Media Narratives and Price Informativeness*

Chukwuma Dim† **Francesco Sangiorgi**‡ **Grigory Vilkov**§

This version: January 30, 2023

Abstract

We theoretically and empirically show that stock return exposure to media narratives' attention, measured with standard methods for extracting topic attention from news text, is linked to lower stock price informativeness about future fundamentals. In the model, narrative exposure proxies for media bias-driven return volatility and is inversely related to price informativeness. Empirically, narrative exposure significantly decreases price informativeness and explains over 82% of idiosyncratic variance in the cross-section. Consequently, idiosyncratic variance and variance related to public information decrease stock price informativeness. Moreover, stocks affected by large average narrative shocks demonstrate elevated trading volume.

Keywords: media narratives, price informativeness, undervaluation, idiosyncratic risk, arbitrage, latent demand.

JEL: G11, G12, G13, G17

*We received helpful comments and suggestions from Yigitcan Karabulut, Emanuel Mönch. We also thank seminar participants at the Frankfurt School of Finance and Management.

†Chukwuma Dim is with George Washington University, USA, cdim@gwu.edu.

‡Francesco Sangiorgi is with Frankfurt School of Finance & Management, Germany, F.Sangiorgi@fs.de.

§Grigory Vilkov is with Frankfurt School of Finance & Management, Germany, vilkov@vilkov.net.

1 Introduction

While slow-moving fundamentals play an essential role in asset pricing models as drivers of risks and risk premiums, in everyday lives, investors are constantly exposed to an intense flow of news containing informative, uninformative, and potentially biased signals from media outlets. Building on Robert Shiller's insights on the link between narratives and economic behavior (see, Shiller, 2020), a growing body of research now extracts narratives from the news and analyzes how attention to various narratives is related to different economic quantities. Our main objective is to study how news media affects the information embedded in stock prices in a typical environment, where news, often considered factual, can in fact be biased, leading to disagreement among investors, some of which are not fully rational to undo such biases.[1] Will individual asset prices covary with the intensity of coverage of specific narratives in the news media? Do stock prices with higher covariance with narrative attention aggregate more or less information about future fundamentals? Does exposure to media narratives create excess volatility in stock prices? We address these and related questions, both theoretically and empirically.

The main results and contributions of this study are four-fold. First, in a stylized dynamic trading model with media bias, we show that asset returns are correlated with media attention to narratives, and the informativeness of asset prices diminishes along with higher narrative exposures, which ultimately increases the non-systematic variance of asset returns. Second, we demonstrate empirically that, indeed, individual stocks' price informativeness diminishes with higher absolute exposure to media narratives identified from *The Wall Street Journal (WSJ)*. In fact, the prices of stocks in the top exposure quartile fail to demonstrate a significant connection to future fundamentals. Third, consistent with the model, narrative exposure is empirically the most prominent cross-sectional explanatory variable for idiosyncratic variance, primarily driven by the firm-specific public information component of return variance. Consequently, high levels of idiosyncratic and public-information-related variances also decrease the price informativeness of individual stocks. Fourth, stocks strongly affected by narrative attention shocks experience

[1] Koijen and Yogo (2019) highlight the importance of sentiment and disagreement for explaining latent demand dispersion across investors. A growing body of research highlights news media biases and their implications for financial markets (e.g., Mullainathan and Shleifer, 2005; Baloria and Heese, 2018; Niessner and So, 2018; Goldman, Gupta, and Israelsen, 2021; Goldman, Martel, and Schneemeier, 2022) and investors' over- and under-reaction to media coverage (e.g., Frank and Sanati, 2018). A substantial body of research has also studied departures from full rationality when agents process information and form beliefs (see Barberis, 2018 for a review).

higher turnover, which supports the role of the media narrative exposure as one of the characteristics explaining dispersion in latent demand across assets. We further document that high exposure to narratives does not lead to 'hyped' overpriced stocks. Instead, such firms are undervalued relative to their industry peers, while their elevated arbitrage risk discourages the exploitation of the undervaluation.

Our empirical analysis is guided by a trading model with time-varying public information that addresses the following questions: Why would stock returns co-move with changes in news media's attention to different narratives? How would these narrative exposures relate to price informativeness in the cross-section? To ensure close alignment with the current state of empirical research, the model maps the Latent Dirichlet Allocation (LDA) algorithm—used in our empirical analysis—to the information process faced by investors. An overview of the setup is as follows. A media outlet publishes news articles around several narratives correlated with firms' fundamentals. The amount of attention accorded to a narrative determines the number of articles on that narrative, and the narrative attention evolves randomly over time. Articles are informative but are also biased, and a fraction of investors do not account for this bias. To derive a clear message about price informativeness, we assume investors are risk neutral and hence shut down any impact of narratives on risk premiums.

The model provides the following insights: (i) When attention to a narrative increases, the associated bias receives more weight in the unsophisticated investors' beliefs. Because asset prices reflect these beliefs, stock returns move in the direction of the narrative bias adjusted for cash flow narrative exposures. In this way, the model provides a mechanism for stock return covariances with changes in narrative attention—that is, for stocks' exposure to media narratives. (ii) The bias-related stock price reaction to changes in narrative attention is unrelated to fundamentals and is, therefore, detrimental to price informativeness. (iii) Narrative exposures proxy for this non-fundamental source of return variation and are negatively related to price informativeness in the cross-section. (iv) Narrative exposures proxy for a significant part of non-systematic return variance despite the fact that shocks to narrative attention explain only a modest fraction of return variance.

We directly test these insights from the theoretical framework using a large archive (more than 300,000) of online WSJ news articles. We use the LDA algorithm to extract an optimal number of narratives (33) for the period 1998 to 2021 and then measure the exposure of individual stocks to attention shocks to each of these and some aggregated narratives. Precisely, we regress each stock's daily excess returns within a year on standard factors (market model, and three-, four-, and five-factor models following Fama and French, 1993, 2015; Carhart, 1997) augmented with a narrative's attention shocks, then take the coefficient for the latter as a narrative beta.[2] Following our model, the absolute narrative betas should be negatively associated with price informativeness. Furthermore, they should be a major driver of the non-systematic variance in the cross-section of individual stocks.

To test the link between absolute narrative betas and price informativeness, we adopt a micro-founded stock-level measure of price informativeness based on Bai, Philippon, and Savov (2016), defined as the predicted variation in cash flows using current market prices. Bai, Philippon, and Savov (2016) demonstrate that this measure is also justified as a welfare measure using Q-theory. In addition to its solid theoretical foundation and empirical support, we prefer this measure to the often-used nonsynchronicity measure—defined as $1 - R^2$ from a market model—because nonsynchronicity ambiguously captures both noise and potentially firm-specific information in stock prices. For example, a decrease (increase) in R^2 (nonsynchronicity) can be entirely due to noisy prices without any improvement in price informativeness, and vice versa. Accordingly, Brogaard, Nguyen, Putnins, and Wu (2022) show that despite the recent increases in R^2, which implies less informative prices based on nonsynchronicity, stock prices have instead increasingly reflected more firm-specific information.[3] Moreover, since our theoretical framework predicts the possibility of firms' exposure to media narratives to distort stock prices by introducing noise, nonsynchronicity could lead to misleading results in our framework since an increase in noise can be misinterpreted as an increase in price informativeness.

We employ a two-stage methodology where we first run an annual cross-sectional regression of future firm fundamentals on current market value and its interaction with absolute narrative

[2]Notably, adding narrative attention shocks barely changes the explanatory power (less than 0.1% adjusted R^2, on average) of the selected models; however, our main object of interest is narrative betas and their connection to non-systematic variance and price informativeness.

[3]They also show that nonsynchronicity yields implausible relationships between price informativeness and several firm characteristics.

exposure (and controls) and then test whether the average coefficients are different from zero in the second stage. We observe a strong decrease in price informativeness for stocks with higher absolute narrative betas, especially when attention to narratives is high. For stocks in the top quartile of absolute narrative exposure each year, current stock prices are *not* significantly informative about future fundamentals. Analyzing the second major prediction of the model, we find that, indeed, absolute narrative betas *alone* explain more than 80% of the cross-sectional variation in stocks' idiosyncratic risk. Decomposing non-systematic variance into private and public firm-specific information and noise using the approach of Brogaard, Nguyen, Putnins, and Wu (2022) (BNPW), we find that narrative exposure is most closely related to the public information component, with noise and private information following closely behind. Testing the connection between the different components of stock variance and price informativeness, we find a consistent picture: high idiosyncratic and public-information-related variances are the strongest 'killers' of the information contained in stock prices.

Media narratives distort the stock prices of highly exposed firms, but it is still not clear whether the mispricing is in the form of overvaluation, undervaluation, or a non-persistent directional price distortion. We, therefore, test whether high narrative exposure is equivalent to the 'hype' surrounding financial markets, which in turn produces overpriced stocks. Using the misvaluation measure of Rhodes-Kropf, Robinson, and Viswanathan (2005), we document that firms with high absolute narrative betas are, on average, undervalued relative to industry peers. A one-standard-deviation increase in absolute narrative exposure is equivalent to about 12% undervaluation and an increase in the probability of undervaluation by 7.3%. The result is consistent with the observation that the news media exhibits negativity bias on average (e.g., Liu and Matthies, 2022; Sacerdote, Sehgal, and Cook, 2020; Niessner and So, 2018), focusing more on negative news that tends to attract more attention. Such negative slant is also pervasive in our sample, as the average WSJ news article has 165% more fraction of negative relative to positive words from the Loughran and McDonald (2011) dictionary. Consequently, firms hugely exposed to media narratives tend to experience more depressing media-induced price shocks, resulting in relative undervaluation.

We analyze why this undervaluation is not exploited by sophisticated investors. Using the arbitrage risk measure of Wurgler and Zhuravskaya (2002), we observe a relatively high arbitrage

risk for stocks highly exposed to media narratives due to their inherently high non-systematic variance. This deters arbitrageurs from exploiting the undervaluation.

Literature review. Our study is related to several developing and mature strands of literature, and we establish new and revealing connections among some research directions.

To quantify price informativeness empirically, we rely on the cross-sectional measure by Bai, Philippon, and Savov (2016), and we give a structural interpretation of this measure in our model. Recent studies have used this measure in various settings: Kacperczyk, Sundaresan, and Wang (2020) use it to analyze the effect of foreign institutional investments on price informativeness; Chen, Kelly, and Wu (2020) use it to measure information spillovers between buy-side and sell-side research, and Cao, Goyal, Ke, and Zhan (2022) use it to study the effect of options trading on stock price informativeness. Farboodi, Matray, Veldkamp, and Venkateswaran (2021) introduce a similar measure to quantify the effects of data abundance on the information content of prices. We contribute to this literature by relating price informativeness to return narrative exposures both theoretically and empirically.

This study also relates to the recent applications of news media text in economics and finance research. As in this study, Bybee, Kelly, Manela, and Xiu (2021) use LDA to quantify the structure of economic news and show that news predicts certain macro variables. Bybee, Kelly, and Su (2022) use LDA to extract latent risk factors from news text, and Hanley and Hoberg (2019) use the algorithm to study emerging risks in the financial sector. Other studies apply supervised or semi-supervised algorithms to infer certain economic quantities from news text. For instance, Baker, Bloom, and Davis (2016) develop an index of policy uncertainty, Manela and Moreira (2017) develop a news-based volatility index, Engle, Giglio, Kelly, Lee, and Stroebel (2020) construct a news-based climate risk measure, Liu and Matthies (2022) quantify investor concerns about economic growth, and Dim, Koerner, Wolski, and Zwart (2022) produce a news-implied sovereign default risk index. All of these studies focus on the role of the media as a valuable source of unstructured data relevant for tracking various economic quantities.

In contrast, we build on research highlighting news media biases (e.g., Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006; Reuter and Zitzewitz, 2006; Baloria and Heese, 2018; Goldman, Gupta, and Israelsen, 2021), as well as biases in investors' belief formation,

such as over- and under-reaction (e.g., De Bondt and Thaler, 1985; Shleifer and Summers, 1990; Barberis, Shleifer, and Vishny, 1998; Frazzini, 2006; Bordalo, Gennaioli, Ma, and Shleifer, 2020), to document three main theoretically motivated results: (i) time-varying attention to specific narratives in the media affects firms heterogeneously; (ii) due to media and investor biases, firms that are disproportionately exposed to media narrative attention shocks have less informative stock prices; and (iii) exposure to high-frequency media attention shocks is a predominant driver of excess volatility in stock returns. Therefore, although the news media can yield useful signals, it distorts some firms' asset prices. We establish the attention to media narratives as a theoretically sound and empirically important channel of disagreement in financial markets.

We also contribute to the literature on news media's effects on the stock market. Tetlock (2007) shows that media pessimism depresses the aggregate market return, consistent with models of noise and liquidity traders. Garcia (2013) shows that this destabilizing impact of the media is magnified in bad times. Calomiris and Mamaysky (2019) show that news predicts aggregate returns in a manner that suggests that news flow mainly captures non-priced risks. Tetlock, Saar-Tsechansky, and Macskassy (2008) show that sentiment in firm-specific news predicts returns. Hillert, Jacobs, and Müller (2014) show that firms particularly covered by the media exhibit stronger momentum, consistent with investor overreaction, and Frank and Sanati (2018) document the stock market's overreaction and underreaction to news with particular tone reported in the media. Dougal, Engelberg, Garcia, and Parsons (2012) abstract away, as we do, from sentiment when studying the impact of journalist-specific biases on market returns. In contrast to these papers, our focus and approach differ markedly. While they primarily focus on the impact of the news media, mainly sentiment, on stock returns, we analyze the biases reflected in media narratives and establish theoretically and empirically the direct destabilizing impact of media narrative exposure on the information content of individual stock prices.

Our results provide important insights for the literature on demand-based asset pricing and the determinants of cross-sectional variance. Recent work (Koijen and Yogo, 2019, p.1488) estimates that changes in latent demand are the most important demand-side determinant of the cross-sectional variance of stock returns, explaining 81 percent of the cross-sectional variance. Gabaix and Koijen (2021) build on De Long, Shleifer, Summers, and Waldmann (1990)'s model that features noisy beliefs driving demand fluctuations. They identify changes in beliefs as one

of the potential determinants of high-frequency flows. We find that stocks' exposure to narrative shocks is one such proxy for changes in beliefs that result in trading, in turn explaining over 82% of the total and idiosyncratic variances, respectively, in the cross-section. Consistent with the proposed theoretical mechanism, we establish narrative exposure as the major characteristic explaining non-systematic variance in the cross-section of stocks, complementing the residual household income risk channel of Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016).

The rest of the study is organized as follows. Section 2 develops a model that motivates the subsequent analysis and sets a number of testable predictions. Section 3 describes the data sources, construction of stock and firm characteristics, the extraction of narratives from news text, and the computation of media narrative exposures. It also contains the summary statistics of the main variables used for analysis in subsequent sections. Section 4 tests the model key predictions. It analyzes how price informativeness is affected by media narrative exposures, then looks at how the different stock return variance components (i.e., proxy for information channels affecting stock returns) relate to narrative exposure and, in turn, price informativeness. It then proceeds to document a link between narrative shocks and turnover. Section 5 completes the analysis with documenting stylized facts that are not directly in the scope of our model, but can be anticipated and potentially derived under additional assumptions. This section established the link between narrative shocks and trading volume, and examines valuations of firms exposed to narratives. Section 6 concludes the analysis, with a short summary of the findings.

2 A Model of Media Narratives and Price Informativeness

2.1 Model Setup

Agents and Assets. Time is discrete, and there are $T+1$ periods. There are N risky assets that are claims to dividends paid at date $T+1$. For asset $n = 1, ..., N$, the final dividend equals

$$D_n = \bar{D}_n + b_n' F + \varepsilon_n, \tag{1}$$

where $\bar{D}_n$ is a constant, F is a $(K \times 1)$ vector of common factors, b_n is a $(K \times 1)$ vector of factor loadings, and ε_n is a residual term independent of F and all other random variables. We assume

that

$$F = \sum_{t=1}^{T} f_t, \tag{2}$$

where f_t is a $(K \times 1)$ vector of factor innovations; $\{f_\tau\}_{\tau=1}^{T}$ are i.i.d. normal with mean vector $\bar{f}$ and variance matrix Σ_f. Without loss of generality, we assume that risky assets are in zero net supply and set $\bar{f}$ equal to zero.

Risky assets are traded at each $t = 1, \ldots, T$ among a continuum of investors. Given our focus on price informativeness, we assume investors are risk neutral to shut down any impact of narratives on risk premiums. Each period, a new cohort of investors is born. Investors live for two periods. In the first period, they trade the N risky securities and a riskless asset with exogenous return normalized to zero; in the second period, investors close all positions, consume, and exit the economy.

We denote $x_{i,t}$ the $(N \times 1)$ vector of risky asset holdings of investor i at time t. Investors have zero wealth when they enter the economy and are subject to holding costs $\frac{1}{2} x'_{i,t} C_i x_{i,t}$, where $C_i = \text{diag}(c_{i,1}, ..., c_{i,N})$ is a diagonal matrix, and each $c_{i,n}$ is a parameter capturing investor i's asset-specific holding costs and preferences.

News and bias. Each period, investors learn about factor innovations from M news articles published in a media outlet. Each news article centers around one of L news topics, or "narratives." We denote z_t the $(L \times 1)$ vector of narratives in period t. Narratives are related to factor innovations as follows:

$$z_t = A f_t + \eta_t, \tag{3}$$

where A is a $(L \times K)$ matrix of constants, and η_t is an $(L \times 1)$ random vector independent of f_t and of all other random variables; $\{\eta_\tau\}_{\tau=1}^{T}$ are i.i.d. normal with mean zero and variance matrix Σ_η. Eq. (3) captures two ideas. First, factors influence each narrative differently through the corresponding row of the matrix A. Second, each narrative has a component which is irrelevant to asset payoffs.

We now describe how news articles are related to narratives in each period. First, the relative attention to the L narratives is determined. This is an $(L \times 1)$ probability vector $\theta_t = (\theta_{1,t}, ..., \theta_{L,t})'$ independently drawn from the same distribution each period. Then, each news article $m = 1, ..., M$ independently selects one of the L narratives at random according to the probability vector θ_t. When article m selects narrative l at time t, its information content is equivalent to the signal

$$s_{m,t} = z_{l,t} + \pi_{l,t} + \zeta_{m,t},$$

where $z_{l,t}$ is the l-th entry of z_t in Eq. (3), $\pi_{l,t}$ is a narrative-specific bias with mean π_l and variance $\pi_l^2\sigma^2$, and the error term $\zeta_{m,t}$ is normally distributed with mean zero and variance M/ω, where ω is a positive constant. $\{\pi_{l,t}\}_{\tau=1}^{T}$ and $\{\zeta_{m,t}\}_{\tau=1}^{T}$ are i.i.d. and independent across narratives and articles.

Thus, the media outlet conveys information to investors that is valuable but biased, and the average values $\pi_1, ..., \pi_L$ capture the persistent components of the media outlet's narrative-specific biases.

