flows lead the official Macroeconomic Climate Index. Conceptually, their paper is the closest cousin to the present study: both treat information flow as a latent driver of observed market behavior and both emphasize state-dependent dynamics. They diverge in the identification strategy. Li et al. (2026) inferred regimes from price dynamics alone using a single asset class (Chinese commodity futures) and obtained latent states as a by-product of the stochastic model. This paper instead measures information transmission across explicitly observed narrative, analyst, and factor channels using quantile-specific connectedness. The findings are structurally analogous. Li et al.'s hidden flows lead macroeconomic indicators just as my narrative and analyst channels lead Fama-French factors. Combining their hidden-regime machinery with the NPDC framework used here would yield regime-conditional connectedness estimates, a natural extension for future work.

Taken together, these studies exemplify prevailing strands in contemporary financial data science:

deep-learning augmentation of classical factor pricing (Afzal et al., 2025), ensemble-based pure forecasting (Dabara and Tahir, 2025), and latent-state stochastic modeling (Li et al., 2026). The present paper contributes a fourth, causally oriented strand: one that preserves the interpretability of Fama-French factors and analyst variables while exposing, through quantile connectedness and structural impulse responses, the directional information architecture that generates them. In doing so, it reframes systematic risk factors as endogenous downstream constructs rather than exogenous risk primitives.

A terminological clarification is warranted regarding the use of the term ”narrative” throughout this study. Behavioral finance distinguishes investor narratives from attention and sentiment. Narratives are coherent, semantically structured stories that coordinate beliefs across agents (Shiller, 2019; Bybee et al., 2023). The proxies employed here, i.e., GDELT daily article counts for the queries ”growth investing” and ”value investing,” Twitter publication volume, and Twitter sentiment averages, do not reconstruct the semantic content of such stories. They are, however, deliberately narrative-targeted rather than generic attention measures: the GDELT series are obtained from thematically pre-specified style queries that isolate discourse about two canonical investment-style narratives, ”growth” and ”value,” each of which corresponds to a long-standing, widely recognized story about how firms are valued and how capital should be allocated. The daily count of articles matching these queries therefore indexes the temporal salience and intensity of each style narrative in global news discourse, not the overall volume of financial news or generic investor attention. This is consistent with the operationalization in Shiller (2015, 2019), where the strength of a narrative is measured by its rate of occurrence in the media, and with the epidemiological view of narratives as contagious stories whose propagation is captured empirically by frequency. Analogously, the Twitter volume and sentiment indices aggregated through the variance-minimizing weights of Section 3.2 are interpreted as proxies for the diffusion and valence of these narratives on social media, not as the narratives themselves. I therefore use ”narrative” to denote the latent, thematically defined object (growth investing, value investing) and ”narrative intensity” or ”narrative salience” to denote its observable footprint in media frequency and social-media discourse. Fully semantic extraction of narrative content, for instance, via topic modeling or embedding-based methods as in Bybee et al. (2023), would refine the measurement of narrative content but would not alter the directional findings of this paper, which concern the timing and transmission of narrative salience shocks into Fama-French factors. Extending the present framework to semantically extracted narratives is a natural avenue for future work and is noted as such in the concluding section.

## 3. Materials and method

### *3.1. Data*

Stock data, including the number of Twitter posts, a Twitter sentiment series, the number of analyst recommendations, the number of buy recommendations, the number of sell recommendations, and stock prices for 503 stocks, were obtained from Bloomberg. The dataset spans January 2015 through September 2024 (Table 1) and combines social-media discourse metrics with traditional analyst coverage variables.

**Table 1.** Summary statistics and first/last non-missing dates per variable.

| variable | first_date | last_date | n_obs | n_missing | mean | median | sd | min | max |
|---|---|---|---|---|---|---|---|---|---|
| TWITTER_PUBLICATION_COUNT | 2015-01-02 | 2024-09-11 | 1241252 | 540374 | 87.2146 | 11.00 | 580.8413 | 0.0000 | 142267.00 |
| TWITTER_SENTIMENT_DAILY_AVG | 2015-01-02 | 2024-09-11 | 1231029 | 550597 | 0.0294 | 0.00 | 0.1732 | -1.0000 | 1.00 |
| best_target_price | 2015-01-01 | 2024-09-11 | 665652 | 1115974 | 137.5009 | 88.50 | 204.3672 | 0.5405 | 8860.00 |
| px_last | 2015-01-02 | 2024-09-11 | 1157382 | 624244 | 130.8029 | 81.11 | 247.6563 | 0.4784 | 9321.27 |
| tot_analyst_rec | 2015-01-01 | 2024-09-11 | 695063 | 1086563 | 23.6857 | 22.00 | 8.8252 | 0.0000 | 79.00 |
| tot_buy_rec | 2015-01-01 | 2024-09-11 | 695063 | 1086563 | 13.3540 | 12.00 | 8.4248 | 0.0000 | 76.00 |
| tot_sell_rec | 2015-01-01 | 2024-09-11 | 695063 | 1086563 | 1.2279 | 1.00 | 1.6395 | 0.0000 | 18.00 |

The sample consists of the 503 common-stock constituents of the S&P 500 index as of September 2024, with multiple share classes (e.g., Alphabet's GOOG and GOOGL) accounting for the count exceeding 500. Because membership is fixed at the end-of-sample snapshot rather than reconstructed point-in-time, the panel inherits a survivorship tilt toward large, liquid firms that remained index-eligible throughout 2015—2024; I discuss the implications of this design choice for the SMB results below. Twitter publication counts average 87.2 daily mentions per stock, with substantial right-skewness reflecting viral attention episodes (maximum: 142,267 mentions). Daily sentiment averages 0.029, indicating mildly optimistic baseline discourse. Analyst coverage data includes consensus target prices (mean: 137.50), stock prices (mean: 130.80), and recommendation distributions showing 23.7 total recommendations per stock on average. The recommendation composition reveals strong buy-side bias, with 13.4 buy recommendations versus 1.2 sell recommendations per stock, consistent with documented analyst optimism in the literature. The panel exhibits considerable missingness across the 503 firms analyzed, totaling 8.8 million missing observations (Tables 2 and 3). On average, 45.08% of potential firm-day observations are missing (median: 44.42%), with heterogeneity ranging from 28.3% to 97.64% across securities. This reflects discontinuous analyst coverage and the episodic nature of Twitter activity for individual stocks.

**Table 2.** Distribution of missing values across stocks.

| total | mean | median | sd | min | max | mean_rate | median_rate | sd_rate | min_rate | max_rate |
|---|---|---|---|---|---|---|---|---|---|---|
| 8835032 | 17564.68 | 17307 | 3535.347 | 11027 | 38044 | 0.4508 | 0.4442 | 0.0907 | 0.283 | 0.9764 |

**Table 3.** Summary statistics (after excluding missing values).

| variable | first_date | last_date | n_obs | mean | median | sd | min | max |
|---|---|---|---|---|---|---|---|---|
| TWITTER_PUBLICATION_COUNT | 2015-01-02 | 2024-09-11 | 567233 | 124.5584 | 15.0000 | 700.2438 | 0.0000 | 142267.00 |
| TWITTER_SENTIMENT_DAILY_AVG | 2015-01-02 | 2024-09-11 | 567233 | 0.0308 | 0.0000 | 0.1795 | -1.0000 | 1.00 |
| best_target_price | 2015-01-02 | 2024-09-11 | 567233 | 139.6738 | 90.0682 | 208.0110 | 0.5405 | 8860.00 |
| px_last | 2015-01-02 | 2024-09-11 | 567233 | 125.7549 | 81.2200 | 189.6050 | 0.4784 | 8641.12 |
| tot_analyst_rec | 2015-01-02 | 2024-09-11 | 567233 | 23.7028 | 22.0000 | 8.8080 | 1.0000 | 79.00 |
| tot_buy_rec | 2015-01-02 | 2024-09-11 | 567233 | 13.4650 | 12.0000 | 8.4070 | 0.0000 | 76.00 |
| tot_sell_rec | 2015-01-02 | 2024-09-11 | 567233 | 1.2136 | 1.0000 | 1.6169 | 0.0000 | 18.00 |

Two distinct notions of missingness need to be distinguished. The raw firm-day panel of 503 constituents across seven Bloomberg variables has an overall missing rate of 45.08% (Table 2), but this panel is not the object of the time-series estimation. The VAR, QVAR, and structural VAR (SVAR) are estimated on aggregate daily indices built from the cross-section of firms reporting on each date, with variance-minimizing weights (Section 3.2) renormalized each day over the available constituents subject to a minimum coverage of five firms. The firm-level covariance matrix feeding the weight optimization is computed from pairwise-complete observations. Each pairwise covariance uses only the dates on

which both stocks report rather than from imputed series, avoiding the distortion that forward-filling or mean-imputation would introduce. No time-dimension imputation, interpolation, or listwise deletion is applied at any stage prior to VAR estimation. The aggregate indices entering the multivariate models are therefore continuous at the daily frequency by construction, and the 45.08% raw-panel missing rate does not translate into gaps or imputed values in the series on which the VAR, QVAR, and SVAR are estimated.

Analyst price target coverage patterns reveal systematic gaps in the data-generating process (Table 4). Over the 3,177-day sample window (mean and median), firms experience an average coverage gap of 667.5 days (median: 683), representing approximately 21% of the total observation period without updated price targets. The interquartile range of gap shares (15.6% to 26.7%) indicates relatively consistent but imperfect coverage across the sample.

**Table 4.** Summary statistics: Analyst price target coverage.

| Firms | Mean days total | Median days total | Mean days gap | Median days gap | Mean share gap | Median share gap | P25 share gap | P75 share gap |
|---|---|---|---|---|---|---|---|---|
| 503 | 3177 | 3177 | 667.5 | 683 | 0.21 | 0.215 | 0.156 | 0.267 |

Table 5 reports the distribution of analyst target gap coverage across firms. The majority of firms (339 out of 503) have target prices on 25% or fewer trading days, indicating sparse and intermittent updates, while 164 firms exhibit moderate coverage (25—50% of trading days). High-coverage firms are rare, reflecting the concentration of analyst attention on a subset of visible firms.