For tractability, we consider the limit where $M \uparrow \infty$ and show in Appendix A that the information published by the media outlet is equivalent to the L signals

$$S_{l,t} = z_{l,t} + \pi_{l,t} + \hat{\zeta}_{l,t}; \qquad \text{for } l = 1, ..., L, \tag{4}$$

where $\hat{\zeta}_{l,t} \sim N\left(0, (\omega\theta_{l,t})^{-1}\right)$. Thus, letting $\Theta_t = \text{diag}(\theta_{1,t}\omega, ..., \theta_{L,t}\omega)$, the $(L \times 1)$ vector of signals $S_t = (S_{1,t}, ..., S_{L,t})$ has precision matrix

$$Var\left(S_t | \, z_t, \pi_t\right)^{-1} = \Theta_t, \tag{5}$$

where π_t is the $(L \times 1)$ vector of a media biases $\pi_t = (\pi_{1,t}, .., \pi_{L,t})'$. Eq. (5) maps the relative narrative attention θ_t into the precision of investor information. When relative attention to a certain narrative increases, that is, when the corresponding element of θ_t goes up, investors learn more about that narrative from the media outlet.

In our empirical analysis, we use the LDA algorithm to extract narratives and the time series $\{\theta_t\}_{\tau=1}^{T}$ of relative attention to those narratives from a large archive of online WSJ news articles (see Section 3.3).

Investor sophistication. Each investor belongs to one of two classes indexed by R and U. Investors in class R are fully rational and are aware of the media bias in each period, whereas investors in class U are unsophisticated and ignore the media bias. The relative proportion of R and U investors is constant across cohorts. We assume that the structure of the economy is common knowledge, and that U investors have dogmatic beliefs.[4] Since all information is public, investor beliefs are the same for all investors in the same class. Thus, for any random variable y, we denote $E_{i,t}(y) = E_{R,t}(y)$ for all $i \in R$ and $E_{i,t}(y) = E_{U,t}(y)$ for all $i \in U$.

2.2 Analysis

Prices and returns. It is convenient to express the dividend Eq. (1) in terms of narratives:

$$D_n = \bar{D}_n + \beta_n' \sum_{t=1}^{T} z_t + \varphi_n, \tag{6}$$

where β_n is the $(L \times 1)$ vector of asset-n dividend sensitivities to the L narratives, and φ_n is a residual term that is uncorrelated with F and with all z_t's[5]. Our assumptions regarding investor sophistication imply the following expectations:

$$E_{R,t}(D_n) = \bar{D}_n + \beta_n' \sum_{\tau=1}^{t} \Phi_\tau (S_\tau - \pi_\tau); \qquad E_{U,t}(D_n) = E_{R,t}(D_n) + \beta_n' \sum_{\tau=1}^{t} \Phi_\tau \pi_\tau, \tag{7}$$

$$\text{where} \quad \Phi_\tau = \left(A\Sigma_f A' + \Sigma_\eta\right)\left(A\Sigma_f A' + \Sigma_\eta + \Theta_\tau^{-1}\right)^{-1}. \tag{8}$$

The $(L \times L)$ matrix Φ_τ depends on the relative attention vector θ_τ via the precision matrix Θ_τ and determines how strongly investor beliefs react to time-τ news. Thus, Φ_τ also determines how strongly the media biases π_τ affect U investor beliefs in Eq. (7).

Proposition 1. *(Asset prices and returns)*

[4] Therefore, R investors know that U investors have biased beliefs, whereas U investors believe R investors have biased beliefs: $E_i(E_j(S_{l,t})) = 0$ and $E_j(E_i(S_{l,t})) = -\pi_l$ for all $i \in R$, $j \in U$, and $l = 1, ..., L$.

[5] See Eqs. (A3)-(A4) in Appendix A.

(i) *The asset price of security n at time t equals*

$$P_{n,t} = (1 - \gamma_n) E_{R,t}(D_n) + \gamma_n E_{U,t}(D_n), \tag{9}$$

where $\gamma_n = \frac{\psi_{U,n}}{\psi_{R,n}+\psi_{U,n}}$ *and* $\psi_{a,n} = \int_a c_{i,n}^{-1} di$, *for* $a = R, U$.

(ii) *The return* $r_{n,t} := P_{n,t} - P_{n,t-1}$ *equals*

$$r_{n,t} = E_{R,t}(\beta_n' z_t) + \Pi_{n,t} \tag{10}$$

where $\Pi_{n,t} = \gamma_n \sum_{j=1}^{L} \pi_{j,t} \beta_n' \phi_{j,t}$ *and* $\phi_{j,t}$ *is the* j*-th column of the matrix* Φ_t.

(iii) *Asset-n's exposure to narrative l's relative attention,* $\beta(n,l) := \frac{Cov(r_{n,t},\theta_{l,t})}{Var(\theta_{l,t})}$, *equals*

$$\beta(n,l) = \frac{\gamma_n}{Var(\theta_{l,t})} \sum_{j=1}^{L} \pi_j \beta_n' cov(\phi_{j,t}, \theta_{l,t}). \tag{11}$$

Proof. See Appendix A. ■

Proposition (1)-(i) shows that an asset price is a weighted average of investor beliefs about the asset payoff, and that the weight of an investor type is dependent upon its trading aggressiveness relative to the other type. This trading aggressiveness is measured by $\psi_{a,n}$, the mass-weighted average of the reciprocal of the holding costs for investors a in asset n.

Proposition (1)-(ii) reveals that an asset return has two parts. The first part, $E_{R,t}(\beta_n' z_t)$, is the rational belief response to time-t news. The second part, $\Pi_{n,t}$, is bias-driven and is due to U investors. The expression for $\Pi_{n,t}$ following Eq. (10) shows that the influence of narrative l's bias on the return on asset n increases in U investors' price impact on asset n, γ^n, and increases in investor beliefs' sensitivity to news about narrative l, $\phi_{l,t}$, weighted by the asset n's cash flow sensitivities to the L narratives, β_n.

Proposition (1)-(iii) derives the assets' narrative exposures that are central to our empirical analysis. The intuition is as follows. When a narrative receives greater attention, U investors' beliefs load more strongly on that narrative's bias (Eqs. (7)-(8)). U investors have price impact,

so the stock return moves in the direction of this narrative's bias, adjusted for cash flow narrative exposures (Eq. (10)). This mechanism leads to the covariance between narrative attention and stock return in Eq. (11).

Price informativeness. We define price informativeness for asset n as

$$I_n = t^{-1} \frac{Cov\left(D_n, P_{n,t}\right)^2}{Var\left(P_{n,t}\right)}. \tag{12}$$

This definition is standard in market microstructure (e.g., Kacperczyk, Nosal, and Sundaresan, 2022) and is consistent with the approach in Bai, Philippon, and Savov (2016), which forms the basis of our empirical analysis. In our model, Eq. (12) measures the reduction in the posterior dividend uncertainty of an investor who learns from the price using a linear model.[6,7]

Proposition 2. *(Narrative exposures and price informativeness)*

(i) *Return variance and price informativeness equal*

$$Var\left(r_{n,t}\right) = SysVar_n + IdVar_n; \qquad I_n = \frac{SysVar_n^2}{SysVar_n + IdVar_n}, \tag{14}$$

where $SysVar_n = Var\left(E_{R,t}\left(\beta_n' z_t\right)\right)$ *is given in Eq. (A13) in Appendix A and*

$$IdVar_n = Var(\Pi_{n,t}) = \gamma_n^2 \sum_i^L \sum_j^L \pi_i \pi_j Cov\left(\beta_n' \phi_{i,\tau}, \beta_n' \phi_{j,\tau}\right) + \gamma_n^2 \sum_l^L \pi_l^2 \sigma^2 E\left(\beta_n' \phi_{l,t}\right)^2. \tag{15}$$

(ii) *The narrative exposures proxy for* $IdVar_n$ *as follows:*

$$\beta\left(n,l\right)^2 = IdVar_n \frac{Corr\left(\Pi_{n,t}, \theta_{l,t}\right)^2}{Var\left(\theta_{l,t}\right)} \tag{16}$$

[6] Consider the linear model $D_n = a_n + b_n P_{n,t} + e_n$. The variance of D_n conditional on $P_{n,t}$ is the variance of the forecast error e_n. Therefore,

$$Var\left(D_n\right) - Var\left(e_n\right) = b_n^2 Var\left(P_{n,t}\right) = tI_n, \tag{13}$$

where the second equality follows from $b_n = \frac{Cov\left(D_n, P_{n,t}\right)}{Var\left(P_{n,t}\right)}$ and the definition of I_n in Eq. (12). In our empirical analysis, we follow Bai, Philippon, and Savov (2016) and estimate $b_n^2 Var\left(P_{n,t}\right)$ from the cross-section.

[7] The "t^{-1}" term in the definition adjusts for the non-stationary nature of our model.

and

$$\sum_{l}^{L} \frac{\beta\left(n,l\right)^{2}}{L} = IdVar_{n} \sum_{l}^{L} \frac{Corr\left(\Pi_{n,t},\theta_{l,t}\right)^{2}}{Var\left(\theta_{l,t}\right)L}. \tag{17}$$

Proof. See Appendix A. ■

Proposition 2-(i) predicts an inverse relationship between asset price informativeness and $IdVar_n$, the media-bias-driven component of return volatility. An empirical test of this prediction requires the identification of $IdVar_n$ separately from other sources of non-fundamental return volatility. However, Proposition 2-(ii) suggests that a stock's narrative exposures can proxy for $IdVar_n$. This is intuitive because both $IdVar_n$ in Eq. (15) and the $\beta\left(n,l\right)$'s in Eq. (11) are driven by media biases weighted by an asset's cash flow narrative exposures. Thus, they carry overlapping information. For example, in the case of independent narratives where asset n loads only on narrative l, Eq. (16) simplifies to $\beta\left(n,l\right)^{2} = IdVar_{n}\kappa_{l}$, where the constant of proportionality κ_l depends on the distribution of $\theta_{l,t}$ and is independent of $\beta_{n,l}, \gamma_n$, and π_l. Therefore, for stocks that load mostly on one narrative, narrative exposures explain most of the cross-sectional variation in $IdVar_n$. For the general case, our empirical analysis in Section 4.2 demonstrates a strong positive cross-sectional relationship between narrative exposures and idiosyncratic variance.

Summary of testable implications. From the outlined model, we deduce several *direct testable implications*, which, if invalidated, falsify the theory: (i) a stock's narrative exposure is negatively related to its price informativeness; (ii) an increase in media's attention to a narrative reduces the price informativeness of exposed stocks; (iii) a stock's narrative exposure is positively related to its idiosyncratic (non-systematic) variance, and is the main determinant of the cross-sectional dispersion in non-systematic variance. From the definition, we also expect a negative link between price informativeness and non-systematic variance. In deriving the model, we made two important assumptions based on existing empirical and theoretical results: the delivery of biased narratives by media outlets and the existence of unsophisticated agents who do not account for the bias in processing information. These features yield an additional implication, namely, the share of unsophisticated agents and the level of bias both have a negative effect on

price informativeness. We defer the tests of this implication to future research, as it requires a thorough analysis of multiple media outlets, quantification of the bias distribution and agents' sophistication.

3 Data and Variable Measurement

This section describes the main data sets and variables used in the study: Section 3.1 covers general stock variables, Section 3.2 defines the sources of news text, Section 3.3 describes the procedures for extracting narratives and measuring narrative exposure, and Section 3.4 provides summary statistics and a preliminary analysis. Our sample period spans from 1998-2021, because our news media data begins in 1998. Table B1 describes all of the variables used in this study.

3.1 Stock and Firm Characteristics

Our sample consists of US common stocks (share codes 10 and 11) listed on the NYSE, AMEX, and NASDAQ stock exchanges. We retrieve daily stock returns, prices, market capitalization, and volume from the daily data files of the Center for Research in Security Prices (CRSP). We obtain firm fundamentals from the Compustat North America Annual File. We exclude firms in the financial sector, firms with year-end market capitalization below $1 million, and filter out stock years with less than 20 observations and stock years in which a stock changed its primary exchange. We use daily factor returns from Kenneth French's Data Library with stock returns to compute factor exposures, idiosyncratic variance, and other characteristics.

We decompose stock return variances into components representing particular information channels using two approaches. First, each year we estimate from daily returns standard linear factor models (market model and three-, four-, and five-factor models by Fama and French (1993), Carhart (1997), and Fama and French (2015)) to decompose excess returns into systematic and idiosyncratic components and compute their respective variances. Second, we decompose stock return variance into components stemming from market information, private information, public information, and noise using the vector autoregression framework of BNPW. We perform the decomposition separately for each stock yearly using daily returns. The details of the procedures for both approaches are provided in the Online Appendix OA.2.

3.2 News Media Text

Public information affecting agents' trading decisions flows primarily through the news media. For our purposes, one requires a news media outlet that is not only widely read by financial market participants but also has a relatively long history and is easily retrievable. We rely on the historical news archive of the WSJ for a large corpus of historical news text and use it to quantify the evolution of different media narratives and firms' exposure to those narratives.

We retrieve the WSJ's historical news archive through its website, spanning from 1998, the first year of availability, to 2021. We apply filters to remove sections of the Journal that are highly unlikely to be relevant to financial markets and that stand the chance of introducing unnecessary noise into our text corpora. These sections include Entertainment, Leisure & Arts, Sports, Lifestyle & Culture, and the like—in total, 37 categories. We further process the news article texts to reduce dimensionality and noise using the `SpaCy` text processing pipeline. We lemmatize words, convert text to lowercase, and exclude stopwords and entities such as persons, geopolitical areas, locations, and nationalities. We also exclude articles shorter than 20 words and end up with 348,649 news articles—averaging 1,206 articles per month—for further analysis.[8]

3.3 Extracting Media Narratives and Computing Narrative Exposures

Procedures for Extracting Media Narratives. Daily news text publications cover various issues that grab agents' attention and potentially shape various economic decisions, including stock trading. Such an information-rich environment has apparent benefits but poses significant challenges related to the extraction of the parsimonious set of narratives behind the news. However, as Shiller (2017) advocates, one can apply recent advances in textual analysis and natural language processing to extract the underlying topical narratives in news text.

We adopt the unsupervised machine learning Latent Dirichlet Allocation (LDA) algorithm of Blei, Ng, and Jordan (2003), which has been successfully used in settings similar to ours (e.g.,

[8] Bybee, Kelly, Manela, and Xiu (2021) and Bybee, Kelly, and Su (2022) also use the WSJ text corpus but have sample periods, starting from 1984, and a different number of news articles, roughly 764,000. The differences arise primarily because the authors obtained their text corpus directly from the Dow Jones Historical News Archive. In contrast, we only have access to digitally accessible online data.

Bybee, Kelly, Manela, and Xiu, 2021; Bybee, Kelly, and Su, 2022; Hanley and Hoberg, 2019). The implementation details are presented in Appendix OA.1.

We find a total of 33 narratives, which we manually label based on the (top-100) uni- and bigrams with the largest rescaled term weights.[9] We aggregate the across-article narrative distribution daily to obtain the level of attention to each narrative on a given day as follows:

$$\theta_{l,\tau} = \frac{\frac{1}{M}\sum_{m=1}^{M}\theta_{m,l,\tau}}{D_\tau}, \tag{18}$$

where $\theta_{l,\tau}$ captures the level of attention to narrative l on day τ, $\theta_{m,l,\tau}$ denotes the level of attention to narrative l in article m on day τ, and $D_\tau = \sum_{l=1}^{L}\frac{\sum_{m=1}^{M}\theta_{m,l,\tau}}{M}$ is a normalization that ensures $\theta_{l,\tau}$ sums to one, so that attention allocation each day is a probability distribution. We also group most of the 33 narratives into a smaller set of 12 based on the similarity of their top terms to broader themes by summing $\theta_{l,\tau}$ across narratives for each sub-group on each day.

Quantifying Exposures to Media Narratives. We quantify firms' exposure to narratives by the co-movement between stock returns and individual narrative attention shocks $\tilde{\theta}_{l,\tau}$, measured (similar to Bybee, Kelly, and Su, 2022) on day τ as the difference between day τ's attention level and the average attention level over the past five days ending on $\tau-1$, i.e., $\tilde{\theta}_{l,\tau} = \theta_{l,\tau} - \frac{1}{5}\sum_{i=1}^{6}\theta_{l,\tau-i}$. We then estimate an augmented factor model for each firm n using daily stock returns in year t:

$$r_{n,\tau} = \alpha + \beta_{n,t}^{\top}F_\tau + \beta_{n,t}^{narr}\tilde{\theta}_{l,\tau} + \varepsilon_{n,\tau}, \tag{19}$$

where $r_{n,\tau}$ is stock n's excess return, and F_τ is the vector of factor realizations (we use the four-factor Carhart (1997) model as the main specification) on day τ in year t.

Our main object of interest in model (19) is the absolute value $|\beta_{n,t}^{narr}|$, which captures the magnitude of stock n's return co-movement with narrative l attention shocks. Stocks with high $|\beta_{n,t}^{narr}|$ are affected by trading decisions that move prices when narrative l witnesses attention shocks. Those trading decisions may be driven by the public information inherent in the atten-

[9] We use the TF-IDF (Term Frequency–Inverse Dense Frequency) weighting, i.e., scale the narrative-term weights such that terms that occur very frequently in a given narrative but less so across all other narratives have high weights for that narrative.

tion shock or may be due to other information sources that are coincidentally manifested in the narrative attention shock.

3.4 Summary Statistics and Preliminary Analysis

	Mean	Std	10%	25%	50%	75%	90%	Obs.
Panel A: Narrative exposure.								
Average $\lvert\beta_{n,t}^{narr}\rvert$	0.267	0.188	0.095	0.135	0.212	0.340	0.522	81,952
Panel B: Variance decomposition.								
$IdVar_{n,t}\times10^3$	1.943	2.703	0.184	0.371	0.898	2.294	4.880	81,952
$SysVar_{n,t}\times10^3$	0.218	0.334	-0.000	0.033	0.105	0.257	0.560	81,952
$MktInfo_{n,t}\times10^3$	0.170	0.270	0.004	0.020	0.068	0.193	0.449	57,974
$PrivateInfo_{n,t}\times10^3$	0.455	0.601	0.038	0.088	0.226	0.567	1.143	57,974
$PublicInfo_{n,t}\times10^3$	0.737	1.038	0.070	0.137	0.333	0.880	1.863	57,974
$Noise_{n,t}\times10^3$	0.852	1.498	0.042	0.104	0.293	0.855	2.180	57,974
Panel C: Factor model betas.								
$Market\ Beta_{n,t}$	0.858	0.528	0.113	0.503	0.889	1.211	1.538	81,952
$Size\ (SMB)\ Beta_{n,t}$	0.706	0.720	-0.206	0.162	0.645	1.179	1.743	81,952
$Value\ (HML)\ Beta_{n,t}$	0.138	0.804	-0.900	-0.326	0.134	0.615	1.116	81,952
$Mom\ (WML)\ Beta_{n,t}$	-0.105	0.578	-0.870	-0.421	-0.074	0.243	0.596	81,952
Panel D: Fundamentals and stock characteristics.								
$\ln(Assets)_{n,t}$	5.756	2.010	3.036	4.186	5.665	7.258	8.582	81,952
$EBIT_{n,t}/Assets_{n,t}$	-0.019	0.218	-0.357	-0.055	0.053	0.107	0.168	81,952
$Debt_{n,t}/Assets_{n,t}$	0.209	0.199	0.000	0.011	0.170	0.347	0.516	81,952
$Cash_{n,t}/Assets_{n,t}$	0.227	0.244	0.010	0.034	0.126	0.349	0.654	81,952
$PP\&E_{n,t}/Assets_{n,t}$	0.237	0.219	0.027	0.064	0.158	0.349	0.623	81,952
$Sales_{n,t}/Assets_{n,t}$	0.971	0.686	0.157	0.447	0.846	1.361	2.001	81,952
$Capex_{n,t}/Assets_{n,t}$	0.047	0.046	0.006	0.015	0.032	0.063	0.113	81,952
$R\&D_{n,t}/Assets_{n,t}$	0.062	0.103	0.000	0.000	0.005	0.082	0.218	81,952
$Turnover_{n,t}$	8.622	9.138	1.445	3.007	6.126	11.184	18.891	81,952
$Illiquidity_{n,t}$	0.319	0.861	0.001	0.002	0.013	0.112	0.951	81,883
$Lottery_{n,t}$	0.277	0.447	0.000	0.000	0.000	1.000	1.000	76,913
Panel E: Misvaluation and arbitrage risk.								
$MISVAL_{n,t}$	0.000	0.576	-0.772	-0.389	-0.018	0.365	0.819	71,891
$ARBRISK_{n,t}\times10^3$	1.767	2.212	0.196	0.389	0.912	2.210	4.422	75,836
Panel F: Institutional variables.								
$DOB_{n,t}$	0.007	0.014	0.001	0.001	0.003	0.007	0.018	23,689
$Inst.\ Ownership_{n,t},\ \%$	0.506	0.320	0.050	0.202	0.536	0.791	0.920	66,887

Table 1: Summary Statistics.
The table shows the summary statistics for selected variables computed from the firm-year panel data. *Average* $\lvert\beta_{n,t}^{narr}\rvert$ is computed as the average absolute exposure for all 33 identified narratives. Each year, all continuous variables are winsorized at 5% and 95% levels.