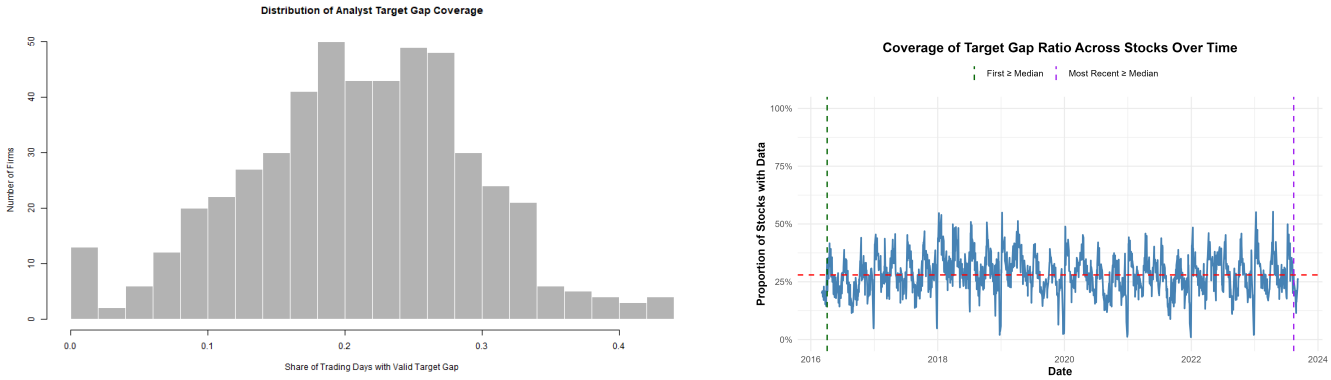
**Table 5.** Distribution of analyst price target gap coverage.

| Price Target Coverage | Firms |
|---|---|
| 25—50% | 164 |
| $\leq 25\%$ | 339 |

Table 6 reports the average coverage of analyst target prices across firms grouped by analyst coverage intensity. Coverage increases monotonically with the number of analysts: firms followed by more than 15 analysts exhibit target prices on about a quarter of the trading days, whereas firms with five or fewer analysts are covered sporadically. This highlights the concentration of analyst attention on a limited set of visible firms.

**Table 6.** Price target coverage by analyst intensity.

| Analyst Coverage | Firms | Mean coverage |
|---|---|---|
| $> 15$ | 402 | 0.232 |
| 11—15 | 75 | 0.139 |
| 6—10 | 20 | 0.080 |
| $\leq 5$ | 6 | 0.041 |




a) Histogram of price target coverage b) Price target coverage

**Figure 1.** Analyst coverage.

Analyst price targets are observed on approximately 21.5% of trading days for the median firm. Price target coverage varies substantially across firms (Figure 1(a)) and is strongly increasing in analyst intensity (number of analyst recommendations), consistent with the infrequent nature of target price revisions.

Figure 1(b) shows the time series of the proportion of stocks with available analyst price target gap data over the sample period. The blue line represents the daily coverage across all tickers. The red dashed line indicates the median coverage, providing a benchmark for the typical level of data availability.

### *3.2. Informative weights and composite indices*

The weighting vector that maximizes the informativeness of price target gap ratios, which corresponds to the minimum-variance solution, is obtained by solving the following optimization problem:

$$\begin{aligned} \text{Minimize:} \quad & \sigma_p^2 = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\ \text{Subject to:} \quad & \sum_{i=1}^{n} w_i = 1 \\ & w_i \geq 0, \quad \forall i, \end{aligned} \tag{1}$$

where $w_i$ represents the weight of target gap $i$ and $\boldsymbol{\Sigma}$ is the $n \times n$ covariance matrix of target gaps. Each stock's statistics are calculated using only its own available observations, allowing the number of data points to vary across stocks. When computing covariances between stock pairs, only overlapping dates where both stocks have observations are used.

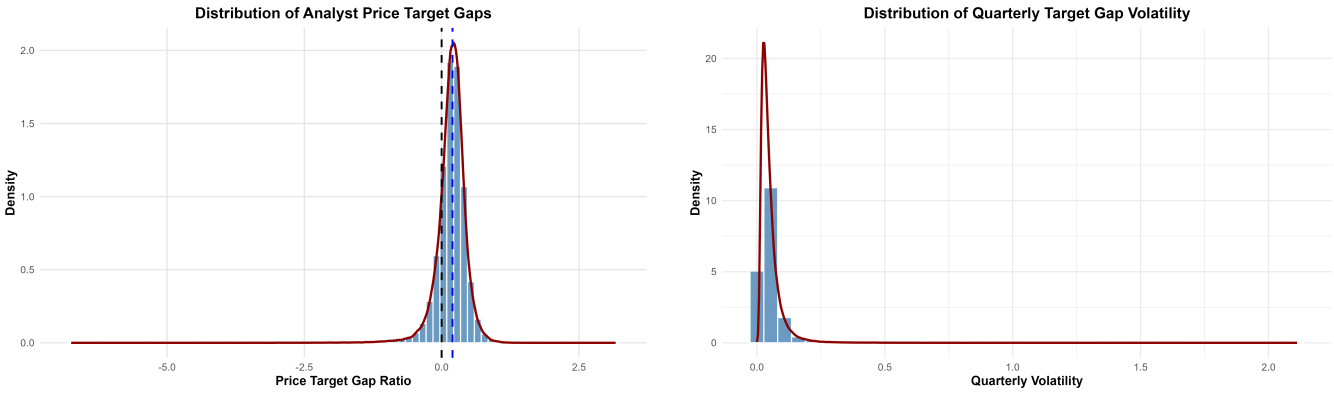
This approach maximizes the use of observed data without requiring interpolation or estimation of missing values. No stocks are excluded due to incomplete data, and calculations are not restricted to a single time window common to all assets. Instead, each calculation adapts to use all available observations for the specific stocks involved, whether computing individual statistics or pairwise covariances.

The optimization implements the Goldfarb-Idnani dual method for solving quadratic programming problems (Goldfarb and Idnani, 1983). The covariance matrix $\boldsymbol{\Sigma}$ is estimated from historical target gap

data using the sample covariance estimator:

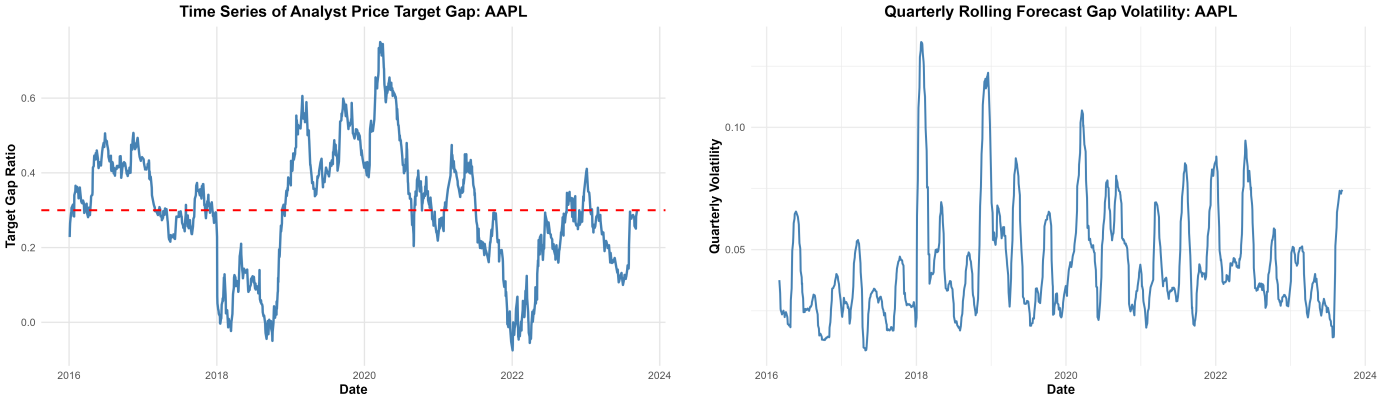
$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{g}_t - \bar{\mathbf{g}})(\mathbf{g}_t - \bar{\mathbf{g}})^{\top} \tag{2}$$

where $\mathbf{g}_t$ denotes the vector of target gaps at time $t$ (Figures 2 and 3), $\bar{\mathbf{g}}$ is the sample mean vector, and $T$ is the number of observations. Expected gaps are estimated as sample means.




a) Histogram of price target gaps b) Histogram of gap volatilities

**Figure 2.** Histograms of analyst price target gaps and gap volatilities.

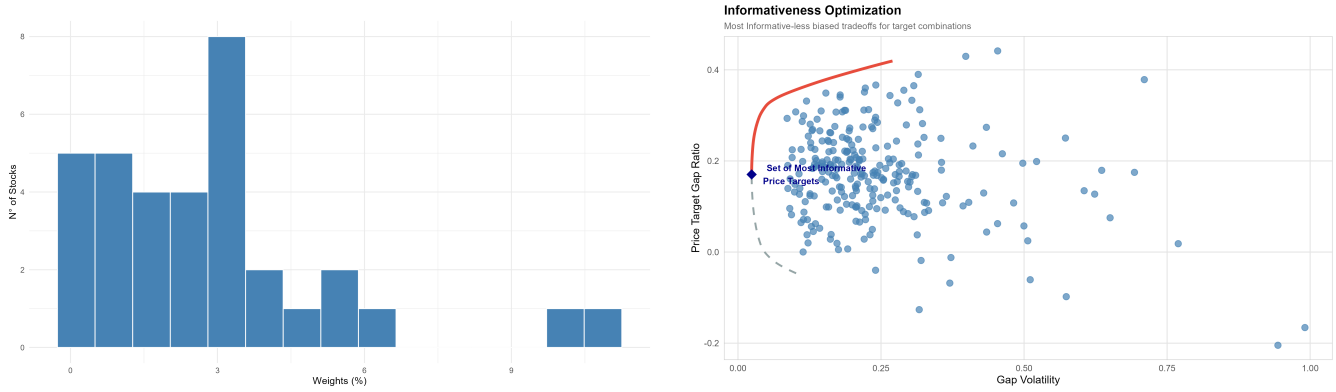



a) Price target gaps (dashed red line: median) b) Gap Volatilities

**Figure 3.** Price target gap and gap volatility for Apple stock.

The resulting precision weights exhibit a sparse structure (Figure 4). Although the optimization assigns formally positive weights to a large number of stocks, economically significant weight is concentrated in a limited subset. Approximately 30 stocks receive substantial weights, while the majority are assigned values on the order of $10^{-14}$ (effectively zero within machine precision), which reflect numerical artifacts of the optimization rather than active information inputs. This decay indicates substantial informational redundancy and heterogeneity in the precision of price target gaps across assets. Beyond the core set of high-precision signals, additional price targets contribute negligibly to

reducing overall dispersion in valuation gaps. This concentration suggests that portfolio construction can safely focus on high-precision signals without meaningful information loss.

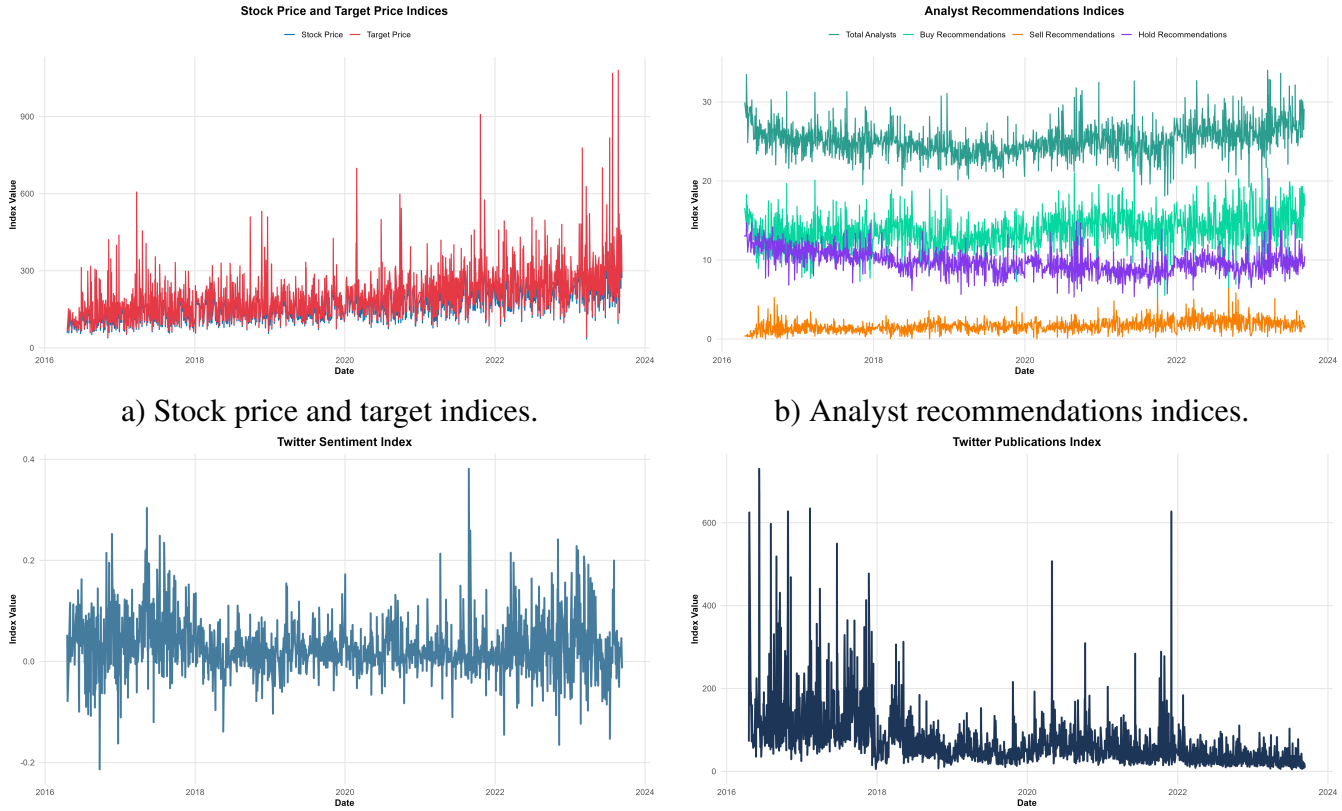



a) Distribution of informativeness weights b) Informative price targets

**Figure 4.** Informativeness weights.

The construction of the composite indices follows a multi-step procedure designed to ensure robustness and consistency across time. First, I filter out stocks with negligible minimum-variance weights (less than 0.0001), retaining only those assets that contribute meaningfully to the optimized portfolio. Second, to address the time-varying nature of data availability in my panel, I implement a dynamic weight-normalization procedure: on each trading date, the baseline minimum-variance weights are re-scaled proportionally such that they sum to unity conditional on the subset of stocks with non-missing observations on that date. This ensures that the index remains fully invested and comparable across time, even as the composition of reporting firms fluctuates. Third, I impose a minimum coverage requirement by excluding dates with fewer than five stocks reporting data, thereby mitigating potential biases arising from extremely sparse cross-sectional coverage. The resulting indices (stock prices, analyst target prices, recommendation counts, and Twitter sentiment metrics in Figure 5 and Table 7) are then normalized to a base value of 100 at the first date meeting the coverage criterion, facilitating straightforward interpretation and cross-index comparisons.

The proposed consistency weighting interprets greater temporal stability in analyst price target gaps as evidence of a higher-quality information environment. Because this information quality is a firm-level characteristic, it is expected to be reflected in other observable signals, including stock prices, analyst recommendations, and social-media sentiment and volume. The minimum-variance weights estimated from analyst price target gaps therefore serve as firm-specific precision measures and are applied uniformly across all constituent series. This logic is analogous to inverse-variance weighting, where precision determines the relative contribution of each observation to an aggregate; the key extension is that precision is treated as a property of the firm rather than of the individual series from which it is estimated, thereby down-weighting firms whose signals are comparatively noisy across every composite index.




a) Stock price and target indices.

b) Analyst recommendations indices.

c) Twitter sentiment index.

d) Twitter publications index.

**Figure 5.** Market indices: Price, analyst activity, and social sentiment.

**Table 7.** Summary statistics of stock indices.

| variable | n_obs | mean | median | sd | min | max |
|---|---|---|---|---|---|---|
| Buy recommendations (raw) | 1651 | 13.7054 | 13.6125 | 2.1349 | 4.5193 | 26.1478 |
| Buy recommendations (robust stand.) | 1651 | 82.7803 | 82.2195 | 12.8948 | 27.2968 | 157.9325 |
| Stock price (raw) | 1651 | 181.7237 | 166.8581 | 88.0353 | 34.1936 | 1058.8081 |
| Stock price (robust stand.) | 1651 | 308.0070 | 282.8111 | 149.2128 | 57.9555 | 1794.5944 |
| Sell recommendations (raw) | 1651 | 1.6722 | 1.5954 | 0.7475 | 0.0000 | 6.3935 |
| Sell recommendations (robust stand.) | 1651 | 472.6626 | 450.9612 | 211.2886 | 0.0000 | 1807.1828 |
| Price target (raw) | 1651 | 201.7818 | 185.0417 | 97.8978 | 41.2896 | 1080.2747 |
| Price target (robust stand.) | 1651 | 294.2693 | 269.8563 | 142.7697 | 60.2149 | 1575.4230 |
| Analyst recommendations (raw) | 1651 | 25.1366 | 24.9689 | 1.9533 | 18.0986 | 34.0127 |
| Analyst recommendations (robust stand.) | 1651 | 84.1123 | 83.5510 | 6.5361 | 60.5616 | 113.8132 |
| Twitter posts (raw) | 1651 | 79.0402 | 58.4341 | 75.5718 | 5.0456 | 730.0027 |
| Twitter posts (robust stand.) | 1651 | 109.6146 | 81.0377 | 104.8046 | 6.9973 | 1012.3829 |
| Twitter sentiment (raw) | 1651 | 0.0281 | 0.0204 | 0.0534 | -0.2135 | 0.3812 |
| Twitter sentiment (robust stand.) | 1651 | 53.2345 | 38.5599 | 100.9223 | -403.7571 | 721.0150 |

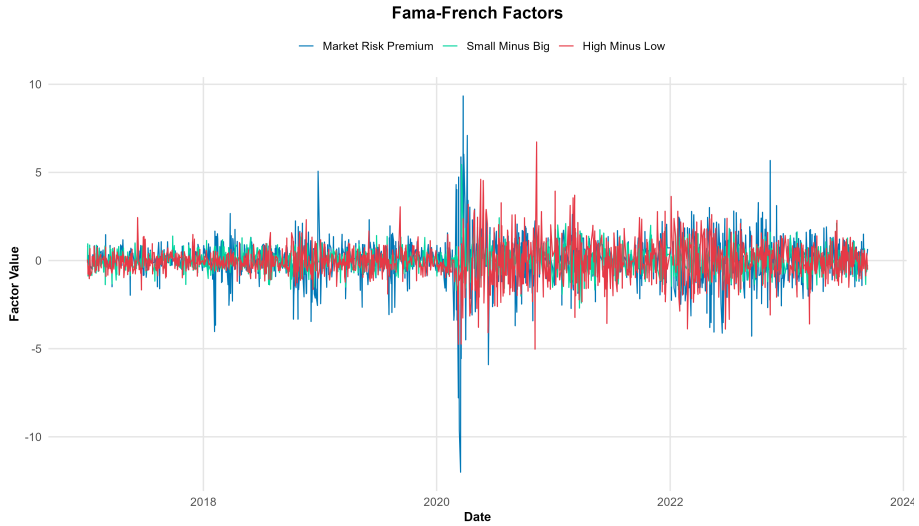
Note: This table presents descriptive statistics for the analyst, social-media and stock-price variables along with their robust standardized versions (*robust stand.*). All values are computed from daily observations over the sample period.

### *3.3. Fama-French factors and factor narratives*

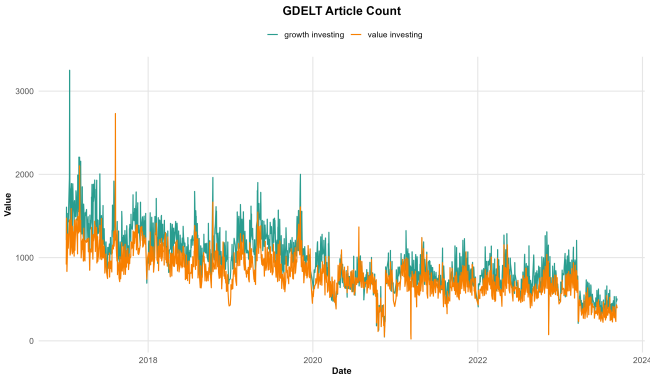
The Fama-French three-factor model data were obtained from Kenneth French's Data Library at the Tuck School of Business, Dartmouth College. The dataset includes daily observations of three key risk factors: the market risk premium (Mkt-RF), representing the excess return of the market portfolio over the risk-free rate; the size factor (SMB, small minus big), capturing the return differential between small and large capitalization stocks; and the value factor (HML, high minus low), measuring the return spread between high and low book-to-market ratio stocks. Over the sample period, the market risk premium exhibited a mean daily return of 0.037% with substantial volatility ranging from -12.01% to 9.34%. The SMB factor showed a near-zero mean return of 0.002% with a range of -3.53% to 5.45%, while the HML factor demonstrated a slightly negative mean return of -0.021%, varying between -5.03% and 6.73%. The risk-free rate (RF) averaged 0.006% daily, reflecting the low-interest rate environment over portions of the sample period. These factors serve as fundamental controls for systematic risk exposures in cross-sectional asset-pricing studies.