Table 1 shows the summary statistics for most of the variables used in the analysis; moreover, Tables OB1 and OB2 in the Online Appendix show correlations among variables of interest and summary statistics for individual narrative exposures.

While we use the levels of the variance components for our analysis, we examine their proportions to determine whether they are comparable to those of BNPW. In factor-based models, the share of the average systematic variance in total variance ranges from 8.5% for the one-factor model to 11.5% for the five-factor model. Clearly, the residual idiosyncratic variance share is, on average, very large. The numbers for the BNPW decomposition are roughly comparable to the original study, even though we have a shorter (and later) sample period (1998 to 2021 compared to 1960 to 2015 in BNPW). We find that market-wide information accounts for 7.4% of the return variance, private information accounts for 20.2% of the variance, public information accounts for 32.3%, and the remaining 40.1% is noise. The respective numbers from an earlier sample in BNPW are 8%, 24%, 37%, and 31%, respectively. Consistent with BNPW, we find a decreasing trend in noise variance for most of the sample period, and an increasing trend for firm-specific information. However, a sharp increase in the noise component and an equivalent drop in the firm-specific (mostly public) variance in 2020-2021 lead to a slight discrepancy in proportions.

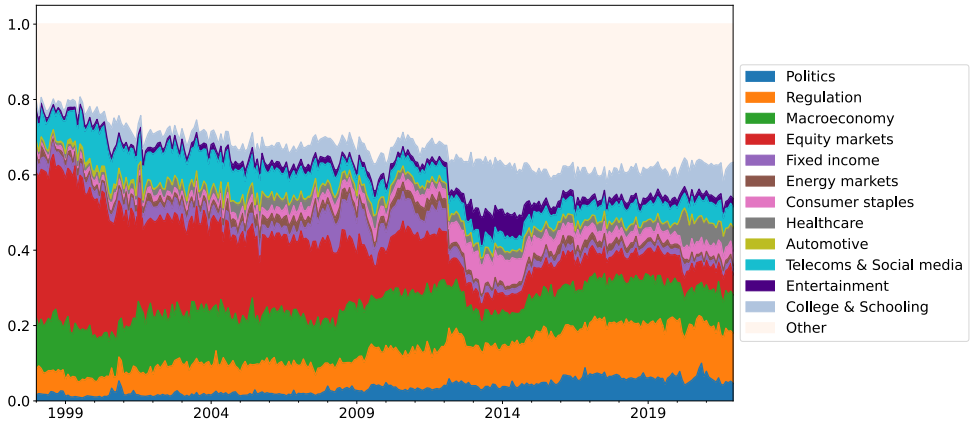

Figure 1: Evolution of Narrative Attention. The figure shows the evolution of attention, from Eq. (18), dedicated to the identified narratives over time after grouping them into 12 themes.

Figure 1 depicts the evolution of the attention level devoted to the identified topical narratives, following Eq. (18), grouped into 12 themes based on manual classification of the individual narratives using their top-100 representative uni- and bigrams. Figure OB1 shows the evolution of data in terms of number of articles and word in these articles, and the model convergence process, and Table OB3 in the Online Appendix lists the top terms and narrative groups. There is substantial variation in the level of attention devoted to each narrative in the WSJ, reflecting the concept that the news media tends to focus on different narratives at different times, due to

changing economic and political conditions, and the changing interests and sentiments of market participants. For instance, the "Equity markets" narrative accounted for a sizeable chunk of the WSJ's attention allocation in the early sample period, but declined over time, while attention to "Regulation" and "Political" narratives grew. Overall, the evident changes in attention allocation to different narratives could impact agents' perspectives regarding the prospects of individual assets, resulting in trading decisions that may or may not distort prices.

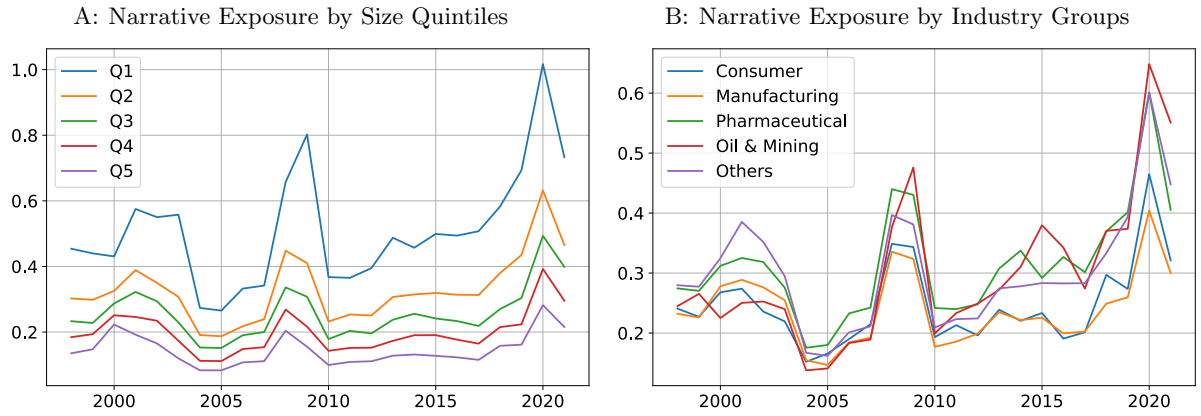

Figure 2: Media Narrative Exposures by Size and Industry. The figure shows the evolution of narrative exposures averaged across all narratives, i.e., *Average* $|\beta_{n,t}^{narr}|$, and then by size quintiles (Panel A) and industry groups (Panel B). In Panel B, the Fama-French 17 industries are collapsed into five major groups to facilitate exposition. The *Consumer* group comprises the Food, Clothing, and Consumer Durables industries; the *Manufacturing* group comprises the Construction, Steel, Fabricated Products, Machinery, and Utilities industries; the *Pharmaceutical* group comprises the Chemicals and Consumer Drugs industries; the *Oil & Mining* group comprises the Mines, Oil, and Steel industries; and the *Others* group comprises the remaining industries.

Finally, Figure 2 depicts the evolution of media narrative exposures averaged across all identified narratives, i.e., *Average* $|\beta_{n,t}^{narr}|$, and then averaged each year by size quintiles and major industry groups. Panel A reveals two striking and persistent patterns: (i) Exposure to media narratives decreases monotonically across size quintiles, which means that smaller firms' stock prices are generally more exposed to media narrative attention shocks. (ii) Exposure to media narratives spikes for firms across all size groups during major stock market downturns, but more so, again, for smaller firms. The first pattern serves as an initial piece of evidence consistent with our theoretical framework. We expect media biases or decisions of agents with biased interpretations of news media coverage to have a more profound impact, for instance, through trading, on the stock prices of smaller firms, leading to the observed higher exposure to narrative attention shocks for such firms. This is because smaller firms are more likely to be traded by investor groups with a higher tendency to exhibit behavioral biases (e.g., Barber and

Odean, 2000), and, at the same time, it is harder for rational agents to exploit such biases due to limits to arbitrage.

Similarly, the spike in narrative exposure across firms in bad times is consistent with existing evidence that news media impacts aggregate stock market prices, particularly in recessions (Garcia, 2013). Here, we further document that in the cross-section, the news media's tendency to distort prices in bad times is likely more pronounced for smaller firms, since such firms are more exposed to media narratives, and their exposures spike even more disproportionately during market downturns.

Panel B of Figure 2 reveals that firms' exposure to media narratives is not driven by some specific industry group. For example, in the early sample period, the Oil & Mining industry group had one of the lowest average exposures but had one of the largest exposures by the end of the sample. The figure further indicates that media narrative exposure exhibits similar time-series trends across industries—again, commonly surging during market downturns. This evidence illustrates that the extracted media narrative exposures are not merely artifacts of estimation error or random fluctuations. Even though they are estimated individually for each firm, we observe strong commonality over time across groups of stocks.

4 Narratives, Information Channels and Price Informativeness

This Section tests direct model predictions: Section 4.1 examines how exposure to media narratives affects price informativeness regarding future firm fundamentals. Section 4.2 tests the claim that idiosyncratic and other types of non-systematic variance are closely related to narrative exposure, and Section 4.3 establishes how the levels of non-systematic variances directly affect price informativeness.

4.1 Narratives and Price Informativeness

Our model in Section 2 predicts in Proposition 2-(i) an inverse relationship between price informativeness and narrative exposure (i.e., the component of return volatility related to media bias). To determine how exposure to media narratives empirically affects the information content

of stock prices, we adopt a stock-level measure of price informativeness based on Bai, Philippon, and Savov (2016), defined as the predicted variation of cash flows using current market prices. More precisely, we test whether, as our theoretical framework predicts, high exposure to media narratives causes current stock prices to be less informative in terms of future firm cash flows. Our main model is specified as the Fama and MacBeth (1973) regression of future earnings h years from today relative to current assets, $E_{n,t+h}/A_{n,t}$, on current earnings, market value relative to assets, $\ln(M_{n,t}/A_{n,t})$, the interaction of market value and particular narrative exposure, and controls:

$$\frac{E_{n,t+h}}{A_{n,t}} = a + b_{0,h}\frac{E_{n,t}}{A_{n,t}} + b_{1,h}\ln\frac{M_{n,t}}{A_{n,t}} + b_{2,h}\ln\frac{M_{n,t}}{A_{n,t}} \times |\beta_{n,t}^{narr}| + b_{x,h}^{\top}X_{n,t} + \varepsilon_{n,t+h}, \tag{20}$$

where h is one or three years, and $|\beta_{n,t}^{narr}|$ denotes the absolute beta of firm n at time t with respect to a particular narrative or the average of the absolute betas of the 33 identified narratives. The vector of controls, $X_{n,t}$, includes the narrative beta used in interaction term, four-factor model betas, fundamental variables $\ln(Assets)$, $Debt/Assets$, $Cash/Assets$, $Ppent/Assets$, $Capex/Assets$, $Sales/Assets$, $R\&D/Assets$, and economic sector dummies (eight one-digit SIC codes after excluding the financial sector). All continuous variables are winsorized at 5% and 95% for each year in the sample period. The market value variable $\ln(M/A)$ is standardized to unit variance each year in the cross-section so that the coefficient, $b_{1,h}$, directly provides the proxy for price informativeness following Bai, Philippon, and Savov (2016). The coefficient $b_{2,h}$, therefore, reveals how price informativeness interacts with a particular narrative exposure.

Table 2 shows that price informativeness significantly decreases for stocks with high absolute narrative exposure for both the one- and three-year future horizons and for all narratives (except for SCHL for the three-year horizon). The pattern does not seem to be dependent upon the perceived relevance of the specific narratives to certain economic fundamentals or industries. This result delivers a profound message: firms whose stock prices co-vary substantially with media narratives, in general, tend to absorb irrelevant information that renders prices uninformative. Consistent with our model, the loss of price informativeness arises from the inherent media bias that, when traded upon, tends to distort affected firms' stock prices.

	AVR	POLY	REGL	MCRO	EQTY	FINC	ENGY	STPL	HLTH	AUTO	TLCO	ENTM	SCHL
Panel A: One-year horizon.													
$\ln(M/A)_{n,t}$	0.030	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln(M/A)_{n,t} \times \lvert\beta_{n,t}^{narr}\rvert$	-0.099	-0.069	-0.112	-0.139	-0.173	-0.059	-0.049	-0.060	-0.055	-0.014	-0.090	-0.044	-0.100
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R^2 (%)	79.32	78.76	78.76	78.77	78.76	78.79	78.78	78.78	78.77	78.77	78.75	78.75	78.78
Obs.	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152
High Attention	–	-0.065	-0.038	-0.009	-0.100	-0.027	-0.031	-0.026	-0.049	-0.006	-0.043	-0.039	-0.056
Marginal Effect		(0.001)	(0.084)	(0.788)	(0.009)	(0.121)	(0.005)	(0.199)	(0.025)	(0.119)	(0.004)	(0.005)	(0.007)
Panel B: Three-year horizon.													
$\ln(M/A)_{n,t}$	0.054	0.024	0.027	0.027	0.027	0.028	0.029	0.028	0.027	0.028	0.022	0.026	0.017
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.103)
$\ln(M/A)_{n,t} \times \lvert\beta_{n,t}^{narr}\rvert$	-0.167	-0.095	-0.196	-0.277	-0.319	-0.109	-0.093	-0.124	-0.105	-0.025	-0.135	-0.078	-0.101
	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.204)
R^2 (%)	60.15	59.51	59.32	59.26	59.43	59.36	59.47	59.31	59.26	59.41	59.29	59.27	59.51
Obs.	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470
High Attention	–	-0.031	-0.087	0.074	-0.126	-0.047	-0.047	-0.102	-0.088	-0.002	-0.124	-0.077	0.107
Marginal Effect		(0.677)	(0.175)	(0.253)	(0.159)	(0.097)	(0.049)	(0.024)	(0.001)	(0.718)	(0.026)	(0.001)	(0.558)

Table 2: Price Informativeness and Narrative Exposure.
The table shows aggregate price informativeness (coefficient for $\ln(M/A)_{n,t}$) and its interaction with absolute exposure $|\beta_{n,t}^{narr}|$ to selected narratives and the average of absolute betas for all 33 narratives, AVR. The model is estimated as the two-stage regression (20) for one- and three-year horizons (Panels A and B, respectively). Below each panel, the mean interaction term coefficient is computed, conditional on high (above the mean) attention to a narrative. Controls include four-factor betas, fundamental variables, and sector dummies. The sample period is from 1998 to 2021, with annual frequency. Each year, all continuous variables before interactions are winsorized at 5% and 95%, and market value $\ln(M/A)$ is standardized to unit standard deviation. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. R^2(%) and the number of observations (Obs.) are average numbers from the cross-sectional stage.

At the end of each panel in Table 2, we estimate the marginal change in the incremental price informativeness, i.e., the interaction term, conditional on periods of high attention level to a particular narrative. For this, we regress the time-series of the interaction term coefficient $b_{2,h}$ from the cross-sectional stage of the Fama-MacBeth procedure on a constant and a dummy variable that equals one for the years of high attention to the specific narrative, defined as periods when attention to the narrative is above its sample mean, and zero otherwise. We report the coefficient on the dummy variable along with its $p-$value. For the majority of narratives for the one-year horizon and for five out of 12 narratives for the three-year horizon, high attention significantly (at 5% level) exacerbates the loss of price informativeness for exposed stocks. Which narratives have a stronger marginal effect is hardly anticipated ex-ante—e.g., the Macroeconomy (MCRO) narrative is insignificant, while Entertainment (ENTM) and Telecoms & Social Media (TLCO) are both significant.

To illustrate the economic magnitude of these effects, we standardize the absolute narrative betas each year in the panel data. For the one-year horizon, the absolute exposure to individual narratives significantly decreases price informativeness by almost identical magnitudes (-0.006 to -0.007) for a standard deviation increase in the exposures. On the other hand, the magnitude

of the decline in price informativeness following a standard deviation increase in the average of absolute exposures to all 33 narratives (AVR) is doubled, at roughly -0.014. For the three-year horizon, we obtain slightly more heterogeneity in economic magnitudes but still find an almost uniform significance of interaction term coefficients, with the exception of the SCHL narrative. Here, again, the economic magnitude of the loss in price informativeness associated with a standard deviation increase in AVR is about two times larger than the average of the individual effects.

The effects are so strong that if we modify model (20) to use an indicator function for high absolute betas in place of the absolute narrative betas, we observe a complete loss of price informativeness for stocks with the highest exposures to media narratives. Precisely, we define the dummy variable for high media narrative exposure, $\mathbf{1}_{H\ |\beta^{narr}_{n,t}|}$, based on whether a stock's $|\beta^{narr}_{n,t}|$ is above the 75th percentile in the cross-section for a given year and narrative. We then obtain the total price informativeness for high media-narrative-exposed firms as the sum of the coefficients of $\ln(M/A)_{n,t}$ and its interaction with $\mathbf{1}_{H\ |\beta^{narr}_{n,t}|}$. Table 3 reveals that the total price informativeness for highly exposed firms is insignificant for *all* individual narratives, even without conditioning on the attention level. Moreover, the total effect for stocks with high average absolute narrative exposure (AVR) is significantly negative.

4.2 Information Channels and Firm Characteristics

In the model in Section 2, non-systematic variance (i.e., variance not generated by the factors driving firm fundamentals) arises because of media narrative exposure. Eqs. (16) and (17) show that absolute narrative exposure is a proxy for non-systematic variance. While we hardly expect that, in reality, only exposure to narratives generates non-systematic variance, we analyze the empirical link between the two concepts in the cross-section of stocks to establish how much of cross-sectional variability in non-systematic variance is explained by exposure to narratives.

In the subsequent analysis, we use two sets of proxies for information channels driving stock returns, clearly separating variance sources into systematic and non-systematic components.[10] The first is a combination of systematic ($SysVar$) and idiosyncratic variances ($IdVar$) estimated

[10]Note that partitions of return variance into components provide us a view of the intensity of information channels driving stock returns.