To quantify media attention surrounding investment styles, I collected news article counts from the GDELT Project (Global Database of Events, Language, and Tone), which monitors global news media across print, broadcast, and web sources in over 100 languages. Using the GDELT API v2 Doc interface, I performed daily queries for the search terms "growth investing" and "value investing" over the period from January 10, 2017, to November 1, 2025. To circumvent API rate-limiting constraints, I partitioned each calendar day into twenty-four one-hour intervals and retrieved up to 250 article records per interval. The data-collection process employed parallel processing with automated retry logic, incorporating exponential backoff delays ranging from 20 to 60 seconds to handle HTTP 429 rate-limit responses. Retrieved articles included metadata on publication title, date, URL, domain, source language, and country of origin. I removed duplicate entries based on unique URLs, retaining only distinct articles, and aggregated the final counts to the daily level to construct time-series measures of media coverage intensity for each investment style. This approach yielded comprehensive coverage of global news discourse on growth and value investing strategies across the sample period (Figure 6).

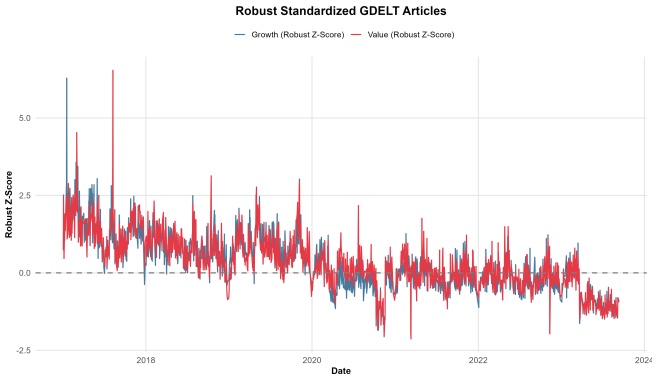
Table 8 presents descriptive statistics for the Fama-French three-factor model variables and media attention proxies. The sample comprises 1488 daily observations. The market risk premium (Mkt-RF) exhibits a positive mean return of 3.73 basis points per day with substantial volatility (SD = 1.25%), ranging from -12.01% to 9.34%. The size factor (SMB) shows a near-zero mean (0.19 basis points), while the value factor (HML) displays a slightly negative average return (-2.13 basis points) with comparable volatility (SD = 1.03%) to the market factor. The risk-free rate (RF) averages 0.62 basis points daily with minimal variation.




a) Fama-French factors




b) GDELT article count




c) GDELT articles robust standardized

**Figure 6.** Fama-French factors and GDELT article counts.

**Table 8.** Descriptive statistics for Fama-French factors and media attention variables.

| variable | N | Mean | Median | SD | Min | Max | Q1 | Q3 |
|---|---|---|---|---|---|---|---|---|
| HML | 1488 | -0.0213 | -0.0600 | 1.0337 | -5.0300 | 6.7300 | -0.5300 | 0.4900 |
| Mkt-RF | 1488 | 0.0373 | 0.0500 | 1.2511 | -12.0100 | 9.3400 | -0.4400 | 0.6625 |
| RF | 1488 | 0.0062 | 0.0000 | 0.0069 | 0.0000 | 0.0200 | 0.0000 | 0.0100 |
| SMB | 1488 | 0.0019 | -0.0100 | 0.6716 | -3.5300 | 5.4500 | -0.3900 | 0.3900 |
| Growth (raw) | 1488 | 973.8817 | 949.0000 | 352.3145 | 46.0000 | 3251.0000 | 719.7500 | 1197.2500 |
| Growth (robust stand.) | 1488 | 0.3545 | 0.2897 | 0.9179 | -2.0630 | 6.2872 | -0.3076 | 0.9365 |
| Value (raw) | 1488 | 807.5047 | 782.0000 | 274.2731 | 22.0000 | 2731.0000 | 626.0000 | 978.2500 |
| Value (robust stand.) | 1488 | 0.3806 | 0.2989 | 0.8783 | -2.1349 | 6.5402 | -0.2007 | 0.9273 |

Note: This table presents descriptive statistics for the Fama-French three-factor model variables (Mkt-RF, SMB, HML, RF) and media attention proxies (Growth, Value) along with their robust standardized versions (*robust stand.*). All values are computed from daily observations over the sample period.

.

Media attention variables show considerable dispersion. The growth-oriented media attention index averages 973.88 mentions (SD = 352.31), ranging from 46 to 3251, while value-oriented coverage averages 807.50 mentions (SD = 274.27), spanning 22 to 2731. The standardized versions (Growth (robust stand.) and Value (robust stand.)) exhibit positive means (0.35 and 0.38, respectively), suggesting periods of above-average media coverage are more prevalent in the sample. Both standardized measures display approximately unit standard deviations and right-skewed distributions, as evidenced by the substantial positive maximum values (6.29 and 6.54, respectively) relative to their negative minimums (−2.06 and −2.13).

The multivariate models (VAR, QVAR, and SVAR) are estimated on the intersection of three data sources merged by date. Because the target gap ratio used for weight estimation requires each observation's stock price to be compared against the analyst target price issued approximately one year prior, the effective end of well-identified index observations is constrained to approximately one year before the last fresh target price issuance in the Bloomberg data. When this stock index series is merged with the Fama-French daily factors and GDELT article counts, the final common sample comprises 1488 daily observations from January 3, 2017, to September 11, 2023. GDELT observations after September 2023 are therefore excluded from estimation, not by design but as a consequence of the date-intersection requirement.

### *3.4. Vector autoregression*

Granger causality (Granger, 1969) tests whether past values of variable $x$ contain statistically significant information for predicting variable $y$ beyond what is contained in $x$'s own lagged values. This concept is operationalized within the vector autoregression (VAR) framework. The reduced form of the VAR model is:

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} A_j \mathbf{y}_{t-j} + \mathbf{u}_t, \tag{3}$$

where $\mathbf{u}_t$ is the vector of estimated residuals and $\mathbf{y}_t$ $(k \times 1)$ is the vector of endogenous variables.

#### *Connectedness*

The net pairwise directional connectedness (NPDC) based on the quantile VAR framework, with a lag length as suggested by the Bayesian information criterion (BIC), identifies the direction of influence between two variables, indicating whether $x$ drives $y$ or is driven by it as in Dash et al. (2024), Chatziantoniou et al. (2021), Long et al. (2024), Khalfaoui et al. (2022), and Gabauer (2021). Connectedness metrics are derived from the $H$-step-ahead generalized forecast error variance decomposition of a quantile vector autoregression, QVAR(p) model as follows:

$$\mathbf{y}_t = \boldsymbol{\mu}(\tau) + \sum_{j=1}^{p} \boldsymbol{\Phi}_j(\tau) \mathbf{y}_{t-j} + \mathbf{u}_t(\tau), \tag{4}$$

where $\tau \in [0, 1]$ is the quantile of interest, $p$ is the lag length of the QVAR model, $\boldsymbol{\mu}(\tau)$ $(k \times 1)$ is the vector of quantile-specific intercepts, and $\boldsymbol{\Phi}_j(\tau)$ $(k \times k)$ is the quantile-specific autoregressive coefficient matrix at lag $j$.

Net pairwise directional connectedness (NPDC) analysis extends the Diebold-Yilmaz connectedness framework (Diebold and Yilmaz, 2012; Diebold and Yılmaz, 2014) by identifying the net direction of

shock transmission between variable pairs. NPDC analysis naturally lends itself to network representation, where nodes represent variables and directed, weighted edges denote the magnitude and direction of net spillover effects. Variables with predominantly positive NPDC values across multiple counterparties emerge as systemically important hubs; net transmitters that propagate shocks throughout the system. Conversely, variables with negative NPDC values function as peripheral receivers. Standard Granger causality tests might establish that analyst upgrades predict subsequent positive returns. However, NPDC analysis could reveal that analyst recommendations are actually net receivers of information from social-media sentiment, suggesting analysts respond to public sentiment rather than generating independent information.

### *3.5. Structural VAR*

A structural vector autoregression (SVAR) is used to identify and interpret causal relationships among multiple time series variables by imposing restrictions on a standard VAR model (Sims, 1980; Sharif et al., 2025). The SVAR is as follows:

$$B\mathbf{y}_t = \Gamma_0 + \Gamma_1 \mathbf{y}_{t-1} + \cdots + \Gamma_p \mathbf{y}_{t-p} + \boldsymbol{\omega}_t, \tag{5}$$

where $\mathbf{y}_t$ is a vector containing nine variables. I employ Cholesky ordering, where the relationship between the structural and the reduced VAR shocks are described by $\mathbf{u}_t = B^{-1}\boldsymbol{\omega}_t$. $\Gamma_1, \ldots, \Gamma_p$ are matrices of coefficients to be estimated, and $\boldsymbol{\omega}_t = \left[\omega_t^1 \cdots \omega_t^9\right]'$ is a vector of serially and mutually uncorrelated structural innovations.

## 4. Results

### *4.1. Analysis of Granger causality relationships*

The multivariate VAR Granger causality tests (Table 9) reveal significant predictive relationships among market indices, analyst recommendations, social-media metrics, and risk factors. Both buy recommendations ($F = 17.52$, $p < .001$) and sell recommendations ($F = 32.03$, $p < .001$) cause the stock price index, suggesting that analyst recommendations contain predictive information about subsequent price movements. The joint test of all predictors ($F = 17.49$, $p < .001$) further confirms the collective importance of these variables in forecasting stock prices.

**Table 9.** Granger causality test results (statistically significant relationships).

| Dependent ⇐ Independent | F | df | *p*-value | $\chi^2$ | df | *p*-value |
|---|---|---|---|---|---|---|
| *Stock Price Index* | | | | | | |
| ⇐ Buy Recommendations | 17.52 | 1, 1475 | <.001*** | 17.52 | 1 | <.001*** |
| ⇐ Sell Recommendations | 32.03 | 1, 1475 | <.001*** | 32.03 | 1 | <.001*** |
| ⇐ ALL | 17.49 | 10, 1475 | <.001*** | 174.94 | 10 | <.001*** |
| *Target Price Index* | | | | | | |
| ⇐ Stock Price Index | 5.67 | 1, 1475 | .017* | 5.67 | 1 | .017* |
| ⇐ Buy Recommendations | 17.26 | 1, 1475 | <.001*** | 17.26 | 1 | <.001*** |
| ⇐ Sell Recommendations | 34.31 | 1, 1475 | <.001*** | 34.31 | 1 | <.001*** |
| ⇐ ALL | 18.65 | 10, 1475 | <.001*** | 186.53 | 10 | <.001*** |
| *Buy Recommendations* | | | | | | |
| ⇐ Sell Recommendations | 7.44 | 1, 1475 | .006** | 7.44 | 1 | .006** |
| ⇐ Twitter Volume | 4.37 | 1, 1475 | .037* | 4.37 | 1 | .037* |
| ⇐ ALL | 14.07 | 10, 1475 | <.001*** | 140.72 | 10 | <.001*** |
| *Sell Recommendations* | | | | | | |
| ⇐ Stock Price Index | 8.10 | 1, 1475 | .004** | 8.10 | 1 | .004** |
| ⇐ Target Price Index | 10.07 | 1, 1475 | .002** | 10.07 | 1 | .002** |
| ⇐ Buy Recommendations | 7.01 | 1, 1475 | .008** | 7.01 | 1 | .008** |
| ⇐ ALL | 10.88 | 10, 1475 | <.001*** | 108.76 | 10 | <.001*** |
| *Twitter Volume* | | | | | | |
| ⇐ Buy Recommendations | 9.40 | 1, 1475 | .002** | 9.40 | 1 | .002** |
| ⇐ Sell Recommendations | 10.71 | 1, 1475 | .001** | 10.71 | 1 | .001** |
| ⇐ Value | 27.39 | 1, 1475 | <.001*** | 27.39 | 1 | <.001*** |
| ⇐ ALL | 14.67 | 10, 1475 | <.001*** | 146.74 | 10 | <.001*** |
| *Twitter Sentiment* | | | | | | |
| ⇐ Stock Price Index | 13.76 | 1, 1475 | <.001*** | 13.76 | 1 | <.001*** |
| ⇐ Target Price Index | 14.26 | 1, 1475 | <.001*** | 14.26 | 1 | <.001*** |
| ⇐ ALL | 3.52 | 10, 1475 | <.001*** | 35.17 | 10 | <.001*** |
| *Mkt-RF* | | | | | | |
| ⇐ Sell Recommendations | 7.97 | 1, 1475 | .005** | 7.97 | 1 | .005** |
| *SMB* | | | | | | |
| ⇐ Sell Recommendations | 4.48 | 1, 1475 | .034* | 4.48 | 1 | .034* |
| ⇐ Mkt-RF | 4.64 | 1, 1475 | .031* | 4.64 | 1 | .031* |
| *HML* | | | | | | |
| ⇐ Buy Recommendations | 7.08 | 1, 1475 | .008** | 7.08 | 1 | .008** |
| ⇐ Sell Recommendations | 4.10 | 1, 1475 | .043* | 4.10 | 1 | .043* |

Granger causality test results (statistically significant relationships) (*continuation*)

| Dependent ⇐ Independent | F | df | *p*-value | $\chi^2$ | df | *p*-value |
|---|---|---|---|---|---|---|
| *Growth* | | | | | | |
| ⇐ Buy Recommendations | 16.39 | 1, 1475 | <.001*** | 16.39 | 1 | <.001*** |
| ⇐ Sell Recommendations | 11.43 | 1, 1475 | <.001*** | 11.43 | 1 | <.001*** |
| ⇐ Twitter Volume | 4.09 | 1, 1475 | .043* | 4.09 | 1 | .043* |
| ⇐ Value | 23.12 | 1, 1475 | <.001*** | 23.12 | 1 | <.001*** |
| ⇐ ALL | 7.61 | 10, 1475 | <.001*** | 76.11 | 10 | <.001*** |
| *Value* | | | | | | |
| ⇐ Target Price Index | 4.69 | 1, 1475 | .031* | 4.69 | 1 | .030* |
| ⇐ Buy Recommendations | 18.94 | 1, 1475 | <.001*** | 18.94 | 1 | <.001*** |
| ⇐ Sell Recommendations | 14.78 | 1, 1475 | <.001*** | 14.78 | 1 | <.001*** |
| ⇐ Growth | 65.19 | 1, 1475 | <.001*** | 65.19 | 1 | <.001*** |
| ⇐ ALL | 13.08 | 10, 1475 | <.001*** | 130.85 | 10 | <.001*** |

Notes: Granger causality test results from a BIC-selected VAR(1), showing only statistically significant relationships. * $p$¡.05; ** $p$¡.01; *** $p$¡.001. ALL denotes a joint test of all predictors.

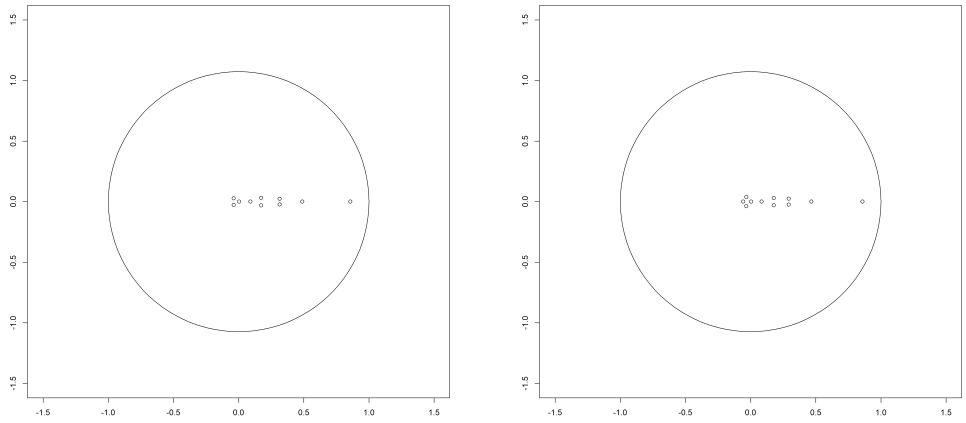
Similarly, the target price index is caused by the stock price index (F = 5.67, p = .017), buy recommendations (F = 17.26, p < .001), and sell recommendations (F = 34.31, p < .001), indicating that analysts adjust their target prices in response to both market movements and their own recommendation changes. Sell recommendations are significantly predicted by the stock price index (F = 8.10, p = .004) and target price index (F = 10.07, p = .002), suggesting that analysts modify their sell ratings in response to market conditions. Interestingly, buy and sell recommendations exhibit mutual Granger causality (F = 7.44, p = .006 and F = 7.01, p = .008, respectively), indicating interdependence in analyst decision-making processes, possibly reflecting herding behavior or consensus formation among financial analysts. Buy recommendations (F = 9.40, p = .002) and sell recommendations (F = 10.71, p = .001) cause the number of Twitter publications, suggesting that analyst activities stimulate social-media discourse. Conversely, Twitter causes buy recommendations (F = 4.37, p = .037), indicating a feedback loop where social-media attention influences analyst behavior. The strong relationship between Twitter and the value narrative (F = 27.39, p < .001) suggests that social-media discussions may be particularly focused on value-oriented stocks or investment strategies.

Twitter sentiment shows different patterns, being predicted by both the stock price index (F = 13.76, p < .001) and target price index (F = 14.26, p < .001), but without significant reverse causality. This asymmetry implies that market movements shape investor sentiment on social-media, but social media sentiment does not significantly predict subsequent market movements, at least at the one-period lag examined here.

The market risk premium (Mkt-RF) is influenced by sell recommendations (F = 7.97, p = .005), while the size factor (SMB) is predicted by both sell recommendations (F = 4.48, p = .034) and Mkt-RF (F = 4.64, p = .031). The value factor (HML) responds to both buy recommendations (F = 7.08, p = .008) and sell recommendations (F = 4.10, p = .043). These findings suggest that analyst recommendations contain information about systematic risk factors beyond idiosyncratic stock-level effects, potentially reflecting analysts' incorporation of broader market conditions into their recommendations.

The growth and value narratives exhibit the strongest mutual relationship (F = 23.12 and F = 65.19, p < .001), consistent with their theoretical construction as opposite investment strategies. Both narratives are significantly predicted by analyst recommendations, with the growth index caused by buy recommendations (F = 16.39, p < .001) and sell recommendations (F = 11.43, p < .001), and the value index similarly predicted by buy recommendations (F = 18.94, p < .001) and sell recommendations (F = 14.78, p < .001). These relationships suggest that analyst recommendations capture style-specific information that predicts differential performance between growth and value strategies. The value narrative is also predicted by the target price index (F = 4.69, p = .031), while the growth index shows sensitivity to Twitter publications (F = 4.09, p = .043), potentially reflecting different information channels for these investment styles, with value investing more responsive to formal analyst projections and growth investing more influenced by social-media attention and momentum.

### *4.2. Interconnectedness*

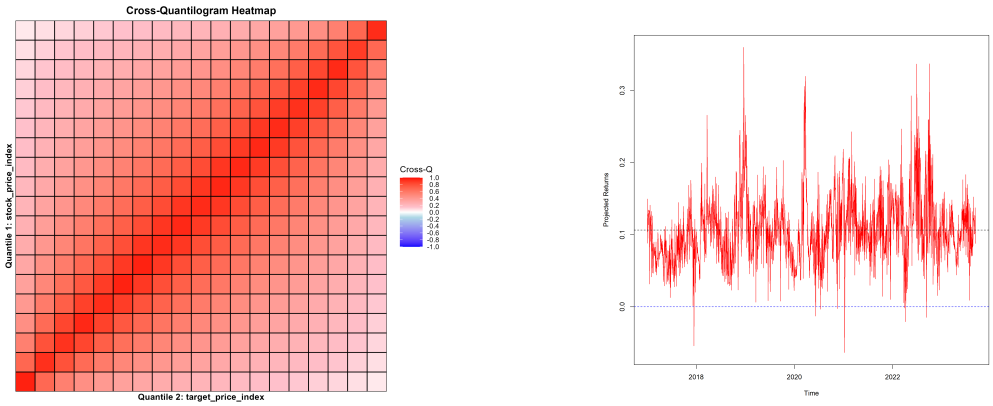


(a) (b)

**Figure 7.** Stationarity. All eigenvalues lie within the unit circle, indicating the stability of the VAR model. The variables included in (a) are: projected returns, buy recommendations, sell recommendations, number of Twitter publications, Twitter sentiment, growth, value and Fama-French factors. The variables included in (b) are: buy recommendations, sell recommendations, number of Twitter publications, Twitter sentiment, growth, value and Fama-French factors.