	AVR	POLY	REGL	MCRO	EQTY	FINC	ENGY	STPL	HLTH	AUTO	TLCO	ENTM	SCHL
Panel A: One-year horizon.													
$\ln(M/A)_{n,t}$	0.013	0.009	0.009	0.009	0.009	0.010	0.010	0.009	0.009	0.009	0.009	0.009	0.009
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln(M/A)_{n,t} \times \mathbf{1}_{H\ \lvert\beta^{narr}_{n,t}\rvert}$	-0.026	-0.012	-0.013	-0.013	-0.012	-0.014	-0.013	-0.012	-0.013	-0.013	-0.012	-0.013	-0.013
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R^2 (%)	79.15	78.71	78.72	78.73	78.71	78.76	78.73	78.73	78.73	78.73	78.71	78.71	78.72
Obs.	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152	3,152
Total for $H\lvert\beta^{narr}\rvert$	-0.013	-0.003	-0.003	-0.003	-0.003	-0.004	-0.004	-0.003	-0.003	-0.003	-0.003	-0.003	-0.004
	(0.001)	(0.183)	(0.191)	(0.159)	(0.231)	(0.041)	(0.077)	(0.244)	(0.219)	(0.163)	(0.249)	(0.176)	(0.141)
Panel B: Three-year horizon.													
$\ln(M/A)_{n,t}$	0.025	0.019	0.019	0.019	0.019	0.020	0.022	0.019	0.020	0.019	0.018	0.019	0.017
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln(M/A)_{n,t} \times \mathbf{1}_{H\ \lvert\beta^{narr}_{n,t}\rvert}$	-0.038	-0.015	-0.019	-0.018	-0.021	-0.024	-0.029	-0.021	-0.023	-0.024	-0.019	-0.022	-0.011
	(0.001)	(0.079)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.334)
R^2 (%)	59.82	59.53	59.29	59.19	59.33	59.29	59.63	59.15	59.17	59.28	59.24	59.21	59.34
Obs.	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470	2,470
Total for $H\lvert\beta^{narr}\rvert$	-0.013	0.005	0.001	0.001	-0.003	-0.004	-0.006	-0.002	-0.003	-0.004	-0.001	-0.003	0.007
	(0.001)	(0.615)	(0.858)	(0.893)	(0.610)	(0.270)	(0.294)	(0.559)	(0.297)	(0.261)	(0.684)	(0.358)	(0.449)

Table 3: Price Informativeness For Highly Exposed Stocks.
The table lists aggregate price informativeness (coefficient for $\ln(M/A)_{n,t}$) and its marginal change for firms with high (above 75th percentile for a given year) absolute exposure to selected narratives ($\mathbf{1}_{H|\beta^{narr}_{n,t}|}$) or high average of absolute betas for all 33 narratives. The model is estimated as the tw-stage regression for one- and three-year horizons (Panels A and B, respectively). Below each panel, the total effect of high absolute narrative exposure is computed. Controls include four-factor betas, fundamental variables, and sector dummies. The sample period is from 1998 to 2021, with annual frequency. Each year, all continuous variables before interactions are winsorized at 5% and 95%, and are then standardized to unit standard deviation. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. $R^2(\%)$ and the number of observations (Obs.) are average numbers from the cross-sectional stage.

from several standard factor models. Systematic variance captures market-wide information that jointly affects all individual firms' stock prices and is not particularly informative regarding an individual firm's future cash flow. Conversely, idiosyncratic variance stems from at least three sources: (i) firm-specific information not reflected in the aggregate market dynamics, (ii) agents' heterogeneous interpretation of how public information deferentially affects firms, and (iii) noise trading unrelated to either public or firm-specific information. The relationship between the level of idiosyncratic variances and the corresponding asset prices' informativeness will likely depend on which of these sources of idiosyncratic price variation is dominant for specific stocks.

Our second set of information channel targets a different and more granular decomposition of stock return variation, allowing for a finer separation of the components of idiosyncratic variance. Precisely, we use the framework of Brogaard, Nguyen, Putnins, and Wu (2022) (BNPW henceforth) to decompose total stock return variance into components stemming from market-wide ($MktInfo$), private ($PrivateInfo$) or public ($PublicInfo$) firm-specific information, and noise ($Noise$). $MktInfo$ is similar to $SysVar$ from a factor model but is identified using vector autoregression as the response of stock returns to market factor shocks only. Private and public firm-specific information is identified as a permanent stock return response to trading volume

and own-return shocks after controlling for market return shocks. Noise absorbs the residual variance.[11]

	$SysVar_{n,t}$	$IdVar_{n,t}$	$MktInfo_{n,t}$	$PrivateInfo_{n,t}$	$PublicInfo_{n,t}$	$Noise_{n,t}$
$SysVar_{n,t}$	1.000	0.043	0.551	0.135	0.096	0.002
$IdVar_{n,t}$	0.043	1.000	0.342	0.783	0.891	0.841
$MktInfo_{n,t}$	0.551	0.342	1.000	0.363	0.407	0.184
$PrivateInfo_{n,t}$	0.135	0.783	0.363	1.000	0.722	0.502
$PublicInfo_{n,t}$	0.096	0.891	0.407	0.722	1.000	0.643
$Noise_{n,t}$	0.002	0.841	0.184	0.502	0.643	1.000

Table 4: Correlation of Information Channels.
The table provides unconditional correlations of information channel proxies for individual stocks: systematic and idiosyncratic variances based on the four-factor model, and BNPW variance decomposition. The sample period is from 1998 to 2021, with annual frequency. All proxies are computed, winsorized at 5% and 95%, and are then standardized to unit variance on an annual basis.

Table 4 shows the correlation of the information channel proxies. The systematic and non-systematic information sources do not overlap much across the two methodologies, but the factor-based systematic variance is somewhat correlated (0.55) with systematic variance from BNPW. On the other hand, the factor-based idiosyncratic variance is highly correlated with all three non-systematic variance components from BNPW (correlations of 0.8-0.9). We see that all of the non-systematic variance components are jointly driven by some common factors or characteristics, and the intensities of the information channels they reflect are strongly connected.

In our model, exposure to media narrative shocks is a common characteristic reflecting public information that deferentially affects agents' perceptions, sentiments, and trading decisions, all of which can distort stock prices. While the news media can be informative for several purposes, our model indicates that due to biases, firms whose stock prices co-move disproportionately with media narratives are predominantly subject to sentiment waves and noise trading, which dampen the informativeness of stock prices. This is consistent with recent empirical studies which argue that factors captured by news flows reflect non-priced risk (e.g., Calomiris and Mamaysky, 2019), that stock investors over- and under-react to news media coverage (Frank and Sanati, 2018), and that news media sentiment distorts aggregate stock market prices (e.g., Tetlock, 2007), especially in bad times (e.g., Garcia, 2013) when agents' decisions are more susceptible to news media content due to their magnified psychological impacts.

[11] BNPW note that "in reality, the distinction between public and private information can at times be blurred," so we refrain from drawing strong conclusions based on this distinction.

Overall, we expect a strong link between absolute narrative exposure and the non-systematic variance measures in the cross-section of stocks. We formally test this hypothesis using a two-stage procedure, in which we annually regress each variance component on stocks' absolute narrative exposure ($Average\ |\beta_{n,t}^{narr}|$) while controlling for a host of stock characteristics and then analyze the time-series average coefficients. A large set of traditional characteristics enables us to isolate the relevance of media narrative exposure from other variables that are potentially relevant to cross-sectional differences in the variance components.

The results are provided in Table 5. The full specification in Panel A includes average absolute narrative exposure, four-factor betas, fundamental variables, stock characteristics, and sector fixed effects. The reduced specification in Panel B contains only the average absolute narrative beta. All continuous variables on both sides are winsorized annually at 5% and 95%, and are then standardized to have a cross-sectional variance of one. The results are truly striking. Comparing the estimates in the specifications in Panels A and B, we see that in terms of economic magnitude, media narrative exposure is the single most important driver of non-systematic variance components in stock returns. More so, media narrative exposure alone explains a whooping 56%-83% of the variation in idiosyncratic variance, variances due to public and private information and noise components.

For example, a one-standard-deviation (STD) increase in the average absolute narrative beta is linked to a $0.91\times STD$ increase in the idiosyncratic variance $IdVar$ in Panel B, and to a $0.76\times STD$ increase after controlling for all other characteristics in Panel A. The reduced specification's R^2 of 83% increases by less than 3% in the full specification. The $PublicInfo$ column shows a similar pattern: a $1\times STD$ increase in the average absolute narrative beta is linked to $0.84\times STD$ and a $0.66\times STD$ increase in the variance due to public information for the reduced and full specifications, respectively. The R^2's are 74% and 71% for the full and reduced specifications, respectively. $Noise$ and $PrivateInfo$ are slightly less strongly related to narrative exposure. $MktInfo$ is statistically linked to narrative exposure, but the economic magnitude is relatively negligible. The factor-based systematic variance is not related to narrative exposure.

Thus, stocks highly exposed to media narratives also have high levels of idiosyncratic variance linked to (and potentially explained by) high variance due to trading on public information and

	$Var_{n,t}$	$SysVar_{n,t}$	$IdVar_{n,t}$	$MktInfo_{n,t}$	$PrivateInfo_{n,t}$	$PublicInfo_{n,t}$	$Noise_{n,t}$
Panel A: Full Specification.							
Average $\lvert\beta_{n,t}^{narr}\rvert$	0.750	-0.008	0.757	0.213	0.644	0.663	0.613
	(0.001)	(0.556)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R^2 (%)	85.21	82.04	85.30	50.56	61.59	74.34	63.70
Obs.	2,260	2,260	2,260	2,260	2,260	2,260	2,260
Factor betas	FF4	FF4	FF4	FF4	FF4	FF4	FF4
Fundamentals	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Reduced Specification.							
Average $\lvert\beta_{n,t}^{narr}\rvert$	0.906	0.049	0.909	0.347	0.742	0.841	0.757
	(0.001)	(0.207)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R^2 (%)	82.07	2.22	82.70	12.95	55.27	70.69	57.46
Obs.	2,260	2,260	2,260	2,260	2,260	2,260	2,260

Table 5: Information Channels and Firm Characteristics.
The table shows the cross-sectional link between the intensity of information channels driving individual stock returns and firm characteristics. Information channels are systematic and idiosyncratic variances based on the four-factor model and BNPW variance decomposition. The coefficients are based on the two-stage regression. Panel A shows results with all regressors and sector dummies control, and Panel B shows a reduced specification without controls. The sample period is from 1998 to 2021, with annual frequency. All continuous variables are winsorized at 5% and 95%, and are then standardized to unit variance in the cross-section on an annual basis. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. $R^2(\%)$ and the number of observations (Obs.) are average numbers from the cross-sectional stage.

noise produced by the news media. In the next section, we directly test whether there is a statistical link between idiosyncratic variance, variances due to public information and noise on the one side, and price informativeness on the other.

4.3 Information Channels and Price Informativeness

We rely on the same methodology in the previous sections to measure price informativeness and analyze the interaction term coefficients $b_{proxy,h}$ in the following specification:

$$\frac{E_{n,t+h}}{A_{n,t}} = a + b_{0,h}\frac{E_{n,t}}{A_{n,t}} + \left[b_{1,h} + b_{proxy,h}^{\top} proxy_{n,t}\right] \times \ln\frac{M_{n,t}}{A_{n,t}} + b_x^{\top} X_{n,t} + \varepsilon_{n,t+h}, \tag{21}$$

where h is one or three years, $proxy_{n,t}$ denotes a vector with information channel proxies of firm n, and the control variables vector X is the same as in the model (20).[12] In terms of information channel proxies, we use the two sets of variance decomposition, factor-based variances, and variances from VAR estimation in BNPW. As before, the coefficient, $b_{1,h}$, provides

[12]We also keep unchanged any processing of variables: All continuous variables are winsorized at 5% and 95% for each year in the sample period. The market value variable $\ln(M/A)$, each information channel captured by $proxy_{n,t}$ and all continuous control variables are standardized to unit-standard deviations in the cross-section.

the proxy for price informativeness, and the vector of coefficients $b_{proxy,h}$ additionally reveals how price informativeness interacts with information channels.

The results in Table 6 demonstrate that while stock prices are, on average, informative about future fundamentals for horizons of one (Panel A) and three (Panel B) years, price informativeness significantly decreases for stocks with high levels of idiosyncratic variance. The effect is economically large, and $1 \times STD$ difference in $IdVar$ decreases the price informativeness by 70% (adjustment of -0.014 applied to the base level of 0.020). The levels of systematic variance in most cases do not significantly affect price informativeness (except for the five-factor model with a borderline $p-$value of 0.056 for an interaction term). In all cases, the interaction term for the $SysVar$ is approximately an order of magnitude smaller than for the $IdVar$. The results for all factor models in the table (market to five-factor models) are similar.[13]

With a more granular variance decomposition (in column BNPW), we observe for both horizons the largest and most significant decrease in price informativeness for stocks with high $PublicInfo$ variance. Keeping market value constant, a $1 \times STD$ change in $PublicInfo$ decreases price informativeness about future one-year fundamentals by around 50% (i.e., by 0.010 compared to the base level of 0.022). $Noise$ also significantly drives price informativeness in the same direction, with the economic magnitude roughly 2.5 times smaller. $PrivateInfo$ and $MktInfo$ are also statistically significant, but economically, their contribution is quite small. For the three-year horizon, $PublicInfo$ is the only information channel significantly interacting with price informativeness, but at a moderate 10% significance level. Interaction with the noise component is economically sizeable but not significant. Overall, price informativeness is negatively associated with non-systematic variance, and the effect is primarily driven by public information.

Thus, consistent with the model, high narrative exposure is linked to lower stock price informativeness, with the effects being stronger in periods with high media narrative attention, and is strongly related to higher non-systematic variance in the cross-section of stocks.

[13] In most of the analysis, we select the Carhart (1997) four-factor model as our benchmark and check the sensitivity to other factor models in terms of robustness.

	MM	FF3	FF4	FF5	BNPW
Panel A: One-year horizon.					
$\ln(M/A)_{n,t}$	0.019	0.020	0.020	0.020	0.022
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln(M/A)_{n,t} \times SysVar_{n,t}$	-0.000	-0.001	-0.001	-0.002	–
	(0.784)	(0.101)	(0.131)	(0.056)	
$\ln(M/A)_{n,t} \times IdVar_{n,t}$	-0.014	-0.014	-0.014	-0.014	–
	(0.001)	(0.001)	(0.001)	(0.001)	
$\ln(M/A)_{n,t} \times MktInfo_{n,t}$	–	–	–	–	-0.002
					(0.001)
$\ln(M/A)_{n,t} \times PrivateInfo_{n,t}$	–	–	–	–	-0.002
					(0.001)
$\ln(M/A)_{n,t} \times PublicInfo_{n,t}$	–	–	–	–	-0.010
					(0.001)
$\ln(M/A)_{n,t} \times Noise_{n,t}$	–	–	–	–	-0.004
					(0.001)
R^2 (%)	79.48	79.49	79.49	79.49	80.18
Obs.	3,152	3,152	3,152	3,152	2,224
Factor betas	FF4	FF4	FF4	FF4	FF4
Fundamentals	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes
Panel B: Three-year horizon.					
$\ln(M/A)_{n,t}$	0.035	0.037	0.037	0.037	0.047
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln(M/A)_{n,t} \times SysVar_{n,t}$	0.001	0.001	0.001	0.001	–
	(0.667)	(0.770)	(0.757)	(0.782)	
$\ln(M/A)_{n,t} \times IdVar_{n,t}$	-0.021	-0.024	-0.025	-0.025	–
	(0.001)	(0.001)	(0.001)	(0.001)	
$\ln(M/A)_{n,t} \times MktInfo_{n,t}$	–	–	–	–	0.012
					(0.436)
$\ln(M/A)_{n,t} \times PrivateInfo_{n,t}$	–	–	–	–	0.019
					(0.431)
$\ln(M/A)_{n,t} \times PublicInfo_{n,t}$	–	–	–	–	-0.029
					(0.090)
$\ln(M/A)_{n,t} \times Noise_{n,t}$	–	–	–	–	-0.024
					(0.198)
R^2 (%)	60.37	60.50	60.50	60.52	62.19
Obs.	2,470	2,470	2,470	2,470	1,736
Factor betas	FF4	FF4	FF4	FF4	FF4
Fundamentals	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes

Table 6: Information Channels and Price Informativeness.
The table shows aggregate price informativeness (coefficient for $\ln(M/A)_{n,t}$) and its interaction with various information channel proxies. The model is estimated as the Fama-MacBeth regression (21) for one- and three-year horizons (Panels A and B, respectively). The first four columns use one-, three-, four-, and five-factor models for variance decomposition into systematic ($SysVar_{n,t}$) and idiosyncratic ($IdVar_{n,t}$) components, and column $BNPW$ uses a decomposition of Brogaard, Nguyen, Putnins, and Wu (2022). Controls include four-factor model betas, a number of fundamental variables, and sector dummies. The sample period is from 1998 to 2021, with annual frequency. Each year, all continuous variables before interactions are winsorized at 5% and 95%, and are standardized to unit variance. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. R^2(%) and number of observations (Obs.) are average numbers from the cross-sectional stage.

5 Additional Analysis

This Section provides additional analysis that is not in the direct scope of our model: Section 5.1 analyzes a link between narrative exposure and trading volume. Section 5.2 investigates how narrative exposure affects the firm valuations relative to industry peers, while Section 5.3 estimates the arbitrage risk of firms with high levels of narrative exposure to provide an explanation of the valuation results.

5.1 Narrative Exposure and Trading Activity

Following the model predictions, shocks to media narrative attention changes the information available to agents, leading to updates in stock return expectations and the subsequent adjustments in portfolio holdings. Thus, stocks affected more strongly by narrative attention shocks should experience higher turnover. We test this claim by relating average turnover to the average stock-specific narrative shock, controlling for a number of other variables that potentially affect market activity.

We continue working on the annual frequency, and use the same two-stage framework as in the previous sections. We quantify the stock-specific average narrative shock by the average product of absolute narrative betas and volatility of daily attention to a given narrative within a year $Average\ |\beta_{n,t}^{narr} \times \sigma_{n,t}^{narr}|$, using all 33 identified narratives. The turnover is computed as the yearly average of the ratio of trading volume (number of shares traded) to the total number of shares outstanding.

The results in Table 7 confirm the model predictions, and the coefficient on the $Average\ |\beta_{n,t}^{narr} \times \sigma_{n,t}^{narr}|$ is positive and significant for all specifications. In the first column for the regression without extra controls (except for sector fixed effects) the average first-stage R^2 is 11.4%, and $1 \times STD$ higher average narrative corresponds to $0.2 \times STD$ higher relative turnover. Adding various controls for the same sample (up to column 4 in the Table) boosts the explanatory power of the cross-sectional stage, and also increases the slope of the average narrative shock, which hints at potential interaction between regressors. For a smaller sample of stocks (in columns 5 and 6), $1 \times STD$ higher average narrative corresponds to 0.6 to $0.7 \times STD$ higher relative turnover, after controlling for all other characteristics.