The estimated VAR model is stable, as all roots lie inside the unit circle (Figure 7). Target prices and stock prices are highly and statistically significantly correlated across all quantiles (Figure 8, panel (a)), leading to near-singular matrices in vector autoregression models. To address this issue, I use the index of projected stock returns, which captures this relationship while retaining a meaningful economic interpretation. This variable is calculated as the difference between the target price and the stock price, divided by the stock price (Figure 8, panel (b)). Granger causality tests are fundamentally designed to detect linear predictive relationships in the conditional mean of a distribution. Quantile vector autoregression (QVAR) models explicitly model the entire conditional distribution of the dependent

variable at specific quantiles. QVAR analysis at $\tau = 0.9$ examines whether past values of $x$ help predict the 90th percentile of $y$'s conditional distribution, effectively testing for causal relationships during periods of extreme positive returns or heightened volatility. This distinction is particularly salient in financial markets, where contagion effects, herding behavior, and risk transmission often manifest asymmetrically during tail events rather than normal market conditions.

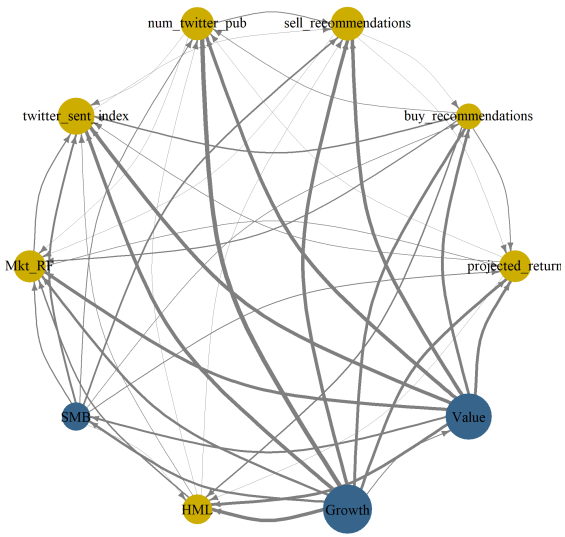



(a) (b)

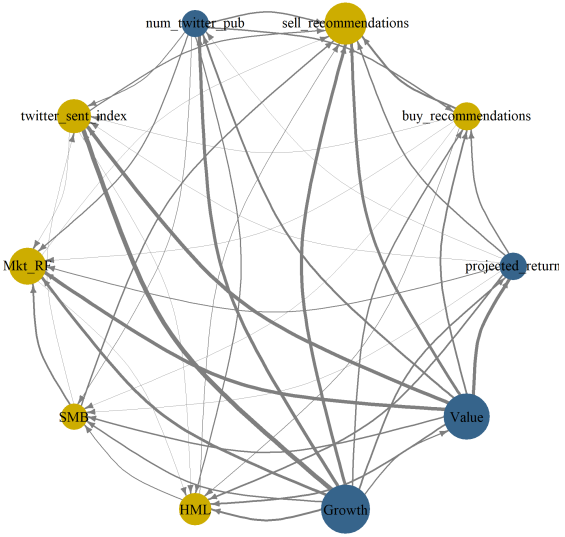
**Figure 8.** (a) Heatmap visualization of the cross-quantilogram matrix. The X- axis represents the quantile values of the price target index and the Y-axis represents the quantile values of the stock price index. The critical values are obtained by stationary bootstrap. (b) Time series plot of the projected stock return.

The NPDC network analysis (Figure 9) reveals a hierarchical information flow structure where narrative-based variables and social-media activity serve as primary information transmitters, while traditional asset-pricing factors and analyst recommendations emerge as net receivers. I interpret the net transmitter status of Twitter publications, projected returns, and both value and growth narratives as evidence that narrative dynamics and attention measures drive subsequent market outcomes rather than merely reflecting them. The net receiver status of the Fama-French factors (HML, SMB, Mkt-RF) suggests that these fundamental risk premia are, at least partially, endogenous to narrative and sentiment shifts, challenging the traditional view of factors as purely exogenous risk sources.

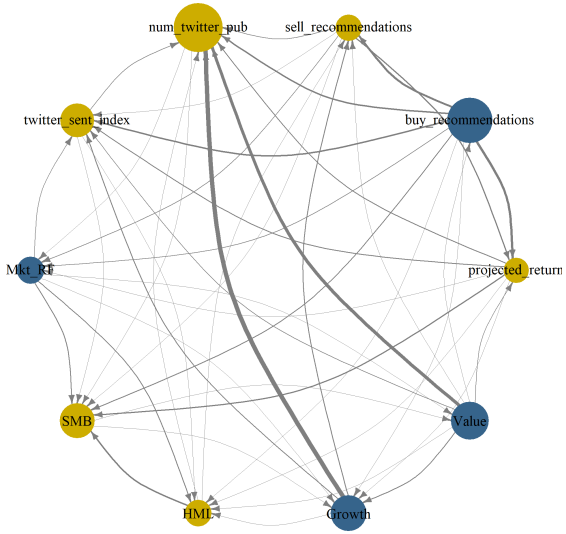
The particularly strong causal links from both growth and value narratives to Twitter sentiment indicate that structured narrative content shapes collective investor sentiment expressed on social-media platforms, supporting a narrative-to-sentiment transmission mechanism. The value narrative's dominant influence on both market equity premium (Mkt-RF) and Twitter sentiment suggests that value-oriented discourse has direct pricing implications for aggregate market returns while simultaneously affecting investor mood. Conversely, the growth narrative's strongest connections to Twitter sentiment and sell recommendations indicate that growth-oriented discourse primarily operates through sentiment channels and influences analyst pessimism, possibly reflecting heightened skepticism during periods when growth narratives intensify. Overall, I interpret this network structure as evidence of a narrative-driven information cascade where textual content and social-media activity precede and potentially cause movements in both market factors and professional analyst opinions.




(a) QVAR at $\tau = 0.1$




(b) QVAR at $\tau = 0.9$




(c) QVAR at $\tau = 0.5$

**Figure 9.** Net pairwise directional connectedness (NPDC) based on the quantile VAR framework with a lag length as suggested by BIC and 6-step-ahead projection with a 150-day rolling window.

To assess whether the hierarchical structure documented at $\tau = 0.9$ is a feature of upper-tail dynamics or a regime-invariant property of the system, I re-estimate the NPDC network at the median ($\tau = 0.5$) and at the lower tail ($\tau = 0.1$).

The narrative-to-factor transmission is robust across quantiles. Growth and value media attention remain net transmitters at all three $\tau$, and HML remains a net receiver throughout. The core claim of the paper, that systematic risk factors are downstream of narrative dynamics, therefore does not hinge on the choice of quantile.

State dependence is concentrated in a specific subset of variables: SMB, the market premium, Twitter publication volume, buy recommendations, and projected returns each switch roles across quantiles. During extreme negative shocks ($\tau = 0.1$), SMB reverses from a net receiver to a net transmitter, indicating that the size factor itself propagates disturbances when small-firm distress dominates, rather than passively absorbing narrative shocks. Mkt-RF is a net receiver at both tails but a net transmitter at the median, consistent with the market factor acting as a conduit in normal conditions and as a shock absorber in extremes. Twitter publication volume and projected returns are net transmitters only at $\tau = 0.9$; at $\tau = 0.1$ and $\tau = 0.5$, they become receivers, implying that attention surges and analyst-implied upside drive the system only during euphoric regimes and follow it otherwise. Sell recommendations remain net receivers at all quantiles.

### *4.3. Impulse-response functions*

To identify the structural VAR model, I impose a recursive identification scheme based on the Cholesky decomposition, which requires ordering variables from most exogenous to most endogenous. Cholesky decomposition requires ordering variables from most exogenous (ordered first, affecting all subsequent variables contemporaneously but not vice versa) to most endogenous (ordered last, contemporaneously affected by all preceding variables). Following the network's net transmitter-receiver structure, I impose the following ordering: (1) Twitter publications (number of posts), (2) projected returns, (3) value narrative, (4) growth narrative, (5) Twitter sentiment, (6) sell recommendations, (7) buy recommendations, (8) Mkt-RF, (9) SMB, and (10) HML.

I position narrative variables and social-media activity in the early positions because they emerge as net transmitters of information, suggesting they represent exogenous shifts in investor attention and discourse. Twitter sentiment is ordered after the narratives but before analyst recommendations, consistent with the strong causal links from narratives to sentiment observed in the network plot. I place analyst recommendations (sell and buy) in intermediate positions, as they appear to respond to both narratives and sentiment while potentially affecting factor returns. Finally, I position the Fama-French factors (Mkt-RF, SMB, HML) in the most endogenous positions, as they emerge as net receivers in the network, suggesting that traditional risk premia respond contemporaneously to narrative shifts, sentiment changes, and analyst opinion revisions. This data-driven identification strategy provides an empirically grounded alternative to ad-hoc orderings, though I acknowledge that the NPDC analysis captures directional information flow rather than strict exogeneity. The baseline Cholesky ordering is motivated by the $\tau = 0.9$ NPDC network, where both SMB and Mkt-RF are unambiguous net receivers; as a robustness check, re-estimating the SVAR with Mkt-RF placed first in the ordering yields a growth-to-SMB impulse response that is virtually identical in sign, magnitude, and significance to the baseline specification (Figure 10(e)).