	$Turnover_{n,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$Average\ \lvert\beta_{n,t}^{narr} \times \sigma_{n,t}^{narr}\rvert$	0.202	0.187	0.369	0.432	0.627	0.669
	(0.011)	(0.012)	(0.001)	(0.001)	(0.001)	(0.001)
$DOB_{n,t}$	–	–	–	–	0.036	0.077
					(0.031)	(0.001)
$Inst.\ Ownership_{n,t},\ \%$	–	–	–	–	–	0.472
						(0.001)
R^2 (%)	11.39	32.94	41.15	43.86	40.49	47.89
Obs.	3,204	3,204	3,204	3,204	961	954
Factor betas	FF4	FF4	FF4	FF4	FF4	FF4
Fundamentals	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 7: Narrative Exposure and Trading Volume.
The table shows the cross-sectional link between the turnover ($Turnover$) defined as the average of the ratio of trading volume relative to shares outstanding, average narrative exposure scaled by the volatility of a given narrative ($Average\ \lvert\beta_{n,t}^{narr} \times \sigma_{n,t}^{narr}\rvert$), and various firm and stock characteristics, including sector dummies (SIC1-code) and in the last two specifications also dispersion of beliefs (DOB) and institutional ownership. The coefficients are based on the two stage regression. The sample period is from 1998 to 2021, with annual frequency. For the specification with the institutional ownership the sample period is from 1999 to 2018. All variables except for industry dummies are winsorized at 5% and 95%, and are then standardized to unit variance in the cross-section on an annual basis. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. R^2(%) and the number of observations (Obs.) are average numbers from the cross-sectional stage.

5.2 Narrative Exposure and Firm Valuation

Given the heterogeneity in narrative exposures within industries, prices of firms with high narrative exposures become less informative regarding future fundamentals and can deviate in valuation from comparable firms. Aabo, Pantzalis, and Park (2017) find that stocks with a high proportion of noise trading, identified by high levels of idiosyncratic volatility, are incorrectly valued according to four different mispricing measures. The previous section indicates that exposure to media narratives is a major source of noise in stock returns, as well as public news that is uninformative about firm fundamentals. Therefore, it is likely that firms that have high exposure to media narratives are predominantly mispriced relative to their peers. Much less clear is whether such firms have overvalued, speculative, and "hyped" stocks (Teeter and Sandberg, 2017), or are instead undervalued. It is also possible for media narrative exposure to have no persistent directional effect on mispricing, such that narrative-driven trading merely increases idiosyncratic variance and distorts valuations equally in both directions.

To identify mispricing, we adopt the methodology of Rhodes-Kropf, Robinson, and Viswanathan (2005) (RKRV), who develop a decomposition that separates the market-to-book ratio into firm-

specific error, industry-specific error, and long-term value-to-book ratio. For our purposes, we are primarily interested in the firm-specific error that measures the deviation in a firm's observed value from the valuation implied by current sector accounting multiples. Each year, we group all firms according to the two-digit SIC classification, estimate the following annual cross-sectional regression for each industry, and save the estimated coefficients:

$$\ln M_{n,t} = \alpha_{0,j,t} + \alpha_{1,j,t} \ln BE_{n,t} + \alpha_{2,j,t} \ln |NI|_{n,t} + \alpha_{3,j,t} I_{(<0)} \ln |NI_{n,t}| + \alpha_{4,j,t} LEV_{n,t} + \varepsilon_{n,t}, \tag{22}$$

where $\ln M_{n,t}$ is the log market value of firm n that belongs to industry j at time t; $\ln BE_{n,t}$ refers to the log book value; $|NI_{n,t}|$ refers to the absolute value of net income; $I_{(<0)}$ indicates where net income is negative; and $LEV_{n,t}$ refers to firm leverage measured as the ratio of long-term debt to the sum of long-term debt and a firm's equity market value. The estimated annual industry accounting multiples, $\widehat{\alpha}_{j,t}$, are then used to compute the fitted value of each firm in the industry for that year. The firm-specific valuation error is then the log difference between the actual and the fitted market values for a given firm in a given year:

$$MISVAL_{n,t} = \ln M_{n,t} / \widehat{M_{n,t}}, \tag{23}$$

and it is negative for undervalued firms and positive for overvalued ones.

We continue to use the two-stage methodology, as in the previous sections, and analyze how absolute narrative exposure is related to misvaluation level and the probability of undervaluation. For the first set of analyses, we directly use $MISVAL_{n,t}$ as the dependent variable. In an alternative specification, we adopt a linear probability model in which the dependent variable is a dummy variable that equals one if $MISVAL_{n,t} < 0$ and zero otherwise.[14] Thus, we estimate the following model:

$$Y_{n,t} = a + b_0 |\beta_{n,t}^{narr}| + b_x^{\top} X_{n,t} + \varepsilon_{n,t}, \tag{24}$$

[14] We also estimate a Probit model using the same controls, and the results are very similar in terms of significance and marginal effects.

where $Y_{n,t}$ is either $MISVAL_{n,t}$ or $\mathbf{1}_{MISVAL_{n,t}<0}$, the $|\beta^{narr}_{n,t}|$ is either the average absolute exposure to all narratives or absolute narrative betas to selected narratives, and vector $X_{n,t}$ includes four-factor betas, standard fundamental controls, and sector (one-digit SIC) dummies.[15] Table 8, Panel A, lists estimates of the association between media narratives and level of mis-

	AVR	POLY	REGL	MCRO	EQTY	FINC	ENGY	STPL	HLTH	AUTO	TLCO	ENTM	SCHL
Panel A: Misvaluation $MISVAL_{n,t}$.													
$\|\beta^{narr}_{n,t}\|$	-0.119	-0.039	-0.041	-0.038	-0.041	-0.044	-0.039	-0.042	-0.041	-0.042	-0.046	-0.044	-0.041
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R^2 (%)	18.82	17.03	17.07	17.02	17.08	17.15	17.00	17.06	17.04	17.08	17.18	17.13	17.07
Obs.	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995
Panel B: Undervaluation dummy $\mathbf{1}_{MISVAL_{n,t}<0}$.													
$\|\beta^{narr}_{n,t}\|$	0.073	0.024	0.024	0.023	0.025	0.026	0.023	0.025	0.021	0.027	0.029	0.027	0.025
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R^2 (%)	11.82	10.96	10.91	10.91	10.95	10.96	10.93	10.93	10.88	10.97	11.02	10.97	10.96
Obs.	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995	2,995

Table 8: Narrative Exposure and Firm Valuation.
The table shows the dependency of firm misvaluation on absolute exposure to selected narratives ($|\beta^{narr}_{n,t}|$), or on average absolute betas for all 33 narratives. The model is estimated as the two-stage regression (24) for misvaluation proxy computed following RKRV ($MISVAL_{n,t}$), with a dummy variable $\mathbf{1}_{MISVAL_{n,t}<0}$ indicating undervaluation (Panels A and B, respectively). Controls include four-factor betas, fundamental variables, and sector (SIC1) dummies. The sample period is from 1998 to 2021, with annual frequency. Each year, all continuous exogenous variables are winsorized at 5% and 95%, and are then standardized to unit standard deviation. $p-$values in parentheses use Newey and West (1987) standard errors with three lags, and are replaced by 0.001 if smaller. R^2(%) and the number of observations (Obs.) are average numbers from the first stage.

valuation, indicating that undervaluation is larger for firms with elevated exposure to all of the selected narratives with about equal magnitude. On the other hand, the magnitude of the effect for the average exposure to all narratives (AVR) is roughly three times higher. The regressors are standardized, and the reported magnitudes have a clear and simple interpretation: a $1 \times STD$ difference in the average absolute narrative beta corresponds to about a 12% lower valuation for a firm compared to industry peers. The coefficients in Panel B can be interpreted as changes in the probability of being undervalued for a $1 \times STD$ difference in the absolute narrative betas. A predicted probability difference for a $1 \times STD$ difference in average exposure (AVR) is 7.3 percentage points, and it is significant, both statistically and economically. The effects for individual narratives are all roughly three times smaller, but all are statistically significant.

Overall, the results show that high media-narrative-exposed firms are predominantly undervalued relative to their peers. This result is consistent with, on the one hand, aspects of our model that feature biases from the media and investors and, on the other hand, studies that highlight limited attention (e.g., Peng and Xiong, 2006; DellaVigna and Pollet, 2009) and news

[15] All continuous variables are winsorized annually at 5% and 95%, and all continuous right-hand variables are standardized annually to arrive at the standard deviation of one.

media's tendency to exhibit a negative slant because it attracts more attention (e.g., Liu and Matthies, 2022; Sacerdote, Sehgal, and Cook, 2020). When taken together, these forces imply that biased investors with limited attention will likely overreact to the more attention-grabbing negative news produced by the media. This, in turn, disproportionately depresses the stock prices of the high media-narrative-exposed firms, leading to their relative undervaluation.

5.3 Arbitrage Risk

As we have seen above, firms exposed to narratives are not typical "hyped" stocks driven by the crowd on Reddit, but are, rather, undervalued stocks with excessive idiosyncratic variance. If this undervaluation is not due to a systematic risk factor omitted in the procedure for computing firm valuation, then the observed deviations in firm values from their fair levels represent a trading opportunity. Why is this opportunity not exploited by arbitrageurs? A likely reason is that an arbitrage portfolio would be exposed to high risk due to elevated variance linked to narrative exposure and narrative-induced trading.

For a formal test, we create an arbitrage risk measure for each stock, following Wurgler and Zhuravskaya (2002). For each stock in our sample, we select, at the end of each year, the three closest *substitute* stocks matched in terms of industry, size, and market-to-book ratio.[16] To select the closest stocks, we compute the sum of the absolute percentage difference of size and market-to-book ratio of each firm relative to all industry peers. The three firms with the smallest percentage difference are then selected as substitute firms. To measure arbitrage risk for stock n each year, we take daily returns for the given year and estimate the time-series regression:

$$r_{n,\tau} - r_{f,\tau} = \beta_{1,n}\left(r_{s1n,\tau} - r_{f,\tau}\right) + \beta_{2n}\left(r_{s2n,\tau} - r_{f,\tau}\right) + \beta_{3n}\left(r_{s3n,\tau} - r_{f,\tau}\right) + \varepsilon_{n,\tau}, \tag{25}$$

where $r_{s1n,\tau}$, $r_{s2n,\tau}$, and $r_{s3n,\tau}$ denote returns on three industry, size, and market-to-book matched substitute stocks, while $r_{f,\tau}$ denotes the risk-free rate. Arbitrage Risk ($ARBRISK_{n,t}$) for stock n in year t is the variance of the residuals $\varepsilon_{n,\tau}$ from this regression. Higher variance indicates poorer substitutes in explaining stock n's returns and, consequently, higher arbitrage risk.

[16] Following Wurgler and Zhuravskaya, we take the classification by Fama and French (1997) with 48 industries.

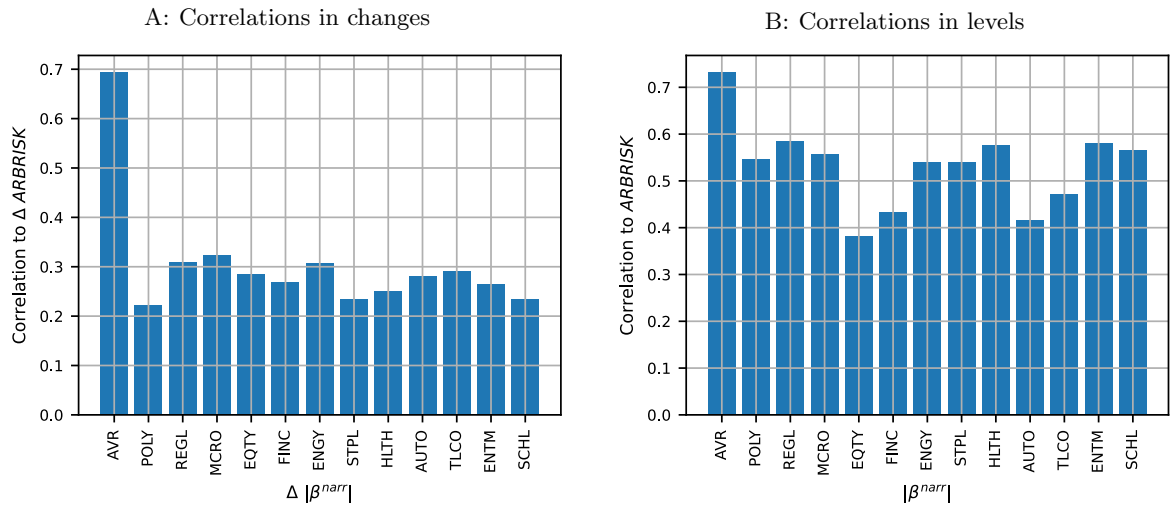

Figure 3: Arbitrage Risk and Narrative Exposure.
The Figure shows the correlations between firm arbitrage risk (computed from Eq. (25)) and absolute exposure $|\beta^{narr}|$ to selected narratives. Panel A shows correlations in changes, and Panel B in levels. The changes are computed from the panel of firm-year arbitrage risk measures and absolute narrative betas (AVR denotes the average of all individual narratives). All variables are winsorized annually at 5% and 95% levels.

Figure 3 delineates the correlations between annual changes (Panel A) and levels (Panel B) in firm arbitrage risk and absolute exposure to narratives, computed from the panel of firm-year arbitrage risk measures and absolute narrative betas (AVR is the average of all individual narrative exposures). Changes in absolute exposure to all individual narratives are highly correlated to changes in arbitrage risk, and the levels of correlation are very similar across all narratives (0.27 on average). The correlation of 0.69 for the average absolute exposure (AVR) and almost uniformly high correlations in levels indicate that stocks with high media narrative exposure (or narrative-induced trading fueling idiosyncratic variance) simultaneously have high arbitrage risk, which deters sophisticated investors from exploiting the apparent undervaluation.

6 Conclusion

We establish theoretically and empirically that attention to media narratives can distort stock prices and decrease their informativeness about future fundamentals. Importantly, we define attention to narratives without measuring their sentiment, that is, in a manner that is consistent with widely used Natural Language Processing methods, such as Latent Dirichlet Allocation (LDA), for extracting topics from news and measuring attention as the proportion of words and phrases associated with a given topic in the total news material. Using a trading model

with time-varying public information production that maps tightly to the LDA methodology employed in our empirical analysis, we demonstrate that in the presence of biased media and investors, attention to the news, which is otherwise not correlated to stock returns, affects stock prices. The weight of biased investors in the economy and the level of attention to a particular narrative distort price informativeness.

Empirically, stock prices of firms with high levels of absolute narrative betas become uninformative regarding future fundamentals, and high attention to certain narratives further deteriorates the information content of exposed firms' stock prices. The model implies that narrative exposure creates non-systematic variances in stock returns, indicating that absolute narrative exposure plays a crucial role in generating excess volatility in returns. Analyzing information channels (identified as components of stock return variance) through which attention to narratives flows to financial markets, we indeed identify absolute narrative exposure as the main characteristic that alone explains 70-80% of the cross-sectional variation in idiosyncratic variance and variance due to firm-specific public information. Adding 15 more accounting and market variables as controls boosts the average cross-sectional R^2 by only three percentage points. Consequently, using two different methods of variance decomposition into systematic and firm-specific (idiosyncratic and noise) components, we reveal that high levels of idiosyncratic variance render market stock prices uninformative regarding future firm fundamentals. Stocks strongly affected by the narrative attention shocks experience higher average trading volume, which indicates that media's narrative attention is feeding into latent demand for individual stocks. Firms highly exposed to narratives are not 'hyped' overvalued firms. Instead, they are strongly undervalued relative to industry peers based on accounting multiples, with a one-standard-deviation difference in average absolute narrative exposure corresponding to about a 12% undervaluation.

Our study complements and extends several major research fields. Abstracting from predictability and risk premium, we show how attention to media narratives interacts with asset return dynamics, creating a bias in the prices of stocks with elevated narrative exposure and distorting their information content. According to existing studies, attention to media narratives can be useful in predicting returns and defining risk premiums. We demonstrate the detrimental media effects on price efficiency, and they are not trivial, both statistically and

economically. Consistent with the proposed theoretical mechanism, we empirically establish narrative exposure as the major characteristic explaining idiosyncratic variance in the cross-section of stocks, complementing the residual household income risk channel of Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016). Linking to the literature on differences in beliefs, we propose attention to media narratives as a theoretically sound and empirically important channel of disagreement in financial markets. Price adjustments resulting from public information flows are one of the major components of non-systematic variance.

References

Aabo, T., C. Pantzalis, and J. C. Park, 2017, "Idiosyncratic volatility: An indicator of noise trading?," *Journal of Banking and Finance*, 75, 136–151.

Amihud, Y., 2002, "Illiquidity and Stock Returns: Cross-section and Time-series Effects," *Journal of Financial Markets*, 5(1), 31–56.

Bai, J., T. Philippon, and A. Savov, 2016, "Have financial markets become more informative?," *Journal of Financial Economics*, 122(3), 625 – 654.

Baker, S. R., N. Bloom, and S. J. Davis, 2016, "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131(4), 1593–1636.

Baloria, V. P., and J. Heese, 2018, "The effects of media slant on firm behavior," *Journal of Financial Economics*, 129(1), 184–202.

Barber, B. M., and T. Odean, 2000, "Trading is hazardous to your wealth: The common stock investment performance of individual investors," *The Journal of Finance*, 55(2), 773–806.

Barberis, N., 2018, "Psychology-based models of asset prices and trading volume," in *Handbook of behavioral economics: applications and foundations 1*. Elsevier, vol. 1, pp. 79–175.

Barberis, N., A. Shleifer, and R. Vishny, 1998, "A Model of Investor Sentiment," *Journal of Financial Economics*, 49(3), 307–343.

Blei, D. M., A. Y. Ng, and M. I. Jordan, 2003, "Latent Dirichlet Allocation," *The Journal of Machine Learning Research*, 3(null), 993–1022.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer, 2020, "Overreaction in macroeconomic expectations," *American Economic Review*, 110(9), 2748–82.

Brogaard, J., T. H. Nguyen, T. J. Putnins, and E. Wu, 2022, "What Moves Stock Prices? The Roles of News, Noise, and Information," *The Review of Financial Studies*, 35(9), 4341–4386.

Bybee, L., B. T. Kelly, A. Manela, and D. Xiu, 2021, "Business News and Business Cycles," Working Paper 29344, National Bureau of Economic Research.

Bybee, L., B. T. Kelly, and Y. Su, 2022, "Narrative Asset Pricing: Interpretable Systematic Risk Factors from News Text," working paper 21-09, Johns Hopkins Carey Business School Research Paper.

Calomiris, C. W., and H. Mamaysky, 2019, "How news and its context drive risk and returns around the world," *Journal of Financial Economics*, 133(2), 299–336.

Cao, J., A. Goyal, S. Ke, and X. Zhan, 2022, "Options trading and stock price informativeness," *Swiss Finance Institute Research Paper*, (19-74).

Carhart, M. M., 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52(1), 57–82.

Chen, Y., B. Kelly, and W. Wu, 2020, "Sophisticated investors and market efficiency: Evidence from a natural experiment," *Journal of Financial Economics*, 138(2), 316–341.

De Bondt, W. F., and R. Thaler, 1985, "Does the stock market overreact?," *The Journal of Finance*, 40(3), 793–805.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, 1990, "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, 98(4), 703–38.

DellaVigna, S., and J. M. Pollet, 2009, "Investor inattention and Friday earnings announcements," *The Journal of Finance*, 64(2), 709–749.

Dim, C., K. Koerner, M. Wolski, and S. Zwart, 2022, "Hot off the press: News-implied sovereign default risk," *EIB Working Papers*.

Dougal, C., J. Engelberg, D. Garcia, and C. A. Parsons, 2012, "Journalists and the stock market," *The Review of Financial Studies*, 25(3), 639–679.

Engle, R. F., S. Giglio, B. Kelly, H. Lee, and J. Stroebel, 2020, "Hedging climate change news," *The Review of Financial Studies*, 33(3), 1184–1216.

Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stock and Bonds," *Journal of Financial Economics*, 33(1), 3–56.

——— , 1997, "Industry costs of equity," *Journal of Financial Economics*, 43(2), 153–193.

——— , 2015, "A five-factor asset pricing model," *Journal of Financial Economics*, 116(1), 1–22.

Fama, E. F., and J. D. MacBeth, 1973, "Risk Return and Equilibrium: empirical Tests," *Journal of Financial Political Economy*, 71, 607–636.

Farboodi, M., A. Matray, L. Veldkamp, and V. Venkateswaran, 2021, "Where Has All the Data Gone?," *The Review of Financial Studies*, 35(7), 3101–3138.

Frank, M. Z., and A. Sanati, 2018, "How does the stock market absorb shocks?," *Journal of Financial Economics*, 129(1), 136–153.

Frazzini, A., 2006, "The disposition effect and underreaction to news," *The Journal of Finance*, 61(4), 2017–2046.

Gabaix, X., and R. Koijen, 2021, "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," NBER Working Paper 28967.

Garcia, D., 2013, "Sentiment during recessions," *The Journal of Finance*, 68(3), 1267–1300.

Gentzkow, M., and J. M. Shapiro, 2006, "Media bias and reputation," *Journal of Political Economy*, 114(2), 280–316.

Goldman, E., N. Gupta, and R. D. Israelsen, 2021, "Political polarization in financial news," *Available at SSRN 3537841*.

Goldman, E., J. Martel, and J. Schneemeier, 2022, "A theory of financial media," *Journal of Financial Economics*, 145(1), 239–258.

Hanley, K. W., and G. Hoberg, 2019, "Dynamic interpretation of emerging risks in the financial sector," *The Review of Financial Studies*, 32(12), 4543–4603.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh, 2016, "The common factor in idiosyncratic volatility: Quantitative asset pricing implications," *Journal of Financial Economics*, 119(2), 249–283.

Hillert, A., H. Jacobs, and S. Müller, 2014, "Media makes momentum," *The Review of Financial Studies*, 27(12), 3467–3501.

Kacperczyk, M., S. Sundaresan, and T. Wang, 2020, "Do Foreign Institutional Investors Improve Price Efficiency?," *The Review of Financial Studies*, 34(3), 1317–1367.

Kacperczyk, M. T., J. Nosal, and S. Sundaresan, 2022, "Market power and price informativeness," *Available at SSRN 3137803*.

Koijen, R. S. J., and M. Yogo, 2019, "A Demand System Approach to Asset Pricing," *Journal of Political Economy*, 127(4), 1475–1515.

Kumar, A., 2009, "Who Gambles in the Stock Market?," *The Journal of Finance*, 64(4), 1889–1933.

Liu, Y., and B. Matthies, 2022, "Long-Run Risk: Is It There?," *The Journal of Finance*.

Loughran, T., and B. McDonald, 2011, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of finance*, 66(1), 35–65.

Manela, A., and A. Moreira, 2017, "News implied volatility and disaster concerns," *Journal of Financial Economics*, 123(1), 137–162.

Mullainathan, S., and A. Shleifer, 2005, "The market for news," *American Economic Review*, 95(4), 1031–1053.

Newey, W. K., and K. D. West, 1987, "A Simple, Positive-semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

Niessner, M., and E. C. So, 2018, "Bad news bearers: The negative tilt of the financial press," *Available at SSRN 3219831*.

Peng, L., and W. Xiong, 2006, "Investor attention, overconfidence and category learning," *Journal of Financial Economics*, 80(3), 563–602.

Reuter, J., and E. Zitzewitz, 2006, "Do ads influence editors? Advertising and bias in the financial media," *The Quarterly Journal of Economics*, 121(1), 197–227.

Rhodes-Kropf, M., D. T. Robinson, and S. Viswanathan, 2005, "Valuation waves and merger activity: The empirical evidence," *Journal of Financial Economics*, 77(3), 561–603.

Sacerdote, B., R. Sehgal, and M. Cook, 2020, "Why is all COVID-19 news bad news?," working paper, National Bureau of Economic Research.

Shiller, R. J., 2017, "Narrative Economics," *American Economic Review*, 107(4), 967–1004.

Shiller, R. J., 2020, *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.

Shleifer, A., and L. H. Summers, 1990, "The noise trader approach to finance," *Journal of Economic Perspectives*, 4(2), 19–33.

Teeter, P., and J. Sandberg, 2017, "Cracking the enigma of asset bubbles with narratives," *Strategic Organization*, 15(1), 91–99.

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

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy, 2008, "More than words: Quantifying language to measure firms' fundamentals," *The Journal of Finance*, 63(3), 1437–1467.

Wiesen, T. F., and P. M. Beaumont, 2020, "A Joint Impulse Response Function for Vector Autoregressive Models," *ERN: Other Econometrics: Applied Econometric Modeling in Forecasting (Topic)*.

Wurgler, J., and E. Zhuravskaya, 2002, "Does Arbitrage Flatten Demand Curves for Stocks?," *Journal of Business*, 75, 583–608.

A Proofs

Proof of Eq. (4) Consider initially the case where the number of articles M is finite. Each article selects a narrative at random according to the probability vector θ_t. We denote $\mathcal{M}_{l_t}$ the index set of all articles about narrative l at time t and we denote $M_{l,t}$ its cardinality. For each narrative $l = 1, \ldots, L$, the set of signals $\{s_{m,t}\}_{m \in \mathcal{M}_{l,t}}$ is equivalent to the sufficient statistic

$$S_{l,t} = \sum_{m \in \mathcal{M}_{l,t}} \frac{s_{m,t}}{M_{l,t}} = z_{l,t} + \pi_{l,t} + \hat{\zeta}_{l,t}; \qquad \text{for } l = 1, ..., L, \tag{A1}$$

where $\hat{\zeta}_{l,t} := \sum_{m \in \mathcal{M}_{l,t}} \frac{\zeta_{m,t}}{M_{l,t}}$. If $M_{l,t} = 0$, then $S_{l,t}$ is pure noise. The precision of $S_{l,t}$ is $Var\left(S_{l,t} \middle| z_{l,t}, \pi_{l,t}\right)^{-1} = \frac{M_{l,t}}{M}\omega$. The Law of Large Numbers implies

$$\lim_{M \uparrow \infty} \frac{M_{l,t}}{M} = \theta_{l,t}. \tag{A2}$$

□

Projection of dividends on narratives in Eq. (6) Eqs. (1)-(3) imply the following projection of f_t onto z_t:

$$f_t = \Sigma_f A' \left(A \Sigma_f A' + \Sigma_\eta\right)^{-1} z_t + \nu_t,$$

where ν_t is uncorrelated with z_t. Therefore, the projection of dividends on narratives in Eq. (6) holds for

$$\beta_n' = b_n' \Sigma_f A' \left(A \Sigma_f A' + \Sigma_\eta\right)^{-1} \tag{A3}$$

and

$$\phi_n = \epsilon_n + \sum_{t=1}^{T} b_n' \nu_t. \tag{A4}$$

□

Proof of Proposition 1-(i) The proof is by induction. First we show that if the vector of asset prices satisfies Eq. (9) at time $t+1$, then it satisfies Eq. (9) at time t. Our model assumptions imply that a newborn investor i at time t maximizes

$$\begin{array}{ll} x_{i,t}' E_{i,t} \left(D - P_t\right) - \frac{1}{2} x_{i,t}' C_i x_{i,t} & \text{for } t = T \\ x_{i,t}' E_{i,t} \left(P_{t+1} - P_t\right) - \frac{1}{2} x_{i,t}' C_i x_{i,t} & \text{for } t < T \end{array}, \tag{A5}$$

where D denotes the $(N \times 1)$ vector of asset payoffs, P_t denotes the $(N \times 1)$ vector of asset prices at time t, and $E_{i,t}$ denotes the time-t conditional expectation of investor i. The solution to (A5) for $t < T$ is

$$x_{i,t} = C_i^{-1} E_{i,t} \left(P_{t+1} - P_t \right),$$

which for asset n reads

$$x_{i,n,t} = c_{i,n}^{-1} E_{i,t} \left[(1 - \gamma_n) E_{R,t+1} \left(D_n \right) + \gamma_n E_{U,t+1} \left(D_n \right) - P_{n,t} \right]. \tag{A6}$$

Market clearing requires

$$\int_R x_{i,n,t} di + \int_U x_{i,n,t} di = 0.$$

Using Eq. (A6) and the definition of $\psi_{R,n}, \psi_{U,n}$ and γ_n in Proposition 1-(i), and solving the market clearing condition for $P_{n,t}$ yields

$$P_{n,t} = (1 - \gamma^n) E_{R,t} \left[(1 - \gamma_n) E_{R,t+1} \left(D_n \right) + \gamma_n E_{U,t+1} \left(D_n \right) \right] + \gamma^n E_{U,t} \left[(1 - \gamma_n) E_{R,t+1} \left(D_n \right) + \gamma_n E_{U,t+1} \left(D_n \right) \right]. \tag{A7}$$

Using Eq. (7) and the fact that $E_R \left(E_U \left(S_{l,t} \right) \right) = 0$ and $E_U \left(E_R \left(S_{l,t} \right) \right) = -\pi_l$ for all $l = 1, ..., L$, we have[17]

$$E_{R,t} \left[(1 - \gamma_n) E_{R,t+1} \left(D_n \right) + \gamma_n E_{U,t+1} \left(D_n \right) \right] = E_{R,t} \left(D_n \right) + \gamma_n \beta_n' \left(E \left(\Phi_{t+1} \right) \pi + \sum_{\tau=1}^{t} \Phi_\tau \pi_\tau \right) \tag{A8}$$

and

$$E_{U,t} \left[(1 - \gamma_n) E_{R,t+1} \left(D_n \right) + \gamma_n E_{U,t+1} \left(D_n \right) \right] = E_{U,t} \left(D_n \right) - (1 - \gamma_n) \beta_n' \left(E \left(\Phi_{t+1} \right) \pi + \sum_{\tau=1}^{t} \Phi_\tau \pi_\tau \right). \tag{A9}$$

Substituting Eqs. (A8)-(A9) into Eq. (A7) and rearranging yields Eq. (9). Next, we show that Eq. (9) holds for $t = T$. The solution to (A5) for $t = T$ is

$$x_{i,T} = C_i^{-1} E_{i,T} \left(D - P_T \right). \tag{A10}$$

[17] Note that by Eq. (9), $\{P_{n,\tau}\}_{\tau=1}^{t}$ reveal $\{\pi_\tau\}_{\tau=1}^{t}$ to U investors who incorrectly interpret media biases as biases in R investors' belief about news media.

Using Eq. (A6) in the market clearing condition, the same steps as above yield

$$P_{n,T} = (1 - \gamma_n) E_{R,T}(D_n) + \gamma_n E_{U,T}(D_n).$$

Therefore, Eq. (9) holds for all t. □

Proof of Proposition 1-(ii) Proposition 1-(i) implies

$$r_{n,t} = E_{R,t}\left(\beta_n' z_{t+1}\right) + \gamma_n \beta_n' \Phi_t \pi_t.$$

Next, we observe that

$$\begin{aligned}
\gamma_n \beta_n' \Phi_t \pi_t &= vec\left(\gamma_n \beta_n' \Phi_t \pi_t\right) \\
&= \gamma_n \left(\pi_t' \otimes \beta_n'\right) vec(\Phi_t) \\
&= \gamma_n \left(\pi_{1,t}\beta_{n,t}^{'}, ..., \pi_{L,t}\beta_n^{'}\right) \left(\phi_{1,t}', ..., \phi_{L,t}'\right)' \\
&= \gamma_n \sum_{l=1}^{L} \pi_{l,t} \beta_n' \phi_{l,t},
\end{aligned}$$

where $\phi_{l,t}$ denotes the l-th column of Φ_t. □

Proof of Proposition 1-(iii) Using Eq. (10) we compute:

$$Cov\left(r_{n,t}, \theta_{l,t}\right) = Cov\left(E_{R,t}\left(\beta_n' z_t\right), \theta_{l,t}\right) + \gamma_n \sum_{j=1}^{L} Cov\left(\pi_{j,t}\beta_n'\phi_{j,t}, \theta_{l,t}\right).$$

We can write

$$Cov\left(E_{R,t}\left(\beta_n' z_t\right), \theta_{l,t}\right) = Cov\left(\beta_n' z_t, \theta_{l,t}\right) + Cov\left(E_{R,t}\left(\beta_n' z_t\right) - \beta_n' z_t, \theta_{l,t}\right).$$

Since $\beta_n' z_t$ and $\theta_{l,t}$ are independent and the expectation error $E_{R,t}(\beta_n' z_t) - \beta_n' z_t$ is orthogonal to time-t information, we conclude that $Cov\left(E_{R,t}\left(\beta_n' z_t\right), \theta_{l,t}\right) = 0$.

Next, we compute

$$\sum_{j=1}^{L} Cov\left(\pi_{j,t}\beta_n'\phi_{j,t}, \theta_{l,t}\right) = \sum_{j=1}^{L} E\left[\pi_{j,t} Cov\left(\beta_n'\phi_{j,t}, \theta_{l,t}\right)\right] + \sum_{j=1}^{L} Cov\left[\pi_{j,t} E\left(\beta_n'\phi_{j,t}\right), E(\theta_{l,t})\right]$$

$$= \sum_{j=1}^{L} \pi_j Cov\left(\beta_n'\phi_{j,t}, \theta_{l,t}\right),$$

where the first equality follows from the law of total covariance.

□

Proof of Proposition 2-(i) First, we prove that $Var\left(r_{n,t}\right) = SysVar_n + IdVar_n$ in Eq. (14). Using Eq. (10) we compute:

$$Var\left(r_{n,t}\right) = Var\left(E_{R,t}\left(\beta_n'z_t\right)\right) + Var\left(\gamma_n \sum_{j=1}^{L} \pi_{j,t}\beta_n'\phi_{j,t}\right) + 2Cov\left(E_{R,t}\left(\beta_n'z_t\right), \gamma_n \sum_{j=1}^{L} \pi_{j,t}\beta_n'\phi_{j,t}\right).$$

We have

$$Cov\left(E_{R,t}\left(\beta_n'z_t\right), \sum_{j=1}^{L} \pi_{j,t}\beta_n'\phi_{j,t}\right) = Cov\left(\beta_n'z_t, \sum_{j=1}^{L} \pi_{j,t}\beta_n'\phi_{j,t}\right) + Cov\left(E_{R,t}\left(\beta_n'z_t\right) - \beta_n'z_t, \sum_{j=1}^{L} \pi_j\beta_n'\phi_{j,t}\right). \tag{A11}$$

Each $\phi_{j,t}$ is a function of θ_t, which is independent of z_t, and so are each $\pi_{j,t}$ and z_t. Thus, the first term in Eq. (A11) is zero. Since the expectation error $E_{R,t}\left(\beta_n'z_t\right) - \beta_n'z_t$ is orthogonal to time-t information, also the second term in Eq. (A11) is zero.

Next, we compute:

$$Var\left(E_{R,t}\left(D_n\right)\right) = Var\left(D_n\right) - E\left[Var_{R,t}\left(D_n\right)\right]$$

$$= t\left[Var\left(\beta_n'z_t\right) - E\left[Var_{R,t}\left(\beta_n'z_t\right)\right]\right]$$

$$= t\left(A\Sigma_f A' + \Sigma_\eta\right) E\left[\left(A\Sigma_f A' + \Sigma_\eta + \Theta_\tau^{-1}\right)^{-1}\right]\left(A\Sigma_f A' + \Sigma_\eta\right),$$

where the first equality follows from the law of total variance, the second equality from the fact that all random variables are independent over time, the third equality from the standard

conditional variance formula for normally distributed random variables:

$$Var_{R,t}\left(\beta_n' z_t\right) = Var\left(\beta_n' z_t\right) - \left(A\Sigma_f A' + \Sigma_\eta\right)\left(A\Sigma_f A' + \Sigma_\eta + \Theta_\tau^{-1}\right)^{-1}\left(A\Sigma_f A' + \Sigma_\eta\right).$$

Therefore,

$$Var\left(E_{R,t}\left(D_n\right)\right) = tSysVar_n, \tag{A12}$$

where

$$SysVar_n = \left(A\Sigma_f A' + \Sigma_\eta\right) E\left[\left(A\Sigma_f A' + \Sigma_\eta + \Theta_\tau^{-1}\right)^{-1}\right]\left(A\Sigma_f A' + \Sigma_\eta\right). \tag{A13}$$

Finally, we compute

$$\begin{aligned} IdVar_n &= Var\left(\sum_{j=1}^{L} \pi_{j,t}\beta_n'\phi_{j,t}\right) \\ &= \sum_{j=1}^{L}\sum_{i=j}^{L} E\left[\pi_{i,t}\pi_{j,t}Cov\left(\beta_n'\phi_{i,t}, \beta_n'\phi_{j,t}\right)\right] + Var\left[\sum_{j=1}^{L}\pi_{j,t}E(\beta_n'\phi_{j,t})\right] \\ &= \sum_{j=1}^{L}\sum_{i=j}^{L} \pi_i\pi_i Cov\left(\beta_n'\phi_{i,t}, \beta_n'\phi_{j,t}\right) + \sum_{j=1}^{L}\pi_j^2\sigma^2 E(\beta_n'\phi_{j,t})^2, \end{aligned}$$

where the second equality follows from the law of total variance and the third from the independence of biases across narratives. Next, we prove the formula for I_n in Eq. (14). Since can write $P_{n,t} = \sum_{\tau=1}^{t} r_{n,\tau}$ and returns are i.i.d. over time, we have

$$Var\left(P_{n,t}\right) = tVar\left(r_{n,t}\right) = t\left(SysVar_n + IdVar_n\right). \tag{A14}$$

Next, using the formula for $P_{n,t}$ in Eq. (12) we compute

$$Cov\left(D_n, P_t\right) = Cov\left(D_n, E_{R,t}\left(D_n\right)\right) + \gamma_n Cov\left(D_n, \sum_{\tau=1}^{t}\beta_n'\Phi_\tau\pi_\tau\right).$$

Since each Φ_τ is a function of θ_τ and D_n and θ_τ are independent, and so are each π_τ and D_n, then $Cov\left(D_n, \sum_{\tau=1}^{t} \beta_n' \Phi_\tau \pi_\tau\right) = 0$. Thus, we are left with

$$Cov\left(D_n, P_{n,t}\right) = Cov\left(D_n, E_{R,t}\left(D_n\right)\right) \tag{A15}$$

$$= Cov\left(D_n - E_{R,t}\left(D_n\right), E_{R,t}\left(D_n\right)\right) + Cov\left(E_{R,t}\left(D_n\right), E_{R,t}\left(D_n\right)\right)$$

$$= Var\left(E_{R,t}\left(D_n\right)\right)$$

$$= tSysVar_n, \tag{A16}$$

where the second equality follows from the fact that the expectation error $D_n - E_{R,t}\left(D_n\right)$ is uncorrelated with $E_{R,t}\left(D_n\right)$ and the third equality follows from Eq. (A12) in the proof of part-(i). Using Eqs. (A14)-(A16) and the definition I_n in Eq. (12) yields the desired result. □