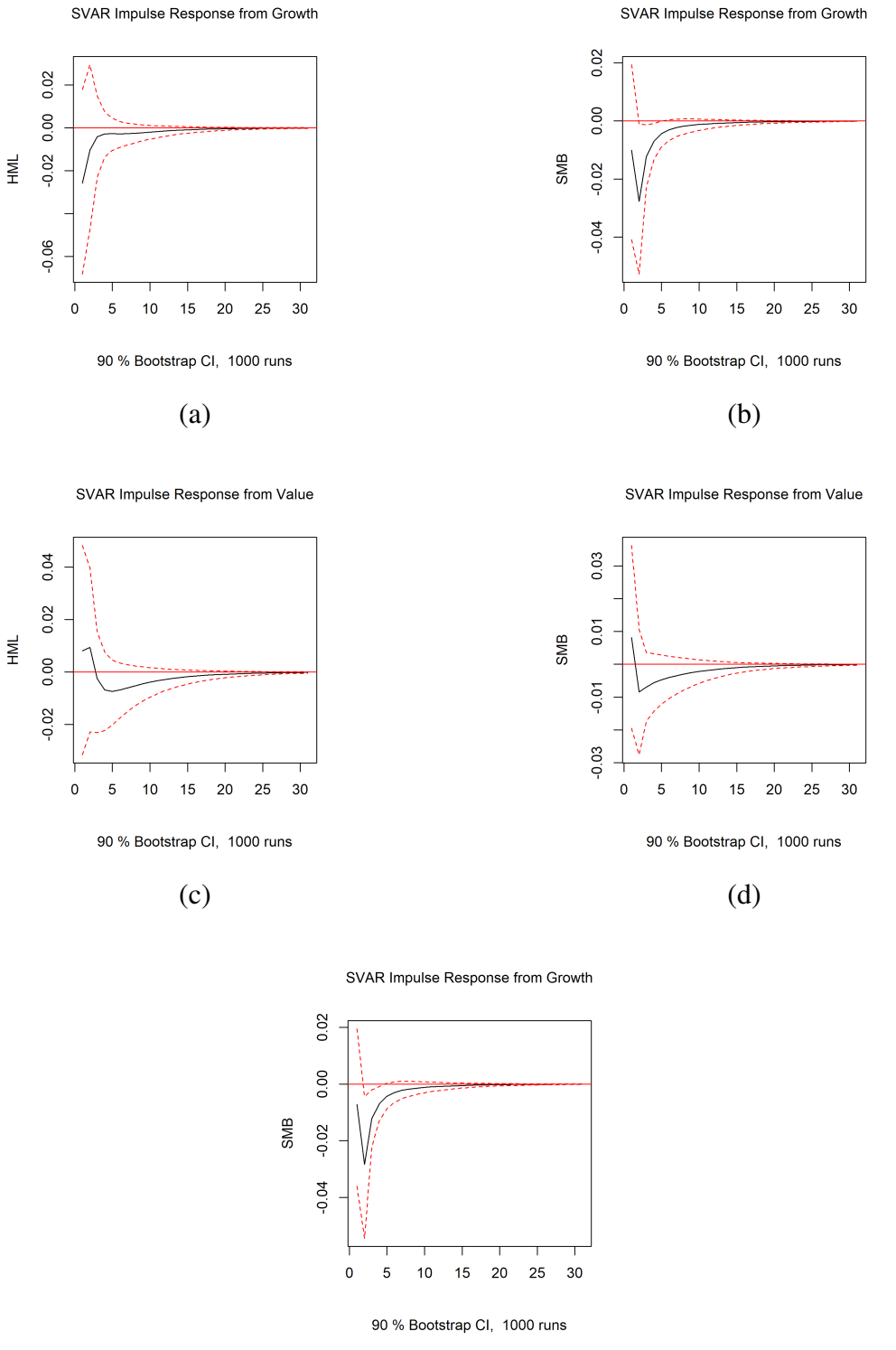
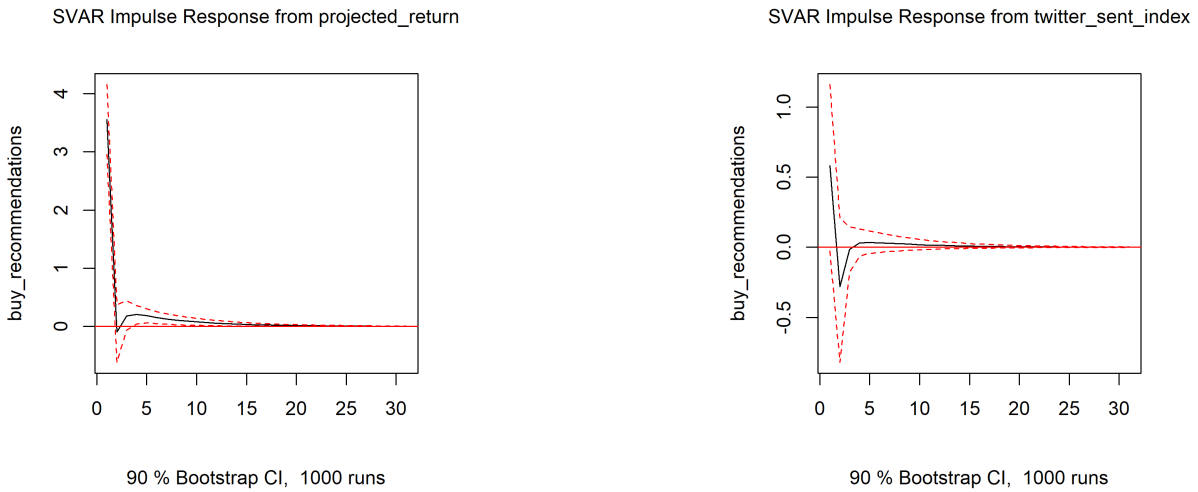



**Figure 10.** (a) Response of HML to an impulse from the growth narrative; (b) response of SMB to an impulse from the growth narrative; (c) response of HML to an impulse from the value narrative; (d) response of SMB to an impulse from the value narrative; (e) response of SMB to an impulse from the growth narrative placing Mkt-RF first in the SVAR ordering.

The impulse response analysis (Figures 11 and 10) reveals several significant relationships between narratives, sentiments, analyst behavior, and factors. Analyst buy recommendations respond significantly and positively to both projected return shocks and Twitter sentiment shocks. The projected returns effect confirms that analysts incorporate fundamental expectations into their recommendations through traditional valuation channels. The Twitter sentiment effect indicates that social-media discourse systematically influences professional analyst output, either through incremental information content or behavioral contagion mechanisms. Among the narrative-factor relationships, I find that the size premium (SMB) responds significantly and negatively to growth narrative shocks, indicating that growth-oriented discourse drives capital flows toward large-cap stocks and compresses small-firm returns.




(a) (b)

**Figure 11.** (a) Response of buy recommendations to an impulse from projected returns; (b) response of buy recommendations to an impulse from Twitter sentiment.

Because my narrative series are aggregated exclusively over large-cap S&P 500 constituents, the negative impulse response of SMB to growth-narrative shocks should not be read as evidence that small-cap sentiment moves the size factor. Rather, I interpret it as a spillover channel: shifts in the growth narrative attached to large, visible firms are transmitted into the large-minus-small return differential that defines SMB, consistent with factor construction loading on the same large-cap universe that dominates investor attention. This reframing narrows the claim but preserves the identification, since the exogenous variation I exploit originates in professional analyst discourse about index-member firms. The remaining narrative effects, i.e., value narratives on HML and SMB, and growth narratives on HML, exhibit theoretically consistent directional patterns but lack statistical significance. To gauge the economic magnitude of the statistically significant growth-to-SMB channel, I read the impulse response in Figure 10(b) in basis points, exploiting the fact that the response is measured in the native units of SMB (percent per day) and that the growth innovation is a one-standard-deviation shock by construction of the robust-standardized series. A one-SD growth shock reduces SMB by approximately 1.0 basis point on day one and by 2.8 basis points at the trough on day two. The response decays monotonically

and the 90% bootstrap confidence band crosses zero by roughly day 12. Cumulating the daily responses over this horizon gives a total SMB displacement of approximately −7 basis points per one-SD Growth shock.

These magnitudes are modest relative to SMB's unconditional daily volatility of 67.16 basis points (Table 8): the peak impact is about 4.2% of one daily SD, and the cumulative 12-day response is about 10% of one daily SD. They are, however, large relative to SMB's near-zero unconditional daily mean of 0.19 basis point, so the channel is economically non-trivial for factor-timing applications even though a single narrative innovation does not, on its own, dominate daily SMB variation. I do not scale these numbers to an annualized strategy return: growth innovations are serially correlated, shocks of one-SD magnitude do not arrive independently every period, and the large-cap S&P 500 constituent universe over which I aggregate narratives biases a live small-minus-big implementation upward through survivorship.

These findings suggest that narratives constitute empirically relevant drivers of both information intermediation and systematic risk premia, with varying degrees of statistical detectability across different transmission channels.

## 5. Discussion and conclusions

This study investigates information transmission mechanisms in financial markets by examining how narratives, analyst recommendations, and social media sentiment interact with systematic risk factors. Using variance-weighted composite indices and quantile vector autoregression with network analysis, I establish that narratives function as systematic information transmitters rather than passive reflectors of market conditions.

Fama-French risk factors are endogenous to narrative dynamics rather than purely exogenous risk exposures. The Granger causality tests reveal that analyst recommendations systematically predict subsequent movements in HML, SMB, and the market risk premium. More critically, the NPDC network analysis identifies all three Fama-French factors as net receivers of information, positioned at the periphery of the information flow network rather than as autonomous systematic risk sources. Traditional frameworks, including the intertemporal capital asset pricing model (ICAPM) (Merton , 1973) and characteristic-based factor models (Fama and French, 1996, 2016), treat factors as exogenous state variables or fundamental risk exposures that determine equilibrium risk premia. My evidence suggests an alternative interpretation: factors are endogenous responses to narrative transmission mechanisms. When analyst recommendations shift, reflecting changes in professional narratives about firm prospects, industry dynamics, or macroeconomic conditions, these narratives propagate through the market, generating systematic movements in factor returns.

The implications extend beyond statistical predictability to theoretical interpretation. Factor risk premia may represent a mixture of fundamental risk compensation and systematic mispricing driven by narrative transmission. This interpretation aligns with recent evidence from Bybee et al. (2023), who demonstrated that narrative-based factors constructed from Wall Street Journal text outperform traditional characteristic-based factors in explaining anomaly returns. My findings suggest a mechanism underlying their result: narrative factors may outperform because they capture the transmission processes that generate endogenous factor dynamics, positioning them higher in the causal chain than downstream characteristic-based constructs like HML and SMB.

A pattern emerges regarding the hierarchical structure of information transmission in financial markets. Professional analyst narratives originate pricing-relevant signals, which subsequently affect both market prices and retail social-media discourse. Analyst recommendations cause stock prices and predict Twitter publication volume. Conversely, Twitter sentiment exhibits no significant predictive power for subsequent price movements, and both stock prices and target prices predict future Twitter sentiment rather than vice versa. This hierarchical structure validates the design choices in recent narrative asset-pricing research. Bybee et al. (2023) constructed systematic risk factors exclusively from professional financial journalism (Wall Street Journal), explicitly excluding retail-generated content. My findings provide empirical justification for this approach: if professional narratives originate information while retail narratives primarily reflect market movements, then factor construction should prioritize professional sources. Had Bybee et al. (2023) instead used Twitter text to construct narrative factors, my evidence suggests the resulting factors would exhibit minimal pricing power, as social-media discourse follows rather than leads market dynamics.