Proof of Proposition 2-(ii) We have:

$$Corr\left(\Pi_{n,t}, \theta_{l,t}\right)^2 = \frac{Cov\left(\Pi_{n,t}, \theta_{l,t}\right)^2}{Var\left(\Pi_{n,t}\right) Var\left(\theta_{l,t}\right)} = \beta\left(n,l\right)^2 \frac{Var\left(\theta_{l,t}\right)}{IdVar_n},$$

where the first equality follows from the definition of correlation and the second from Eq. (11). Rearranging terms gives Eq. (16). Summing over $\beta\left(n,l\right)^2$ gives Eq. (17). □

B Additional Tables

Table B1: Variable Definitions

Variable	Years	Definition
Selected Narratives		
POLY	1998-2021	Politics
REGL	1998-2021	Regulation
MCRO	1998-2021	Macroeconomy
EQTY	1998-2021	Equity markets
FINC	1998-2021	Fixed income markets
ENGY	1998-2021	Energy markets
STPL	1998-2021	Consumer staples
HLTH	1998-2021	Healthcare
AUTO	1998-2021	Automotive
TLCO	1998-2021	Telecommunications and social media
ENTM	1998-2021	Entertainment
SCHL	1998-2021	College and schooling
Narrative Exposure		
$\beta_{n,t}^{narr}$	1998-2021	Narrative beta estimated at the end of each year using daily excess returns, factor realizations and attention shocks to the particular narrative over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary, CRSP, Own computations.
Average $\lvert\beta_{n,t}^{narr}\rvert$	1998-2021	Average absolute narrative beta $\lvert\beta_{n,t}^{narr}\rvert$ to all 33 identified narratives. Source: K. French's DataLibrary, CRSP, Own computations.
Average $\lvert\beta_{n,t}^{narr} \times \sigma_{n,t}^{narr}\rvert$	1998-2021	Average absolute narrative beta scaled by the standard deviation of attention to each of 33 identified narratives in a given year. Source: K. French's DataLibrary, CRSP, Own computations.
Fundamentals and Stock Characteristics		
$Market\ Cap_{n,t}$	1998-2021	A stock's market capitalization. Source: CRSP.
$Assets_{n,t}$	1998-2021	Total assets (Compustat item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$Debt/Assets_{n,t}$	1998-2021	Sum of the book value of long-term debt (Compustat data item DLTT) and the book value of current liabilities (DLC) divided by total assets (Compustat data item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$Cash/Assets_{n,t}$	1998-2021	Cash and short-term investments (Compustat data item CHE) divided by total assets (Compustat data item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$PP\&E/Assets_{n,t}$	1998-2021	Property, plant, and equipment (Compustat data item PPENT) divided by total assets (Compustat data item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$EBIT/Assets_{n,t}$	1998-2021	Earnings before interest and taxes (Compustat data item EBIT) divided by total assets (Compustat data item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$EBIT_{n,t+h}/Assets_{n,t}$	1998-2021	Earnings before interest and taxes (Compustat data item EBIT) h years from the current year divided by total assets (Compustat data item AT). Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$Capex/Assets_{n,t}$	1998-2021	Capital expenditures divided by assets. Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$R\&D/Assets_{n,t}$	1998-2021	R&D expenditures (Compustat data item XRD) divided by total assets (Compustat data item AT). Missing values set to zero. Winsorized annually at 5% and 95%. Source: Compustat NA Annual.
$Turnover_{n,t}$	1998-2021	Turnover relative to shares outstanding. Computed as the daily volume over shares outstanding averaged over all days in a given year. Source: CRSP.

Variable	Years	Definition
Fundamentals and Stock Characteristics		
$Illiquidity_{n,t}$	1998-2021	Amihud (2002) Illiquidity measure computed as the daily absolute return over traded volume ratio averaged over all days in a given year (for stocks with at least 63 observations). Winsorized annually at 5% and 95%. Source: CRSP.
$Lottery_{n,t}$	1998-2021	Kumar (2009) Lottery measure computed for each stock at the end of each year as an indicator variable equal to one for stocks with small price (below median in the cross-section), high idiosyncratic skewness and high four-factor idiosyncratic volatility (both above median in the cross-section). Idiosyncratic skewness and volatility are estimated from daily returns for a given year (for stocks with at least 63 valid observations). Source: K. French's DataLibrary, CRSP.
$Inst.\ Ownership_{n,t}, \%$	1999-2018	Quarterly institutional ownership averaged to the annual level for each year and firm. Source: Thomson Reuters 13F.
Factor Betas		
$Market_{n,t}$	1998-2021	Market beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
$Size\ (SMB)_{n,t}$	1998-2021	Size factor beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
$Value\ (HML)_{n,t}$	1998-2021	Value factor beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
$Momentum\ (WML)_{n,t}$	1998-2021	Momentum factor beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
$Profitability\ (RMW)_{n,t}$	1998-2021	Profitability factor beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
$Investment\ (CMA)_{n,t}$	1998-2021	Investment factor beta estimated for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 valid return observations. Source: K. French's DataLibrary.
Variance Decomposition Variables		
$IdVar_{n,t}$	1998-2021	Idiosyncratic variance for several factor models (market, three-, four-, and five-factor models) for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 return observations. Computed as the mean-squared error of the fitted model residual. Source: K. French's DataLibrary.
$SysVar_{n,t}$	1998-2021	Systematic variance for several factor models (market, three-, four-, and five-factor models) for each year at the end of December using daily excess returns and factor realizations over the past 252 trading days for stocks with at least 63 return observations. Computed as the total variance of daily returns minus the respective idiosyncratic variance. Source: K. French's DataLibrary.
$MktInfo_{n,t}$	1998-2021	Stock variance due to market-wide information. Estimated for each year at the end of December using daily market returns, daily stock signed dollar volume and daily stock returns for the given year. Details are provided in Online Appendix OA.2. Source: CRSP.
$PrivateInfo_{n,t}$	1998-2021	Stock variance due firm-specific private information. Estimated for each year at the end of December using daily market returns, daily stock signed dollar volume and daily stock returns for the given year. Details are provided in Online Appendix OA.2. Source: CRSP.
$PublicInfo_{n,t}$	1998-2021	Stock variance due to public information. Estimated for each year at the end of December using daily market returns, daily stock signed dollar volume and daily stock returns for the given year. Details are provided in Online Appendix OA.2. Source: CRSP.
$Noise_{n,t}$	1998-2021	Stock variance due to noise. Estimated for each year at the end of December using daily market returns, stock signed dollar volume and stock returns for the given year. Details are provided in Online Appendix OA.2. Source: CRSP.

Online Appendix

to

"Media Narratives and Price Informativeness"

OA Data Processing and Construction of Variables

OA.1 News Text Processing

We provide a brief summary of the algorithm and refer interested readers to the original paper (Blei, Ng, and Jordan, 2003) for a detailed description. LDA gives text a hierarchical structure, where documents (news articles) are composed of topical narratives containing words. Precisely, each document has a probability distribution over latent narratives, with parameter $\alpha > 0$, and each narrative is defined by a probability distribution over words with parameter $\beta > 0$. α controls the sparsity of narratives in a document, while β controls the sparsity of words in a narrative. LDA treats a document as a mixture of narratives and a narrative as a mixture of words, such that documents overlap each other rather than being separated into discrete groups.

Training the LDA algorithm boils down to finding the optimal number of latent narratives L that best fit the data. Fitting the LDA algorithm on a corpus of documents with a chosen L yields two outputs: the distribution of word frequencies for each narrative, and the distribution of narratives across documents. For each document, the narrative distribution is a vector of loadings that reflect how much attention is devoted to each narrative in the document, such that higher loading for a particular narrative indicates that the document is more likely associated with that narrative.

We train the LDA algorithm using standard cross-validation and grid search procedures. We first convert the processed text corpus into a document term matrix whose rows are the news articles and columns the unique single words (unigrams) and two-word combinations (bigrams) in the text corpus, excluding terms that occur in less than 0.5% of the text corpus to reduce noise. These unigrams and bigrams constitute the feature space for grouping articles into topical narratives. Next, We use each article's WSJ section name and year of appearance in the WSJ archive as a group variable to split the text corpus into five equal train-test folds for cross-validation. This allows us maintain similar proportion of articles in each section each year throughout training and validation samples. Finally, we search for the number of narratives, L, that minimizes (maximizes) the average test set perplexity (log-likelihood) score.

Figure OB1 summarizes the WSJ news text corpus, our machine learning model training, and the evolution of attention to narratives over time. Panel A shows, on the left axis, the monthly number of news articles in our WSJ historical web archive, and, on the right axis, the number of words in these articles. We observe substantial variations in both the volume of publications and the length of publications over time. Panel B depicts the convergence of the average test set log-likelihood (in millions) to its maximum across the number of narratives,

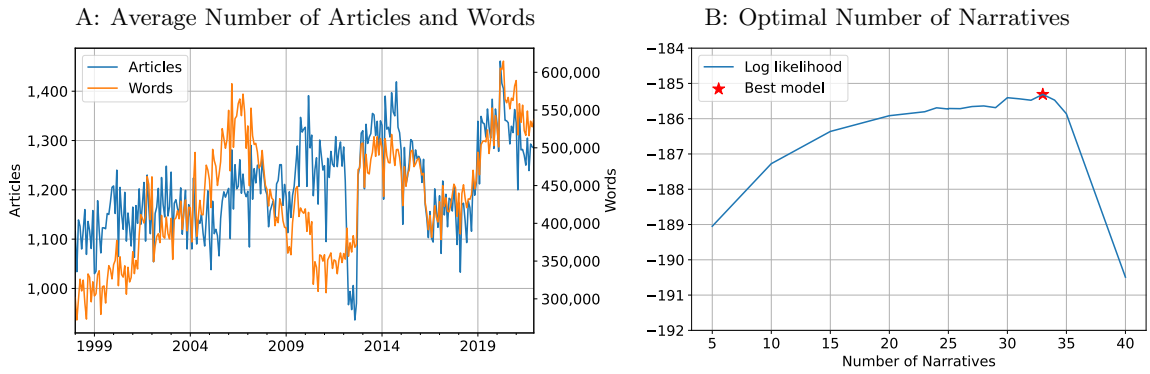

Figure OB1: Article Counts and Model Training. In Panel A, the figures show the total number of articles in our WSJ news corpus per month (left y-axis) and the total number of words in those articles per month (right y-axis) after our preprocessing procedure. Panel B depicts the number of topical narratives in the LDA model that best characterize our news corpus.

L, during the LDA model training. The figure indicates that 33 topical narratives optimally characterize our WSJ text corpus.

OA.2 Information Channels via Variance Decomposition

We obtain the information channels affecting stock returns by two different methods of variance decomposition. We perform estimation separately for each firm and each year using daily returns and factor realizations within the year. First, we estimate several linear factor models of the form

$$r_{n,\tau} = \alpha_{n,t} + \beta_{n,t}^{\top} F_{\tau} + \varepsilon_{n,\tau}, \tag{OA1}$$

where $r_{n,\tau}$ is stock n's excess return on day τ in year t, F is the vector of factor realizations on day τ. We use the market model, the three-factor Fama and French (1993) model, the four-factor Carhart (1997) model, and the five-factor Fama and French (2015) model. After estimating each model for firm n in year t we compute the idiosyncratic variance $IdVar_{n,t}$ as the mean-squared error of the residuals, and the systematic variance $SysVar_{n,t}$ as the total variance minus idiosyncratic variance.

Second, we decompose the total stock return variance following the procedure outlined in "Appendix A: Estimation of the structural VAR" in Brogaard, Nguyen, Putnins, and Wu (2022). For the full procedure, we refer our readers to the original paper. Below we outline the major steps of the procedure (freely copying some parts of the original paper) and specific decisions

we made in our analysis. The stock return is decomposed into the following parts:

$$r_\tau = \underbrace{\mu}_{\text{discount rate}} + \underbrace{\theta_{r_m}\varepsilon_{r_m,\tau}}_{\text{market-wide info}} + \underbrace{\theta_x\varepsilon_{x,\tau}}_{\text{private info}} + \underbrace{\theta_r\varepsilon_{r,\tau}}_{\text{public info}} + \underbrace{\Delta s_\tau}_{\text{noise}}, \tag{OA2}$$

where $\varepsilon_{r_m,\tau}$ is the unexpected innovation in the market return and $\theta_{r_m}\varepsilon_{r_m,\tau}$ is the market-wide information incorporated into stock prices, $\varepsilon_{x,\tau}$ is an unexpected innovation in signed dollar volume and $\theta_x\varepsilon_{x,\tau}$ is the firm-specific information revealed through trading on private information, and $\varepsilon_{r,\tau}$ is the innovation in the stock price producing the $\theta_r\varepsilon_{r,\tau}$ that is the remaining part of firm-specific information not captured by trading on private information. The components above are obtained from a structural vector autoregression (VAR) model with five lags estimated for market returns $r_{m,\tau}$, signed dollar volume of trading in the given stock x_τ, and stock returns r_τ:

$$\begin{aligned}
r_{m,\tau} &= \sum_{l=1}^{5} a_{1,l} r_{m,\tau-l} + \sum_{l=1}^{5} a_{2,l} x_{\tau-l} + \sum_{l=1}^{5} a_{3,l} r_{\tau-l} + \varepsilon_{r_m,\tau} \\
x_\tau &= \sum_{l=0}^{5} b_{1,l} r_{m,\tau-l} + \sum_{l=1}^{5} b_{2,l} x_{\tau-l} + \sum_{l=1}^{5} b_{3,l} r_{\tau-l} + \varepsilon_{x,\tau} \\
r_\tau &= \sum_{l=0}^{5} c_{1,l} r_{m,\tau-l} + \sum_{l=1}^{5} c_{2,l} x_{\tau-l} + \sum_{l=1}^{5} c_{3,l} r_{\tau-l} + \varepsilon_{r,\tau}
\end{aligned} \tag{OA3}$$

The required parameters are obtained by first estimating a reduced-form VAR

$$\begin{aligned}
r_{m,\tau} &= a_0^* + \sum_{l=1}^{5} a_{1,l}^* r_{m,\tau-l} + \sum_{l=1}^{5} a_{2,l}^* x_{\tau-l} + \sum_{l=1}^{5} a_{3,l}^* r_{\tau-l} + e_{r_m,\tau} \\
x_\tau &= b_0^* + \sum_{l=1}^{5} b_{1,l}^* r_{m,\tau-l} + \sum_{l=1}^{5} b_{2,l}^* x_{\tau-l} + \sum_{l=1}^{5} b_{3,l}^* r_{\tau-l} + e_{x,\tau} \\
r_\tau &= c_0^* + \sum_{l=1}^{5} c_{1,l}^* r_{m,\tau-l} + \sum_{l=1}^{5} c_{2,l}^* x_{\tau-l} + \sum_{l=1}^{5} c_{3,l}^* r_{\tau-l} + e_{r,\tau}
\end{aligned} \tag{OA4}$$

and then using the reduced form error covariances to recover the structural VAR parameters, including variances of the residuals $\sigma^2_{r_m}$, σ^2_x, and σ^2_r.

Parameters $\theta_{r_m},\ \theta_x,\ \theta_r$ are defined as the long-run cumulative return response functions in the structural model and are computed by feeding through the reduced model the equivalent reduced form shocks. We use for this purpose the joint impulse response function derived in Wiesen and Beaumont (2020).

The variance components are then computed as follows:

$$MktInfo = \theta_{r_m}\sigma^2_{r_m},\ PrivateInfo = \theta_x\sigma^2_x,\ PublicInfo = \theta_r\sigma^2_r, \tag{OA5}$$

$$Noise = Total\ Variance - MktInfo - PrivateInfo - PublicInfo.$$

OB Additional Tables

	$Average\ \lvert\beta^{narr}_{n,t}\rvert$	$IdVar_{n,t}$	$SysVar_{n,t}$	$MktInfo_{n,t}$	$PrivateInfo_{n,t}$	$PublicInfo_{n,t}$	$Noise_{n,t}$
Panel A: Narrative exposure.							
$Average\ \lvert\beta^{narr}_{n,t}\rvert$	1.000	0.740	0.234	0.381	0.673	0.695	0.620
Panel B: Variance decomposition.							
$IdVar_{n,t}$	0.740	1.000	0.136	0.421	0.768	0.904	0.882
$SysVar_{n,t}$	0.234	0.136	1.000	0.642	0.168	0.182	0.062
$MktInfo_{n,t}$	0.381	0.421	0.642	1.000	0.409	0.499	0.254
$PrivateInfo_{n,t}$	0.673	0.768	0.168	0.409	1.000	0.726	0.523
$PublicInfo_{n,t}$	0.695	0.904	0.182	0.499	0.726	1.000	0.687
$Noise_{n,t}$	0.620	0.882	0.062	0.254	0.523	0.687	1.000
Panel D: Factor model betas.							
$Market\ Beta_{n,t}$	-0.071	-0.081	0.442	0.276	-0.002	-0.047	-0.119
$Size\ (SMB)\ Beta_{n,t}$	0.187	0.154	0.351	0.269	0.196	0.176	0.070
$Value\ (HML)\ Beta_{n,t}$	-0.037	0.055	-0.027	0.027	0.019	0.070	0.076
$Mom.\ (WML)\ Beta_{n,t}$	-0.129	-0.180	-0.137	-0.154	-0.152	-0.190	-0.153
Panel D: Fundamentals and market characteristics.							
$EBIT_{n,t}/Assets_{n,t}$	-0.539	-0.484	-0.059	-0.230	-0.456	-0.485	-0.373
$\ln(Assets)_{n,t}$	-0.453	-0.499	0.124	-0.116	-0.399	-0.459	-0.443
$Debt_{n,t}/Assets_{n,t}$	0.004	-0.000	0.015	-0.021	-0.012	0.006	0.026
$Cash_{n,t}/Assets_{n,t}$	0.228	0.129	0.099	0.131	0.172	0.128	0.039
$PP\&E_{n,t}/Assets_{n,t}$	-0.117	-0.075	-0.012	-0.069	-0.088	-0.079	-0.037
$Sales_{n,t}/Assets_{n,t}$	-0.104	-0.025	-0.106	-0.080	-0.060	-0.037	0.028
$Capex_{n,t}/Assets_{n,t}$	-0.069	0.031	0.039	0.031	0.027	0.039	0.026
$R\&D_{n,t}/Assets_{n,t}$	0.319	0.232	0.071	0.150	0.267	0.232	0.133
$Turnover_{n,t}$	0.336	0.116	0.304	0.236	0.251	0.111	0.004
$Illiquidity_{n,t}$	0.171	0.381	-0.185	-0.046	0.153	0.295	0.487
$Lottery_{n,t}$	0.506	0.509	0.001	0.182	0.431	0.486	0.430
Panel E: Misvaluation and arbitrage risk.							
$MISVAL_{n,t}$	-0.120	-0.135	0.106	0.025	-0.064	-0.126	-0.164
$ARBRISK_{n,t}$	0.731	0.978	0.210	0.446	0.756	0.903	0.842

Table OB1: Correlations for Selected Variables.
The table shows the unconditional correlations among selected variables computed from the firm-year panel data. $Average\ \lvert\beta^{narr}_{n,t}\rvert$ is computed as the average absolute exposure for all 33 identified narratives. Each year, all continuous variables are winsorized at 5% and 95% levels.

	Mean	Std	10%	25%	50%	75%	90%	Obs.
Average $\lvert\beta_{n,t}^{narr}\rvert$	0.267	0.188	0.095	0.135	0.212	0.340	0.522	81,952
POLY	0.141	0.157	0.014	0.036	0.087	0.186	0.330	81,952
REGL	0.075	0.079	0.008	0.020	0.048	0.099	0.176	81,952
MCRO	0.055	0.055	0.006	0.015	0.036	0.074	0.128	81,952
EQTY	0.045	0.047	0.005	0.013	0.029	0.061	0.106	81,952
FINC	0.160	0.171	0.016	0.042	0.101	0.214	0.400	81,952
ENGY	0.187	0.200	0.020	0.050	0.118	0.249	0.440	81,952
STPL	0.142	0.159	0.014	0.036	0.088	0.188	0.343	81,952
HLTH	0.178	0.216	0.017	0.043	0.103	0.227	0.428	81,952
AUTO	0.587	0.621	0.064	0.164	0.382	0.788	1.400	81,952
TLCO	0.084	0.084	0.009	0.024	0.056	0.115	0.204	81,952
ENTM	0.207	0.229	0.021	0.052	0.127	0.275	0.510	81,952
SCHL	0.098	0.114	0.010	0.024	0.059	0.129	0.232	81,952

Table OB2: Summary Statistics for Narrative Exposure Topics.
The table shows the summary statistics for media narrative exposure for selected narratives computed from the firm-year panel data. *Average* $\lvert\beta_{n,t}^{narr}\rvert$ is computed as the average absolute exposure for all 33 identified narratives. Each year, all variables are winsorized at 5% and 95% levels.

Meta Theme	Politics	Regulation		Macroeconomy				Equity Markets					Fixed Income		Energy Markets	Consumer Staples	Healthcare
Narrative	1	2	3	4	5	6	7	8	9	10	11	12	12	14	15	16	17
1	senate	request comment	court	unemployment	state own	central bank	fed	street journal	join conversation	common stock	franc	third quarter	treasurys	loan	oil	restaurant	pandemic
2	voter	plan would	justice	gdp	renewable	tariff	monetary policy	dow	nasdaq composite	value million	samsung	cent share	top list	lender	barrel	museum	patient
3	sen	trade deal	attorney	labor market	follower	ecb	euro zone	jones	composite index	com complete	thomson	net income	year treasury	creditor	crude	ice	coronavirus
4	senator	commission say	white house	unemployment rate	world bank	imf	fannie	dow jones	index rise	million cash	reuters	million cent	nyse	borrower	natural gas	tree	dr
5	obama	state legislature	prosecutor	labor department	life insurance	finance minister	freddie	newswires	full list	regulatory filing	pence	quarter end	year note	tokyo	gasoline	christie	vaccine
6	ballot	deny wrongdoing	criminal	consumer price	corporate governance	international monetary	raise rate	jones newswires	blue chip	amortization	ftse	quarter net	biggest	subprime	crude oil	chicken	health care
7	nominee	federal state	supreme	jobless	bnp paribas	monetary fund	economist say	stoxx	index fall	capital partners	nobel	revenue rise	bond price	foreclosure	million barrel	kitchen	covid
8	bipartisan	rule would	supreme court	consumer spending	privatization	ministry say	treasury note	stoxx index	technology stock	deal value	samsung electronics	quarter profit	percentage price	write andrew	opec	mall	cancer
9	caucus	rule require	justice department	home sale	public private	large economy	investment officer	dow industrial	index gain	thomson financial	financial times	quarter result	list biggest	billion yen	metric	dining	medicare
10	pelosi	clause	allege	say economist	economic development	trading partner	rate rise	online journal	markets data	buy stake	cac	share compare	biggest percentage	mitsubishi	exxon	chef	fda
11	majority leader	would require	allegation	statistics	subsidy	central banker	mae	dutch	go markets	earning interest	franc franc	quarter revenue	year bond	sovereign debt	barrel day	painting	public health
12	presidential candidate	regulate	attorney general	global growth	taxpayer	policy meeting	sarah	jones stoxx	center wsjmarkets	early company	write sarah	income million	yield year	nomura	oil company	beautiful	coronavirus pandemic
13	cia	school district	law enforcement	annual rate	city state	finance ministry	home price	royal dutch	wsjmarkets com	depreciation amortization	swiss	early period	corporate bond	toyota motor	gallon	milk	irs
14	homeland security	proposal would	guilty	trade deficit	welfare	late dollar	term interest	dollar term	wsjmarkets	gs	abn	earning cent	benchmark year	motor corp	mercantile	gallery	medicaid
15	polling	appeals	cuomo	consumer confidence	state government	dollar yen	fannie mae	world stock	index close	interest taxis	market committee	analyst expectation	investment grade	yen yen	mercantile exchange	bathroom	lockdown
16	senate majority	law require	jury	seasonally adjust	poverty	late euro	freddie mac	jones global	track stock	million stock	scotland	operating profit	among common	subprime mortgage	brent	cream	clinical
17	president elect	deadline	allegedly	headwind	budget	yuan	fed chairman	jones world	decliners	share outstanding	cac index	report net	write michael	current fiscal	refinery	cheese	physician
18	obama administration	court ruling	alleged	department report	per year	trade organization	bear market	indexes	price large	expect close	federal open	thomson reuters	treasury yield	repayment	drilling	borough	illness
19	presidential campaign	bargaining	impeachment	job growth	sand	foreign exchange	bull market	newswires editor	gainers	stock new	amro	quarter loss	treasury bond	commercial bank	fossil	pizza	deduction
20	midterm	trade agreement	testify	previous month	private sector	monetary	fed official	sector dow	close high	cash stock	abn amro	revenue fall	year yield	collateral	grain	breakfast	prescription
21	house senate	union say	plead	industrial production	taxis	foreign currency	million euro	global indexes	point nasdaq	value company	swiss franc	quarter company	investors service	lending	york mercantile	chocolate	surgery
22	aide say	say federal	administration official	jobless claim	charter	communist party	cut interest	wall street	major index	ipo	ftse index	first call	moody investors	refinance	mobil	fiction	nursing
23	election day	proposal	convict	straight month	insurance company	ese	say economy	est	stock average	stake company	amsterdam	say net	yield rise	toyota	metric ton	renovation	income tax
24	minority leader	federal agency	defendant	manufacturing sector	entitlement	economic crisis	rate increase	wall	index lose	company board	penny	report earning	write david	fitch	countries	sandwich	tax credit
25	turnout	new law	plead guilty	survey thomson	tax	foreign investment	low interest	journal	index add	public offering	ing	current quarter	bond issue	financial institution	oil future	coat	say dr
26	centrist	regulation	qaeda	seasonally	exceptional	exchange rate	interest rate	laggard	tech stock	deal expect	france	analyst poll	yield fall	repay	katrina	office building	dr say
27	bernie sanders	authorization	plaintiff	month ahead	budget	currency	raise interest	jones	street	initial public	deutsche bank	analyst survey	yield move	bank also	heating	fish	therapy
28	sens	would force	indictment	create job	debt crisis	competitiveness	rate fall	exxon mobil	index end	stock option	korea	quarter compare	inversely	deposit	oil production	bike	food drug
29	candidacy	times report	federal court	datum release	federal government	global financial	rate would	asian	close see	sell stake	deutsche	share earning	move inversely	bailout	nymex	garden	drug administration
30	nonpartisan	comply	subpoena	commerce department	offshore	global economy	cut rate	edt	stock end	data center	sterling	quarter say	bond market	balance sheet	oil field	flower	health official
31	national committee	compliance	clinton	wage	rebuild	rest world	economist expect	poor stock	stock gain	go public	british	quarter sale	euro yen	austerity	oil market	clothe	disease control
32	envoy	chamber commerce	prosecution	unemployed	infrastructure	economic policy	slow pace	three quarter	index climb	company billion	paris	fiscal first	inversely price	lend	gasoline price	square	control prevention
33	write john	would also	wrongdoing	durable	waste	geopolitical	chief investment	tim	benchmark index	regulatory approval	insurer	revenue increase	market action	honda	exporting	glass	centers disease
34	house representatives	would allow	judicial	employment	percent	foreign investor	market go	rocky	point close	blackstone	seoul	fiscal fourth	bond yield	say bank	organization petroleum	portrait	epidemic
35	electoral	eu	federal judge	report show	project	strong dollar	rate cut	index track	chip stock	public company	london	early revenue	stock nasdaq	mortgage	exporting countries	shirt	medication
36	second term	federal regulator	lawyer say	slow growth	program say	global economic	low rate	flock	share change	agree sell	open market	quarter up	bond	cent dollar	petroleum exporting	cow	care act
37	nomination	unanimously	federal prosecutor	fast pace	casino	imbalance	rate low	jones industrial	index slip	billion cash	nv	fiscal third	high yield	big bank	oil producer	tall	affordable care
38	poll show	local government	misconduct	median	large part	economic recovery	bull	quarterly earning	technology share	trade company	wind down	quarter million	back security	pay down	matthew	fruit	tax return
39	presidential election	free trade	overturn	payroll	corrupt	easing	job market	times	profit taking	ipos	pretax profit	fiscal second	mortgage back	investment trust	crude price	rent	clinical trial
40	liberal	state federal	verdict	growth slow	scarce	central	rate say	milestone	bovespa	closing price	expect report	yield	sheet	barrel oil	menu	medical center	
41	election	would receive	judge say	contraction	corruption	election campaign	many investor	industrial average	stock advance	new company	pound	consortium	say earning	borrow	sweet crude	vegetable	prescription drug
42	presidential	uniform	uphold	growth rate	government plan	bilateral	quantitative	europe	nasdaq index	company raise	plc	say earning	government bond	estate investment	light sweet	thick	health organization
43	committee chairman	audit	dissent	purchasing	sustainable	dollar trade	year date	also provide	issue lead	agree buy	suisse	previous quarter	municipal	banking system	oil natural	festival	side effect
44	rep	eligible	inspector general	labor	funding	government spending	risky	rattle	volume billion	shareholder	credit suisse	post net	basis point	borrowing	price oil	art	please include
45	campaign say	would seek	federal bureau	hourly	wealthy	treasury secretary	curve	reassure	general index	filing	ab	poll thomson	junk	liquidity	commodity market	meal	disease
46	party leader	levy	bureau investigation	minimum wage	incentive	policy maker	investment strategist	news	ge	spinoff	electronics	quarter last	issuance	bankruptcy	cent barrel	grill	symptom
47	party say	exemption	appeal court	many economist	write joseph	say speech	selloff	overweight	small cap	combined	foreign company	earning estimate	face value	distressed	comex division	bread	medicine
48	tax cut	adhere	statute	labor cost	public policy	grim	market could	think people	index drop	financial statement	suisse first	earning forecast	bondholder	default	barrel new	deck	genetic
49	congressman	would need	prosecute	chief economist	move ahead	intervention	short term	euro	close low	company own	press report	report third	fix income	banking sector	say yesterday	cook	enrollment
50	ic	submit	district attorney	economist	initiative	economy say	relatively low	consumer electronic	gain point	holdings inc	hsbc	report second	maturity	financial system	greenhouse	smell	clinic

Panel B: Set 2

Meta Theme	Automotive	Telecoms & Social Media			Entertainment	College & Schooling	Others									
Narrative	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	euros	mobile	software	google	film	student	dowjone	incorrectly	de	violence	goldman	flight	vaccination	coal	gm	gold
2	fiat	wireless	microsoft	facebook	disney	kid	dowjone com	version article	blasio	immigration	private equity	airline	egg	rise cent	sanders	steel
3	car maker	smartphone	app	amazon	studio	teacher	mutual	article incorrectly	de blasio	troop	sachs	plane	walker	tobacco	mart	metal
4	mainland	sprint	patent	social medium	song	wine	etf	com corrections	pesos	terrorist	goldman sachs	airport	sound like	fall cent	wal	mining
5	nissan	broadband	iphone	tweet	youtube	high school	mutual fund	prosecutor say	class share	peace	climate change	boeing	co author	rand	wal mart	carbon
6	volkswagen	mobile phone	pc	twitter	musical	feel like	fund manager	chancellor	cola	taliban	equity firm	passenger	naturally	gain cent	chrysler	rail
7	container	high speed	tech company	web site	viewer	baseball	ira	incorrectly say	coca	democracy	fargo	jet	prioritize	morris	ibm	aluminum
8	euro euros	service provider	hacker	amazon com	music	fun	fidelity	board member	mexico	soldier	wells fargo	airlines	president george	complete coverage	motor co	miner
9	suv	communications inc	microsoft corp	blog	videogame	public school	pension fund	coca cola	racist	state department	aig	cellphone	stroke	down cent	finance chief	european union
10	hk	internet service	windows	two people	olympics	athlete	vanguard	conflict interest	journal editor	protester	enron	aviation	systemic	aspl	sedan	aerospace
11	bmw	handset	apple inc	alphabet	walt	flavor	bond fund	current former	newsedito wsj	gen	factset	sugar	global warming	lose cent	motors co	railway
12	seng	cellular	computing	facebook inc	walt disney	classroom	trade fund	member say	newsedito	militant	sachs group	flu	etc	add cent	mart stores	imperial
13	hong	long distance	operating system	jennifer	dvd	laugh	fund say	front runner	nations	investment banking	airbus	virus	surely	drop cent	stores inc	gold price
14	dax	verizon	software company	search engine	disney co	stadium	stock fund	sec corrections	editor newseditor	terrorism	morgan chase	cargo	virtue	cent new	motors corp	royal bank
15	kong	phone company	packard	google inc	broadcasting	exhibition	market fund	life say	ipc	rebel	wealth management	airways	narrative	shed cent	annual sale	toronto
16	billion euro	nortel	personal computer	write jennifer	inc top	list go	amy	corrections amplifications	grupo	united nations	citigroup inc	mayor say	parade	climb cent	store open	mining company
17	daimlerchrysler	federal communications	hewlett	dot com	sony corp	school student	equity fund	amplifications item	ipc index	diplomatic	accord research	earthquake	inevitably	decline issue	supply chain	group rise
18	sport utility	communications commission	hewlett packard	wow	movie	funny	fund invest	former chairman	miller	diplomat	merger acquisition	kill people	precisely	rip cent	pay cent	copper
19	utility vehicle	lucent	machines	ventures	broadcast	school say	fund company	amplifications	lion	refugee	jpm	northwest	supposedly	cigarette	auto industry	ounce
20	hong kong	optic	business machines	york times	comedy	tell story	say fund	say person	banco	human right	street firm	delta	wisdom	cent cent	square foot	iron
21	frankfurt	research motion	desktop	yesterday	sports	girl	fund investor	conspiracy	del	foreign policy	security firm	federal fund	go wrong	philip	new car	mill
22	hsbc holdings	fiber optic	intel corp	click	streaming	money see	fund rate	person say	city	al qaeda	accord thomson	see fall	excerpt	cent	department store	mine
23	dax index	nortel networks	computer system	database	documentary	old man	investment fund	landlord	peso	commander	lynch co	boeing co	one could	five cent	auto maker	gas price
24	fall euro	aol	machines corp	com inc	game	old son	taxable	one people	cement	treaty	salomon smith	write paul	good idea	two cent	motors	rio
25	auto	motorola	computer maker	washington post	musician	college student	individual investor	include former	class	terror	familiar matter	lines	point view	beer	ford	ore
26	motor	internet access	jose calif	email	football	humor	pension	drone	beverage	nato	wells	air lines	op	diamond	company spokesman	canada
27	tire	communications	palo	com say	cbs	husband	retirement	step down	bellwether	killing	investment banker	infection	obvious	three cent	home depot	india
28	automotive	time warner	palo alto	sad	league	mom	tax free	heather	metro	embassy	hedge fund	vessel	somehow	four cent	rental	environmental
29	car	subscriber	packard co	shipping	news corp	mother	investing	public relation	worth million	homeland	people familiar	cruise	columnist	africa	depot	steelmaker
30	hang seng	nokia	apple computer	web	theater	coach	outflow	cfo	bottle	bombing	chase co	outbreak	one thing	group ltd	brand name	commodity price
31	vehicle	comcast	microsystems	card	concert	young people	money market	former vice	real	security council	accord people	pilot	memoir	skid	apparel	brussels
32	seng index	grid	variant	fundraising	sony	nasty	fund	former head	hunting	quarantine	underwriting	airplane	solve problem	sag	brands	locally
33	rise euro	worldcom	apple	advertiser	singer	teenager	put money	senior executive	financial group	kremlin	matter say	missile	grasp	advancer	state control	comex
34	truck	frontier	san jose	site	basketball	graduate	planner	ethic	share worth	extremist	hedge	port	perspective	consolidated	stores	forest
35	new model	cable	alto	chat	bowl	daughter	money manager	employee say	fix line	communist	person familiar	southwest	good thing	shed	bedroom	pan
36	machinery	telecommunication company	oracle	ad	fan	bed	allocation	also call	se	nuclear weapon	citigroup	aboard	presumably	foods	lease	australia
37	rubber	phone	computer	fake	actor	inspiration	charity	respond request	surrender	condemn	financial firm	aircraft	sing	index shed	detroit	freight
38	spa	satellite	fbi	glitch	abc	boy	financial adviser	trump	ordinary	civil war	international group	temperature	universe	world big	battery	sub
39	engine	customer service	hardware	online	actress	teaching	funds	say former	share market	militia	steams	evacuate	teenage	leap	chain	material
40	luxury	warner	software maker	mail	camera	marriage	inflow	rape	broadcaster	terrorist attack	accord person	crew	observe	jewelry	industry analyst	resources
41	diesel	warner inc	ebay	publication	fox	smile	buy stock	tenure	turnover	detention	merrill	accident	incumbent	outpace	brand	packaging
42	hang	spectrum	laptop	advertise	roster	grow up	investments	say try	enterprises	riot	bear stearns	coastal	wake up	diversified	general motors	european
43	highway	frequency	intellectual property	classified	star	drink	investment company	chairwoman	soft	immediately respond	familiar	traveler	moral	american	grocery store	pacific
44	ag	provider	server	advertisement	franchise	gay	institutional investor	karen	total million	humanitarian	leveraged	death toll	arguably	information technology	gbt	rose
45	convertible	dial	printer	people use	audience	dream	portfolio	runner	worth billion	security force	stanley	fleet	correctly	edge up	hybrid	commodity
46	emission	telephone	high tech	anonymous	recording	father	heir	injury	low income	citizenship	lynch	helicopter	ensure	decliner	time company	industries
47	holdings ltd	footprint	intel	yahoo	dance	love	raise money	succession	local	two country	morgan stanley	taxi	theory	board	bidder	group also
48	maker	turner	clara	los angeles	tooth	birthday	retiree	chief operating	go effect	insurgent	large bank	province	philosophy	shares	traction	rise price
49	suspension	bundle	gadget	angeles	console	school	payout	operating officer	also know	diplomacy	firm also	fly	honest	sink	outsourcing	month high
50	drive	network	jose	kelly	visual	college	billion asset	staffer	fix	nuclear program	lehman	crash	brilliant	precious	cut cost	northern

Table OB3: Most Relevant Terms for the Identified Narratives.
This table shows the top 50 terms with the largest rescaled weights for each narrative and groups (i.e., broader themes) manually assigned to narratives based on the similarity of the top terms.