My findings regarding professional narratives also connect to the experimental evidence of Johnson and Tuckett (2022), who demonstrated that even professional financial analysts are susceptible to narrative-based expectation formation, projecting past trends into the future in ways inconsistent with rational expectations. My time-series evidence reveals the aggregate market consequences of this individual-level behavior. When professional analysts' narrative-driven expectations manifest in recommendation changes, these recommendations cause both prices and systematic factors. The combination of Johnson and Tuckett's micro-level experiments with my macro-level time-series evidence provides convergent validation across methodological approaches and levels of analysis. The growth and value investment style narratives exhibit particularly rich dynamics in my network analysis. Both narratives emerge as net transmitters of information, and they are mutually causal, indicating bidirectional predictive relationships. The growth narrative predicts Twitter publication volume, while the value narrative shows strong predictive power for Twitter sentiment. The structural VAR impulse response analysis reveals that shocks to the growth narrative generate significant negative responses in the SMB factor, consistent with growth-oriented discourse directing capital flows toward large-capitalization stocks. If growth and value narratives partially drive the factor returns themselves, then HML and SMB may represent endogenous constructs that emerge from collective narrative sense-making rather than purely reflecting underlying economic risks. Recent theoretical work on narrative economics (Hirshleifer, 2020; Shiller, 2019, 2015) emphasized how stories become viral and coordinate expectations across economic agents. My evidence provides an empirical demonstration of this coordination mechanism in the specific context of investment style narratives and systematic factor dynamics.

Constructing variance-minimizing composite indices draws on insights from the analyst forecast consistency literature, particularly Hilary and Hsu (2013)'s demonstration that consistently biased forecasts provide more valuable information than erratic predictions. By weighting constituent signals according to their consistency properties (inverse variance weighting), I construct aggregate indices that isolate systematic information transmission from idiosyncratic noise. The resulting indices exhibit substantial weight concentration: approximately 30 stocks receive economically meaningful weights in the price target gap index, while the majority are assigned near-zero weights, reflecting heterogeneity in signal precision across securities. This concentration pattern has several implications. First, it suggests that information quality varies dramatically across the cross-section of stocks, with a limited

subset of highly covered, liquid securities providing high-precision signals that dominate composite indices. Second, it validates the approach to handling missing data: rather than excluding stocks with incomplete coverage or interpolating missing values, I compute covariances using available overlapping observations and allow the optimization to down-weight imprecise signals. This data-adaptive approach maximizes information utilization while maintaining statistical rigor.

The application of quantile vector autoregression to the connectedness analysis addresses a key limitation of standard Granger causality tests, which are restricted to predictability in the conditional mean. Financial market dynamics frequently exhibit asymmetries and nonlinearities that manifest differently across the return distribution, and the QVAR framework makes these state-dependent patterns directly observable. Estimating the NPDC network at $\tau = 0.1$, $\tau = 0.5$, and $\tau = 0.9$ produces two results that would be invisible in a mean-conditional analysis.

First, the narrative block is regime-invariant. Growth and value media attention remain net transmitters, and HML remains a net receiver, at all three quantiles. The central claim of the paper, that narrative variables sit upstream of the Fama-French factors in the information flow, therefore does not depend on the market state. This symmetry is itself informative: the forcefulness of narrative transmission does not scale with the sign of the shock, which argues against interpreting narrative spillovers through an asymmetric-processing lens at the market-aggregate level.

Second, the asymmetry that the data do support is located in the factor block rather than in the narrative block. SMB switches from a net receiver at $\tau = 0.5$ and $\tau = 0.9$ to a net transmitter at $\tau = 0.1$, indicating that during extreme downturns, the size factor originates rather than absorbs systemic information. The market premium is a net receiver at both tails but a net transmitter at the median, consistent with Mkt-RF acting as a conduit in normal conditions and as an absorber in extremes. Twitter publication volume and projected returns transmit only at $\tau = 0.9$ and become receivers at the median and lower tail, so attention surges and analyst-implied upside drive the system only under euphoric conditions and follow it otherwise. Sell recommendations are net receivers at every quantile, while buy recommendations transmit at the tails but not at the median.

Taken together, these results refine rather than reject the asymmetry documented in the mean-conditional Granger tests. The sell-versus-buy asymmetry, i.e., sell recommendations cause Mkt-RF, SMB, and HML with larger F-statistics than buy recommendations on several channels, captures a real feature of downside transmission in the conditional mean. What the quantile network adds is a locus: the downside amplification operates through a reconfiguration of the factor block (most clearly, SMB at $\tau = 0.1$), not through a strengthening of narrative channels.

For portfolio construction, the endogeneity of traditional factors to narrative dynamics suggests that factor exposure management should incorporate narrative monitoring. If growth investment narratives cause SMB, then investors can improve factor timing by tracking media attention to investment styles and adjusting size exposures accordingly. Similarly, the strong predictive power of analyst recommendations for factor returns suggests that systematic incorporation of recommendation changes into factor momentum strategies could enhance performance.

For risk modeling, the evidence that factors are net receivers of information from narratives implies that traditional factor models may misattribute systematic risk. If factor risk premia partially reflect narrative-driven capital flows rather than purely compensate for fundamental risks, then historical factor performance may not reliably guide future risk assessments.

## Use of AI tools declaration

The author declares he has used Artificial Intelligence (AI) tools in the creation of this article. AI tools used: Google Scholar for collection and analysis of references, ChatGPT for grammar and spelling.

## Acknowledgments

The author would like to thank Michael González for his assistance in automating GDELT queries.

## Conflict of interest

The author declares no conflicts of interest related to this paper.

## References

Afzal T, Afridi MA, Jan MN (2025) Integrating LSTM with Fama-French six factor model for predicting portfolio returns: Evidence from Shenzhen stock market China. *Data Sci Financ Econ* 5: 177–204. https://doi.org/10.3934/DSFE.2025009

Bybee L, Kelly B, Su Y (2023) Narrative asset pricing: Interpretable systematic risk factors from news text. *Rev Financ Stud* 36: 4759–4787. https://doi.org/10.1093/rfs/hhad042

Chatziantoniou I, Gabauer D, Stenfors A (2021) Interest rate swaps and the transmission mechanism of monetary policy: A quantile connectedness approach. *Econ Lett* 204: 109891. https://doi.org/10.1016/j.econlet.2021.109891

Dabara KA, Tahir IM (2025) Leveraging hybrid ensemble models in stock market prediction: A data-driven approach. *Data Sci Financ Econ* 5: 355–386. https://doi.org/10.3934/DSFE.2025015

Dash SR, Gabauer D, Goel G, Subramaniam S (2024) Investigating dynamic connectedness of global equity markets: the role of investor attention. *Appl Econ* 56: 7222–7243. https://doi.org/10.1080/00036846.2023.2277705

Diebold FX, Yilmaz K (2012) Better to give than to receive: Predictive directional measurement of volatility spillovers. *Int J Forecast* 28: 57–66. https://doi.org/10.1016/j.ijforecast.2011.02.006

Diebold FX, Yilmaz K (2014) On the network topology of variance decompositions: Measuring the connectedness of financial firms. *J Econom* 182: 119–134. https://doi.org/10.1016/j.jeconom.2014.04.012

Fama EF, French KR (1996) Multifactor explanations of asset pricing anomalies. *J Finance* 51: 55–84. https://doi.org/10.1111/j.1540-6261.1996.tb05202.x

Fama EF, French KR (2016) Dissecting anomalies with a five-factor model. *Rev Financ Stud* 29: 69–103. https://doi.org/10.1093/rfs/hhv043

Gabauer D (2021) Dynamic measures of asymmetric and pairwise connectedness within an optimal currency area: Evidence from the ERM I system. *J Multinatl Financ Manag* 60: 100680. https://doi.org/10.1016/j.mulfin.2021.100680

Goetzmann WN, Kim D, Shiller RJ (2022) Crash narratives. *NBER Working Paper* No. 30195. https://doi.org/10.3386/w30195

Goldfarb D, Idnani A (1983) A numerically stable dual method for solving strictly convex quadratic programs. *Math Program* 27: 1–33. https://doi.org/10.1007/BF02591962

Granger CW (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424–438. https://doi.org/10.2307/1912791

Hilary G, Hsu C (2013) Analyst forecast consistency. *J Finance* 68: 271–297. https://doi.org/10.1111/j.1540-6261.2012.01800.x

Hirshleifer D (2020) Presidential address: Social transmission bias in economics and finance. *J Finance* 75: 1779–1831. https://doi.org/10.1111/jofi.12906

Johnson SG, Tuckett D (2022) Narrative expectations in financial forecasting. *J Behav Decis Mak* 35: e2245. https://doi.org/10.1002/bdm.2245

Khalfaoui R, Jabeur SB, Dogan B (2022) The spillover effects and connectedness among green commodities, Bitcoins, and US stock markets: Evidence from the quantile VAR network. *J Environ Manag* 306: 114493. https://doi.org/10.1016/j.jenvman.2022.114493

Li Z, Chen Z, Huang Z (2026) Modelling the data-generating mechanism of China's commodity market by identifying hidden information flow regimes. *Financ Innov* 12: Article 28. https://doi.org/10.1186/s40854-025-00804-w

Long S, Chatziantoniou I, Gabauer D, Lucey B (2024) Do social media sentiments drive cryptocurrency intraday price volatility? New evidence from asymmetric TVP-VAR frequency connectedness measures. *Eur J Finance* 30(: 1470–1489. https://doi.org/10.1080/1351847X.2024.2314085

Merton RC (1973) An intertemporal capital asset pricing model. *Econometrica* 41: 867–887. https://doi.org/10.2307/1913811

Sharif T, Bouteska A, Abedin MZ, Cotturone S (2025) An enquiry into the monetary policy and stock market shocks in the US. *Int Rev Econ Finance* 98: 103925. https://doi.org/10.1016/j.iref.2025.103925

Shiller RJ (2019) *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.

Shiller RJ (2015) *Irrational Exuberance*, 3rd ed. Princeton University Press.

Sims CA (1980) Macroeconomics and reality. *Econometrica* 48: 1–48. https://doi.org/10.2307/1912017




AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)